A METHOD TO ISOLATE THE CTD OF RNA POLYMERASE II

FOR PROTEOMICS ANALYSIS

Nada S. Alakhras

Submitted to the faculty of the University Graduate School
In partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Biochemistry and Molecular Biology
Indiana University

December 2014

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Master's Thesis Committee

Amber L. Mosley, Ph.D., Chair

Mark G. Goebl, Ph.D.

Thomas D. Hurley, Ph.D.

DEDICATION

To my mother, Aysheh.

Thank you for your love and support.

AKNOWLEDGEMENTS

This thesis was completed through guidance from my committee members, and support from my friends and family. I would like to express my sincere gratitude to the following people:

- Dr. Amber Mosley, for accepting me in her lab and for her mentorship, patience guidance and friendship. I learned a lot while working in her lab; I would never have learned Molecular Biology techniques and concepts without her help.
- Dr. Mark Goebl for helping me to get into the Biochemistry program and for his guidance and support.
- Dr. Thomas Hurley for serving on my committee and for his guidance and advice on my research.
- Dr. Sherry Queener for helping me to transfer into the Biochemistry program.
- Members of the Mosley laboratory Melanie Fox, Jerry Hunter, and Whitney Smith-Kinnaman for their help, equipment training, and friendship.
- My friends and co-workers for their support and encouragement.
- My family for encouraging me throughout this experience and for believing in me I would never have completed my thesis without your support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

2D-PAGE	Two-dimensional poly acrylamide gel electrophoresis
3C	Human Rhinovirus 3C
5-FOA	5-fluoroorotic acid
βME	β-mercaptoethanol
CBP	Calmodulin binding protein
ChIP	Chromatin Immunoprecipitation
CLONAT	Nourseothricin
CTD	C-terminal domain of RNA polymerase subunit Rpb1
EDTA	Ethylenediamine tetraacetic acid
EGTA	Ethlyene glycol tetraacetic acid
IgG	Immunoglobulin G
MudPIT	Multi-dimensional protein identification technology
MS	Mass spectrometry
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PIC	Pre-initiation complex
PSM	Peptide spectrum matches
RNAPII	RNA polymerase II complex
RP	Reverse phase chromatography
S. cerevisiae	*Saccharomyces cerevisiae*
SCX	Strong cation exchange

SDS-PAGE	Sodium dodecyl sulfate poly acrylamide gel electrophoresis
TAP	Tandem affinity purification
TBE	Tris borate EDTA
TBS	Tris buffered saline
TCA	Tricholoracteic acid
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
YNB	Yeast nitrogen base
YPD	Yeast peptone dextrose
WT	Wild type

INTRODUCTION

I. RNA Polymerase II (RNAPII) and the Transcription Cycle

In eukaryotes, there are three multi-subunit RNA polymerase enzymes that are responsible for gene transcription, RNA polymerase I, II and III. RNA polymerase II (RNAPII) is responsible for synthesizing all protein-coding RNAs and many non-coding RNAs including small nuclear RNAs and microRNAs. RNAPII is a 514 kDa complex consisting of 10-core subunits and two heterodimer subunits Rpb4 and Rpb7 that associate with the 10-core subunits (Bushnell and Kornberg, 2003). Some of the subunits are conserved and share structural and sequence homology with subunits of RNAPI and RNAPIII (reviewed in Cramer et al., 2008). Five of the subunits are common to all three eukaryotic RNA polymerases. A crystal structure of the complete 12 subunits of RNAPII is shown in Figure 1.

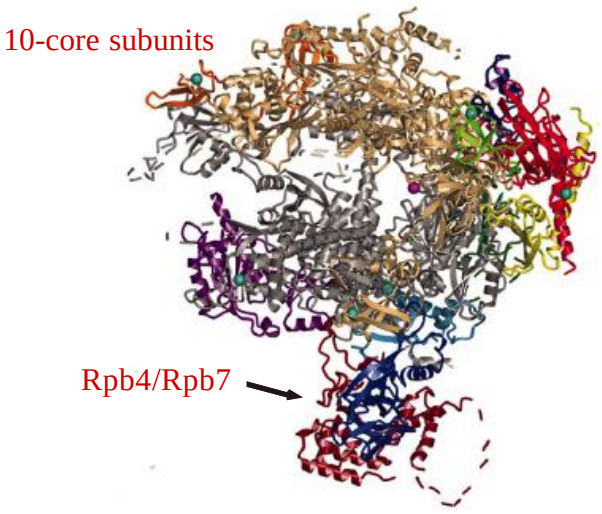


Figure 1. A ribbon representation of the 12-subunits of RNAPII. Black arrow indicates the position of Rpb4 and Rpb7 heterodimer.

RNAPII is subject to regulation at all stages of the transcription cycle including initiation, elongation and termination. In particular, regulatory factors (activators or repressors) bind to specific gene sequences at the promoter near the transcription start site and initiate or repress transcription by either interacting with transcription machinery or by recruiting factors that modify chromatin structure (reviewed in Hahn, 2004). The core promoter acts as platform for the assembly of transcription pre-initiation complex (PIC). PIC includes general transcription factors TFIID, TFIIB, TFIIE, TFIIF and RNAPII (reviewed in Thomas and Chiang, 2006). PIC assembly begins with the binding of TATA binding protein that promotes the binding of other transcription factors and the recruiting of RNAPII at the gene promoter. There are two proposed pathways for the order of entry of the general factors. First pathway describes a sequential assembly of the factors and RNAPII at the promoter (Buratowski et al., 1989) whereas the second pathway describes a preassembled holoenzyme RNAPII (Kim el at., 1994). After PIC formation, DNA strands are separated and the template strand is inserted into RNAPII active site resulting in open promoter complex that allows the synthesis of nascent RNA from the transcription start site (Treutlein et al., 2012).

After the nascent RNA reaches ~ 30 nucleotides in length, RNAPII escapes the promoter to enter early elongation stage. The transition from early elongation to productive elongation is accompanied with promoter-proximal pausing in mammals, which is a regulatory process of transcription. RNAPII can escape the pausing state to productive elongation through the action of positive elongation factor (P-TEFb). Several RNA processing factors and histone modifying proteins are recruited to elongation complex and interacted with RNAPII such as TREX complex and FACT complex

respectively (reviewed in Guo et al.2013; Kwak et al., 2013; Zhou et al., 2012). The termination of the transcription cycle occurs when RNAPII and the newly synthesized RNA are released from the DNA template. Termination of transcription requires poly (A) signal and is coupled with RNA 3′ end processing. Termination is an important regulatory process because it prevents interference with downstream elements for example downstream promoters and it promotes RNAPII recycling (reviewed in Kuehner, 2011).

II. The C-terminal Domain (CTD) of RNAPII

Although RNAPII shares structural homology with RNAPI and RNAPIII, it possesses a unique C-terminal domain (CTD) at its largest subunit Rpb1. CTD comprises multiple tandem repeats of heptapeptide with the consensus: Tyr-Ser-Pro-Thr-Ser-Pro-Ser, the amino acids are numbered one to seven: $Y_1S_2P_3T_4S_5P_6S_7$ (reviewed in Corden 1990). The number of the heptapeptide repeats varies among different organisms, for example yeast *S. cerevisiae* has 27 repeats whereas mammals and Drosophila have 52 and 42 repeats respectively (reviewed in Prelich, 2002). It has been established that the importance of the number of the repeats and the critical function of each amino acid among the repeats differ among organisms (reviewed in Jasnovidova and Stefl, 2013). Truncation studies demonstrated that removing more than half of the heptapeptide repeats in yeast *S. cerevisiae* results in growth defects, for example cells containing 10-12 complete heptapeptide repeats are conditionally viable whereas cells containing fewer than 10 repeats are not viable (Nonet et al., 1987). In addition, substitution of Ser2 or

Ser5 to alanine or glutmatae is lethal in yeast *S. cerevisiae* as well as the substitution of Tyr1 to phenylalanine (West et al., 1995). In contrast, mutation of Ser2 to alanine is not lethal in yeast *S. pombe* during vegetative growth, but is required for mating (Schwer et al., 2011).

During transcription, the CTD undergoes a cycle of dynamic post-translational modifications most frequently phosphorylation with five potential sites for phosphorylation (Y_1, S_2, T_4, S_5, and S_7). However, the main phosphorylation sites that have been identified are Ser2, Ser5 and Ser7 (reviewed in Jasnovidova and Stefl, 2013). Several protein kinases have been identified as responsible for CTD phosphorylation in mammals and yeast (Table 1), the kinases are TFIIH cyclin-dependent kinase (Cdk7/Kin28), mediator-associated kinase (Cdk8/Srb10), and P-TEFb kinase (Cdk9/Ctk-1 and Bur1) (reviewed in Buratowski, 2009; Phantnani and Greenleaf, 2006; Prelich, 2002). Cdk7/Kin28 phosphorylates Ser5 after PIC formation. In yeast TFIIH kinase is required for normal transcript levels *in vivo*, data from microarray experiments have shown that the inhibition of Kin28 results in a dramatic reduction of mRNA levels and Ser5-P phosphorylation levels (Hong et al., 2009).

As the RNAPII proceeds in transcription, the level of Ser5-P phosphorylation is reduced and the CTD becomes phosphorylated at Ser2 by Cdk9 (mammals) and Ctk-1/ Bur1 (yeast). Chromatin immunoprecipitation (ChIP) experiments have shown that the greatest level of Ser5-P phosphorylation is observed at the promoter regions whereas the greatest level of Ser2-P is observed at the coding regions (Kanin et al., 2007). In addition, recent studies have discovered that Ser7 is phosphorylated *in vivo* by TFIIH kinases in mammals and yeast, ChIP signals show enrichment of Ser5-P and Ser7-P at the 5′ end

and inhibition of TFIIH kinases *in vivo* results of loss of both marks (Kim et al., 2009; Akhtar et al., 2009).

P-TEFb was first identified as an elongation factor that promotes the release of paused RNAPII at the promoter proximal region by phosphorylating Spt4/Spt5 complex a subunit of DRB sensitivity inducing factor (DSIF)(reviewed in Guo et al., 2013; Kwak et al., 2013). However, recent studies have proposed that Cdk9 kinase of P-TEFb has a dual function during elongation. Besides its role in phosphorylating the Spt4/Spt5 complex, it also phosphorylates Ser2 of the CTD (reviewed in Guo et al. 2013). In yeast, there are two kinases with sequence and functional homology to Cdk9, the kinases are CTDK-1 (Ctk-1) and Bur 1. It was discovered that Ctk1 primary phosphorylates Ser2 (Cho et al., 2001) whereas Bur1 phosphorylates Spt4/5 complex as well as Ser2 (Liu et al., 2009).

Three protein phosphatases (Ssu72, Fcp1, and Rtr1) have been implicated in the removal of the phosphate group from the CTD in yeast (Table 1) (reviewed in Hsin and Manley, 2012; Mosley et al., 2009). Ssu72 is an essential conserved termination factor that has been characterized as a Ser5 phosphatase in the protein tyrosine phosphatase superfamily (Ganem et al., 2003). Ssu72 localizes at the 5′ and 3′ end of the gene with the highest enrichment at the 3′ end. Depletion of Ssu72 results with increased Ser5-P at the 3′ end but not at the 5′ end (Bataille et al., 2012). In addition, Ssu72 depletion impairs transcription *in vitro* (Krishnamurthy et al., 2004). Ssu72 has also been characterized as Ser7 phosphatase because it is depletion results in increased Ser7-P at the 3′ end (Bataille et al., 2012; Zhang et al., 2012)

Rtr1 an RNA binding protein in yeast was identified as a CTD phosphatase that specifically dephosphorylates Ser5-P (Mosley et al., 2009). Experimental analyses *in vivo*

and *in vitro* have shown that Rtr1, and its human homology, RPAP2 regulate the transition of high Ser5-P level to high Ser2-P (Mosley et al., 2009; Ni Z et al., 2014; Hsu et al., 2014). ChIP analysis has proven that Rtr1 localizes at the region proceeding Ser2-P and not enriched at the promoter region (Mosley et al., 2009). Deletion of Rtr1 causes an accumulation of Ser5-P level throughout the coding gene (Mosley et al., 2009).

Fcp1 interacts with the elongation complex to specifically dephosphorylate Ser2-P. ChIP data have indicated that Fcp1 cross-links at the promoter and the coding region and Fcp1 mutants cause an increased level of Ser2-P in the coding region (Cho et al., 1999, 2001). Additionally, Fcp1 has an important role in recycling RNAPII to the hypo-phosphorylated form for PIC assembly (Cho et al., 1999).

Table 1. List of the proposed CTD kinases and phosphatases in *S. cerevisiae* and mammals (reviewed in Hsin and Manley, 2012; Palancade and Bensaude, 2003; Buratowski, 2009).

CTD Site	Protein/complex	Function	*S. cerevisiae*	Mammals
Ser5, Ser7	TFIIH	Kinase	Kin28	Cdk7
Ser5	Mediator	Kinase	Srb10	Cdk8
Ser2	P-TEFb	Kinase	Ctk1 Bur1	Cdk9
Ser5, Ser7		Phosphatase	Ssu72	Ssu72
Ser5				Scp1
Ser5		Phosphatase	Rtr1	RPAP2
Ser2		Phosphatase	Fcp1	Fcp1

III. CTD and Its Binding Proteins

During transcription cycle several factors and multi-subunits proteins interact with RNAPII to regulate transcription and many of these interactions are associated with CTD modifications. The CTD serves as a flexible platform for the recruitment of these factors and proteins.

Early in transcription hypo-phosphorylated CTD interacts with TBP a subunit of TFIID general factor and multi-subunit Mediator complex, the interaction of the CTD with these factors helps position RNAPII at the promoter for PIC assembly (Imasaki et al., 2011). Phosphorylation of Ser5 by TFIIH kinase disrupts the bonds stabilizing PIC complex and triggers the release of RNAPII from the complex (Sogaard and Svejstrup, 2007).

The modified CTD mediates the coupling of transcription with RNA processing. Multiple RNA processing factors have been described which interact with different modified states of the CTD (McCracken et al., 1997). For example Ser5-P CTD interacts with RNA capping enzyme; this interaction facilitates the capping of nascent RNA when it exists from the active site (Cho et al., 1997). It has been proposed that the capping enzyme is recruited to the promoter in TFIIH kinase dependent manner and was demonstrated through ChIP assay in *S. cerevisiae.* ChIP assay has showed that capping enzyme is localized to 5′ end of the gene but not in the coding region (Komarnitsky et al., 2000). The coupling of the transcription and capping protects nascent RNA from degradation and ensures that is accurately capped.

In addition to the capping enzyme, The CTD interacts with RNA splicing and processing factors with different phosphorylated states for example yeast splicing factor Prp40 interacts with hyper-phosphorylated CTD (Morris and Greenleaf, 2000) whereas the polyadenylation cleavage factor (Pcf11) interacts with CTD containing only Ser2-P (Meinhart and Cramer, 2004).

Many of the transcription elongation factors that interact with modified CTD mediate the coupling of messenger RNA (mRNA) synthesis with mRNA nuclear export

such as TREX complex, which interacts with phosphorylated Ser5 and Ser2 (Meinel et al 2013). In addition several elongation factors facilitate RNAPII elongation through chromatin such as PAF and FACT complexes (reviewed in Jaehning, 2010; Belotserkovskaya and Reinberg, 2004). However, the precise binding of many of these elongation factors with CTD and how these interactions control transcriptional regulation is still unknown. Therefore, the work presented here is focused on developing a method to isolate the CTD and its binding proteins for proteomics analysis in order to contribute of revealing the basis of the interactions of the CTD with these factors.

Introduction to Quantitative Proteomics

Proteomics in general can be described as the study of all the proteins expressed by the genome and the characterization of their features that includes but is not limited to expressions, structures, activities, interactions, and post-translational modifications. In the post-genomic era, several analytical methods have been developed to provide tools for analyzing proteome and the complexity associate with it, among these analytical methods is mass spectrometry based-proteomics. Mass spectrometry has proved to be a significant tool in identifying protein within a complex as well as investigating protein-protein interaction because mass spectrometry is very sensitive, reproducible and versatile (reviewed in Aebersold and Mann, 2003; Bauer and Kuster, 2003).

Fundamentally, mass spectrometry measures mass-to-charge ratio (m/z) of ions in gas phase. Most mass spectrometer consists of three components including an ion source that ionizes analyte molecules, a mass analyzer that separates the ions based on m/z ratio, and a detector that records the number of ions (reviewed in Aebersold and Mann, 2003).

Electrospray ionization (ESI) and matrix-assisted desorption/ ionization (MALDI) are the most common ionization techniques that are capable of ionizing peptides or proteins (Fenn et al., 1989; Karas et al., 1988). In mass spectrometry based-proteomics, enzymatic digestion of protein mixture followed by separation methods are required prior to mass spectrometer analysis to simplify the mixture in order to achieve identification of the proteome with high resolution. Previously, two-dimensional protein polyacrylamide gel electrophoresis (2D-PAGE) followed with mass spectrometry analysis was the common method that provides quantitative analysis for proteomics (reviewed in Hanash, 2000). In 2D-PAGE method, proteins are separated in first dimension according to isoelectric point (pI) by isoelectric focusing and in the second dimension according to the molecular weight by SDS-PAGE (O'Farrell, 1975). Following separation, the proteins of interest are excised from the gel, treated with enzymatic digestion, and analyzed by mass spectrometry (Shevchenko et al., 1996). Although 2D-PAGE is a good method for quantitative proteomics, there are several limitations associated with it such as the difficulty in detection every proteins in the sample and the incompatibility in detection of high pI proteins as well as high molecular weight proteins (Gygi et al., 2000, Oh-Ishi et al., 2000).

Consequently, to overcome the limitations that are associated with 2D-electrophoresis method, alternative two-dimensional chromatography separation methods were developed by the Yates lab: Link et al., 1999; and Washburn et al., 2001; to provide a high-throughput analysis of proteomics. Multi-dimensional protein identification technology (MudPIT) is a two-dimensional liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS) (Link et al., 1999; Washburn et al., 2001). In

MudPIT, a protein mixture is subjected to enzymatic digestion to yield a mixture of peptides, the mixture is then loaded in microcapillary column packed with two or three chromatography phases (Link et al., 1999; McDonald et al., 2002). Once the peptide mixture is loaded into the column, the peptides are fractionated through the chromatography phases (strong cation exchange and reverse- phase matrix) and then the peptide fragments are sprayed through high voltage into tandem mass spectrometer for analysis. Subsequently, the obtained MS/MS spectra are searched against available FASTA protein database using a database search algorithms such as SEQUEST (Link et al., 1999). MudPIT coupled with tandem mass spectrometry has been used successfully in identifying large-scale of *S. cerevisiae* proteome including proteins with large molecule weight as well as proteins with a high pI (Washburn et al., 2001; McDonald et al., 2002; Mosley et al., 2011).

In quantitative proteomics, tandem mass spectra (MS/MS) are searched using a database search engine such as MASCOT, SEQUEST, Phenyx, and X!Tandem. The basic function of these programs is to score the correlation between an experimental fragment ion spectrum of a peptide against a theoretical peptide spectrum generated by the database search engine (Reviewed in Nesvizhskii et al., 2007). Typically, the output of the database-search is a list of proteins and peptides with score presents the number of peptide spectrum matches (PSMs) to the theoretical proteins within a set false-discovery rate threshold (typically $\leq 5\%$). Following peptide spectra matching, the proteins and peptides are analyzed manually or using program such as Scaffold to provide quantitative information including the normalized PSMs values (NSAF) and the level of protein enrichments (Mosley et al., 20011and 2013).

Previous analyses of the RNAPII protein interacting network using affinity purification coupled with MudPIT quantitative analysis have shown that the highest protein enrichment levels are assigned to one of 12 subunits of RNAPII when using Rpb3, Rpb7, or Rpb11 as a bait of the affinity tag (Mosley et al., 2009, 2011 and 2013). These analyses have demonstrated the efficiency of using this approach in identifying high abundance proteins that interact with RNAPII during multiple stages of the transcription cycle including the elongation factor Spt5/Spt4 and subunits of the general transcription factor TFIIF (Tfg1/Tfg2/Taf14) (Mosley et al., 2013). However, they also reveal a limitation of this approach in identifying low abundance RNAPII interacting proteins that only interact with RNAPII at certain state during transcription cycle. As shown in Figure 2 several of low abundance proteins were not detected in MudPIT analysis such as Set2, Asr1, Cdc73, Npa3 and Leo1 (Figure 2 is adapted from Mosley et al., 2013).

Consequently, our main objective in the work presented here was to devise a method to facilitate the identification of low abundance RNAPII interacting proteins that are recruited to the CTD during transcription cycle and participate in controlling regulation of transcription. In order to accomplish that, we devised a purification scheme that specifically isolate the CTD-associated and CTD-less RNAPII to identify and characterize the proteins that interact with both the CTD and the globular core of RNAPII including low abundance proteins.

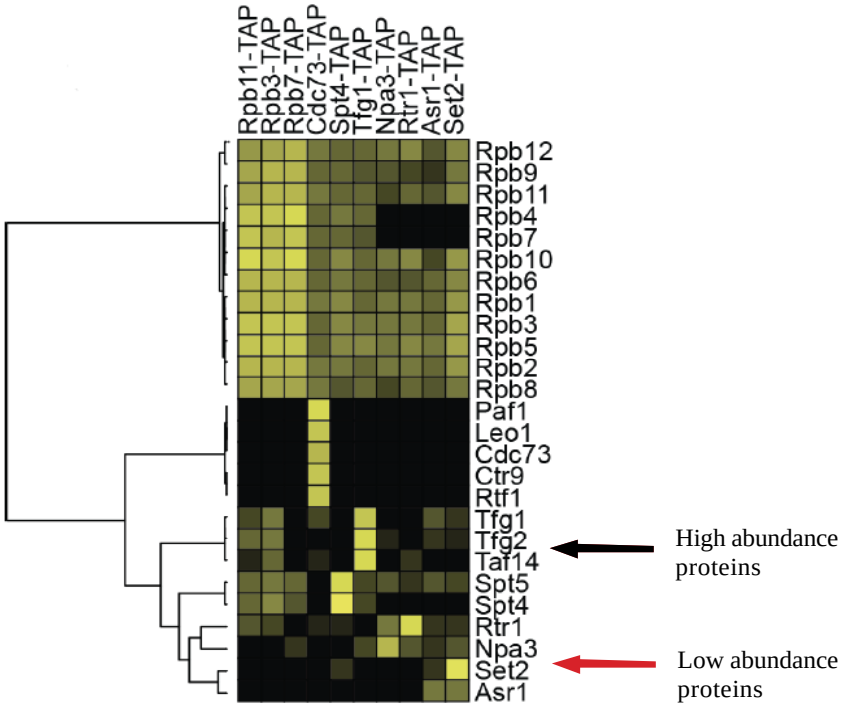


Figure 2. Hierarchical clustering analysis by MudPIT of RNAPII interacting proteins. The protein identified by mass spectrometry are listed to the right of the cluster and the baits tag proteins are labeled to the top. The yellow squares represent the NSAF (normalized PSMs) value and the black squares indicate no protein detection. Black arrow to the right indicates the position of high abundance proteins and the red arrow indicates the position of low abundance proteins.

I. Method development and Experimental Design

In the presented work our research objective was designing a novel method to isolate CTD of the largest subunit of RNAPII and the CTD interacting proteins in *S. cerevisiae* to improve the identification of low abundance RNAPII interacting proteins. Consequently, we have introduced Tobacco etch virus (TEV) protease cleavage site (ENLYFGQG) into the flexible linker region of Rpb1 upstream of the CTD using site directed-mutagenesis PCR-based method. The site for TEV insertion at the linker region was chosen based on work previous done by Kornberg lab. Kornberg and co-workers

prepared a CTD-less version of RNAPII by replacement of five amino acids from the junction between CTD and the rest of the subunits with Factor Xa (FXa) cleavage site (IEGR) (Li and Kornberg, 1994). They used the isolated CTD-less RNAPII for *in vitro* transcription analysis. In addition, they used the CTD-less RNAPII to generate a 3-dimensional structure of RNAPII by electron crystallography at 16 A˚ resolution (Meredith et al., 1996).

The next approach in our method, was constructing yeast strains express the modified Rpb1-TEV gene using various molecular biology methods including polymerase chain reaction (PCR) DNA amplification, gene knockout, homologous recombination, and 5-fluoroorotic acid (5-FOA) plasmid shuffle. Further to isolate CTD and its associated proteins, we performed tandem affinity purification (TAP) to specifically purify RNAPII complex and to perform on beads TEV cleavage of the CTD. Tandem affinity purification (TAP) developed by Rigaut et al. 1999, is a method to isolate protein complexes expressed at their endogenous levels under native conditions. The basic concept of TAP purification involves fusing TAP tag to protein of interest then introducing the construct into the host cell (Puig et al., 2001). When expressed in the cell, the TAP Tagged-protein forms complex with the endogenous partners (Bauer and Kuster, 2003). Subsequently, the protein complex along with the associated proteins is purified using two consecutive affinity steps. The TAP tag consists of two immunoglobulin (IgG) binding domains of *Staphylococcus aureus* protein A (ProtA) and calmodulin binding peptide (CBP). The two affinity domains are separated by TEV protease cleavage site (Puig et al., 2001; Rigaut et al., 1999).

The general procedure of TAP purification is involved the following: the protein complex along with the associated proteins is recovered on IgG-coupled resins through a strong interaction of IgG with protein A. Further, the protein complex is released from the IgG resins by proteolytic cleavage using TEV protease. TEV protease cleavage allows the release of the protein A portion of the TAP tag with IgG-coated resins effectively eluting the protein of interest and its associated proteins (Rigaut et al., 1999). The protein complex is then incubated with calmodulin-coated resins where it becomes immobilized to calmodulin resins through calmodulin binding protein (CBP) in the presence of calcium. Subsequent to multiple washing steps, the protein complex along with the associated proteins is released by using the chelating agent EGTA (Rigaut et al., 1999; Puig et al., 2001; Bauer and Kuster, 2003).

The TAP two affinity steps strategy has advantages over regular one-step affinity purifications because it significantly reduces non-specific background protein binding therefore simplifying the identification and characterization of a protein complex by mass spectrometry (Rigaut et al., 1999; Bauer and Kuster, 2003). Another advantage of TAP is performing the purification near native conditions avoids the loss of any of the associated proteins, which could occur if more stringent conditions are used (Bauer and Kuster, 2003). In addition, TAP is a valuable method for purifying protein complexes from different cellular compartments as well as purifying large protein complexes (Gavin et al., 2002).

In our method, we tagged yeast Rpb3 in strains expressing Rpb1-TEV gene with a modified TAP tag. The modified TAP tag has the TEV cleavage site replaced with HR 3C protease (Human Rhinovirus 3C, hereafter referred to as 3C protease) cleavage site

(LEVLFQGP). Additionally, we modified the original TAP purification described above to take advantage of TEV cleavage site upstream of the CTD. In general, we performed on bead cleavage with TEV protease while the RNAPII complex was immobilized on either IgG resins or calmodulin resins. The CTD-less RNAPII complex was released from IgG beads by 3C protease cleavage and further was purified through incubation with calmodulin beads and eluted with EGTA. The CTD flow-through and the CTD-less RNAPII elutions were analyzed in mass spectrometry using the MudPIT method followed with quantitative analysis. A cartoon schematic of our purification model is shown in Figure 3.

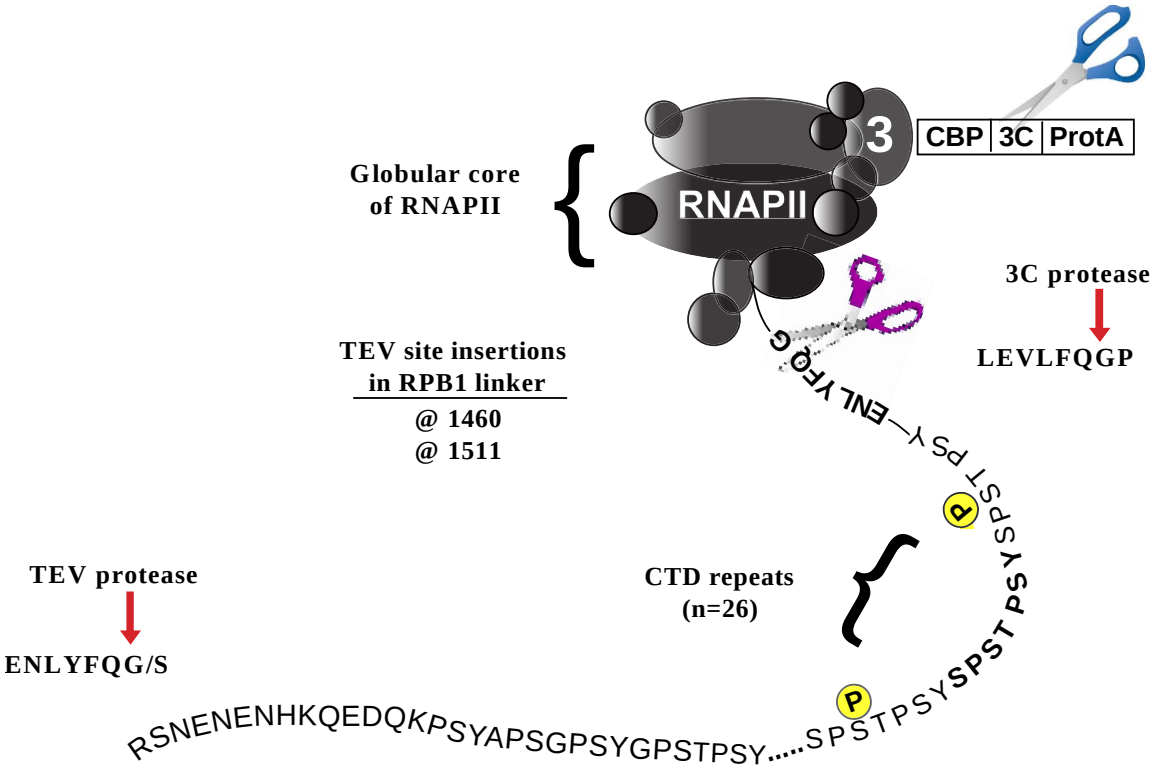


Figure 3. A cartoon representation of the proposed Rpb3-3C-TAP Rpb1-TEV purification scheme. TEV and 3C cleavage sites are shown as following: bottom left and top right respectively. The red arrows indicate the protease optimal cleavage site.

MATERIALS AND METHODS

I. Reagent and Solution Recipe

Table 2. Recipes of the reagents and solutions used in in this work.

Reagent	Recipe
LB- kanamycin media	1% bacto-tryptone, 0.5% yeast extract, 1% NaCl, 50 µg/ml kanamycin
LB-kanamycin Plate	1% bacto-tryptone, 0.5% yeast extract, 1% NaCl, 1.5% agar, ,50 µg/ml kanamycin
LB- ampicillin media	1% bacto-tryptone, 0.5% yeast extract, 1% NaCl, 100 µg/ml ampicillin
LB-ampicillin Plate	1% bacto-tryptone, 0.5% yeast extract, 1% NaCl, 1.5% agar, ,100 µg/ml ampicillin
10x TE buffer	100 mM Tris-HCl pH 7.5, 10 mM EDTA pH 8
TE-LiAc solution pH 7.5	100 ul of 10x TE buffer pH 7.5, 100 ul of LiAc, 800 ml sterile H_2O
PEG-LiAC solution	100 ul of 10x TE buffer pH 7.5, 100 ul of LiAc, 800 ml sterile polyethylene glycol (PEG)
YPD media	2 % bacto peptone, 1% yeast extract, 2% glucose
YNB-His$^-$ Ura$^-$ plate	0.67 % yeast without amino acid extract, 0.077% CSM drop out mix (histidine$^-$ uracil$^-$),), 2% glucose, 2% agar
YNB-Leu$^-$ Ura$^-$ plate	0.67 % yeast without amino acid extract, 0.077% CSM drop out mix (leucine$^-$ uracil$^-$), 2% glucose, 2% agar
YNB-His$^-$ Plate	0.67 % yeast nitrogen base without amino acid extract, 0.077% CSM drop out mix (histidine$^-$), 2% glucose, 2% agar
YNB-Leu$^-$ Plate	0.67 % yeast nitrogen base without amino acid extract, 0.077% CSM drop out mix (leucine$^-$), 2% glucose, 2% agar
5-fluoroorotic acid (5-FOA plate	0.67 % yeast without amino acid extract, 0.077% CSM drop out mix (*URA3*$^-$), uracil 25 mg, 2% glucose, 2% agar, 0.1 % 5-FOA
1x TBE	100 mM Tris-HCl pH 7.5, 0.5% boric acid, 10 mM EDTA pH 8
1 % agarose gel	1g agarose in 100 ml 1x TBE
12 % SDS-PAGE	8 ml 1M Tris-HCl pH 8.8, 0.108 ml SDS (20%), 6 ml polyacrylamide (40%), 0.108 ml 10% APS, 0.02 ml TEMED, 5.8 ml ddH_2O
2 % SDS-PAGE stacking gel	1.14 ml 1M Tris-HCl pH 6.8, 0.04 ml SDS (20%), 0.7 ml polyacrylamide (40%), 0.076 ml 10% APS, 0.01 ml TEMED, 5.7 ml ddH_2O
1x TGS	25 mM Tris-HCl pH 8, 192 mM glycine, 0.1% SDS (20%)
Transfer buffer	25mM tris base, 190 mM glycine, 10% methanol and 0.1% SDS (20%)

1x Tris-Buffered Saline (TBS)	25 mM Tris-HCl pH 7.5, 136 mM NaCl
5% Blocking Buffer	5 g milk in TBS buffer
TAP Lysis Buffer	40mM Hepes-KOH, 10% glycerol, 350 mM NaCl, 0.1% Tween-20, added prior to use: 1x Yeast protease inhibitors (Sigma), and 1mM β-ME
TEV Cleavage Buffer	10 mM Tris-HCl pH 8, 10% glycerol, 150 mM NaCl, 0.1% IPEGAL, 0.5 mM EDTA, added prior to use: 1x Yeast protease inhibitors (Sigma), and 1mM β-ME
3C Cleavage Buffer	50 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA, added prior to use: 1x Yeast protease inhibitors (Sigma), and 1mM β-ME
Calmodulin Binding Buffer	10 mM Tris-HCl pH 8, 1 mM MgOAc, 1 mM imidazole, 2 mM $CaCl_2$, 100 mM NaCl, 10 % glycerol, added prior to use: 1x Yeast protease inhibitors (Sigma), and 1mM β-ME
Calmodulin Elution Buffer	10 mM Tris-HCl pH 8, 1 mM MgOAc, 1 mM imidazole, 2 mM EGTA, 100 mM NaCl, 10 % glycerol, added prior to use: 1x Yeast protease inhibitors and 1mM β-ME
CTD Cleavage Buffer	10 mM Tris-HCl pH 8, 150 mM NaCl, 0.1% IPEGAL, 2 mM $CaCl_2$, added prior to use: 1x Yeast protease inhibitors (Sigma), and 1mM β-ME
Fixing Solution	30% ethanol, 10 % acetic acid, 60 % MilliQ H_2O
Ethanol Wash	30% ethanol, 70 % MillQ H_2O
Sensitizer Solution	0.02 Sodium thiosulfate in MilliQ H_2O (w/v)
Silver Nitrate Solution	0.1% silver nitrate, 0.02% formaldehyde, 99.9% MilliQ H_2O
Developing Solution	2.5% sodium carbonate, 0.05 % formaldehyde, 0.005 % sodium thiosulfate in MilliQ H_2O (w/v)
Stop Solution	0.5 glycine in 100 ml MilliQ H_2O (w/v)
Buffer A	0.5% acetonitrile. 0.1% formic acid

II. Construction of *RPB1*-TEV-CTD Plasmids

TEV Insertion Reaction

To introduce a TEV cleavage site at amino acid 1511 into the linker region of *RPB1* gene by PCR based site directed mutagenesis, we used a DNA template of *S. cerevisiae RPB1* gene in pENTR-D-TOPO (Life Technologies, Mosley Lab) (Table 4) and *RPB1*-TEV-1511 forward and reverse primers that were designed in-house (Table 3). The insertion reaction was performed using Quick change II XL site directed mutagenesis kit (Agilent Technologies, catalog #200521) following a modified procedure of the

manufacture instructions. PCR reactions (50 µl) contained 100 ng of p*RPB1* template, 3 µM of each primer, 3µM of dNTP mix, 1 µl of QuickSolution and 2.5 units of *PfuUltra* high fidelity DNA polymerase. The PCR cycles were: 95 °C for 2 minutes followed by 18 cycles. Each amplification cycle consisted of: 1 minute denaturation at 95 °C, 1 minute annealing at 55 °C, and 16 minutes extension (2 minutes per kb of template) at 68 °C. The PCR reaction was completed with final extension for 7 minutes at 68 °C. The PCR product was treated with 10 units of *Dpn I* restriction enzyme at 37 °C for 2 hours to digest the parental template, and then 2 µl of *Dpn* I-treated PCR product was transformed to XL10-Gold Ultracompent cells by heat shock for 30 seconds at 42 °C. Subsequently, 0.5 ml of S.O.C medium (Invitrogen, catalog # 15544-034) was added to the transformation reaction followed by incubation for 1 hour at 37 °C with shaking. After the incubation was complete, the transformed cells were spread onto LB- kanamycin plate and incubated overnight at 37 °C. Following overnight incubation, cells from the plate were inoculated in 3 ml of LB-kanamycin media and grown overnight at 37 °C with shaking. The next day, the plasmids were purified using QlAPrep spin miniprep kit (Qiagen, catalog # 27106) following manufacturer's instructions and quantified using spectrophotometer. And then the purified plasmids were sent to Genewiz, Inc. for sequences analysis using primers just upstream of TEV insertion site.

LR reaction

An LR reaction was performed to introduce Rpb1-TEV-CTD construct into a destination vector with a selective marker: GPD413-*HIS3* (Addgene) and GPD415-*LEU2* (Addgene) (Table 4). Fifty nanograms of *RPB1*-TEV-CTD construct was mixed with 150 ng of the destination vector and 1 µl of TE buffer (pH 8), followed by the addition of 1x

LR Clonase II reaction mix. The reaction mixtures were centrifuged for a minute and then incubated overnight at room temperature. After overnight incubation, 1 µl of proteinase K solution was added to the reaction mixtures with vortex mixed to terminate the reaction and then incubated at 37 °C for 10 minutes. One microliter of each reaction mixture was transformed to TOP 10 cells by heat shock for 30 seconds at 42°C. After heat shock transformation, 250 µl of S.O.C. media was added to the transformed cells followed by incubation at 37 °C for an hour with shaking. The transformed cells were spread onto LB- ampicillin plates and incubated overnight at 37 °C. In the following day, cells from the plates were inoculated in two separate flasks of 3 ml LB-ampicillin media and grown overnight at 37 °C with shaking. The next day, the plasmids were purified using QIAPrep spin miniprep kit (Qiagen, Catalog # 27106) and then they were sent to Genewiz, Inc. for further sequences analysis.

Introduce TEV site at amino acid 1460 into RPB1 linker region

To investigate the optimal CTD cleavage location, we introduced TEV site at amino acid 1460 into Rpb1 linker region using *RPB1*-template (Table 4) and *RPB1*-TEV-1460 forward and reverse primers (Table 3). TEV insertion reaction and LR reaction described above were repeated in this experiment. TEV insertion was confirmed by DNA sequence analysis (Genewiz) using a primer to the *RPB1* gene.

Construction of RPB1-3C-CTD Plasmid

To examine the cleavage specificity of 3C protease and TEV protease at their cleavage sites we introduced 3C cleavage site into Rpb1 linker region at amino acid 1460 using *RPB1*-template (Table 4) and *RPB1*-3C-CTD forward and reverse primers (Table 3). 3C insertion was confirmed by DNA sequences analysis as outlined above.

Table 3. List of primers used in construction yeast *S. cerevisiae* strains.

Primer	Sequences*	T_m°C
RPB1-TEV-1511 (F)	5´-GCTGGAGGATTTACAGCGTAC**GAGAATTT GTATTTTCAGGGT**GGTGGTGCTGATTATGG TGAAGCCACGTCTCCATTTGGTGCTTATGGT GAAGCACCTACATCTCCCGGATTTGGAGTCT CC-3´	71.9
RPB1-TEV-1511 (R)	5´-TTCACCATAATCAGCACCACC**ACCCTGAA AATACAAATTCTC**GTACGCTGTAAATCCTC CAGCCATAGCGTCATTTGAACCCGAATCAAC CAGAGGTGAAAACATTAGCTCATCTTTAACG TC-3´	71.5
RPB1-TEV-1460 (F)	5´-TGAGGAGTCACTGGTAAAATACATGCCAG AACAAAAAATA**GAGAATTTGTATTTTCAGG GT**ACTGAGATTGAAGACGGACAAGATGGTGG CGTCACACCAT-3´	70.7
RPB1-TEV-1460 (R)	5´-ATGGTGTGACGCCACCATCTTGTCCGTCT TCAATCTCAGT**ACCCTGAAAATACAAATTC TC**TATTTTTTGTTCTGGCATGTATTTTACCAG TGACTCCTCA-3´	70.7
RPB1-3C-CTD (F)	5´-TTTGATGTGATGATCGATGAGGAGTCACT GGTAAAATACATGCCAGAACAAAAAATA**TT GGAGGTTTTGTTTCAGGGACCT**ACTGAGAT TGAAGACGGACAAGATGGTGGCGTCACACC A-3´	69.1
RPB1-3C-CTD (R)	5´-TGGTGTGACGCCACCATCTTGTCCGTCTT CAATCTCAGT**AGGTCCCTGAAACAAAACC TCCAA**TATTTTTTGTTCTGGCATGTATTTTA CCAGTGACTCCTCATCGATCATCACATCAAA-3´	67.8
RPB3-TAP (F)	5´- CCTTACTCCAATGCATCTCAAATGGGTA ATACTGGATCAGGAGGGTATGATAATGCTT GGTCCATGGAAAAGAGAAG-3´	69.1
RPB3-TAP (R)	5´- TTCTCTTATTATTTTCGGTTCGTTCACTTG TTTTTTTTCCTC TATTACGCC CACTTGAGAA TACGACTCACTATAGGG-3´	67.8

*The insertion sequences (TEV site and 3C site) are in the bold font.

III. Construction of Yeast Strains

Yeast Transformation

To introduce *RPB1*-TEV construct into yeast strain that express a wild type Rpb1

gene from a plasmid (the native *RPB1* gene was deleted with a *CLONAT* resistance cassette) using standard lithium acetate transformation method, cells from *RPB1 GPD URA3* strain (Table 4) were inoculated in 25 ml of YPD media and grown at 30 °C overnight. Cells were collected by centrifugation at 4 °C after reaching OD_{600} ~ 0.5, they were washed with 1 ml of sterile water and then cell pellets were re-suspended with 500 µl of TE-LiAc solution. The yeast transformation reaction was prepared by mixing 5 µl of 10 mg/ml of single stranded carrier DNA (Sigma-Aldrich, catalog # D7656-1ML), 500 ng of the *RPB1*-TEV-CTD plasmid and 100 µl of cell suspension. Subsequently, 600 µl of PEG-LiAC solution was added to transformation reaction followed by vortex mixing. The transformed cells were incubated at 42 °C for 30 minutes. Following the incubation, cells were spin down for few seconds and cell pellets were re-suspended with 200 µl of sterile water. The cells were plated in the appropriate selective plates: YNB-His^- Ura^- plate, or YNB-Leu^- Ura^- plate. The plates were incubated at 30 °C for 4 days.

5-Fluoroorotic Acid (5-FOA) Selection

5-fluoroorotic acid (5-FOA) was used to select for yeast strains that express the modified *RPB1* gene (Rpb1-TEV-CTD) and to evict the WT *RPB1 URA3* plasmid. Six single colonies of the yeast strains that were grown in the plates above were isolated and re-streaked each onto their appropriate plates YNB-His^- Ura^- plate or YNB-Leu^- Ura^- plate to obtain single uniformly white colonies and incubated at 30 °C to grow for 2 days. After the cells grew, they were re-streaked onto replica 5-FOA plates and incubated at 30 °C for 2 days. Resulting the 5-FOA resistant colonies were re-streaked onto YNB-His^- Ura^- plate or YNB-Leu^- $Ura,^-$ plate to test if they lost their ability to grow in media

lacking uracil. In addition, same colonies were re-streaked onto YNB-His$^-$ or YNB-Leu$^-$ to confirm their ability to grow in media lacking histidine and leucine.

Tagging Yeast Strains with Rpb3-3C-TAP Tag

PCR Reaction

Plasmid containing 3C-TAP-*URA3* cassette gene (Mosley lab) was amplified by PCR using *RPB3*-TAP primer pair (Table 3). PCR reaction mixtures (100 µl) contained 100 pg of *RPB3*-3C-TAP plasmid, 0.3 mM of dNTP mix, 0.3 mM of primer each, 10 µl of *pfx50* PCR mix, and 5 units of *pfx50* DNA polymerase. PCR amplification was carried out as follows: initial denaturation at 94 °C for 2 minutes, 35 cycles of denaturation at 94 °C for 30 seconds, annealing 52 °C for 30 seconds, and extension at 68 °C for 3 minutes. The PCR reaction was completed with final extension at 68 °C for 5 minutes. Following the amplification, PCR product was mixed with DNA loading dye and was loaded onto 1% agarose gel containing ethidium bromide, 8 µl of DNA ladder (Thermo Scientific) was loaded alongside the PCR product. The gel was run at 100 V in 1x TBE buffer until the dye line was 80% down the gel. DNA bands were visualized with ChemiDoc molecular imaging system (Bio-Rad). Subsequently, DNA bands were excised and purified using QlAquick gel extraction kit (Qiagen, Catalog #28704) and then quantified using spectrophotometer.

Yeast Transformation

Two micrograms of PCR product obtained in the above step was transferred to *RPB1*-TEV-CTD strains (constructed in section II) using Lithium acetate transformation method. Briefly, yeast strain was cultured in 30 ml YPD media and grown to OD_{600} ~ 0.5. Cells were collected by centrifugation at 4°C and were re-suspended with 500 µl of TE-

LiAc solution. Transformation reaction was prepared similar as in section II. After the incubation of transformed cells at 42 °C for 30 minutes, cells were incubated overnight at 4 °C to increase transformation efficiency. In the next day, the cells were spin down for 30 seconds, the supernatant was removed and the cell pellets were re-suspended with 200 µl of sterile water. And then the transformed cells were plated onto YNB-Leu$^-$ Ura$^-$ plate or YNB-His$^-$ Ura$^-$ plate. Cells were incubated at 30 °C for 4 days.

Western-Blot

Yeast Strains: Rpb3-3C-TAP Rpb1-TEV-CTD (constructed above), BY4741 (Open Biosystems), and Rpb3-TAP (Mosley lab) were inoculated each in 30 ml of YPD media and grown overnight at 30 °C. In the next day, cells were pelleted with centrifugation at 4 °C and then washed with sterile water. Cell pellets were re-suspended with TAP lysis buffer and they were lysed with glass beads on a disruptor genie for 30 minutes at 4°C and then the lysates were clarified at 14,000 RPM at 4 °C for 30 minutes. A 30 µl aliquot of each prepared lysates was mixed with same volume of 2x Laemmli loading buffer with BME (Bio-Rad, catalog # 161-0731), and the mixture was subjected to denaturation by boiling for five minutes at 100 °C. The denatured samples were loaded onto 12% SDS polyacrylamide gel alongside 5 µl of Precision Plus Protein Dual Color Standard molecular weight marker (Bio-Rad, catalog #161-0374). The gel was electrophoresed in 1x Tris-Gly-SDS (TGS) buffer at 200 V until the dye front reached the bottom of the gel. The proteins were transferred from the gel to nitrocellulose membrane in a wet transfer conditions overnight at 30 V. After overnight transferring, non- specific protein binding to the nitrocellulose membrane was prevented by blocking with 5% milk in 1x TBS buffer for 45 minutes. Subsequently, the membrane was probed overnight at 4 °C with

anti-TAP polyclonal primary antibody (Thermo, 1:1000 diltuion). After overnight incubation, the membrane was washed three times with 1x Tris-buffered saline (TBS) (10 minutes each). And then the membrane was incubated with anti-rabbit IgG horseradish peroxidase secondary antibody (GE Healthcare, 1:5000 dilution). After incubation with secondary antibody the membrane was washed three times with 1x TBS buffer (10 minutes each). Proteins were visualized ECL (GE Healthcare) and were detected using Fujfilm imager FLA-5000 at 473 nm and the LBP blue filter.

Table 4. List of plasmids and yeast strains used or constructed in construction of Rpb1-TEV-CTD strains.

Plasmid	**Description**	**Source or reference**
RPB1-template	*RPB1*-PENTR-D TOPO	Mosley Lab
RPB1-TEV construct	*RPB1*-TEV-CTD	This work
pAG413GAL-ccdB	GPD413-*HIS3*	Addgene
pAG415-ccdB	GPD415-*LEU2*	Addgene
*RPB1*GPD415-*LEU2*	*RPB1*-TEV-CTD-*LEU2*	This work
*RPB1*GPD413-*HIS3*	*RPB1*-TEV-CTD-*HIS3*	This work
pBS159 with 3C-TAP	pBS159 with 3C-TAP	Mosley lab
Yeast Strain	**Description**	**Source or reference**
BY4741	*MATa HIS3-1 LEU2-0 met1-50 URA3-0*	Open Biosystems
RPB1 GPD URA3	*MATa HIS3-1 LEU2-0 met1-50 URA3-0 [pAG416-URA3-RPB1] rpb1::CLONAT*	Mosley lab
Rpb1-TEV-CTD *HIS3*	*MATa HIS3Δ1LEU2 Δ 0 met15Δ0 URA3Δ0 [p RPB1-TEV-CTD-HIS3] rpb1::CLONAT*	This work
Rpb1-TEV-CTD *LEU2*	*MATa HIS3Δ1LEU2 Δ 0 met15Δ0 URA3Δ0 [p RPB1-TEV-CTD-LEU2] rpb1::CLONAT*	This work
Rpb3-3C-TAP Rpb1-TEV-CTD	*MATa HIS3Δ1LEU2 Δ 0 met15Δ0 URA3Δ0*	This work

[p RPB1-TEV-CTD-LEU2] Rpb3::3C-TAP rpb1::CLONAT

IV. Tandem Affinity Purification (TAP) of Rpb3-3C-TAP Rpb1-TEV-CTD

Yeast Cell extract

Rpb3-3C-TAP Rpb1-TEV-CTD strain (Table 4) was cultured in 6 liters of YPD media and grown at 30 °C to OD_{600} ~ 2.5. Cell pellets were collected by centrifugation at 4°C, washed with sterile water, and re-suspended with TAP lysis buffer (1 to 10 of cells to lysis buffer). The cell extract mixture (250 ml) was transferred to stainless steel bead beating chamber, and then glass beads were added to the chamber until the mixture reached to the top of the chamber. Cell extract was lysed by bead beating in the stainless steel chamber surrounded by an ice bath in one minute on and off cycles for 30 minutes. Subsequently, the cell lysate was treated with 100 units of DNAse and 300 µg of heparin sulfate for 10 minutes to solubilize chromatin (Mosley et al., 2009). After solubilzation, cell lysate was cleared by centrifugation for an hour at 4°C, and then the supernatant was transferred to a 100 ml glass beaker to proceed with affinity purification.

Purification I

Purification I was conducting following the steps illustrated in Figure 4A. An aliquot of IgG beads (500 µl) was washed with TAP lysis buffer and then the beads were re-suspended with 500 µl of TAP lysis buffer. The 500 µl of IgG beads were added to the cell lysate and incubated overnight at 4°C on stir plate. In the following day, cell lysate was transferred to a 30 ml Bio-Rad Econoprep column and drained by gravity flow. The beads were washed with 30 ml of TAP lysis buffer followed by washing with 3 ml of

TEV cleavage buffer. Subsequently, IgG beads were re-suspended in 1 ml TEV cleavage buffer and transferred to a new 1.5 ml centrifuge tube. The beads were treated with 10 µl of AcTEV protease (Invitrogen) and incubated overnight at 4°C on the rotating wheel. After overnight cleavage, the cell suspension-beads mixture was transferred to the same column to elute TEV cleaved product by gravity flow, TEV elution presumably has CTD and its associated proteins. The retained beads in the column were washed with 5 ml of 3C cleavage buffer. After washing step, the beads were re-suspended with 1 ml of 3C cleavage buffer and were transferred to 1.5 ml centrifuge tube. The beads were treated with 10 µl of Prescission 3C protease (GE Healthcare) and incubated over night at 4 °C on the rotating wheel. Following overnight cleavage, the cell-beads mixture was transferred into the same column and the 3C elution was collected in 15 ml conical tube. The beads were washed with 3 ml of calmodulin binding buffer and the flow through was collected into the same 15 ml conical tube followed by the addition of 3 µl of 1M $CaCl_2$ to 3C elution in the 15 ml conical tube to allow the binding of CBP protein to the beads. An aliquot of calmodulin sepharose resins (500 µl) was washed with calmodulin binding buffer and were re-suspended with 500 µl of calmodulin binding buffer. Calmodulin sepharose (500 µl, Agilent) was added to 3C elution and incubated for 5 hours at 4°C on the rotating wheel. Following the incubation, 3C elution-resins mixture was transferred to a new Bio-Rad 30 ml Econoprep column and the flow through was drained by gravity flow. After washing the column with 30 ml of calmodulin binding buffer, RNAPII subunits and the associated proteins were eluted by incubated the column for 5 minutes with 500 µl of calmodulin elution buffer with EGTA to chelate $CaCl_2$, then drained by

gravity flow. The elutions were collected in 1.5 ml centrifuge tube, the step was repeated for 6 times to collect 6 elutions numbered E1-E6.

Silver Nitrate Stain Analysis of Purification I Calmodulin Elutions

Prior to the silver nitrate stained SDS-PAGE analysis, calmodulin elutions (E1-E6) obtained in purification I were TCA precipitation in order to concentrate the proteins (see section V for detailed TCA precipitation procedure). The precipitated samples were treated with 8 M urea in 100mM Tris-HCl and then 5% aliquot of each urea treated samples was mixed with 2x loading buffer and then boiled for 5 minutes at 100 °C. The samples were loaded onto pre-cast SDS-PAGE gel (Bio-Rad) and the gel was run in 1x Tris-Gly-SDS buffer at 200 V until the dye front reached the bottom of the gel. Following SDS-PAGE separation, the gel was incubated with 100 ml fixing solution overnight to improve the sensitivity of the staining and to restrict the movement of the proteins out of the gel. In the next day, the gel was washed with 100 ml methanol wash for 10 minutes followed by washing with 100 ml MilliQ water for 10 minutes. After water wash, the gel was incubated with 100 ml of sensitizer solution for 10 minutes and then washed with 100 ml MilliQ water for 10 minutes. Succeeding water wash, the gel was incubated with 100 ml of silver nitrate solution for 10 minutes followed by washing with 100 ml MilliQ water for 5 minutes. One hundred milliliters of developer solution was added to the gel and the developing reaction was stopped with 100 ml of the stopping solution when the protein bands reached the desired intensity. The developed gel was imaged using Epson perfection V700 photo scanner.

Purification II

Purification II was conducting following the steps illustrated in Figure 4B. In this purification, the clarified cell extract was treated with 500 µl of IgG beads and incubated overnight at 4°C on stir plate. In the following day, cell-beads mixture was transferred to a 30 ml Bio-Rad Econoprep column and was drained by gravity flow. After extensive washing with TAP lysis buffer, the beads were washed with 3 ml of 3C cleavage buffer. Subsequently, IgG beads were re-suspended in 1 ml 3C cleavage buffer, transferred to 1.5 ml centrifuge tube and then they were treated with 10 µl of prescission 3C protease and incubated over night at 4 °C on the rotating wheel. After overnight cleavage, the mixture was transferred to the same column and the 3C cleaved product was allowed to drain by gravity flow and was collected in 15 ml conical tube. The beads remaining in the column were washed with 4 ml of calmodulin binding buffer and the flow through was collected in the same 15 ml conical tube. After the addition of 3 µl of 1M $CaCl_2$ to 3C elution in the 15 ml conical tube, a 500 µl aliquot of calmodulin sepharose resins was washed with calmodulin binding buffer and was re-suspended with 500 µl of calmodulin binding buffer. Subsequently, the calmodulin sepharose resins were added to 3C elution and incubated overnight at 4°C on the rotating wheel. Following overnight incubation, the 3C elution-calmodulin resins mixture was transferred to a new 30 ml Bio-Rad Econoprep column and the buffer was drained by gravity flow. After extensive wash of the resins with calmodulin binding buffer, they were re-suspended with 1 ml of CTD cleavage buffer and the mixture was transferred to 1.5 ml centrifuge tube. And then the mixture was treated with 10 µl of AcTEV protease (Invitrogen) and incubated for overnight cleavage at 4 °C. In the next day, the mixture was returned to the same column and TEV cleaved product was allowed to drain by gravity and was collected in separate 1.5 ml

centrifuge tube, the resins were washed with 1 ml of calmodulin binding buffer and the wash was collected in another centrifuge tube. TEV elutions were stored in -80 ° prior MudPIT analysis. The remaining resins in the column were washed with 30 ml of calmodulin binding buffer and then RNAPII subunits and the associated proteins were eluted from the column using calmodulin elution buffer with EGTA to chelate $CaCl_2$. The elution was performed as following: the resins in the column were incubated for 5 minutes with 500 µl of calmodulin elution buffer, the elution was drained by gravity flow and collected in 1.5 ml centrifuge tube. The elution step was repeated for 5 times to collect 5 elutions numbered E1-E5.

Silver Nitrate Stain Analysis of Purification II Calmodulin Elutions and CTD Flow-through

SDS-PAGE silver stained procedure was performed in order to detect the proteins purified in purification II and the CTD elutions from both purifications. Concisely, an aliquot (20 µl) of calmodulin elutions (E1-E5) obtained in purification II, CTD elutions obtained from both purification was mixed each with 20 µl of 2x loading buffer and then denatured for 5 minutes at 100 °C. After denaturation, the samples were loaded onto pre-cast SDS-PAGE gel (Bio-Rad) and the gel was run in 1x Tris-Gly-SDS buffer at 100 V for an hour. Subsequently, the gel was fixed with fixing solution (100 ml) overnight. After extensive wash of the gel with methanol and MilliQ water wash (10 minutes each), the gel was incubated with sensitizer solution (100 ml) for 10 minutes followed by another 10 minutes round of MilliQ water wash. The gel was treated with 100 ml of silver nitrate solution for 10 minutes followed by gentle wash with MilliQ water for 5 minutes. The gel was then transferred to the developing solution (100 ml) and the

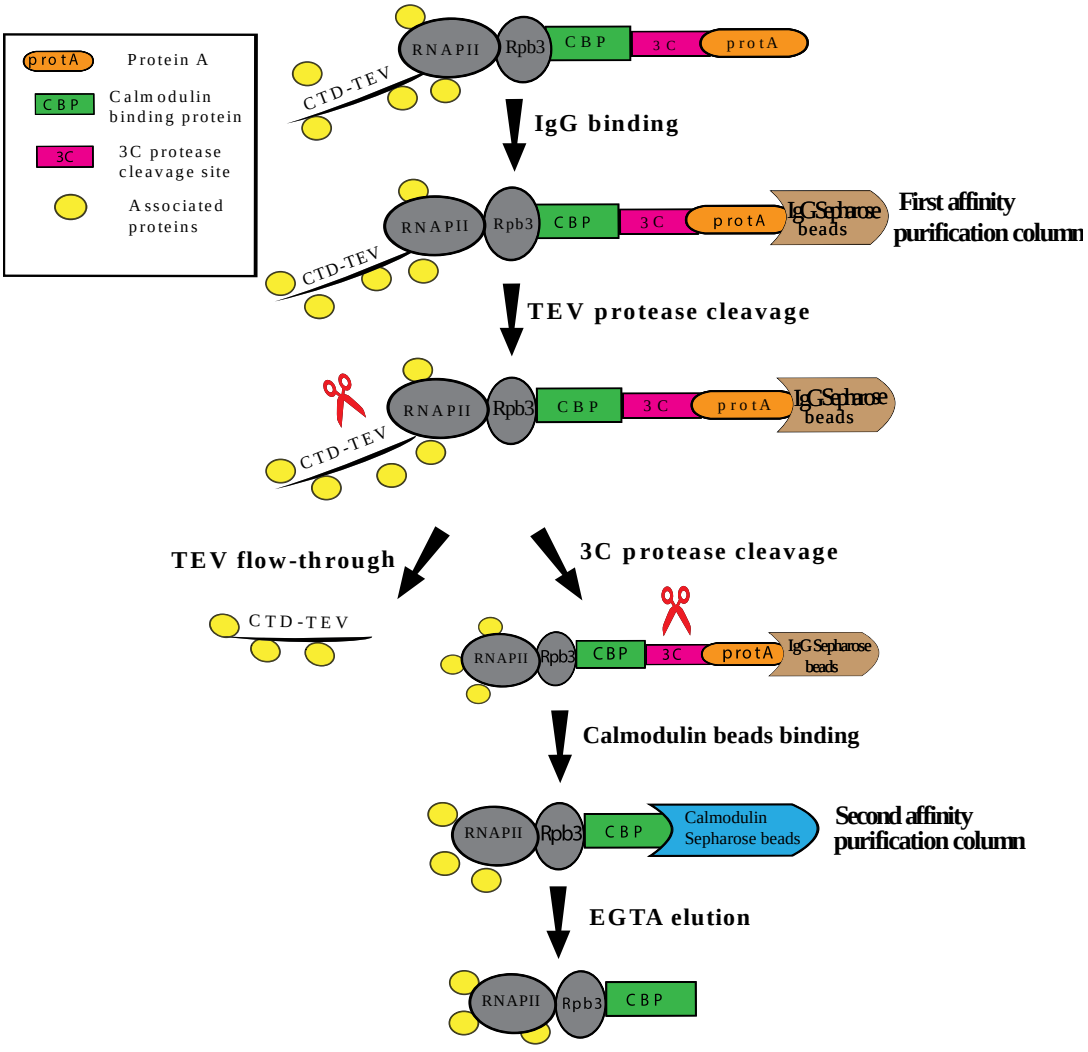
staining was stopped with 100 ml of stopping solution after the protein bands reached the desired intensity. The proteins bands were imaged using Epson perfection V700 photo scanner.

Western-Blot to Examine the CTD Cleavage

To confirm that TEV protease cleaved CTD from RNAPII, western-blot analysis was performed using calmodulin elutions 1 to 5 (Purification II), CTD elutions obtained from both purifications. Briefly, an aliquot of each sample was mixed with same volume of 2x Laemmli loading buffer with BME (Bio-Rad); the mixture was denatured at 100 °C for five minutes. The denatured samples were loaded onto 12% SDS polyacrylamide gel alongside 5 µl of Precision Plus Protein Dual Color Standard molecular weight marker (Bio-Rad). The proteins were separated by electrophoresis in 1x Tris-Gly-SDS (TGS) buffer at 200 V and then they were transferred from the gel to nitrocellulose membrane at 30 V overnight. Following overnight transferring, the nitrocellulose membrane was blocked with 5% milk in 1x TBS buffer for 45 minutes, and then the membrane was probed overnight at 4 °C with anti- mouse-H14 monoclonal primary antibody which recognizes the phosphorylated serine 5 version of RNAPII (Covance, 1:1000 dilution). The next day, the membrane was washed three times with 1x TBS buffer (10 minutes each) followed by incubation for an hour at room temperature with anti-mouse-IgM secondary antibody (Sigma, 1:5000 dilution). After incubation with secondary antibody the membrane was washed three times with 1x TBS buffer (10 minutes each). Subsequently, the proteins were visualized using ECL (GE Healthcare) and were imaged using Fujfilm imager FLA-5000 at 473 nm and the LBP blue filter.

In order to detect the cleavage of 3C-TAP tag in the samples, western-blot procedure conducted above was repeated using same set of the samples except probing the samples with anti-TAP polyclonal primary antibody (Thermo, 1:1000 dilution) and anti-rabbit IgG horseradish peroxidase secondary antibody (GE Healthcare, 1:5000 dilution).

A



B

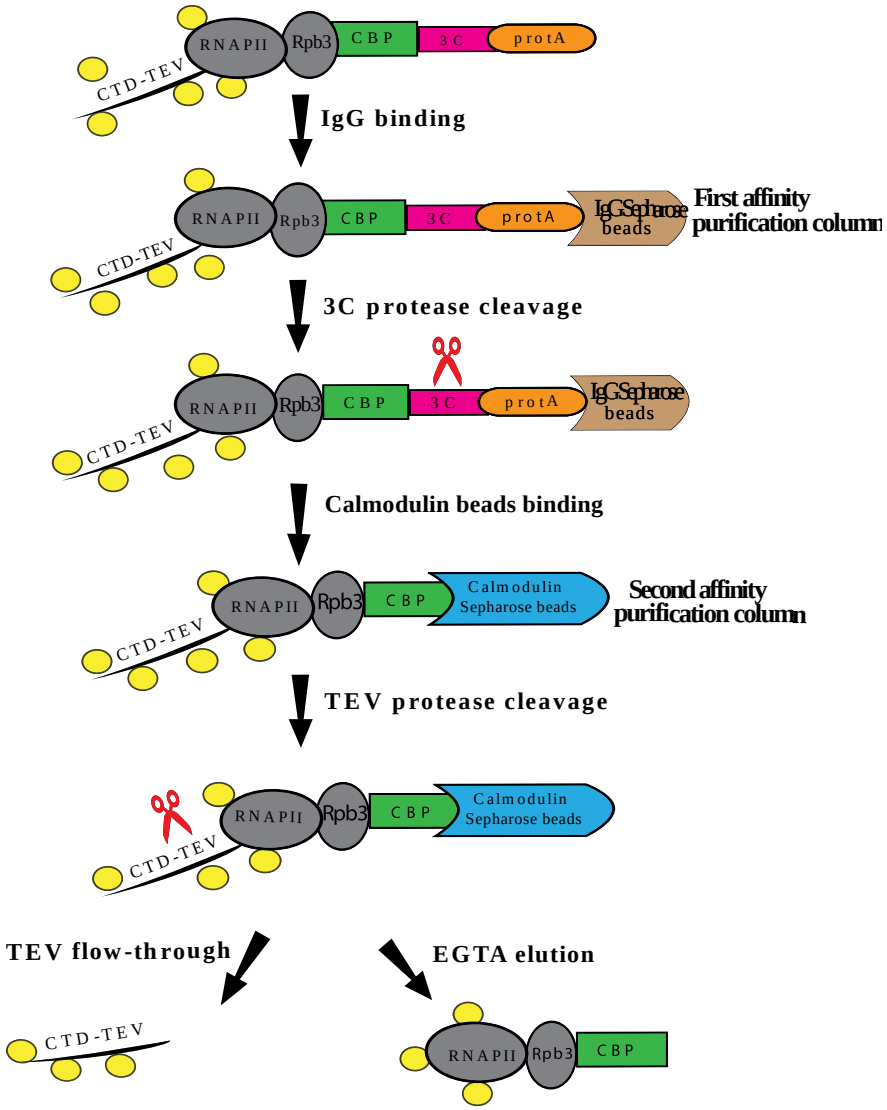


Figure 4. A schematic representation of the modified TAP purification used to isolate the CTD of RNAPII. A. An overview of purification I scheme: in this purification CTD was cleaved on-bead by TEV protease while RNAPII complex was immobilized on IgG beads. B. An overview of purification II scheme: in this purification CTD was cleaved on-bead while RNAPII complex was immobilized on calmodulin beads.

V. MudPIT Analysis

The MudPIT protocol performed in this work (Figure 5) was based on the original MudPIT protocol (Florens and Washburn, 2006).

Preparation of the samples

Two hundreds microliters of the elutions obtained from the purifications performed in the previous section was mixed each with 200 µl of 100mM Tris-HCl (pH 8) and 100 µl of trichloroacetic acid. Following vortex mixing, the samples were incubated overnight at 4 °C for TCA precipitation. The following day, the samples were centrifuged for 30 minutes at 4 °C at 14,000 RPM. The supernatant was discarded and the pellets were washed with acetone cold (500 µl) at 4 °C at 14000 rpm, acetone wash was repeated twice. After the wash, the acetone wash was removed and any remaining acetone was let to evaporate in speed vac. Subsequently, 8 M urea in 100mM Tris-HCl and 5mM of TCEP was added to the precipitated proteins for protein denaturation and reduction, and then the samples were incubated for 30 minutes at room temperature. Following protein denaturation and reduction, 10mM chloroacetamide (CAM) was added to the samples for protein alkylation and the samples were incubated in dark for 30 minutes at room temperature. Following the incubation, the denatured, reduced and carboxy methylated samples were treated with 0.3 µg of endoproteinase LysC and incubated overnight at 37 °C with shaking. After overnight digestion, 100mM Tris-HCl (pH 8) was added to the samples to dilute the urea to 2 M followed by addition of 2mM of $CaCl_2$. 0.5 µg trypsin cold was added to the samples and then the samples were incubated at 37 °C with shaking for overnight digestion. In the next day, the digestion reaction was quenched with 5% formic acid and the samples were stored at -80 prior to MudPIT analysis.

Microcapillary Column Packing and Sample Loading

Microcapillary column was packed in three-phase using pressurization vessel as previously described (Florens and Washburn, 2006; McDonald et al., 2002), the column was packed in the following order: 8 cm Aqua C18 reverse phase (RP) (Phenomenex), followed by 2.5 cm strong cation exchange (SCX)(Phenomenex), and finally with 2 cm reverse phase (RP). Following washing the packed column with methanol for 10 minutes, the column was equilibrated with buffer A for 30 minutes. Prior to loading the sample in the column, the sample was centrifuged at 14,000 for 30 minutes to clarify the sample from any particles. And then the sample was loaded overnight into the column. The following day, the column was washed with buffer A until it was transferred to a Nanoflex nanospray source in-line with a Velos Pro ion trap mass spectrometer (Thermo). This process was repeated for all the samples obtained from the two purifications described in the previous section.

High Pressure Liquid Chromatography, Tandem Mass Spectrometry and Database Search

The packed, loaded and washed column was placed in-line with prexeon nano-liquid chromatography. And then the sample was subjected to 10-MudPIT analysis as previously described (Washburn et al., 2001). Subsequently, the chromatography data was eluted in LTQ- Velos Pro ion trap for mass spectrometer analysis. Following mass spectrometer analysis, the tandem spectra (MS/MS) were searched in SEQUEST algorithm using Proteome Discoverer 1.3 version (Thermo). In SEQUEST search engine the following parameters were selected: MS/MS tolerance 0.8 Da, peptide tolerance 1.4 dalton, static modifications +57 dalton on cysteine to account for carbamidomethylation

from CAM (alkylation), and dynamic modifications + 16 dalton on methionine to account for oxidation. The spectra were searched against FASTA database containing 5819 *S. cerevisiae* protein sequences from National Center for Biotechnology Information (NBCI) and 281 common contaminated proteins including human keratins, immunoglobins and proteolytic enzymes. In the data analysis, high confidence peptide was chosen to calculate for false discovery rate (FDR), which was kept equal to 1% or less in all the analysis performed in this work.

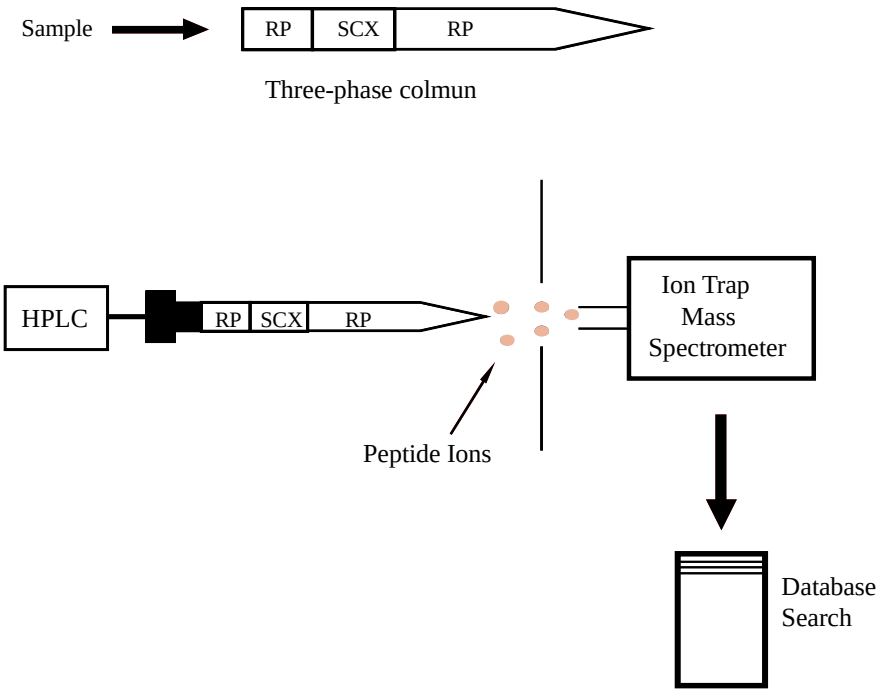


Figure 5. A schematic representation of MudPIT analysis. In this protocol: proteins are subjected to enzymatic digestion and then the sample is loaded into three-phase column (RP/SCX/RP), the column is transferred to LC/MS, proteins are fractionated and sprayed into mass spectrometer, spectra are recorded and then are searched against FASTA database.

RESULTS

I. Construction of Rpb1-TEV-CTD Plasmids

The C-terminal domain (CTD) of the largest subunit of RNAPII acts as a platform in cooperating the binding of transcription factors and accessories proteins involved in transcription, mRNA processing and chromatin modification. In order to develop a method to selectively isolate the CTD of RNAPII and its binding proteins for proteomics analysis, we inserted the seven amino acids of TEV protease cleavage site (ENLYFGQG) into the unstructured linker region of Rpb1 gene using a modified QuickChange site-directed mutagenesis PCR-based method. QuickChange II XL site- directed mutagenesis PCR-based method developed by Agilent Technologies (Santa Clara, CA) works by utilizing supercoiled DNA plasmid contains gene of interest as a template and pair of complementary primers with the desired mutation (TEV site insert in our experiment). Annealing of the mutant primers to DNA template and the extending of the mutant primers by *PfuUltra* high fidelity DNA polymerase produces double strand mutant PCR product with a staggered nicks (Figure 6B). The PCR product is then treated with *Dpn* 1 endonuclease to selectively digest the methylated (parental DNA template) and hemi-methylated (parental-mutant template) to increase transformation efficiency. Subsequently, the un-methylated mutant nicked PCR product is transformed into ultracompetent cells for nick ligation (Figure 6B). Despite that QuickChange PCR- based mutagenesis method is a powerful tool in modifying gene sequences in molecular biology studies, the procedure has some limitations that could impact the mutagenesis efficiency for example the method is restricted with using complementary synthetic oligonucleotide

primers of 25-45 bases in length with melting temperature of $T_m \geq 78°C$ and the mutation should be in the middle of the primers. Using complementary primers in PCR-based mutagenesis reaction could result in a low mutation efficiency especially when introducing multiple or large mutations because the probability for primers to overlap and self- annealing is high, therefore primer-dimer formation becomes more favorable than primer-template formation (Wang and Malcolm, 1999).

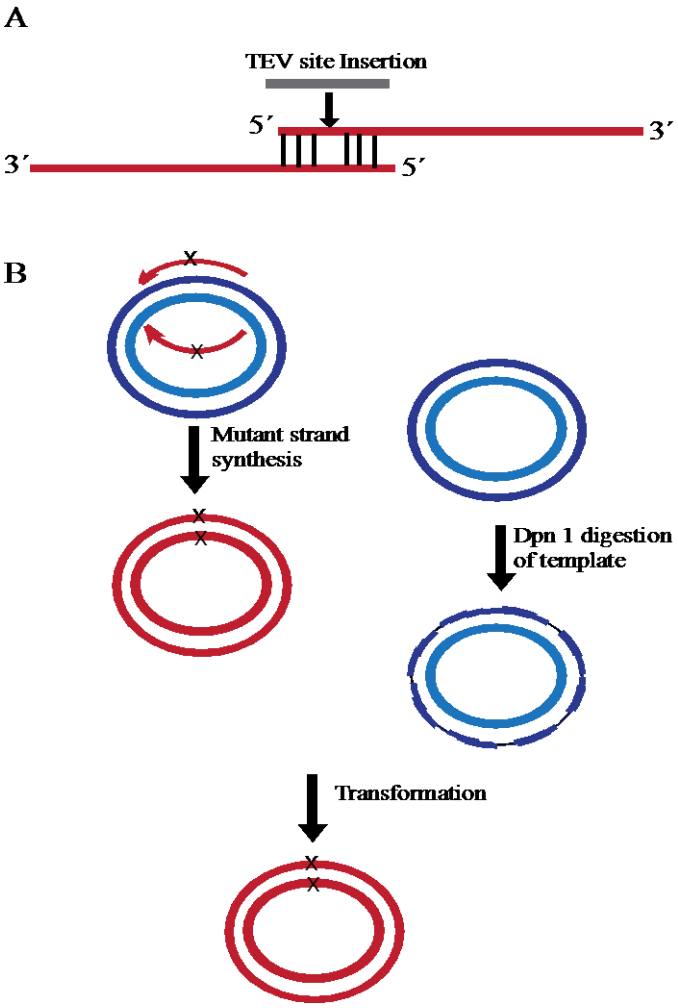


Figure 6. An overview of QuickChange site-directed mutagenesis PCR-based method. A. A diagram of the new primer pair design, TEV insertion is placed in the middle of 5′ complementary region. B. A scheme of QuickChange three steps: synthesis of mutant strands by PCR, digestion of parental templates by *Dpn*1, and transformation of mutant plasmids into ultracompetent cells.

In the beginning of our TEV insertion experiment following QuickChange site-directed mutagenesis protocol we encountered problems in inserting the TEV cleavage site (21 bases in length) (Figure 7A) into *RPB1-HIS3* template (size = 12019 bases) due to several reasons such as the large size of DNA template we initially used in the reaction and the difficulties of satisfying QuickChange primers' requirements described above for instance: primers length and melting temperature. After performing several unsuccessful troubleshooting attempts of increasing DNA concentrations and optimizing PCR thermal cycle parameters, we eventually were able to generate *RPB1*-TEV construct following a modified QuickChange method developed by Liu and Naismith, 2008. In this method, Liu and Naismith designed a primer pair with complementary (overlapping) sequences at 5′ end and larger non-overlapping sequences at 3′ end. They constructed various mutations including insertion of 18 bases and deletions of 21 bases using their new primer design and QuickChange PCR-based mutagenesis method (Liu and Naismith, 2008). Based on Liu and Naismith developed method, we designed a primer pair with complementary sequences at 5′ end and larger non-overlapping sequences at 3′ end, TEV insertion site was placed in the middle of the complementary region (Figure 6A). Using the newly designed primers (T_m 71.9°C) and a smaller DNA template (*RPB1*-template: 7801 bases) we were able to introduce TEV site upstream of CTD with high efficiency; all the analyzed colonies had the mutation. The main advantage of using this method was reducing the probability of primers self-annealing because the region of non-overlapping end of the new primers is long enough to reach and bind to the nick in the newly synthesized DNA therefore use it as a template (Liu and Naismith, 2008). In addition, the method eliminated the problem associated with introducing the mutation at low GC

region. Furthermore, *RPB1*-TEV construct was cloned into vector with selective markers using LR reaction.

To test the optimal location of CTD cleavage, we generated *RPB1*-TEV construct, which has the TEV site at amino acid 1460 into Rpb1 linker region (Figure 7C)

TEV and 3C proteases recognize and cleave distinct but closely related sequences (See discussion). Therefore to examine the cleavage specificity of both proteases at their cleavage sites in our developed method, we generated *RPB1*-3C-CTD construct (Figure 7D).

Please note that only *RPB1*-TEV-CTD construct with TEV insertion at amino acid 1511 was used to conduct the work presented here including: construction of yeast strains, tagging the strain with Rpb3-3C-TAP, TAP purification and MudPIT analysis. The construct with TEV insertion at amino acid 1460 was used by students at DePauw University to generate wild type and mutant yeast strains. The *RPB1*-3C-CTD construct will be used for future analysis by Mosley laboratory.

A TEV protease cleavage site

GAGAATTTGTATTTCAGGGT

B

```
RPB1_wt          AATGACGCTATGGCTGGAGGATTTACAGCGTACG--------------------GTGGT
Rpb1_TEV-1_E     AATGACGCTATGGCTGGAGGATTTACAGCGTACGAGAATTTGTATTTTCAGGGTGGTGGT
                 *********************************                  *****

RPB1_wt          GCTGATTATGGTGAAGCCACGTCTCCATTTGGTGCTTATGGTGAAGCACCTACATCTCCC
Rpb1_TEV-1_E     GCTGATTATGGTGAAGCCACGTCTCCATTTGGTGCTTATGGTGAAGCACCTACATCTCCC
                 ************************************************************
```

C

```
RPB1_wt          TTTGATGTGATGATCGATGAGGAGTCACTGGTAAAATACATGCCAGAACAAAAAATA---
Rpb1_TEV_P1      TTTGATGTGATGATCGATGAGGAGTCACTGGTAAAATACATGCCAGAACAAAAAATAGAG
                 ************************************************************

RPB1_wt          ------------------ACTGAGATTGAAGACGGACAAGATGGTGGCGTCACACCATAC
Rpb1_TEV_P1      AATTTGTATTTTCAGGGTACTGAGATTGAAGACGGACAAGATGGTGGCGTCACACCATAC
                                    ****************************************
```

D 3C protease cleavage site

TTGGAGGTTTTCAGGGACCT

E

```
RPB1_wt          TTTGATGTGATGATCGATGAGGAGTCACTGGTAAAATACATGCCAGAACAAAAAATA---
Rpb1_3C_CTD_2    TTTGATGTGATGATCGATGAGGAGTCACTGGTAAAATACATGCCAGAACAAAAAATATTG
                 ************************************************************

RPB1_wt          ---------------------ACTGAGATTGAAGACGGACAAGATGGTGGCGTCACACCA
Rpb1_3C_CTD_2    GAGGTTTTGTTTCAGGGACCTACTGAGATTGAAGACGGACAAGATGGTGGCGTCACACCA
                                       ***************************************
```

Figure 7. Construction of *RPB1*-TEV-CTD plasmids. A. A diagram showing TEV cleavage site gene sequence. B and C. ClustalW alignments of the WT *RPB1* and *RPB1*-TEV-CTD plasmid sequencing data (performed by Genewiz) showing the TEV cleavage site insertion at amino acid 1511 and 1460 of *RPB1* respectively. D. A diagram showing gene sequence of 3C cleavage site. E. ClustalW alignment of the WT *RPB1* and *RPB1*-TEV-CTD plasmid sequencing data (performed by Genewiz) showing 3C site insertion at amino acid 1460 of Rpb1.

II. **Construction of Yeast Strains**

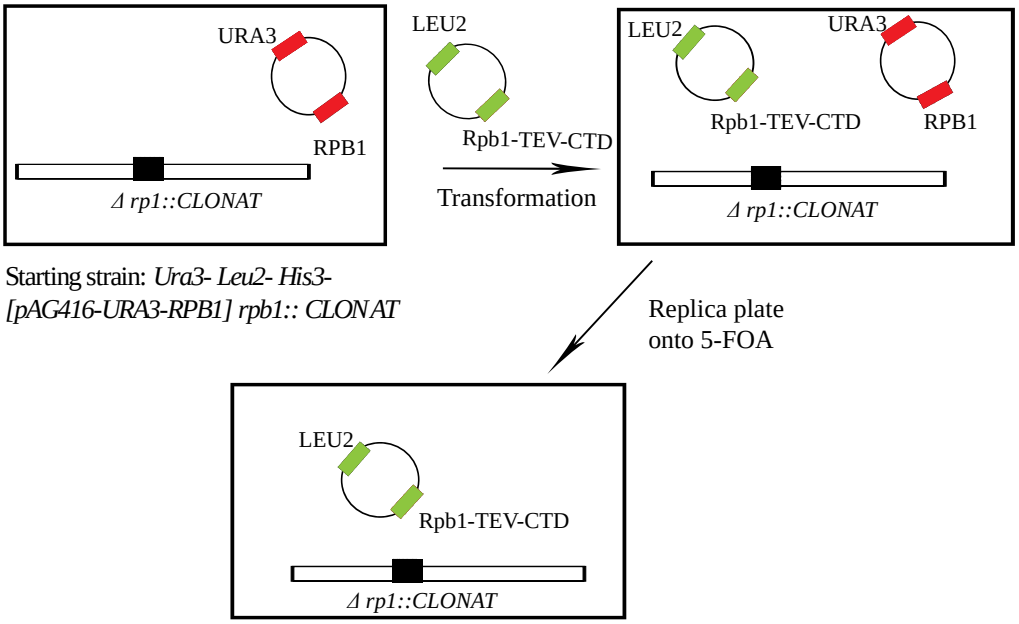
The next approach in our developed method was to construct yeast strain bearing the modified copy of *RPB1* gene (*RPB1*-TEV-CTD). In order to accomplish that, three

experiments were performed. In the first experiment, *RPB1* open reading frame in BY4741 strain (Table 4) was deleted with cassette carrying *CLONAT* marker. The cassette was obtained by PCR amplification method from the pAG25 plasmid with primers had 65 nucleotide homology regions upstream and downstream of RPB1 gene. PCR product was transformed into BY4741 strain (Table 4) to replace the native *RPB1* gene with *CLONAT* cassette by homologous recombination. Since *RPB1* is an essential gene, BY4741 cells were transformed with a *RPB1* containing plasmid driven by a glyceraldehyde 3-phosphate dehydrogenase (GPD) promoter that also contained a *URA3* selectable marker (Table 4). Following transformation with the *CLONAT* cassette, *rpb1::CLONAT* mutant strains were selected on YNB Ura-CLONAT+ plate.

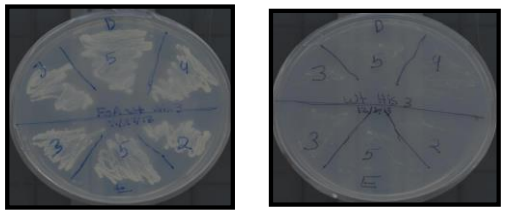
In the second experiment, Rpb1-TEV construct was transformed into *RPB1 GPD URA3* by lithium acetate transformation. The strains were selected on YNB-Leu^- Ura^- plate or YNB-His^- Ura^- plate. The third experiment was performing plasmid shuffle using 5-FOA. 5-FOA is a pyrimidine analogue drug that is commonly used for selection of strains borne mutation in *URA3* gene. In the cell, 5-FOA is converted by orotate phosphoribosyl transferase into 5-fluorotidine monophosphate (5-FOMP). And then, 5-FOMP is further converted by 5-monophosphate decarboxylase (encoded by *URA3* gene) to 5-fluorouridine monophosphate (5-FUMP) (Boeke et al., 1984, 1987). 5-FUMP is a toxic metabolite that inhibits thymidylate synthase thus blocking pyrimidine synthesis and causing cell death (Boeke et al., 1984, 1987). Therefore, mutant cells that lack *URA3* gene are unable to convert 5-FOA to its toxic form and hence are resistant to the drug whereas wild type cells are not resistant. In our experiment, we used 5-FOA to select for the plasmid harboring *RPB1*-TEV-CTD construct and evict the WT *RPB1 URA3* plasmid

(see Figure 8A for the plasmid shuffle scheme). Furthermore, colonies that only were able to grow on 5-FOA (Figure 8B) were isolated and used in the preceding section.

A



B



Replica plate onto FOA plate YNB Ura- His- plate

Figure 8. Construction of yeast strains. A. A schematic representation of plasmid shuffle selection for *RPB1*-TEV-CTD strains. B. Growth of the colonies on 0.1% FOA replica plate but not on YNB Ura-His- plate.

III. **Tagging Yeast Strains With Rpb3-3C-TAP Tag**

Plasmid construct containing a 3C-TAP was amplified using PCR method with primers had 65 nucleotide homology upstream of the stop codon of Rpb3 and just downstream of the Rpb3 gene (this will remove the stop codon in Rpb3 allowing for fusion with the 3C-TAP tag). The 3C-TAP tag is consisting of: Calmodulin binding peptide (CBP), Protein A, and 3C cleavage site replacing TEV cleavage site. Agarose gel electrophoresis of the amplified DNA is shown in Figure 9A. Subsequently, lithium acetate yeast transformation method was performed to integrate the PCR product into the genome of *RPB1*-TEV-CTD *HIS3* and *RPB1*-TEV-CTD *LEU2* strains in order to replace RPB3 wild type gene. Following several unsuccessful troubleshooting attempts of yeast transformation, we were able to integrate the PCR product into the strain and obtained colonies grown in one of the selective plates. To confirm the presence of 3C-TAP tag, chosen colonies from the plate were analyzed by western-blot alongside Rpb3-TAP strain as a positive control and BY4741 strain as a negative control using polyclonal antibody against TAP tag (Figure 9B). Western-blot analysis confirmed the tagging of our constructed strain as shown in Figure 9B.

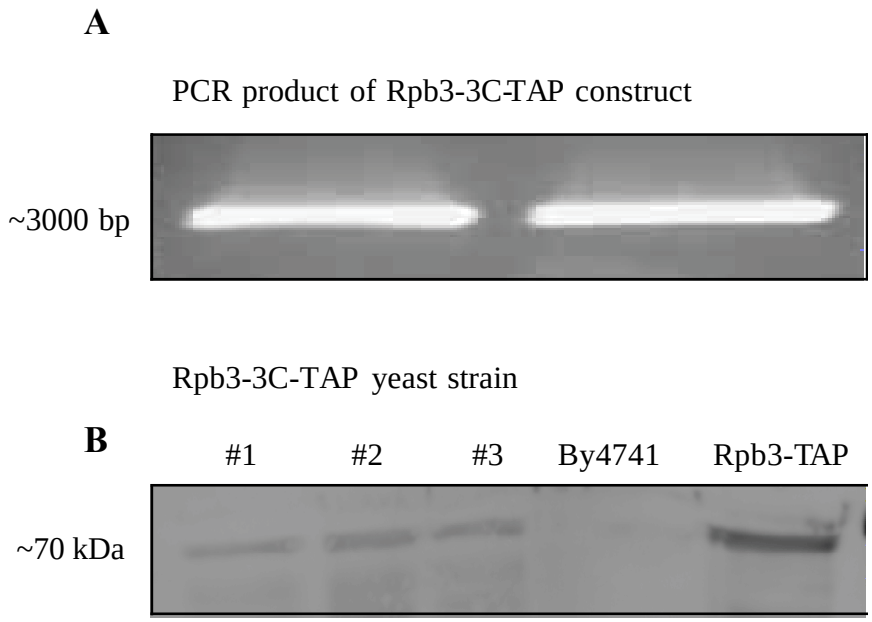


Figure 9. Tagging yeast strains with Rpb3-3C-TAP. A. DNA image of PCR product around 3000 bp. B. Western-blot analysis confirms the presence of the 3C-Tap tag: Rpb3-3C-TAP Rpb1-TEV-CTD and Rpb3-TAP (positive control) were detected around 70 kDa.

IV. TAP Purification of Rpb3-3C-TAP *RPB1*-TEV-CTD

In our modified TAP purification method for isolating the CTD of Rpb1 and its associated proteins we modified the original TAP purification protocol (Rigaut et al., 1999; Puig et al., 2001). Consequently, we performed purification I as presented in Figure 4A scheme. In this purification, cell extract was subjected to first affinity column where protein A in TAP tag bound to IgG resins thus immobilizing RNAPII and the associated proteins on the beads. And then, we performed on-bead CTD cleavage with TEV protease while the RNAPII subunits were immobilized on IgG resins. Subsequently, the CTD-less RNAPII subunits were released from IgG through incubation and cleavage with 3C protease. The CTD-less RNAPII subunits and their associated proteins were then

subjected to second affinity column where calmodulin-binding peptide (CBP) in the tag bound to calmodulin sepharose in the presence of calcium ions. Then, EGTA was used to elute CTD-less RNAPII subunits and their associated proteins from the calmodulin sepharose.

Calmodulin elutions (E1-E6) obtained in purification I were subjected to TCA precipitation in order to concentrate the proteins and then were analyzed by silver stained SDS-PAGE gel to detect the proteins and to evaluate the purification yield. Silver staining analysis showed that elutions E1 and E2 had the most recovery (Figure 10); the remaining elutions E3-6 had low recovery yield (data not shown). Additionally, all the 12 subunits of RNAPII were detected as shown in Figure 10, which was determined based on previous silver staining analysis of RNAPII (Mosley et al., 2011). Moreover, Rpb1 exists in two forms: hyper-phosphorylated CTD and hypo-phosphorylated CTD, thus, migrate differently on SDS-PAGE gel as previously described (Dubois et al., 1997; Svejstrup et al., 1997). However, only one form of Rpb1 was detected in silver stained gel around 170 kDa (Figure 10). We speculated that the absence of the second Rpb1 form could be resulted from the loss of the CTD.

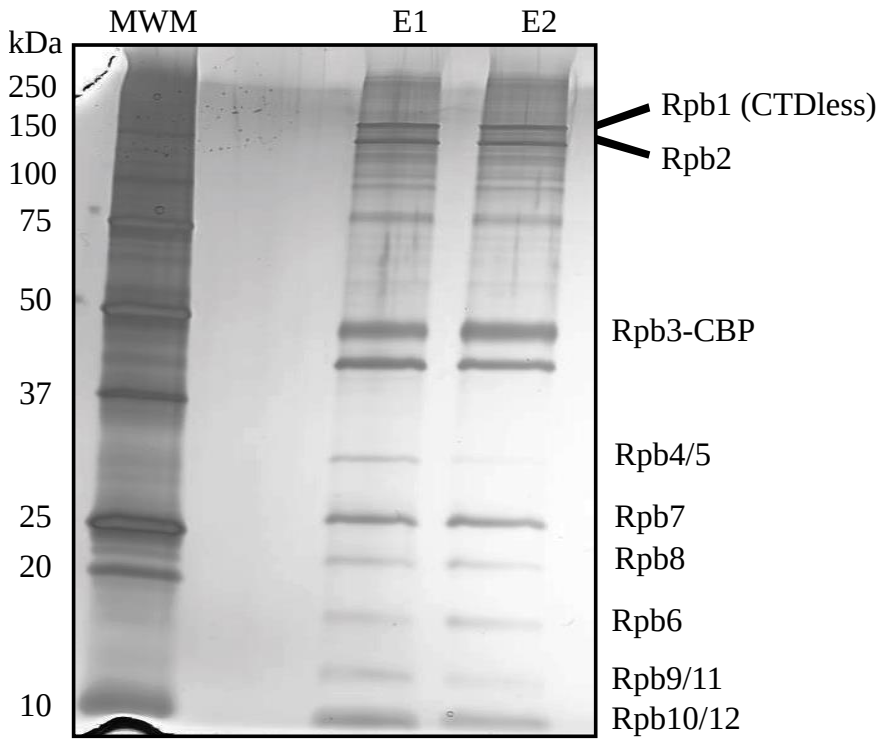


Figure 10. Silver staining SDS-PAGE analysis of purification I. The gel showing the detection of RNAPII 12 subunits in calmodulin elutions as indicated in the left of the gel. Molecular weight marker (MWM) was loaded in the first lane to the right side of the gel.

CTD elution collected post TEV protease cleavage was analyzed in LC/MS reverse phase. The preliminary data we obtained from LC/MS analysis showed that the CTD sample was contaminated with RNAPII all twelve subunits (Table 5). These data indicated that TEV protease partially cleaved 3C site thus releasing part of RNAPII subunits from the IgG resins.

Table 5. RNAPII 12 subunits were identified in CTD elution by LC/MS.

Accession	Description	Sequence Coverage %	# Peptide	# PSMS	# AAs	MW
6320061	Rpb1	27.35	37	197	1733	191.5
398365417	Rpb2	24.02	8	119	1224	138.7
398364469	Rpb3	56.92	12	43	318	40.3
6322321	Rpb4	50.68	9	34	221	25.4
398366567	Rpb7	52.63	6	24	171	19.0
6324798	Rpb8	43.84	5	23	146	16.5
6325445	Rpb6	39.35	4	14	155	17.9
6321368	Rpb9	38.52	2	14	144	14.3
6319630	Rpb5	27.91	6	13	215	25.1
398365775	Rpb10	54.29	3	8	144	14.3
6324569	Rpb11	42.50	3	6	120	13.6
6321937	Rpb12	28.57	1	5	70	7.7

In order to prevent the cleavage of 3C site by TEV protease, we modified our TAP purification protocol as shown in Figure 4B scheme (purification II). In this purification, we initially performed 3C protease cleavage and the releasing of RNAPII and the associated proteins from the IgG resins. And then, RNAPII subunits and the associated proteins were subjected to a second affinity column followed by on bead TEV protease cleavage while the RNAPII subunits were immobilized on calmodulin sepharose. Subsequently, the CTD-less RNAPII subunits were released from the calmodulin sepharose with 2mM EGTA.

Furthermore, calmodulin elutions (E1-E5) obtained from purification II CTD elutions obtained from both purifications were analyzed in silver stained SDS-PAGE to

evaluate the purification yield and to examine the CTD cleavage (Figure 11). Although, the purification yield was lower than purification I yield, similar pattern of RNAPII subunits were detected in this purification as in the previous purification, for example only one form of Rpb1 was detected around 170 kDa corresponded with purification I. In addition, protein around 30 kDa was detected in CTD elutions but not in calmodulin elutions and would likely be TEV protease.

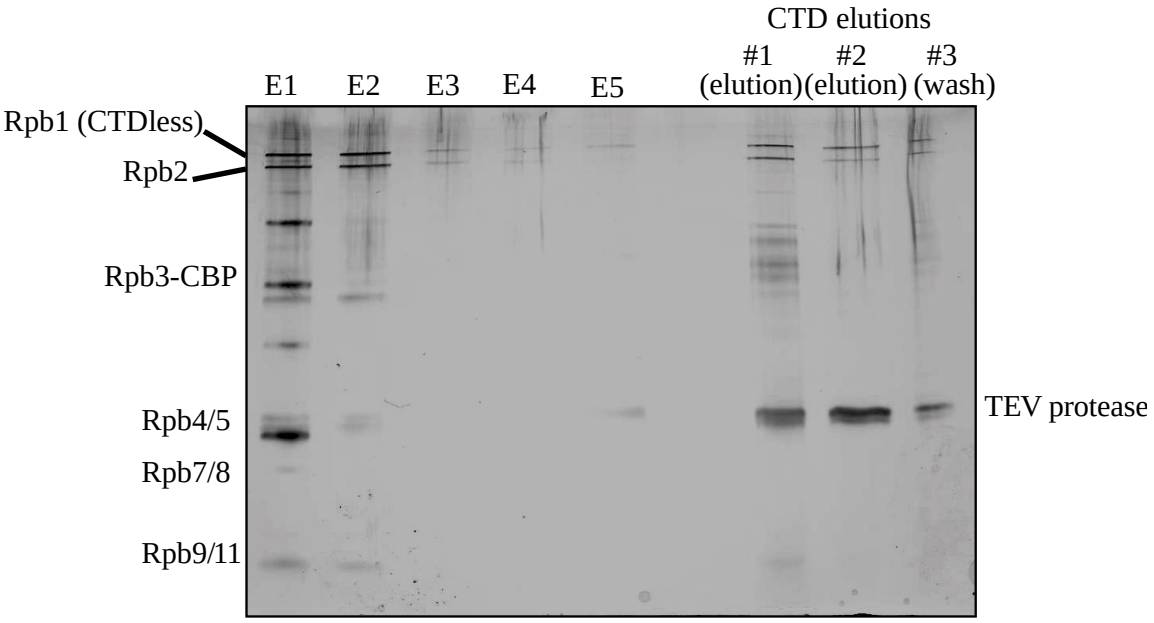


Figure 11. Silver staining SDS-PAGE analysis of purification II. The gel showing RNAPII subunits detected in calmodulin elutions as indicated in the right side of the gel. CTD elutions (elutions of the CTD that released by TEV protease cleavage) are labeled as following: Sample #1 is the CTD elution obtained from purification I. Sample #2 is the CTD elution obtained from purification II, and sample #3 is the wash elution collected post the elution of the cleaved CTD (purification II).

To further confirm the cleavage of CTD, two western-blot analyses were performed. First western-blot was performed to detect the presence of CTD in calmodulin elutions and CTD elutions using antibody against phosphorylated CTD (Anti-H14)

(Figure 12 top). In western-blot analysis, CTD was observed in CTD elutions (sample #3 and #1). On the other hand, CTD was not observed in calmodulin elutions (Figure 12 top). The absence of CTD in calmodulin elution samples indicates that the CTD was cleaved from Rpb1 linker region by TEV protease.

Second western-blot was performed to determine the cleavage of TAP tag in the calmodulin elutions as well as to examine if the cleaved tag is present in the CTD elutions using polyclonal antibody against TAP tag. Western-blot analysis confirmed the presence of CBP-Rpb3 around 50 kDa in calmodulin elutions. Conversely, CBP-Rpb3 was not identified in the CTD elutions (Figure 12 bottom). These data showed that RNAPII subunits were not detected in CTD elutions.

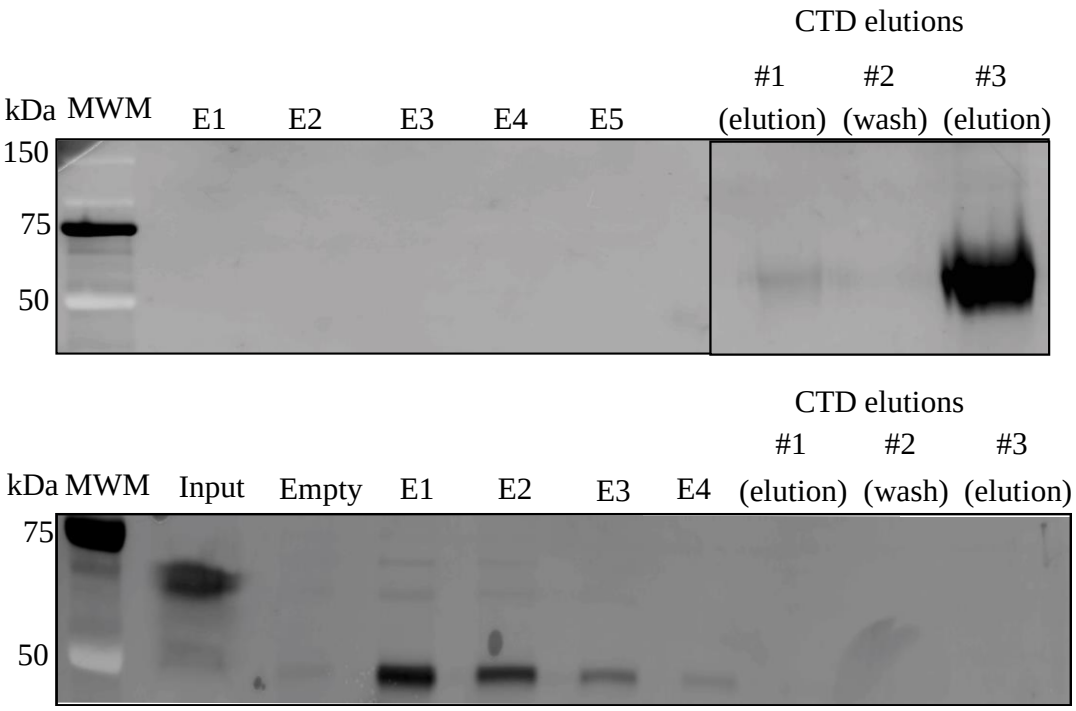


Figure 12. Detection of CTD cleavage by western-blot analysis. Top: western-blot analysis with anti-H14 to detect the presence of the CTD in calmodulin elutions and CTD elutions. Bottom: western-blot analysis with anti-TAP to detect the presence of cleaved TAP tag in calmodulin and CTD elutions. E1-E5 (calmodulin elutions from purification II). CTD elutions (elutions of the CTD that released by TEV protease cleavage) are labeled as following: Sample #1 is the CTD elution obtained from purification II. Sample #2 is the wash elution collected post the elution of the cleaved CTD (purification II), and sample #3 is the CTD elution obtained from purification I.

MudPIT Analysis

Calmodulin elutions and CTD elutions (hereafter referred to as CTD-less RNAPII and CTD-associated respectively) obtained from aforementioned purifications were analyzed in mass spectrometry using MudPIT method and database search algorithms SEQUEST. In CTD-less RNAPII, all the twelve subunits of RNAPII were detected and listed in Table 6. A total of twenty-one transcription-relevant proteins were identified in CTD-less RNAPII samples with at least one unique peptide detected (Table 7). The common TAP purification contaminants in *S. cerevisiae* and the cytoplasmic ribosomal

subunits were excluded from the results as described before (Korgan et al., 2006, 2004). Additionally, the proteins that are not transcription-relevant and were detected with at least 2 unique peptides and 100 PSMs are also included in (Table 7).

Table 6: RNAPII 12 subunits detected in CTD-less RNAPII samples by MudPIT.

Accession	Description	Sequence Coverage %	# Peptide	# PSMS	# AAs	MW
6320061	Rpb1	62.55	112	7108	1733	191.5
398365417	Rpb2	66.58	105	6412	1224	138.7
398364469	Rpb3	78.30	25	1502	318	40.3
6319630	Rpb5	61.86	17	689	215	25.1
6322321	Rpb4	86.43	18	499	221	25.4
6324798	Rpb8	44.52	10	379	146	16.5
6325445	Rpb6	58.71	8	269	155	17.9
398365775	Rpb10	54.29	4	259	151	8.3
6321368	Rpb9	61.48	10	237	144	14.3
6324569	Rpb11	96.67	18	170	120	13.6
6321937	Rpb12	60.00	4	88	70	7.7
398366567	Rpb7	55.48	9	50	171	19.0

Table 7. Proteins identified in CTD-less RNAPII samples by MudPIT.

Accession	Description	Sequence Coverage %	# Peptide	# PSMS	# AAs	MW
Transcription-relevant proteins						
gi6323632	Spt5p	5.64	7	10	1063	115.6
gi6321552	Spt6p	5.69	2	7	1451	168.2
gi398365155	Npa3p	9.35	3	3	385	43.2
gi39836484	Rtr1p	7.78	2	3	226	26.27
gi6320041	Dhh1p	7.71	3	3	506	57.5
gi6321013	Pab1p	6.58	3	3	577	64.3

gi6321295	Dst1p	4.21	3	3	309	34.8
gi6321911	Set1p	1.68	2	3	1080	123.8
gi6324761	Rap1p	5.26	2	2	433	48.9
gi37362706	Asr1p	4.86	2	2	288	33.3
gi63233999	Rpd3p	3.70	2	2	437	47.9
gi39836603	Pbp1p	3.46	2	2	722	78.79
gi39836667	Rba50p	3.19	2	2	439	50.29
gi39836659	Npl3p	3.14	2	2	414	45.41
gi6323676	Arp9p	2.57	2	2	467	53.0
gi33044360	Set2p	2.05	2	2	733	84.45
gi6324879	Isw2p	1.52	2	2	1120	130.2
gi6323839	Rtp1p	1.12	2	2	981	113.2
gi39836660	Ada2p	3.00	1	1	434	50.55
gi39836582	Caf130	1.07	1	1	1122	128.7
gi10383811	Cdc39	0.57	1	1	2108	240.1
Not transcription-relevant proteins						
gi398364951	Vma2p	53.77	35	135	517	57.7
gi6323278	Yef3p	16.67	26	101	1044	115.9
gi6324707	Eft1p	16.39	20	99	842	93.2

In CTD-associated, a total of fifty-one transcription-relevant proteins were identified with at least one unique peptide detected, the proteins are listed in (Table 8). The proteins that are not transcription-relevant and were detected with at least 2 unique peptides and 100 PSMs are included in (Table 8).

Table 8. Proteins identified in CTD-associated samples by MudPIT.

Accession	Description	Sequence Coverage %	# Peptide	# PSMS	# AAs	MW
Transcription-relevant proteins						
gi6320119	Sub2p	29.70	11	40	577	64.3

gi39836659	Npl3p	26.52	12	21	414	45.41
gi39836667	Rba50p	22.32	12	21	439	50.29
gi6324761	Rap1p	14.19	8	14	433	48.9
gi6321013	Pab1p	13.34	8	14	577	64.3
gi6321295	Dst1p	16.83	7	12	309	34.8
gi6321333	Arc1p	13.30	6	12	376	42.1
gi39836515	Npa3p	8.05	5	11	385	43.25
gi39836484	Rtr1p	13.72	4	7	226	26.27
gi39836506	Sup45p	8.92	6	7	437	49.01
gi6323738	Npl6p	6.67	4	6	49.6	4.96
gi6323632	Spt5p	4.42	4	6	1063	115.6
gi39836603	Pbp1p	4.29	4	6	722	78.79
gi6321231	Spt16p	2.8	6	6	1035	118.6
gi323144	Cft2p	11.71	2	5	859	96.2
gi6321265	Xrn1p	4.29	5	5	1528	175.3
gi6321552	Spt6p	2.76	4	5	1451	168.2
gi6321265	Xrn1p	4.29	5	5	1528	175.3
gi632089	Snu13p	25.40	3	4	126	13.6
gi6320041	Dhh1p	10.47	3	4	506	57.5
gi39865953	Sup35p	4.96	4	4	685	76.5
gi6324761	Pan2	4.48	4	4	1115	127
gi6321911	Set1p	2.04	4	4	1080	123.8
gi6319772	Isw1p	3.01	3	3	1129	131
gi39836545	Spt4p	18.63	2	2	102	11.27
gi6324613	Ckb2p	8.91	2	2	258	29.8
gi6323676	Arp9p	6.85	2	2	467	53
gi6320589	Yra1p	5.75	1	2	226	24.9
gi6323336	Pob3p	4.89	2	2	553	63.0
gi37362669	Ess1p	4.71	2	2	170	19.4
gi6321625	Tfg1p	3.13	2	2	735	82.1
gi6314675	Taf5p	3.13	2	2	798	88.9

gi6323518	Nab6p	2.82	2	2	1134	126.1
gi6325028	Cet1p	2.00	2	2	549	61.8
gi63198890	Taf2p	1.56	2	2	1470	161.4
gi10383811	Cdc39	1.44	2	2	2108	240.1
gi6323307	Ysh1p	1.41	2	2	779	87.6
gi39836582	Caf130	1.07	2	2	1122	128.77
gi323892	Taf9p	7.64	1	1	157	17.3
gi39836488	Ckb1p	6.12	1	1	278	32.27
gi39836592	Taf10p	4.37	1	1	20.6	23.01
gi6322710	Ctk1p	3.79	1	1	528	60.5
gi39836491	Spt15p	2.91	1	1	240	27.01
gi6325291	Arp7p	2.73	1	1	477	53.8
gi6323524	Ctk3p	2.70	1	1	296	34.8
gi6323255	Bur2p	2.53	1	1	395	45.7
gi6325246	Taf3p	2.27	1	1	353	40.3
gi6321326	Taf6p	2.13	1	1	57.9	5.43
gi6325066	Nab3p	1.98	1	1	459	51.9
gi6323572	Taf8p	1.96	1	1	510	40.3
gi6321012	Chd1p	1.57	1	1	656	75.1
gi6322334	Spt10p	1.25	1	1	640	72.9
Not transcription-relevant proteins						
gi6325337	Tef1p	58.73	26	1066	408	50.0
gi6324707	Eft1p	16.39	20	216	842	93.2
gi6323278	Yef3p	16.67	26	187	1044	115.9
gi6324265	Ydj1p	56.23	16	132	409	132
gi398364951	Vma2p	47.0	35	113	517	57.7
gi6323013	Tsa1p	48.98	12	103	196	21.6

DISCUSSION

In an effort to advance the methodology in analyzing RNAPII protein-protein interaction network and to determine the role of the CTD in controlling RNAPII transcription, we devised a method to specifically isolate the CTD-associated and CTD-less RNAPII to identify the proteins that interact with both the CTD and the globular core of RNAPII using novel purification scheme coupled to quantitative proteomics.

The first approach in this method was placing TEV protease cleavage site at the flexible linker region of RNAPII upstream of CTD using a modified site directed mutagenesis PCR-based method generating a construct with *RPB1*-TEV gene (Figure 7). Similar Rpb1 modified constructs have been made before to prepare a CTD-less RNAPII for *in vitro* transcription studies (Li and Kornberg, 1993; Nair et al., 2005). The site of TEV insertion at the linker region of *RPB1* was determined based on crystal structures of RNAPII. In the crystal structures, the linker emerges from the RNAPII core below the clamp and connects to the CTD of RNAPII as shown in Figure 13.

Further, the modified Rpb1-TEV gene was introduced in *S. cerevsisiae* strains expressing wild type *RPB1* gene from a plasmid (the native *RPB1* gene was deleted with a *CLONAT* resistance cassette) to replace the wild type gene using 5-FOA plasmid shuffle yielding strains dependent on the modified *RPB1*-TEV gene for growth (See Figure 8 for the plasmid shuffle scheme). The next approach in our method was tagging the strain with a modified TAP tag that has TEV cleavage site replaced by 3C cleavage site. A plasmid construct expresses the 3C-TAP was amplified by PCR to increase the homology region upstream and downstream of *RPB3* gene. And then the PCR fragment was integrated at the C-terminal of *RPB3* gene of the *RPB1*-TEV strain by homologues

recombination. The presence of the 3C-TAP tag was confirmed by western-blot analysis, as shown in (Figure 9B), Rpb3-3C-TAP Rpb1-TEV strain and the positive control (Rpb3-TAP) were detected at ~70 kDa whereas the negative control (BY4741) was not detected.

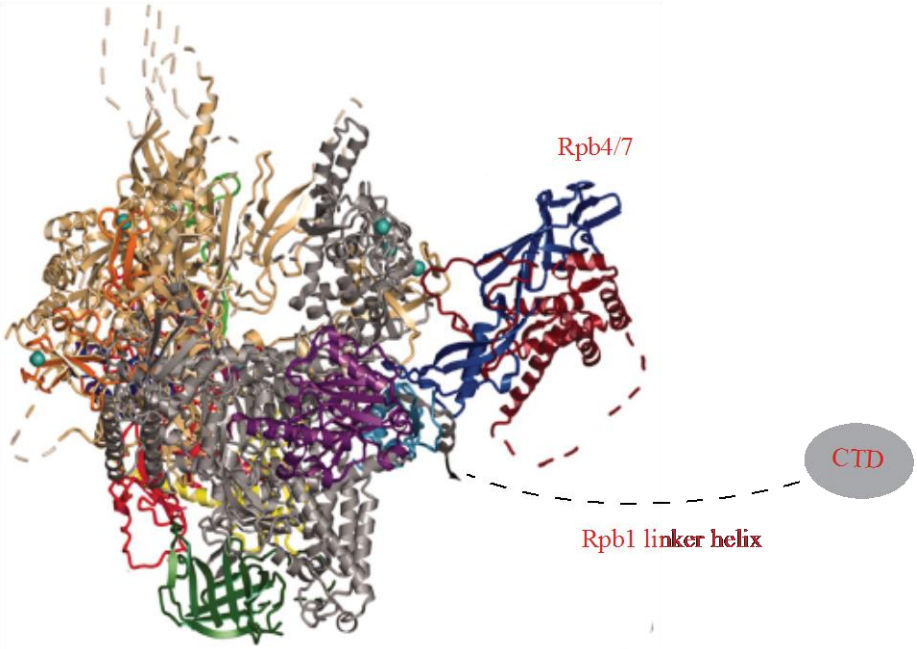


Figure 13. A ribbon model of RNAPII showing the linker helix region as visualized upon Rpb4/Rpb7 binding. In the right side of the figure a cartoon schematic of the CTD and Rpb1 linker helix, linker helix is shown as a dashed line. Figure is adapted from (Armache et al., 2003; 2005).

To isolate the CTD-associated and CTD-less RNAPII, we designed a novel TAP purification scheme (Figure 3 and 4) by modifying the original TAP protocol (Rigaut et al., 1999; Puig et al., 2001) to take advantage of TEV cleavage site upstream of CTD. In the modified purification, we performed on-bead cleavage of the CTD using TEV protease while the RNAPII and its associated proteins are retained on the IgG beads through association with Rpb3-3C-TAP. The CTD-less RNAPII along with the associated proteins were released from IgG resins by the action of 3C protease followed by incubation with calmodulin resins and the release of the CTD-less RNAPII and the

associated proteins with EGTA. The isolated CTD-less RNAPII (calmodulin elutions) was analyzed by silver stained SDS-PAGE gel to evaluate the efficiency of our designed purification method. The on-bead cleavage of the CTD by TEV protease did not prevent RNAPII subunits detection; all the 12 subunits were detected in the silver stained SDS-PAGE analysis (Figure 10). In previous TAP purification silver staining analysis of wild type *S. cerevisiae*, two forms were observed for Rpb1 subunit corresponded to hyper-phosphorylated CTD and hypo-phosphorylated CTD (Mosley et al., 2011). Conversely, we only observed one form of Rpb1 subunit around 170 kDa we hypothesized that the mobility change of the Rpb1subunit on the gel caused by truncation from TEV protease cleavage. Additionally, similar CTD-less RNAPII mobility trend on SDS-PAGE was observed before (Li and Kornberg, 1993; Nair et al., 2005).

In preliminary data from MudPIT analyses of the CTD-associated (CTD elution) it was observed that the purification was contaminated with RNAPII subunits, all the twelve subunits were detected (Table 5). However, the number of peptide spectra matches (PSMs) for the twelve subunits were below the average of detected PSMs values of RNAPII subunits purified from wild type *S. cerevsisiae* as previously observed (Mosley et al., 2013). Therefore, we hypothesized that TEV protease cleaved a portion of the 3C cleavage sites thus releasing part of RNAPII subunits from the IgG resins.

In general, the TEV protease recognition site is consisting of seven amino acids (ENLYFQ↓G/S) (reviewed in Waugh, 2011). Previous experimental studies investigating the specificity of TEV protease have shown a strong preference for glutamic acid and tyrosine in the P6 and P3 position respectively and a moderate preference for phenylalanine, leucine and asparagine in the P2, P4, P5 position respectively (Daugherty

et al., 1989; Carrington et al., 1988). These data lead to define the consensus for TEV protease cleavage site as "Glu-Xaa-Xaa-Tyr-Xaa-Gln ↓Ser/Gly" with the assumption that Xaa position could be substituted with other residues (Daugherty et al., 1989; Carrington et al., 1988). However, recent study has shown that various residues can be accommodated in P1′ position with little impact on TEV protease cleavage efficiency (Kapust et al., 2002). In contrast, the 3C protease recognition site is consisting of eight amino acids (LEVLFQ↓PG) where the presence of Gly-Pro dipeptide after the scissile bond is restricted (reviewed in Waugh, 2011). Unlike TEV protease, the specificity of 3C protease has not been extensively investigated in the literature.

Consequently, we propose that the specificity tolerance of TEV protease for residues in P1′, P2, P4 and P5 positions caused cleavage of a portion of the 3C sites of our modified TAP tag. In an attempt to eliminate the cleavage of 3C by TEV protease, we revised our purification method (Figure 4B). In the revised purification, we performed 3C protease cleavage and the releasing of RNAPII and the associated proteins from the IgG resins. Following the incubation of RNAPII subunits with calmodulin resins, TEV protease cleavage was performed while the RNAPII subunits were immobilized on calmodulin sepharose resin. Subsequently, aliquots of the isolated CTD-less RNAPII and CTD-associated were analyzed by SDS-PAGE and visualized with silver staining (Figure 11). It was noted from SDS-PAGE analysis that purification II yield was significantly less than purification I (Figure 11 and 10 respectively). Yet, a similar pattern of RNAPII subunits was detected including Rpb1 mobility change on SDS-PAGE. Additionally, we observed unique protein in CTD-associated samples around 30 kDa, based on the size of the protein we proposed that this protein is TEV protease (Figure 11). The cleaved CTD

remnant appears to run as a smear just above the 37 kDa marker and can be detected using monoclonal antibodies against the CTD (Figure 12).

In order to increase the confidence of CTD cleavage, two western-blot analyses were performed using second aliquots of the isolated CTD-less RNAPII and CTD-associated (Figure 12). It was concluded from the first western-blot analysis that the CTD was present in CTD-associated samples (#3 and #1) but not in the CTD-less RNAPII samples (Figure 12 top). In the second western-blot analysis, the cleaved TAP tag (CBP-Rpb3) was not observed in CTD-associated samples conversely it was observed in CTD-less RNAPII samples (Figure 12 bottom). In summary, western-blot data confirmed the isolation of CTD from RNAPII. Besides, the data has indicated that the CTD-associate samples from purification II were not contaminated with significant amounts of RNAPII subunits.

The last approach in our designed method was analyzing the CTD-less RNAPII and CTD-associated samples in mass spectrometry using MudPIT followed by quantitative analysis using database search alogrithms SEQUEST through the software program proteome Discovere 1.4 (Thermo). It was observed before in (unpublished data from Mosley lab) that the native CTD peptides especially in *S. cerevisiae* are undetectable in mass spectrometry analysis due to the lack of basic residues in its primary amino acid sequence for trypsin digestion. Therefore, in an attempt to clarify if the isolation of CTD through our developed method could be detectable in mass spectrometry, MS data was screened for the peptide that was cleaved by TEV protease. However, the peptide was not detectable because the region of the Rpb1 linker where the TEV site was placed has only one basic lysine residue (Figure 14). In an assumption that

trypsin cleaved the peptide bond between lysine and aspartic acid upstream TEV site then the resulting peptide is 30-mer which is a suboptimal peptide to be detected by mass spectrometry (Figure 14).

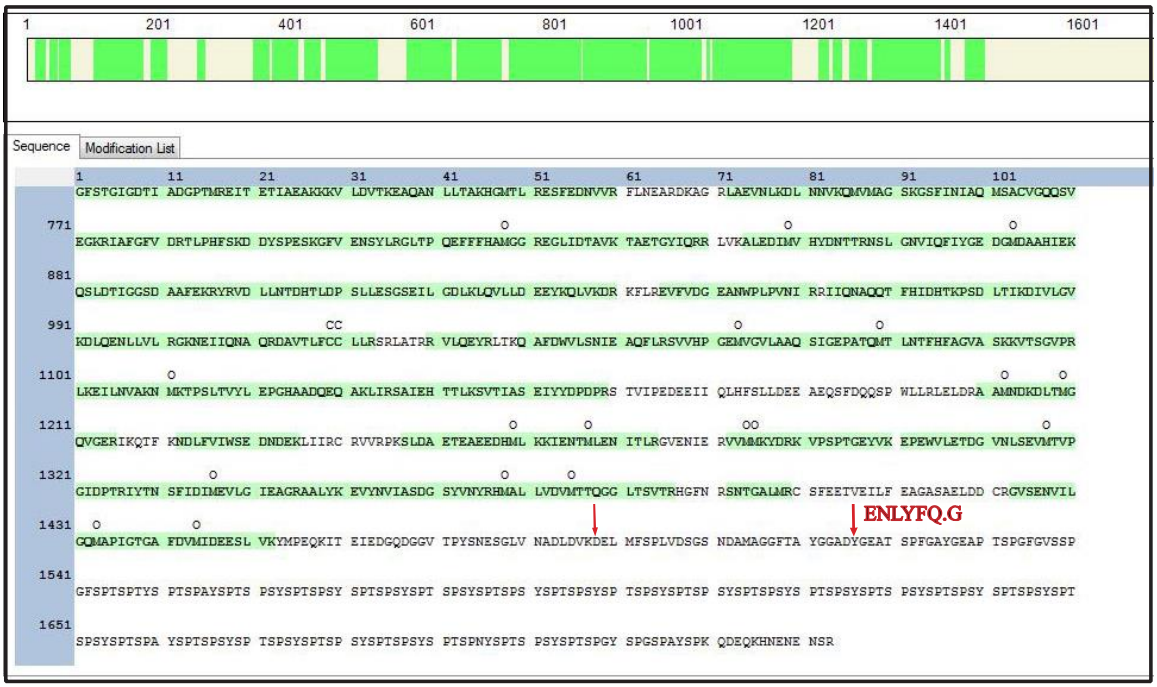


Figure 14. Peptide sequences of Rpb1 subunit of RNAPII detected by MudPIT. The sections highlighted in green were identified by MudPIT analysis. The red arrows indicate the position of TEV protease cleaved peptide that was not detected by mass spectrometry; TEV site is shown in red font. The peptide sequence is: DELMFSPLVDSGSNDAMAGGFTAY(ENLYFQ.G).

It was noted that the PSMs values for all the proteins detected in CTD-less RNAPII and CTD-associated from purification II were significantly low, we speculated that the reason for that was the poor yield of purification II. Therefore, the mass spectrometry data presented in Table 7 and 8 were combined from purifications I and II.

In addition to the 12 subunits of RNAPII, a total of twenty-one transcription-relevant proteins were detected in CTD-less RNAPII. In CTD-associated a total of fifty-one transcription- relevant proteins were detected. Some of these proteins were detected

in both CTD-associated and CTD-less RNAPII; these proteins are summarized in (Figure 15) according to the PSMs values. Among these proteins, there were two proteins Spt5 and Spt6 detected with higher PSMs values in CTD-less RNAPII than CTD-associated. Spt5 is a component of elongation factor DSIF (Spt4/spt5) involved in controlling transcription elongation whereas Spt6 is an elongation factor involved in nucleosomes remolding, Spt5 and Spt6 are essential for growth in *S. cerevisiae* (Wada et al., 1998). In addition, genetic analysis has suggested that Sp5 and Spt6 are functionally related but no evidence if they are physically associated (Hartzog et al., 1998). It has been reported before that Spt5 is the most enriched RNAPII interacting protein identified in mass spectrometry when using one of RNAPII subunits such as Rpb3, Rpb7 or Rpb11 as bait for TAP purification (Mosley et al., 2013, 2011). In correlation with this observation, Spt5 was detected in both CTD-less RNAPII and CTD-associated. Interestingly, three of the low abundance proteins that were hardly detected in mass spectrometry quantitative analysis as observed before (Mosley et al., 2013) Set2, Asr1, and Npa3 were detected in CTD-less RNAPII.

A total of nine transcription-relevant proteins were identified in CTD-associated with higher PSMs values than in CTD-less RNAPII (Figure 15), among these proteins is Dst1. Dst1 (TFIIS) is an elongation factor that is conserved in bacteria and eukaryotes, it composes of three domains N-terminal domain I, C-terminal domain II and zinc loop fold domain III. Dst1 stimulates the intrinsic transcripts cleavage activity of RNAPII following backtracking (reviewed in Fish and Kane, 2002; Matrinez-Rucobo and Cramer, 2013). RNAPII backtracking is a process of translocation that is involved in transcription through difficult nucleosomes and in promoter-proximal pausing. In backtracking,

RNAPII moves backward along the elongation complex that leads to the disengagement of the 3′ end of the transcript from the active site causing a transcription arrest (reviewed in Fish and Kane, 2002; Matrinez-Rucobo and Cramer, 2013).

Structural studies have revealed the mechanism of Dst1 in stimulating RNAPII intrinsic cleavage activity and releasing RNAPII from the arrested state. Dst1 binds to RNAPII with its C-terminal domain and extending its domain III (zinc ribbon fold) into the active site with three charged residues one basic residue and two acidic residues (Wang et al., 2009; Cheung and Cramer 2011). The binding of Dst1 to the arrested RNAPII stimulates the recruitment of Mg^{+2} ion and water molecule to RNAPII active site. Consequently, Dst1 facilitates the alignment of 3′ end of the backtracked transcript at the active site and promotes the cleavage of phosphodiester bond of the backtracked transcript by RNAPII through stabilizing the charged transition state and through promoting proton transfer (Wang et al., 2009; Cheung and Cramer 2011).

In addition to stimulate transcription *in vitro*, Dst1 was found to stimulate transcription past an artificial arrest *in vivo* (Kulish and Struhl, 2001). Kulish and Stuhl designed an assay to analyze transcription elongation *in vivo* involves an artificial arrest site (ARTAR). Using ARTAR they observed discontinuous in transcription elongation and they reported that the transcriptional reactivation is strongly related to Dst1 function (Kulish and Struhl, 2001). Interestingly, recent reports suggested that Dst1 contributes to transcription initiation beside its role as an elongation factor. Recent study using ChIP assay demonstrated a cross-linking of Dst1 and GAL1 activation sequences, it also showed that the Dst1 was required for optimal binding of TATA binding protein (TBP) and RNAPII to GAL1 promoter (Prather et al., 2005). In addition Dst1 was identified in

quantitative mass spectrometry analysis of PIC and that could suggest a role of Dst1 in initiation of transcription (Ranish et al., 2003). Yet, there is no direct biochemical evidences to support its role during transcription initiation.

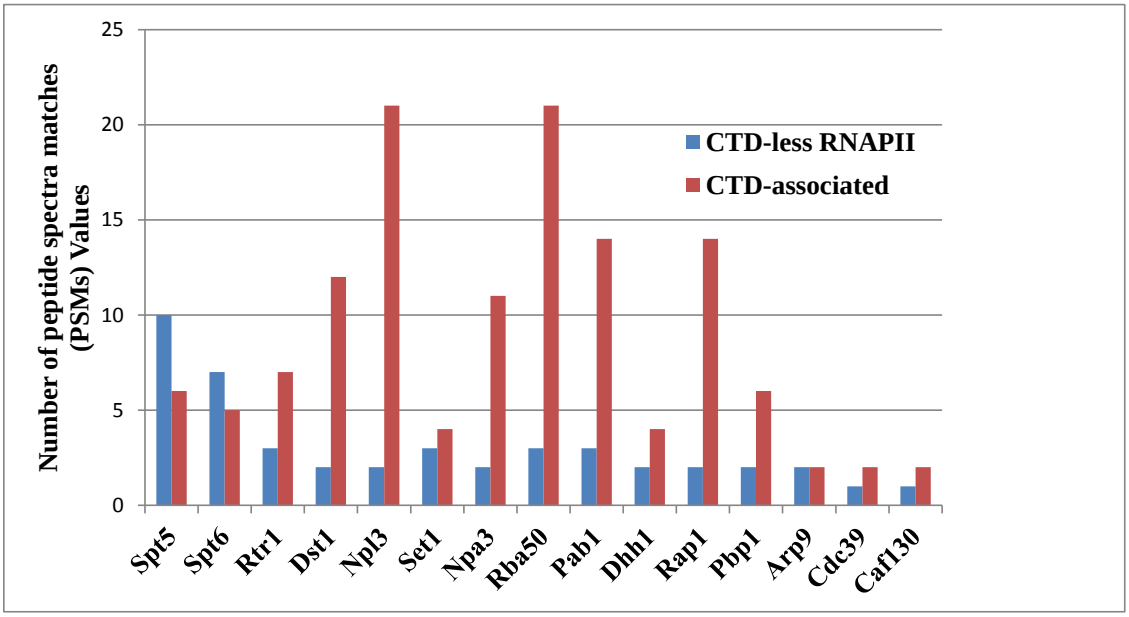


Figure 15. Proteins identified by MudPIT analyses in both CTD-associated and CTD-less RNAPII following TAP purification. The difference in detection level (number of peptide spectra matches) versus the proteins names.

Besides Dst1, there were seven proteins detected in CTD-associated with significant difference of PSMs than CTD-less RNAPII (Figure 15), the proteins are Npl3, Rab50, Pab1, Pbp1, Npa3, Rap1, and Rtr1. Npl3 (also termed Nob1 and Nop3) is an RNA binding protein in *S. cerevisiae* that exports poly (A) mRNA from the nucleus to the cytoplasm, mutation in Npl3 has caused nuclear accumulation of mRNA (Lee et al., 1996). Rba50 is a protein involved in transcription through interaction with Rpb2, Rpb3 and Rpb11of RNAPII (Saccharomyces Genome Database: http://www.yeastgenome.org). Pab1 binds to poly (A) tail of mRNA and function with other factor to couple 3′ end processing with mRNA export (Dunn et al., 2005). Pbp1 is a Pab1 binding protein that

regulates mRNA polyadenylation (Saccharomyces Genome Database: http://www.yeastgenome.org). Npa3 is a cytoplasmic protein important of transporting RNAPII into the nucleus, genetic study has shown that depletion of Npa3 results in accumulation of RNAPII in the cytoplasm (Staresinic et al., 2011). Npa3 as mentioned earlier is a low abundance protein that is usually not detectable in mass spectrometry analysis. Rtr1 is a CTD Ser5 phosphatase that regulates the transition of high Ser5-P level to high Ser2-P (Mosley et al., 2009). Rap1 is a DNA-binding protein that has dual functions either activate or repress transcription (Hardy et al., 1992).

In addition to the proteins mentioned above, four proteins were detected at the same level in CTD-associated and CTD-less RNAPII (Set1, Arp9, Cdc39 and Caf130). Set1 is a histone H3-lysine 4 methyltransferase (H3-K4) that recruited to RNAPII at the 5′ end of the coding region. Set 1 interacts with Ser5 phosphorylated CTD form of RNAPII but not phosphorylated Ser2 (Ng et al., 2003). Cdc39 and Caf130 are components of yeast CCR4-NOT complex that is important for gene regulation during the transcription cycle (Saccharomyces Genome Database: http://www.yeastgenome.org).

Sub2 a component of TREX complex was detected only in CTD-associated with high PSMs value. TREX complex mediates the coupling of mRNA synthesis and the nuclear mRNA export (Strasser et al., 2002). TREX complex consists of THO complex that interacts with phosphorylated Ser2 and Ser5, and Sub2/Yra1 mRNA export factors. ChIP-chip analysis has shown that the occupancy of THO, Sub2 and Yra1increases from 5′ end to 3′ end of the gene (Meinel et al 2013). Meinel and colleagues used ChIP-based assay with a reporter containing self-cleaving ribozyme, through ChIP data they proposed that TREX recruitment to transcription machinery is dependent upon the interaction of

THO complex to Ser2-P and Ser5-P and partially RNA dependent. Interestingly, Yra1 was detected in CTD-associated with a significant lower PSMs value than Sub2.

Subunits of the general transcription factor TFIID were identified in CTD-associated sample including Taf2/Taf3/Taf5/Taf6 and Taf8-10. In general, TFIID comprises TBP and 13 TBP-associated factors (TAFs). TBP and Taf2/Taf3/Taf8 recognize and bind to the TATA box at the core promoter prior PIC assembly (reviewed in Thomas and Chiang, 2006). Some of the TFIID subunits Taf5/Taf9/Taf10 are present in SAGA complex and involved in chromatin modification (Grant et al., 1998). Previous studies have shown that mutations in three of TFIID subunits (Taf5, Taf10 and Taf12) impact SAGA composition and integrity proposing an important role of these subunits for SAGA structure (Durso et al., 2001; Grant et al., 1998).

Ctk1 and Ctk2 are subunits of CTDK-I the kinase complex that phosphorylates Ser2. Both subunits were identified in CTD-associated with only one unique peptide identified for each. In addition, the following proteins were identified with only one unique peptide for each: Nab3, Chd1, Yra1 and Bur2. Nab3 is an RNA binding protein that associates with Nrd1 and Sen1 to form a complex to interact with exosome and stimulate RNA degradation activity be exosome (reviewed in Kuehner et al., 2011). The function of Nrd1-Nab3 complex is linked to CTD phosphorylation. Increasing of CTD Ser2-P phosphorylation by mutation in Fcp1 leads to decrease in Nrd1 complex recruitment and termination defects (Gudipati et al., 2008).

Components of FACT complex were detected in CTD-associated including Sp16 and Pob3. FACT complex is an elongation factor that interacts specifically with nucleosomes and H2A/H2B dimers and facilitates nucleosomes disassembly ahead of

RNAPII (Orphanides et al., 1999). It has been proposed that FACT affects transcription elongation fidelity by linking initiation to elongation. Mutation in Spt16 results of decreased PIC at the promoter complex and suppressed transcription of some of the coding genes (Mason and Struhl, 2003).

In summary, our results have indicated that the devised purification method is an efficient method in analyzing RNAPII protein interaction network. Through this method we successfully were able to isolate the CTD and CTD-less RNAPII (which has been demonstrated by western-blot analyses) without affecting the detection of the 12 subunits of RNAPII as was shown in silver stained SDS-PAGE and mass spectrometry analyses. In addition through this method we were able to detect various proteins that involved in transcription including high abundance proteins such as elongation factors Spt4/Spt5, Dst1, components of FACT complex and Spt6. In addition, the low abundance proteins that only interact with RNAPII during certain stage in transcription were detected such as Asr1, Set2, and Npa3.

In the modified TAP purification, two purification schemes have been used. In the first scheme (Figure 4A), on-bead TEV cleavage of the CTD was performed while the RNAPII complex was immobilized on IgG beads. Although this purification scheme resulted in good purification yield as well as good detection level of RNAPII interacting proteins by MudPIT analyses, the isolated CTD-associated was contaminated with RNAPII subunits as discussed earlier. This limitation could be overcome by incubating the CTD supernatant following TEV cleavage with calmodulin beads to remove the contaminant and ensure that the RNAPII subunits are cleared from the CTD-associated sample or by scaling up the starting amount of protein for RNAPII isolation.

Another approach to specifically isolate the CTD and eliminate the cleavage of 3C site by TEV protease is to perform the CTD cleavage with 3C protease rather than TEV protease. As been demonstrated with our data, 3C is more specific and more restricted for cleaving its site than TEV protease. Therefore, a suggested future path is to use the Rpb1-3C-CTD construct that was generated through this work (Figure 7E) to construct yeast strains express Rpb1-3C-CTD gene and tagged with the original TAP tag or a 3X FLAG tag. Therefore the purification scheme would be similar as purification I scheme except performing on-bead 3C cleavage of the CTD followed by releasing RNAPII complex from IgG beads with TEV protease.

The protein detection level by mass spectrometry in CTD-associated and CTD-less RNAPII from second purification scheme was significantly lower than expected as described earlier. In the second purification scheme TEV cleavage was performed overnight while RNAPII and associated protein are immobilized on calmodulin beads. The interaction of calmodulin with CBP may not be as strong as the interaction of IgG-coated beads with prot A and could be disrupted under mild condition causing a loss of RNAPII subunits along with the associated proteins. Therefore, we concluded that the first purification scheme is more efficient than the second scheme and it will be further tested to overcome its limitations and to validate its performance in detection protein level in wild type and mutant *S. cerevisiae.*

CONCLUSIONS

The CTD of the largest subunit of RNAPII serves as a platform for the recruitment of various transcription factors and accessory proteins. The recruitment of these factors and proteins is controlled by the phosphorylation state of the CTD that marks each stage of transcription and mediates the coupling of RNAPII transcription with RNA processing as well as controlling chromatin modification.

Mosley and colleagues use TAP purification and MudPIT method followed by quantitative proteomics to analyze RNAPII protein interaction network to gain better understanding on the regulation of transcription elongation dynamics. Through this approach they were able to identify novel regulatory pathways for example identifying Rtr1 as a CTD phosphatase that regulates the transition of high Ser5-P level to high Ser2-P (Mosley et al., 2009).

Although this approach has proved to be an efficient tool in analyzing RNAPII protein interaction network, some limitations have been associated with it including the identification of low abundance proteins. Consequently, the focus in the work presented here was to improve the quantitative proteomics method used by Mosley laboratory in analyzing RNA protein interaction through developing a novel purification method to isolate the CTD of RNAPII.

Wild type yeast strain expressing Rpb1-TEV-CTD and tagged with Rpb3-3C-TAP was generated in this work as discussed earlier. Subsequently, the CTD-associated and CTD-less RNAPII were specifically isolated through the novel purification method, which has been demonstrated in western-blot analyses (Figure 12). In addition several

proteins involved in transcription have been detected in CTD-associated and CTD-less RNAPII by MudPIT analyses including high and low abundance proteins (Table 7 and 8).

Our data has shown that the novel purification method is an efficient tool for quantitative proteomics analysis. Yet, more experiments need to be conducted to validate the method and to measure proteins enrichment level in wild type and mutant *S. cerevisiae* using Scaffold quantitative analysis.

REFERENCES

Aebersold, Ruedi and Matthias Mann. "Mass Spectrometry-Based Proteomics." *Nature* 422, no. 6928 (2003): 198-207.

Akhtar, M. S., M. Heidemann, J. R. Tietjen, D. W. Zhang, R. D. Chapman, D. Eick and A. Z. Ansari. "TfIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II." *Mol Cell* 34, no. 3 (2009): 387-93.

Armache, K. J., H. Kettenberger and P. Cramer. "Architecture of Initiation-Competent 12-Subunit RNA Polymerase II." *Proc Natl Acad Sci U S A* 100, no. 12 (2003): 6964-8.

Armache, K. J., S. Mitterweger, A. Meinhart and P. Cramer. "Structures of Complete RNA Polymerase II and Its Subcomplex, Rpb4/7." *J Biol Chem* 280, no. 8 (2005): 7131-4.

Bataille, Alain R, Célia Jeronimo, Pierre-Étienne Jacques, Louise Laramée, Marie-Ève Fortin, Audrey Forest, Maxime Bergeron, Steven D Hanes and François Robert. "A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes Along Genes." *Molecular Cell* 45, no. 2 (2012): 158-170.

Bauer, A. and B. Kuster. "Affinity Purification-Mass Spectrometry. Powerful Tools for the Characterization of Protein Complexes." *Eur J Biochem* 270, no. 4 (2003): 570-8.

Belotserkovskaya, Rimma and Danny Reinberg. "Facts About Fact and Transcript Elongation through Chromatin." *Current Opinion in Genetics & Development* 14, no. 2 (2004): 139-146.

Boeke, J. D., F. LaCroute and G. R. Fink. "A Positive Selection for Mutants Lacking Orotidine-5'-Phosphate Decarboxylase Activity in Yeast: 5-Fluoro-Orotic Acid Resistance." *Mol Gen Genet* 197, no. 2 (1984): 345-6.

Boeke, Jef D., Joshua Trueheart, Georges Natsoulis and Gerald R. Fink. "[10] 5-Fluoroorotic Acid as a Selective Agent in Yeast Molecular Genetics." In *Methods in Enzymology*, edited by Lawrence Grossman Ray Wu, Volume 154, 164-175: Academic Press, 1987.

Buratowski, S., S. Hahn, L. Guarente and P. A. Sharp. "Five Intermediate Complexes in Transcription Initiation by RNA Polymerase II." *Cell* 56, no. 4 (1989): 549-61.

Buratowski, Stephen. "Progression through the RNA Polymerase II CTD Cycle." *Molecular Cell* 36, no. 4 (2009): 541-546.

Bushnell, D. A., C. Bamdad and R. D. Kornberg. "A Minimal Set of RNA Polymerase II Transcription Protein Interactions." *J Biol Chem* 271, no. 33 (1996): 20170-4.

Bushnell, D. A. and R. D. Kornberg. "Complete, 12-Subunit RNA Polymerase II at 4.1-a Resolution: Implications for the Initiation of Transcription." *Proc Natl Acad Sci U S A* 100, no. 12 (2003): 6969-73.

Carrington, J. C., S. M. Cary and W. G. Dougherty. "Mutational Analysis of Tobacco Etch Virus Polyprotein Processing: Cis and Trans Proteolytic Activities of Polyproteins Containing the 49-Kilodalton Proteinase." *J Virol* 62, no. 7 (1988): 2313-20.

Cheung, A. C. and P. Cramer. "Structural Basis of RNA Polymerase II Backtracking, Arrest and Reactivation." *Nature* 471, no. 7337 (2011): 249-53.

Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt and S. Buratowski. "Opposing Effects of Ctk1 Kinase and Fcp1 Phosphatase at Ser 2 of the RNA Polymerase II C-Terminal Domain." *Genes Dev* 15, no. 24 (2001): 3319-29.

Cho, E. J., T. Takagi, C. R. Moore and S. Buratowski. "Mrna Capping Enzyme Is Recruited to the Transcription Complex by Phosphorylation of the RNA Polymerase II Carboxy-Terminal Domain." *Genes Dev* 11, no. 24 (1997): 3319-26.

Cho, H., T. K. Kim, H. Mancebo, W. S. Lane, O. Flores and D. Reinberg. "A Protein Phosphatase Functions to Recycle RNA Polymerase II." *Genes Dev* 13, no. 12 (1999): 1540-52.

Corden, J. L. "Tails of RNA Polymerase II." *Trends Biochem Sci* 15, no. 10 (1990): 383-7.

Cramer, P., K. J. Armache, S. Baumli, S. Benkert, F. Brueckner, C. Buchen, G. E. Damsma, S. Dengl, S. R. Geiger, A. J. Jasiak, A. Jawhari, S. Jennebach, T. Kamenski, H. Kettenberger, C. D. Kuhn, E. Lehmann, K. Leike, J. F. Sydow and A. Vannini. "Structure of Eukaryotic RNA Polymerases." *Annu Rev Biophys* 37, (2008): 337-52.

Dougherty, W. G., S. M. Cary and T. D. Parks. "Molecular Genetic Analysis of a Plant Virus Polyprotein Cleavage Site: A Model." *Virology* 171, no. 2 (1989): 356-64.

Dubois, M. F., M. Vincent, M. Vigneron, J. Adamczewski, J. M. Egly and O. Bensaude. "Heat-Shock Inactivation of the Tfiih-Associated Kinase and Change in the Phosphorylation Sites on the C-Terminal Domain of RNA Polymerase II." *Nucleic Acids Res* 25, no. 4 (1997): 694-700.

Dunn, E. F., C. M. Hammell, C. A. Hodge and C. N. Cole. "Yeast Poly(a)-Binding Protein, Pab1, and Pan, a Poly(a) Nuclease Complex Recruited by Pab1, Connect Mrna Biogenesis to Export." *Genes Dev* 19, no. 1 (2005): 90-103.

Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse. "Electrospray Ionization for Mass Spectrometry of Large Biomolecules." *Science* 246, no. 4926 (1989): 64-71.

Fish, Rachel N. and Caroline M. Kane. "Promoting Elongation with Transcript Cleavage Stimulatory Factors." *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1577, no. 2 (2002): 287-307.

Florens, Laurence and Michael P. Washburn. "Proteomic Analysis by Multidimensional Protein Identification Technology." 328, 159-175, 2006.

Ganem, C., F. Devaux, C. Torchet, C. Jacq, S. Quevillon-Cheruel, G. Labesse, C. Facca and G. Faye. "Ssu72 Is a Phosphatase Essential for Transcription Termination of Snornas and Specific Mrnas in Yeast." *Embo j* 22, no. 7 (2003): 1588-98.

Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga. "Functional Organization of the Yeast Proteome by Systematic Analysis of Protein Complexes." *Nature* 415, no. 6868 (2002): 141-7.

Grant, P. A., D. Schieltz, M. G. Pray-Grant, D. J. Steger, J. C. Reese, J. R. Yates, 3rd and J. L. Workman. "A Subset of TAF(II)S Are Integral Components of the Saga Complex Required for Nucleosome Acetylation and Transcriptional Stimulation." *Cell* 94, no. 1 (1998): 45-53.

Gudipati, R. K., T. Villa, J. Boulay and D. Libri. "Phosphorylation of the RNA Polymerase II C-Terminal Domain Dictates Transcription Termination Choice." *Nat Struct Mol Biol* 15, no. 8 (2008): 786-94.

Guo, Jiannan and David H. Price. "RNA Polymerase II Transcription Elongation Control." *Chemical Reviews* 113, no. 11 (2013): 8583-8603.

Gygi, S. P., G. L. Corthals, Y. Zhang, Y. Rochon and R. Aebersold. "Evaluation of Two-Dimensional Gel Electrophoresis-Based Proteome Analysis Technology." *Proc Natl Acad Sci U S A* 97, no. 17 (2000): 9390-5.

Hahn, S. "Structure and Mechanism of the RNA Polymerase II Transcription Machinery." *Nat Struct Mol Biol* 11, no. 5 (2004): 394-403.

Hanash, S. M. "Biomedical Applications of Two-Dimensional Electrophoresis Using Immobilized Ph Gradients: Current Status." *Electrophoresis* 21, no. 6 (2000): 1202-9.

Hardy, C. F., L. Sussel and D. Shore. "A Rap1-Interacting Protein Involved in Transcriptional Silencing and Telomere Length Regulation." *Genes Dev* 6, no. 5 (1992): 801-14.

Hartzog, G. A., T. Wada, H. Handa and F. Winston. "Evidence That Spt4, Spt5, and Spt6 Control Transcription Elongation by RNA Polymerase II in Saccharomyces Cerevisiae." *Genes Dev* 12, no. 3 (1998): 357-69.

Hong, S. W., S. M. Hong, J. W. Yoo, Y. C. Lee, S. Kim, J. T. Lis and D. K. Lee. "Phosphorylation of the Rna Polymerase Ii C-Terminal Domain by Tfiih Kinase Is Not Essential for Transcription of Saccharomyces Cerevisiae Genome." *Proc Natl Acad Sci U S A* 106, no. 34 (2009): 14276-80.

Hsin, J. P. and J. L. Manley. "The RNA Polymerase II CTD Coordinates Transcription and RNA Processing." *Genes Dev* 26, no. 19 (2012): 2119-37.

Hsu, P. L., F. Yang, W. Smith-Kinnaman, W. Yang, J. E. Song, A. L. Mosley and G. Varani. "Rtr1 Is a Dual Specificity Phosphatase That Dephosphorylates Tyr1 and Ser5 on the RNA Polymerase II Ctd." *J Mol Biol* 426, no. 16 (2014): 2970-81.

Imasaki, T., G. Calero, G. Cai, K. L. Tsai, K. Yamada, F. Cardelli, H. Erdjument-Bromage, P. Tempst, I. Berger, G. L. Kornberg, F. J. Asturias, R. D. Kornberg and Y. Takagi. "Architecture of the Mediator Head Module." *Nature* 475, no. 7355 (2011): 240-3.

Jaehning, J. A. "The Paf1 Complex: Platform or Player in RNA Polymerase II Transcription?" *Biochim Biophys Acta* 1799, no. 5-6 (2010): 379-88.

Jasnovidova, O. and R. Stefl. "The CTD Code of RNA Polymerase II: A Structural View." *Wiley Interdiscip Rev RNA* 4, no. 1 (2013): 1-16.

Kanin, E. I., R. T. Kipp, C. Kung, M. Slattery, A. Viale, S. Hahn, K. M. Shokat and A. Z. Ansari. "Chemical Inhibition of the TFIIH-Associated Kinase Cdk7/Kin28 Does Not Impair Global Mrna Synthesis." *Proc Natl Acad Sci U S A* 104, no. 14 (2007): 5812-7.

Kapust, R. B., J. Tozser, T. D. Copeland and D. S. Waugh. "The P1' Specificity of Tobacco Etch Virus Protease." *Biochem Biophys Res Commun* 294, no. 5 (2002): 949-55.

Karas, M. and F. Hillenkamp. "Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons." *Anal Chem* 60, no. 20 (1988): 2299-301.

Kim, M., H. Suh, E. J. Cho and S. Buratowski. "Phosphorylation of the Yeast Rpb1 C-Terminal Domain at Serines 2, 5, and 7." *J Biol Chem* 284, no. 39 (2009): 26421-6.

Kim, Y. J., S. Bjorklund, Y. Li, M. H. Sayre and R. D. Kornberg. "A Multiprotein Mediator of Transcriptional Activation and Its Interaction with the C-Terminal Repeat Domain of RNA Polymerase II." *Cell* 77, no. 4 (1994): 599-608.

Komarnitsky, P., E. J. Cho and S. Buratowski. "Different Phosphorylated Forms of RNA Polymerase II and Associated Mrna Processing Factors During Transcription." *Genes Dev* 14, no. 19 (2000): 2452-60.

Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey. "Ssu72 Is an RNA Polymerase II CTD Phosphatase." *Mol Cell* 14, no. 3 (2004): 387-94.

Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili and J. F. Greenblatt. "Global Landscape of Protein Complexes in the Yeast Saccharomyces Cerevisiae." *Nature* 440, no. 7084 (2006): 637-43.

Krogan, N. J., W. T. Peng, G. Cagney, M. D. Robinson, R. Haw, G. Zhong, X. Guo, X. Zhang, V. Canadien, D. P. Richards, B. K. Beattie, A. Lalev, W. Zhang, A. P. Davierwala, S. Mnaimneh, A. Starostine, A. P. Tikuisis, J. Grigull, N. Datta, J. E. Bray, T. R. Hughes, A. Emili and J. F. Greenblatt. "High-Definition Macromolecular Composition of Yeast Rna-Processing Complexes." *Mol Cell* 13, no. 2 (2004): 225-39.

Kuehner, J. N., E. L. Pearson and C. Moore. "Unravelling the Means to an End: RNA Polymerase II Transcription Termination." *Nat Rev Mol Cell Biol* 12, no. 5 (2011): 283-94.

Kulish, D. and K. Struhl. "TfIIS Enhances Transcriptional Elongation through an Artificial Arrest Site in Vivo." *Mol Cell Biol* 21, no. 13 (2001): 4162-8.

Kwak, H. and J. T. Lis. "Control of Transcriptional Elongation." *Annu Rev Genet* 47, (2013): 483-508.

Lee, M. S., M. Henry and P. A. Silver. "A Protein That Shuttles between the Nucleus and the Cytoplasm Is an Important Mediator of RNA Export." *Genes Dev* 10, no. 10 (1996): 1233-46.

Li, Y. and R. D. Kornberg. "Interplay of Positive and Negative Effectors in Function of the C-Terminal Repeat Domain of RNA Polymerase II." *Proc Natl Acad Sci U S A* 91, no. 6 (1994): 2362-6.

Link, A. J., J. Eng, D. M. Schieltz, E. Carmack, G. J. Mize, D. R. Morris, B. M. Garvik and J. R. Yates, 3rd. "Direct Analysis of Protein Complexes Using Mass Spectrometry." *Nat Biotechnol* 17, no. 7 (1999): 676-82.

Liu, H. and J. H. Naismith. "An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol." *BMC Biotechnol* 8, (2008): 91.

Liu, Y., L. Warfield, C. Zhang, J. Luo, J. Allen, W. H. Lang, J. Ranish, K. M. Shokat and S. Hahn. "Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the Paf Complex." *Mol Cell Biol* 29, no. 17 (2009): 4852-63.

Martinez-Rucobo, Fuensanta W. and Patrick Cramer. "Structural Basis of Transcription Elongation." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1829, no. 1 (2013): 9-19.

Mason, P. B. and K. Struhl. "The Fact Complex Travels with Elongating RNA Polymerase II and Is Important for the Fidelity of Transcriptional Initiation in Vivo." *Mol Cell Biol* 23, no. 22 (2003): 8323-33.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S. D. Patterson, M. Wickens and D. L. Bentley. "The C-Terminal Domain of RNA Polymerase II Couples Mrna Processing to Transcription." *Nature* 385, no. 6614 (1997): 357-61.

McDonald, W. Hayes, Ryoma Ohi, David T. Miyamoto, Timothy J. Mitchison and John R. Yates Iii. "Comparison of Three Directly Coupled Hplc Ms/Ms Strategies for Identification of Proteins from Complex Mixtures: Single-Dimension Lc-Ms/Ms, 2-Phase Mudpit, and 3-Phase Mudpit." *International Journal of Mass Spectrometry* 219, no. 1 (2002): 245-251.

Meinel, D. M., C. Burkert-Kautzsch, A. Kieser, E. O'Duibhir, M. Siebert, A. Mayer, P. Cramer, J. Soding, F. C. Holstege and K. Strasser. "Recruitment of Trex to the Transcription Machinery by Its Direct Binding to the Phospho-CTD of RNA Polymerase II." *PLoS Genet* 9, no. 11 (2013): e1003914.

Meinhart, A. and P. Cramer. "Recognition of RNA Polymerase II Carboxy-Terminal Domain by 3'-Rna-Processing Factors." *Nature* 430, no. 6996 (2004): 223-6.

Meredith, G. D., W. H. Chang, Y. Li, D. A. Bushnell, S. A. Darst and R. D. Kornberg. "The C-Terminal Domain Revealed in the Structure of RNA Polymerase II." *J Mol Biol* 258, no. 3 (1996): 413-9.

Morris, D. P. and A. L. Greenleaf. "The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-Terminal Domain of RNA Polymerase II." *J Biol Chem* 275, no. 51 (2000): 39935-43.

Mosley, A. L., G. O. Hunter, M. E. Sardiu, M. Smolle, J. L. Workman, L. Florens and M. P. Washburn. "Quantitative Proteomics Demonstrates That the RNA Polymerase II Subunits Rpb4 and Rpb7 Dissociate During Transcriptional Elongation." *Mol Cell Proteomics* 12, no. 6 (2013): 1530-8.

Mosley, A. L., S. G. Pattenden, M. Carey, S. Venkatesh, J. M. Gilmore, L. Florens, J. L. Workman and M. P. Washburn. "Rtr1 Is a Ctd Phosphatase That Regulates RNA Polymerase II During the Transition from Serine 5 to Serine 2 Phosphorylation." *Mol Cell* 34, no. 2 (2009): 168-78.

Mosley, A. L., M. E. Sardiu, S. G. Pattenden, J. L. Workman, L. Florens and M. P. Washburn. "Highly Reproducible Label Free Quantitative Proteomic Analysis of RNA Polymerase Complexes." *Mol Cell Proteomics* 10, no. 2 (2011): M110 000687.

Nair, D., Y. Kim and L. C. Myers. "Mediator and TFIIH Govern Carboxyl-Terminal Domain-Dependent Transcription in Yeast Extracts." *J Biol Chem* 280, no. 40 (2005): 33739-48.

Nesvizhskii, A. I., O. Vitek and R. Aebersold. "Analysis and Validation of Proteomic Data Generated by Tandem Mass Spectrometry." *Nat Methods* 4, no. 10 (2007): 787-97.

Ng, H. H., F. Robert, R. A. Young and K. Struhl. "Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity." *Mol Cell* 11, no. 3 (2003): 709-19.

Ni, Z., C. Xu, X. Guo, G. O. Hunter, O. V. Kuznetsova, W. Tempel, E. Marcon, G. Zhong, H. Guo, W. H. Kuo, J. Li, P. Young, J. B. Olsen, C. Wan, P. Loppnau, M. El Bakkouri, G. A. Senisterra, H. He, H. Huang, S. S. Sidhu, A. Emili, S. Murphy, A. L. Mosley, C. H. Arrowsmith, J. Min and J. F. Greenblatt. "Rprd1a and Rprd1b Are Human Rna Polymerase II C-Terminal Domain Scaffolds for Ser5 Dephosphorylation." *Nat Struct Mol Biol* 21, no. 8 (2014): 686-95.

Nonet, M., D. Sweetser and R. A. Young. "Functional Redundancy and Structural Polymorphism in the Large Subunit of RNA Polymerase II." *Cell* 50, no. 6 (1987): 909-15.

O'Farrell, P. H. "High Resolution Two-Dimensional Electrophoresis of Proteins." *J Biol Chem* 250, no. 10 (1975): 4007-21.

Oh-Ishi, M., M. Satoh and T. Maeda. "Preparative Two-Dimensional Gel Electrophoresis with Agarose Gels in the First Dimension for High Molecular Mass Proteins." *Electrophoresis* 21, no. 9 (2000): 1653-69.

Orphanides, G., W. H. Wu, W. S. Lane, M. Hampsey and D. Reinberg. "The Chromatin-Specific Transcription Elongation Factor Fact Comprises Human Spt16 and Ssrp1 Proteins." *Nature* 400, no. 6741 (1999): 284-8

Palancade, B. and O. Bensaude. "Investigating RNA Polymerase II Carboxyl-Terminal Domain (Ctd) Phosphorylation." *Eur J Biochem* 270, no. 19 (2003): 3859-70.

Phatnani, H. P. and A. L. Greenleaf. "Phosphorylation and Functions of the RNA Polymerase II CTD." *Genes Dev* 20, no. 21 (2006): 2922-36.

Prather, D. M., E. Larschan and F. Winston. "Evidence That the Elongation Factor TFIIS Plays a Role in Transcription Initiation at Gal1 in Saccharomyces Cerevisiae." *Mol Cell Biol* 25, no. 7 (2005): 2650-9.

Prelich, G. "RNA Polymerase II Carboxy-Terminal Domain Kinases: Emerging Clues to Their Function." *Eukaryot Cell* 1, no. 2 (2002): 153-62.

Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm and B. Seraphin. "The Tandem Affinity Purification (Tap) Method: A General Procedure of Protein Complex Purification." *Methods* 24, no. 3 (2001): 218-29.

Ranish, J. A., E. C. Yi, D. M. Leslie, S. O. Purvine, D. R. Goodlett, J. Eng and R. Aebersold. "The Study of Macromolecular Complexes by Quantitative Proteomics." *Nat Genet* 33, no. 3 (2003): 349-55.

Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann and B. Seraphin. "A Generic Protein Purification Method for Protein Complex Characterization and Proteome Exploration." *Nat Biotechnol* 17, no. 10 (1999): 1030-2.

Schwer, B. and S. Shuman. "Deciphering the Rna Polymerase II CTD Code in Fission Yeast." *Mol Cell* 43, no. 2 (2011): 311-8.

Shevchenko, A., M. Wilm, O. Vorm and M. Mann. "Mass Spectrometric Sequencing of Proteins Silver-Stained Polyacrylamide Gels." *Anal Chem* 68, no. 5 (1996): 850-8.

Sogaard, T. M. and J. Q. Svejstrup. "Hyperphosphorylation of the C-Terminal Repeat Domain of RNA Polymerase II Facilitates Dissociation of Its Complex with Mediator." *J Biol Chem* 282, no. 19 (2007): 14113-20.

Staresincic, L., J. Walker, A. B. Dirac-Svejstrup, R. Mitter and J. Q. Svejstrup. "Gtp-Dependent Binding and Nuclear Transport of RNA Polymerase II by Npa3 Protein." *J Biol Chem* 286, no. 41 (2011): 35553-61.

Strasser, K., S. Masuda, P. Mason, J. Pfannstiel, M. Oppizzi, S. Rodriguez-Navarro, A. G. Rondon, A. Aguilera, K. Struhl, R. Reed and E. Hurt. "Trex Is a Conserved Complex Coupling Transcription with Messenger RNA Export." *Nature* 417, no. 6886 (2002): 304-8.

Svejstrup, J. Q., Y. Li, J. Fellows, A. Gnatt, S. Bjorklund and R. D. Kornberg. "Evidence for a Mediator Cycle at the Initiation of Transcription." *Proc Natl Acad Sci U S A* 94, no. 12 (1997): 6075-8.

Thomas, M. C. and C. M. Chiang. "The General Transcription Machinery and General Cofactors." *Crit Rev Biochem Mol Biol* 41, no. 3 (2006): 105-78.

Treutlein, B., A. Muschielok, J. Andrecka, A. Jawhari, C. Buchen, D. Kostrewa, F. Hog, P. Cramer and J. Michaelis. "Dynamic Architecture of a Minimal RNA Polymerase II Open Promoter Complex." *Mol Cell* 46, no. 2 (2012): 136-46.

Wada, T., T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G. A. Hartzog, F. Winston, S. Buratowski and H. Handa. "Dsif, a Novel Transcription Elongation Factor That Regulates RNA Polymerase II Processivity,

Is Composed of Human Spt4 and Spt5 Homologs." *Genes Dev* 12, no. 3 (1998): 343-56.

Wang, D., D. A. Bushnell, X. Huang, K. D. Westover, M. Levitt and R. D. Kornberg. "Structural Basis of Transcription: Backtracked RNA Polymerase II at 3.4 Angstrom Resolution." *Science* 324, no. 5931 (2009): 1203-6.

Wang, W. and B. A. Malcolm. "Two-Stage PCR Protocol Allowing Introduction of Multiple Mutations, Deletions and Insertions Using Quikchange Site-Directed Mutagenesis." *Biotechniques* 26, no. 4 (1999): 680-2.

Washburn, M. P., D. Wolters and J. R. Yates, 3rd. "Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology." *Nat Biotechnol* 19, no. 3 (2001): 242-7.

Waugh, D. S. "An Overview of Enzymatic Reagents for the Removal of Affinity Tags." *Protein Expr Purif* 80, no. 2 (2011): 283-93.

West, M. L. and J. L. Corden. "Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations." *Genetics* 140, no. 4 (1995): 1223-33.

Zhang, D. W., A. L. Mosley, S. R. Ramisetty, J. B. Rodriguez-Molina, M. P. Washburn and A. Z. Ansari. "Ssu72 Phosphatase-Dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-Terminal Domain Is Essential for Viability and Transcription Termination." *J Biol Chem* 287, no. 11 (2012): 8541-51.

Zhou, Q., T. Li and D. H. Price. "RNA Polymerase II Elongation Control." *Annu Rev Biochem* 81, (2012): 119-43.

CURRICULUM VITAE

Nada S. Alakhras

EDUCATION

Indiana University, Indianapolis, IN December 2014

Master of Science in Biochemistry and Molecular Biology

Thesis Project: A method to isolate the CTD of RNA polymerase II for proteomics analysis.

Indiana University, Bloomington, IN August 2010

Bachelor of Science in Biochemistry

PROFESSIONAL EXPERINCE

Lilly Research Laboratories, Eli Lilly and Company (Contract by Advanced Testing Laboratories), Indianapolis, Indiana April 2011-present

Biologist

Biologist at Investigative Toxicology

Key responsibilities:

- Evaluate, optimize and validate Immunoassays for various Biomarkers.
- Conduct Immunoassay ELISA to measure Biomarker level in mouse and rat serum, plasma and urine.
- Analyze HepG2 human cell lysate and supernatant on ELISA based-assay in a process to develop an ELISA method to analyze Biomarkers involved in identifying cell death mechanisms and inflammation.
- Perform Immunoblotting to detect Biomarker level in HepG2 human cells.

- Analyze, interpret, and report data in form of reports and presentations to external groups.
- Write assay procedure and perform assay troubleshooting.
- Document experiments data in research electronic notebook.
- Participate with Toxicology Biomarker group in developing and designing Immunoassays for target Biomarkers.
- Participate with Toxicology Cell segregate group in identifying protein markers involved in cell death mechanisms and inflammation.

Indiana University, Bloomington, Indiana May 2010-October 2010

Lab Assistant

Lab assistant at School of Public and Environmental Affairs

Key responsibilities:

- Performed extraction and Silica Column Chromatography to prepare and separate mixture of various environmental samples following precise established procedure.
- Performed data analysis using Gas Chromatography/Mass Spectrometry (GC/MS).