STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF RTF1 AND INSIGHT INTO ITS ROLE IN TRANSCRIPTIONAL REGULATION

by

Adam Douglas Wier

B.S., Molecular Biology and Biochemistry, Lebanon Valley College, 2009

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Adam D. Wier

It was defended on

August 20, 2014

and approved by

James M. Pipas, Ph.D., Professor, Biological Sciences

Karen M. Arndt, Ph.D., Professor, Biological Sciences

John M Rosenberg, Ph.D., Professor, Biological Sciences

Martin C. Schmidt, Ph.D., Associate Professor, Microbiology and Molecular Genetics

Committee Chair: Andrew P. VanDemark, Ph.D., Associate Professor, Biological Sciences

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF RTF1 AND INSIGHT INTO ITS ROLE IN TRANSCRIPTIONAL REGULATION

Adam D. Wier, PhD

University of Pittsburgh, 2014

Originally discovered in a search for RNA polymerase II-associated factors, the Paf1 complex (Paf1C) is best characterized for its roles in regulating transcription elongation. The complex co-localizes with RNA polymerase II from the promoter to the 3' end of genes and has been linked to a growing list of transcription-related processes including: elongation through chromatin, histone modifications, and recruitment of factors important in transcript maturation. The complex is conserved throughout eukaryotes and is comprised of the proteins Paf1, Ctr9, Cdc73, Rtf1, and Leo1. The domain structures of Paf1C subunits are largely undefined and have few clear homologs, making it difficult to postulate for or localize functions to the individual subunits. To understand mechanistically how Paf1C coordinates its functions and interactions, I took an approach utilizing biochemical, biophysical, and structural techniques to characterize proteins within Paf1C, specifically focusing on the Rtf1 subunit.

The goal of my thesis work was to determine the molecular mechanism by which Rtf1 influences transcription and chromatin structure. To this end I focused on studying different functional domains within Rtf1. I provided a molecular description of how Rtf1 mediates Paf1C recruitment to elongating RNA polymerase II. Recruitment of Rtf1 is controlled by its centrally located Plus3 domain and a direct interaction with the conserved elongation factor Spt5. I solved the co-crystal structure of the human Plus3 domain bound to a phosphorylated C-terminal repeat

of Spt5. The structure revealed the basis for recognition of the repeat motif of Spt5, an important component in the recruitment of regulatory factors to RNA polymerase II. I have performed further structural characterization of Rtf1, studying the N-terminal histone modification domain. To gain insight into the molecular mechanism underlying the domain's function, I successfully crystallized and solved the structure of a minimal region of Rtf1 that is necessary and sufficient for Rtf1-mediated histone modifications. The structure, paired with conservation analysis and genetic phenotype data, have allowed us to identify important surfaces on Rtf1 that function in regulating chromatin structure. Taken together these studies shed new light onto the mechanism by which Paf1C influences the complex network of regulatory interactions required for eukaryotic transcription.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Now that my time as a graduate student here at the University of Pittsburgh is drawing to a close I need to thank many people for helping me on my journey. However, there are far too many people to properly thank in this preface, so I will endeavor to keep things brief.

First and foremost I need to thank my thesis advisor Dr. Andrew VanDemark for giving me the opportunity to conduct research in his laboratory. I am extremely grateful for all of his guidance and support over the years, particularly his advice in all aspects of conducting scientific research from designing thoughtful experiments to expressing my ideas clearly in writing. I am also very thankful for his willingness to allow me to explore teaching and the chance to teach several of his biochemistry lab sessions.

I also need to thank all the past and present members of the VanDemark lab for making my time in lab each day an enjoyable and memorable experience. Swarna, I need to thank for all of her support in lab and her friendship. And for putting up with me calling her by the wrong name for years. Who knew it wasn't Shwarna? Without Aubrey's diligence as the lab manager, it is safe to say the lab would have fallen into chaos. Her concern over me never bringing a lunch has always been appreciated even if I still opted not to eat anything most of the time. Thank you Chris for being a great friend and always keeping me abreast of the new and funny happenings on the internet. Our epic confrontations in Team Fortress 2 will not soon be forgotten. I need to

thank Jenna for being a wonderful bay mate in lab. I will miss our random discussions about homeownership, camping, and all manner of other subjects. Her noble quest to silence the vent above our heads in lab cannot be understated.

I need to express extreme gratitude to Dr. Karen Arndt and all of the members of her lab for their help throughout the years. My success as a graduate student was helped immeasurably by Karen's advice and scientific insight. I cannot envision a better collaborator. Peggy too has been a constant source of advice and a fantastic resource for anything science related. I also need to thank Manasi for her partnership in tackling the mechanistics of Paf1 complex recruitment to chromatin through the Plus3 domain of Rtf1! Branden's help laboring over the histone modification domain and its constant mysteries has been a joy. Much thanks goes to Christine for all of the great scientific discussions and musings on the roles of her collection of histone mutants. Her commitment to making me accept and love coffee has also been appreciated.

I would like to thank my wonderful collaborator Dr. Kara Bernstein and her lab for allowing me to help in their studies of the Shu Complex. Also, Kara's advice and assistance in my quest finding a post-doctoral position has been invaluable. Stephen has my thanks for his company and help learning the technique of fluorescence anisotropy.

To my thesis committee: Jim Pipas, Karen Arndt, John Rosenberg, and Martin Schmidt thank you all for your advice and encouragement throughout my thesis work.

I would be remiss if I did not thank my two brothers Greg and Eric for keeping me going through all of the trials and tribulations of graduate school. All of those long drives to and from Pittsburgh with Greg will be remembered fondly. I could not ask for better friends.

Now that I am about to break onto the third page in this preface, I must remind myself to be brief. Thank you all!

1.0 INTRODUCTION

Gene expression is a fundamental cellular process allowing cells to utilize information stored within DNA to respond to the environment. A cell's gene expression patterns can have profound consequences, as altered transcription levels are characteristic in numerous diseases, most notably cancers. The initial stage of gene expression is transcription, where an RNA polymerase catalyzes the synthesis of an RNA message that is used to generate a functional gene product. For the majority of eukaryotic genes, RNA polymerase II is the enzyme that catalyzes transcription. The polymerase's progression through the transcription cycle is facilitated by a host of accessory factors that aid in every aspect of the process from recruitment of the polymerase to the promoter to termination. The Paf1 complex is one of these accessory factors that plays a vital role in gene expression, mediating a broad range of events, including transcription-coupled histone modifications, maintenance of chromatin, and recruitment of factors important in transcript maturation. In the following chapter, I will touch on important aspects of the transcription cycle and describe the role that the Paf1 complex plays in regulating transcription.

1.1 TRANSCRIPTION

Life relies on the accurate and timely expression of genes. Gene expression begins with the process of transcription and the generation of messenger RNA (mRNA). Transcription revolves around the activity of RNA polymerase II (Pol II) transcribing a gene and making an mRNA transcript of it. While highly complex and multifaceted, transcription can be broken down in a cycle with three main steps, initiation, elongation, and termination (Figure 1, 2) [1]. The cycle starts with initiation, where general transcription factors (GTFs) are recruited to the target gene promoter, which in turn recruit RNA polymerase II (Pol II). These lead to the formation of the pre-initiation complex (PIC) and the synthesis of the 10-12 phosphodiester bonds of the mRNA transcript by RNA Pol II [2, 3]. Once productive mRNA production begins and the promoter is cleared, elongation starts. During elongation the transcribing polymerase travels through the body of the gene generating an mRNA transcript. Here the polymerase relies heavily on the aid of elongation factors, including histone modifying enzymes, chaperones, and remodelers to bypass the obstacles of chromatin structure. As transcription progresses, the growing mRNA transcript undergoes capping and processing events in its maturation process. When RNA Pol II finishes transcribing the last of the coding region of the target gene, the cycle shifts into termination where the transcript is processed and the polymerase is released to begin a new round of transcription. In the coming section, I will go into more detail about each stage of the transcription cycle, describing several of the defining features of each. Further, I will review chromatin structure and the role it plays in transcription with a focus on the mechanisms utilized by the transcribing polymerase to overcome chromatin structure.

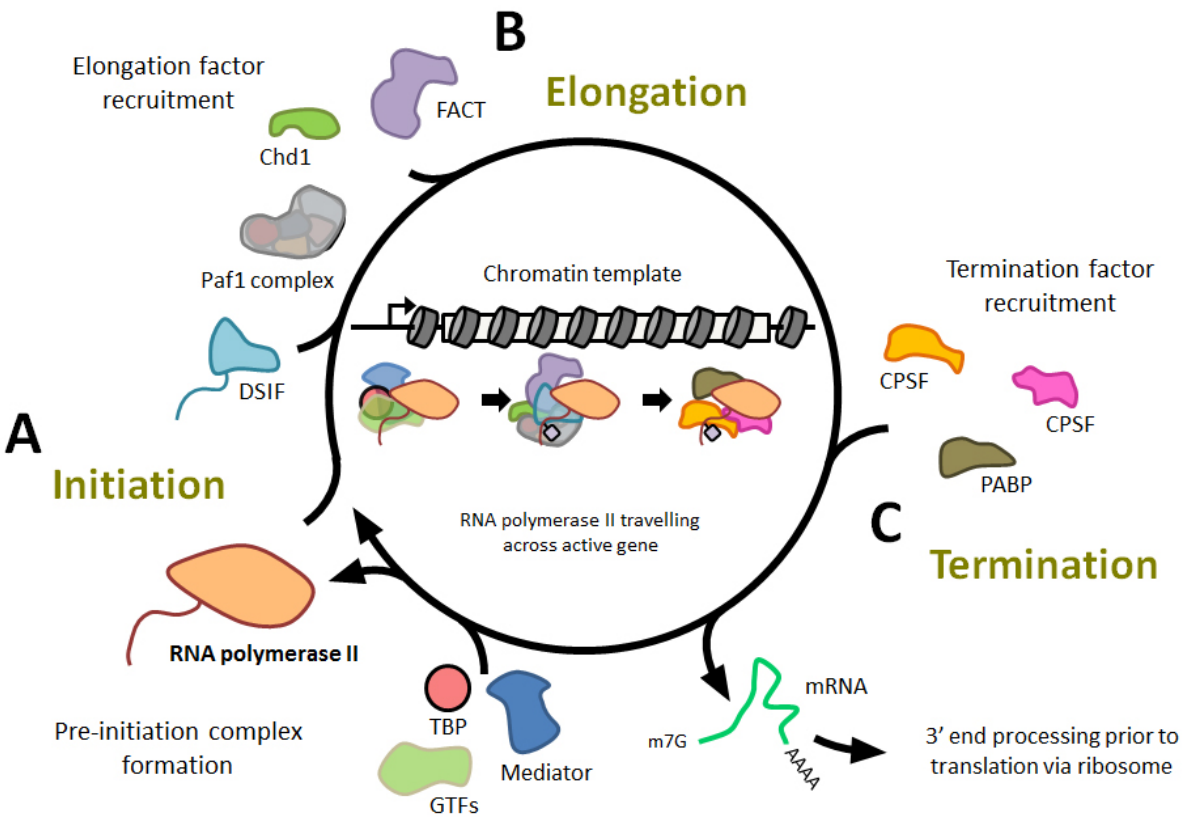


Figure 1. The transcription cycle

The major stages of the transcription cycle are represented in a circular schematic. The progression of RNA polymerase II on a chromatin template is depicted in the center of the circular diagram. A) Initiation phase of transcription: TBP (TATA-binding protein) along with the general transcription factors aid in the recruitment of RNA polymerase II and the formation of a stable pre-initiation complex capable of initiating transcription. B) Elongation phase of transcription: As the polymerase transitions into productive transcription, it is aided by a range of elongation factors including DSIF (DRB sensitivity-inducing factor; Spt4-Spt5 in yeast) and Paf1C (Paf1 complex). Chromatin remodelers like Chd1 (chromodomain helicase DNA binding protein 1) and histone chaperones like FACT (facilitates chromatin transcription; Spt16-Pob3 in yeast) modify chromatin structure to facilitate the polymerase's passage through the gene body. C) Termination phase of transcription: Once the polymerase finishes transcribing the last of the coding region of the target gene, termination and RNA processing factors are recruited. CPSF (cleavage and polyadenylation specificity factor), CstF (cleavage stimulation factor), and PABP (poly(A) binding protein) are important in the proper maturation of mRNA transcripts.

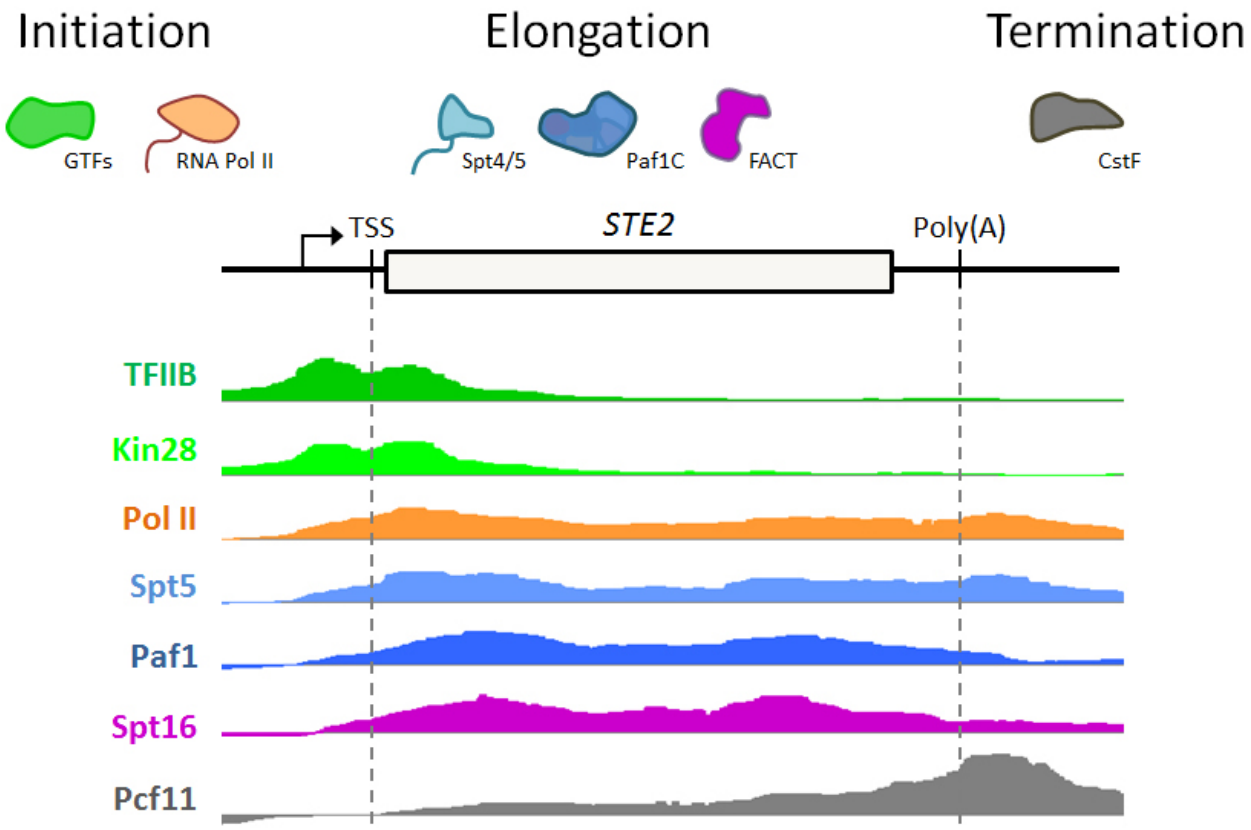


Figure 2. Occupancy profile of RNA Pol II and transcription factors at an active gene

Occupancy profiles for RNA Pol II and transcription factors important in regulating different stages of the transcription cycle are depicted below a diagram of the highly transcribed gene *STE2* in *S. cerevisiae*. The location of the promoter, transcription start site (TSS), gene body, and poly(A) tail site are indicated. Genome-wide ChIP-Seq data from Mayer *et al.* [4] was downloaded from the Saccharomyces Genome Database and viewed in the Integrated Genome Browser. Profiles for each of the transcription factors and RNA Pol II at *STE2* were generated in the Integrated Genome Browser.

1.1.1 Initiation

The transcription cycle begins with initiation, where RNA Pol II is recruited to a target gene along with the general transcription factors to form the pre-initiation complex (Figure 1A, 2),

followed by promoter clearance and the start of productive mRNA synthesis by the polymerase. The process begins at the level of the gene and the specific elements characterizing its core promoter. The RNA pol II core promoter is characterized as the minimal sequence needed to direct the initiation of transcription at a gene (reviewed in [5-7]). The composition of a core promoter varies widely between genes and organisms. While structurally and functionally diverse, they all are comprised of a collection of sequence motifs that are recognized and serve as a nucleation point for the formation of a functional pre-initiation complex composed of RNA Pol II and associated general transcription factors (GTFs; TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH) [8]. In the following section, I will discuss in more detail some of the key elements influencing initiation and some of the regulatory aspects involved.

1.1.1.1 The core promoter and core promoter elements

The core promoter of each gene is composed of a collection of core promoter elements that aid in PIC formation and regulation. In most cases, the core promoter includes the transcription start site and approximately 40 base pairs on either side. The first eukaryotic core promoter element identified was the TATA box which is located -31 to -30 base pairs from the transcription start site and is recognized and bound by TATA-binding protein (TBP) of TFIID [9, 10] (Goldberg. M. L. (1979). PhD thesis, Stanford University, California). Another common element that is recognized by TFIID is the Initiator element (Inr) which encompasses the transcription start site of some promoters [11, 12]. Inr often functions with the Downstream promoter element (DPE) as a counterpart of the TATA box in core promoters [13, 14]. In the case of most mammalian promoters, the TATA box and DPEs are uncommon and are instead replaced by DNA stretches

of CG dinucleotides, called CpG islands [15, 16]. These promoters are typically larger and contain multiple transcription start sites, unlike the focused TATA box promoters.

In addition to the core promoter, there are several other *cis*-acting DNA elements that further regulate gene expression. These include proximal and distal promoter elements, silencers, enhancers, and insulators, all of which represent binding sites for sequence-specific transcription factors (reviewed in [17-19]). Together these *cis*-regulatory elements work hand in hand with *trans*-acting factors, such as the PIC, chromatin remodelers, and histone modifying factors to ensure the target gene's expression pattern is properly executed.

1.1.1.2 Formation of the pre-initiation complex

In vitro experiments using purified factors suggest that the PIC is formed through a stepwise recruitment and assembly of the GTFs and RNA Pol II [8, 20]. First to be recruited in the PIC's sequential assembly is TFIID, a multiprotein complex containing TBP and 12-15 distinct TBP-associated factors (TAFs). TFIID recognizes and associates specifically with different elements of the core promoter including the TATA box, Inr, and DPEs [9, 21]. Once bound, TFIID acts like a scaffold to facilitate recruitment of the remaining members of the functional PIC. In addition to aiding in recruitment, components of TFIID possess other functions to promote productive initiation, including protein kinase and histone modification activity [22]. After TFIID, next in the sequence are TFIIA and TFIIB, both of which aid in stabilizing the promoter-bound TFIID [23, 24]. This allows for the subsequent recruitment of a co-complex of RNA Pol II-TFIIF, further stabilizing the assembling PIC. Finally, the general transcription factors TFIIH and TFIIE along with the Mediator complex join and complete the PIC. TFIIH is critical in shifting the PIC into the open complex formation, the state required for productive transcription

[25]. In an ATP-dependent manner, TFIIH uses its helicase activity (XPB) to melt the promoter and its kinase activity to phosphorylate the largest RNA Pol II subunit, triggering promoter clearance [26]. In contrast to the stepwise assembly model for PIC formation, PIC formation may alternatively occur in a single recruitment step of a large preassembled holoenzyme, including RNA Pol II, GTFs, Mediator subunits, and other regulatory components of the transcription cycle [27-31].

Once a functional PIC is formed on the promoter of a target gene, it must transition into productive transcription. RNA transcripts shorter than 10 nucleotides are unstable and result in a high frequency of premature termination (abortive initiation). Once 10 nucleotides of the transcript are formed, promoter escape is favored and the polymerase transitions into the early stages of elongation. During this transition the transcribing polymerase dissociates from TFIID, TFIIA, TFIIB, and the Mediator complex, leaving them bound to the promoter and free to initiate a new round of transcription (Figure 2) [3, 32].

1.1.1.3 RNA polymerase II C-terminal domain

One of the most important regulatory elements in the transition from initiation into elongation and the regulation of the entire transcription cycle is the post-translational modification state of the C-terminal domain (CTD) of the largest subunit of RNA Pol II, Rpb1 (RNA polymerase II subunit B1). The CTD is comprised of tandem hepta-peptide repeats (25-26 repeats in yeast, 52 in mammals) of the consensus sequence Tyr_1–Ser_2–Pro_3–Thr_4–Ser_5–Pro_6–Ser_7 (YSPTSPS). The motifs are capable of receiving a host of post-translational modifications, including glycosylation (Ser and Thr), peptidyl-prolyl isomerization (Pro), and phosphorylation (Ser, Thr, Tyr)[33]. These modified motifs then serve as a scaffold to recruit factors responsible in regulating each

stage of transcription and regulating co-transcriptional processes such as RNA capping and maturation. Perhaps best documented is the stage specific phosphorylation of serines 2, 5, and 7 of the CTD [34, 35]. During initiation, RNA Pol II is recruited and loaded on the promoter with a hypo-phosphorylated CTD. Formation of a functional PIC catalyzes the kinase activity of TFIIH (CDK7) to phosphorylate serine 5 of the Pol II CTD [36]. This Ser5-phosphorylated CTD is recognized and bound by the 5' mRNA capping enzyme which adds a methylguanosine cap to the end of the nascent RNA transcript, signaling productive transcription [37]. As transcription progresses the phosphorylation status of the CTD changes, seeing a gradual loss of Ser5 phosphorylation along with a concomitant rise in Ser2 phosphorylation. These marks help facilitate RNA Pol II interactions with proteins important for elongation such as Set1, Set2, and Elongator [38-40].

1.1.2 Elongation

After escaping the promoter, the RNA Pol II transitions into the elongation phase of the cycle (reviewed in [41, 42]). During this stage, the polymerase travels through the body of the target gene producing an mRNA transcript of the template. The transition into elongation results in significant changes in the composition of the transcribing RNA Pol II complex. Several of the GTFs dissociate from the polymerase, some including TFIID, TFIIA, and TFIIH remain bound to the promoter to aid in reinitiation (Figure 2) [43]. Others like TFIIF can remain bound to Pol II and serve as elongation factors [44, 45]. The elongation phase presents the transcribing polymerase with a set of obstacles unique from those seen during initiation, including RNA Pol II pausing and chromatin architecture throughout the gene body. To overcome these barriers, the

transcribing polymerase is aided by numerous elongation factors, chromatin remodelers, and histone modification complexes (Figure 1B).

1.1.2.1 Promoter pausing of RNA polymerase II

In higher eukaryotes, early elongation is often characterized by promoter-proximal and distal-nucleosome pausing, where the polymerase pauses after transcribing 20-80 nucleotides of the target gene [46-48]. Pausing within the promoter-proximal region (+40) correlates with the strength of the core promoter elements (similarity to consensus motifs) [49, 50]. Distal pausing (+80) is caused by a physical barrier created by the first nucleosome in the polymerase's path [51]. Once paused, pausing factors NELF (negative elongation factor) and DSIF (DRB sensitivity–inducing factor) associate with Pol II and stabilize it *in vivo* [52]. Mechanistically how these pausing factors negatively influence transcription is still uncertain; however the ability of both factors to binding RNA may be involved. DSIF is capable of binding the emerging nascent RNA from RNA Pol II. Further, the crystal structure of the archaeal RNA polymerase bound to DSIF shows it binds the RNA Pol II clamp domain, closing the polymerase's active center cleft and interacting with the RNA transcript, possibly modulating RNA Pol II processivity [53].

Genome-wide analysis using GRO-seq (global run-on sequencing) indicates that the RNA Pol II pausing is widespread, with an estimated Pol II pausing seen on 30-40% of active genes in humans and mice [46, 54]. With Pol II pausing being so prevalent in mammalian cells, proper regulation of the release of the paused polymerase must be critical for proper gene expression. The transition of paused Pol II to one engaged in productive elongation requires the activity of the protein kinase complex P-TEFb (positive transcription elongation factor b).

Composed of a heterodimer of CYC-T (cyclin T) and CDK9 (cyclin-dependent kinase 9) subunits [55] P-TEFb's major function is to phosphorylate the CTD of Pol II at Ser2 [56], and the pausing factors DSIF and NELF [57-59]. Phosphorylation of NELF and DSIF relieves their negative influence on transcription, causing NELF to be released and DSIF to become a positive elongation factor, signaling productive transcription to begin anew.

1.1.3 Termination

The final stage in the transcription cycle is termination where the RNA transcript is finished being transcribed and RNA Pol II is released from the template DNA and is free to begin subsequent rounds of transcription (reviewed in [60, 61]). The process is coupled to 3'-end processing of the pre-RNA [62-65] (Figure 1C, 2). Proper termination is critical for proper expression of the target gene and ensures the polymerase does not interfere with downstream transcriptional units [66]. Termination can be accomplished through several pathways, dictated by the gene in question and the termination factors present [61, 67]. Two of the best characterized pathways utilized are the poly(A)-dependent pathway and the Nrd1-Nab3-Sen1-dependent pathway. In the following section I will briefly discuss the poly(A)-dependent and the Nrd1-Nab3-Sen1-dependent termination pathways, as well as touch on the role of gene looping in transcription termination.

1.1.3.1 Poly(A)-dependent termination

In eukaryotes, the majority of protein-coding genes are terminated in the poly(A)-dependent pathway, where RNA Pol II termination is directly coupled with 3'-processing of the

transcript. In the pathway, the 3’ end of the nascent RNA transcript must be cleaved and a poly(A) tail attached, along with degradation of the downstream Pol II-associated RNA [68, 69]. In eukaryotes, the pre-mRNA of most protein coding genes contains a conserved hexanucleotide, AAUAAA sequence followed by a loosely conserved U- or GU-rich region. These *cis*-acting RNA elements are specifically recognized by the 3’-end processing machinery and help direct cleavage and polyadenylation. Two of the primary protein complexes facilitating RNA processing are the cleavage and polyadenylation specificity factor (CPSF; CPF in yeast) and the cleavage stimulation factor (CstF; CFIA in yeast). Both complexes are recruited to the 3’-end of genes through an interaction with the polymerase itself, CstF specifically recognizing the Ser2 phosphorylated CTD [70, 71]. Once recruited, CPSF recognizes the AAUAAA motif, while CstF directly recognizes the U/GU-rich element. Endonucleolytic cleavage is then catalyzed by CPSF [72, 73] and the poly(A)-binding protein (PABP) is recruited to synthesize the poly(A) tail on the maturing RNA transcript [74].

Soon after the AAUAAA sequence in the nascent RNA is transcribed and recognized by CPSF, RNA Pol II processivity decreases dramatically and it pauses [75, 76]. Once the nascent RNA transcript has been cleaved by CPSF and a poly(A) tail incorporated, the RNA fragment downstream of the cleavage site must be degraded and RNA Pol II released to complete termination. CPSF cleavage of the nascent RNA results in the exposure of a new 5’ RNA end, allowing the 5’-3’ exonuclease RNA-trafficking protein 1 (Rat1; Xrn2 in humans) to enter and begin degrading the RNA transcript [77-79]. It is thought that upon reaching the paused polymerase, Rat1 collides with Pol II, forcing Pol II to dissociate and hence promoting termination [80-82].

1.1.3.2 Sen1-dependent termination

In yeast, an alternative termination pathway, distinct from the poly(A)-dependent one, exists for most non-coding RNAs, focusing around the Nrd1-Nab3-Sen1 complex. Contrary to what is seen in protein-coding genes, most non-coding genes, including small nuclear RNAs (snRNA) and small nucleolar RNAs (snoRNA), do not contain a poly(A) tail when fully matured. Instead, their 3'-ends are created through endoribonucleolytic cleavage and/or exoribonucleolytic trimming by the nuclear exosome–TRAMP complex [83]. This alternative termination pathway requires the essential RNA-binding protein Nrd1 (Nuclear pre-mRNA down-regulation), the associate DNA/RNA helicase Sen1 (Splicing endonuclease 1), and Nab3 (Nuclear polyadenylated RNA-binding 3) [84-87]. Similar to what is seen with CPSF and CstF in the poly(A)-termination pathway, Nrd1 and Nab3 recognize specific RNA sequences at the 3'-end of the nascent RNA transcript. Once bound to the nascent RNA, it is thought that the complex promotes termination by unwinding the RNA-DNA duplex in the active site of RNA Pol II [85].

1.1.4 Chromatin architecture

In eukaryotes transcription and all DNA templated processes (recombination, replication, DNA repair) take place in the context of chromatin [88]. Chromatin is a highly ordered protein-DNA assembly that acts to fold and compact DNA to fit within the boundaries of a cell's nucleus (reviewed in [89]). The first level of chromatin's organizational hierarchy is the nucleosome. Functioning as the basic repeating unit of chromatin, the nucleosome is composed of the nucleosome core particle, 145-147 base pairs of DNA, and the linker histone H1. The nucleosome core particle contains an octamer of four different histone proteins (H2A, H2B, H3,

H4) around which 145-147 base pairs of DNA are wrapped in 1.65 turns of a left-handed superhelix (Figure 3A) [90, 91]. Each of the histones in the core particle shares a conserved structural motif, the histone fold (Figure 3B). The fold averages 70 amino acids and is comprised of three alpha helices (α1, α2, α3) connected by short loops (L1, L2). The central α2 helix is longer than the neighboring α1 and α2 helices, and together they form a shallow groove that allows for dimerization with other histone folds [90].

Each of the core histone proteins pair with one another to create H2A-H2B and H3-H4 heterodimers. When dimerized, two histone folds align antiparallel to one another forming a large interaction surface with their aligned α2 helices (Figure 3C). These heterodimers assemble into the nucleosome octamer through the formation of four-helix bundle arrangements created with the C-terminal halves of the α2 helices and the α3 helices of H3, H2B, and H4 [90]. First a tetramer of H3-H4 dimers is formed through a single association between both H3 histones. The H2A-H2B dimers are then able to associate with the H3-H4 tetramer through an interaction between H4 and H2B. Under physiological conditions, the histone octamer is not stable without DNA. Instead, the H3-H4 tetramer and the H2A-H2B dimers are the biological subunits, only assembling into the octamer when DNA is present [89].

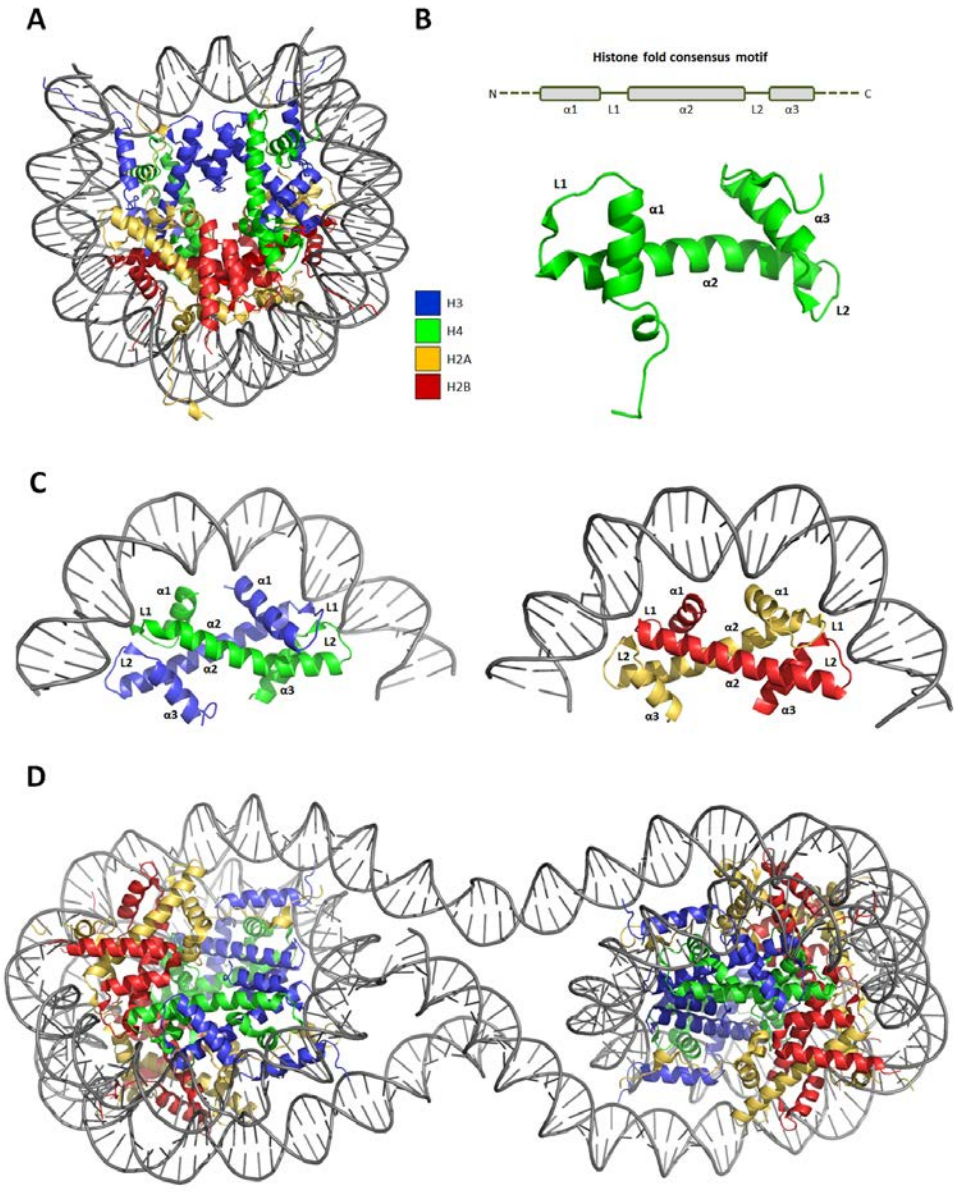


Figure 3. The structure of the nucleosome core particle

A) X-ray crystal structure of the nucleosome (PDB id: 1AOI). The histones and DNA (grey) are represented in cartoon form (H3, blue; H4, green; H2A, yellow; H2B, red). B) Diagram of a consensus histone fold motif along with the structure of the histone fold of H4. C) Structure of H3-H4 and H2A-H2B histone pairs bound to DNA. The N- and C- termini for each of the histones were omitted for clarity. D) X-ray crystal structure of the tetranucleosome (PDB id: 1ZBB).

The primary DNA binding interface on the nucleosome lies along the histone fold dimers in a continuous basic groove (Figure 3C) [90]. Each of the heterodimers organizes roughly 30 base pairs of DNA, utilizing three DNA-binding motifs. When arranged as a heterodimer, two of the DNA-binding sites are created on either end of the dimer, formed by the juxtaposition of the L1 loop of one histone with the L2 loop of the other (Figure 3C). The third binding site is centrally located in the dimer and is formed by the alignment of the N-termini of the α1 helices of the dimerized histones (Figure 3C) [90]. These DNA-binding interfaces are capable of binding DNA irrespective of the DNA sequence. There are no direct interactions between the histones and the DNA bases; instead contacts are made through hydrogen bonding of the protein's main chain amide atoms with the phosphate backbone of the DNA helix. In addition to L1L2 and α1α1 DNA-binding motifs on the histone fold dimers, the entry and exit points of nucleosomal DNA are recognized by the N-terminal extensions of histone H3 [90].

Outside of the conserved histone fold, each of the core histones contains more variable structured extensions and flexible tails. The structured extensions are typically responsible for protein-protein interactions within the octamer [90, 91]. The C-terminal extension of H2A (also known as the docking domain) for instance, acts to stabilize the octamer through bridging another interaction with the H3-H4 tetramer and positioning the H3 extension to interact with nucleosomal DNA. Unlike the extensions, the flexible histone tails do not form an integral part of the nucleosome octamer and extend outside the body of the nucleosome [90, 91]. The tails are typically found on the N-terminus of the core histones and are the site of an extensive array of post-translation modifications, including acetylation, phosphorylation, ubiquitylation, methylation, and ADP-ribosylation (reviewed in [92]).

1.1.4.1 Higher order chromatin structure

Nucleosomes are connected to one another through short stretches of linker DNA, forming nucleosomal arrays (primary structure, 10-nm fiber). Linker DNA is recognized by linker histone H1, and through an interaction with the histone octamer, H1 helps facilitate chromatin arrangement and compaction [93-95]. Nucleosomal arrays undergo further compaction, through short-range interactions between neighboring nucleosomes forming 30-nm chromatin fibers (secondary structure) [93, 95-97]. In the 30-nm fiber, a tetranucleosome structure is formed where four nucleosomes zigzag back and forth, creating two stacks of two nucleosome cores (Figure 3D) [98, 99]. These tetranucleosomal units stack together end on end to form the 30-nm fiber [99]. The main interaction mediating tetranucleosomal stacking is thought to be through internucleosomal contacts between the N-terminal H4 tail of one stack and the H2A-H2B interface (H2A-H2B acidic patch) of another [90, 99, 100]. Subsequent fiber-fiber interactions continue to compact the genome, forming the complex three-dimensional assemblages of higher-order structures seen in chromatin (tertiary structure) [93].

1.1.4.2 RNA Pol II elongation through chromatin

Chromatin structure acts as a significant barrier for the transcribing RNA Pol II during transcription [101, 102]. The polymerases progression through the body of the gene is assisted by a variety of factors that act to weaken or remove the nucleosome obstacle, including nucleosome remodeling, histone modifications, and the incorporation of histone variants into nucleosomes. In the following section I will briefly discuss some of these factors and how they affect transcription.

Chromatin remodelers

ATP-dependent chromatin remodeling complexes play an important role allowing RNA Pol II to bypass the nucleosome barrier during transcription. These remodeling complexes utilize the energy generated by ATP hydrolysis to modify chromatin structure through assembling, sliding, and displacing nucleosomes. All chromatin remodelers contain a core catalytic subunit with a conserved Swi2/Snf2 ATPase domain from superfamily 2 (SF2) DNA translocases [103]. Chromatin remodeling complexes are highly conserved from yeast to humans and can be broadly classified into four main families based on additional domains and motifs within their ATPase subunits. These chromatin remodeling families are SWI-SNF, ISWI, CHD, and INO80/SWR (reviewed in [104]).

Remodelers in the SWI-SNF (SWItching and Sucrose Non-Fermenting) family are large complexes of 8-14 subunits that typically exceed a molecular weight of over 1 MDa [104, 105]. In *S. cerevisiae*, the SWI-SNF remodeling complex exists in two forms, SWI-SNF and RSC (Remodel the Structure of Chromatin). Humans contain an ortholog of RSC known as PBAF (Polybromo and BRG1-Associated Factors) [104]. Remodelers of this family typically function by sliding and ejecting nucleosomes, affecting processes such as DNA replication, transcription, and DNA repair. RSC for instance, is capable of stimulating RNA Pol II elongation *in vitro* through a reconstituted transcription system containing mononucleosomes [106]. Further, RSC is known to be recruited and required for RNA Pol II transcription through the body of stress-activated genes *in vivo* [107].

The ISWI (Imitation SWItch) family of remodelers are much smaller than those of the SWI-SNF family, containing only 2 to 4 subunits [104, 105]. ISWI remodelers contain one or two catalytic subunits containing a conserved Swi2/Snf2 ATPase domain along with a SANT

(SWI3, ADA2, NCoR, TFIIIB) and SLIDE (SANT-like ISWI) domain characteristic of the family. Most eukaryotes utilize multiple ISWI remodeling complexes that differ in the inclusion of specialized accessory proteins that impart different functional domains. Some of the ISWI variants in humans include CHRAC (Chromatin Accessibility Complex), ACF (ATP-utilitzing Chromatin assembly and remodeling Factor), and NURF (Nucleosome Remodeling Factor). In yeast ISWI remodeling complexes include ISW1a, ISW1b, and ISW2. Many ISWI remodelers are capable of evenly spacing nucleosomes in an array and play important roles in chromatin condensation and gene silencing [104, 105].

The CHD (Chromodomain Helicase DNA-Binding) family of remodelers combines 1 to 10 subunits depending upon the organism and the specific complex [104, 105]. The founding and the best characterized member of the family is the monomeric Chd1 [108, 109]. Chd1 and the other catalytic subunits in the family contain two tandemly arranged chromodomains followed by an ATPase domain. Similar to ISWI remodelers, Chd1 is able to generate evenly space nucleosomal arrays by sliding nucleosomes [110, 111]. Additionally, Chd1 in known to physically interact with the elongation factors DSIF, FACT, and the Paf1 complex (Polymerase-associated factor 1 complex) and is found localized to sites of active transcription [112-116].

Initially discovered in yeast [117, 118], chromatin remodelers from the INO80 (Inositol Requiring 80) family are comprised of 10 or more subunits [104, 119]. Each complex in the family shares a conserved catalytic core containing a characteristic ‘split’ ATPase domain, with a long insertion in the middle of the ATPase domain. This insertion functions as a binding site for other members of the remodeling complexes including AAA-ATPases (Rvb1/2) and actin-related proteins (ARP) [104, 119]. Remodelers of the INO80 family are important in a diverse range of functions, including transcriptional regulation, DNA repair, and DNA replication. In yeast, one

of the family members, SWR1 (Swi2/Snf2-related ATPase), is able to restructure nucleosomes by removing canonical H2A-H2B dimers and depositing H2A.Z-H2B dimers in their place [104, 120, 121].

Histone chaperones

In addition to chromatin remodelers, there is a well-established role of histone chaperones in stimulating transcription elongation. One of these critical in elongation is the histone chaperone complex, FACT (facilitates chromatin transcription) [122]. Comprised of two subunits, SPT16 and SSRP1 (Spt16 and Pob3 in yeast), FACT promotes the dissociation of the H2A-H2B dimer from the nucleosome, allowing the polymerase to bypass the remaining histone hexamer. Once Pol II traverses the barrier, FACT can reassemble the nucleosome in the polymerase's wake [123, 124]. Another important chaperone that travels with the transcribing Pol II is Spt6. Spt6 interacts with histones H3 and H4 and is important for maintenance of chromatin structure [125]. Loss of Spt6 activity in yeast results in reduced nucleosome occupancy over the coding region of several genes. This nucleosome depletion results in aberrant transcription initiation starting within gene coding regions from cryptic initiation start sites [126]. The precise mechanisms utilized by histone chaperones to remove and redeposit nucleosomes on DNA are not clear, however several possibilities exists. One possibility is that they may directly modulate nucleosome stability, promoting nucleosome disassembly. Alternatively, they may function as a reservoir to soak up histones released during elongation or as a pool of histones when reassembling nucleosomes post-elongation.

Covalent histone modification

In addition to histone remodelers and chaperones, nucleosome structure is further modulated through covalent modification of histones themselves (reviewed in [92]). Histones are subject to a wide range of post-translation modifications including, acetylation, methylation, ubiquitylation, phosphorylation, and SUMOylation. The majority of these post-translational modifications are localized to the unstructured amino-terminal tails of the histones that lie outside the body of the nucleosome. These modifications are known to affect elongation by acting as a binding surface to recruit regulatory factors and by directly altering chromatin structure through modifying electrostatics and internucleosomal contacts.

Histone acetylation was the first modification to be discovered and is typically associated with active transcription [127, 128]. Occurring on lysine residues, the mark is thought to effectively weaken charge-dependent interactions between histones and DNA through charge neutralization [129-131]. This in turn increases DNA accessibility to RNA Pol II and any other regulatory factor. Studies using combinations of lysine-to-arginine substitutions in H4 and H3 tails support the idea that the cumulative charge neutralization of multiple lysine acetylation marks primarily determines the transcriptional effect of acetylation [130, 131]. An increase in histone acetylation in the body of a gene is typically correlated with transcription activity [132]. This increase in histone acetylation often pales in comparison to the acetylation signal seen in the promoter region of active genes [133, 134]. Despite this, acetylation is critical for proper elongation as indicated by the enrichment of histone acetyltransferases (HATs) and histone deacetylases (HDACs) in the coding regions of genes [135-137]. Further, diminishing histone acetylation levels in yeast by removing the Gcn5 and Elp3 HATs results in reduced transcription

levels at constitutively active genes and correlates well with lower acetylation levels in their coding region [138].

Histone methylation is also another prevalent histone modification associated with transcriptional elongation. Lysine residues are capable of receiving up to three methyl groups, however unlike acetylation the marks do not affect the residue's charge, making it more difficult to access the modification's effect on nucleosome dynamics. Histone methylation's role in transcriptional regulation is complex, regulating both activation (H3K4 and H3K36) and repression (H3K9 and H3K27) of transcription. The methylated lysine moiety is known to serve as a tag for effector proteins containing methyl-binding domains such as Tudor, MBP, chromo, and PHD domains [139, 140], suggesting the modification serves as a regulatory module. This notion is supported by the well-characterized recruitment of the Rpd3S HDAC complex in yeast to di- and trimethylated H3K36 [141], where it catalyzes the elongation-associated deacetylation of nucleosomes [135, 142-144]. Further, H3K9 and H3K27 methylation function in heterochromatin formation and Polycomb silencing through aiding in recruitment of HP1 (Heterochromatin Protein 1) and Polycomb [145].

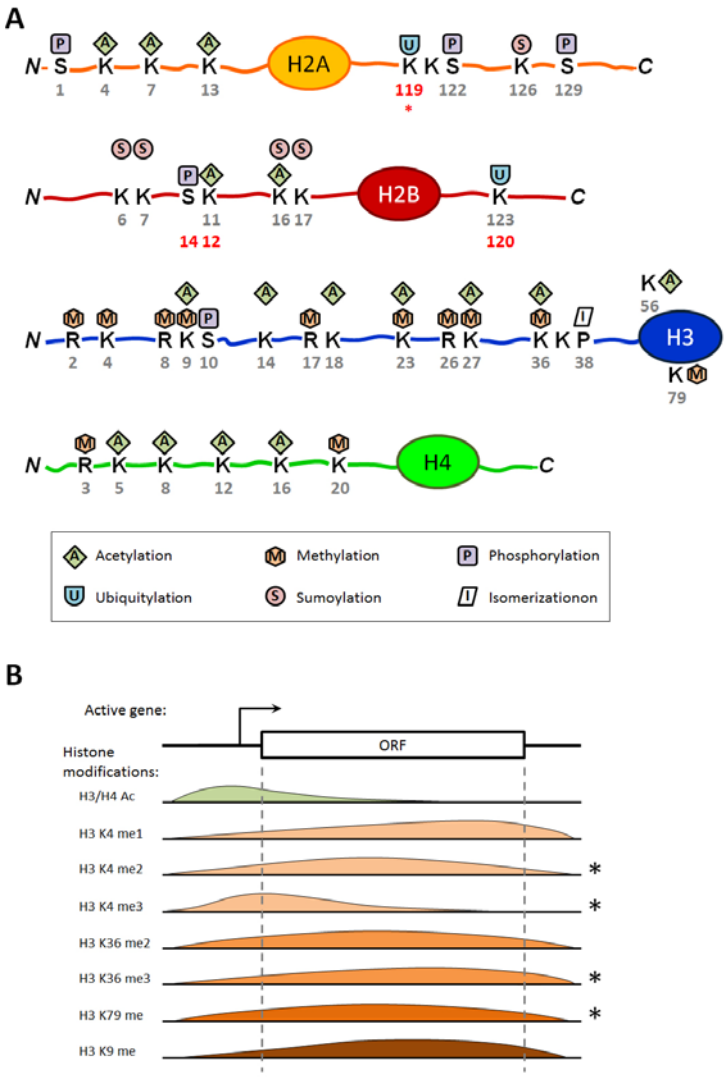


Figure 4. Histone post-translational modifications

A) Schematic of post-translational modifications on the *S. cerevisiae* histones. The globular histone fold within each of the proteins is represented by an oval. The specific residues that are modified are labeled and numbered. Alternative residue numbers that refer to mammalian histones are indicated (red). Ubiquitination of histone H2A on K119 is not observed in yeast (*). Figure adapted from [146]. B) Genome-wide distribution patterns of select histone modifications. The distribution of each of the histone modifications is mapped under the schematic of an arbitrary gene, depicting its promoter region and open reading frame (ORF). Histone modifications requiring Paf1C are marked with an asterix (*). Figure adapted from [147].

Histone monoubiquitylation also plays an important role in regulating histone dynamics. Unlike protein polyubiquitylation which serves as a signal for protein degradation by the 26S proteasome, monoubiquitylation acts as a signal to regulate many processes including transcription and DNA repair (reviewed in: [148]). In transcriptional regulation, H2B lysine 123 (K123 in S. cerevisiae; K120 in mammals) monoubiquitylation has been well characterized, and is known to play roles in both transcriptional activation and repression [148-151]. The mark is found at the promoter and coding regions of genes and is coupled with RNA Pol II transcription [152-154]. In *S. cerevisiae*, monoubiquitylation is carried out by the Rad6–Bre1 ubiquitin conjugase–ligase complex [155-157]. The mark is a prerequisite for the downstream di- and trimethylation of H3 K4 and K79 by the Set1/COMPASS and Dot1 methyltransferases, respectively [158-160]. These di- and trimethylation marks facilitate the recruitment and stimulate the activity of histone acetyltransferase and deacetylase complexes, controlling gene expression by modulating the histone acetylation patterns of genes [161-164].

1.2 THE POLYMERASE-ASSOCIATED FACTOR 1 COMPLEX

The Paf1 complex was initially identified in a search for accessory factors that associate with RNA Pol II in *S. cerevisiae*. RNA Pol II was affinity purified from a yeast transcription extract using an antibody directed against the CTD of Rpb1, the largest subunit of RNA Pol II. Surprisingly, a fraction of polymerase-associated proteins distinct from the fraction containing the activator responsive holoenzyme was discovered. In addition to containing known elongation factors like TFIIS, the fraction also contained novel proteins including Paf1 (polymerase

associated factor 1) and Cdc73 (cell division cycle protein 73) [165, 166]. Additional proteins associated in complex with Paf1 were discovered in subsequent studies and include Ctr9, Rtf1, Cdc73, and Leo1 [114, 167-169]. Each of these proteins was identified earlier via yeast genetic studies and was thought to have a role in transcription. In human Paf1C, an additional protein, Ski8, was found to be part of the complex [170]; Ski8 is known to play a role in exosome-mediated mRNA decay when part of the Ski complex [171]. Paf1C is evolutionarily conserved in eukaryotes and colocalizes with RNA Pol II at active genes from the transcriptional start site to the poly(A) site before dissociating (Figure 2) [4, 172]. Consistent with its association with RNA Pol II and localization along the body of active genes, the complex is known to regulate transcription elongation (Figure 5).

Initial studies of the complex showed that its disruption resulted in numerous genetic phenotypes associated with transcription elongation defects. For instance, deletion of any Paf1C subunit causes growth sensitivity to 6-azauracil (6-AU), a nucleotide base analog that diminishes intracellular ribonucleotide levels [169, 173]. A similar phenotype is seen when Paf1C subunits are deleted and mutants are grown in the presence of mycophenolic acid (MPA) [169, 173], a metabolite that interferes with guanine nucleotide synthesis [174]. In the presence of 6-AU and MPA, Pol II processivity becomes more reliant on elongation factors to overcome low nucleotide levels. Paf1C mutants are more sensitive to 6-AU and MPA than wild type yeast, suggesting a role in elongation. Further, deletion of Paf1C subunits results in growth sensitivity to a host of compounds that evoke cellular stress responses, such as high temperature, caffeine, rapamycin, and hygromycin [173, 175, 176].

Aside from its association with RNA Pol II, Paf1C has strong connections to numerous transcriptional regulatory factors and elongation factors. For instance, Paf1C has genetic and

physical interactions with elongation complexes such as FACT (Spt16-Pob3 in yeast) and DSIF (Spt4-Spt5 in yeast) [114, 169, 173]. Additionally, Paf1C interacts with the monomeric, ATP-dependent chromatin remodeler Chd1 [113, 177], which associates along the bodies of active genes [112, 113] and is important for fast and efficient nucleosome sliding [178-180]. In humans, Paf1C physically interacts with the general transcriptional elongation factor TFIIS through the Paf1 and Leo1 subunits [181]. Together Paf1C and TFIIS act synergistically to facilitate transcription elongation on chromatin templates *in vitro* [181]. Mutations in genes encoding Paf1C subunits cause a range of synthetic phenotypes, including lethality, when deleted in combination with genes, *SRB5*, *CTK1*, *FCP1*, and *POB3* [169, 173].

While Paf1C was initially thought to act primarily as a transcription elongation complex, in recent years, it has become increasingly evident that it functions more broadly, affecting nearly every aspect of RNA Pol II transcriptional regulation (Figure 5). Some of these functions include regulating proper phosphorylation levels of the CTD of RNA Pol II [182, 183], recruitment of termination and mRNA maturation factors [183-185], maintenance of chromatin during transcription [186], and transcription-coupled posttranslational modifications of histones [187-191].

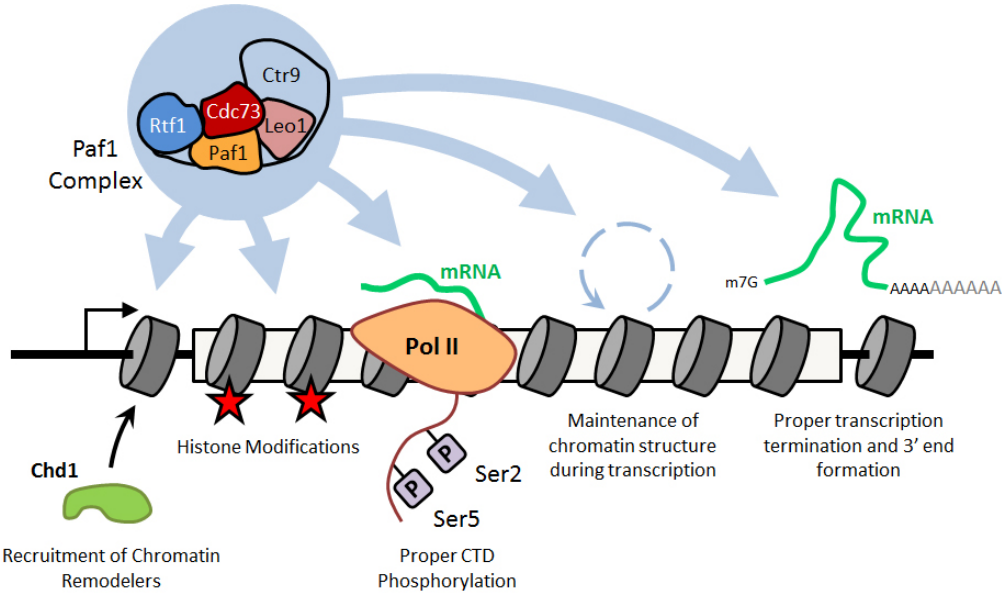


Figure 5. Summary of the functional roles of Paf1C in transcription

Paf1C is required for recruitment of chromatin remodelers, several transcription-coupled histone posttranslational modifications, proper phosphorylation levels of the CTD of RNA Pol II, maintenance of chromatin structure during transcription, and the recruitment of termination and mRNA maturation factors.

Consistent with the expanding role of Paf1C in regulating gene expression, loss of Paf1 in yeast results in positive or negative expression changes in approximately 15-20% of genes [192]. For those genes affected, there is an overrepresentation of essential genes, many of which are required for rRNA processing [192]. In both yeast and humans, Paf1C is required for full expression of a subset of cell cycle-regulated genes [193, 194], including *CLN1* and *RNR1*. The Paf1 complex's importance in regulating gene expression is further emphasized by its connection to a growing list of cancers (thyroid, breast, liver, pancreatic, and others) associated with alterations of subunits within the complex [195, 196]. Additionally, Paf1C is involved in cellular responses to viral infection [197, 198], and has an important role in maintaining embryonic stem cell pluripotency and preventing lineage specification [199-201].

1.2.1 The role of Paf1C in transcription regulation

The Paf1 complex is implicated in regulating processes in all stages of the transcription cycle. Surprisingly, none of the complex members have any known enzymatic activity. This suggests the complex's regulatory role is that of a scaffold mediating protein-protein interactions and recruitment of other regulatory factors. Here I summarize some of the major functions attributed to Paf1C and briefly touch on their importance in proper gene expression.

1.2.1.1 Paf1 complex-dependent histone modifications

One of the Paf1 complex's best characterized roles is its requirement for a number of transcription-coupled histones posttranslational modifications (Figure 4). These marks act as important transcriptional regulators, influencing chromatin accessibility at genes by facilitating the recruitment and activities of different proteins, including chromatin remodelers, histone-modifying enzymes, and elongation factors. Here I summarize the histone modifications regulated by Paf1C.

H2B K123 ubiquitylation and H3 K4 and K79 di- and trimethylation

The Paf1 and Rtf1 subunits of Paf1C are required for the monoubiquitylation of histone H2B on K123 in yeast (K120 in human) by the Rad6 ubiquitin conjugase and Bre1 ubiquitin ligase [153, 191, 202]. Monoubiquitylation of H2B K123 is a prerequisite for the di- and trimethylation of histone H3 K4 and K79 by the methyltransferases Set1 (component of the COMPASS complex) and Dot1, respectively [158-160, 203]. Paf1C is thought to mediate H2B ubiquitylation by facilitating recruitment of the Rad6-Bre1 ubiquitin ligase to chromatin [153,

162, 189, 191]. In yeast H2B K123 ubquitylation is believed to be primarily controlled by the Rtf1 subunit of Paf1C. Yeast strains lacking Paf1 see a substantial reduction in Rtf1 protein levels, suggesting that Paf1's role in the modification is a structural one, maintaining Paf1C integrity and Rtf1 localization [182]. The specific region of Rtf1 required for these histone modifications has been narrowed down to amino acids 62-152. Known as the histone modification domain (HMD), this region is necessary and sufficient for H2B K123 ubiquitylation *in vivo* [177, 204, 205]. The HMD is one of the most highly conserved regions in Rtf1 amongst eukaryotes and contains several invariant residues that are required for H2B K123 ubiquitylation, including glutamate 104 (E104) [204].

Similar to what is observed in yeast, human Paf1C also regulates proper gene expression through mediating H2B K120 monoubiquitylation and the subsequent di- and tri-methylation of H3 K4 and K79 [154, 206, 207]. Human Paf1C also helps recruit Rad6 and Bre1 to monoubiquitylate H2B on lysine 120 [154, 206]. Paf1C's role in recruiting the Rad6-Bre1 complex is thought to be mediated through a direct interaction of Paf1 with Bre1 [206]. H2B K120 ubiquitylation then facilities the downstream methylation of H3 K4 and K79 by Set1 and Dot1, respectively [206, 207].

H2B K123 ubiquitylation and H3 K4 and K79 methylation are enriched along the coding regions of active genes [153, 208, 209] where they play both positive and negative roles in regulating transcription. During elongation, the positive role of H2B ubiquitylation on transcription is mediated in part by a connection to the histone chaperone FACT and its ability to control nucleosome dynamics [154, 210]. *In vitro* studies using a fully reconstituted chromatin-transcription system have shown that H2B ubiquitylation enhances RNA Pol II elongation rate through a chromatin template, in a manner requiring the histone chaperone FACT [154]. *In vivo*,

H2B ubiquitylation regulates nucleosome reassembly in the wake of transcribing RNA Pol II at galactose-inducible genes, in a mechanism requiring the histone chaperone Spt16 (component of FACT complex) [210]. Further, in humans H2B ubiquitylation is known to stimulate the expression of developmental HOX genes [211] and genes regulating pluripotency in embryonic stem cells [199].

A microarray study in yeast utilizing a strain unable to support H2B ubiquitylation indicates that the modification is important for regulating proper gene expression of a subset of protein coding genes (*htb1*-K123R substitution). For those genes affected, nearly 75% saw an increase in expression, suggesting H2B ubiquitylation has a predominantly negative influence on transcription [212]. This notion of repression is supported by studies showing that the H2B K123 deubiquitylating enzyme, Ubp8, is needed for full expression of several genes, including *ADH2*, *GAL1*, *GAL10*, and *SUC2* [149, 152, 212, 213]. H2B K123 ubiquitylation is also known to be important in mediating Paf1C-dependent repression of a subset of yeast genes, including the gene *ARG1* [214]. Furthermore, Paf1C and histone modifications dependent upon the complex play a role in mediating heterochromatic gene silencing, in a mechanism involving silent information regulator proteins (Sir) [159, 177, 189, 215-218].

H3 K36 trimethylation

The Paf1 and Ctr9 subunits of Paf1C are required for trimethylation of H3 K36 [187]. H3 K36 methylation is found throughout the bodies of active genes and is enriched towards their 3'-ends [219-222], where it is associated with an inhibitory role in transcription. The mark is catalyzed by the histone methyltransferase Set2, which is recruited to actively transcribed genes through a direct interaction with the hyperphosphorylated form of the transcribing polymerase

[220, 223]. H3 K36 methylation plays an important role in regulating histone acetylation levels at active genes by facilitating the recruitment and activity of the HDAC Rpd3S. Specifically, H3 K36 is recognized by the chromodomain of Eaf3, a subunit of the Rpd3S HDAC [135, 142-144]. Once Rpd3S is recruited to the actively transcribing polymerase, the HDAC mediates histone deacetylation within the 3' coding regions of genes and in turn represses improper transcription at cryptic promoters. Interestingly, recruitment and activation of Rpd3S only requires dimethylation of H3 K36 [143], making the exact role of Paf1C and H3 K36 trimethylation in repressing internal cryptic transcription uncertain. The two may function independently of the Set2-Rpd3S pathway, utilizing a pathway involving the cyclin-dependent protein kinase, Bur1-Bur2 [187].

1.2.1.2 RNA Pol II CTD phosphorylation

Paf1C is also important in mediating proper levels of RNA Pol II CTD phosphorylation. Specifically, in yeast loss of Paf1, Ctr9 or to a lesser extent, Cdc73, and Rtf1 results in a reduction in CTD phosphorylation at Ser2 [182, 183]. While the current mechanism for this requirement is unclear, the effect may be related to proper recruitment of the major Ser2 kinase, Ctk1. Both Paf1C and Ctk1 share similar occupancy profiles along transcriptionally active genes. Due to the fact that Ctk1's recruitment to transcribing polymerase lies downstream of Paf1C, it may be possible that Paf1C loss affects Ctk1 activity or recruitment [172, 224]. This loss in Ser2 phosphorylation levels likely affects downstream recruitment of termination factors and RNA 3'-end processing factors [183]. Further it likely deregulates the acetylation-deacetylation cycle during transcription, as Ser2 phosphorylation aids in recruiting the methyltransferase Set2 which methylates H3 K36 [141, 220, 223, 225], a mark that is recognized by different HDACs [143].

1.2.1.3 Transcription termination and RNA 3’ end processing

Aside from Paf1C’s roles in regulating transcription elongation, the complex is also tied to proper transcription termination and RNA 3’-end formation. Loss of Paf1C subunits results in changes in utilization of 3’-end formation sites in a subset of yeast genes [192]. These changes result in read-through errors at poly(A) sites, leading to extended transcripts that are subject to nonsense-mediated decay [192]. In addition to improper 3’-end site selection, loss of Paf1C subunits also results in shorter poly(A) tail lengths [182]. These errors in mRNA 3’-end formation can be partially attributed to the observation that loss of Paf1C reduces chromatin association of the cleavage and polyadenylation factor Pcf11 [182]. This effect could however be indirect because Pcf11 is known to associate with the Ser2 phosphorylated form of RNA Pol II [226], a mark which is reduced by the loss of Paf1C [224, 227]. In addition to Pcf11, Paf1C also associates with the cleavage and polyadenylation factor Cft1, facilitating its recruitment to the Ser5 phosphorylated form of RNA Pol II [183]. Akin to what is seen in yeast, human Paf1C is known to associate with subunits of the 3’ mRNA processing factors CPSF (CPSF-30, 73, 100, 160) and CstF (CstF-66, 77) [184]. Further, loss of Paf1C in humans also results in a reduction of mRNA polyadenylation [228].

In addition to mediating mRNA processing, Paf1C is also important for proper termination and 3’-end formation of non-polyadenylated RNA transcripts, including a class of non-coding RNAs known as small nucleolar RNAs (snoRNAs) [185]. When properly terminated and processed these snoRNAs go on to complex with proteins to form small nucleolar ribonucleoproteins (snoRNPs) and catalyze ribosomal RNA processing and ribosome biogenesis (reviewed in [229]). Similar to what is seen with mRNA transcripts, when Paf1C is compromised snoRNA gene transcripts do not terminate properly, leading to RNAs with extended 3’ ends

[185, 204]. Of these snoRNAs, *SNR13* and *SNR47* have been best characterized and their proper termination is known to be further affected by Paf1C-dependent histone modifications [204]. These defects in proper snoRNA termination are associated with reduced levels of recruitment of the Nrd1-Nab3-Sen1 complex. This 3'-end processing complex recognizes non-coding RNA transcripts and prepares them for cleavage and maturation by the nuclear exosome–TRAMP complex [83]. By disrupting recruitment and function of the Nrd1-Nab3-Sen1 complex and other termination factors, snoRNA transcripts often fail to terminate and continue into the body of downstream genes.

1.2.1.4 Maintenance of chromatin during transcription

Studies in the Marten's lab, uncovered a previously uncharacterized role for Paf1C in controlling transcription-coupled nucleosome occupancy [186]. Specifically they found that members of Paf1C are required to maintain repression of the *SER3* gene though maintenance of chromatin structure over the gene's promoter [186].

In *S. cerevisiae* the expression of the gene *SER3* is controlled by a unique regulatory mechanism involving transcription interference. *SER3* codes for an enzyme that catalyzes the first step in serine biosynthesis. Expression of *SER3* is repressed in serine-rich conditions and rapidly activated in serine-starved conditions [230]. Repression is controlled by the serine-dependent transcription of the non-coding RNA *SRG1,* which lies upstream of *SER3* and extends across *SER3*'s promoter to its transcriptional start site [230-232]. Intergenic transcription of *SRG1* helps promote nucleosome occupancy across the *SER3* promoter, effectively preventing the binding of transcription factors to activate *SER3* expression. *SRG1* transcription-coupled nucleosome occupancy over *SER3*'s promoter requires the activity of several factors including

the chromatin remodeling complex Swi/Snf, the histone chaperones Spt6 and Spt16, and the HMG (high mobility group)-like protein Spt2 [233-235]. In a serine-dependent manner Swi/Snf is recruited to *SRG1* by the Cha4 activator protein, where it facilitates RNA Pol II transcription of *SRG1* by remodeling nucleosomes. *SRG1* transcription positions and maintains nucleosomes over the *SER3* promoter in a mechanism requiring Spt6, Spt16, and Spt2 [235, 236].

Pruneski *et al.* found that *SRG1* transcription-dependent repression of *SER3* is also dependent on Paf1C, in a mechanism where Paf1C mediates nucleosome occupancy over the *SER3* promoter [186]. Deletion of Paf1C subunits Paf1 and Ctr9 strongly derepresses *SER3* expression (8-10 fold), while deletion of Leo1, Cdc73, and Rtf1 show only modest *SER3* derepression (2-3 fold). This derepression does not appear to be controlled through Paf1C-dependent histone modifications or RNA Pol II CTD phosphorylation. Under *SER3* repressive conditions, Paf1C colocalizes with RNA Pol II along the *SRG1* transcribed region. Strains lacking Paf1 and Ctr9 exhibit reduced nucleosome occupancy and reduced recruitment of Spt16 at the promoter of *SER3* [186]. Together, these lead to a model where Paf1C controls transcription-coupled nucleosome occupancy at the *SER3* promoter by maintaining localization and activity of the histone chaperone FACT. Future studies are needed to determine whether Paf1C functions more broadly across the genome in regulating gene expression through nucleosome occupancy.

1.2.2 The subunits of the Paf1 complex

In the following section, I will briefly discuss each of the Paf1C subunits and describe what the community has learned about them. First I will summarize the discovery of the subunit and some

of the known functional roles of the subunit. Finally, I will touch on the current structural information on the subunits.

1.2.2.1 Paf1

As mentioned earlier, the Paf1 subunit in yeast was initially discovered in an search for RNA Pol II associated proteins, where RNA Pol II was affinity-purified from yeast using an antibody directed against the CTD tail of RNA Pol II. From those initial immunoprecipitations, the protein Paf1 along with Cdc73 was discovered to associate with the polymerase in a complex distinct from the Srb-containing mediator complex [165, 166]. The other members of the complex were found to associate with Paf1 in subsequent studies through immune-affinity and tandem affinity purification [167, 237]. In humans the Paf1 subunit was discovered as a novel protein interacting with the human homolog of Cdc73, parafibromin [238, 239].

Recently, the crystal structure of a subcomplex of human Paf1 and Leo1 was solved (Figure 6), giving the first structural information on the intersubunit interactions within Paf1C [240]. The Paf1 segment is composed of residues 161-250 and forms a structure consisting of three β-strands flanked on either side by an α-helix. The Leo1 fragment is comprised of residues 370-462 and consists of a single α-helix and six β-strands (Figure 6A). The two fragments form a tight heterodimer, creating a mixed antiparallel β-sheet (Figure 6B). Co-immunoprecipitation experiments suggest that the Paf1-Leo1 interaction observed in the structure is critical in Leo1 recruitment to Paf1C. Further, deletion of the Ctr9 subunit prevented Leo1 and Paf1 association with the remaining Paf1C complex members, strongly suggesting that Ctr9 acts as the key scaffold mediating Paf1C formation. Finally, through *in vitro* pull-down assays it was shown that both Paf1 and Leo1 are capable of interacting with the N-terminal tail of histone H3 (residues 1-

28), the histone octamer, and the nucleosome. Proper nucleosome binding is mediated by the C-terminal portion of Paf1 (residues 376-531) and the N-terminal portion of Leo1 (residues 1-369), and requires Paf1-Leo1 heterodimerization [240].

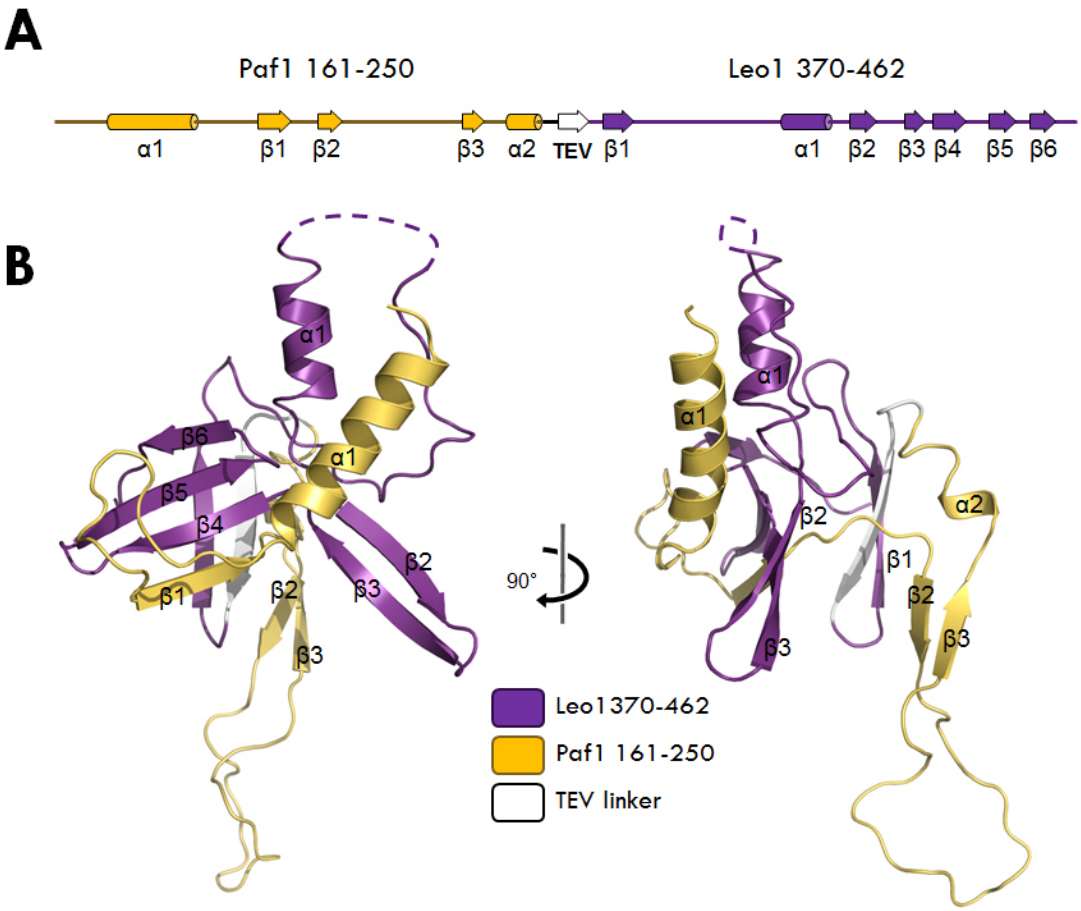


Figure 6. Crystal structure of the human Paf1-Leo1 subcomplex

A) The observed secondary structure in the human Paf1 161-250 – Leo1 370-462 complex structure (PDB id: 4M6T) [240]. The protein used for crystallization was a fusion protein with a linker containing a TEV cleavage site. The TEV cleavage site forms a β-strand in the structure (white arrow). B) Ribbon diagram of the Paf1-Leo1 complex structure. Paf1 is shown in yellow, Leo1 in violet, and the TEV linker in white. Residues 399–406 in Leo1 are not visible in the crystal structure and are represented as a dashed violet line.

1.2.2.2 Ctr9

Ctr9 is the largest member of the Paf1 complex and was discovered around the same time in both yeast and human. In yeast, Ctr9 (Cln three (*CLN3*) requiring 9) was first identified genetically by its connection to the cell cycle, where it is required for proper expression of G1 cyclin genes, including *CLN2* [237, 241]. Additionally, Ctr9 was discovered independently in another genetic screen looking for factors affecting chromosome segregation and named Cdp1 (Cbf1p-dependent mutant 1) [242]. Subsequently, Ctr9 was determined to be identical to Cdp1 and was found to associate with the other members of the Paf1 complex [167, 237]. In humans, Ctr9 was first identified from a collection of cDNA clones [243], and found to be a nuclear protein involved in protein-protein interactions between TPR-containing and SH2-containing proteins [244]. It was only until much later, in a study looking for Cdc73 associated proteins was Ctr9 determined to be a member of Paf1C, analogous to that seen in yeast [238].

Currently no NMR or x-ray crystallography data for Ctr9 exist to aid in functional characterization. *In silico* analysis of the primary amino acid sequences of both yeast and human Ctr9 reveal that both contain numerous tandem tetratricopeptide repeats (TPR) motifs (Figure 7) [242, 244]. TPR motifs are typically thought of as protein-protein interaction motifs and are involved in regulating a wide range of cellular functions (reviewed in [245]). The TPR motif is a degenerate, 34 amino acid repeat that is often arranged in tandem arrays of 3-16 repeats [246]. The consensus sequence of the TPR motif is characterized by a pattern of small and large hydrophobic amino acids, with the highest conservation at residues 8, 20, and 27 (typically glycine or alanine). Structurally, the TPR motif is composed of two antiparallel α–helices in a helix-turn-helix arrangement [247]. Adjacent TPR motifs pack together in a parallel fashion, creating a helical array of repeating antiparallel α–helices [247-250]. As more co-crystal

structures of TPR helical arrays have been solved, it has become evident that TPR arrays are versatile in their ability to recognize and bind different protein binding partners. Some TPR arrays, like those seen in the Hsp70/Hsp90 organizing protein (Hop), utilize the concave surface formed from the helical array of multiple TPR motifs to bind interacting partners [248]. Others, like those seen in p67phox (component of NADPH oxidase), create a binding surface along the edge of the TPR arrays utilizing the connecting loops between adjacent TPR motifs [249]. In the case of Ctr9, the presence of tandem arrays of TPR motifs strongly suggests the subunit acts primarily as a molecular scaffold mediating the formation of Paf1C. Future studies are needed to uncover mechanistically how Ctr9 utilizes its TPR motifs to bind with members of Paf1C.

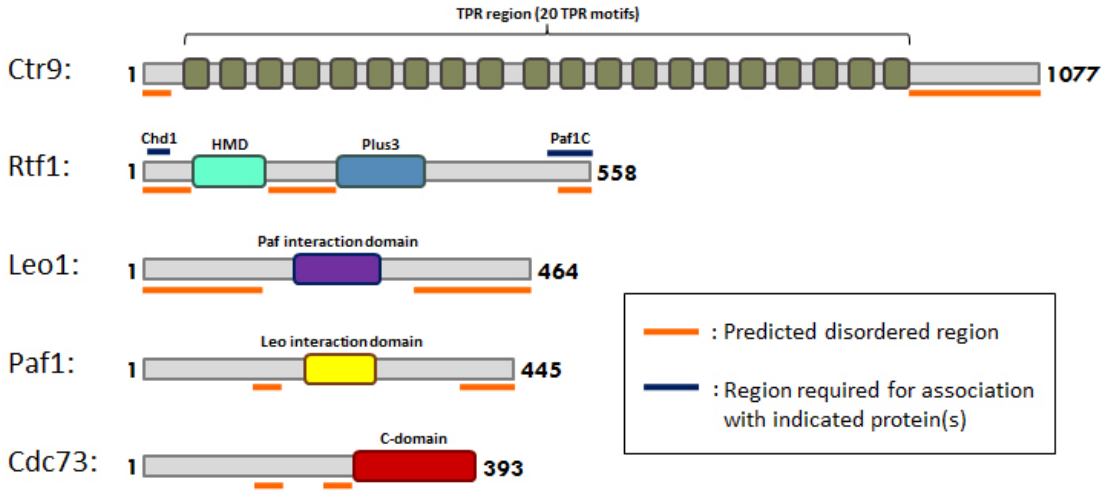


Figure 7. Domain structure of the *S. cerevisiae* Paf1 complex

Each of the Paf1 complex subunits are represented by grey bars with their amino acid residue length indicated. Colored boxes indicate predicted or known structural domains in *S. cerevisiae*. Ctr9 has 8 predicted tetratricopeptide repeat (TPR) motifs [251], however secondary structural predictions suggest it contains many more, forming a TPR region. The structured regions indicated in Leo1 and Paf1 indicate the regions homologous to those co-crystallized from the human Paf1 complex [240]. Blue bars above the subunits indicate regions required for subunit association with the indicated protein(s). Orange bars below the subunits indicate regions predicted to be disordered [252].

1.2.2.3 Leo1

Leo1 (Left open reading frame 1) was initially identified in yeast when a 5.6 kp DNA segment of chromosome XV was sequenced [253]. The gene was briefly characterized and found to be constitutively expressed at low levels and non-essential for viability. Further they found that the protein was extremely hydrophilic and likely interacted with other basic proteins [253]. It was only much later that the protein was determined to be a component of Paf1C, and its deletion suppresses several *paf1Δ* phenotypes [167]. In humans, Leo1 was identified as a member of Paf1C when it along with Ctr9 and Paf1 was found to co-immunoprecipitate with the human homolog of Cdc73, parafibromin [238, 239]. Around the time Leo1 was determined to be a component of human Paf1C, the gene was identified by another lab and named *RDL* (replicative senescence down-regulated Leo1-like gene). They found the gene was down-regulated upon replicative senescence in human 2BS fibroblasts and that exogenous expression of the C-terminal fragment of the protein accelerated senescence and shortened the lifespan of 2BS fibroblasts [254].

In yeast, Leo1 has been shown to play a role in recruiting Paf1C to chromatin [255]. Deletion of *LEO1* in yeast reduces occupancy of the remaining members of Paf1C on active genes (*YEF3*, *PYK1*). This is thought to be in part mediated by Leo1's RNA binding capabilities, as both Leo1 and Rtf1 are capable of binding RNA *in vitro*. RNA binding of Paf1C *in vivo* does not appear to be affected by loss of *RTF1*, suggesting Leo1 is the major contributor in Paf1C's RNA binding function [255]. Further when *LEO1* is deleted in yeast, the ability of Paf1C to co-precipitate with cross-linked mRNA transcripts is greatly reduced, suggesting Leo1 is important for the stable association of mRNA with Paf1C *in vivo* [255]. Interestingly, strains lacking *LEO1* do not exhibit strong phenotypes related to chromatin modifications or elongation [167, 169],

suggesting Leo1 may cooperate with Rtf1 and Cdc73 in the recruitment and stability of Paf1C with elongating RNA Pol II. Future studies are needed to fully establish Leo1's RNA binding capabilities and its biological significance in Paf1C recruitment and stability at active genes.

As detailed previously, a region of Leo1 heterodimerized with a fragment of Paf1 was recently crystallized. The structure revealed the two proteins form a tight interaction that is important for proper Paf1C formation (Figure 6) [240]. Further, both Paf1 and Leo1 were shown to associate with the N-terminal tail of histone H3 and nucleosomes *in vitro*, suggesting another mechanism to aid in Paf1C recruitment and stable association with Pol II on a chromatin template [240].

1.2.2.4 Cdc73

In yeast Cdc73 (Cell division cycle 73) was initially identified as a protein involved in mating signaling pathways and cell division [256], before it was found to co-purify with Paf1 and RNA Pol II [168]. In humans, Cdc73 was first identified and known as Parafibromin and recognized as a tumor suppressor gene, that when mutated in the germ line often results in hyperparathyroidism-jaw tumor syndrome [257-259]. The C-terminal region of Parafibromin was recognized to display modest homology (27%) to the yeast version of Cdc73 and later found to co-purify with human homologs of Paf1C [238, 239]. Parafibrombin's role as a tumor suppressor is in part regulated by recruiting histone methyltransferases to reduce cyclin D1 expression [260, 261]. It has also been found to act as an oncogenic factor, regulating nuclear Wnt signaling through a direct interaction with β-catenin [262]. Interestingly, the regions responsible for these seemingly antagonistic functions are not found within the yeast sequence [262, 263].

Cdc73 along with Rtf1 plays an important role in recruiting Paf1C to chromatin. Deletion of either subunit reduces Paf1C occupancy levels at transcribed genes and reduces the levels that immunoprecipitate with RNA Pol II [177, 182, 183, 255]. The region of Cdc73 responsible for recruitment is contained within the highly conserved C-terminal region, known as the C-domain (Figure 7) [264, 265]. Removal of the C-domain in yeast results in a significant reduction of Paf1C association with active genes, but importantly does not affect Paf1C assembly [264]. Additionally, *in vitro* studies have shown the C-terminal region is capable of binding diphosphorylated RNA Pol II CTD repeats and phosphorylated Spt5 C-terminal repeats (CTR) [265]. Furthermore, substitution mutations in the C-terminal region of Cdc73 that reduce *in vitro* binding to CTD and CTR peptides also reduce Paf1C recruitment to coding regions of several active genes [265]. Taken together, these data suggest that the C-domain of Cdc73 mediates Paf1C recruitment to chromatin through a direct interaction with the phosphorylated forms of RNA Pol II and Spt5. This also partially explains the requirement of the cyclin-dependent kinases, Bur1(phosphorylates the CTR and Ser5 of the CTD) and Kin28 (phosphorylates Ser5 and 7 of the CTD) in Paf1C recruitment to active genes [265-268].

In yeast, the C-domain of Cdc73 (amino acids 236-393) has been crystallized and found to form a stable, globular domain (Figure 8). Remarkably despite having little sequence identity, the C-domain of Cdc73 adopts a fold with significant structural similarity to small GTPases of the Ras superfamily [264, 269]. Most proteins in the Ras superfamily are small nucleotide-binding proteins that utilize the energy in nucleotide hydrolysis to alternate between two different conformations (reviewed in [270]). In the case of GTPases, they toggle between GDP- and GTP-bound states, each of which supports association with different sets of effector proteins. This effectively allows nucleotide binding to serve as a molecular switch to regulate protein-

protein interactions. Despite its high sequence similarity to small GTPases, the C-domain does not contain a discernible binding pocket for a nucleotide. The critical loops that aid in nucleotide recognition and hydrolysis (P-loop, Switch I, Switch II) differ in conformation and are not well conserved [264]. Mutational analysis of conserved surface exposed residues of the C-domain identified a signal tryptophan residue (W321A) that when mutated exhibits sensitivity to 6-AU [264]. This suggests that the residue is functionally important and may represent a protein-protein interaction site important for Paf1C recruitment to chromatin. Further, parafibromin has been shown to interact with β-catenin in a manner important for proper activation of the Wnt signal-transduction pathway, a commonly altered pathway in many cancers [262].

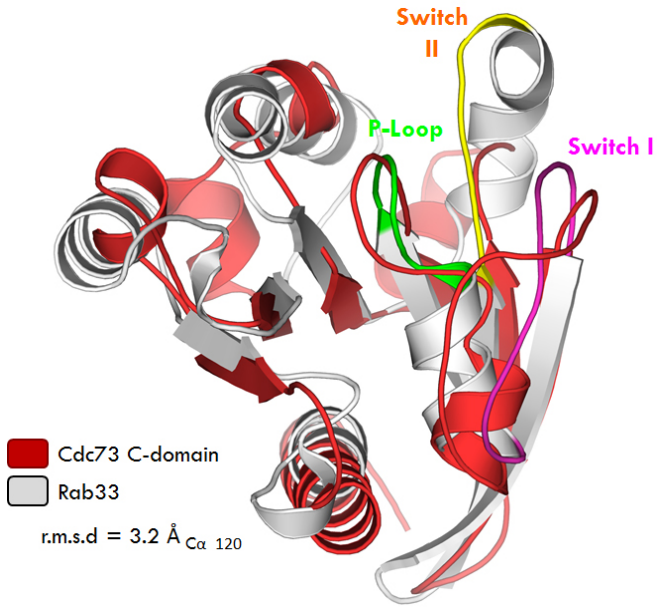


Figure 8. The Cdc73 C-domain shares structural similarity with small GTPases

Structural overlap of the S. cerevisiae Cdc73 C-domain (red; 3V46) with Rab33 (white; PDB id: 1Z06). The two structures deviated by a r.m.s.d of 3.2 Å over 120 Cα atoms. Within the Rab33 structure, the P-loop (green), and Switch I (magenta) and II (yellow) are indicated.

1.2.2.5 Rtf1

Rtf1 was first identified in an *S. cerevisiae* suppressor screen looking for factors that restored TATA box-binding protein (TBP) function (restores TBP function) in a TBP mutant with relaxed DNA binding specificity [175]. Shown to be a member of Paf1C [167], Rtf1's role in transcriptional regulation has been greatly expanded over the years to include regulation of co-transcriptional histone modifications, recruitment of Paf1C to active open reading frames, and recruitment of the chromatin remodeler Chd1[188, 189, 205, 271]. Structural and functional analysis in the Arndt lab has shown that each of these functions is mediated by discrete non-overlapping regions of Rtf1 (Figure 7) [177]. Rtf1's N-terminal region (amino acids 3-30) is required for physical interaction with the chromatin-remodeling protein Chd1 [271, 272]. The histone modification domain (HMD; animo acids 62-152) is required for the mono-ubiquitylation of histone H2B (K123) and the subsequent di- and trimethylation of histone H3 K4 and K79 [188, 189, 205, 271]. The centrally located Plus3 domain (amino acids 201-395) is needed for Rtf1 recruitment to active open reading frames. Finally, Rtf1's association with Paf1C is mediated by a C-terminal region (amino acids 491-558) [177]. While the regions of Rtf1 important for its diverse functions have been established, the molecular mechanisms driving each of these functions are still unclear.

The Paf1C-dependent histone modifications H2B K123 ubiquitylation and the downstream di- and trimethylation marks of H3 K4 and K79 require the HMD of Rtf1 in yeast. Comprised of only 90 amino acids (residues 62-152), when expressed as the only source of Rtf1 in yeast, the HMD is capable of promoting H2B K123 ubiquitylation and H3 K4 and K79 methylation [177, 205]. Several conserved residues within the domain have been identified that impair the histone modification function of Rtf1, including glutamate 104 and phenylalanine 123

[204]. Sequence conservation analysis suggests that the HMD is evolutionarily conserved [177]. This is supported experimentally in *S. cerevisiae*, where its HMD sequence can be replaced by the predicted HMD sequences of other organisms (*S. pombi*, *D. melanogaster*, *D. rerio*, *H. sapiens*) and reconstitute Rtf1-dependent histone modifications to varying degrees, strongly suggesting the HMD histone modification functions are evolutionarily conserved [205].

The HMD can also support Rtf1-dependent histone modifications in the absence of other Paf1C members [205], including Paf1 and Ctr9 which are known to be important in Paf1C stability and integrity [181, 182]. Remarkably, when the HMD is expressed as the only source of Rtf1 in yeast, the HMD is still capable of localizing to chromatin *in vivo* and associating with nucleosomes *in vitro* [205]. Amino acid substitutions in E104 and F123 of the HMD, both residues known to reduce histone modification activity of full-length Rtf1 [204], reduce the HMD's ability to associate with nucleosomes *in vitro*. This suggests the HMD's ability to associate with chromatin is important for proper histone modifications [205]. Interestingly, when the HMD is expressed in the absence of the rest of Rtf1, the HMD and HMD-mediated histone modifications are found enriched at transcriptionally inactive loci, suggesting the other regions of Rtf1 act to direct HMD association and function to transcribed genes [205].

As mentioned earlier, both Rtf1 and Cdc73 are critical for regulating recruitment of Paf1C to chromatin, as deletion of either results in reduced occupancy levels of Paf1C at active genes [177, 182, 183, 255]. Rtf1's role in recruiting Paf1C to chromatin is localized to its Plus3 domain (named for the domain's three invariant basic residues), also known as the ORF-Association Region (OAR) [177]. The molecular mechanism detailing the Plus3 domain's involvement in Paf1C recruitment to chromatin was recently shown to be mediated through a direct interaction with the essential transcriptional regulator Spt5 [273]. Spt5 forms a complex

with the zinc finger protein Spt4 (Spt4-Spt5, known as DSIF in humans), and together are recruited to elongating RNA pol II through Spt5's NusG N-terminal domain and KOW (Kyprides, Ouzounis, Woese) domains (Figure 9) [53, 274]. Akin to the CTD in the Rpb1 subunit of RNA Pol II, Spt5 contains a set of short repeats at its C-terminus (CTR) that are capable of being phosphorylated (Figure 9) [58, 266, 275, 276]. Phosphorylation of the CTR by the kinase Bur1/P-TEFb, regulates recruitment of Paf1C to elongating RNA pol II [58, 266, 267]. The direct interaction between the Plus3 domain and Spt5 is through the CTR region of Spt5 and association is stimulated by CTR phosphorylation [273]. Additionally, the Plus3 domain can localize to chromatin independently of other Paf1C members and exhibit occupancy patterns similar to what is seen for full-length Rtf1. The ability of the Plus3 domain to occupy chromatin is significantly reduced in strains missing the Spt5 CTR or Bur2 [273], the cyclin in the Bur1-Bur2 protein kinase responsible for phosphorylating the Spt5 CTR [266, 267, 277]. Mutation of the invariant Plus3 residues (R251, R273, K299) of the domain, particularly when mutated simultaneously, caused transcription-related mutant phenotypes and reduced the ability of Rtf1 to associate with chromatin [273]. Taken together, these data strongly suggest that the Plus3 domain of Rtf1 acts to recruit Paf1C to the elongating polymerase through a direct interaction with the CTR region of Spt5.

An NMR structure of the internal Plus3 domain of human Rtf1 marked the first structural information on any of the members of Paf1C [278]. The NMR study revealed the Plus3 domain forms a globular fold comprised of six α-helices and six β-strands aligned in a mixed α/β topology. The six β-strands arrange in a four-stranded antiparallel β-sheet, forming a β-subdomain. The β-subdomain is surrounded by the N- and C-terminal α-helices. Utilizing the DALI server [279], de Jong *et al.* found that the β-subdomain has structural homology to a wide

variety of nucleic acid binding proteins, including the PAZ domains of Dicer and Argonaute and tutor domains [278]. Based on the Plus3 domain's structural homology to DNA binding motifs and the prevalence of surface exposed basic amino acids, the investigators suggested the domain might mediate nucleic acid binding. This notion was supported by ^{15}N-HSQC titration experiments and electrophoretic mobility shift assays showing the Plus3 domain was capable of binding single-stranded DNA substrates. The domain's DNA binding surface was narrowed down to one surface of the β-subdomain, including conserved residues R402, R430, and R436 [278]. The biological role for the Plus3 domain's single-stranded DNA binding is currently uncertain, however it has been suggested the interaction could help in structural organization of the elongating transcription bubble.

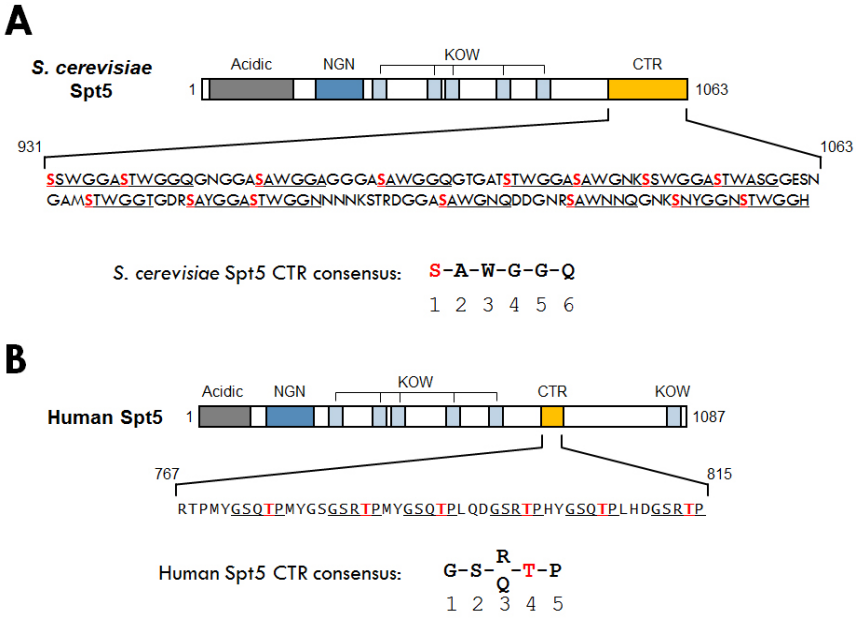


Figure 9. Elongation factor Spt5

A) Diagram of functional regions within *S. cerevisiae* Spt5 (NGN, NusG N-terminal domain; KOW, Kyprides Ouzounis Woese; CTR, C-terminal repeat). The CTR region is expanded to show the primary sequence and individual CTR motifs are underlined. The serine residues that are phosphorylated in the CTR region are highlighted in red. A consensus CTR sequence is shown. B) Diagram of the functional regions within the human Spt5.

1.2.3 Composition of the Paf1 complex

Our understanding of the interaction networks within Paf1C are slowly emerging. In yeast, overexpression or deletion of individual Paf1C subunits can affect the levels of other members of the complex. Overexpression of Paf1 or Cdc73 can stabilize and increase the expression of one another [168]. Deletion of *PAF1* results in a dramatic decrease in cellular levels of Ctr9, Cdc73, and Rtf1 by over ten-fold. Similarly a three-fold decrease in the cellular levels of Paf1, Rtf1, and Leo1 results from the deletion of *CDC73* [182]. Rtf1's association with Paf1C is significantly compromised in strains lacking *CDC73*. Deletion of *RTF1*, however does not appear to affect the Cdc73's association with the rest of Paf1C [183].

In humans, the composition and intersubunit interactions within Paf1C are more well defined. *In vitro* work by the Roeder lab has established an elaborate network of interactions between Paf1C subunits [181]. Paf1C subunits were co-expressed in pairs recombinantly in insect cells and direct interactions monitored via co-immunoprecipitation. Further they determined the effect of individual subunit deletion on complex formation. They concluded that the Paf1 subunit acts as the main scaffold in the complex because it is capable of binding every member of Paf1C with the exception Ski8. Ctr9 also appears to play an important scaffolding role interacting with Paf1, Cdc73, and mediating the only interaction with Ski8. Cdc73 was found to interact with Ctr9, Paf1, Leo1, and weakly with Rtf1. Rtf1 binds to Paf1 and weakly associates with Cdc73. Interestingly; in humans it appears as if the Rtf1 subunit is less stably associated with Paf1C [170, 239], and may exist in several different splice forms [181]. The recent co-crystal structure of the Paf1 and Leo1 heterodimer and corresponding co-immunoprecipitation studies support the network established by the Roeder lab [240].

1.2.4 The Paf1 complex's role in disease state and cancer progression

Due to the Paf1 complex's broad range of functions in transcriptional regulation, it is not surprising that alterations in its subunits result in numerous diseases. Here I will briefly touch on some of the known connections between Paf1C and disease states.

The Paf1 complex has been shown to play a role in the innate defense against viral infection [197, 198]. Components of Paf1C were identified in a screen to find anti-viral factors acting at the early events of infection and integration of HIV-1 into the host genome [197]. Knockdown of all members of Paf1C results in enhanced HIV-1 infection. Additionally, expression of Paf1C restricts infection of HIV-1, HIV-2, and SIV (Simian immunodeficiency virus). Paf1C does not exert its anti-viral activity through up regulating expression of known anti-HIV restriction factors (p21, APOBEC3G, Tetherin, TRIM5α, SAMHD1, TREX1). Further it does not appear to be mediated through controlling the cell cycle or affecting viral genome stability [197]. Paf1C is known to associate with the HIV-1 protein Tat (trans-activator of transcription). Tat forms a stable complex with P-TEFb, Paf1C, and several MLL-fusion partners (mixed lineage leukimia) and is thought to enhance transcription of the HIV-1 promoter [280]. Subunits of Paf1C are also implicated in suppressing infection of influenza virus [198]. Upon influenza virus infection, influenza non-structural protein 1 (NS1) interacts with the infected cell's epigenome, hijacking host proteins. Interestingly, the NS1 protein of influenza A possesses a histone H3-like sequence, mimicking the first four residues of H3. This histone-mimic sequence is able to target and bind Paf1, preventing Paf1C from activating antiviral genes to combat the infection. Moreover, loss of Paf1-NS1 binding attenuates influenza infection [198].

Paf1C is also known to be involved in MLL fusion-mediated leukemogenesis [281, 282]. The protein MLL (mixed-linage leukemia) is a histone methyltransferase that regulates *HOX* gene expression through H3 K4 methylation via its C-terminal SET domain [283, 284]. Translocations in the *MLL* gene are associated with aggressive forms of leukemia, where the N-terminal amino acids of MLL are fused in frame with over 60 translocation partners [283]. These fusion proteins lack the C-terminal SET domain of MLL and deregulate transcriptional activity of MLL target genes, leading to immortalization. In most cases, transcriptional activation is thought to be mediated by the MLL fusion protein and a complex of other proteins including P-TEFb and the histone methyltransferase Dot1 [285-287]. Components of Paf1C have been shown to bind MLL at the region immediately upstream of the fusion point, containing CxxC and RD2 domains [281]. Paf1C is also capable of stimulating MLL and MLL-AF9 (fusion protein) transcriptional activation of a *Hoxa9* reporter construct, in a manner requiring the CxxC and RD2 domains. Moreover, knockdown of Paf1C reduces localization of MLL to target genes and reduces MLL fusion-mediated transcriptional activation. Finally, deletion of the MLL-Paf1C interaction region appears to prevent MLL-AF9 mediated immortalization of bone marrow cells, strongly suggesting Paf1C is critical for leukemogenesis [281].

Paf1C is strongly linked to a growing number of cancers, including thyroid, breast, liver, and pancreatic (reviewed in [195]). Expression of the Paf1 subunit is found upregulated in pancreatic cancer cell lines and overexpression of Paf1 in NIH 3T3 (mouse embryonic fibroblast) cells results in enhanced growth rates *in vitro* and tumor formation *in vivo* [288]. Of all of Paf1C subunits, Cdc73/parafibromin's role in cancer progression has been best studied and characterized. Encoded by the gene *HRPT2* (hyperparathyroidism 2), Cdc73/parafibromin is associated with suppression of the hyperparathyroidism-jaw tumor (HPT-JT) syndrome [257-

259]. Two-thirds of patients with parathyroid carcinomas have mutations in Cdc73/parafibromin that result in lower expression levels, truncations, or inactivation of the protein [289]. Many of these truncation mutants expressed in tumors lose their ability to associate with Paf1 and RNA Pol II, suggesting that tumorigenesis is in part mediated by its separation from Paf1C [239]. More research is needed to fully define Paf1C's role in cancer development. Its close ties to transcriptional regulation and gene expression ensure many more connections to cancer progression exist.

1.3 DISSERTATION AIMS

The ability of the Paf1 complex to mediate such a broad range of functions throughout the transcription cycle makes Paf1C a very interesting subject for study. Currently, however, for the majority of these functions, the molecular mechanisms and the specific subunits required are unknown. Analysis of the domain structure of Paf1C subunits has yielded minimal clues to how it performs its range of functions. Few of the Paf1C subunits have recognized structural domains, making it difficult to assign functions to individual subunits (Figure 7). Rtf1 has the clearest domain structure of Paf1C subunits. Structural and functional analysis of Rtf1 has revealed several domains, including the histone modifying domain (HMD) and Plus3 domain (Figure 7) [177, 278]. While we appreciate the HMD is required for several Paf1C-dependent histone modifications and the Plus3 domain is required for proper Paf1C recruitment to chromatin, the molecular mechanism and physical basis for both are uncertain. The goal of my thesis work is to determine the molecular mechanism by which Rtf1 influences transcription and chromatin

structure. To this end I have focused on studying the HMD and the Plus3 domain of Rtf1 in both *S. cerevisiae* and *H. sapiens*, utilizing biochemical, biophysical, and structural techniques. Additionally I have studied Paf1C as a whole, focusing on identifying the regions of the complex critical in maintaining a stable core complex structure.

Aim 1: The structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin

Recruitment of Rtf1 to actively transcribing RNA pol II is controlled by its internally located Plus3 domain [177]. The goal of this aim is to uncover the molecular and physical basis controlling Plus3-dependent recruitment of Paf1C to chromatin.

Aim 2: Structural characterization of the histone modification domain of Rtf1

In yeast, the HMD is necessary and sufficient for histone H2B monoubiquitylation of K123 and the subsequent di- and trimethylation of histone H3 K4 and K79 [177, 205]. The goal of this aim is to uncover the molecular and structural basis for HMD-mediated histone modifications.

Aim 3: Structural characterization and identification of the core of the Paf1 complex

A comprehensive understanding of Paf1C function requires detailed knowledge of the spatial arrangement of Paf1C subunits. Currently very little is known about how Paf1C organizes its functional and structural domains. The goal of this aim is to identify the domain structure required for complex formation and characterize the complex and sub-complexes with biophysical and biochemical techniques.

2.0 THE STRUCTURAL BASIS FOR SPT5-MEDIATED RECRUITMENT OF THE PAF1 COMPLEX TO CHROMATIN

2.1 INTRODUCTION

Eukaryotic gene expression utilizes numerous accessory factors to aid progression of RNA polymerase (Pol) II through the transcription cycle and to regulate transcript levels. Additionally, some of these factors, including the Polymerase Associated Factor 1 Complex (Paf1C), influence gene expression through their ability to modify chromatin structure or regulate RNA processing [290, 291]. Originally identified in yeast during a search for RNA Pol II-associated factors [165, 166], Paf1C is conserved throughout eukaryotes and is minimally composed of the subunits Paf1, Ctr9, Cdc73, Rtf1, and Leo1 [114, 167, 169]. Human Paf1C also contains an additional subunit Ski8 [170], which functions to regulate exosome-mediated mRNA degradation [171]. The complex co-localizes with RNA Pol II from the promoter to the 3' end of genes [4, 172] and mediates a diverse set of events, including transcription-coupled histone modifications [187-191], phosphorylation of the RNA Pol II C-terminal domain (CTD) [182, 183], recruitment of termination and RNA 3'-end processing factors [183-185], and maintenance of chromatin during transcription [186] (Figure 5). Proper transcription-coupled ubiquitylation of histone H2B at lysine 123 in yeast is dependent on Paf1C recruitment to chromatin [273]. Phenotypes caused by mutations that dissociate Paf1C from active genes argue that the ability of Paf1C to carry out its

diverse set of activities in a controlled manner requires proper recruitment of Paf1C to actively transcribing RNA pol II [177, 264, 265, 273].

Several Paf1C-interacting factors have been shown to play a role in the recruitment of Paf1C to transcribed genes. In humans, Paf1C promotes transcription elongation through direct interactions with RNA Pol II as well as the elongation factor SII/TFIIS [181]. Paf1C also co-purifies with the elongation factor Spt4-Spt5/DSIF complex and the histone chaperone Spt16-Pob3/FACT [114, 167, 169, 292], all of which travel with the elongating polymerase and could thus be mediating Paf1C recruitment. Spt5 is of particular note since it is the only RNA pol II associated factor known to be conserved in all three kingdoms of life (NusG in bacteria, RpoE in archaea) [53]. Outside of bacteria, Spt5 is found in complex with the small zinc binding protein Spt4 [293], which stabilizes Spt5 and improves Pol II processivity [274]. Recruitment of Spt4-Spt5 to polymerase is mediated by a direct interaction between the Pol II clamp domain and the Spt5 NusG N-terminal domain (NGN) [53, 274]. Similar to the RNA Pol II CTD, Spt5 contains a series of short repeating sequence motifs at or near its C-terminus, known as C-terminal repeats (CTR; Figure 9). Also similar to the RNA Pol II CTD, the Spt5 CTR has been proposed to serve as a scaffold for the recruitment of elongation factors, such as TFIIF, FACT, and Paf1C [169, 266, 267, 294], as well as RNA 3'-end processing factors [295]. Phosphorylation of the Spt5 CTR by the kinase Bur1/P-TEFb [58, 266, 275, 276] regulates its interactions with these factors.

The subunits Rtf1, Cdc73, and Leo1 have all been shown to play a role in recruitment of Paf1C to chromatin at active genes [177, 182, 183, 255]. In Cdc73, a highly conserved C-terminal domain containing a Ras-like fold mediates binding to both phosphorylated CTR and CTD peptides *in vitro* and facilitates chromatin occupancy *in vivo* (Figure 8) [264, 265, 269]. Within Rtf1, a central region called the ORF-Association Region (OAR) was shown to be

required for recruitment of Paf1C to chromatin (Figure 7) [177]. The OAR, defined through a series of internal deletions within Rtf1, is a conserved domain also named the Plus3 domain after the presence of three invariant positively charged amino acids. NMR structural analysis and chemical perturbation studies revealed that the Plus3 domain adopts a fold similar to Tudor and PAZ domains and identified residues on the Plus3 surface required for interactions with DNA substrates that mimic the transcription bubble [278]. A direct role for the Plus3 domain in recruitment of Paf1C to chromatin had remained elusive until a recent report demonstrating that the yeast Plus3 domain is both necessary and sufficient for binding to the C-terminal repeats of Spt5 [273]. Further, deletion of the Spt5 CTR or loss of Bur1 activity impairs recruitment of full-length Rtf1 [202, 265-267, 296] and the isolated Plus3 domain to chromatin [273], demonstrating that modification of the Spt5 CTR is crucial for recruitment of Paf1 to active genes.

Here we provide the first structural insight into how the Spt5 CTR mediates recruitment of a transcription elongation factor to RNA Pol II through the structure of the Plus3 domain from human Rtf1 in complex with a phosphorylated Spt5 CTR motif. The structure reveals the basis for recognition of the CTR motif as well as the molecular basis for phosphate specific recognition. Peptide binding is driven by two interfaces, one that discriminates phosphorylation at the T4 position within a CTR repeat while a second, hydrophobic surface interacts with position P5 and the Spt5 sequence between canonical CTR repeats. Both interaction surfaces observed in the Plus3-Spt5 crystal structure are critical for CTR peptide binding in vitro, and together contribute to Paf1C recruitment to chromatin in yeast.

2.2 RESULTS

2.2.1 A distinct conserved surface distinguishes the Plus3 domain from other Tudor-containing proteins.

To elucidate the mechanism by which the Plus3 domain of Rtf1 recognizes Spt5 and mediates recruitment of Paf1C, we crystallized and determined the structure of the Plus3 domain from human Rtf1 (residues 353-484) (Figure 10), refining to an R_{free} value of 24.0% against data to 2.12 Å resolution (Table 1). The Plus3 domain crystallized with two structurally similar monomers in the asymmetric unit (0.6 Å r.m.s.d. over 125 Cα atoms). A six-stranded antiparallel β-sheet anchors the core of the Plus3 domain and is encircled by six α-helices to form a compact, globular fold (Figure 11A). Overall the fold is similar to that observed using NMR (1.0 Å r.m.s.d. over 123 pairs of Cα atoms) [278], as is disorder within loop residues 392-397 observed in one of the two molecules in the asymmetric unit. Comparison with the structural database using DALI [297] identified similarity between the β-sheet core of the OAR and members of the Tudor superfamily including Tudor, MBT, chromo-, PWWP, and the related PAZ domains [278]. Tudor domains are best known as methyl-arginine and methyl-lysine binding domains that play roles in a broad range of biological functions including histone modifications [298], chromatin remodeling [299], and RNA maturation [300]. Recognition of methyl modifications is accomplished through a cage of two to four aromatic residues, typically contained within the Tudor domain itself, with enhanced peptide specificity achieved through neighboring sequences in Tandem [301, 302] or Extended [303-305] Tudor domains. While adopting a similar fold, the

aromatic residues required for methyl recognition are not conserved in the Rtf1 Plus3 domain (Figure 11B), suggesting it is unlikely to bind these substrates.

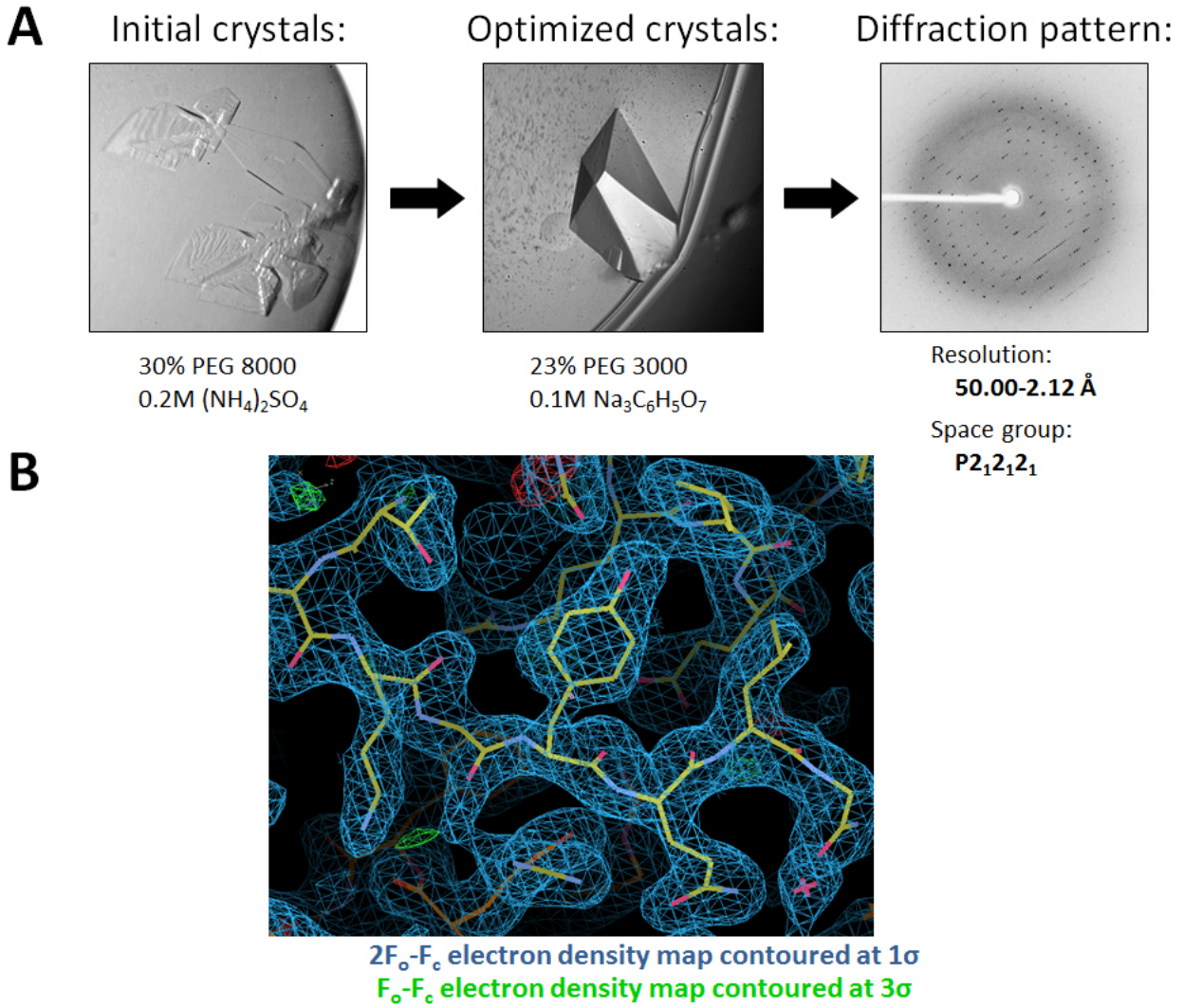


Figure 10. Crystallization and structure determination of the Plus3 domain

A) Schematic of the crystallization and data collection process for the Plus3 domain. Representative images from the crystallization process are shown, including the buffer conditions used in each case. A frame from the x-ray diffraction data is shown along with the resolution range and experientially determined space group. B) Section of the electron density map of the human Rtf1 Plus3 crystal structure.

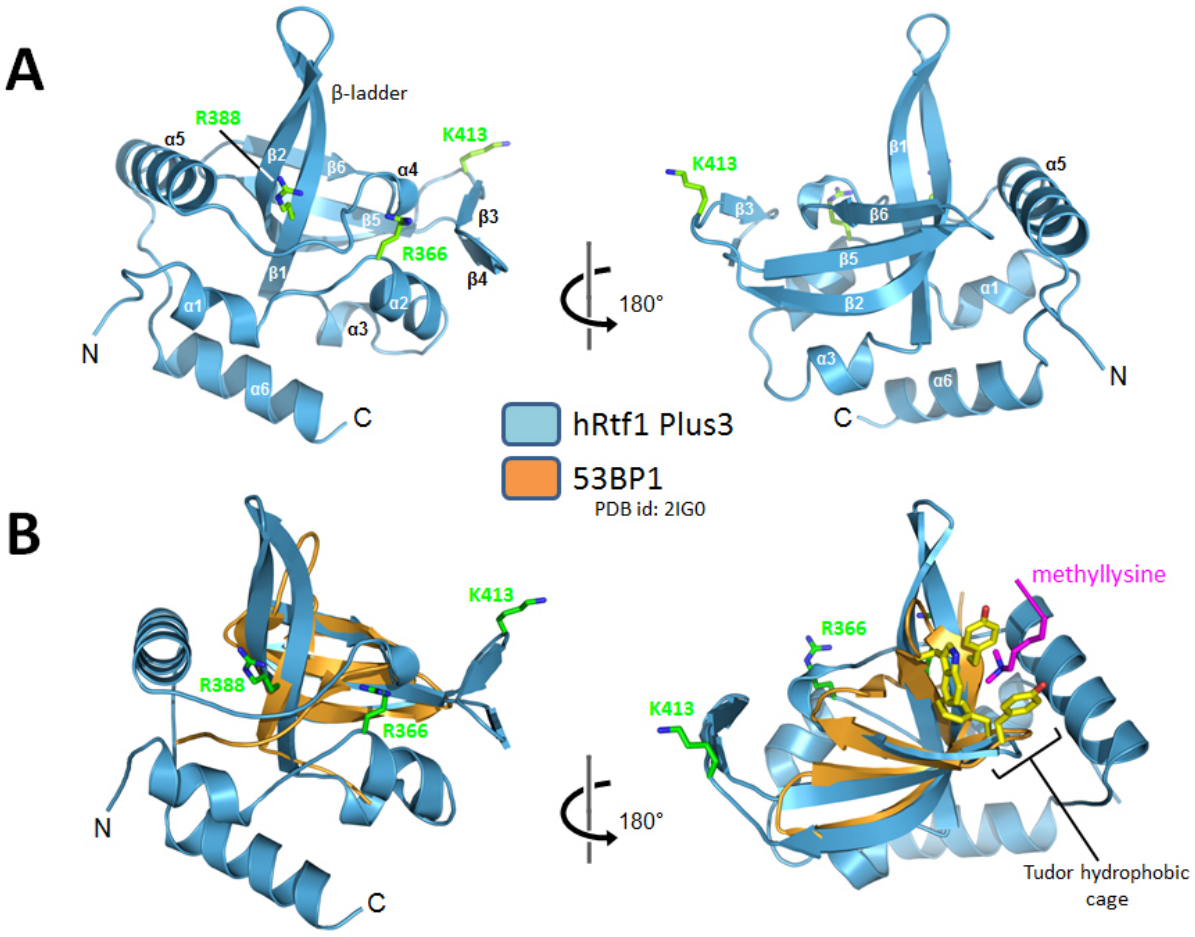


Figure 11. Structure of the human Rtf1 Plus3 domain

A) Ribbon diagram of the human Rtf1 Plus3 crystal structure. Secondary structural elements are labelled and the invariant Plus3 residues are shown as sticks (green). B) Structural overlap of the Rtf1 Plus3 domain (blue) with a Tudor-histone H4 methyllysine complex from p53-Binding Protein 1 (53BP1) (orange; PDBid 2IG0) [301]. Aromatic cage residues (yellow) in 53BP1 are indicated.

We mapped sequence conservation onto the surface of the Rtf1 Plus3 domain from an alignment of nine Rtf1 sequences (Figure 12, 13). A surface on Rtf1 that functions in the recognition of DNA substrates that mimic the transcription bubble has been reported [278]. This surface is located near the methyl recognition cage in Tudor domains (Figure 13B) and contains a number of residues that are conserved within Rtf1. Our conservation mapping also identified an additional conserved surface on Rtf1, which contains two of the three Plus3 residues (R366 and R388; Figure 13A). This newly identified surface is completely separate from any known Rtf1 interaction surface, and we hypothesized that it may represent an additional binding site within the Plus3 domain.

Table 1. Data collection and refinement statistics for human Rtf1 Plus3 domain

	Human Rtf1 Plus3	Plus3-pCTR complex
Data collection		
Space group	$P2_12_12_1$	$C2_1$
Cell dimensions		
a, *b*, *c* (Å)	33.6, 68.0, 117.0	112.9, 172.5, 58.48
α, β, γ (°)	90.0, 90.0, 9.0	90.0, 107.0, 90.0
Unique Reflections	14,681	40,278
Resolution (Å)	50.0 - 2.12 (2.16-2.12)	50.0 - 2.43 (2.52-2.43)
R_{merge} (%)	8.5 (62.4)	6.8 (37.6)
$I / \sigma I$	9.82 (2.60)	14.6 (2.31)
Completeness (%)	91.9 (94.1)	99.9 (99.9)
Redundancy	4.6 (4.5)	3.6 (3.4)
Refinement		
Resolution (Å)	29.2-2.12 (2.19-2.12)	46.9-2.42 (2.51-2.43)
R_{work} / R_{free} (%)	20.6/24.0 (26.3/33.4)	17.7/22.4 (22.6/29.4)
Number of. atoms		
Protein	2109	6484
Peptide		235
Solvent	101	280
B-factors (Å^2)		
Protein	51.1	49.5
Peptide		69.2
Solvent	53.8	68.5
R.m.s. deviations		
Bond lengths (Å)	0.006	0.004
Bond angles (°)	0.94	0.92
Ramachandran		
Outliers (%)	0.00	0.00
Allowed (%)	1.19	1.74
Favored (%)	98.8	98.3

Values in parentheses are for highest-resolution shell.

$R_{merge} = (|(\Sigma I - \langle I \rangle)|)/(\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.

$R_{work} = \Sigma_{hkl}||F_{obs}(hkl)|| - F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$.

R_{free} represents the cross-validation *R* factor for 5% of the reflections against which the model was not refined.

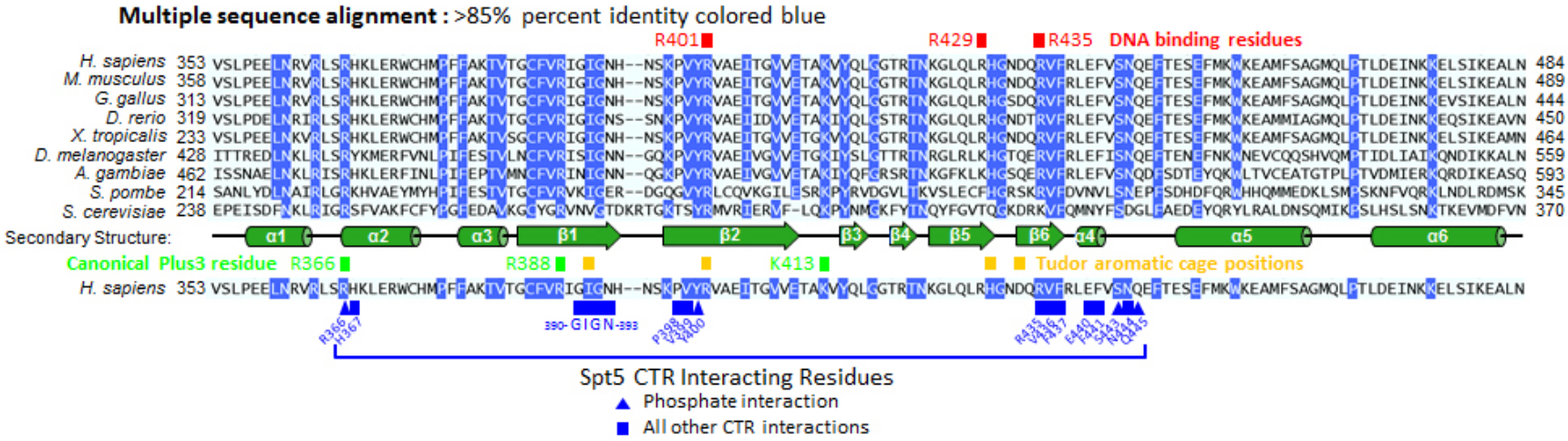

Multiple sequence alignment : >85% percent identity colored blue
R401
R429
R435
DNA binding residues
H. sapiens 353 VSLPEELNRVRLSRHKLERWCHMPFFAKTVTGCFVRIGIGNH--NSKPVYRVAETTGVVETAKVYQLCGTRTNKGLQLRHCNDQRVFRLEFVSNQEFTESEFMKKWKEAMFSAGMQLPTLDEINKKELSIKEALN 484
M. musculus 358 VSLPEELNRVRLSRHKLERWCHMPFFAKTVTGCFVRIGIGNH--NSKPVYRVAETTGVVETAKVYQLCGTRTNKGLQLRHCNDQRVFRLEFVSNQEFTESEFMKKWKEAMFSAGMQLPTLDEINKKELSIKEALN 489
G. gallus 313 VSLPEELNRVRLSRHKLERWCHMPFFAKTVTGCFVRIGIGNH--NSKPVYRVAETTGVVETAKVYQLCGTRTNKGLQLRHCSDQRVFRLEFVSNQEFTESEFMKKWKEAMFSAGMQLPTLDEINKKEVSIKEALN 444
D. rerio 319 VSLPDELNRIRLSRHKLERWCHMPFFAKTVTGCFVRIGIGNS--SNKPVYRVAETIDVVETAKIYQLCSTRTNKGLQLRHCNDTRVFRLEFVSNQEFTESEFMKKWKEAMMIAGMQLPTLDEINKKEQSIKEAVN 450
X. tropicalis 233 VSLPEELNKVRLSRHKLERWCHMPFFAKTVSGCFVRIGIGNH--NSKPVYRVAETTGVVETGKVYQLCGTRTNKGLQLRHCNDQRVFRLEFVSNQEFTESEFMKKWKEAMFSAGMQLPTLDEINKKELSIKEAMN 464
D. melanogaster 428 ITTREDLNKLRLSRYKMERFVNLPIFESTVLNCFVRISIGNN--GQKPVYRVAETVGVVETGKIYSLCTTRTNRGLRLKHCTQERVFRLEFISNQEFTENEFNKKWNEVCQQSHVQMPTIDLIAIKQNDIKKALN 559
A. gambiae 462 ISSNAELNKLRISRHKLERFINLPIFEPTVMNCFVRINIGNN--QGKPVYRVAETVGVVETAKIYQFCRSRTNKGFKLKHCSQERVFRLEFVSNQDFSDTEYQKWLTVCEATGTPLPTVDMIERKQRDIKEASQ 593
S. pombe 214 SANLYDLNAIRLGRKHVAEYMYHPIFESTVTGCFVRVKIGER--DGQGVYRLCQVKGILESRKPYRVDGVLTKVSLECFHCRSKRVFDVNVLSNEPFSDHDFQRWHHQMMEDKLSMPSKNFVQRKLNDLRDMSK 345
S. cerevisiae 238 EPEISDFKKLRIGRSFVAKFCFYPGFEDAVKGCYGRVNVCTDKRTGKTSYRMVRTERVF-LQKPYNMCKFYTNQYFGVTQCKDRKVFQMNYFSDGLFAEDEYQRYLRALDNSQMIKPSLHSLSNKTKEVMDFVN 370
Secondary Structure:
α1
α2
α3
β1
β2
β3
β4
β5
β6
α4
α5
α6
Canonical Plus3 residue R366
R388
K413
Tudor aromatic cage positions
H. sapiens 353 VSLPEELNRVRLSRHKLERWCHMPFFAKTVTGCFVRIGIGNH--NSKPVYRVAETTGVVETAKVYQLCGTRTNKGLQLRHCNDQRVFRLEFVSNQEFTESEFMKKWKEAMFSAGMQLPTLDEINKKELSIKEALN
390-GIGN-393
Spt5 CTR Interacting Residues
Phosphate interaction
All other CTR interactions

Figure 12. Alignment of Rtf1 Plus3 sequences

Multiple Rtf1 sequences were aligned using Clustal Omega and colored with Jalview at an identity cutoff of 85%. Invariant Plus3 residues (green) and positions that form the hydrophobic cage in Tudor domains (magenta) are indicated above the alignment. Rtf1 secondary structure observed in the Plus3-pCTR structure is indicated below the alignment and residues making contacts with the Spt5 pCTR peptide are indicated (blue squares). Residues making direct hydrogen bonds with the pT784 are indicated (blue triangles).

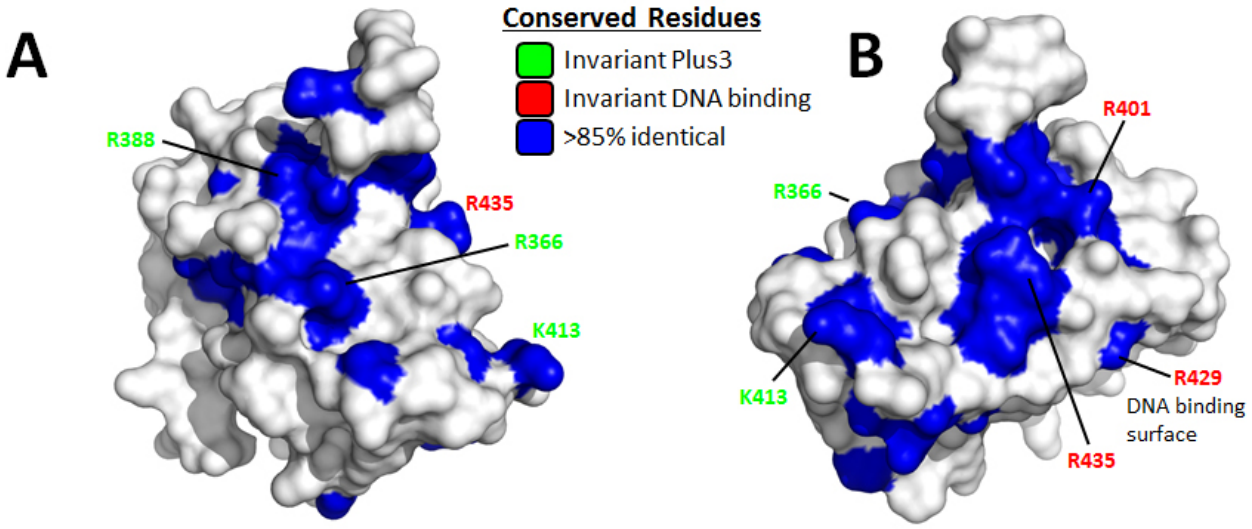


Figure 13. Sequence conservation on the Rtf1 Plus3 domain surface

A and B) Surface representation of the Rtf1 Plus3 domain with residues that contain >85% sequence identity in an alignment of Rtf1 sequences colored blue. Residues involved in DNA recognition are labeled red. The invariant Plus3 residues are labeled green.

2.2.2 Phosphospecific recognition of a Spt5 CTR motif by the Rtf1 Plus3 domain.

The transcription elongation factor Spt5 exists as a dimer with Spt4 in a complex known as DSIF in humans [306] and contains between one and fifteen CTR repeats (Figure 9). The CTR motifs can be phosphorylated by the kinase Bur1/P-TEFb [266, 267, 292], which was shown to enhance

recruitment of Paf1C to chromatin [202, 265-267, 273, 296], and enhanced RNA Pol II processivity during transcription elongation [58, 292]. Recent biochemical and in vivo evidence has demonstrated a direct interaction between the Rtf1 Plus3 domain and the CTR of Spt5 in yeast [273]. We sought to characterize the human Plus3-Spt5 interaction and determine whether this interaction was phospho-specific. To clarify the interpretation of biochemical results and provide a unique binding register for our structural work, we synthesized two Spt5 peptides corresponding to residues 779 to 790 of human Spt5. These peptides contain one complete CTR repeat including its spacer and were constructed either with or without a phosphothreonine at position T784 (Figure 14A). We labeled both peptides with 6-carboxyfluorescein at their N-terminus and monitored binding of recombinant human Rtf1 Plus3 domain to these peptides using fluorescence anisotropy (Figure 14B). We found that the Plus3 domain binds the phosphorylated form of the Spt5 CTR (pCTR) with a calculated equilibrium dissociation constant of 5.45 ± 0.38 μM (Figure 14B). Binding to unphosphorylated peptide resulted in a negligible change in anisotropy, with an estimated Kd of 1.57 ± 3.78 mM. These data suggest that the Plus3 domain can recognize and directly bind a single phosphorylated repeat of the Spt5 CTR and strongly suggest Paf1C recruitment to RNA Pol II is mediated through a direct interaction with Spt5.

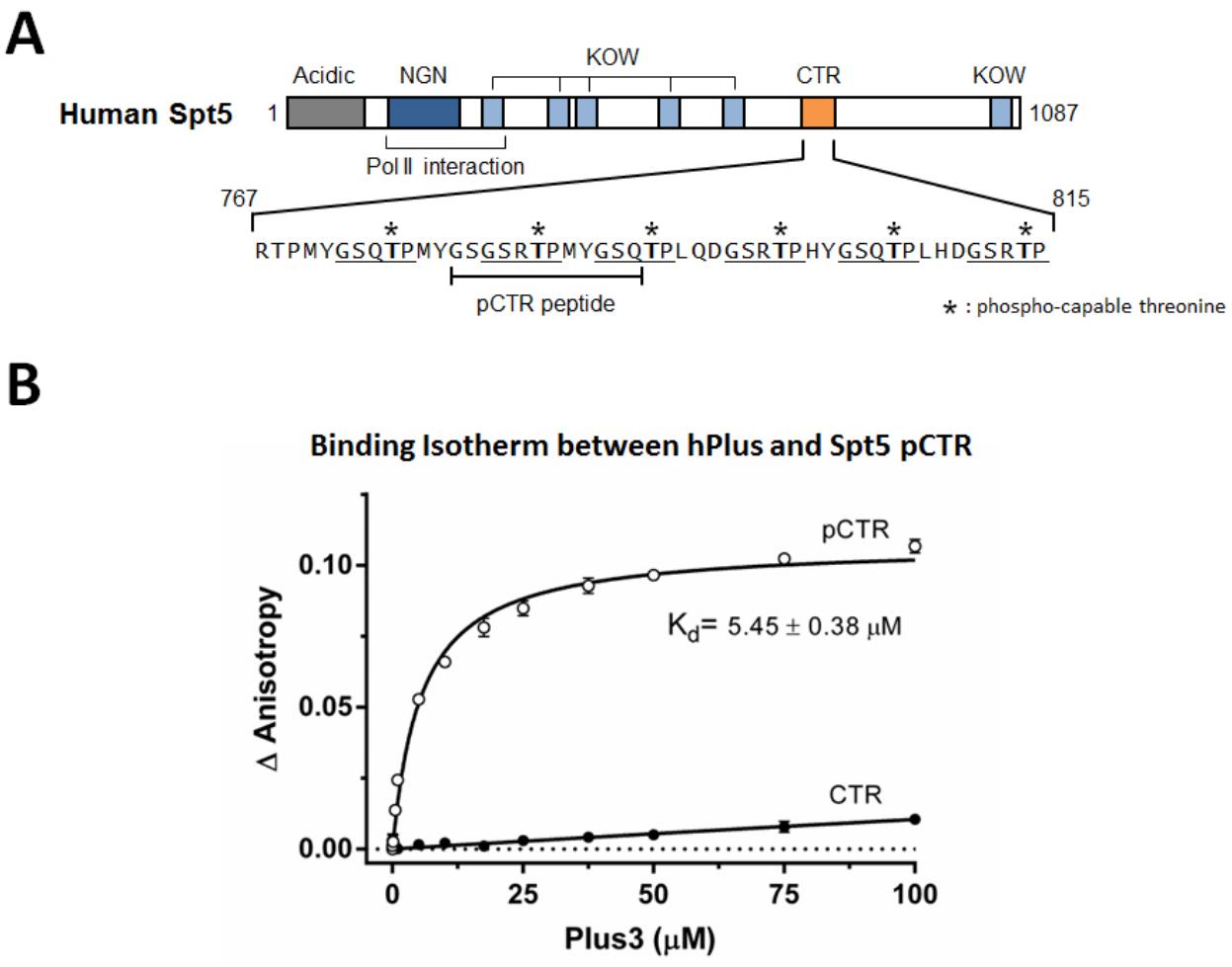


Figure 14. Human Rtf1 Plus3 domain recognizes phosphorylated Spt5 CTR

A) Domain architecture of human Spt5. Spt5 sequence (767-815) containing the CTR repeats is expanded, with pentapeptide CTR motifs underlined and phospho-capable threonine residues indicated with an asterisk. B) Increasing concentrations of purified Plus3 were titrated into 20 nM of fluorescein-labeled pCTR or CTR peptide and binding was assessed via fluorescence anisotropy. Binding measurements were performed in triplicate and the standard deviation plotted. The K_d was extrapolated using nonlinear regression.

2.2.3 Structure of the Plus3-Spt5 pCTR complex.

To determine the molecular basis for phosphospecific recognition of a Spt5 CTR repeat by the Rtf1 Plus3 domain, we crystallized and determined the structure of a Plus3-pCTR complex (Figure 15). Phases were determined using molecular replacement and the Plus3 structure as a search model. The final model was refined at 2.43 Å resolution with R_{work} and R_{free} values of 17.7 and 22.4 %, respectively (Table 1). Six Plus3 molecules are contained within the asymmetric unit, four of which are in complex with a CTR motif (Figure 16). For clarity we are describing the complex that is best resolved and whose model is most complete, however all four complexes adopt a similar conformation (maximum r.m.s.d of 0.8 Å over 131 Cα atoms). The two remaining Plus3 proteins are making crystal contacts that occlude the peptide binding cleft. The conformation of the Plus3 domain is not significantly altered as a result of peptide binding, with an r.m.s.d. of 0.8 Å over 133 Cα atoms. Similar to what was seen in the unbound structure, the loop between β1 and β2 was disordered in two of the molecules in the asymmetric unit.

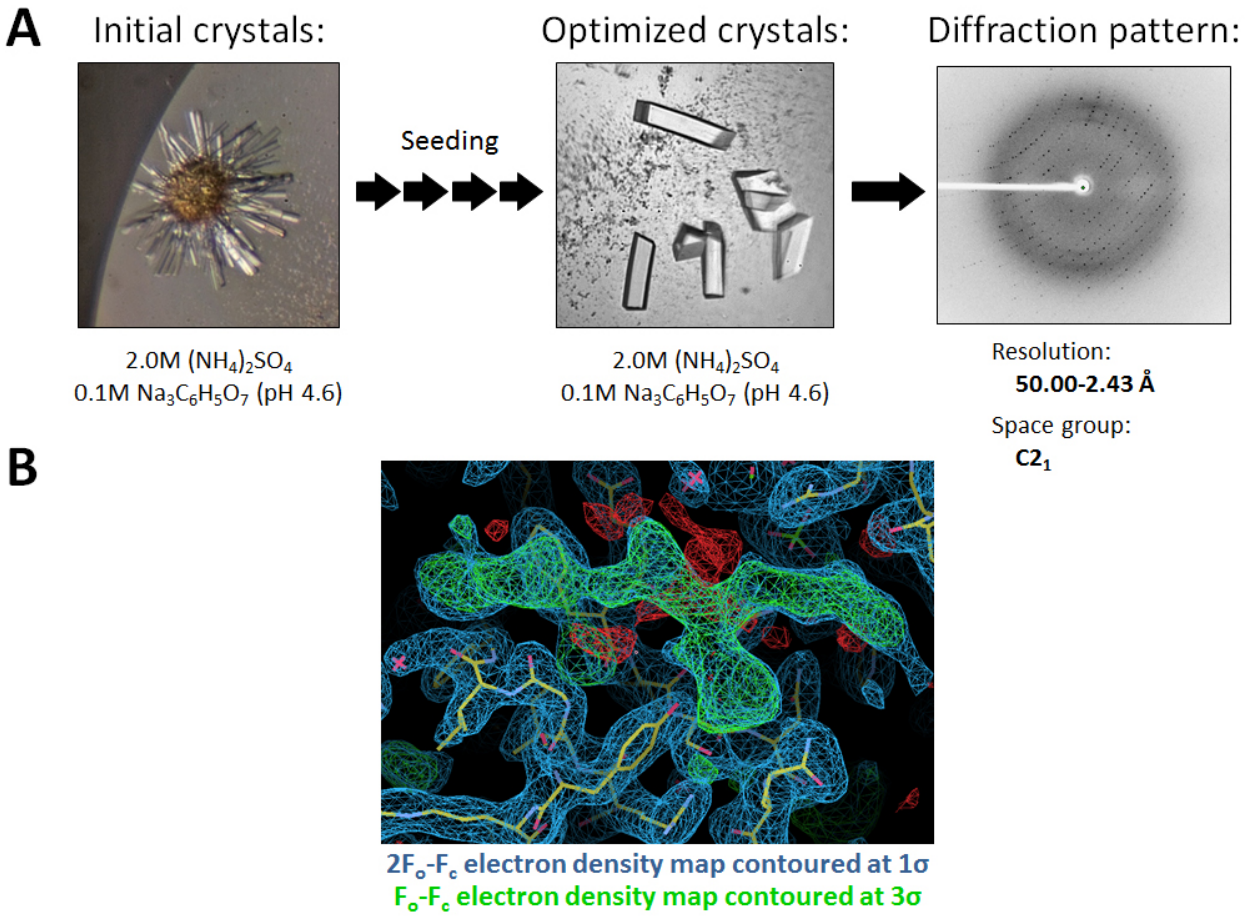


Figure 15. Crystallization and structure determination of the human Rtf1 Plus3-Spt5 pCTR complex

A) Schematic of the crystallization and data collection process for the Plus3-Spt5 pCTR complex. Representative images from the crystallization process are shown, including the buffer conditions used in each case. A frame from the x-ray diffraction data is shown along with the resolution range and experientially determined space group. B) Section of the electron density map of the human Rtf1 Plus3-Spt5 pCTR complex structure. There is clear positive difference density for the Spt5 pCTR peptide (green mesh).

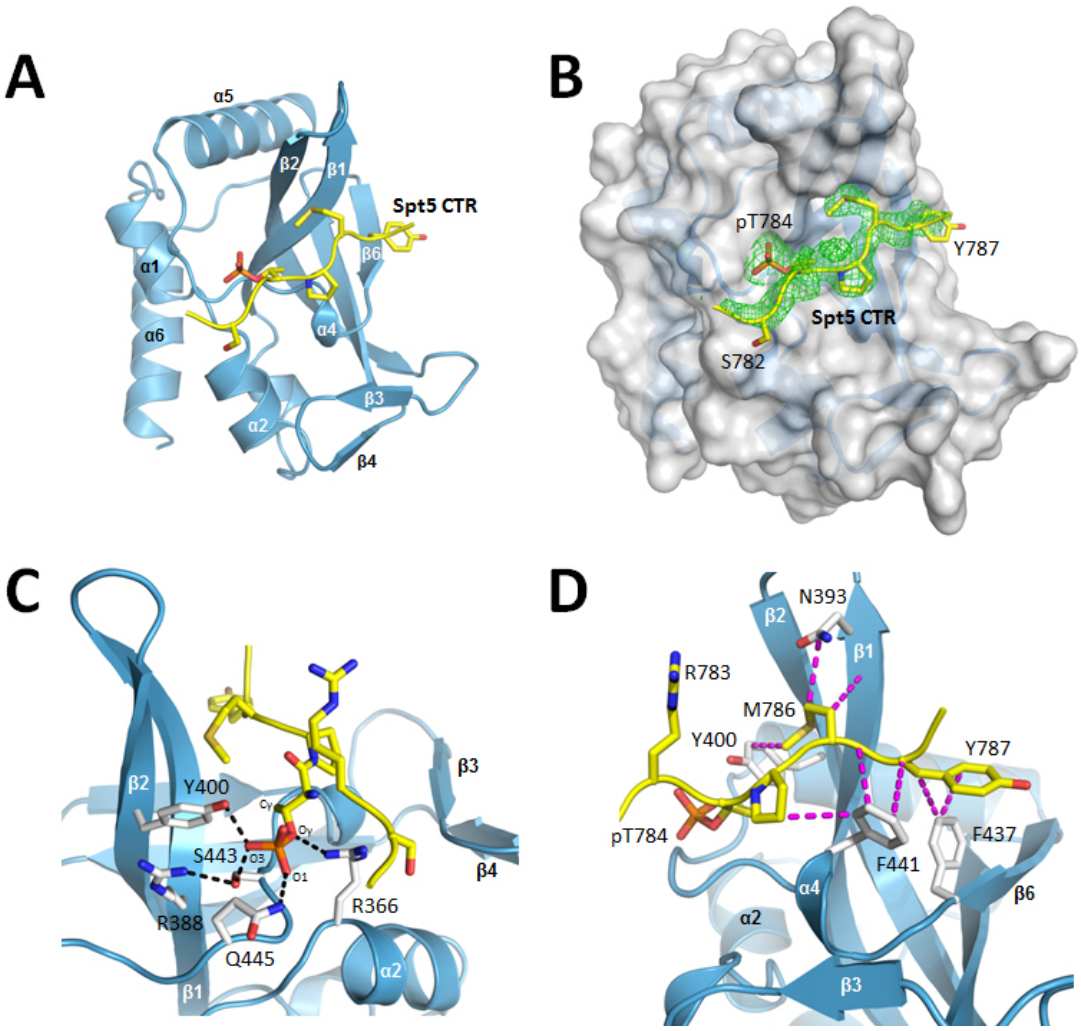


Figure 16. Human Rtf1 Plus3 uses two surfaces to recognize the Spt5 pCTR motif

A) Ribbon diagram of the Rtf1 Plus3 domain (blue) in complex with the Spt5 pCTR (yellow). B) Surface view of the Plus3 domain (white) in complex with the Spt5 pCTR peptide (yellow). Omit map density (Fo-Fc, contoured at 4.0 σ, green) is shown for the peptide. The side chain of Spt5 R783 has been removed for clarity. C) The phosphothreonine recognition pocket. The Spt5 pCTR peptide is shown in yellow. Critical Plus3 residues involved in peptide recognition (white sticks) and the hydrogen bonds involved in recognition (black) are indicated. D) The hydrophobic Spt5 pCTR binding surface. Van der Waal interactions (magenta) between Plus3 residues (white) and the pCTR peptide (yellow) are shown. Spt5 residue pT784 and P785 are contained within the CTR motif while residues M786 and Y787 are found within the spacer between CTR repeats and are making significant contacts with the Rtf1 Plus3 domain.

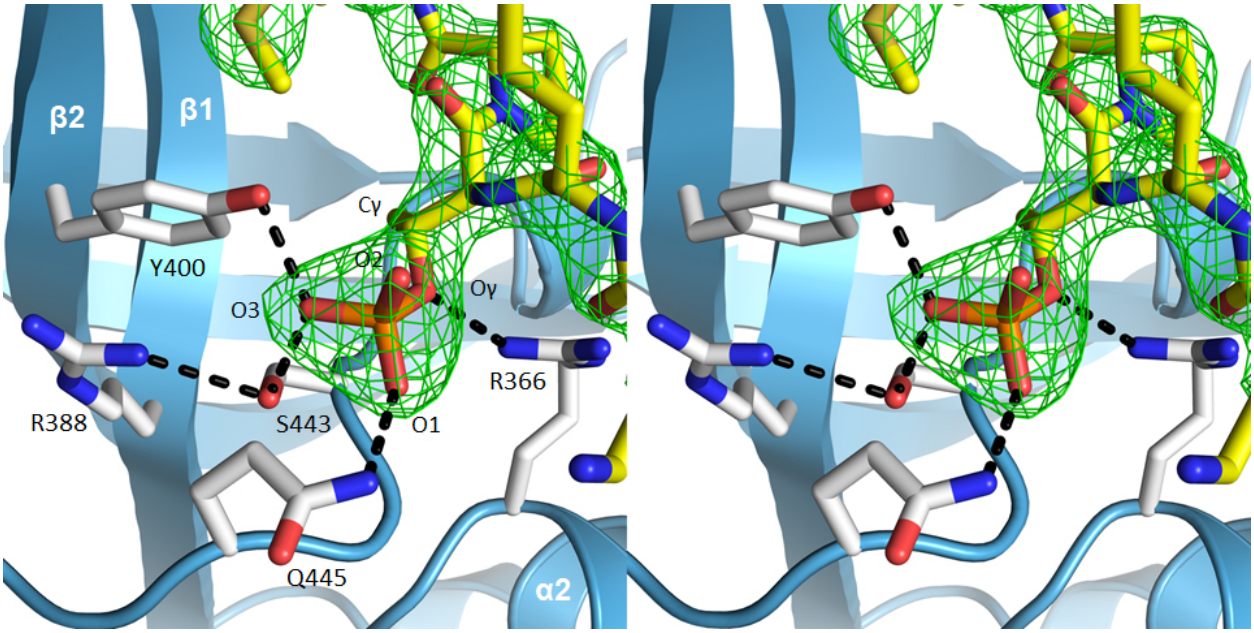


Figure 17. Stereo image of the phosphothreonine recognition pocket

The Spt5 pCTR peptide is shown in yellow with omit map density (Fo-Fc contoured at 4.0 σ) in green. Critical Plus3 residues involved in peptide recognition (white sticks) and the hydrogen bonds involved in recognition (black) are indicated.

The CTR peptide lies across a shallow groove formed by the β1- β2 central hairpin and the loop connecting β6 and α5 (Figure 16A,B), burying 443 Å2 of surface area on the Plus3 domain. The Plus3-Spt5 interface is distant from the location of methyl binding in Tudor domains, and largely overlaps the conserved surface observed earlier (Figure 13). Interactions between the Plus3 and the Spt5 pCTR peptide can be divided into two sections, a phosphothreonine recognition interface and a hydrophobic peptide binding interface. Recognition of the CTR phosphothreonine is accomplished through a series of hydrogen bonding interactions that form a cage around the phosphate. Two of the conserved positively charged residues from which the Plus3 domain derives its name are found in this interface. One of these, R366, makes a hydrogen bond with the gamma oxygen of pT784, orienting the position of the

threonine residue within the binding pocket (Figure 16C, 17). The positioning of pT784 is also conferred by hydrophobic interactions between the gamma carbon of the pThr residue and the aromatic ring of Y400 and the Cα positions of S443 and F441. The other canonical Plus3 residue in this interface, R388 stabilizes the positions of Y400 and S443 both of which are making direct hydrogen bonds to the O3 position of the phosphate moiety (Figure 16C, 17). Combined with residue Q445, which is making a direct hydrogen bond to the O2 oxygen, this network of hydrogen bonding interactions recognizes 3 of the 4 oxygens within the pT784, providing a mechanism for phospho-specific recognition by the Plus3 domain.

The second Plus3-pCTR interface is hydrophobic, and unexpectedly, we found it primarily contained Spt5 residues from the spacer region between repeats (Figure 14A, 16D). The size of this spacer in humans is often conserved and typically contains residues with hydrophobic character (Figure 18). In the complex we crystallized, Plus3 residues Y400, P398, G390, and G392 combine to form a small pocket that accommodates Spt5 spacer residue M786. The first spacer position is occupied by a large hydrophobic residue such as Met, His, or Leu, any of which could be accommodated within this pocket. The second spacer position, Y787 in our peptide, is interacting through its aromatic ring with a hydrophobic surface on Rtf1 formed by residues F441, F437, and I391. While the tyrosine in our peptide is the consensus residue at this position, it seems quite reasonable that the Plus3 surface could accommodate glutamine and histidine residues which are also found in this position. Outside of the phosphothreonine recognition surface, the central organizing residue on the Plus3 surface is F441. This residue makes van der Waals contacts with Spt5 residues P785, M786, and Y787 (Figure 16D). Interactions with P785 are with the Cβ and Cγ positions explaining its preference for this cyclic residue, while interactions with M786 and Y787 are largely through the peptide backbone

allowing for some flexibility in CTR sequence recognition. Together the structure of the Rtf1 Plus3 domain in complex with a pCTR peptide reveals the basis for phosphothreonine specific recognition of Spt5 while interactions outside of pT784 can accommodate the observed variation in CTR sequence.

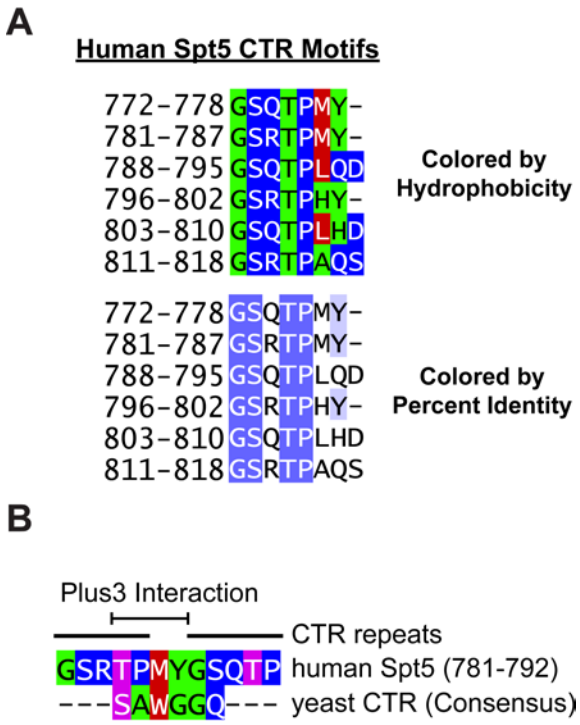


Figure 18. Conservation within the human Spt5 CTR motif

A) The six recognized CTR motifs from human Spt5 were aligned and colored by hydrophobicity [307] or percent identity. Hydrophobicity color scheme: Hydrophobic - Red>Green>Blue – Hydrophilic B) Alignment of human and *S. cerevisiae* Spt5 CTR repeats, highlighting conservation of chemical properties. Repeats are aligned based on their phosphorylated CTR residues. Residues are colored based on hydrophobicity, using the scheme in panel A. Phosphorylated residues are highlighted in magenta.

2.2.4 Identification of critical residues within the Plus3 domain that mediate pCTR binding.

The binding interface observed in our Plus3-pCTR complex is separate from both the DNA binding surface identified by NMR chemical perturbation studies [278] and the canonical methyllysine binding site in Tudor domains [301], suggesting the Plus3 domain has evolved a separate binding surface specifically to interact with Spt5 (Figure 13B). We next tested the importance of residues at the Plus3-pCTR interface for peptide binding in solution. Guided by our structural data, we used site-directed mutagenesis to alter amino acids that lie on the Plus3 surface, avoiding residues that may play a significant role in conferring protein stability. We tested the ability of purified wild-type and mutant Plus3 domains to interact with pCTR peptide using differential scanning fluorometry [308-310]. In this assay, a hydrophobic fluorescent dye is used to monitor the change in melting temperature conferred by the addition of ligand. We observe a ΔTm of 1.5°C when using wild-type Plus3 and pCTR peptide as compared to a ΔTm of 0.5°C when using wild-type Plus3 and an unphosphorylated Spt5 peptide (Figure 19A). Introducing Plus3 variants into this assay, we found that a R366A substitution abolished binding, suggesting an important role in peptide binding for this residue. This role may be to orient the side chain of Spt5 T784, thereby positioning the attached phosphate group into the hydrogen bonding cage, and suggesting why the effect of a R366A mutant was stronger than that of Q445A which is also hydrogen bonding to pT784 (Figure 19A). Alanine substitutions within the hydrophobic interface at positions F437 and F441 significantly compromised peptide binding, while a R435A variant had a subtle effect, all consistent with their role in peptide recognition observed in the structure. Substitutions in conserved residues K413 and F440, located outside the

binding interface had no effect (Figure 19A), further validating predictions based on the structure.

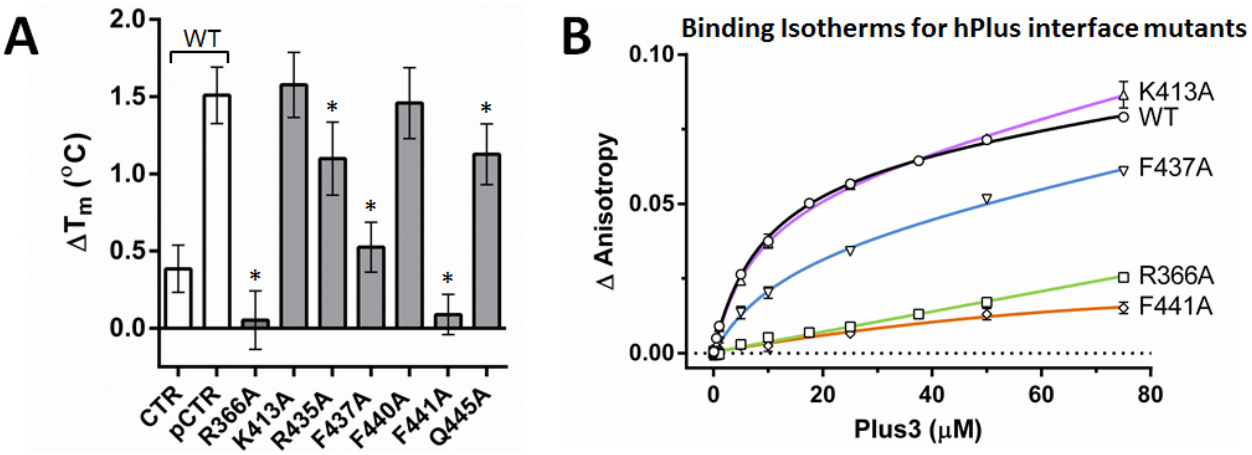


Figure 19. Substitutions of residues at the Plus3-pCTR interface affect peptide binding

A) Increase in the unfolding temperature for Plus3 and the indicated Plus3 variants upon the addition of pCTR peptide as measured by differential scanning fluorometry. Asterisks denote p-values of <0.05. Error bars represent standard deviation from 12 replicates. B) Binding isotherms for human Plus3 or the indicated Plus3 variant to 6-CF labeled pCTR peptide. Increasing concentrations of purified Plus3 variants were titrated into 20 nM of pCTR peptide and binding was assessed via fluorescence anisotropy. Experiments were performed in triplicate and the standard deviation plotted.

We used fluorescence anisotropy to test the importance of residues at the Plus3-pCTR interface using an alternative approach. R366A and F441A substitutions both severely reduced pCTR binding (Figure 19B), in agreement with the results from differential scanning fluorometry. These results suggest that both hydrogen bonding and hydrophobic surfaces are critical for proper pCTR binding. Titrations using a F437A variant of Plus3 resulted in reduced affinity for pCTR, consistent with an important but not critical role in peptide recognition.

Together, these data demonstrate that the pCTR binding surface observed in the crystal is also being utilized for peptide binding in solution. Further it demonstrates that both the phosphothreonine recognition cage and the hydrophobic binding interface centered on F441 are critical determinants for Spt5 binding.

2.2.5 Rtf1 Plus3 residues are important for proper chromatin association.

My structural and biochemical studies of human Plus3 identified several residues critical for Plus3-pCTR interaction *in vitro* (R366, F441). Next we wanted to test if these residues were important *in vivo* in the recruitment of Rtf1 to chromatin. Due to the high degree of conservation of many of these residues in *Saccharomyces cerevisiae* (Figure 12), in collaboration with Manasi Mayekar of the Arndt lab, we analyzed the consequence of making alanine substitutions to Plus3 residues R251 and Y327 in *S. cerevisiae*, which correspond to human Plus3 residues R366 and F441 respectively. To test the importance of these residues in Rtf1 recruitment to chromatin, Manasi performed chromatin immunoprecipitation (ChIP) assays where alanine substitutions of Rtf1 residues R251 and Y327 were introduced together and separately. When incorporated individually or simultaneously, alanine substitutions at residues R252 and Y327 caused a decrease in Rtf1 occupancy on two active genes, *PYK1* and *PMA1* [311]. The R252/Y327 double alanine substitution caused a more severe reduction of Rtf1 recruitment than either of the individual substitutions, causing an approximately additive reduction in chromatin occupancy. None of the alanine substitutions assayed affected the stability of Rtf1 as assessed by western analysis [311]. These results strongly suggest that the Plus3-pCTR interface identified in our structural studies acts as a critical mediator of Rtf1 and Paf1C recruitment to chromatin.

Knowing that the carboxy-terminal domain of Cdc73 (Cdc73 C-domain) is important in Paf1C recruitment to chromatin [264, 265, 273], we reasoned that simultaneous removal of the Cdc73 C-domain and alanine substitutions of Plus3 residues R252/Y327 would result in an even more severe reduction in Paf1C localization to chromatin. To test this, Manasi created yeast strains that contained a *cdc73* C-domain deletion mutation, the *rtf1*-R251A, Y327A Plus3 domain mutation, or both mutations. Using ChIP analysis of Cdc73 occupancy, we found that simultaneous removal of the Cdc73 C-domain and the Plus3 Spt5 CTR-interacting residues (R252/Y327) resulted in an even greater defect in Paf1C chromatin localization than either of the single mutant conditions [311]. The stability of each mutant protein was assessed by western analysis and found to be at wild-type levels [311]. Taken together, these results support the role of our identified Plus3-Spt5 CTR interface in Paf1C recruitment and further support the notion that Paf1C's association with RNA Pol II is facilitated by a dual attachment mechanism.

2.2.6 Role of the Rtf1 Plus3 domain in binding the phosphorylated CTD of RNA Pol II

Recent work in *S. cerevisiae* has demonstrated that Rtf1 is capable of binding RNA Pol II CTD peptides diphosphorylated on serine 2 and 5 (S2P,S5P) [265]. Given the Plus3 domain's affinity for phosphorylated Spt5 CTR motifs, I decided to investigate the possibility that the domain was also capable of binding the Pol II CTD. Given Rtf1's apparent binding specificity for S2P,S5P peptides [265], I used fluorescence anisotropy to test human Plus3's ability to bind an RNA Pol II CTD S2P,S5P peptide (pCTD, Figure 20A). The peptide was fluorescently labeled at its N-terminus with 6-carboxyfluorescein and contained two tandem CTD repeats each diphosphorylated on serine 2 and 5. Surprisingly, I found that the human Plus3 domain binds the

S2P, S5P CTD peptide with an estimated K_d of 4.83 ± 0.46 μM (Figure 20B). Directly comparing this K_d to the Spt5 pCTR K_d (5.45± 0.38 μM) is difficult owing to the differences in the number of repeats and phosphorylation marks used in the two peptides. It is very likely that our calculated K_d for Plus3-pCTD binding is misleadingly tight when compared to Plus3-pCTR binding because it contains an additional 3 phosphorylated moieties.

It is possible that the Plus3 domain is capable of mediating interactions with both phosphorylated motifs, either through individual or overlapping binding surfaces. In an effort to tease apart Plus3's CTR and CTD binding surfaces, I performed a set of fluorescence anisotropy experiments using Plus3 R366 and F441 alanine substitutions, both mutants which largely abolish pCTR binding (Figure 19) [311]. We found that F441 alanine substitution did little to affect pCTD binding, suggesting it is not important for mediating pCTD binding (Figure 20C). The R366 substitution on the other hand significantly reduced pCTD binding, resulting in an estimated K_d of 88.4 ± 14.1 μM (Figure 20D). These data suggest the under the conditions tested the two peptides utilize different peptide binding surfaces that share a common phospho-recognition surface.

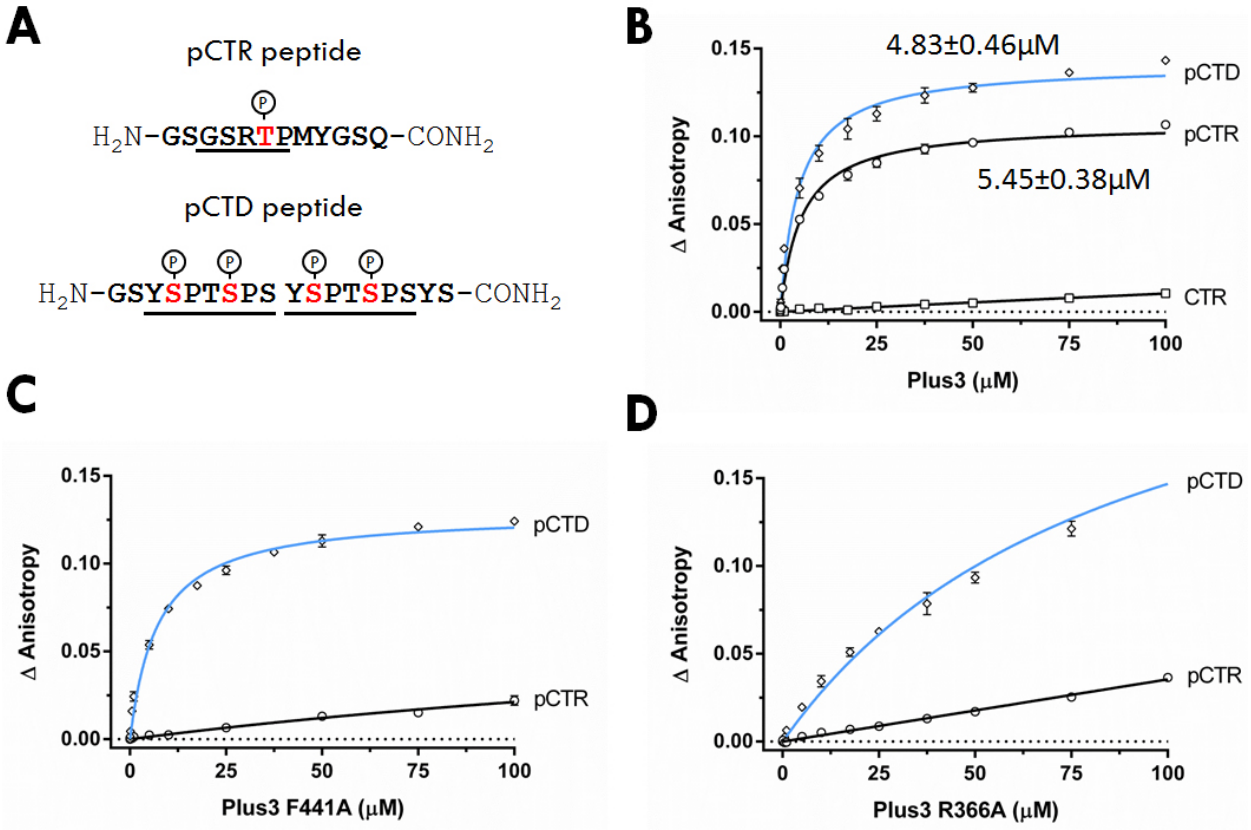


Figure 20. The human Plus3 domain recognizes phosphorylated RNA Pol II CTD

A) Fluorescently labeled Spt5 phospho-CTR peptide and RNA Pol II phospho-CTD peptide used in anisotropy studies. B) Increasing concentrations of purified Plus3 were titrated into 20 nM of fluorescein-labeled pCTD, pCTR or CTR peptide and binding was assessed via fluorescence anisotropy. Binding measurements were performed in triplicate and the standard deviation plotted. The K_d values for pCTD and pCTR were extrapolated using nonlinear regression. C) Binding isotherms for human Plus3 F441A to pCTD and pCTR peptide. Increasing concentrations of purified Plus3 F441A was titrated into 20 nM of peptide and binding was assessed via fluorescence anisotropy. Binding measurements were performed in triplicate and the standard deviation plotted. D) Binding isotherms for human Plus3 R366A to pCTD and pCTR peptide. Increasing concentrations of purified Plus3 R366A was titrated into 20 nM of peptide and binding was assessed via fluorescence anisotropy. Binding measurements were performed in triplicate and the standard deviation plotted.

The Plus3-pCTD interactions I observed were under buffer conditions containing a very low salt concentration (50 mM NaCl) and could simply represent non-specific interactions. Thus before becoming too invested in narrowing down the Plus3-pCTD interaction surface, I wanted to ensure the interaction was resilient to increasing NaCl concentrations. To test this, I titrated increasing amounts of NaCl into a reaction containing Plus3-pCTR or Plus3-pCTD complex and measured dissociation via fluorescence anisotropy. Surprisingly, I found that the Plus3-pCTD interaction dissipates rapidly as the NaCl concentration increases, dropping below 30% bound at 150 mM NaCl. The Plus3-pCTR interaction is much more stable, retaining over 60% bound at 150 mM NaCl. These data suggest that the Plus3-pCTD interaction observed *in vitro* may not be biologically relevant and is likely due to the fact that both peptides share a common phosphate moiety.

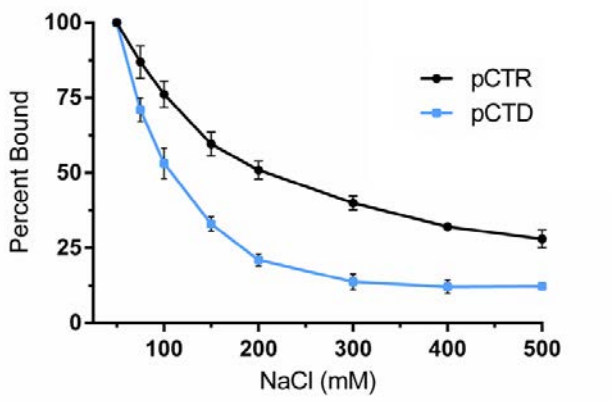


Figure 21. The Plus3-pCTD interaction is labile to NaCl concentration.

Increasing concentrations of NaCl were titrated into 20 nM of fluorescein-labeled pCTR or pCTD bound to 20 μM Plus3 domain and binding was assessed via fluorescence anisotropy. Binding measurements were performed in triplicate and graphed as percent bound with the standard deviation plotted.

2.2.7 Role of the Plus3 domain in binding nucleic acids

When the structure of the human Plus3 domain was first determined, its structural homology to nucleic acid binding proteins, including the Dicer and Argonaute PAZ (Piwi Argonaut and Zwille) domains, led researchers to ask whether the domain was capable of binding DNA [278]. Using electrophoretic mobility shift assays (EMSA), they found the domain was capable of binding a bubble DNA substrate (Figure 22A) with an apparent affinity of approximately 3 μM. Using ^{15}N-HSQC (heteronuclear single quantum coherence) spectroscopy and mutational analysis, they partially mapped the DNA recognition surface on a highly conserved region of the human Plus3 domain, including residues R401, R429, and R435 [278]. This DNA binding surface lies on a distinct surface from the Plus3-Spt5 interface I identified, presenting the possibility that the domain could mediate both interactions.

To begin investigating Plus3's DNA binding properties, I designed DNA substrates mimicking those used in the previous NMR study of Plus3 (Figure 22A) [278]. Minor changes to the single stranded probe were made to reduce internal hairpin formation. I used fluorescence anisotropy to measure human Plus3 affinity to the different DNA substrates. I found that the Plus3 domain preferentially binds the single-stranded DNA substrate (ssDNA) with a calculated K_d of 11.51 ± 0.99 μM (Figure 22B). Plus3 affinity to the double-stranded bubble DNA (dsBubble) and double-stranded DNA (dsDNA) fall off markedly compared to the single-stranded substrate. These data conflict with the previous study showing the Plus3 domain preferentially associates with the bubble DNA substrate [278]. These discrepancies may be partially explained by internal hairpin formation in the single-stranded substrate used in the NMR studies (T_m : 42°C). Additionally, all of the EMSAs reported by de Jong *et al.* were

performed in a buffer containing 75 mM NaCl [278], a concentration that could permit non-specific protein-DNA interactions that are not biologically relevant.

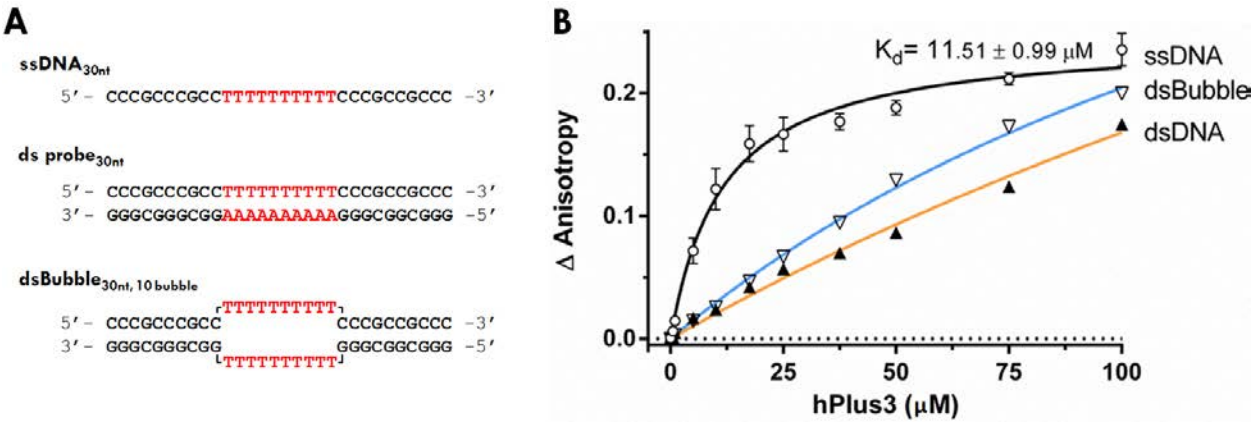


Figure 22. The human Plus3 domain recognizes single-stranded DNA

A) Fluorescently labeled DNA probes used to assess the human Plus3's DNA binding affinity. Probes were adapted from those by de Jong *et al.* [278]. B) Increasing concentrations of purified Plus3 were titrated into 20 nM of fluorescein-labeled ssDNA, dsDNA, or dsBubble DNA and binding was assessed via fluorescence anisotropy. Binding measurements were performed in triplicate and the standard deviation plotted. The K_d for ssDNA was extrapolated using nonlinear regression.

To test whether the Plus3-DNA interactions I observed were stable at biologically relevant NaCl concentrations (100-150 mM), I performed a competition experiment where I titrated increasing amounts of NaCl into a reaction containing human Plus3 bound to fluorescein-labeled ssDNA, dsDNA, or dsBubble DNA and monitored the loss in anisotropy. Similar to what I observed for the Plus3-pCTD interaction, the Plus3 domain's association with all of the DNA substrates drops off dramatically with increasing NaCl concentrations when compared to the interaction with the Spt5 pCTR (Figure 23). These data suggest that the interaction between the

Plus3 domain and DNA is possibly a consequence of the domains highly basic surface and not likely to be functionally relevant.

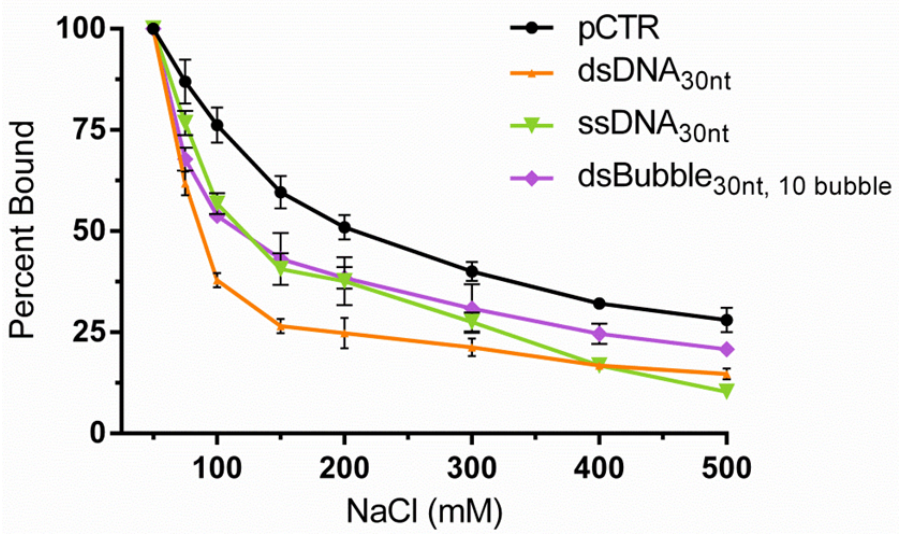


Figure 23. The Plus3-DNA interactionsare labile to NaCl concentration

Increasing concentrations of NaCl were titrated into 20 nM of fluorescein-labeled ssDNA, dsDNA, and dsBubble DNA bound to 20 μM Plus3 domain and binding was assessed via fluorescence anisotropy. Binding measurements were performed in triplicate and graphed as percent bound with the standard deviation plotted.

2.2.8 Rtf1 Plus3 domain may recognize multiple phospho-Spt5 CTR motifs

In the structural studies of the Plus3-Spt5 CTR complex described earlier, only one of five CTR repeats found within human Spt5 was contained within the peptide used during crystallization [311]. Each of the annotated CTR repeats (Figure 14A, 18; consensus: G-S-R/Q-T-P) are arranged sequentially, separated by a short hydrophobic spacer ranging from two to four amino

acids [58]. Due to the close proximity of each CTR motif and that each contains a phospho-capable threonine, it is intriguing to speculate that multiple CTR motifs could work together to enhance binding strength and specificity to the Plus3 domain of Rtf1. In support of this notion, the Plus3 domain has several highly conserved surface-exposed pockets that could potentially accommodate additional CTR motifs (Figure 24). Further, several of these conserved surface exposed regions contain charged and polar residues that could support hydrogen-bonding interactions with CTR phosphothreonines (Figure 24). To begin exploring the possibility that the Rtf1 Plus3 domain is capable of binding multiple Spt5 CTR motifs simultaneously, I conducted molecular dynamic simulations in collaboration with Dr. Michael Grabe and Dr. Pushkar Pendse.

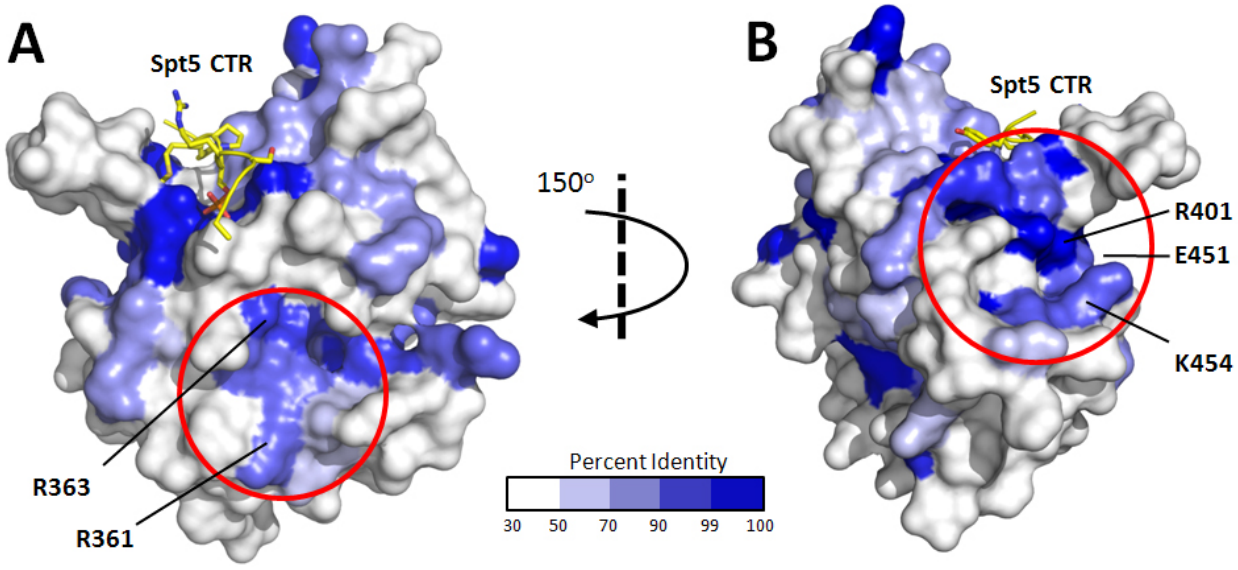


Figure 24. Predicted Spt5 CTR binding sites on human Rtf1 Plus3

A,B) Surface representation of Plus3-Spt5 CTR complex. The surface of the Plus3 domain is colored based on amino acid conservation from a multiple sequence alignment of model organisms (Figure 12). Predicted Spt5 CTR binding surfaces are highlighted with red circles and predicted residues are labeled.

To assess whether multiple Spt5 CTR motifs are capable of binding the Plus3 domain of Rtf1, we conducted a set of molecular dynamic simulations, assessing whether several CTR motifs could associate with the Plus3 domain surface simultaneously. The Plus3-Spt5 CTR structure (PDB id: 4L1U) [311] was used to construct a model containing three tandem Spt5 CTR motifs in Coot [312]. Since the Plus3-Spt5 CTR crystal structure contains a single complete CTR motif, we manually built in the additional residues required to reconstitute another CTR motif to the N and C-terminus of the crystallized CTR peptide. We phosphorylated the threonine residues in the newly built CTR motifs, since they are critical determinates in Plus3-Spt5 binding [311]. Finally, the extended CTR peptide model was checked for any irregular amino acid geometry or steric clashes and modified accordingly.

In human Spt5, the phospho-capable threonine residues of the CTR region are separated from one another by 7-9 amino acids, differing based on the length of the hydrophobic spacer separating CTR motifs (Figure 14A, 18). In a fully extended form, each threonine residue would be separated by roughly 25 Å (assuming an amino acid is 3.5 Å in length). In the crystal structure the visible CTR motif (7 residues), including the hydrophobic spacer residues (M786 and Y787), measures approximately 18 Å. With these distance considerations in mind, one can predict two surface exposed patches on the Plus3 surface that could accommodate another phosphorylated CTR motif (Figure 24). The first patch, spaced roughly 20 Å from T784, contains residues R361 and R363 (Figure 24A). The second is composed of residues R401, E451, and K454 and clusters approximately 24 Å from T784 (Figure 24B). Both of these surface exposed patches are highly conserved and contain polar residues suitable for hydrogen-bonding with CTR phosphothreonines.

Over the course of the 20 ns simulation, the Plus3 domain remained relatively unchanged, maintaining its overall globular fold (r.m.s.d. of 1.04 Å). Unlike the Plus3 domain, the Spt5 peptide deviated from its starting position greatly resulting in RMSD values greater than 5.14 Å. Much of this variation was confined to the N-terminal region of the Spt5 peptide, which contains the first CTR motif (G772-Y778). This region remained solvated for the duration of the simulation, making no notable contacts with the surface of Plus3. The internal CTR motif in the peptide, and that seen in the crystal structure (G781-Y787), remained relatively stationary over the simulation with a maximum RMSD of 1.61 Å. This was due to favorable hydrogen-bonding and hydrophobic interactions of CTR residues with the Plus3 domain. In particular, phosphothreonine 784 remained locked in the hydrogen-bonding cage seen in the crystal structure (Figure 16C, 17). Further, Spt5 residues M786 and Y787 stayed in place due to hydrophobic interactions on the Plus3 surface seen in the structure (Fig. 1E, 3C).

Within the first nanosecond of the simulation, the C-terminal residue of the Spt5 peptide locks itself into position on a surface exposed basic patch of Plus3 comprised of residues K413, Y415, and R438 (Figure 25). This interaction was mediated by the carboxylic acid of D795 hydrogen bonding with Plus3 residues K413 and R438. Van der Waal interactions between Plus3 residue Y415 and Spt5 D795 further strengthen the association (Figure 25B). While the interaction is stable throughout the course of the simulation, it is unlikely to be biologically relevant because CTR phosphothreonine 791 is positioned away from the surface of Plus3. It is interesting to note that K413 is one of the three invariant positively charged residues the Plus3 domain was named after (R366, R388, K413) [278]. The other two invariant Plus3 residues (R366, R388) are utilized to bind CTR 2 in the co-crystal structure (Figure 16, 17). Had the C-terminal D795 residue not contacted this basic patch (K413, Y415, R438) it is possible that

phosphothreonine 791 could have taken its place. Further molecular dynamic simulations are needed to more conclusively assess whether multiple phosphor-CTR motifs can associate the Plus3 domain simultaneously. These simulations would be aided by increasing the simulation times and altering the starting position of the extended Spt5 CTR peptide.

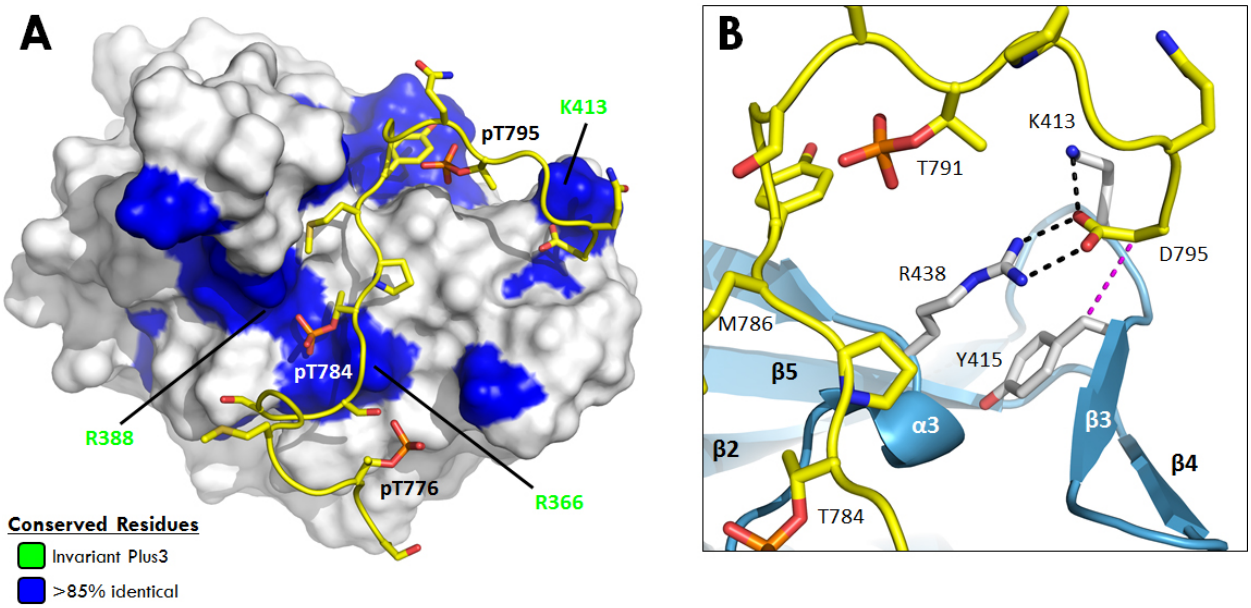


Figure 25. Molecular dynamics simulation of an extended Spt5 CTR peptide binding the Plus3 domain

A) Surface representation of the Rtf1 Plus3 domain bound to an extended Spt5 CTR peptide (yellow sticks) after a 20 ns molecular dynamic simulation. The phosphothreonine in each of the CTR motifs are indicated. Plus3 residue positions that contain >85% sequence identity in an alignment of Rtf1 sequences (Figure 12) are colored blue. The invariant Plus3 residues are labeled green. B) The Plus3 recognition surface for Spt5 D795 (CTR repeat 3). The Spt5 CTR peptide is shown in yellow. Critical Plus3 residues (white sticks) involved in D795 recognition are depicted. The hydrogen bonding (black) and van der Waal (magenta) interactions are indicated as dotted lines.

2.3 CONCLUSIONS

The universally conserved transcription elongation factor Spt4-Spt5 plays an important role in the recruitment of transcription elongation and RNA processing factors to RNA Pol II. The recruitment of these factors is regulated by Bur1/P-TEFb-mediated phosphorylation of short repeated CTR motifs in Spt5 [294, 295]. The CTR could then serve as a scaffold for the phosphorylation-dependent recruitment of factors to transcribed genes in a manner analogous to the CTD of RNA Pol II. Structural data of RNA Pol II CTD peptides in complex with several different interacting proteins [226, 313, 314], has shown a remarkable diversity in the manner in which these proteins recognize specific residues within the CTD motifs. In contrast, the molecular basis for phospho-specific Spt5 CTR motif recognition has remained elusive.

Here we present the first structural insight into this important region of Spt5 in complex with the Plus3 domain of human Rtf1. Interestingly, CTR recognition is contained within a conserved surface distinct from its DNA binding surface [278]. The CTR binding surface is also distinct from binding surfaces utilized by its structural homologs and we believe this represents a new variation in peptide recognition by a Tudor domain. Since the Plus3-pCTR interface is separate from the DNA binding surface, it is possible that the Plus3 domain of Rtf1 could bind DNA and the Spt5 CTR simultaneously. Interactions with DNA could then serve a regulatory role to either enhance or hinder Spt5 binding. A precedent for this type of regulation can be found in the chromodomain of MSL3, which requires recognition of both double stranded DNA and histone H4 mono-methylated on lysine 20 to facilitate chromatin recruitment [315]. While some of our data argues against the biological relevance of the DNA binding observed by the Plus3 domain, further work is needed to show this conclusively.

The human CTR motif (GS[Q/R]TP) used in this study is also conserved in a diverse range of organisms including mouse, zebrafish, mosquito, *S. pombe*, and Arabidopsis [58]. The structural data described here suggest an expansion of the CTR motif to include sequences between traditional repeats. Spt5 residues G782 and S783 do not make significant contacts with the Plus3 domain in our structure; however these residues are strongly conserved within the CTR motif. One explanation for this apparent discrepancy is that while these two positions are not important for Plus3 binding, they are recognized by other CTR binding proteins providing evolutionary pressure to avoid changes at those positions within the CTR. Alternatively, the highly flexible GS pair may provide a malleable linkage between CTR motifs. This could be important for factors that may be simultaneously recognizing multiple CTR motifs. While we cannot rule out this possibility for the Rtf1 Plus3 domain, we note that binding affinity for a single CTR motif already exceeds that of Scaf8, Pcf11, Nrd1, and Rtt103 for a tandem CTD repeat [313, 316], suggesting that increased affinity provided by multiple CTR motifs may not be necessary at least for Rtf1. Lastly, it is possible that the conserved GS residues allows Spt5 sequences containing multiple CTR motifs to adopt additional structure such as has been observed with phosphorylated CTD peptides [226, 313, 317, 318].

Substitution of residues critical for CTR peptide binding disrupted the recruitment of Rtf1 to chromatin in yeast, demonstrating the importance of this conserved interaction surface and providing a molecular mechanism for CTR-mediated recruitment of Paf1C to chromatin. Studies in humans have demonstrated that Spt5 is capable of interacting with Paf1C without the need of CTR phosphorylation or the CTR itself [292] suggesting that there is an additional contact point within Paf1C for Spt4-Spt5. Future effort will be needed to reveal the details of this interaction. Our results on the Plus3-pCTR interaction shed new light on the diverse mechanisms by which

eukaryotes coordinate RNA synthesis with important co-transcriptional processes, including histone modification and recruitment of RNA processing factors.

Moving forward, it will be important to delve deeper into the mechanism behind Spt5-mediated recruitment of Paf1C to chromatin. Of particular interest is determining if the Plus3 domain is able to interact with multiple phosphorylated CTR repeats simultaneously and whether there is an apparent specificity towards one phosphorylated repeat motif over another. Particularly in the case of yeast, where the CTR repeats are more numerous and variable in nature (Figure 9) [266], there is a distinct possibility that the Rtf1 Plus3 domain preferentially recognizes specific repeats.

The Plus3 domain's affinity towards multiple Spt5 CTR repeats could easily be assessed via fluorescence anisotropy and differential scanning fluorometry with larger peptides containing multiple phospho-CTR repeats. Alternatively, the larger Spt5 fragments could be recombinantly expressed in insect cells with a baculovirus expression system, which should correctly modify the CTR repeats.

Aside from assessing the effects of multiple CTR motifs on Plus3-Spt5 association, it would also be interesting to further investigate the possibility that other regions of Spt5 enhance this interaction. Spt5 contains numerous KOW repeats, some of which are known to be important in the interaction of Spt5 with RNA Pol II [274, 319]. Intriguingly, the Spt5 KOW motifs share structural homology to the Plus3 domain [278]. This homology is confined to the β-sheet core of Plus3, specifically aligning with β-strands 1, 2, 5, and 6 with an r.m.s.d of 2.3 Å over 48 Cα atoms. Functionally what this structural homology means is uncertain, but a more detailed investigation of the effect of the KOW motifs on Rtf1-Spt5 interaction could be informative.

3.0 STRUCTURAL CHARACTERIZATION OF THE HISTONE MODIFICATION DOMAIN OF RTF1

3.1 INTRODUCTION

Rtf1 is required for the transcription-coupled monoubiquitylation of H2B K123 and the downstream di- and trimethylation marks of H3 K4 and K79 [177, 205]. Specifically, these marks require a small N-terminal region of Rtf1 known as the histone modification domain (HMD) [177, 205]. Remarkably, the HMD is capable of supporting Rtf1-dependent histone modifications in the absence of the other Paf1C members. In *S. cerevisiae* several residues within the evolutionarily conserved domain have been identified as being required for proper levels of Rtf1-dependent histone modifications, including E104 and F123 [204]. In humans, Paf1C is known to aid in the recruitment of the Rad6–Bre1 ubiquitin conjugase–ligase responsible for the monoubiquitylation mark [154, 206].

The molecular mechanism underlying the role of the HMD in H2B K123 monoubiquitylation and the downstream methylation marks on H3 is currently unknown. To begin understanding the mechanism and physical basis behind HMD function, I determined the crystal structure of a region of the HMD spanning residues 74-139. I have found that the HMD adopts a simple α-helical fold, with several highly conserved surface exposed residues likely serving as interaction surfaces for binding partners. Several of the conserved residues are

required for proper levels of H2B K123 ubquitylation as well as H3 K4 and K79 methylation. The fold of the HMD structurally resembles the high-mobility group protein Nhp6 and the canonical histone fold, potentially providing clues to how the HMD functions.

3.2 RESULTS

3.2.1 Identifying HMD constructs suitable for crystallography

To facilitate expression, purification, and the eventual crystallization of the *S. cerevisiae* HMD, we initially sought to refine the domain's boundaries. The original boundaries of the HMD (amino acids 62-152) were determined using a series of sequential *rtf1* deletion mutations spanning the entire coding region of the gene [177]. To refine these boundaries, Aubrey Lowen in our lab used a collection of techniques including conservation analysis, secondary structural predications, and protein NMR. Secondary structural predictions using the PSIPRED server [320, 321] in combination with conservation analysis of full length Rtf1 suggest the boundaries for the domain fell somewhere between residues 74 through 184 (Figure 26). The amino acid sequences outside these boundaries drop off rapidly in sequence conservation and are not predicted to have secondary structure (until reaching the Plus3 domain). The predicted secondary structure of the region between 74 and 184 is comprised entirely of α-helices and loops. Sequence conservation of the residues aligns well with the regions predicted to be α-helical and can roughly be broken into two regions, one ranging from residues 74-131 and another from 149-178 (Figure 26). Deletion of Rtf1 residues 151-202, results in a moderate Spt$^-$ phenotype,

suggesting its deletion results in a transcriptional defect [177]. This phenotype combined with the sequence conservation and predicted secondary structure of residues 149-178 convinced us that we should focus on two constructs. The first construct was designed to contain the conserved α3 helix and span residues 74-184, while the second from 74-139 would lack the α3 helix.

The Rtf1 sequence from residues 74-139 and 74-184 were cloned into the pET151/D-TOPO (Invitrogen) vector and used to recombinantly express the two fragments in *E.coli* Codon+ (RIPL) cells (Stratagene) using ZY autoinduction media [322]. Both constructs expressed well and were amenable to purification, eluting from an S200 size exclusion column as monodisperse samples with no visible contaminants. After concentrating both proteins (between 15-100 mg/ml), they were put into crystallization trials using the lab's Crystal Phoenix crystallography robot (Art Robbins). For each protein multiple conditions yielded quasi- or pseudo-crystals, appearing as amorphous bodies of protein. Despite extensive efforts to optimize the crystallization conditions in the range of initial hits from the robot screens, both proteins resisted growth of ordered and periodic crystals suitable for data collection. This led Aubrey to establish a collaboration with Dr. Angela Gronenborn's lab (Department of Structural Biology, University of Pittsburgh) to perform NMR spectroscopy analysis of the HMD 74-184 construct. The protein NMR analysis (data not shown) led to the conclusion that the HMD 74-184 construct contains two well folded and structured regions that move independently of one another, one composed of residues 74-139 and the second of residues 140-184.

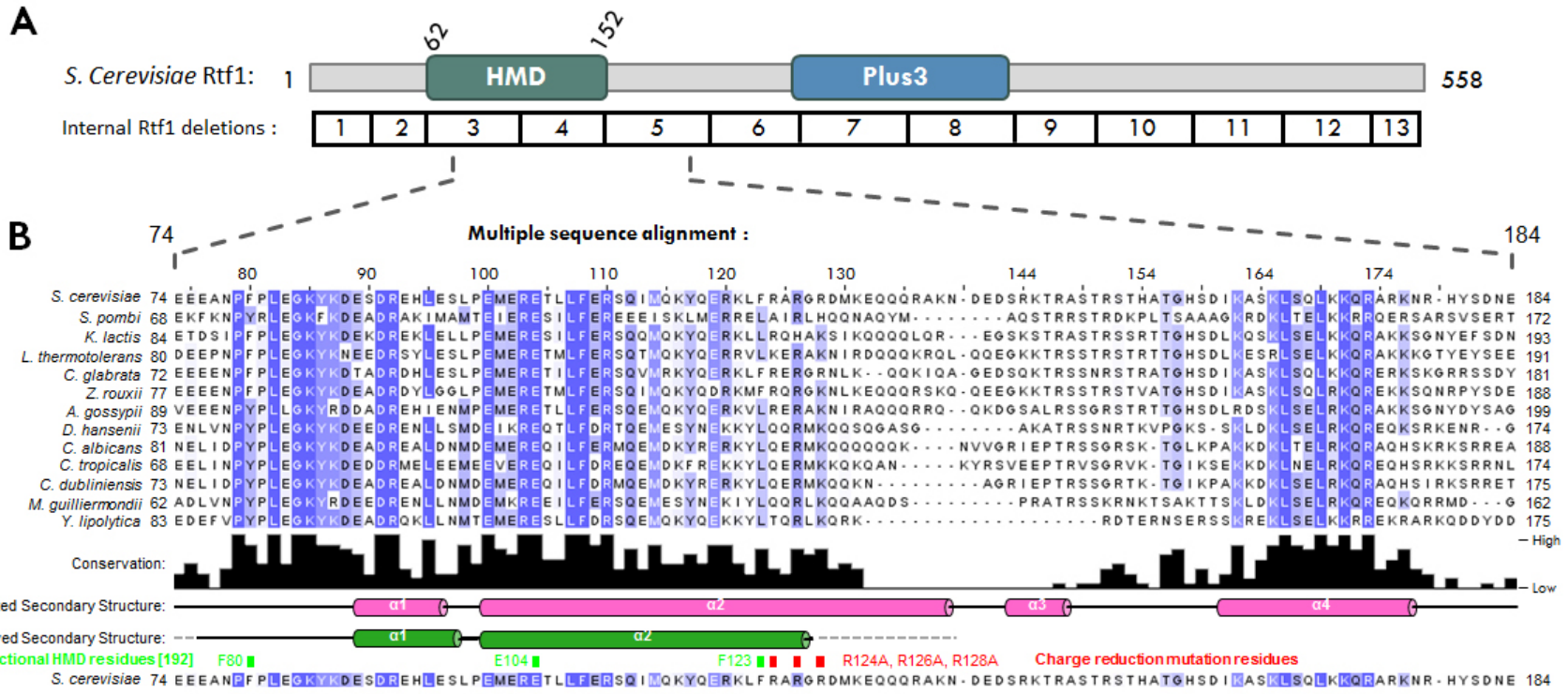

A
S. Cerevisiae Rtf1:
1
62
152
HMD
Plus3
558
Internal Rtf1 deletions :
1
2
3
4
5
6
7
8
9
10
11
12
13
B
74
Multiple sequence alignment :
184
80
90
100
110
120
130
144
154
164
174
S. cerevisiae 74 EEEANPFPLEGKYKDESDREHLESLPEMERETLLFERSQIMQKYQERKLFRARGRDMKEQQQRAKN-DEDSRKTRASTRSTHATGHSDIMASKLSQLKKQRARKNR-HYSDNE 184
S. pombi 68 EKFKNPYRLEGKFKDEADRAKIMAMTEIERESILFERSQQMQKYQERKLLRQHAKSIKQQQQLQR---EGSKSTRASTRSSRTTGHSDLMQSKLSELKKQRAKKSGNYEFSDN 193
K. lactis 84 ETDSIPFPLEGKYKDEKDREKLELLPEMERESILFERSQQMQKYQERKLLRQHAKSIKQQQQLQR---EGSKSTRASTRSSRTTGHSDLMQSKLSELKKQRAKKSGNYEFSDN 193
L. thermotolerans 80 DEEPNPFPLEGKYKNEEDRSYLESLPEMERETMLFERSQTMQKYQERRVLKERAKNIRDQQQKRQL-QQEGKKTRSSTRSTRTTGHSDLMESRLSELKKQRAKKKGTYEYSEE 191
C. glabrata 72 EEEENPFPLEGKYKDTADRDHLESLPEMERETILFERSQVMRKYQERKLFRERGRNLK--QQKIQA-GEDSQKTRSSNRSTRATGHSDIMASKLSQLKKQRERKSKGRRSSDY 181
Z. rouxii 77 EEEENPFPLEGKYKDEADRDYLGGLPEMERETMLFERSQIMQKYQDRKMFRQRGKNLKEQQQRSKQ-QEEGKKTRSSTRSTVATGHSDIMASKLSQLRKQRERKKSQNRPYSDE 188
A. gossypii 89 VEEENPYPLLGKYRDDADREHIENMPEMERETLLFERSQEMQKYQERKVLRERAKNIRAQQQRRQ--QKDGSALRSSGRSTRTTGHSDLRDSKLSELRKQRAKKSGNYDYSAG 199
D. hansenii 73 ENLVNPYPLEGKYKDEEDRENLLSMDEIKREQTLFDRTQEMESYNEKKYLQQRMKQQSQGASG--------AKATRSSNRTKVPGKS-SMLDKLSELRKQREQKSRKENR--G 174
C. albicans 81 NELIDPYPLEGKYKDEADREALDNMDEMEREQILFERMQEMDKYRERKYLQERMKQQQQQQK----NVVGRIEPTRSSGRSK-TGLKPAMKDKLTELRKQRAQHSKRKSRREA 188
C. tropicalis 68 EELINPYPLEGKYKDEDDRMELEEMEEVEREQILFDREQEMDKFREKKYLQERMKKQKQAN-----KYRSVEEPTRVSGRVK-TGIKSEMKDKLNELRKQREQHSRKKSRRNL 174
C. dubliniensis 73 NELIDPYPLEGKYKDEADREALDNMDEMEREQILFDRMQEMDKYRERKYLQERMKQQKN---------AGRIEPTRSSGRTK-TGIKPAMKDKLSELRKQRAQHSIRKSRRET 175
M. guilliermondii 62 ADLVNPYPLEGKYRDEEDRENLLNMDEMKREEILFERSQEMESYNEKIYLQQRLKQQAAQDS---------PRATRSSKRNKTSAKTTSMLDKLSELRKQREQKQRRMD---G 162
Y. lipolytica 83 EDEFVPYPLEGKYKDEADRQKLLNMTEMERESLLFDRSQEMQKYQEKKYLTQRLKQRK--------------------RDTERNSERSSMREKLSELKKRREKRARKQDDYDD 175
Conservation:
High
Low
Predicted Secondary Structure:
α1
α2
α3
α4
Observed Secondary Structure:
α1
α2
Functional HMD residues [192]
F80
E104
F123
R124A, R126A, R128A
Charge reduction mutation residues
S. cerevisiae 74 EEEANPFPLEGKYKDESDREHLESLPEMERETLLFERSQIMQKYQERKLFRARGRDMKEQQQRAKN-DEDSRKTRASTRSTHATGHSDIMASKLSQLKKQRARKNR-HYSDNE 184

Figure 26. *S. cerevisiae* Rtf1 domain diagram and sequence alignment of its internal HMD

A) Domain architecture of *S. cerevisiae* Rtf1, showing the original boundaries of the HMD (62-152) in slate [177]. Below the domain architecture of the protein is a schematic showing the internal deletion mutants of Rtf1 used to identify the HMD and other functional region [177]. B) A region of Rtf1 (74-184) is expanded to show sequence conservation and predicted secondary structure. Multiple Rtf1 sequences were aligned using Clustal Omega and colored with Jalview at an identity cutoff of 30%. Amino acid conservation is displayed as a histogram below the alignment. The predicted secondary structure of the *S. cerevisiae* HMD, estimated with PSIPRED, is shown below the alignment. The secondary structure observed in the HMD_{74-139} R124/6/8A structure is indicated below the alignment, with unobserved residues marked with a dashed line (grey). Functional HMD residues (green) [204] and the residues targeted for surface entropy reduction mutations (red) are indicated below the alignment.

The NMR studies refocused our attention on crystallizing the shorter HMD_{74-139} construct. The protein's ability to form pseudo-crystals in a wide range of buffers and protein concentrations suggested the protein was on the verge of crystallizing but needed one more modification to promote favorable crystal contacts and lattice formation. Despite being only 66 amino acids in length, HMD_{74-139} is extensively charged, containing 16 negatively charged (13 Glu, 3 Asp) and 14 positively charged residues (8 Arg, 6 Lys; Figure 26). This suggested that it might be possible to promote protein-protein crystal contacts by neutralizing the charge of regions of the protein sequence through alanine substitutions. Two sets of charge reduction mutants were designed, one targeting residues E100, E102, and E104 and another targeting residues R124, R126 and R128 (Figure 26). Both sets of residues were substituted to alanines via site-directed mutagenesis and recombinantly expressed in bacteria. The E100/102/104A HMD mutant suffered from low expression levels and was difficult to purify, owing to poor TEV protease cleavage and a tendency to precipitate in solution. Our difficulties working with the E100/102/104A HMD mutant are interesting in light of the functional importance of residue

E104 in yeast [204, 205]. The R124A, R126A, R128A HMD mutant (R124/6/8A) on the other hand expressed and purified normally, eluting from a gel filtration column as a single monodisperse sample. Unlike the wild-type HMD constructs, the R124/126/18A mutant generated crisp well-defined crystals in several conditions without the need for further optimization. Crystals grown in 0.1M phosphate-citrate (pH 4.2) and 40% PEG 300 were looped directly from the crystallography robot screening tray and flash frozen with liquid nitrogen for data collection.

In addition to the wild-type HMD_{74-139} R124/6/8A crystals, crystals with selenomethionine-substituted HMD_{74-139} R124/6/8A were grown in similar conditions (0.1M phosphate-citrate (pH 4.2) and 38% PEG 300). These crystals were cryoprotected by increasing the concentration of PEG 300 to 44% before being flash frozen in liquid nitrogen. The selenomethionine-substituted HMD crystals were used to collect single wavelength-anomalous dispersion (SAD) data to estimate the phases for the diffraction data for the HMD_{74-139} R124/6/8A crystal.

3.2.2 Determining the crystal structure of HMD_{74-139} R124/6/8A

In x-ray crystallography, one measures the diffraction of an x-ray beam by a crystal formed from protein or other macromolecules. These diffracted x-rays are measured by a detector and visualized as spot or reflections, making up a diffraction pattern. The information contained in the diffraction pattern, specifically the indices and intensities of each spot, are used to generate an electron density map for the protein in the crystal. This is done by solving the electron density equation, which requires the indices, the intensities, and phases of each unique reflection in the

x-ray diffraction data. Unlike the indices and intensities, the phases cannot be measured directly from the diffraction data. Instead they must be estimated from another similar protein structure, or an experimental technique like single wavelength-anomalous dispersion (SAD), where anomalous scattering of selenium atoms are used to calculate phases.

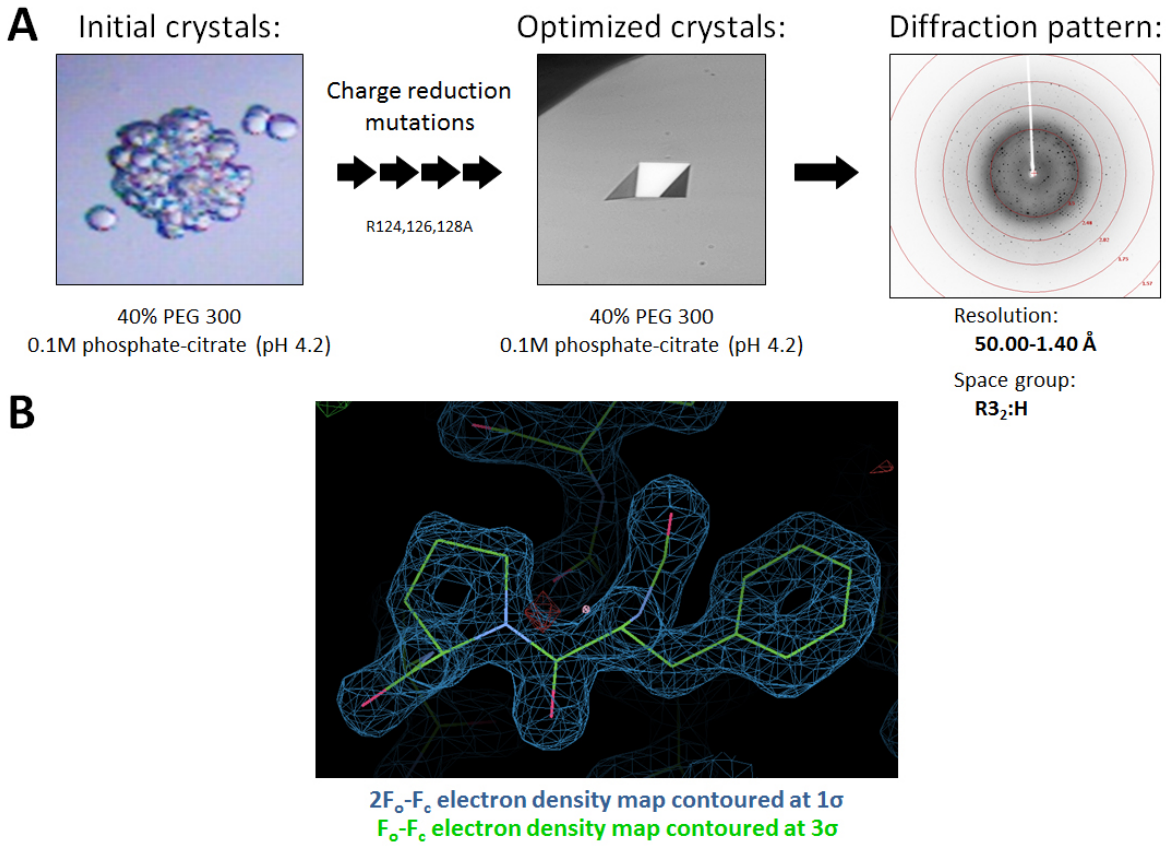


Figure 27. Crystallization and structural determination of *S. cerevisiae* Rtf1 HMD_{74-139} R124/6/8A

A) Schematic of the crystallization and data collection process for HMD_{74-139} R124/6/8A. Representative images from the crystallization process are shown, including the buffer conditions used in each case. The charge reduction mutations critical for crystal optimization are indicated. A frame from the x-ray diffraction data is shown along with the resolution range and experientially determined space group. B) Section of the electron density map of the *S. cerevisiae* HMD_{74-139} R124/6/8A crystal structure.

Native and SAD data for the HMD_{74-139} R124/6/8A crystals were collected by Annie Héroux at the National Synchrotron Light Source on beamline X25A. We were successful in determining the crystal structure of the R124/6/8A HMD mutant, spanning residues 74-139. Phases were determined using the SAD dataset and an interpretable electron density map was generated and used to build an initial model. This was then used as a search model for molecular replacement [323, 324] into the native dataset. The final model was refined at 1.4 Å resolution with R_{work} and R_{free} values of 15.8 and 17.3 %, respectively (Table 2). The crystal's asymmetric unit contains two molecules that are tightly packed together.

In the HMD_{74-139} R124/6/8A structure, residues 77-126 are ordered and are clearly visible in the electron density map. The protein is composed of a short N-terminal α-helix (E89-S97) followed by a longer C-terminal α-helix (E100-R126), assembled anti-parallel to one another at a ~60^o angle (Figure 28A). The HMD's fold is relatively planar, exposing the majority of the residues to the surface, and confining buried surface area to the interface between the two helices. This interface helps maintain the protein's fold through several conserved nonpolar residues (F80, L95, L106), forming a small hydrophobic core (Figure 28B). There are also two highly conserved residues that form electrostatic interactions in our structure (D91, R110). Only two of the charge reduction mutations are visible in the structure (R124A and R126A) (Figure 28A). Both are situated at the end of the larger C-terminal helix of the HMD and are not involved in maintaining the domain's fold. This strongly suggests the overall structure of the domain was not affected by the alanine substitution mutations. It is clear however, that the substitutions were critical in facilitating crystal formation by generating an additional crystal contact (Figure 29). Specifically, alanine substitution at R126 creates several favorable van der Waal interactions with residues E102, T105, and L106 on a neighboring HMD molecule in the

lattice. R124A on the other hand is not close enough to neighboring HMD molecules to form favorable contacts. Since R128A is not visible in the electron density map, one can infer that it is highly flexible in the lattice and is unlikely to be playing a major role in facilitating crystal formation under these conditions.

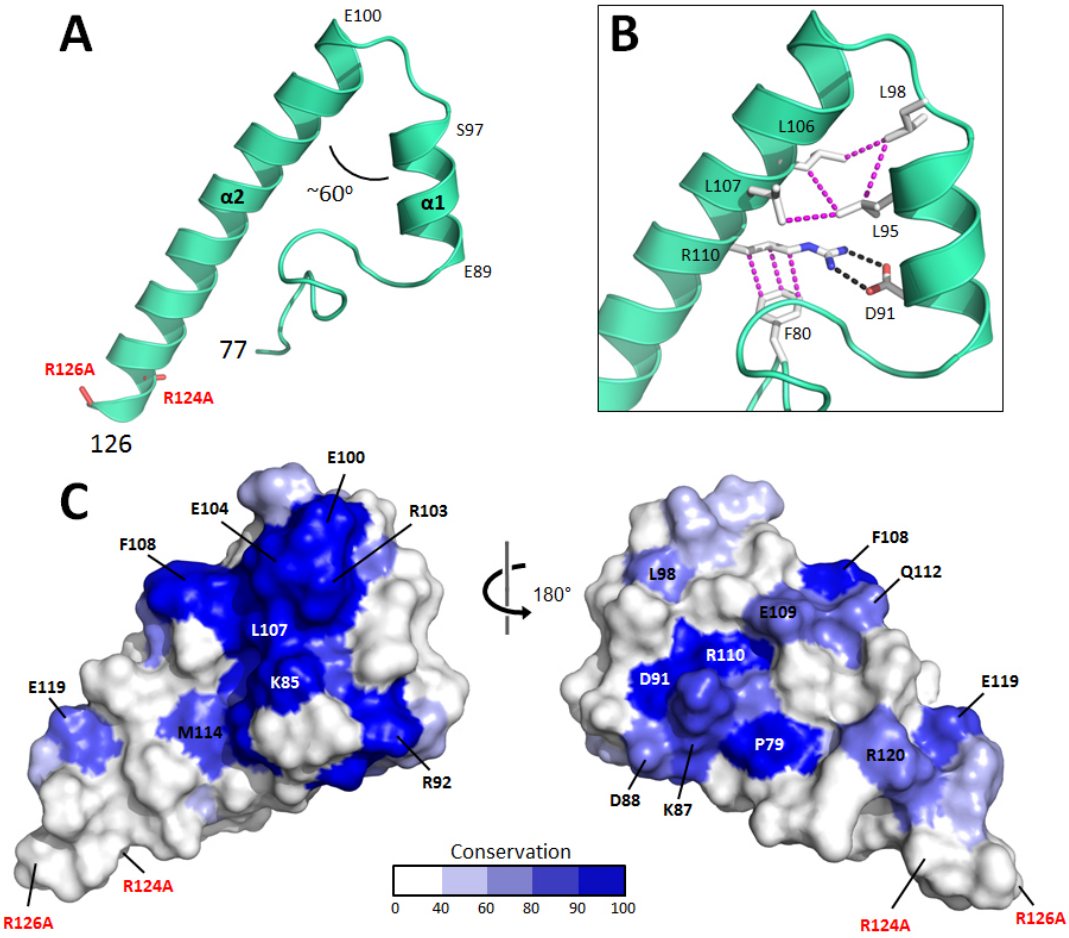


Figure 28. Crystal structure of the *S. cerevisiae* Rft1 HMD$_{74-139}$ R124/6/8A

A) Ribbon diagram of the Rtf1 HMD$_{74-139}$ R124/6/8A crystal structure. Charge reduction mutations critical in crystallization are indicated (red). B) HMD buried surface area. Residues important in maintaining the HMD fold are shown as white sticks. Van der Waal (magenta) and H-bonding (black) interactions important in maintaining the domain's fold are indicated as dashed lines. C) Sequence conservation from a multiple sequence alignment (Figure 26) of Rtf1 sequences was scored and displayed on the HMD surface with low conservation in white and invariant residues in blue. Highly conserved surface exposed residues and charge reduction mutations (red) are indicated.

Table 2. Data collection and refinement statistics for *S. cerevisiae* Rtf1 HMD_{74-139} R124/6/8A

	SeMet	**Native**
Data collection		
Space group	$R3_2$:H	$R3_2$:H
Cell dimensions		
a, *b*, *c* (Å)	94.3, 94.3, 76.4	94.3, 94.3, 76.8
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Unique Reflections	13,524	30,553
Resolution (Å)	50.0 - 1.74 (1.77-1.74)	50.0 - 1.32 (1.34-1.32)
R_{merge} (%)	11.6 (54.7)	5.1 (62.6)
$I / \sigma I$	32.57 (2.40)	77.79 (1.30)
Completeness (%)	100.0 (100.0)	98.3 (78.5)
Redundancy	10.2 (7.2)	13.7 (3.5)
Refinement		
Resolution (Å)		20.00-1.40 (1.46-1.40)
R_{work} / R_{free} (%)		15.8/17.3 (27.4/29.8)
Number of. atoms		
Protein		1831
Water		88
B-factors (Å^2)		
Protein		23.2
Solvent		39.8
R.m.s. deviations		
Bond lengths (Å)		0.009
Bond angles (°)		1.206
Ramachandran		
Outliers (%)		0.00
Allowed (%)		3.19
Favored (%)		96.8

Values in parentheses are for highest-resolution shell.

$R_{merge} = (|(\Sigma I - \langle I \rangle)|)/(\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.

$R_{work} = \Sigma_{hkl}||F_{obs}(hkl)|| - F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$.

R_{free} represents the cross-validation *R* factor for 5% of the reflections against which the model was not refined.

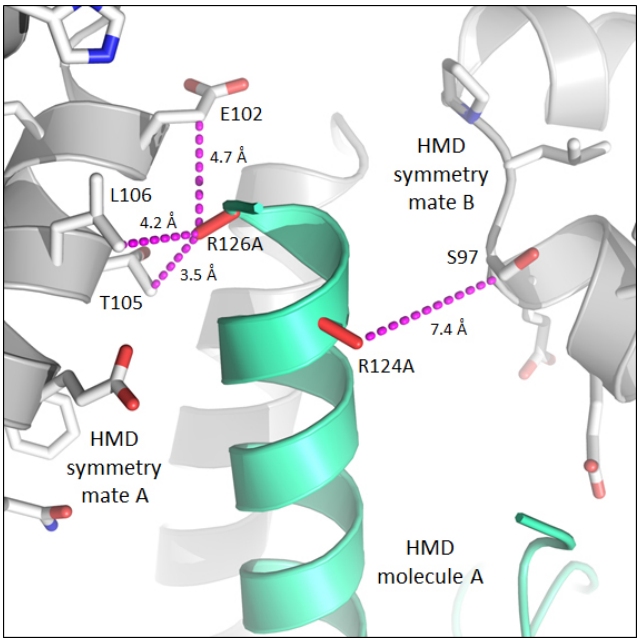


Figure 29. The R126A substitution in HMD 74-139 introduces an additional crystal contact

The R126A substitution in the HMD creates several favorable crystal contacts with neighboring HMD molecules in the crystal lattice. HMD R124/6/8A is shown (cyan) with neighboring symmetry mates in the crystal lattice shown in white. R124A and R126A substitutions are labelled and highlighted as red sticks. Van der Waal interactions important in mediating crystal contacts are shown as dashed lines (magenta) with the distances indicated. The interaction distance between R124A and S97 (7.4 Å, symmetry mate B) is too great for a proper van der Waal interaction.

To begin identifying functionally important residues within the HMD, I performed a multiple sequence alignment of different yeast species (acquired from the Saccharomyces Genome Database) and mapped the alignment on the surface of the HMD as a heat map (Figure 28C). There are several highly conserved surface exposed residues on the HMD, with the most prominent being a patch on one face of the protein containing residues K85, R92, E100, R103, E104, L107, and F108. Based on the requirement of residue E104 for HMD-histone modifications [204], this conserved region likely acts as a protein-protein interaction surface to

mediate Rtf1-dependent histone modifications. On the opposite face of the protein, the two residues D91and R110 are highly conserved and are partially surface exposed, along with residues P79, D88, K87, E109, and Q112 (Figure 28C). Finally, toward the C-terminal region of the domain, there is a moderately well conserved patch comprised of residues E119, R120, and F123. It is interesting to note that simultaneous alanine substitution mutations of residues 120-121 were shown to have little effect on HMD function [204].

3.2.3 Designing HMD_{74-139} mutations for genetic phenotype analysis

Having determined the structure of the R124/6/8A HMD mutant, I designed a set of mutants to probe the residues important for HMD function. The protein's small size (66 residues) and the prevalence of surface exposed residues, allowed me to base candidate mutations primarily off of sequence conservation (Figure 28C). All of the mutants designed target a single amino acid and can be broadly broken up into four groups, based on their location on the structure (Figure 30A, Table 3). Residues in group 1 make up the largest conserved surface on the HMD (E83, K85, R92, E96, P99, E100, R103, E104, L107) and contains residue E104 which is known to be important for HMD function [204]. On the opposite face of the HMD is another conserved surface that contains residues in group 2 (D91, E102, L106, R110). Group 3 residues (F108, Q112, Q115, K116) are located within the center of the long C-terminal helix and form a small basic patch on the HMD's surface (Figure 30A). One final residue, not falling in any of the 4 groups, was chosen for mutation, the internally buried residue L95. The residues targeted for charge reduction mutations in the crystal structure make up group 4 (R124, R126, R128). All of the mutations designed change the target residue to alanine, with the exception of E102 and

R110 which are also mutated to lysine and glutamic acid, respectively. Based on the location and conservation of these residues, we predicted many of the amino acid substitutions would have a profound effect on Rtf1 function.

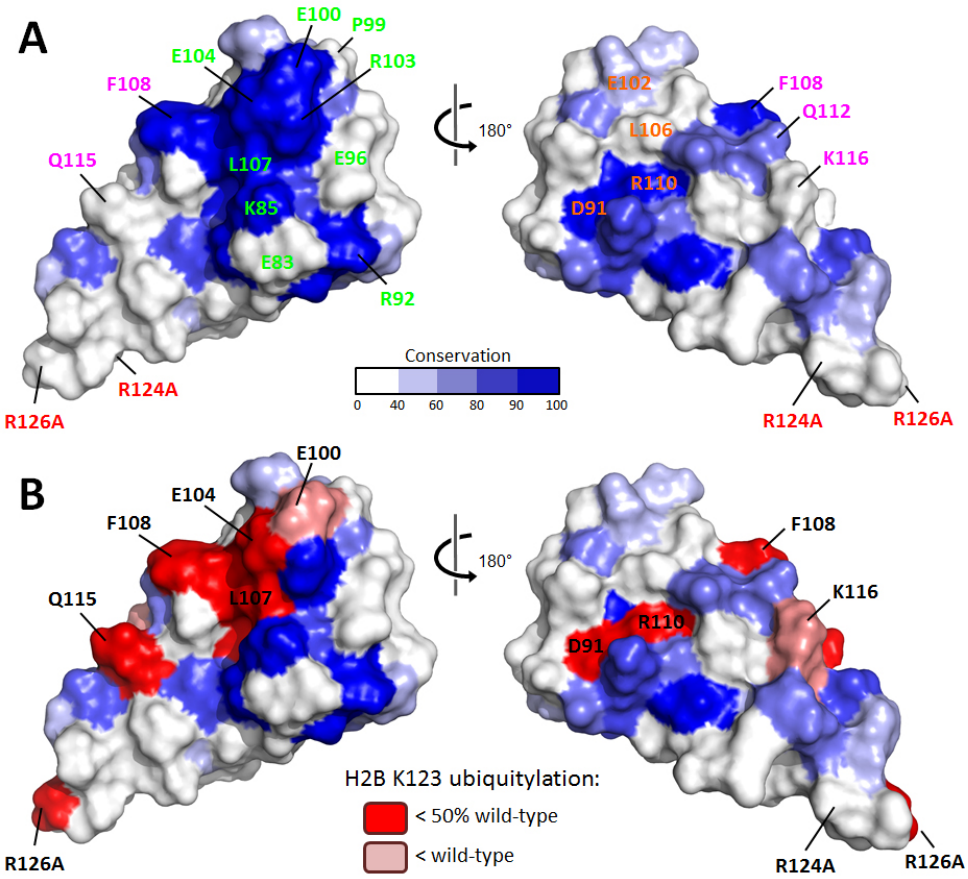


Figure 30. Rtf1 HMD amino acid substitution mutations and H2B K123 ubiquitylation defects

A) Sequence conservation from a multiple sequence alignment (Figure 26) of Rtf1 sequences was scored and displayed on the HMD surface with low conservation in white and invariant residues in blue. Amino acid residues targeted for mutation are indicated and colored based on the group they fall into (Group 1 : green, Group 2 : orange, Group 3 : magenta, Group 4 : red). B) Amino acid substitutions that result in H2B K123 monoubiquitylation defects are highlighted in red. Amino acid substitutions that result in monoubiquitylation levels less than 50% of wild-type are colored in dark red, while those with levels less than wild-type are colored in light red. The phenotypic analysis depicted here was carried out by Branden Van Oss of the Arndt lab.

Table 3. *S. cerevisiae* Rtf1 HMD_{74-139} amino acid substitution mutations

Residue(s)	Substitution(s)	Notes
E83	alanine	Surface exposed, conserved
K85	alanine	Surface exposed, invariant
D91	alanine	Salt bridge residue, invariant
R92	alanine	Surface exposed, invariant
L95	alanine	Maintains HMD fold, conserved
E96	alanine	Surface exposed
P99	alanine	Surface exposed
E100	alanine	Surface exposed, invariant
E102	alanine, lysine	Surface exposed, conserved
R103	alanine	Surface exposed, invariant
E104	alanine, lysine	Functionally important [204], invariant
L106	alanine	Maintains HMD fold
L107	alanine	Maintains HMD fold, invariant
F108	alanine	Surface exposed, invariant
R110	alanine, glutamic acid	Salt bridge residue, invariant
Q112	alanine	Surface exposed, conserved
Q115	alanine	Surface exposed
K116	alanine	Surface exposed
R124	alanine	Charge reduction mutation
R126	alanine	Charge reduction mutation
R128	alanine	Charge reduction mutation
R124, R126, R128	alanine	Charge reduction mutation

In collaboration with the Arndt lab, Branden Van Oss introduced each of the HMD amino acid substitutions into full-length *S. cerevisiae RTF1* on a low-copy expression level vector (pLS21-5). After being transformed into yeast, he assessed the effects of the HMD amino acid substitutions for growth phenotypes associated with transcription-related defects, including the Spt^- phenotype, sensitivity to the base analog 6-azauracil, cryptic initiation within coding regions, and telomeric silencing. He also tested protein stability for each of the mutants via western analysis to avoid confounding effects of the mutations. Importantly, using western analsis he determined the effects of the mutations on the HMD-dependent histone modifications

H2B K123 mono-ubiquitylation and di- and trimethylation of H3 K4 and K79. Branden also performed all of the assays with the R124/6/8A mutant to determine if the charge reduction mutations disrupt HMD function.

As expected, Branden found that several of the residues mutated on the HMD surface have strong defects on HMD function. Substitution mutants in all four groups (Figure 30A) show a range of defects in H2B monoubiquitylation (Figure 30B). The majority of these defects fall along an extended surface of the HMD created by the domain's longer α2 helix and the elbow region between the α1 and α2 helix (Figure 30B; E100, E104, L107, F108, Q115, K116). The two strongest defects seen within this patch are from substitutions of residues E104 and F108. F108 is particularly interesting because bulky hydrophobic residues like phenylalanine are typically confined within the interior of a protein. It is possible that this exposed phenylalanine is critical in mediating protein-protein interactions with a binding partner. Both of the conserved residues D91 and R110, forming electrostatic interactions with one another in the HMD, also exhibit strong defects in H2B monoubiquitylation when mutated (Figure 30B). Interestingly, the charge reduction mutations (R124/6/8A), critical for HMD crystallization, show strong H2B K123 monoubiquitylation defects. Branden has found that the majority of the defect is localized to substitution R126A. It is important to note that while R124, R126, and R128 are required for proper HMD function, their mutation is unlikely to have affected the overall fold observed in the $HMD_{74\text{-}139}$ R124/6/8A structure. In addition to causing H2B K123 ubiquitylation defects, the majority of the HMD substitutions caused defects in H3 K4 and K79 methylation along with other growth phenotypes suggesting transcription-related defects. Together, these data support the idea that conserved residues on the surface of the HMD's fold are needed for HMD function.

3.2.4 HMD structural alignments suggest possible functional mechanisms

To gain more insight into how the HMD functions, I performed a structural comparison of the HMD to all of the structures in the PBD (Protein Data Bank) using PDBeFold [325]. Based on the simplicity of the overall HMD structure (two α-helices), I had to be wary of each of the structural homologs identified. Any protein with an extended α-helix would likely be recognized, making analysis difficult and reliant on other structural or functional similarities. As expected, I received several hundred hits from the structural homology search, the majority of which were recognized for their alignment to the larger α2 helix of the HMD. In the following section I will briefly describe some of the interesting structural similarities found. It is difficult however to draw any definitive conclusions from any of the alignments due to the small size and lack of complexity seen in the HMD fold.

3.2.4.1 The HMD has structural similarities to Nhp6A

HMD R124/6/8A shares structural similarity to the non-sequence specific DNA binding protein Nhp6A (Non-histone protein 6A) in *S. cerevisiae*. The two proteins align with an r.m.s.d of 2.47 Å over 43 Cα atoms (PDB id: 1J5N; Figure 31A). Nhp6A belongs to a class of chromatin-associated proteins known as high-mobility group (HMG) proteins that are important in mediating a wide range of functions including altering chromatin structure and V(D)J junction recombination [326, 327]. Nhp6A, like other HMG proteins, functions by binding and bending the DNA helix. HMG proteins bind DNA with a three helix HMGB motif that forms a characteristic L-shaped fold that sits in the minor groove of DNA [326, 328, 329]. The N-

terminus of Nhp6A wraps around and contacts the major groove of DNA with numerous basic residues. DNA bending occurs at two conserved hydrophobic amino acids that act as wedges to distort the DNA towards the major groove. In yeast, Nhp6A has been suggested to play a role in recruiting the chromatin chaperone complex FACT (Spt16-Pob3) to chromatin (reviewed in [330]).

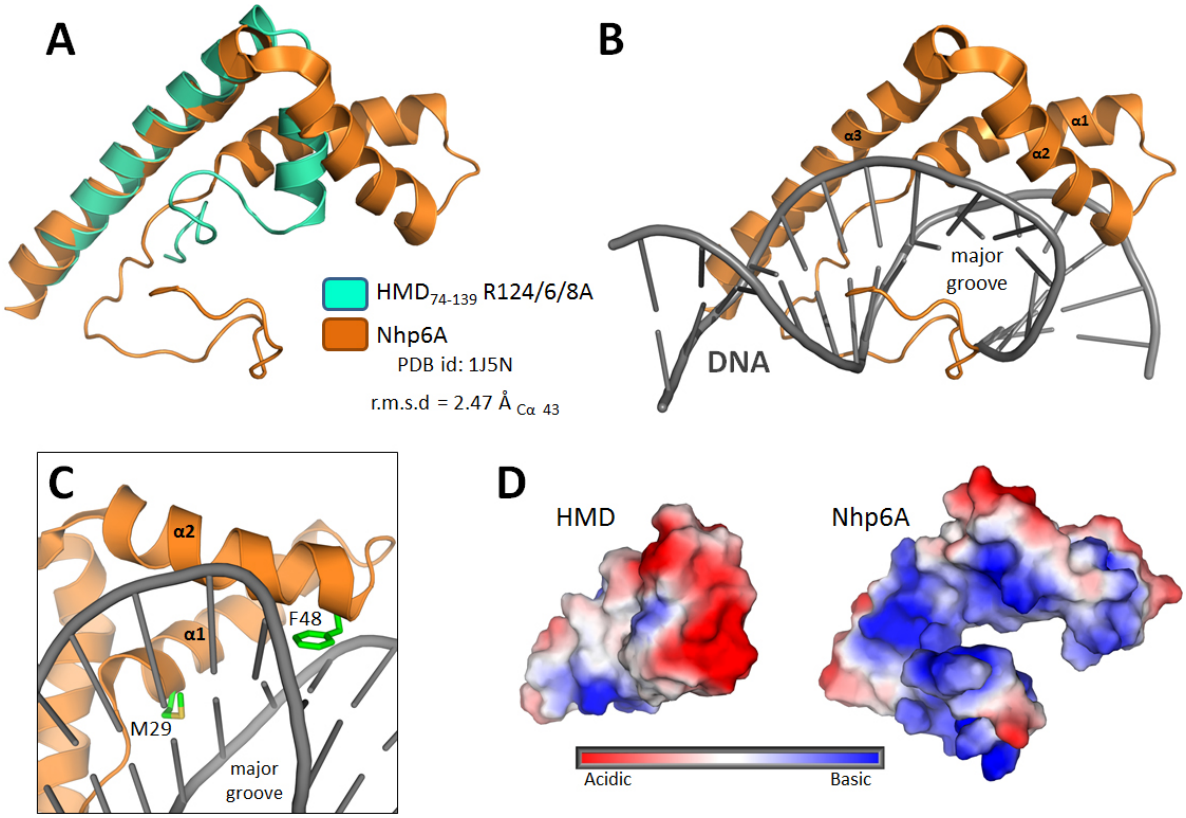


Figure 31. The HMD shares structural similarity with Nhp6A

A) Structural overlap of the Rtf1 HMD_{74-139} R124/6/8 structure (cyan) with Nhp6A (orange; PDB id: 1J5N). The two structures deviated by a r.m.s.d of 2.47 Å over 43 Cα atoms. B) Structure of Nhp6A (orange) bound to double stranded DNA (grey). The secondary strutural elements of Nhp6A are labelled along with the major groove of the bound DNA. C) The M29 and F48 hydrophobic wedges used to distort and bend the DNA helix. The critical hydrophobic wedge residues (M29, F48) are indicated (green sticks). D) Electrostatic surface representation of the HMD and Nhp6A generated in PyMol. Notice how Nhp6A has a basic internal channel to mediate interaction with DNA.

While HMD R124/6/8A aligns reasonably well with Nhp6A, most of the alignment is confined to the HMD's longer α2 helix with the α3 helix of Nhp6A. Functionally Nhp6A makes extensive hydrophobic and electrostatic contacts with DNA (Figure 31B). Of particular importance are two hydrophobic ridges referred to as the M29 wedge and the F48 wedge which are required to distort and bend DNA (Figure 31C) [331]. These hydrophobic wedges are not conserved in the HMD. Significant differences also exist in the surface electrostatics of the proteins, with the HMD lacking a basic cleft to accommodate DNA like that seen in Nhp6A (Figure 31D). Collectively, these structural differences between Nhp6A and the HMD strongly suggest the two proteins utilize their folds in a functionally different manner.

3.2.4.2 The HMD has structural similarities to the histone fold

Another intriguing possibility is that the HMD acts as a histone mimic to regulate nucleosome accessibility, promoting histone modifications including H2B K123 monoubiquitylation. When super-imposed on the structures of crystallized histones, the HMD aligns reasonably well. For instance, the HMD aligns with *S. cerevisiae* histone H2A with an r.m.s.d of 2.1 Å over 36 Cα atoms (PDB id: 1ID3; Figure 32A). The C-terminal helix in our structure (E100-R126) aligns well with the central helix of the histone fold. The smaller HMD helix (E89-S97), while similar to what is seen in a canonical histone fold (Figure 3), does not align properly with the α1 helix of H2A, differing by roughly 40^{o} (Figure 32A). Interestingly, the predicted secondary structure for an extended HMD construct to residue 152 includes another small helix C-terminal of $HMD_{74\text{-}139}$ protein construct (Figure 26), a feature reminiscent of the canonical three helix histone fold. Before determining more structural information or binding data, however, it is difficult to make too much from these alignments. Not surprisingly, based on its structural similarities to the

histone fold, the HMD also shows similarities to several of the TATA-binding protein associated factors (TAF) of the general transcription factor TFIID (Figure 32B).

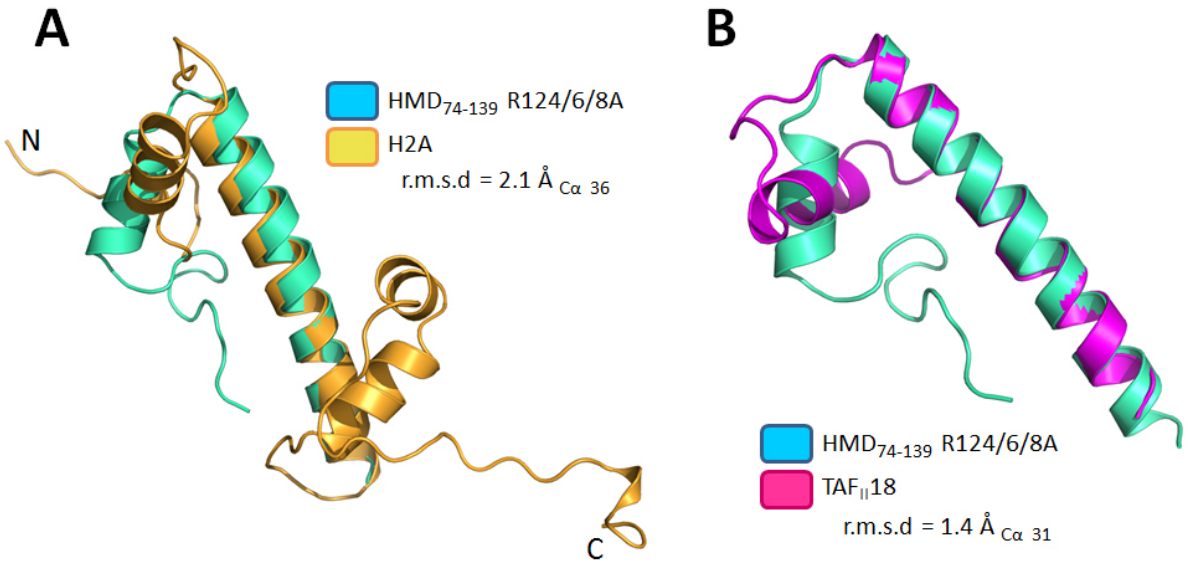


Figure 32. The HMD shares structural similarity with histones

A) Structural overlap of the Rtf1 HMD_{74-139} R124/6/8 structure (cyan) with *S. cerevisiae* H2A (orange; PDB id: 1ID3). The two structures deviated by a r.m.s.d of 2.1 Å over 36 Cα atoms. B) Structural overlap of the Rtf1 HMD_{74-139} R124/6/8 structure (cyan) with $TAF_{II}18$ (magenta; PDB id: 1.4). The two structures deviated by a r.m.s.d of 2.47 Å over 43 Cα atoms.

3.2.5 Determining the crystal structure of HMD_{74-139} R126A

By examining the neighboring HMD molecules in our HMD R124/6/8A crystal structure, it is clear that only R126A is making productive van der Waal interactions to facilitate crystallization (Figure 29). Knowing this and the R124/6/8A substitution mutations are defective in HMD function in yeast; I wanted to attempt crystallization of HMD_{74-139} with only a single charge

reduction mutation at R126 (R126A). To do this, I generated a new expression construct containing $HMD_{74\text{-}139}$ R126A using site-directed mutagenesis. Additionally, I designed and cloned a collection of extended HMD constructs for crystallization trials. These constructs included those extending the HMD boundaries to residue 152 and 184, both of which add additional elements of predicted secondary structure. Further, I designed each of the constructs to include one of the following charge reduction mutations: R126A, R126/128A, or R124/126/128A (Figure 33).

I have expressed, purified, and performed crystallization trials on several of the constructs designed (summarized in Table 4). Similar to what I have observed in my crystallization attempts with earlier HMD constructs, all of them form pseudo-crystals in a wide collection of conditions. Only the shortest HMD constructs (74-139) with substitutions containing R126A (R126A, R126/8A, or R126/6/8A), however, form crystals with well-defined and crisp edges.

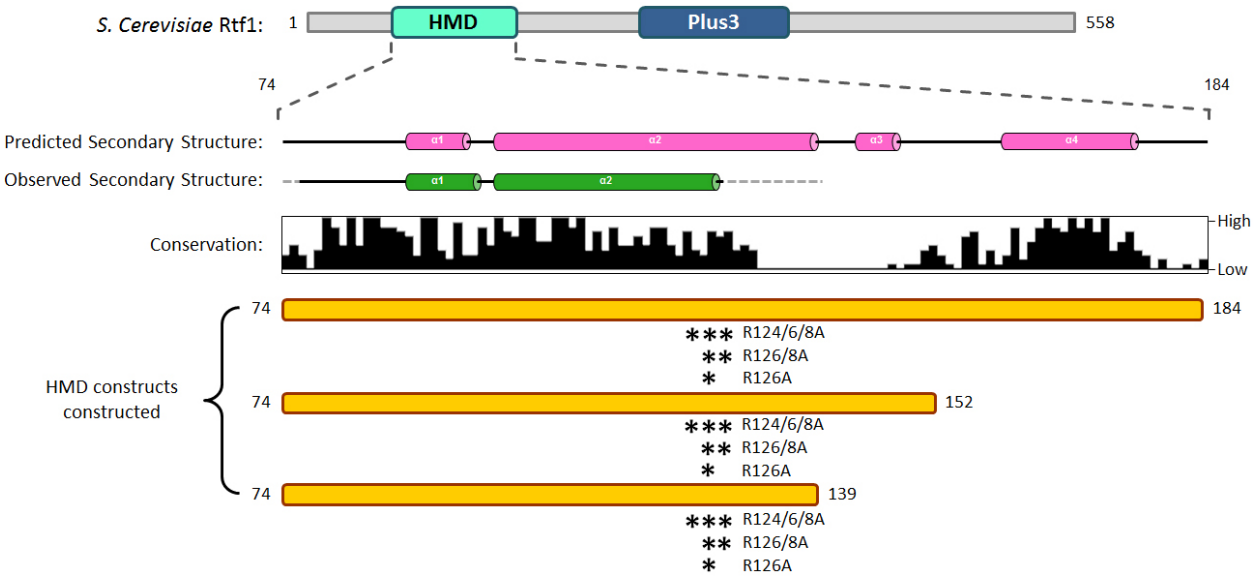


Figure 33. Extended HMD constructs generated

Domain architecture of *S. cerevisiae* Rtf1. The HMD is expanded, showing its predicted secondary structure (PSIPRED) and percent identity. The observed secondary structure for HMD_{74-139} R124/6/8A is shown below the predicted secondary structure. New HMD constructs are indicated below the secondary structure and percent identity as yellow bars. Charge reduction mutants for each construct are marked by asterisks.

Initially I focused my attention on crystallizing HMD_{74-139} R126A. As expected, HMD_{74-139} R126A crystals grow in the same conditions that generated the HMD R124/6/8A crystals (40% PEG 300, 0.1M phosphate-citrate (pH 4.2)) which we used for data collection and structural determination. Unlike what was seen with HMD R124/6/8A, the R126A mutant protein formed hundreds of small crystals (Figure 34A) as opposed to two or three large crystals. To collect a high resolution x-ray data set it is important to have large well defined crystals. This can often be accomplished by slightly altering the crystal growth conditions, typically through lower the precipitant concentration (in my case PEG300). Using the HMD R124/6/8A crystallization conditions (40% PEG 300, 0.1M phosphate-citrate (pH 4.2)) as a starting point, I went about systematically optimizing each of the different components of the crystallization experiment. Despite extensive efforts varying pH, temperature (4^oC and 21^oC), precipitant concentrations, and protein concentrations, HMD R126A continued to form hundreds of tiny protein crystals.

Table 4. *S. cerevisiae* Rtf1 HMD constructs in pQLinkH

Construct	Vector	Notes	DH5α
HMD 74-139	pQLinkH	Expressed/purified (2L growth » ~12 mg protein)	AV608
HMD 74-139 R124A	pQLinkH	Expressed/purified (2L growth » ~12 mg protein)	AV657
HMD 74-139 R126A	pQLinkH	Expressed/purified (2L growth » ~14 mg protein)	AV611
HMD 74-139 R128A	pQLinkH	Cloned	AV663
HMD 74-139 R126A, R128A	pQLinkH	Expressed/purified (2L growth » ~14 mg protein)	AV612
HMD 74-139 R124A, R126A, R128A	pQLinkH	Expressed/purified (2L growth » ~28 mg protein)	AV613
HMD 74-152	pQLinkH	Expressed/purified (2L growth » ~4 mg protein)	AV609
HMD 74-152 R124A	pQLinkH	Expressed/purified (2L growth » ~3 mg protein)	AV659
HMD 74-152 R126A	pQLinkH	Expressed/purified (2L growth » ~22 mg protein)	AV614
HMD 74-152 R128A	pQLinkH	Cloned	AV665
HMD 74-152 R126A, R128A	pQLinkH	Expressed/purified (2L growth » ~18 mg protein)	AV615
HMD 74-152 R124A, R126A, R124A	pQLinkH	Expressed/purified	AV616
HMD 74-184	pQLinkH	Expressed/purified (2L growth » ~40 mg protein)	AV610
HMD 74-184 R124A	pQLinkH	Expressed/purified (2L growth » ~15 mg protein)	AV661
HMD 74-184 R126A	pQLinkH	Expressed/purified (2L growth » ~15 mg protein)	AV617
HMD 74-184 R128A	pQLinkH	Cloned	AV667
HMD 74-184 R126A, R128A	pQLinkH	Cloned	AV618
HMD 74-184 R124A, R126A, R128A	pQLinkH	Expressed/purified (2L growth » ~14 mg protein)	AV619

In addition to varying the buffer conditions and the protein concentrations, sometimes protein crystallization can be facilitated by the addition of an additive. An additive can be any small molecule such as an organic acid, a nucleotide, or a small peptide. Typically one screens through a collection of additives, looking for those that establish stabilizing intermolecular interactions with your protein and promote lattice formation and crystallization. Several companies make large, commercially available additive screens, for optimizing protein crystals.

Hampton Research for instance makes the Silver Bullets Bio additive screen that is made up of a collection of biologically important small molecules including things like small peptides, co-factors, and biochemical pathway intermediates [332].

Having little success optimizing the crystallization conditions for HMD_{74-139} R126A, I decided to perform additive screens, focusing first on using Hampton's Silver Bullets Bio screen. Strikingly, I found four conditions in the Silver Bullets screen that grew beautiful crystals that were both larger in size and more uniform in definition (Figure 34A). Looking through the components of each of the four Silver Bullets conditions identified, I found that they all shared a common component, benzamidine. Of the four different Silver Bullets Bio conditions identified, condition 7 resulted in the largest and most uniform crystals. Using Silver Bullets condition 7 as an additive, I was able to optimize crystal growth in 28% PEG 300 and 0.1M phosphate-citrate (pH 4.2). Crystals were cryoprotected by increasing the concentration of PEG 300 to 44% before being flash frozen in liquid nitrogen.

I collected a native dataset at our home source and used molecular replacement to estimate phases using my HMD R124/6/8A structure as a model. This allowed me to generate an interpretable electron density map and build an initial model of the protein. The model was improved with manual building using Coot [312] and refined at 1.62 Å resolution with an R_{work} and R_{free} values of 19.4 and 22.8%, respectively (Table 5). The crystal of HMD R126A belongs to the $R3_2$:H space group like the HMD R124/6/8A crystal and contains two molecules in its asymmetric unit. The resulting crystal structure of HMD R126A is nearly identical to the original R124/6/8A structure, with an r.m.s.d of 0.31 Å over 52 Cα atoms (Figure 35A). The additive benzamidine interacts with HMD residues on the domain's smaller α1 helix, forming pi stacking interactions with H94 and H-bonding interactions with S90 (Figure 35B). This interaction serves

to create a new crystal contact with a neighboring HMD molecule, mediated through residues H94 and S90 of both HMD molecules (Figure 35C).

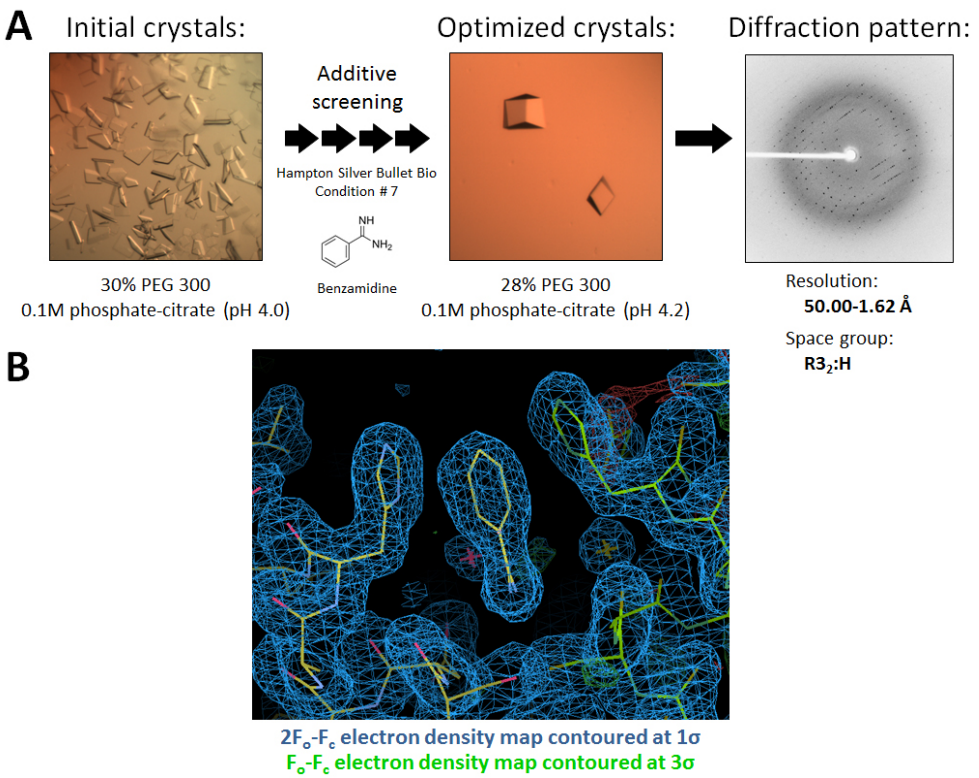


Figure 34. Crystallization and structural determination of *S. cerevisiae* Rtf1 HMD_{74-139} R126A

A) Schematic of the crystallization and data collection process for HMD_{74-139} R126A. Representative images from the crystallization process are shown, including the buffer conditions used in each case. The additive screen condition used in crystal optimization is indicated, including a diagram of the critical component of the additive, benzamidine. A frame from the x-ray diffraction data is shown along with the resolution range and experientially determined space group. B) Section of the electron density map of the *S. cerevisiae* HMD_{74-139} R126A crystal structure.

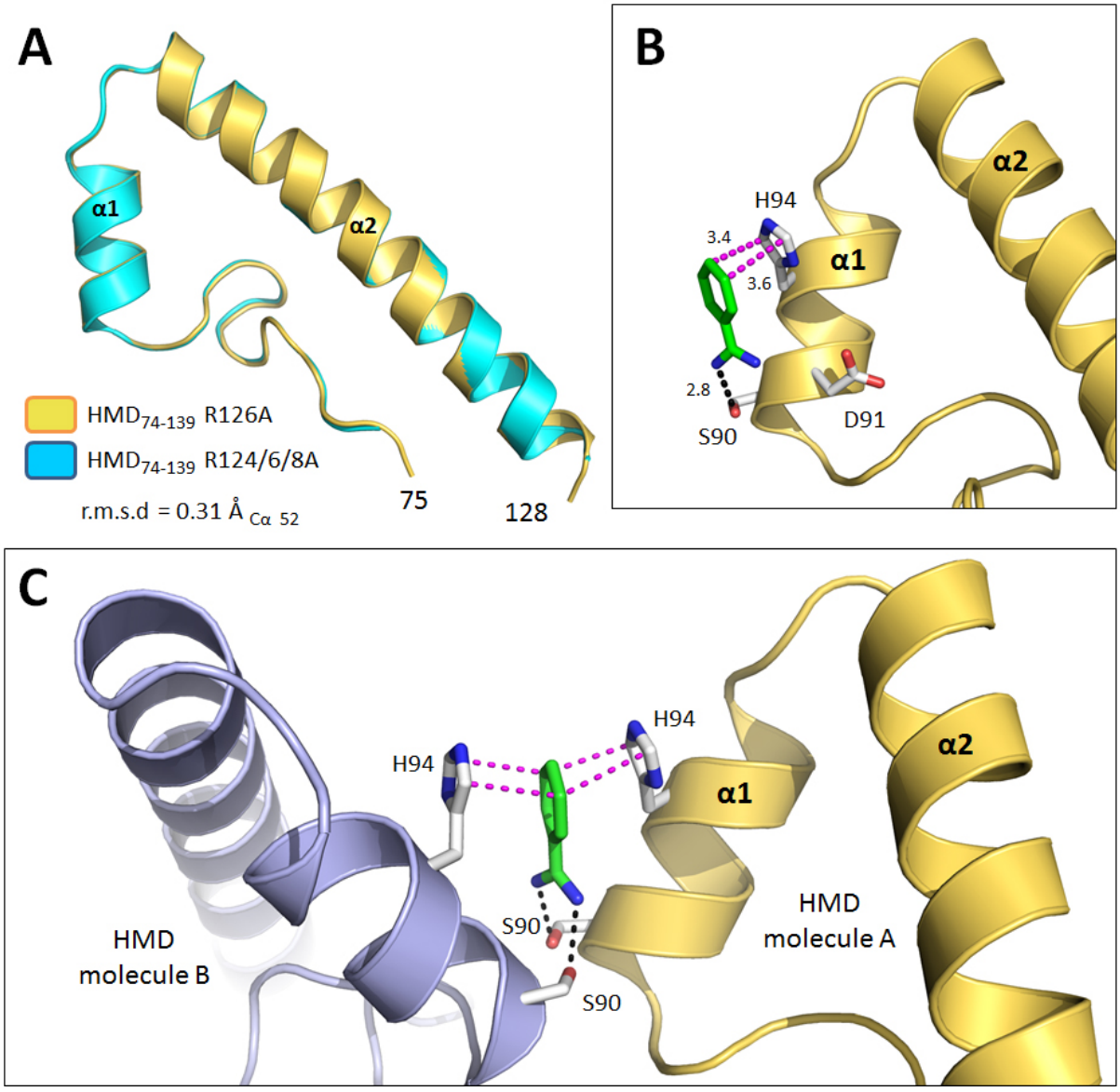


Figure 35. Crystal structure of the *S. cerevisiae* Rtf1 HMD_{74-139} R124A

A) Overlap of the Rtf1 HMD 74-139 R126A structure (orange) with the HMD_{74-139} R124/6/8A structure (blue). The two structures deviated by a r.m.s.d of 0.31 Å over 52 Cα atoms. B) Benzamidine binding surface important in a mediating crystal contact. Benzamidine is shown in green with HMD residues critical in its recognition shown as white sticks. Hydrogen bonding (black) and van der Waal interactions (magenta) involved in the recognition are indicated as dashed lines. C) Crystal contact generated with the addition of benzamidine. The neighboring symmetry mate in the crystal lattice is shown in blue (molecule B).

Table 5. Data collection and refinement statistics for *S. cerevisiae* Rtf1 HMD_{74-139} R126A

	Native
Data collection	
Space group	$R3_2$:H
Cell dimensions	
a, *b*, *c* (Å)	93.9, 93.9, 75.3
α, β, γ (°)	90.0, 90.0, 120.0
Unique Reflections	16,294
Resolution (Å)	50.0 - 1.62 (1.65-1.62)
R_{merge} (%)	5.7 (50.3)
$I / \sigma I$	43.0(6.77)
Completeness (%)	99.6 (100.0)
Redundancy	20.1 (15.2)
Refinement	
Resolution (Å)	30.00-1.62 (1.78-1.62)
R_{work} / R_{free} (%)	19.3/22.8 (20.1/25.3)
Number of. atoms	
Protein	976
Other	9
Water	75
B-factors (Å^2)	
Protein	31.2
Other	28.6
Solvent	41.3
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.05
Ramachandran	
Outliers (%)	0.00
Allowed (%)	2.53
Favored (%)	97.5

Values in parentheses are for highest-resolution shell.

$R_{merge} = (|(\Sigma I - \langle I \rangle)|)/(\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.

$R_{work} = \Sigma_{hkl}||F_{obs}(hkl)|| - F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$.

Having identified an additive that promotes HMD crystallization by creating a new crystal contact, I decided to revisit crystallization attempts with wild-type constructs. Each of the wild-type HMD constructs (74-139, 74-184) readily form pseudo-crystals in a wide range of conditions, making it possible that the additional crystal contact afforded by benzamidine would be just enough to promote crystallization. At around this time, Branden Van Oss from the Arndt lab finished his analysis of the effects the HMD charge reduction mutations on H2B K123 monoubiquitylation. He found that the defect in H2B K123 monoubiquitylation observed upon R124/6/8A substitution was localized primarily to the R126A substitution. Based on this information I decided to clone versions of each of the HMD constructs containing singular charge reduction substitutions at R124A and R128A (Table 4).

I have expressed and purified several of the wild-type and single charge reduction substitution HMD constructs and subjected them to crystallization trials. Unfortunately it does not appear that the addition of benzamidine has a significant effect on facilitating crystallization of the wild-type constructs or the R124A and R128A substitutions. Pseudo-crystals still form in a broad range of diverse crystallization conditions. I have attempted to further optimize several of the different conditions identified for both wild-type and R124A substitution constructs (Table 4), but have been unable to grow diffraction quality crystals.

3.2.6 DNA and RNA binding capabilities of the HMD

When looking for Rtf1 HMD interaction partners, Tony Piro in the Arndt lab made an interesting observation that HMD_{74-184} co-purified with a large amount of nucleic acid. The nucleic acid was susceptible to RNase, suggesting it was primarily RNA. Upon further investigation he found that

the shorter HMD construct (74-139) was not able to co-purify with RNA. This suggests that the HMD, specifically residues 140-184, contains a previously undescribed RNA binding domain. Both Rtf1 and Leo1 have been shown to be capable of binding RNA *in vitro* [255]. In the case of Rtf1, this interaction could be in part localized to the Plus3 domain which has been shown to interact with DNA [278].

To more fully characterize the HMD's nucleic acid binding properties, I performed a series of electrophoretic mobility shift assays (EMSA) using a range of different DNA and RNA probes and Rtf1 constructs. In the work by Dermondy *et al* where they showed Rtf1 was able to bind RNA *in vitro*, they used large RNA oligonucleotides (67-74 nucleotides) created through *in vitro* transcription assays. For my assays, I designed a series of smaller 14 nucleotide probes (Table 6), including a double stranded, single stranded, and hairpin RNA. Additionally I designed a double and single stranded DNA probe. The sequences of each of the probes were designed to minimize unwanted secondary structure (OligoAnalyzer, Integrated DNA Technologies). I tested three Rtf1 constructs for nucleic acid binding, full-length Rtf1, HMD_{74-139}, and HMD_{74-184}.

Surprisingly, I found that nearly all of the Rtf1 constructs had little to not affinity to the nucleic acid probes tested. When adding an excess (25-50 μM) of either the full-length Rtf1 or the larger HMD construct (74-184), I would observe a disappearance of the free probe for the double stranded and single stranded RNA probes. This behavior suggests that the protein and RNA are aggregating and fail to enter the gel, making it difficult to ascertain specific RNA binding of either construct. Even if the disappearance of free RNA probe represents specific binding for full-length Rtf1 and HMD 74-184, the apparent dissociation constant (K_d) would fall well above the ~1.0 μM K_d observed in the literature [255]. This difference in affinity may be

partially attributed to the size difference in oligonucleotide probes. Further, Dermondy *et al* cross-linked their reactions with UV irradiation prior to resolving them via SDS-PAGE [255]. I have confidence in the reaction conditions I used because full-length Rtf1 was able to bind double stranded and single stranded DNA with an estimated K_d in the micromolar range. This interaction is likely mediated through the Plus3 domain of the protein, which has been shown to bind DNA *in vitro* (Figure 13B) [278]. Further studies are required to more firmly establish the role, if any, the Rtf1 HMD has in nucleic acid binding. Currently however, I feel that if the HMD is capable of binding RNA it is not likely to be biologically relevant. It is possible that proper RNA binding by the HMD requires other regions of Rtf1 or another protein altogether.

Table 6. Probes used in HMD nucleic acid EMSAs

Name	Sequence	Storage (-20^{o}C)	Notes
ssDNA14-F (ADW31)	5’ – CTTCTTGTCCGTCGA – 3’	ADW oligos box 1	
ssDNA14-R (ADW32)	5’ – TCGACGGACAAGAAG – 3’	ADW oligos box 1	
ssRNA14-F (ADW33)	5’ - ACGCGACUCAGGAG - 3’	ADW oligos box 1	Adapted from NCBI ID: 1C4L
ssRNA14-R (ADW34)	5’ - CUCCUGAGUCGCGU - 3’	ADW oligos box 1	Adapted from NCBI ID: 1C4L
hpRNA13 (ADW35)	5’ – GCUCUCAGUGAGC – 3’	ADW oligos box 1	Adapted from NCBI ID: 1Q75

3.3 CONCLUSIONS

The N-terminal histone modification domain of Rtf1 is a critical regulator of H2B monoubiquitylation of K123 in *S. cerevisiae* [177, 205]. Catalyzed by the ubiquitin-conjugating enzyme Rad6 and the ubiquitin-protein ligase Bre1 [155-157], H2B monoubiquitylation is a prerequisite for the downstream di- and trimethylation of histone H3 K4 and K79 by the

Set1/COMPASS and Dot1 methyltransferases, respectively [158-160]. Strikingly, while only composed of 90 amino acids (62-152), the HMD is necessary and sufficient for H2B K123 ubiquitylation and H3 K4 and K79 methylation [177, 205]. Further still, the HMD is capable of localizing to chromatin *in vivo* when present as the only source of Rtf1 in *S. cerevisiae* [205]. How a domain of such size can accomplish all these feats is still a source of much debate.

To aid in our mechanistic understanding of how the HMD functions, I determined the crystal structure of a region of the HMD spanning residues 74-139. The domain forms a deceptively simple structure composed of two α-helices arranged anti-parallel to one another, forming a planar and highly solvent accessible surface (Figure 28,35). Across this surface are a series of evolutionarily conserved residues, several of which form a patch on the domain (E100, E104, L107, F108, K116, Q115, Figure 30B) that is required for Rtf1-dependent histone modifications. This patch is very likely acting as a protein-protein interaction surface with an unknown binding partner to facilitate Rtf1-dependent histone modifications.

Despite extensive efforts by the Arndt lab, the HMD's binding partner has remained elusive. Logical interaction partners like the ubiquitin-conjugating enzyme Rad6 and the ubiquitin-protein ligase Bre1 do not appear to interact with the domain through *in vitro* and *in vivo* immunoprecipitations (Arndt lab, data unpublished). Structural homology searches with our crystal structure against the entire Protein Data Bank (PDB, http://www.rcsb.org/) reveal few definitive clues about the mechanism by which the domain functions. The domain shares loose structural homology with several chromatin-associated protein, including TAF proteins from the transcription initiation factor TFIID and the DNA binding protein Nhp6A. Further the domain exhibits structural similarities to the canonical histone fold, suggesting a possible mechanism

where the domain acts as a histone mimic to loosen chromatin structure and facilitate access for chromatin modifying enzymes.

Along the lines of acting as a histone mimic, it is also possible that the HMD interacts directly with the nucleosome. A particularly intriguing idea is that the HMD binds the nucleosome on the acidic patch formed at the H2A-H2B interface on the nucleosome surface (Figure 36). This interaction would thereby compete for acidic patch binding with the H4 N-terminal tail, an interaction critical for chromatin compaction and higher order structure [333-335]. There is already precedence for such an interaction in nucleosome-binding proteins [336-340], exemplified in the co-crystal structures of the LANA peptide [337], the Sir3 BAH domain [336], and RCC1 [338] with the nucleosome. In each of these structures, the proteins specifically recognize the acidic patch of the nucleosome with a strategically located arginine residue forming hydrogen bonding interactions with H2A residues E62, D91, and E93 of the acidic patch (residue nomenclature is from *S. cerevisiae*, residue numbers and identities differ slightly between organisms). In the case of the HMD, two surface exposed invariant arginine residues (R92, R103) exist that could plausibly interact with the acidic patch (Figure 36). By mediating an interaction with the acidic patch of the nucleosome, the HMD would prevent the nucleosome-nucleosome interactions important for intrinsic chromatin condensation. This in turn could lead to an alternate more open chromatin architecture accessible to chromatin modifying enzymes like Rad6-Bre1 and downstream methyltransferases. Future studies are needed to more full address this possibility. Preliminary work by Branden Van Oss of the Arndt lab shows that R92 and R103 alanine substitutions have modest effects on downstream H3 K4 and K79 methylation marks but little effect on K123 monoubiquitylation. *In vitro* HMD binding studies using

recombinant nucleosome could be particularly valuable. These studies could be coupled with crystallography efforts.

Further efforts to determine the structure of larger HMD constructs could be tremendously informative. Any additional structural elements will add complexity to the HMD fold and make structural homology searches easier and more meaningful. While I have performed extensive crystallization trials with several of the shorter HMD constructs generated, focused efforts on some of the larger constructs (Table 4), particularly those with the R128A substitutions could be beneficial. Crystallization of these larger constructs could be greatly aided with the addition of additives such as benzamidine, like that seen in my HMD R126A structure (Figure 35). Along these lines, generation of new HMD constructs could also facilitate crystallization. *S. cerevisiae* residues 134-146 are considerably more variable than the rest of the extended HMD sequence (74-184; Figure 26). It is possible that their removal could promote crystallization of a larger Rtf1 fragment containing the HMD.

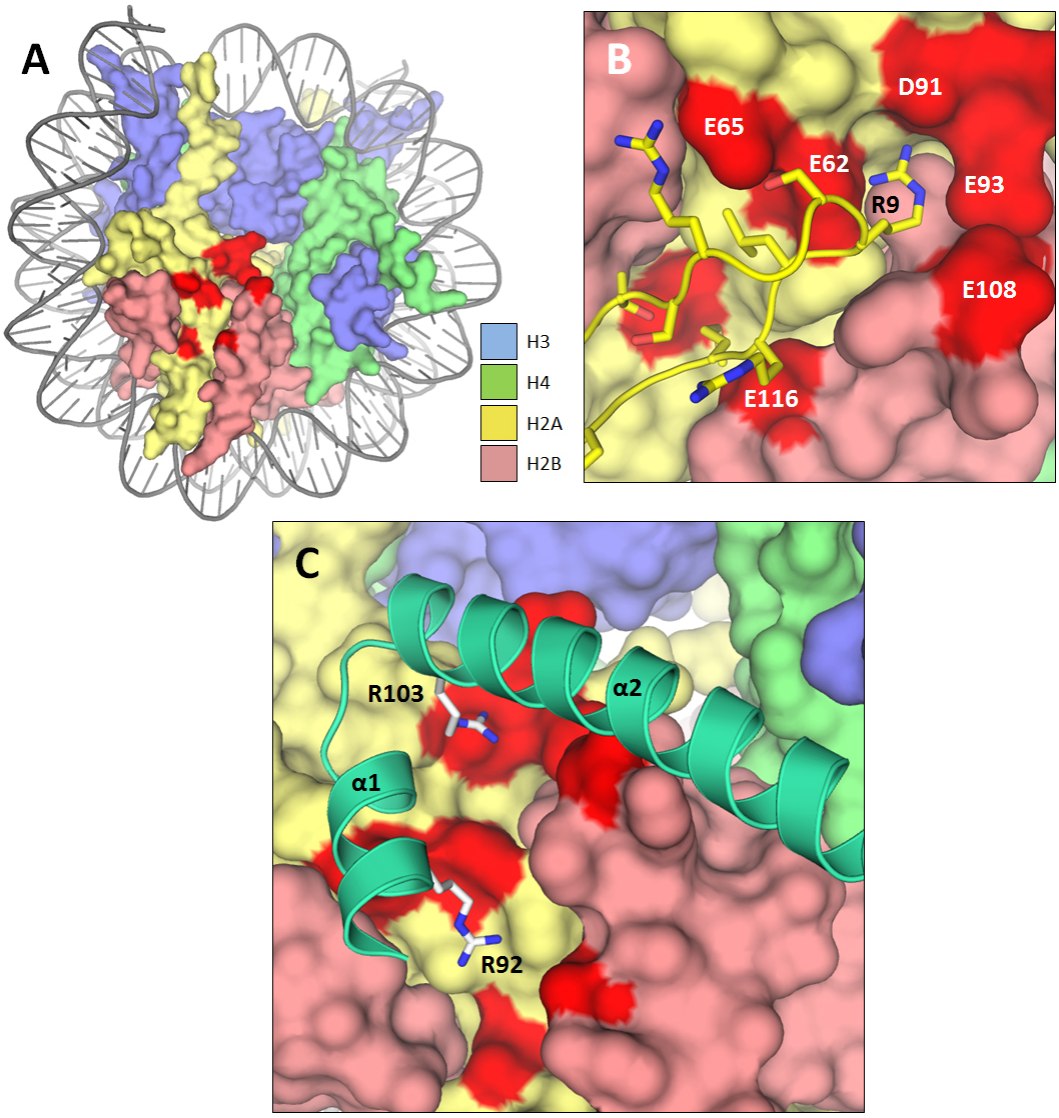


Figure 36. The H2A-H2B acidic patch of the nucleosome

A) H2A-H2B acidic patch on nucleosome surface. X-ray crystal structure of the nucleosome (PDB id: 1ZLA). The histones are shown as surface representations (H3, blue; H4, green; H2A, yellow; H2B, pink). H2A and H2B residues making up the acidic patch are highlighted in red. B) LANA (latency-associated nuclear antigen) peptide recognition surface on the *X. laevis* nucleosome (PDB id: 1ZLA). Critical LANA peptide residues and acidic H2A patch residues are indicated. Labels correspond to the residues in *S. cerevisiae*. C) Proposed HMD association with the H2A-H2B acidic patch. Conserved HMD arginine residues proposed to mediate binding are indicated (white sticks).

4.0 STRUCTURAL CHARACTERIZATION AND IDENTIFICATION OF THE CORE PAF1 COMPLEX

4.1 INTRODUCTION

While the description and identification of individual regions or domains within Paf1C is incredibly valuable, a comprehensive understanding of Paf1C function requires an understanding of the spatial arrangement of Paf1C subunits. Subunit arrangement will provide a window into how Paf1C organizes functional domains and how it coordinates associations with RNA Pol II and other elongation factors. The best strategy to elucidate Paf1C's spatial orientation is through structural analysis with techniques like x-ray crystallography, small angle x-ray scattering (SAXS), and electron microscopy.

To gain insight into how Paf1C coordinates its functions and interactions, I have worked towards determining the crystal structure of Paf1C and any sub-complexes via x-ray crystallography. Success in obtaining protein crystals and quality x-ray diffraction data is heavily dependent upon choosing the correct constructs to work with. Orderly formation of a crystal lattice is greatly facilitated by removal of a protein's flexible sequences, while retaining its rigid core sequences, those locked together through intrasubunit interactions. Flexible sequences of a protein can include N- and C-terminal tails, internal loops within the primary sequence, and even stable domains that are only tethered to the protein core. Removal of these flexible sequences

becomes even more critical when dealing with a protein complex, because any one could prevent crystallization of the entire complex. It is important to note that sequences required for biological functions can be found within both flexible and core sequences of a protein, but from a crystallographer's standpoint, any sequences outside of the core would be best studied in isolation.

While diffraction quality crystals of Paf1C have eluded me, I have been successful in establishing two recombinant expression systems for Paf1C sub-complexes. Notably, I have found that co-expression of the soluble N-terminus of Cdc73 solubilizes Ctr9, Paf1, and Leo1 in *S. cerevisiae*. This sub-complex is well expressed in *E. coli* and is amenable to a battery of purification techniques. Most importantly, the sub-complex co-elutes from a size exclusion chromatography column as a single peak, strongly suggesting the subunits are forming a complex.

4.2 RESULTS

4.2.1 Co-expression system for *S. cerevisiae* Ctr9, Paf1, Leo1, and Cdc73

Recombinant expression for multi-protein complexes typically utilizes one of three possible strategies. The first involves a multi-plasmid approach in which each subunit is expressed from an individual plasmid. The second involves a single expression plasmid coding for a polycistronic message containing each protein of the complex. Finally you can reconstitute the protein complex with separately expressed and purified subunits, a strategy successfully used

with the exosome [341]. Each approach has its own advantages and disadvantages. The use of multiple vectors provides increased flexibility to modify the protein construct and individual vectors are often easier to construct. On the downside, individual vectors often show widely different expression levels, which may be undesirable. The polycistronic approach has the advantage of providing matched levels of protein expression under the control of a single promoter. Additionally, each subunit is expressed at a centralized location, promoting complex formation. In turn, it suffers from difficult vector construction. The reconstitution approach shares the advantages of multiple vectors, but is labor intensive and requires that each subunit is soluble for purification. For my first Paf1C expression system, I chose an approach that balanced these concepts, utilizing both polycistronic expression and multiple expression vectors.

This Paf1C expression system was created with two compatible expression vectors (pMCSG7, pMCSG23) that were readily available. The first plasmid (CPL), cloned by Dr. Margaret Shirra, was a pMCSG7 vector coding for polycistronic expression of full-length Ctr9, full-length Leo1, and a truncated version of Paf1 (Paf1Δ5, residues 1-385). The second (HM73), cloned by Chris Amrich, utilized a pMCSG23 backbone and encoded an N-terminal fragment of Cdc73 (residues 1-230). The CPL vector encoded for a TEV cleavable His_6-tagged Ctr9, while Paf1Δ5 and Leo1 remained untagged (Figure 37). The HM73 vector was engineered to produce a Cdc73 fusion protein with N-terminal His_6 and maltose binding protein (MBP) tags, both of which could be removed by TEV protease (Figure 37).

When separately transformed into CodonPlus *E.Coli* (Stratagene), both vectors were capable of expressing the cloned proteins. Expression of Ctr9, Leo1, and Paf1Δ5 however was very poor and was strictly insoluble. Remarkably, co-expression of HM73 with CPL (CPL73) successfully pulled Ctr9, Leo1, and Paf1Δ5 into the soluble fraction along with Cdc73 (Figure

38, lane 3). Solubilizing Ctr9, Leo1, and Paf1Δ5 with the addition of Cdc73 was a major advance because it allowed for bacterial expression of a large, nearly intact Paf1C sub-complex. This success focused my attention on utilizing bacterial expression vectors to obtain Paf1C for structural studies, in lieu of purifying physiological Paf1C from yeast.

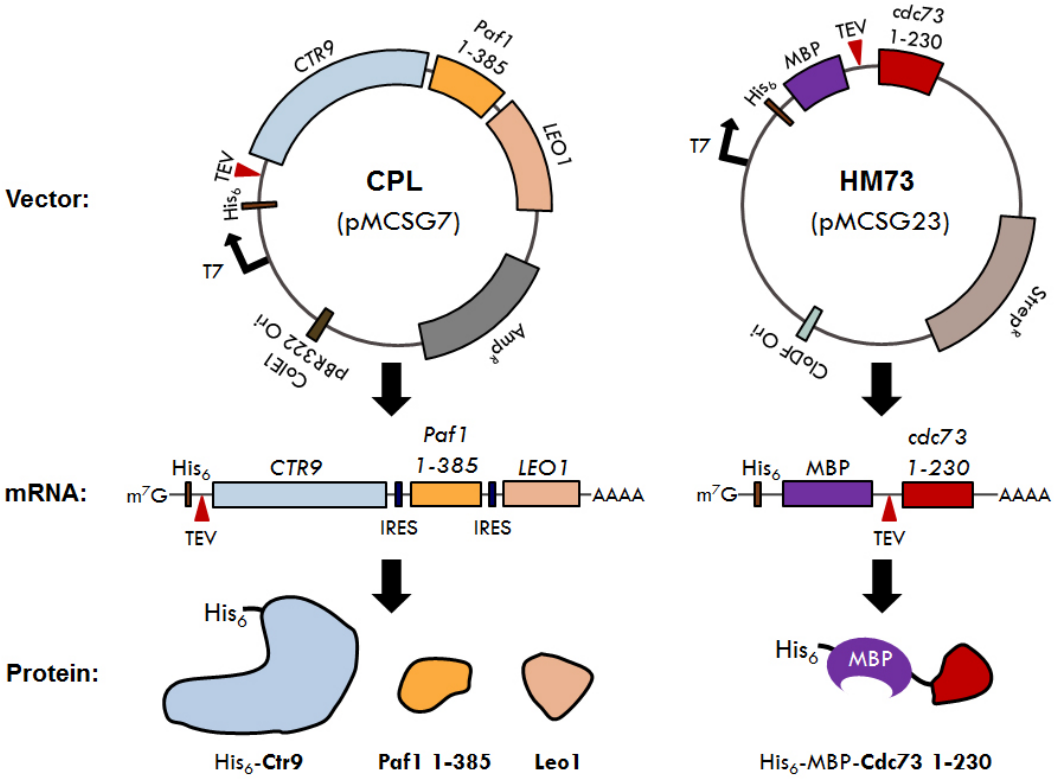


Figure 37. Schematic of the CPL73 expression system

Ctr9, Paf1 1-385, and Leo1 were cloned into a pMCSG7 vector under the control of a single T7 promoter (CPL). The resulting polycistronic mRNA message contains internal ribosome entry sites (IRES), allowing the message to be translated into an N-terminally His_6-tagged Ctr9 and untagged versions of Paf11-385 (Δ5) and Leo1. Cdc73 1-230 was cloned into a pMCSG23 vector, downstream of a His_6-MBP tag. Both vectors have different selection factors and compatible origins of replication, allowing for their co-expression in *E.coli*.

Taking advantage of the N-terminal His_6 tags imparted on Ctr9 and Cdc73 1-230 by their vectors, I used affinity chromatography (Ni-NTA) to begin complex purification. Consistent with

complex formation, untagged members of the complex (Leo1, Paf1Δ5) purified along with the tagged members (Ctr9, Cdc73) (Figure 38, lane 4). Following affinity chromatography, eluted complex was incubated at room temperature with TEV protease to remove tags on Ctr9 and Cdc73 (Figure 38, lane 5). Digested complex was sent over a 2nd Ni-NTA column to recapture nonspecific contaminants, while Paf1C was recovered in the flow through (FT) (Figure 38, lane 6). Shifts in migration on SDS-PAGE upon TEV digestion support the identities of Ctr9 and Cdc73 1-230 (Figure 38, lane 4+6). As a final test for complex integrity and homogeneity, the complex was analyzed with size exclusion chromatography. Complex members co-eluted off an S500 sizing column in a single peak, further supporting complex formation (Figure 38, lanes 7,8).

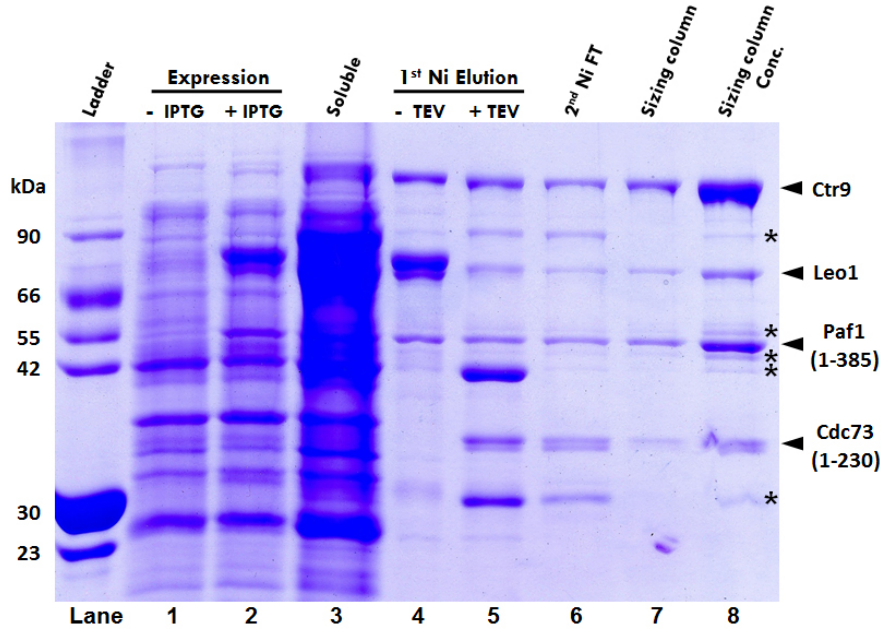


Figure 38. Purification of a four member Paf1C subcomplex using the CPL73 expression system

The CPL and HM73 expression vectors were co-transformed and expressed with IPTG induction (lane 2) and purified by two rounds of Ni affinity chromatography (lanes 4,6) followed by size exclusion chromatography (lane 7). Contaminating proteins are marked with an asterisk (*).

The purified complex was able to be concentrated to approximately 6.4 mg/ml without precipitation, a range suitable for structural studies. I conducted a series of crystallization trials with the purified Paf1 subcomplex using our lab's crystallography robot. The robot allowed me to screen for Paf1 subcomplex crystallization conditions in a 96-well format, using commercially available sparse-matrix crystallization screens. These screens failed to produce any initial crystal hits suitable for further optimization. This failure of the Paf1 subcomplex to form crystals was likely caused in part by a non-homogeneous protein sample. When viewed via SDS-PAGE, the concentrated protein samples used for crystallization contained several contaminates (Figure 38, lane 8). Many of these impurities are likely breakdown products of Paf1C subunits, supporting the notion that we have not identified the stable core components of Paf1C. Further it suggested that we needed to develop a new expression system able to be quickly modified to add or subtract regions of Paf1C subunits.

4.2.2 Paf1C co-expression using the pQLink expression system

Both the CPL and HM73 expression vectors were not originally designed for use in a multi-vector co-expression system. This was exemplified by the fact that both vectors contained an N-terminal His_6 tag, resulting in the expression of His_6 tagged Cdc73 and Ctr9. Additionally, neither vector was engineered to be easily manipulated after the initial cloning. This final point was critically important in my efforts to crystallize the core Paf1C. Successful crystallization of the complex would require screening through multiple constructs for each subunit to identify those comprising a stable Paf1C core. For this to be feasible, we needed an expression system

that was rapidly and easily amenable to iterative addition/subtraction of constructs, while not sacrificing overall protein expression.

Many of those criteria were met by the pQLink expression system (Figure 39) [342]. The pQLink system is a unique expression system that allows for the generation of co-expression plasmids through ligation-independent cloning (LIC) of three standard pQLink plasmids (Figure 39B). Each of the vectors contain two LINK sequences (LINK1, LINK2) [342], flanking the vector's expression cassette (promoter, multiple cloning site, transcriptional terminator). These LINK sequences allow for the insertion of one expression cassette into a second pQLink vector through LIC, effectively allowing you to link together an unlimited number of genes (Figure 39C). The vectors share a common vector backbone (pQTEV2), differing only in their affinity tag, or lack thereof (pQLinkN), imparted on the expressed protein. The pQLinkH vector for instance, imparts a His_7 tag, whereas the pQLinkG vector imparts a GST tag. All of the pQLink vectors are capable of expressing single proteins or can be converted to co-expression plasmids through LIC reaction of other pQLink expression cassettes. After cloning your constructs into the desired pQLink vectors, you can quickly and easily combine them in any combination desirable. This system was perfectly suited for structural studies of Paf1C, because it allowed for simple generation of new co-expression plasmids as the different boundaries of the Paf1C core were determined.

In collaboration with Dr. Margaret Shirra in the Arndt lab, we cloned a series of Paf1C subunits and subcomplexes into the pQLink vectors (summarized in Table 7). We cloned full-length constructs of each Paf1C subunit into pQLink vectors, save Paf1 and Rtf1, which were cloned in truncated forms. In addition to the full-length constructs, truncated forms of Leo1, Ctr9, and Cdc73 were cloned into pQLink vectors (Table 7). Ctr9 and Cdc73 were cloned into

pQLinkH and pQLinkG vectors, respectively, while the other subunits were cloned into the pQLinkN vector. This allowed for selection of two subunits (Ctr9, Cdc73) during complex purification, using two different affinity tags (His7, GST), expediting complex purification and accurate complex stoichiometry.

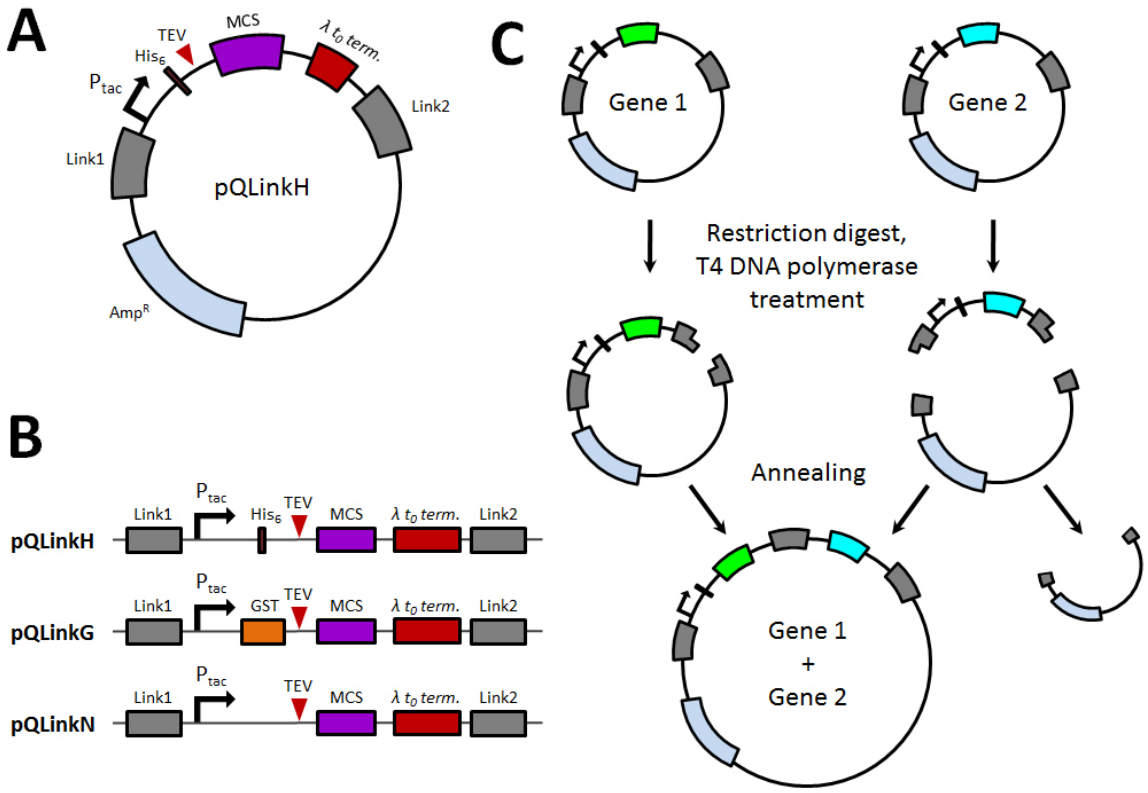


Figure 39. pQLink expression system and construction of co-expression plasmids

A) Vector map of pQLinkH. Multiple cloning site (MCS), TEV protease cleavage site (TEV), Transcription terminator (term.) B) pQLink expression cassettes in each of the parent pQLink vectors. The vectors differ only in the tag imparted on the expressed protein. C) General overview of the construction of a co-expression plasmid from two pQLink plasmids containing different cDNA inserts (Gene1, Gene2).

I have tested the expression of each of the individual Paf1C subunits in a range of *E.coli* expression strains (RIL, RIPL, BL21). Subunit expression remained relatively consistent between expressions, but varied significantly between constructs. Ctr9 and Cdc73 for example, expressed well in all constructs tested, while Paf1Δ5 and full-length Leo1 were not detectable by coomassie staining. The truncated version of Leo1 (109-379) however, expressed very well, presenting the intriguing possibility that constructs could be initially screened based entirely on *E.coli* expression levels.

In an effort to improve Paf1Δ5 and Leo1 expression, and test the co-expression ability of the pQLink vectors, we generated several co-expression vectors through LIC. Among the vectors created was a vector (qCPL73) mimicking my original CPL73 expression system, containing Paf1Δ5, full-length Ctr9, full-length Leo1, and Cdc73 (1-230). Based on my co-expression studies with several of the combined expression cassette vectors, I have found that subunits that express well on their own (Ctr9, Cdc73) continue to express strongly. For those subunits that express poorly, the results are more variable. For instance, dual expression of Paf1Δ5 with any of the other Paf1C subunits appears to have no effect on Paf1Δ5 expression levels, resulting in undetectable expression of Paf1Δ5 by coomassie staining. Interestingly when Paf1Δ5 is expressed in conjunction with full-length Cdc73 and full-length Ctr9 (qCP73), we see induction of a band that is sized appropriately to be Paf1Δ5 (Figure 40A). Counterintuitively, when Paf1Δ5 is expressed with full-length Cdc73, full-length Ctr9, and full-length Leo1 (qCPL73), there is no clear induction band for Paf1Δ5 while Ctr9, Leo1, and Cdc73 express strongly (Figure 40B).

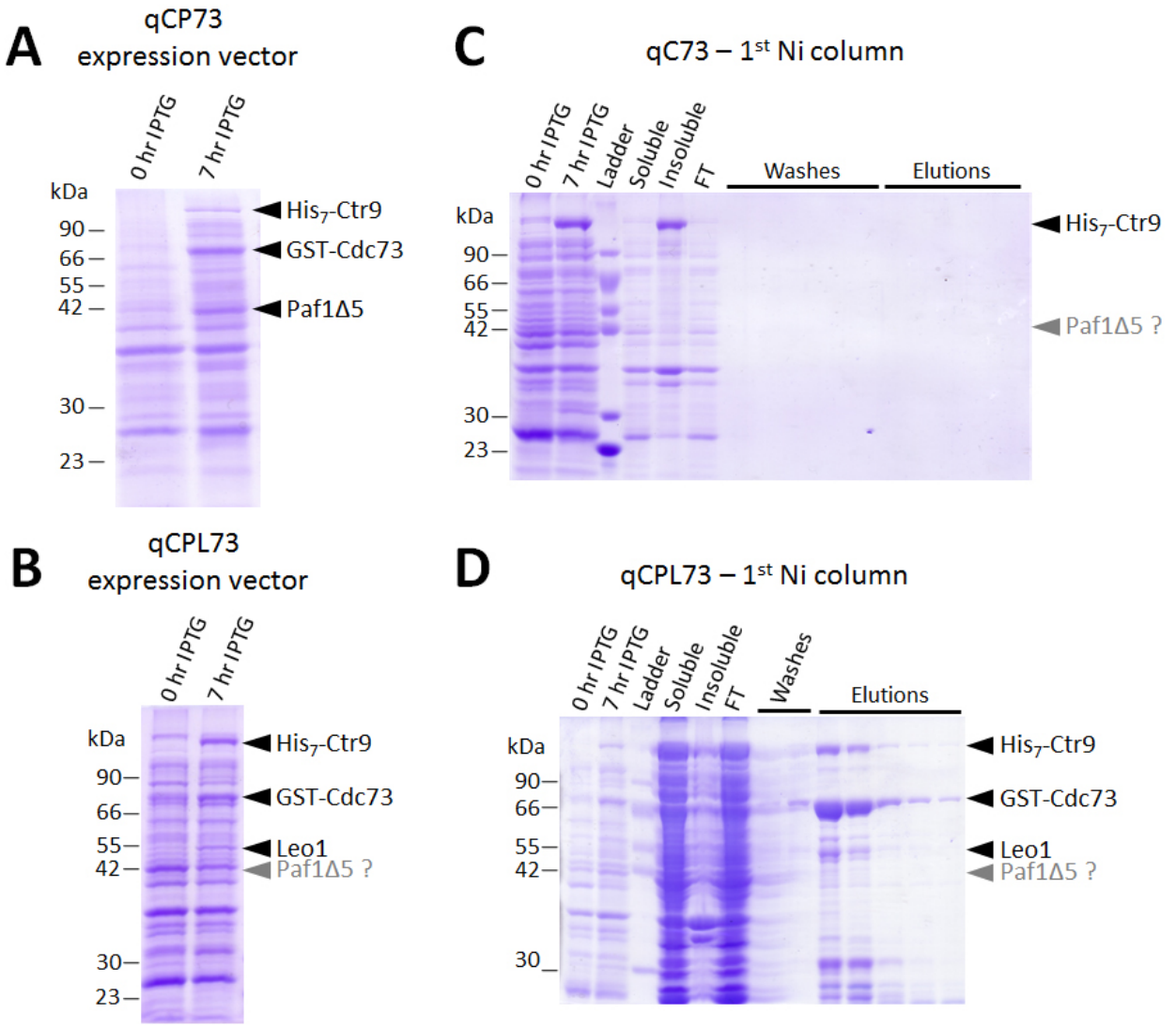


Figure 40. Expression and purification of pQLink Paf1 subcomplexes

A) qCP73 IPTG expression. Expected subunits sizes: His_7-Ctr9 ~125 kDa, GST-Cdc73 ~67 kDa, Paf1Δ5 ~45 kDa B) qCPL73 IPTG expression. Paf1Δ5 is not expressed (grey text). Expected subunits sizes: His_7-Ctr9 ~125 kDa, GST-Cdc73 ~67 kDa, Paf1Δ5 ~45 kDa, Leo1 ~53.8 C) 1st Ni affinity chromatography column for qC73 . His_7-Ctr9 remains in the insoluble fraction. Paf1Δ5 is not expressed (grey text). D) 1st Ni affinity chromatography column for qCPL73. His_7-Ctr9, GST-Cdc73, and Leo1 co-elute off of the Ni-column. Paf1Δ5 is not expressed (grey text).

I have performed preliminary protein purifications with several of the IPTG co-inductions, including the co-induction of Paf1Δ5 with full-length Ctr9 and the co-induction of

qCPL73. In the case of Paf1Δ5 and Ctr9, Paf1Δ5 is not induced and Ctr9 remains in the insoluble fraction (Figure 40C). The retention of Ctr9 in the insoluble fraction agrees well with our earlier purification trials using Ctr9 produced from other expression systems lacking Cdc73. Work with the pCPL73 co-expression vector has proven to be more successful. His_7-Ctr9, GST-Cdc73, and Leo1 are all confined to the soluble fraction of the bacterial lysate. Importantly, both GST-Cdc73 and Leo1 co-elute off of a Ni-affinity chromatography column, suggesting the three proteins are forming a complex together (Figure 40D). Paf1Δ5 expression is still problematic. There are no apparent bands for Paf1Δ5 in the induced gel sample or any of the subsequent samples throughout the affinity purification. The Ctr9-Cdc73-Leo1 subcomplex that co-elutes from the 1st Ni-affinity column remains intact through subsequent purification steps, including a 2nd Ni-affinity column after TEV protease digestion and a gel filtration column (data not shown). Interestingly, a large portion of the Ctr9-Cdc73-Leo1 subcomplex remains bound to the Ni-affinity resin after TEV cleavage, suggesting the tagged constructs are mildly resistant to TEV protease digestion. Future work is needed to more completely characterize the pQLink vectors currently cloned, specifically focusing on their expression levels and the amenability of the expressed proteins to purification.

Table 7. Paf1C pQLink vectors

Construct	Vector	Notes	VanDemark storage	Arndt lab storage
Rtf1 230-558	pQLinkN	No HMD		KB965
Paf1 2-385	pQLinkN	Δ5		KB966
Paf1' 2-385	pQLinkN	Codon optimized		
Leo1 2-464	pQLinkN	Full-length		KB967
Leo1 109-379	pQLinkN			KB968
Ctr9 2-1077	pQLinkH	Full-length		KB969
Ctr9 40-870	pQLinkH			KB989
Ctr9 870-1077	pQLinkH		AV371	
Cdc73 1-393	pQLinkG	Full-length	AV356	
Cdc73 1-205	pQLinkG		AV357	
Paf1 2-385 + Rtf1 230-558	pQLinkN			KB974
Paf1 2-385 + Leo1 2-464	pQLinkN			KB975
Paf1 2-385 + Leo1 109-379	pQLinkN			KB976
Paf1 2-385 + Ctr9 2-1077	pQLinkN	His7-Ctr9		KB977
Paf1 2-385 + Ctr9 40-870	pQLinkN	His7-Ctr9		KB982
Paf1 2-385 + Cdc73 1-393	pQLinkN	GST-Cdc73		KB996
Cdc73 1-393 + Leo1 2-464	pQLinkG	GST-Cdc73		KB997
Cdc73 1-393 + Ctr9 2-1077	pQLinkG	His7-Ctr9, GST-Cdc73		KB998
Rtf1 230-558 + Leo1 109-379	pQLinkN			KB1002
Paf1' 2-385 + Cdc73 1-393	pQLinkN	Paf1 codon optimized		KB1104
Paf1 2-385 + Leo1 2-464 + Ctr9 40-870	pQLinkN	His7-Ctr9		KB988
Paf1 2-385 + Leo1 2-464 + Cdc73 1-393	pQLinkN	GST-Cdc73		KB994
Paf1 2-385 + Leo1 2-464 + Cdc73 1-205	pQLinkN	GST-Cdc73		KB995
Paf1' 2-385 + Cdc73 1-393 + Ctr9 2-1077	pQLinkN	Paf1 codon optimized, His7-Ctr9, GST-Cdc73		KB1106
Paf1 2-385 + Leo1 2-464 + Cdc73 1-393 + Ctr9 2-1077	pQLinkN	His7-Ctr9, GST-Cdc73		KB999

4.2.3 Limited proteolysis of a Paf1C subcomplex

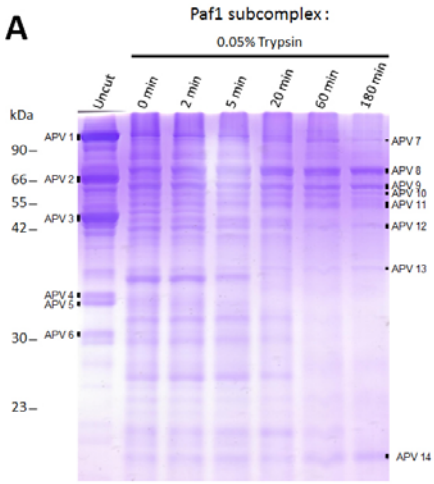
Limited proteolysis is a technique commonly used to identify stable domains of a protein or a protein complex. In the technique one exposes a purified protein to a protease at an extremely low concentration. When protease levels are limited, regions of a protein that are most accessible will be digested first, including surface loops, linkers, and the N- and C-termini. Buried regions of the protein however, like those found in a protein's core, will be shielded from digestion.

Typically the digestion is done over an extended period of time (1-3 hours) and samples are taken throughout to visualize via SDS-PAGE. Protein bands that are still visible at the end of the digestion, via SDS-PAGE, are considered proteolytically stable fragments of the protein. Such a technique is ideal for identifying and refining the boundaries of a protein or complex of proteins for structural studies.

To begin probing Paf1C stability and identifying the core complex, I performed a number of limited proteolysis experiments on a recominantly expressed and purified Paf1 subcomplex composed of Paf1Δ5, full-length Ctr9, full-length Leo1, and an N-terminal fragment of Cdc73 (expressed from the pCPL73 expression system, Figure 37,38). The complex was digested with either trypsin or Subtilisin A and analyzed via SDS-PAGE and coomassie staining (Figure 41A,C). Several bands appeared during the digestions that were stable throughout the trypsin and Subtilisin A time courses, potentially representing stable core fragments. These bands were excised from the gels and their identities determined by complete in-gel tryptic digestion combined with mass spectrometry in collaboration with Dr. Rich Gardner at the University of Washington (Department of Pharmacology). Several of the bands from the undigested Paf1 subcomplex sample were also excised and sent for mass spectrometry analysis. This was done to ensure that our starting sample contained all of the expected subunits and confirm our assignments of the different subunits observed via SDS-PAGE.

Analysis of the different peptide fragments identified by the mass spectrometry analysis confirmed that our starting Paf1 subcomplex contained all 4 of the expected subunits, Paf1Δ5, full-length Ctr9, full-length Leo1, and an N-terminal fragment of Cdc73 (Figure 41). The identities of the proteolytically stable fragments from the trypsin and Subtilisin A time courses were much more surprising. Instead of containing truncated versions of all of the members of th

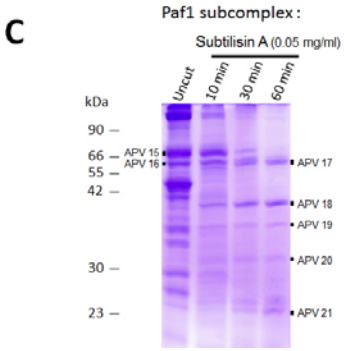
Paf1 subcomplex, we found that all of the proteolytically stable fragments excised from the gels were primarily composed of Ctr9 (Figure 41B,D). These data suggest that all of the subunits from the purified Paf1 subcomplex aside from Ctr9 are protease sensitive. One explanation for this observation is that the purified Paf1 subcomplex is not folded properly and hence is susceptible to proteases even at limiting amounts. It is also possible that lower protease concentrations are need to be used to generate meaningful data for the other subunits of the subcomplex. Future work is required to differentiate between these possibilities.



B

Tandem-mass spec analysis of peptide fragments

	Peptide fragment hits			
Gel sample	Ctr9	Cdc73	Paf1	Leo1
APV1	245	7	31	15
APV2	64	6	23	105
APV3	40	4	85	54
APV4	32	21	11	20
APV5	19	38	7	13
APV6	9	9	4	5
APV7	41	1	2	6
APV8	62	3	1	0
APV9	61	3	3	0
APV10	45	3	1	0
APV11	47	2	2	1
APV12	28	2	2	0
APV13	16	0	0	0
APV14	15	8	1	2



D

Tandem-mass spec analysis of peptide fragments

	Peptide fragment hits			
Gel sample	Ctr9	Cdc73	Paf1	Leo1
APV15	98	21	39	265
APV16	75	16	57	152
APV17	83	2	6	3
APV18	44	3	5	2
APV19	56	2	6	3
APV20	83	4	9	3
APV21	56	7	20	5

Figure 41. Limited proteolysis and mass spectrometry analysis of a Paf1 subcomplex

A) Purified Paf1 subcomplex was subjected to limited proteolysis with 0.05% trypsin over a period of 3 hours, with samples taken at different time points and stopped with SDS loading dye before being run on an 15% SDS-polyacrylamide gel. Bands to be excised for in-gel tryptic digestion followed by mass spectrometry analysis are labelled (APV1-14). B) Bands were excised from the gel (APV1-14) in panel A and subjected to in-gel tryptic digestion before being analyzed by tandem-mass spectrometry analysis to identify peptide fragments. The peptide hits for each of the Paf1 subcomplex members are listed for each gel band analyzed. C) Purified Paf1C subcomplex was subjected to limited proteolysis with Subtilisin A (0.05 mg/ml) over a period of 1 hour. Samples were taken at different time points and stopped with SDS loading dye before being run on an 15% SDS-polyacrylamide gel. Bands to be excised for in-gel tryptic digestion followed by mass spectrometry analysis are labelled (APV15-21). D) Bands were excised from the gel (APV1-14) in panel C and subjected to in-gel tryptic digestion before being analyzed by tandem-mass spectrometry analysis to identify peptide fragments. The peptide hits for each of the Paf1 subcomplex members are listed for each gel band analyzed.

4.3 CONCLUSIONS

Despite some of the initial setbacks, the pQLink expression system still holds great promise for exploring Paf1C architecture and determining the structure of the core complex. For the system to be successful, a more comprehensive and detailed expression study of the available constructs is needed. If Paf1 and Leo1 expression still prove to be problematic, their coding sequences can be replaced by ones optimized for expression in *E.coli*, both of which our lab currently has available. In the case of Paf1, Dr. Shirra has already cloned the optimized version (Paf1') into the pQLink system individually, in combination with full-length Cdc73, and in combination with Cdc73 and full-length Ctr9 (Table 7). Paf1 could also be introduced on a separate expression vector compatible with the pQLink vectors. The pRSF vector would be a good candidate for such studies. I have already tested co-expression of pQLinkG-expressed Cdc73 and pRSF-expressed Rtf1 (HMD) and found the two vectors are compatible (data not shown).

After successfully establishing a robust expression system and purification scheme, focus can be shifted to examining one Paf1C subunit at a time and systematically work through all of the cloned constructs to determine which behaves best in the context of the other subunits. This should allow us to narrow down Paf1C crystallization candidates, and loosely define the core of Paf1C. These results can be further refined through limited proteolysis studies coupled with mass spectrometry. Combining limited proteolysis of Paf1C with Native PAGE could also be particularly informative. The core Paf1 subcomplex remaining after limited proteolysis should travel as a single band when run on Native-PAGE, allowing for mass-spectrometry analysis and identification of the stabilized fragments physically associated with one another.

In addition to recombinant expression of Paf1C in bacteria, there is great promise in using other recombinant or endogenous sources to answer structural questions about Paf1C. While more costly and time consuming than using bacteria, recombinant expression of Paf1C in insect cells, using a baculovirus expression system, promises to be particularly powerful. Since expression occurs in a eukaryotic source, many of the post-translational modifications normally seen in endogenously expressed proteins are maintained. Incorporation of the proper post-translation modification may be particularly important in formation of Paf1C, as evidence by my limited proteolysis work on the Paf1, Ctr9, Leo1, and Cdc73 subcomplex. The baculovirus expression system in insect cells has already been successfully used to reconstitute human Paf1C [181]. Endogenous expression and purification of the Paf1C could also be incredibly valuable as a source of complex for structural techniques such as electron microscopy and small-angle x-ray scattering (SAXS). Paf1C has already been successfully purified endogenously from several sources including humans [181], *S.cerevisiae* [114], and S. pombi [343].

5.0 DISCUSSION

My research in Dr. VanDemark's lab has focused on revealing the mechanisms underlying DNA templated processes, including transcription and DNA repair. My transcriptional focus has been centered on the Paf1 complex, specifically the Rtf1 subunit of the complex and its role in regulating RNA polymerase II-mediated transcription. Using biochemical and structural techniques, I provided a molecular description of how Rtf1 mediates Paf1 complex recruitment to elongating RNA polymerase II. Recruitment of Rtf1 is controlled by its centrally located Plus3 domain and a direct interaction with the evolutionarily conserved elongation factor Spt5. I solved the co-crystal structure of the human Plus3 domain of Rtf1 bound to a phosphorylated C-terminal repeat of Spt5. The structure revealed the basis for recognition of the repeat motif of Spt5, an important component in the recruitment of regulatory factors to RNA Pol II [311]. I have performed further structural characterization of Rtf1, studying the N-terminal histone modification domain. To gain insight into the molecular mechanism underlying the domain's function, I successfully crystallized and solved the structure of a minimal region of the domain necessary and sufficient for Rtf1-mediated histone modifications. Further, I helped identify a conserved surface on the domain's surface that is required for proper HMD function.

In collaboration with the Bernstein lab, I have studied the *S. cerevisiae* Shu complex and its role in Rad51-dependent homologous recombination. Using fluorescence anisotropy and

electrophoretic mobility shift assays, I helped determine that the Csm2-Psy3 heterodimer of the Shu complex preferentially binds forked and 3' DNA substrates, both substrates used in the homologous recombination pathway. Further we showed that Csm2 interacts with Rad51 through the Rad55-Rad57 heterodimer, promoting Rad51 filament formation and stability. This suggests Csm2-Psy3 acts to recruit the Shu complex to sites of double-stand breaks where it promotes error-free homologous recombination [344].

My work here has just scratched the surface of all that can be learned about the roles of Rtf1 in regulating transcription. Work in the Arndt lab using *S. cerevisiae* has shown that the C-terminal region of Rtf1 is required for its association with the rest of Paf1C [177]. Refining the boundaries of this region and establishing which members of Paf1C are involved in the interaction is an important standing question. In the case of the human version of Rtf1, its connection to Paf1C appears to be much more transitory, partially dissociating from Paf1C upon gel filtration [170, 181, 238, 239]. Studies using recombinant human Rtf1 suggests the subunit associates with Paf1C through the Paf1 and Cdc73 subunits, specifically requiring residues 250-585 [181]. It remains to be determined if these interaction points are direct or indirect. The N-terminus of Rtf1 (residues 3-30) is also known to be involved in the recruitment of the monomeric chromatin remodeler Chd1 [113, 177]. The phenylalanine at residue 11 is known to play some role in this because its substitution to alanine results in similar defect in Chd1 recruitment [177]. Defining the specific mechanism for Rtf1-dependent Chd1recruitment will be important in more fully understanding Rtf1's role in regulating chromatin structure.

Moving away from the Rtf1 subunit, the remaining members of the Paf1 complex present even more questions. Several subunits of the complex, namely Leo1 and Ctr9, are still relatively uncharacterized both structurally and functionally. Sequence analysis of Ctr9 strongly suggests

that it is comprised almost entirely of TPR motifs (Figure 7) [242, 244]. These motifs have been crystallized in many different proteins [345-348], making Ctr9 a favorable target for structural studies. The biggest hurdle that needs to be overcome for Ctr9 structural studies is generating a soluble sample to work with. I have recently cloned the human form of Ctr9 into the bacterial expression vector pQLinkH. Preliminary expression tests and purifications using the vector are promising. Human Ctr9 is soluble; however it does not bind efficiently to Ni-NTA resin. This may suggest the protein is forming a soluble aggregate. Alternatively, the His_7-tag may be obscured and benefit from being moved to the C-terminal end of the protein. If Ctr9 is forming a soluble aggregate, we may be able to get around this by expressing the protein with a different tag that has molecular chaperone characteristics, like maltose-binding protein [349, 350]. Additionally, recombinant expression of Ctr9 in insect cells using a baculovirus expression system could be a potential solution, because the Roeder lab has already successfully done so with the human form (of which we also have the expression plasmid) [181].

Leo1 is perhaps the biggest mystery of all of the members of the Paf1 complex, as its loss in *S. cerevisiae* does not result in strong transcription-related phenotypes [167, 169]. It is known to play a role in recruiting Paf1C to chromatin through an interaction with RNA [255]. Looking deeper into this interaction promises to greatly expand our understanding of Leo1's role in Paf1C. Like with Ctr9, one of the biggest challenges will be generating soluble Leo1 for biochemical studies. In our hands, the yeast form of Leo1 is insoluble when recombinantly expressed by itself in *E. coli*. The most direct course of action would be to shift to using the human form of the protein. Dermody *et al.* successfully co-expressed and purified full-length human Leo1 and Paf1 from *E. coli* [255]. With a soluble and well behaved version of Leo1, RNA binding can be assessed with EMSAs and fluorescent anisotropy.

Another intriguing avenue of study is investigating the molecular mechanisms underlying Cdc73's role in cell cycle progression. While Cdc73 is often thought of as a tumor suppressor in humans [257], work by Iwata *et al.* has shown that Cdc73 can also promote cell growth and entry into S phase when overexpressed in cell lines expressing SV40 large T antigen (TAg) [351]. TAg is a protein encoded for by the small DNA virus SV40 (Simian Vacuolating Virus 40) upon infection (Reviewed in [352, 353]). The protein alters gene expression and growth of host cells by binding a range of transcription factors and components of the replication machinery, including p53 and retinoblastoma proteins like pRb, p107, and p130 [352]. These binding interactions are important for transformation and tumorigenesis [354]. Iwata *et al.* found that the enhanced cell proliferation caused by Cdc73 overexpression requires a direct interaction of Cdc73 with TAg [351]. The interaction was mapped to the central region of Cdc73 (residues 218-263) and residues 361-481 of TAg's helicase domain. Interestingly, the TAg-binding region on Cdc73 matches closely to its binding site for β-catenin [262], suggesting that Cdc73 has a common binding site for transcriptional activators. Further, the Cdc73-binding site on TAg colocalizes to one of the known binding sites of p53 [355].

The direct interaction between human Cdc73 and TAg represents a promising candidate for structural studies and has the promise to greatly inform the mechanism used by Cdc73 to regulate the cell cycle. Several crystal structures exist of the helicase region of TAg [355, 356], including a co-crystal structure of the helicase region bound to the DNA binding region of p53 [357]. These structures were all solved using protein recombinantly expressed in *E.coli* and purified following a simple scheme using affinity chromatography (GST) and gel filtration. All of these qualities make the domain very tractable for structural work in our lab. In the case of the TAg-binding domain of Cdc73 (residues 218-263) [351], little structural data exists. Secondary

structural predictions based on the protein's primary sequence suggest the region is well ordered, forming a mixed α/β topology (DISOPRED [320]). Human Cdc73 residues 214-264 would serve as a reasonable construct for preliminary structural studies with the helicase domain of TAg.

Aside from structural studies focusing on Cdc73-TAg binding, it will be important to more fully establish the molecular mechanisms underlying Cdc73's role in recruiting Paf1C to chromatin. In yeast, the C-terminal region of Cdc73 from residues 201-393 are capable of binding diphosphorylated RNA Pol II CTD repeats and phosphorylated Spt5 CTR repeats [265]. Using fluorescence anisotropy, Chris Amrich from our lab confirmed yeast Cdc73 is capable of binding to a CTD peptide containing two repeats phosphorylated on Serines 2 and 5. It remains to be determined if the human form of Cdc73 has a similar affinity towards the phosphorylated RNA Pol II CTD and/or the Spt5 CTR. These can be easily examined with fluorescence anisotropy using the fluorescently labeled CTD and CTR peptides currently available in the lab. Chris Amrich has performed similar experiments using a C-terminal fragment of human Cdc73 (residues 353-531) corresponding to the yeast C-domain and found no appreciable binding to either the CTD or CTR peptides. This agrees well with what is seen in yeast because the residues critical in binding are located right upstream of the C-domain (residues 201-229) [265]. To fully assess human Cdc73's binding affinity towards the CTD and CTR peptides, larger Cdc73 constructs need to be tested. The two best candidates are full-length Cdc73 (residues 1-531) and an extended C-terminal fragment from residues 241-531. I have recently cloned both versions of Cdc73 into the pQLinkH and pQLinkG2 *E.coli* expression vectors. Preliminary expression tests suggest that both versions of Cdc73 express well in the pQLinkG2 vector. After a simple purification, both versions of the protein can be tested for peptide binding with fluorescence anisotropy. Any interactions identified can then be accessed for specificity with NaCl completion

anisotropy experiments. Interactions that are stable at NaCl concentration between 100 and 150 mM are likely biologically relevant.

It will also be important to begin determining mechanistically how Paf1C mediates many of its broadly recognized functions. While we understand a great deal about how the complex is recruited to chromatin, we know much less about how the loss of Paf1C results in the reduction of levels of serine 2 phosphorylation of the CTD of RNA Polymerase II. Further, our understanding of how Paf1C mediates proper RNA 3'-end formation is limited. These functions are controlled by a complex network of interactions involving multiple members of Paf1C, necessitating an understanding of the complex as a whole rather than simply the individual subunits. This is supported by the recent work of Chu *et al.* showing the crystal structure of a human Paf1-Leo1 subcomplex and how heterodimerization of Paf1 and Leo1 is required for their association with the N-terminal tail of histone H3 [240]. Structural information detailing the spatial arrangement of Paf1C or other subcomplexes of Paf1C will be essential in understanding the physical basis for Paf1C's different functions. Towards this end, I have helped develop several recombinant expression systems that are worth pursuing in more depth in the future.

The initial efforts in studying Paf1C structurally should focus on using the CPL73 expression system (Figure 37), due to the robust and reproducible expression of four of the Paf1C subunits. A closer examination and optimization of each of the stages during purification of the complex will be important. Further, the inclusion of additional purification steps could prove to be beneficial in generating better complex with proper stoichiometry. The most obvious inclusion would be adding an ion exchange column between Ni-affinity purification and gel filtration steps. I have successfully used several of these columns (Heparin, Q) to purify the CPL73 Paf1 subcomplex but never in combination with gel filtration.

Generating a monodisperse and homogeneous purification of the Paf1 subcomplex will be essential for any structural studies. Optimizing the purification conditions and the addition of ion exchange chromatography should hopefully accomplish all of this for the CPL73 generated complex. However due to the dynamic nature of protein complexes, it is still possible that the resulting complex will exhibit structural heterogeneity (compositional and/or conformational). Some of this heterogeneity may be able to be removed through redesigning the expression constructs and removing flexible or disordered regions. For Paf1C this will be difficult without some other piece of information to direct what segments can be removed. Alternatively, a technique could be used to stabilize the existing Paf1 subcomplex. Mild chemical crosslinking of the purified complex with formaldehyde or glutaraldehyde is one viable option. The technique GraFix (Gradient and Fixation) could be particularly valuable for our purposes. The technique involves sedimentation of your protein complex of interest in a combined density (glycerol, sucrose, etc) and crosslinker gradient [358, 359]. The technique was designed to generate stabilized monodisperse complexes for electron microscopy and has be used successfully for many complexes, including the yeast SWR1 chromatin-remodeling complex [360]. The density gradient helps separate and maintain monodisperse complexes while the chemical crosslinker gradient acts to stabilize the protein complex. After performing GraFix on the Paf1 subcomplex, all of the resulting fractions could be analyzed with structural techniques such as electron microscopy, SAXS, and hydrogen/deuterium exchange experiments. Formaldehyde would be a good candidate for a crosslinker because the crosslinks can be reversed and the complex analyzed by SDS-PAGE to assure complex stoichiometry is correct.

While work is being done with the CPL73 expression system, it would be valuable to try and identify Paf1C complexes in other organisms that are predicted to be more amenable to

structural studies, like crystallography. Based solely on primary amino acid sequence, there are numerous properties that can be used as predictors of crystallizability, including secondary structural predictions, disorder predictions, a protein's isoelectric point, and hydrophobicity [361]. These properties and others can be examined for each member of Paf1C over a diverse range of different organisms. The results from this analysis can be compared and an organism chosen that has exhibits the most favorable characteristics. This analysis can be done manually but it is very time consuming. Alternatively the analysis can be done in a more comprehensive and high throughput manner using the XtalHunter program in development in our lab.

One potential candidate organism to use for structural studies of Paf1C is the thermophilic fungus *Chaetomium thermophilum* [362]. Proteins from thermophilic organisms are typically understood to be more stability folded than those of mesophilic relatives [363]. This makes them more attractive targets for structural and biochemical studies. The genome of *C. thermophilum* was recently sequenced and successfully utilized to study the nuclear pore complex through electron microscopy [362]. I have identified homologs of each of the members of Paf1C in *C. thermophilum* and performed preliminary RNA extractions from *C. thermophilum*. If we wish to pursue work with *C. thermophilum*, it may be more time and cost effective to have the genes synthesized.

6.0 MATERIALS AND METHODS

6.1 PROTEIN EXPRESSION AND PURIFICATION

6.1.1 Human Rtf1 Plus3 domain

The coding sequence for the human Rtf1 Plus3 domain (amino acids 353-484) was amplified from a full-length construct (kindly provided by R. Roeder) and subcloned into the pLC3 bacterial expression vector (J. Sacchettini). The resulting vector expresses the Plus3 domain tagged with N-terminal His_6- and MBP- tags, both of which can be removed via digestion with TEV protease. Plus3 protein was expressed using *E.coli* Codon+ (RIPL) cells (Agilent Technologies) grown in ZY auto-induction media [322] at room temperature for 16-24 hours. Cells were harvested by centrifugation, lysed with homogenization in 20 mM Tris (pH 8.0), 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 1 mM β-mercaptoethanol, and the lysates cleared by centrifugation at 30,000xg. Plus3 protein was purified by nickel affinity chromatography (Qiagen) followed by an overnight digestion with TEV protease. After digestion, any uncleaved protein, His_6-MBP tag, and His_6-tagged TEV protease were removed by a second round of nickel affinity chromatography. The protein was then dialyzed at 4°C overnight against a buffer containing 20 mM Tris pH 8.0, 20 mM NaCl, 8% glycerol, and 1 mM β-mercaptoethanol, prior to cation exchange chromatography and gel filtration using a Sephacryl S-200 column (GE

Healthcare). Peak fractions were concentrated to 8.5 mg/ml in 10 mM HEPES (pH 7.6), 80 mM NaCl, and 1 mM β-mercaptoethanol using a Vivaspin concentrator (Millipore) prior to crystallization.

Plasmids expressing mutant Plus3 proteins were generated by site-directed mutagenesis (QuickChange-Stratagene) and the resulting protein expressed and purified in a manner similar to wild-type. After purification, Plus3 variants were concentrated to 6.0 mg/ml in 20 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol, and 1 mM β-mercaptoethanol using a Vivaspin concentrator for use in differential scanning fluorometry and fluorescence anisotropy assays. All peptides were synthesized at the University of Pittsburgh Peptide Synthesis Core.

6.1.2 *S. cerevisiae* Rtf1 HMD

S. cerevisiae Rtf1 residues 74-184 and 74-139 (HMD) were amplified from genomic DNA by PCR and cloned into the pET151/D-TOPO (Invitrogen) vector. The HMD 74-139 construct was then used as a template for site-directed mutagenesis to simultaneously replace residues R124, R126, and R128 with alanines (R124/6/8A) using the QuikChange method (Stratagene). The R124/6/8A construct was transformed into *E.coli* Codon+ (RIPL) cells (Stratagene) and R124/6/8A protein was expressed by growth in ZY auto-induction media [322] at room temperature for 16-24 hours. Cells were harvested by centrifugation, lysed with homogenization in 20 mM Tris (pH 8.0), 500 mM NaCl, 10% glycerol, 5 mM imidazole, 5 mM β-mercaptoethanol, and the lysates cleared by centrifugation at 10,000xg. R124/6/8A protein was purified by nickel affinity chromatography (Qiagen) followed by an overnight digestion with TEV protease dialyzed against a buffer containing 20 mM Tris (pH 8.0), 500 mM NaCl, 8%

glycerol, 5 mM β-mercaptoethanol. After digestion, any uncleaved protein and the His-tagged TEV protease were removed by a second round of nickel affinity chromatography. Cleaved R124/6/8A protein was further purified through sequential use of HiTrap SP FF and HiTrap Q FF (GE Healthcare) ion exchange chromatography. As a final purification step, protein was run over a size exclusion Sephacryl S-200 column (GE Healthcare), with peak fractions eluting as an apparent monomer. Protein was concentrated to 15 mg/ml in 10 mM Tris, 150 mM NaCl, 1 mM β-mercaptoethanol using a Vivaspin concentrator (Millipore). Selenomethione-substituated R124/6/8A was expressed using PASM media [322] and purified following the same procedure for the native protein.

S. cerevisiae Rtf1 residues 74-184, 74-152, and 74-139 were also cloned into the pQLinkH vector [342]. These constructs were then used as a template for site-directed mutagenesis to create a series of charge reduction substitutions (detailed in Table 4) targeting residues R124, R126, and R128. Expression and purification of each of these proteins followed the same procedure as the R124/6/8A protein.

6.1.3 Paf1 subcomplexes

Paf1 subcomplexes from both the CPL73 and pQLink expression systems were transformed into *E.coli* Codon+ (RIL or RIPL) cells (Stratagene) and protein was expressed by IPTG induction at room temperature for 6-8 hours. Cells were harvested by centrifugation, lysed with homogenization in 25 mM Tris (pH 8.0), 175 mM NaCl, 10% glycerol, 5 mM imidazole, 1 mM β-mercaptoethanol, and the lysates cleared by centrifugation at 10,000xg. Paf1 subcomplexes were purified by nickel affinity chromatography (Qiagen) followed by an overnight digestion

with TEV protease dialyzed against a buffer containing 25 mM Tris (pH 8.0), 175 mM NaCl, 10% glycerol, 1 mM β-mercaptoethanol. Digested complex was sent over a 2nd nickel affinity column to recapture nonspecific contaminants, uncleaved protein, and the His-tagged TEV protease. As a final purification step, subcomplexes were run over a size exclusion Sephacryl S-500 column (GE Healthcare), with peak fractions eluting as an apparent monomer. Protein was concentrated to 6 mg/ml in 20 mM Tris, 170 mM NaCl, 10% glycerol, and 1 mM β-mercaptoethanol using a Vivaspin 100,000 MWCO concentrator (Millipore).

6.1.4 *S. cerevisiae* Csm2-Psy3 complex

Full-length *S. cerevisiae* Csm2 and Psy3 were amplified from genomic DNA by PCR and cloned into the bacterial co-expression plasmid pCDF Duet-1 (EMD Millipore). Protein expression was performed in *E. coli* Codon+(pRIL) cells (Stratagene) via IPTG induction at room temperature for 6-12 hours. Cells were harvested by centrifugation, lysed in 20 mM Tris (pH 8.0), 300 mM NaCl, 10% glycerol, 5 mM imidazole and 1 mM β-mercaptoethanol and the lysates cleared by centrifugation at 30 000g. Csm2 and Psy3 were co-purified by nickel affinity chromatography (Qiagen) via the His_6-tag on Csm2, followed by an overnight digestion with TEV protease. The Csm2–Psy3 complex was then further purified using HiTrap Heparin HP (GE Healthcare) affinity chromatography and size-exclusion chromatography using a Sephacryl S-200 column (GE Healthcare) with peak fractions eluting as an apparent heterodimer verified by SDS-PAGE. The peak fractions were dialysed into a buffer containing 20 mM Tris (pH 8.0), 300 mM NaCl, 8% glycerol and 1 mM dithiothreitol and concentrated to 1.6 mg/ml using a Vivaspin concentrator (Millipore).

6.2 CRYSTALLIZATION AND STRUCTURE DETERMINATION

6.2.1 Human Rft1 Plus3 domain

Plus3 crystals were grown at room temperature over 1-2 weeks using the sitting drop vapor diffusion method against a reservoir solution containing 0.1M sodium citrate pH 5.5 and 18% PEG 3000. Crystals were cryoprotected by transition of the crystal into reservoir solution supplemented to 26% PEG3000 and 20% glycerol followed by flash freezing with liquid nitrogen. To generate crystals of the Plus3-pCTR complex, Plus3 at 7.0 mg/ml was mixed with a 5-fold molar excess of pCTR peptide, and crystallized using sitting drop vapor diffusion as before using a reservoir solution containing 0.1 M sodium acetate pH 4.2 and 2.0 M ammonium sulfate. Crystals were grown at room temperature over a two-week period and were optimized using microseeding. Crystals were cryoprotected by transitioning the crystal into mother liquor supplemented to 3.5M ammonium sulfate and flash freezing under liquid nitrogen.

Diffraction data for the crystal were collected at our home source, using an FR-E rotating anode generator with VariMax optics and RAXIS IV ++ image plate detector. Diffraction data were integrated, scaled, and merged using HKL2000 [364]. Crystals of human Plus3 belong to the space group $P2_12_12_1$ (a=33.6 Å, b=68.0 Å, c=117.0 Å; α=β=γ= 90.0°) and contain two molecules in the asymmetric unit. Initial phases were estimated via molecular replacement using Phaser [365] and a search model derived from structural analysis of the Plus3 domain by NMR [278]. The model was then refined against data to 2.12 Å resolution and the model improved by manual rebuilding within COOT [312] combined with simulated annealing, positional, B-factor, and TLS refinement [366] within PHENIX [367].

Diffraction data for Plus3-pCTR crystals were collected at the National Synchrotron Light Source on beamline X25 using a Pilatus 6M detector. Plus3-pCTR crystals belong to the space group $C2_1$ (a=112.9 Å, b=172.5 Å, c=58.5 Å; β= 107.0°) and contain six Rtf1 molecules in the asymmetric unit. Phases were determined using molecular replacement with PHASER but using our Plus3 structure as a starting model. Refinement restraints for the phosphothreonine residues were generated using eLBOW [368] as implemented in PHENIX. The model was refined against diffraction data at 2.4 Å resolution and improved through the use of simulated annealing and refinement using positional, B-factor, TLS, and NCS parameters. Model quality for both structures was assessed using MolProbity [369]. All structural figures were generated using PyMol (PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC.).

Models and structure factors for both the human Rtf1 Plus3 domain and the Plus3-pCTR complex presented in this manuscript have been deposited in Protein Data Bank (www.rcsb.org) under the codes 4L1P and 4L1U.

6.2.2 *S. cerevisiae* Rtf1 HMD 74-139

Native HMD R124/6/8A crystals were grown at room temperature over 2-3 weeks using sitting drop vapor diffusion against a reservoir solution containing 0.1M phosphate-citrate (pH 4.2) and 40% PEG 300. The drops contained one part protein solution concentrated to 15 mg/ml and one part reservoir solution. Single crystals were taken directly from the mother liquor and flash frozen with liquid nitrogen. Selenomethione-substitiuted R124/6/8A crystals were grown over a one week after period after seeding the mother liquor with native R124/6/8A crystals using a cat whisker. The reservoir solution contained 0.1M phosphate-citrate (pH 4.2) and 38% PEG 300.

Crystals were cryoprotected by increasing PEG 300 concentration to 44% in 2% intervals followed by flash freezing with liquid nitrogen.

Single wavelength-anomalous dispersion (SAD) data were collected at the National Synchrotron Light Source on beamline X25A using a Pilatus 6M detector. The quality of both native and selenomethionine-substituted crystals was high, with low mosaicity and high resolution, 1.74 Å, 1.32 Å respectively. Both crystals belong to the $R3_2$:H space group (a=b=94.31 Å, c=76.98 Å, α=β= 90.0°, γ= 120.00°) and contain two molecules in the asymmetric unit. HKL2000 was used to integrate, scale, and merge the data [364].

Four of the six possible selenomethiones in the two molecules of the asymmetric unit were located using SOLVE [370] within AutoSol and an interpretable electron density map was generated [367]. An initial model was built into this experimental density using AutoBuild and the model was used for molecular replacement [323, 324] into the native dataset. Subsequent model building was performed in COOT [312], and the model was improved using simulated annealing, positional, and anisotropic B-factor refinement within PHENIX [367]. Model quality for the structure was assessed using MolProbity [369].

Native HMD R126A crystals were grown at room temperature over the period of one week using sitting drop vapor diffusion against a reservoir solution containing 0.1M phosphate-citrate (pH 4.2) and 28% PEG 300. Each drop contained 1 µl protein solution concentrated to 18 mg/ml, 1 µl reservoir solution, and 0.3µl of Silver Bullets Bio condition 7 (Hampton). Crystals were cryoprotected by increasing PEG 300 concentration to 44% in 2% intervals followed by flash freezing with liquid nitrogen.

Diffraction data for the HMD R126A crystals were collected at our home source, using an FR-E rotating anode generator with VariMax optics and RAXIS IV ++ image plate detector.

Diffraction data were integrated, scaled, and merged using HKL2000 [364]. HMD R126A crystals belong to the $R3_2$:H space group (a=b=93.9 Å, c=75.3 Å, α=β= 90.0°, γ= 120.00°) and contain two molecules in the asymmetric unit. Initial phases were determined via molecular replacement using Phaser [365] and our HMD R124/6/8A structure as a starting model. Benzamidine was identified and fit into the electron density map with Ligand Identification within PHENIX [371, 372]. The model was refined against data to 1.62 Å resolution and improved with manual building in COOT [312], along with simulated annealing, positional, and anisotropic B-factor refinement within PHENIX [367].

6.3 DIFFERENTIAL SCANNING FLUOROMETRY

6.3.1 Human Rtf1 Plus3 domain

Differential scanning fluorometry assays were performed following a method outlined in [309]. Each Plus3 variant was assayed at a final protein concentration of 33 µM, in a reaction buffer containing 20 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol and 1.7X SYPRO® Orange (Invitrogen, S-6650) in the presence or absence of 100 µM of peptide (pCTR, CTR). Protein unfolding was assessed by monitoring SYPRO Orange fluorescence at 570 nm as a function of temperature. Fluorescence data was analyzed using the Protein Thermal Shift Software v1.1 (Applied Biosystems), utilizing the Boltzmann model to calculate the T_m.

6.4 FLUORESCENCE ANISOTROPY

6.4.1 Human Rtf1 Plus3 domain

Fluorescence anisotropy was measured using a Cary Eclipse Fluorescence Spectrophotometer (Varian) equipped with an automated polarizer and multicell holder with a peltier thermostat. Titration experiments were measured at 4°C with an excitation wavelength of 498 nm (slit-width 5 nm) and an emission wavelength of 520 nm (slit-width 5 nm) using a photomultiplier tube voltage of 780 V. Purified Plus3 was titrated into a 400 µl reaction (10 mM HEPES, 50 mM NaCl) containing 20 nM fluorescein-labeled peptide (pCTR, pCTD, CTR) and equilibrated for 2.5 minutes before measuring its anisotropic signal. Binding curves for each protein were performed in triplicate and average values and standard deviation for each measurement was calculated. The Plus3-peptide dissociation constants (K_d) were determined in PRISM (GraphPad) using non-linear regression analysis. Anisotropy curves using the mutant forms of Plus3 were carried out as detailed above using a modified reaction buffer (20 mM Tris pH 8.0, 100 mM NaCl, 5% glycerol, and 1 mM β-mercaptoethanol). Fluorescence anisotropy experiments using fluorescein-labeled DNA substrates (Figure 22) were carried out exactly like those using peptide using a reaction buffer containing 20 mM Tris pH 8.0, 50 mM NaCl, 5% glycerol, and 1 mM β-mercaptoethanol.

For NaCl dissociation curves, fluorescence experiments were carried out at 20^{o}C with 20 nM fluorescein-labeled probe (pCTR, pCTD, CTR, ssDNA, dsDNA, dsBubble) and 20 µM purified Plus3 in a 400 µl reaction (10 mM HEPES, 50 mM NaCl). 5M NaCl was titrated into the reaction increasing concentrations from 50 mM to 1M, allowing a two minute equilibration

period before taking its anisotropic signal. NaCl dissociation data was graphed in PRISM (GraphPad).

6.4.2 *S. Cerevisiae* Csm2-Psy3

All experiments were performed using a Cary Eclipse Fluorescence Spectrophotometer (Varian) fitted with a peltier thermostatted multicell holder and automated polarizer. Fluorescent anisotropy/polarization measurements were collected with the excitation wavelength of 498 nm (slit-width 5 nm) and the emission wavelength at 520 nm (slit-width 5 nm), with a photomultiplier tube voltage of 780 V. Reactions were carried out at 30°C in a standard reaction buffer (20 mM Tris, pH 8, and 100 mM NaCl) with a total volume of 400 µl. A fluorescein-labelled DNA fork substrate was used in each of the experiments at a concentration of 20 nM. Purified Csm2–Psy3 was titrated into the reaction volume to the indicated concentration and allowed 7.5 min for equilibration before the anisotropy measurement. Protein was titrated until the anisotropy signal plateaued, indicating saturation of the labelled probe. The K_d and Hill coefficient were calculated in PRISM (GraphPad) using a global non-linear regression from the three Csm2–Psy3 binding isotherms. For the competition curves, fluorescence experiments were carried out with 25 nM fluorescein-labelled DNA fork and 546.6 nM Csm2–Psy3. The unlabelled competitors were added at increasing concentrations and allowed 7.5 min equilibration time before each measurement. Unlabelled DNA probe was added until polarized fluorescence stabilized, indicating saturation of the reaction with unlabelled DNA probe. Experiments were done in triplicate. The apparent K_i for each competitor was calculated in

PRISM using a non-linear regression assuming a single-binding site, which is consistent with our EMSA results.

6.5 ELECTROPHORETIC MOBILITY SHIFT ASSAYS

6.5.1 *S. cerevisiae* Rtf1 HMD

25 ng of nucleic acid (detailed in Table 6) was incubated on ice for 20 minutes with purified HMD (74-139, 74-189) with a final concentration ranging from 0.016-50 μM in EMSA reaction buffer (20 mM Tris pH 7.3, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.05 mg/ml BSA) in a reaction volume of 10 μl. Equilibrated samples were loaded on a pre-cooled and pre-run 5% native polyacrylamide gel containing 0.5X TBE and run at 70 volts for 3.5 hours at 4ºC. The resulting gel was stained with 1x SYBR Gold (Invitrogen) for 10 minutes at room temperature before being visualized using a LAS-3000 Imager (Fujifilm).

6.5.2 *S. cerevisiae* Csm2-Psy3

For competition EMSAs, fluorescein-labeled DNA fork (260 nM), detailed in Table 6, with or without an unlabeled competitor DNA oligonucleotide (1300 nM), was incubated on ice for 30 minutes with purified Csm2-Psy3 complex (3.0 μM) in EMSA reaction buffer (10 mM Tris pH 8.0, 50 mM NaCl, 2 mM DTT, 4 mM MgCl2) in a reaction volume of 10 μl. Equilibrated samples were loaded on a pre-cooled and pre-run 5% native polyacrylamide gel containing 0.5X

TBE and run at 200 volts for 2 hours at 4ºC. The resulting gel was visualized by fluorescence using a FLA-5100 Fluorescent Image Analyzer (FujiFilm).

6.6 LIMITED PROTEOLYSIS

6.6.1 Paf1 subcomplex – Ctr9, Paf1 1-385, Leo1, Cdc73 1-230

Purified Paf1 subcomplex (1.8 mg/ml) was subjected to limited proteolysis with 0.05% trypsin over a period of 3 hours. Samples were taken at different time points and stopped with 2 µl of phenylmethanesulfonylfluoride (PMSF) and SDS loading dye before being loaded on a pre-cooled and pre-run 15% SDS-polyacrylamide gel and run at 200 volts for 3 hours at 4^{o}C. Bands to be analyzed by in-gel tryptic digestion coupled with mass spectrometry were excised from the gel with a razor blade. Gel slices were washed three times in Destain buffer I (40% methanol, 7% acetic acid) with gentle agitation for 30 minutes, followed by three wash with Destain buffer II (5% methanol, 7% acetic acid), and a final three washes with deionized water.

Limited proteolysis of purified Paf1 subcomplex using Subtilisin A was performed exactly like those with trypsin, with the exception that digestions were performed over a one hour time period and Subtilisin A final concentrations were 0.05 mg/ml in the digestions.

6.7 MASS SPECTROMETRY

6.7.1 Paf1 subcomplex – Ctr9, Paf1 1-385, Leo1, Cdc73 1-230

All mass spectrometry work and analysis was performed in collaboration with Dr. Richard Gardner at the University of Washington (Department of Pharmacology). Excised gel bands from limited proteolysis experiments were subjected to in-gel tryptic digestion before being analyzed by tandem-mass spectrometry analysis using an Orbitrap mass spectrometer (Fred Hutchinson Cancer Research Center Proteomics Facility). Peptide Prophet [373] was used to evaluate the validity of peptide identifications, eliminating peptides with a score of <0.85 (<2.5% false discovery rate). Values represent the sum of identified peptides for each sample.

6.8 MOLECULAR DYNAMIC SIMULATIONS

6.8.1 Human Rtf1 Plus3 domain

All molecular dynamic simulations were performed in collaboration with Dr. Michael Grabe and Dr. Pushkar Pendse. Using the co-crystal structure of the human Plus3 domain of Rtf1 bound to a single Spt5 CTR repeat (PDB ID: 4L1U) as a starting point [311], the CTR peptide (GSGSRTPMYGSQ) was extended in Coot [312] to include an additional CTR repeat on either end (772-GSQTPMY-778, 791-TPLQD-795). The model was then solvated and ionized (0.15M NaCl) using CHARMM [374, 375], with an approximate system size of 88 Å x 88 Å x 88 Å, consisting of 59777 atoms. The model was parameterized using the CHARM36 force field [376]

with the CHARMM phosphothreonine parameter, and the TIP3P explicit water model [377]. The resulting model was minimized, heated, and equilibrated using NAMD2.7 [378] and the coordinates and velocities after heating were moved to Stampede at the Texas Advanced Computing Center (TACC) to carry out 20 ns of production.

APPENDIX A

THE CSM2-PSY3 HETERODIMER PREFERENTIALLY BINDS FORKED OR 3' OVERHANG DNA SUBSTRATES

I worked in collaboration with the Bernstein lab on a project focusing on the *S. cerevisiae* Shu complex. The Shu Complex promotes Rad51-dependent homologous recombination (HR) and consists of Shu1, Shu2, Csm2, and Psy3 [379, 380]. Recently two studies determined the structure of Csm2 complexed with Psy3 and found they are structurally homologous to Rad51. Additionally, both studies showed the Csm2-Psy3 heterodimer was capable of binding DNA [381, 382]. Neither study however examined DNA substrates more typical to what would be seen in HR, like 3' overhang DNA, making the preferred physiological substrate unclear.

To address the preferred DNA substrate of the Csm2-Psy3 heterodimer, I worked hand in hand with Stephen Godin of the Bernstein lab conducting a number of biochemical assays. After recombinantly expressing and purifying the *S. cerevisiae* Csm2-Psy3 heterodimer from *E. coli*, we used the purified complex in fluorescence anisotropy and electrophoretic mobility shift assays (EMSA), to determine binding affinities to different DNA substrates. First we assessed Csm2-Psy3's affinity to a fluorescein-labeled forked DNA substrate using fluorescence anisotropy, estimating an equilibrium dissociation constant, K_d, of 599 ± 105.4 nM (Figure 42A).

Next to begin defining the preferred binding substrate for the Csm2-Psy3 heterodimer, I performed a competition assay in the form of an EMSA. In this assay a mobility shift is observed when an unlabeled DNA substrate (DNA fork, 5' overhang, 3' overhang, dsDNA, ssDNA) competes off the fluorescently labeled fork substrate from the Csm2-Psy3 heterodimer. We found the Csm2/Psy3 heterodimer preferentially binds forked and 3' overhang DNA substrates (Figure 42B). Using fluorescence anisotropy, a more quantitative assay, we confirmed the results of the EMSA (Figure 42C).

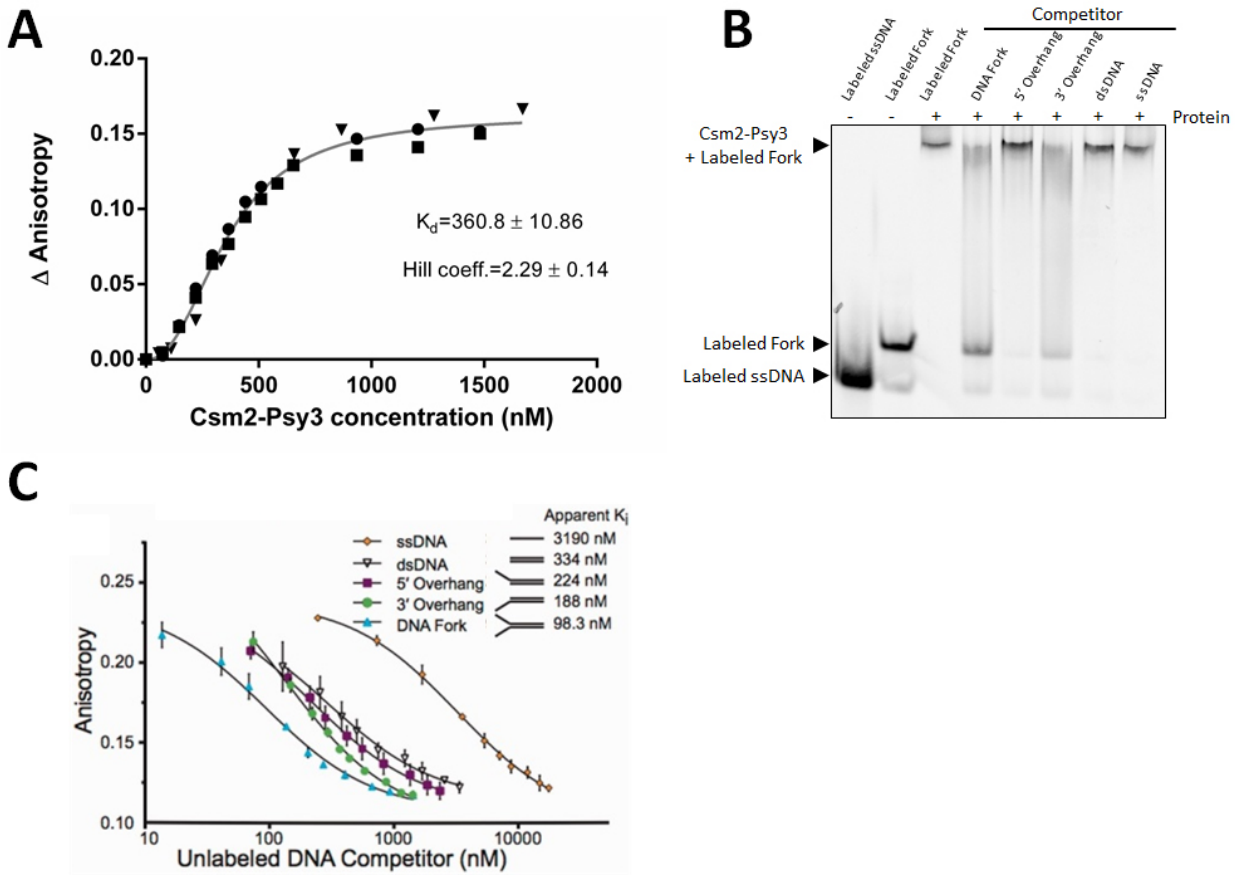


Figure 42. The Csm2-Psy3 heterodimer preferentially binds forked DNA substrates

A) DNA binding affinity of the Csm2–Psy3 heterodimer. Increasing concentrations of Csm2–Psy3 were added to a reaction mixture containing 25 nM fluorescein-labelled DNA fork in a fluorescence spectrophotometer. The binding isotherm was performed on three separate days, and each data set is shown. These data were fit using a global non-linear regression to obtain the K_d and a Hill coefficient. B) Fluorescein-labeled DNA fork (2.6 pmol) was incubated for 20 minutes on ice with Csm2-Psy3 heterodimer (30 pmol) either alone or in the presence of 13 pmol unlabeled DNA competitors (DNA Fork, 5' Overhang, 3' Overhang, dsDNA, ssDNA). Sample were then visualized via Native-PAGE. C) Complexes of Csm2–Psy3 bound to a fluorescein-labelled fork substrate were generated and fluorescence anisotropy was then measured with increasing concentrations of the indicated unlabelled competitors. Competition curves were fit using non-linear regression to calculate the apparent K_i. The experiment was done in triplicate, and standard deviations are plotted.

BIBLIOGRAPHY

1. Shandilya, J. and S.G. Roberts, *The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling.* Biochim Biophys Acta, 2012. **1819**(5): p. 391-400.
2. Liu, X., et al., *Initiation complex structure and promoter proofreading.* Science, 2011. **333**(6042): p. 633-7.
3. Sainsbury, S., J. Niesser, and P. Cramer, *Structure and function of the initially transcribing RNA polymerase II-TFIIB complex.* Nature, 2013. **493**(7432): p. 437-40.
4. Mayer, A., et al., *Uniform transitions of the general RNA polymerase II transcription complex.* Nat Struct Mol Biol, 2010. **17**(10): p. 1272-8.
5. Butler, J.E. and J.T. Kadonaga, *The RNA polymerase II core promoter: a key component in the regulation of gene expression.* Genes Dev, 2002. **16**(20): p. 2583-92.
6. Juven-Gershon, T., et al., *The RNA polymerase II core promoter - the gateway to transcription.* Curr Opin Cell Biol, 2008. **20**(3): p. 253-9.
7. Baumann, M., J. Pontiller, and W. Ernst, *Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview.* Mol Biotechnol, 2010. **45**(3): p. 241-7.
8. Orphanides, G., T. Lagrange, and D. Reinberg, *The general transcription factors of RNA polymerase II.* Genes Dev, 1996. **10**(21): p. 2657-83.
9. Burley, S.K. and R.G. Roeder, *Biochemistry and structural biology of transcription factor IID (TFIID).* Annu Rev Biochem, 1996. **65**: p. 769-99.
10. Carninci, P., et al., *Genome-wide analysis of mammalian promoter architecture and evolution.* Nat Genet, 2006. **38**(6): p. 626-35.
11. Smale, S.T. and D. Baltimore, *The "initiator" as a transcription control element.* Cell, 1989. **57**(1): p. 103-13.
12. Purnell, B.A., P.A. Emanuel, and D.S. Gilmour, *TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes.* Genes Dev, 1994. **8**(7): p. 830-42.
13. Burke, T.W. and J.T. Kadonaga, *Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters.* Genes Dev, 1996. **10**(6): p. 711-24.
14. Kadonaga, J.T., *The DPE, a core promoter element for transcription by RNA polymerase II.* Exp Mol Med, 2002. **34**(4): p. 259-64.
15. Kim, T.H., et al., *A high-resolution map of active promoters in the human genome.* Nature, 2005. **436**(7052): p. 876-80.

16. Bajic, V.B., V. Choudhary, and C.K. Hock, *Content analysis of the core promoter region of human genes.* In Silico Biol, 2004. **4**(2): p. 109-25.
17. Blackwood, E.M. and J.T. Kadonaga, *Going the distance: a current view of enhancer action.* Science, 1998. **281**(5373): p. 60-3.
18. Bulger, M. and M. Groudine, *Looping versus linking: toward a model for long-distance gene activation.* Genes Dev, 1999. **13**(19): p. 2465-77.
19. West, A.G., M. Gaszner, and G. Felsenfeld, *Insulators: many functions, many mechanisms.* Genes Dev, 2002. **16**(3): p. 271-88.
20. Buratowski, S., et al., *Five intermediate complexes in transcription initiation by RNA polymerase II.* Cell, 1989. **56**(4): p. 549-61.
21. Albright, S.R. and R. Tjian, *TAFs revisited: more data reveal new twists and confirm old ideas.* Gene, 2000. **242**(1-2): p. 1-13.
22. Wassarman, D.A. and F. Sauer, *TAF(II)250: a transcription toolbox.* J Cell Sci, 2001. **114**(Pt 16): p. 2895-902.
23. Kraemer, S.M., et al., *TFIIA interacts with TFIID via association with TATA-binding protein and TAF40.* Mol Cell Biol, 2001. **21**(5): p. 1737-46.
24. Deng, W. and S.G. Roberts, *TFIIB and the regulation of transcription by RNA polymerase II.* Chromosoma, 2007. **116**(5): p. 417-29.
25. Laine, J.P. and J.M. Egly, *When transcription and repair meet: a complex system.* Trends Genet, 2006. **22**(8): p. 430-6.
26. Lu, H., et al., *Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II.* Nature, 1992. **358**(6388): p. 641-5.
27. Malik, S. and R.G. Roeder, *Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells.* Trends Biochem Sci, 2000. **25**(6): p. 277-83.
28. Ossipow, V., et al., *A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation.* Cell, 1995. **83**(1): p. 137-46.
29. Chao, D.M., et al., *A mammalian SRB protein associated with an RNA polymerase II holoenzyme.* Nature, 1996. **380**(6569): p. 82-5.
30. Maldonado, E., et al., *A human RNA polymerase II complex associated with SRB and DNA-repair proteins.* Nature, 1996. **381**(6577): p. 86-9.
31. Wilson, C.J., et al., *RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling.* Cell, 1996. **84**(2): p. 235-44.
32. Saunders, A., L.J. Core, and J.T. Lis, *Breaking barriers to transcription elongation.* Nat Rev Mol Cell Biol, 2006. **7**(8): p. 557-67.
33. Egloff, S. and S. Murphy, *Cracking the RNA polymerase II CTD code.* Trends Genet, 2008. **24**(6): p. 280-8.
34. Gomes, N.P., et al., *Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program.* Genes Dev, 2006. **20**(5): p. 601-12.
35. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7.* Science, 2007. **318**(5857): p. 1780-2.
36. Ohkuma, Y. and R.G. Roeder, *Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation.* Nature, 1994. **368**(6467): p. 160-3.

37. Komarnitsky, P., E.J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.* Genes Dev, 2000. **14**(19): p. 2452-60.
38. Gerber, M. and A. Shilatifard, *Transcriptional elongation by RNA polymerase II and histone methylation.* J Biol Chem, 2003. **278**(29): p. 26303-6.
39. Hampsey, M. and D. Reinberg, *Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation.* Cell, 2003. **113**(4): p. 429-32.
40. Otero, G., et al., *Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation.* Mol Cell, 1999. **3**(1): p. 109-18.
41. Kwak, H. and J.T. Lis, *Control of transcriptional elongation.* Annu Rev Genet, 2013. **47**: p. 483-508.
42. Selth, L.A., S. Sigurdsson, and J.Q. Svejstrup, *Transcript Elongation by RNA Polymerase II.* Annu Rev Biochem, 2010. **79**: p. 271-93.
43. Yudkovsky, N., J.A. Ranish, and S. Hahn, *A transcription reinitiation intermediate that is stabilized by activator.* Nature, 2000. **408**(6809): p. 225-9.
44. Bengal, E., et al., *Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II.* Mol Cell Biol, 1991. **11**(3): p. 1195-206.
45. Tan, S., R.C. Conaway, and J.W. Conaway, *Dissection of transcription factor TFIIF functional domains required for initiation and elongation.* Proc Natl Acad Sci U S A, 1995. **92**(13): p. 6042-6.
46. Min, I.M., et al., *Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells.* Genes Dev, 2011. **25**(7): p. 742-54.
47. Nechaev, S., et al., *Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila.* Science, 2010. **327**(5963): p. 335-8.
48. Rasmussen, E.B. and J.T. Lis, *In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 7923-7.
49. Rhee, H.S. and B.F. Pugh, *Genome-wide structure and organization of eukaryotic pre-initiation complexes.* Nature, 2012. **483**(7389): p. 295-301.
50. Kwak, H., et al., *Precise maps of RNA polymerase reveal how promoters direct initiation and pausing.* Science, 2013. **339**(6122): p. 950-3.
51. Mavrich, T.N., et al., *Nucleosome organization in the Drosophila genome.* Nature, 2008. **453**(7193): p. 358-62.
52. Lee, C., et al., *NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila.* Mol Cell Biol, 2008. **28**(10): p. 3290-300.
53. Martinez-Rucobo, F.W., et al., *Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity.* EMBO J, 2011. **30**(7): p. 1302-10.
54. Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* Science, 2008. **322**(5909): p. 1845-8.
55. Price, D.H., *P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II.* Mol Cell Biol, 2000. **20**(8): p. 2629-34.
56. Ramanathan, Y., et al., *Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences.* J Biol Chem, 2001. **276**(14): p. 10913-20.
57. Kim, J.B. and P.A. Sharp, *Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase.* J Biol Chem, 2001. **276**(15): p. 12317-23.

58. Yamada, T., et al., *P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation.* Mol Cell, 2006. **21**(2): p. 227-37.
59. Fujinaga, K., et al., *Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element.* Mol Cell Biol, 2004. **24**(2): p. 787-95.
60. 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, 2011. **12**(5): p. 283-94.
61. Richard, P. and J.L. Manley, *Transcription termination by nuclear RNA polymerases.* Genes Dev, 2009. **23**(11): p. 1247-69.
62. Birse, C.E., et al., *Coupling termination of transcription to messenger RNA maturation in yeast.* Science, 1998. **280**(5361): p. 298-301.
63. Hirose, Y. and J.L. Manley, *RNA polymerase II and the integration of nuclear events.* Genes Dev, 2000. **14**(12): p. 1415-29.
64. Yonaha, M. and N.J. Proudfoot, *Transcriptional termination and coupled polyadenylation in vitro.* EMBO J, 2000. **19**(14): p. 3770-7.
65. Proudfoot, N., *New perspectives on connecting messenger RNA 3' end formation to transcription.* Curr Opin Cell Biol, 2004. **16**(3): p. 272-8.
66. Shearwin, K.E., B.P. Callen, and J.B. Egan, *Transcriptional interference--a crash course.* Trends Genet, 2005. **21**(6): p. 339-45.
67. Lykke-Andersen, S. and T.H. Jensen, *Overlapping pathways dictate termination of RNA polymerase II transcription.* Biochimie, 2007. **89**(10): p. 1177-1182.
68. Logan, J., et al., *A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene.* Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8306-10.
69. Whitelaw, E. and N. Proudfoot, *Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene.* EMBO J, 1986. **5**(11): p. 2915-22.
70. McCracken, S., et al., *The C-terminal domain of RNA polymerase II couples mRNA processing to transcription.* Nature, 1997. **385**(6614): p. 357-61.
71. Barilla, D., B.A. Lee, and N.J. Proudfoot, *Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2001. **98**(2): p. 445-50.
72. Ryan, K., O. Calvo, and J.L. Manley, *Evidence that polyadenylation factor CPSF-73 is the rnRNA 3 ' processing endonuclease*. Rna-a Publication of the Rna Society, 2004. **10**(4): p. 565-573.
73. Mandel, C.R., et al., *Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease.* Nature, 2006. **444**(7121): p. 953-6.
74. Minvielle-Sebastia, L., et al., *The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation.* Proc Natl Acad Sci U S A, 1997. **94**(15): p. 7897-902.
75. Orozco, I.J., S.J. Kim, and H.G. Martinson, *The Poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template.* Journal of Biological Chemistry, 2002. **277**(45): p. 42899-42911.

76. Nag, A., K. Narsinh, and H.G. Martinson, *The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase.* Nature Structural & Molecular Biology, 2007. **14**(7): p. 662-669.
77. Teixeira, A., et al., *Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination.* Nature, 2004. **432**(7016): p. 526-30.
78. Ghazal, G., et al., *Yeast RNase III triggers polyadenylation-independent transcription termination.* Mol Cell, 2009. **36**(1): p. 99-109.
79. Nabavi, S. and R.N. Nazar, *Pac1 endonuclease and Dhp1p 5'-->3' exonuclease are required for U3 snoRNA termination in Schizosaccharomyces pombe.* FEBS Lett, 2010. **584**(15): p. 3436-41.
80. Kim, M., et al., *The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II.* Nature, 2004. **432**(7016): p. 517-22.
81. West, S., N. Gromak, and N.J. Proudfoot, *Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites.* Nature, 2004. **432**(7016): p. 522-5.
82. Connelly, S. and J.L. Manley, *A Functional Messenger-Rna Polyadenylation Signal Is Required for Transcription Termination by Rna Polymerase-Ii.* Genes & Development, 1988. **2**(4): p. 440-452.
83. Houseley, J. and D. Tollervey, *The many pathways of RNA degradation.* Cell, 2009. **136**(4): p. 763-76.
84. Kim, H.D., J. Choe, and Y.S. Seo, *The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase.* Biochemistry, 1999. **38**(44): p. 14697-710.
85. Steinmetz, E.J. and D.A. Brow, *Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1.* Mol Cell Biol, 1996. **16**(12): p. 6993-7003.
86. Steinmetz, E.J., et al., *RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts.* Nature, 2001. **413**(6853): p. 327-31.
87. Steinmetz, E.J., et al., *Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase.* Mol Cell, 2006. **24**(5): p. 735-46.
88. Kornberg, R.D., *Structure of chromatin.* Annu Rev Biochem, 1977. **46**: p. 931-54.
89. Luger, K., M.L. Dechassa, and D.J. Tremethick, *New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?* Nature Reviews Molecular Cell Biology, 2012. **13**(7): p. 436-447.
90. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
91. Davey, C.A., et al., *Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution.* J Mol Biol, 2002. **319**(5): p. 1097-113.
92. Zentner, G.E. and S. Henikoff, *Regulation of nucleosome dynamics by histone modifications.* Nat Struct Mol Biol, 2013. **20**(3): p. 259-66.
93. Finch, J.T. and A. Klug, *Solenoidal model for superstructure in chromatin.* Proc Natl Acad Sci U S A, 1976. **73**(6): p. 1897-901.
94. Carruthers, L.M., et al., *Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding.* Biochemistry, 1998. **37**(42): p. 14776-87.

95. Thoma, F., T. Koller, and A. Klug, *Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin.* J Cell Biol, 1979. **83**(2 Pt 1): p. 403-27.
96. Garcia-Ramirez, M., F. Dong, and J. Ausio, *Role of the histone "tails" in the folding of oligonucleosomes depleted of histone H1.* J Biol Chem, 1992. **267**(27): p. 19587-95.
97. Moore, S.C. and J. Ausio, *Major role of the histones H3-H4 in the folding of the chromatin fiber.* Biochem Biophys Res Commun, 1997. **230**(1): p. 136-9.
98. Schalch, T., et al., *X-ray structure of a tetranucleosome and its implications for the chromatin fibre.* Nature, 2005. **436**(7047): p. 138-41.
99. Song, F., et al., *Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units.* Science, 2014. **344**(6182): p. 376-80.
100. Dorigo, B., et al., *Nucleosome arrays reveal the two-start organization of the chromatin fiber.* Science, 2004. **306**(5701): p. 1571-3.
101. Wasylyk, B. and P. Chambon, *Transcription by eukaryotic RNA polymerases A and B of chromatin assembled in vitro.* Eur J Biochem, 1979. **98**(2): p. 317-27.
102. Izban, M.G. and D.S. Luse, *Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing.* Genes Dev, 1991. **5**(4): p. 683-96.
103. Fairman-Williams, M.E., U.P. Guenther, and E. Jankowsky, *SF1 and SF2 helicases: family matters.* Curr Opin Struct Biol, 2010. **20**(3): p. 313-24.
104. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes.* Annu Rev Biochem, 2009. **78**: p. 273-304.
105. Gangaraju, V.K. and B. Bartholomew, *Mechanisms of ATP dependent chromatin remodeling.* Mutat Res, 2007. **618**(1-2): p. 3-17.
106. Carey, M., B. Li, and J.L. Workman, *RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation.* Mol Cell, 2006. **24**(3): p. 481-7.
107. Mas, G., et al., *Recruitment of a chromatin remodelling complex by the Hog1 MAP kinase to stress genes.* EMBO J, 2009. **28**(4): p. 326-36.
108. Delmas, V., D.G. Stokes, and R.P. Perry, *A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain.* Proc Natl Acad Sci U S A, 1993. **90**(6): p. 2414-8.
109. Marfella, C.G. and A.N. Imbalzano, *The Chd family of chromatin remodelers.* Mutat Res, 2007. **618**(1-2): p. 30-40.
110. Lusser, A., D.L. Urwin, and J.T. Kadonaga, *Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly.* Nat Struct Mol Biol, 2005. **12**(2): p. 160-6.
111. Stockdale, C., et al., *Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes.* J Biol Chem, 2006. **281**(24): p. 16279-88.
112. Stokes, D.G., K.D. Tartof, and R.P. Perry, *CHD1 is concentrated in interbands and puffed regions of Drosophila polytene chromosomes.* Proc Natl Acad Sci U S A, 1996. **93**(14): p. 7137-42.
113. Simic, R., et al., *Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes.* Embo Journal, 2003. **22**(8): p. 1846-1856.
114. Krogan, N.J., et al., *RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach.* Mol Cell Biol, 2002. **22**(20): p. 6979-92.

115. Kelley, D.E., D.G. Stokes, and R.P. Perry, *CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin.* Chromosoma, 1999. **108**(1): p. 10-25.
116. Smolle, M., et al., *Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange.* Nat Struct Mol Biol, 2012. **19**(9): p. 884-92.
117. Ebbert, R., A. Birkmann, and H.J. Schuller, *The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex.* Mol Microbiol, 1999. **32**(4): p. 741-51.
118. Shen, X., et al., *A chromatin remodelling complex involved in transcription and DNA processing.* Nature, 2000. **406**(6795): p. 541-4.
119. Conaway, R.C. and J.W. Conaway, *The INO80 chromatin remodeling complex in transcription, replication and repair.* Trends Biochem Sci, 2009. **34**(2): p. 71-7.
120. Kobor, M.S., et al., *A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin.* PLoS Biol, 2004. **2**(5): p. E131.
121. Krogan, N.J., et al., *A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1.* Mol Cell, 2003. **12**(6): p. 1565-76.
122. Orphanides, G., et al., *FACT, a factor that facilitates transcript elongation through nucleosomes.* Cell, 1998. **92**(1): p. 105-16.
123. Belotserkovskaya, R., et al., *FACT facilitates transcription-dependent nucleosome alteration.* Science, 2003. **301**(5636): p. 1090-3.
124. Xin, H., et al., *yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement.* Mol Cell, 2009. **35**(3): p. 365-76.
125. Bortvin, A. and F. Winston, *Evidence that Spt6p controls chromatin structure by a direct interaction with histones.* Science, 1996. **272**(5267): p. 1473-1476.
126. Kaplan, C.D., L. Laprade, and F. Winston, *Transcription elongation factors repress transcription initiation from cryptic sites.* Science, 2003. **301**(5636): p. 1096-9.
127. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis.* Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.
128. Phillips, D.M., *The presence of acetyl groups of histones.* Biochem J, 1963. **87**: p. 258-63.
129. Hong, L., et al., *Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA.* J Biol Chem, 1993. **268**(1): p. 305-14.
130. Dion, M.F., et al., *Genomic characterization reveals a simple histone H4 acetylation code.* Proc Natl Acad Sci U S A, 2005. **102**(15): p. 5501-6.
131. Martin, A.M., et al., *Redundant roles for histone H3 N-terminal lysine residues in subtelomeric gene repression in Saccharomyces cerevisiae.* Genetics, 2004. **167**(3): p. 1123-32.
132. Kouskouti, A. and I. Talianidis, *Histone modifications defining active genes persist after transcriptional and mitotic inactivation.* EMBO J, 2005. **24**(2): p. 347-57.
133. Workman, J.L. and R.E. Kingston, *Alteration of nucleosome structure as a mechanism of transcriptional regulation.* Annu Rev Biochem, 1998. **67**: p. 545-79.

134. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast.* Cell, 2005. **122**(4): p. 517-27.
135. Carrozza, M.J., et al., *Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription.* Cell, 2005. **123**(4): p. 581-92.
136. Wang, A., S.K. Kurdistani, and M. Grunstein, *Requirement of Hos2 histone deacetylase for gene activity in yeast.* Science, 2002. **298**(5597): p. 1412-4.
137. Govind, C.K., et al., *Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions.* Mol Cell, 2007. **25**(1): p. 31-42.
138. Kristjuhan, A., et al., *Transcriptional inhibition of genes with severe histone h3 hypoacetylation in the coding region.* Mol Cell, 2002. **10**(4): p. 925-33.
139. Daniel, J.A., M.G. Pray-Grant, and P.A. Grant, *Effector proteins for methylated histones: an expanding family.* Cell Cycle, 2005. **4**(7): p. 919-26.
140. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nature Structural & Molecular Biology, 2007. **14**(11): p. 1025-1040.
141. Li, B., et al., *The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II.* J Biol Chem, 2003. **278**(11): p. 8897-903.
142. Keogh, M.C., et al., *Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex.* Cell, 2005. **123**(4): p. 593-605.
143. Li, B., et al., *Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription.* J Biol Chem, 2009. **284**(12): p. 7970-6.
144. Li, B., et al., *Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin.* Science, 2007. **316**(5827): p. 1050-4.
145. Fischle, W., et al., *Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains.* Genes Dev, 2003. **17**(15): p. 1870-81.
146. Smolle, M. and J.L. Workman, *Transcription-associated histone modifications and cryptic transcription.* Biochim Biophys Acta, 2013. **1829**(1): p. 84-97.
147. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.
148. Weake, V.M. and J.L. Workman, *Histone ubiquitination: triggering gene activity.* Mol Cell, 2008. **29**(6): p. 653-63.
149. Henry, K.W., et al., *Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8.* Genes Dev, 2003. **17**(21): p. 2648-63.
150. Chandrasekharan, M.B., F. Huang, and Z.W. Sun, *Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation.* Epigenetics, 2010. **5**(6): p. 460-8.
151. Schulze, J.M., et al., *Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123.* Genes Dev, 2011. **25**(21): p. 2242-7.
152. Kao, C.F., et al., *Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B.* Genes Dev, 2004. **18**(2): p. 184-95.

153. Xiao, T., et al., *Histone H2B ubiquitylation is associated with elongating RNA polymerase II.* Mol Cell Biol, 2005. **25**(2): p. 637-51.
154. Pavri, R., et al., *Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II.* Cell, 2006. **125**(4): p. 703-17.
155. Hwang, W.W., et al., *A conserved RING finger protein required for histone H2B monoubiquitination and cell size control.* Mol Cell, 2003. **11**(1): p. 261-6.
156. Robzyk, K., J. Recht, and M.A. Osley, *Rad6-dependent ubiquitination of histone H2B in yeast.* Science, 2000. **287**(5452): p. 501-4.
157. Wood, A., et al., *Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter.* Mol Cell, 2003. **11**(1): p. 267-74.
158. Briggs, S.D., et al., *Gene silencing: trans-histone regulatory pathway in chromatin.* Nature, 2002. **418**(6897): p. 498.
159. Sun, Z.W. and C.D. Allis, *Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast.* Nature, 2002. **418**(6893): p. 104-8.
160. Dover, J., et al., *Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6.* J Biol Chem, 2002. **277**(32): p. 28368-71.
161. Govind, C.K., et al., *Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes.* Mol Cell, 2010. **39**(2): p. 234-46.
162. Kim, T. and S. Buratowski, *Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions.* Cell, 2009. **137**(2): p. 259-72.
163. Martin, D.G., et al., *The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3.* Mol Cell Biol, 2006. **26**(21): p. 7871-9.
164. Taverna, S.D., et al., *Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs.* Mol Cell, 2006. **24**(5): p. 785-96.
165. Shi, X., et al., *Paf1p, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may have both positive and negative roles in transcription.* Mol Cell Biol, 1996. **16**(2): p. 669-76.
166. Wade, P.A., et al., *A novel collection of accessory factors associated with yeast RNA polymerase II.* Protein Expr Purif, 1996. **8**(1): p. 85-90.
167. Mueller, C.L. and J.A. Jaehning, *Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex.* Mol Cell Biol, 2002. **22**(7): p. 1971-80.
168. Shi, X., et al., *Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme.* Mol Cell Biol, 1997. **17**(3): p. 1160-9.
169. Squazzo, S.L., et al., *The Paf1 complex physically and functionally associates with transcription elongation factors in vivo.* EMBO J, 2002. **21**(7): p. 1764-74.
170. Zhu, B., et al., *The human PAF complex coordinates transcription with events downstream of RNA synthesis.* Genes Dev, 2005. **19**(14): p. 1668-73.
171. Anderson, J.S. and R.P. Parker, *The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex.* EMBO J, 1998. **17**(5): p. 1497-506.

172. Kim, M., et al., *Transitions in RNA polymerase II elongation complexes at the 3' ends of genes.* EMBO J, 2004. **23**(2): p. 354-64.
173. Costa, P.J. and K.M. Arndt, *Synthetic lethal interactions suggest a role for the Saccharomyces cerevisiae Rtf1 protein in transcription elongation.* Genetics, 2000. **156**(2): p. 535-47.
174. Bentley, R., *Mycophenolic Acid: a one hundred year odyssey from antibiotic to immunosuppressant.* Chem Rev, 2000. **100**(10): p. 3801-26.
175. Stolinski, L.A., D.M. Eisenmann, and K.M. Arndt, *Identification of RTF1, a novel gene important for TATA site selection by TATA box-binding protein in Saccharomyces cerevisiae.* Mol Cell Biol, 1997. **17**(8): p. 4490-500.
176. Betz, J.L., et al., *Phenotypic analysis of Paf1/RNA polymerase II complex mutations reveals connections to cell cycle regulation, protein synthesis, and lipid and nucleic acid metabolism.* Mol Genet Genomics, 2002. **268**(2): p. 272-85.
177. Warner, M.H., K.L. Roinick, and K.M. Arndt, *Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification.* Mol Cell Biol, 2007. **27**(17): p. 6103-15.
178. Ryan, D.P., et al., *The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains.* EMBO J, 2011. **30**(13): p. 2596-609.
179. McKnight, J.N., et al., *Extranucleosomal DNA binding directs nucleosome sliding by Chd1.* Mol Cell Biol, 2011. **31**(23): p. 4746-59.
180. Patel, A., et al., *Identification of residues in chromodomain helicase DNA-binding protein 1 (Chd1) required for coupling ATP hydrolysis to nucleosome sliding.* J Biol Chem, 2011. **286**(51): p. 43984-93.
181. Kim, J., M. Guermah, and R.G. Roeder, *The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS.* Cell, 2010. **140**(4): p. 491-503.
182. Mueller, C.L., et al., *The Paf1 complex has functions independent of actively transcribing RNA polymerase II.* Mol Cell, 2004. **14**(4): p. 447-56.
183. Nordick, K., et al., *Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II.* Eukaryot Cell, 2008. **7**(7): p. 1158-67.
184. Rozenblatt-Rosen, O., et al., *The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 755-60.
185. Sheldon, K.E., D.M. Mauger, and K.M. Arndt, *A Requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation.* Mol Cell, 2005. **20**(2): p. 225-36.
186. Pruneski, J.A., et al., *The Paf1 complex represses SER3 transcription in Saccharomyces cerevisiae by facilitating intergenic transcription-dependent nucleosome occupancy of the SER3 promoter.* Eukaryot Cell, 2011. **10**(10): p. 1283-94.
187. Chu, Y., et al., *Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes.* EMBO J, 2007. **26**(22): p. 4646-56.
188. Krogan, N.J., et al., *The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation.* Mol Cell, 2003. **11**(3): p. 721-9.

189. Ng, H.H., S. Dole, and K. Struhl, *The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B.* J Biol Chem, 2003. **278**(36): p. 33625-8.
190. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity.* Mol Cell, 2003. **11**(3): p. 709-19.
191. Wood, A., et al., *The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p.* J Biol Chem, 2003. **278**(37): p. 34739-42.
192. Penheiter, K.L., et al., *A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets.* Mol Cell, 2005. **20**(2): p. 213-23.
193. Porter, S.E., et al., *The yeast Paf1-RNA polymerase II complex is required for full expression of a subset of cell cycle-regulated genes.* Eukaryotic Cell, 2002. **1**(5): p. 830-842.
194. Moniaux, N., et al., *The human RNA polymerase II-associated factor 1 (hPaf1): a new regulator of cell-cycle progression.* PLoS One, 2009. **4**(9): p. e7077.
195. Chaudhary, K., et al., *Human RNA polymerase II-associated factor complex: dysregulation in cancer.* Oncogene, 2007. **26**(54): p. 7499-507.
196. Selvarajan, S., et al., *Parafibromin expression in breast cancer: a novel marker for prognostication?* J Clin Pathol, 2008. **61**(1): p. 64-7.
197. Liu, L., et al., *A whole genome screen for HIV restriction factors.* Retrovirology, 2011. **8**: p. 94.
198. Marazzi, I., et al., *Suppression of the antiviral response by an influenza histone mimic.* Nature, 2012. **483**(7390): p. 428-33.
199. Ding, L., et al., *A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity.* Cell Stem Cell, 2009. **4**(5): p. 403-15.
200. Ponnusamy, M.P., et al., *RNA polymerase II associated factor 1/PD2 maintains self-renewal by its interaction with Oct3/4 in mouse embryonic stem cells.* Stem Cells, 2009. **27**(12): p. 3001-11.
201. Xiong, L., et al., *Mass spectrometric studies on epigenetic interaction networks in cell differentiation.* J Biol Chem, 2011. **286**(15): p. 13657-68.
202. Laribee, R.N., et al., *BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex.* Curr Biol, 2005. **15**(16): p. 1487-93.
203. Nakanishi, S., et al., *Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1*. J Cell Biol, 2009. **186**(3): p. 371-7.
204. Tomson, B.N., et al., *Identification of a role for histone H2B ubiquitylation in noncoding RNA 3'-end formation through mutational analysis of Rtf1 in Saccharomyces cerevisiae.* Genetics, 2011. **188**(2): p. 273-89.
205. Piro, A.S., et al., *Small region of Rtf1 protein can substitute for complete Paf1 complex in facilitating global histone H2B ubiquitylation in yeast.* Proc Natl Acad Sci U S A, 2012. **109**(27): p. 10837-42.

206. Kim, J., et al., *RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells.* Cell, 2009. **137**(3): p. 459-71.
207. Minsky, N., et al., *Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells.* Nat Cell Biol, 2008. **10**(4): p. 483-8.
208. Bernstein, B.E., et al., *Methylation of histone H3 Lys 4 in coding regions of active genes.* Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8695-700.
209. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3.* Nature, 2002. **419**(6905): p. 407-411.
210. Fleming, A.B., et al., *H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation.* Molecular Cell, 2008. **31**(1): p. 57-66.
211. Zhu, B., et al., *Monoubiquitination of human histone H2B: The factors involved and their roles in HOX gene regulation.* Molecular Cell, 2005. **20**(4): p. 601-611.
212. Mutiu, A.I., et al., *The role of histone ubiquitylation and deubiquitylation in gene expression as determined by the analysis of an HTB1(K123R) Saccharomyces cerevisiae strain.* Mol Genet Genomics, 2007. **277**(5): p. 491-506.
213. Daniel, J.A., et al., *Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription.* J Biol Chem, 2004. **279**(3): p. 1867-71.
214. Crisucci, E.M. and K.M. Arndt, *The Paf1 Complex Represses ARG1 Transcription in Saccharomyces cerevisiae by Promoting Histone Modifications.* Eukaryotic Cell, 2011. **10**(6): p. 712-723.
215. Krogan, N.J., et al., *COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression.* J Biol Chem, 2002. **277**(13): p. 10753-5.
216. Nislow, C., E. Ray, and L. Pillus, *SET1, a yeast member of the Trithorax family, functions in transcriptional silencing and diverse cellular processes.* Molecular Biology of the Cell, 1997. **8**(12): p. 2421-2436.
217. Huang, H.H., et al., *The ubiquitin-conjugating enzyme Rad6 (Ubc2) is required for silencing in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 1997. **17**(11): p. 6693-6699.
218. Ng, H.H., et al., *Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association.* Genes Dev, 2002. **16**(12): p. 1518-27.
219. Li, J., D. Moazed, and S.P. Gygi, *Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation.* J Biol Chem, 2002. **277**(51): p. 49383-8.
220. Xiao, T., et al., *Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast.* Genes Dev, 2003. **17**(5): p. 654-63.
221. Krogan, N.J., et al., *Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II.* Mol Cell Biol, 2003. **23**(12): p. 4207-18.
222. Kizer, K.O., et al., *A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation.* Mol Cell Biol, 2005. **25**(8): p. 3305-16.
223. Krogan, N.J., et al., *Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II.* Molecular and Cellular Biology, 2003. **23**(12): p. 4207-4218.

224. Ahn, S.H., M. Kim, and S. Buratowski, *Phosphorylation of serine 2 within the RNA polymerase IIC-terminal domain couples transcription and 3 ' end processing.* Molecular Cell, 2004. **13**(1): p. 67-76.
225. Schaft, D., et al., *The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation.* Nucleic Acids Res, 2003. **31**(10): p. 2475-82.
226. Meinhart, A. and P. Cramer, *Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors.* Nature, 2004. **430**(6996): p. 223-6.
227. Licatalosi, D.D., et al., *Functional interaction of yeast pre-mRNA 3 ' end processing factors with RNA polymerase II.* Molecular Cell, 2002. **9**(5): p. 1101-1111.
228. Nagaike, T., et al., *Transcriptional activators enhance polyadenylation of mRNA precursors.* Mol Cell, 2011. **41**(4): p. 409-18.
229. Matera, A.G., R.M. Terns, and M.P. Terns, *Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs.* Nat Rev Mol Cell Biol, 2007. **8**(3): p. 209-20.
230. Martens, J.A., P.Y. Wu, and F. Winston, *Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae.* Genes Dev, 2005. **19**(22): p. 2695-704.
231. Martens, J.A., L. Laprade, and F. Winston, *Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene.* Nature, 2004. **429**(6991): p. 571-574.
232. Thompson, D.M. and R. Parker, *Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae.* Mol Cell Biol, 2007. **27**(1): p. 92-101.
233. Martens, J.A. and F. Winston, *Evidence that Swi/Snf directly represses transcription in S-cerevisiae.* Genes & Development, 2002. **16**(17): p. 2231-2236.
234. Hainer, S.J. and J.A. Martens, *Identification of histone mutants that are defective for transcription-coupled nucleosome occupancy.* Mol Cell Biol, 2011. **31**(17): p. 3557-68.
235. Thebault, P., et al., *Transcription Regulation by the Noncoding RNA SRG1 Requires Spt2-Dependent Chromatin Deposition in the Wake of RNA Polymerase II.* Molecular and Cellular Biology, 2011. **31**(6): p. 1288-1300.
236. Hainer, S.J., et al., *Intergenic transcription causes repression by directing nucleosome assembly.* Genes Dev, 2011. **25**(1): p. 29-40.
237. Koch, C., et al., *A role for Ctr9p and Paf1p in the regulation G(1) cyclin expression in yeast.* Nucleic Acids Research, 1999. **27**(10): p. 2126-2134.
238. Rozenblatt-Rosen, O., et al., *The parafibromin tumor suppressor protein is part of a human Paf1 complex.* Molecular and Cellular Biology, 2005. **25**(2): p. 612-620.
239. Yart, A., et al., *The HRPT2 tumor suppressor gene product parafibromin associates with human PAF1 and RNA polymerase II.* Mol Cell Biol, 2005. **25**(12): p. 5052-60.
240. Chu, X., et al., *Structural insights into Paf1 complex assembly and histone binding.* Nucleic Acids Res, 2013. **41**(22): p. 10619-29.
241. Di Como, C.J., H. Chang, and K.T. Arndt, *Activation of CLN1 and CLN2 G1 cyclin gene expression by BCK2.* Mol Cell Biol, 1995. **15**(4): p. 1835-46.
242. Foreman, P.K. and R.W. Davis, *CDP1, a novel Saccharomyces cerevisiae gene required for proper nuclear division and chromosome segregation.* Genetics, 1996. **144**(4): p. 1387-1397.
243. Nagase, T., et al., *Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1.* DNA Res, 1995. **2**(4): p. 167-74, 199-210.

244. Malek, S.N., et al., *p150(TSP), conserved nuclear phosphoprotein that contains multiple tetratricopeptide repeats and binds specifically to SH2 domains.* Journal of Biological Chemistry, 1996. **271**(12): p. 6952-6962.
245. D'Andrea, L.D. and L. Regan, *TPR proteins: the versatile helix.* Trends Biochem Sci, 2003. **28**(12): p. 655-62.
246. Sikorski, R.S., et al., *A Repeating Amino-Acid Motif in Cdc23 Defines a Family of Proteins and a New Relationship among Genes Required for Mitosis and Rna-Synthesis.* Cell, 1990. **60**(2): p. 307-317.
247. Das, A.K., P.W. Cohen, and D. Barford, *The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions.* EMBO J, 1998. **17**(5): p. 1192-9.
248. Scheufler, C., et al., *Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine.* Cell, 2000. **101**(2): p. 199-210.
249. Lapouge, K., et al., *Structure of the TPR domain of p67phox in complex with Rac.GTP.* Mol Cell, 2000. **6**(4): p. 899-907.
250. Gatto, G.J., et al., *Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5.* Nature Structural Biology, 2000. **7**(12): p. 1091-1095.
251. Schultz, J., et al., *SMART, a simple modular architecture research tool: identification of signaling domains.* Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5857-64.
252. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004. **20**(13): p. 2138-9.
253. Magdolen, V., et al., *The gene LEO1 on yeast chromosome XV encodes a non-essential, extremely hydrophilic protein.* Biochim Biophys Acta, 1994. **1218**(2): p. 205-9.
254. Zhao, L., T. Tong, and Z. Zhang, *Expression of the Leo1-like domain of replicative senescence down-regulated Leo1-like (RDL) protein promotes senescence of 2BS fibroblasts.* FASEB J, 2005. **19**(6): p. 521-32.
255. Dermody, J.L. and S. Buratowski, *Leo1 subunit of the yeast paf1 complex binds RNA and contributes to complex recruitment.* J Biol Chem, 2010. **285**(44): p. 33671-9.
256. Reed, S.I., J. Ferguson, and K.Y. Jahng, *Isolation and characterization of two genes encoding yeast mating pheromone signaling elements: CDC72 and CDC73.* Cold Spring Harb Symp Quant Biol, 1988. **53 Pt 2**: p. 621-7.
257. Carpten, J.D., et al., *HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome.* Nat Genet, 2002. **32**(4): p. 676-80.
258. Howell, V.M., et al., *HRPT2 mutations are associated with malignancy in sporadic parathyroid tumours.* J Med Genet, 2003. **40**(9): p. 657-63.
259. Shattuck, T.M., et al., *Somatic and germ-line mutations of the HRPT2 gene in sporadic parathyroid carcinoma.* N Engl J Med, 2003. **349**(18): p. 1722-9.
260. Woodard, G.E., et al., *Parafibromin, product of the hyperparathyroidism-jaw tumor syndrome gene HRPT2, regulates cyclin D1/PRAD1 expression.* Oncogene, 2005. **24**(7): p. 1272-6.
261. Yang, Y.J., et al., *The tumor suppressor, parafibromin, mediates histone H3 K9 methylation for cyclin D1 repression.* Nucleic Acids Res, 2010. **38**(2): p. 382-90.
262. Mosimann, C., G. Hausmann, and K. Basler, *Parafibromin/Hyrax activates Wnt/Wg target gene transcription by direct association with beta-catenin/Armadillo.* Cell, 2006. **125**(2): p. 327-41.

263. Newey, P.J., M.R. Bowl, and R.V. Thakker, *Parafibromin--functional insights.* J Intern Med, 2009. **266**(1): p. 84-98.
264. Amrich, C.G., et al., *Cdc73 subunit of Paf1 complex contains C-terminal Ras-like domain that promotes association of Paf1 complex with chromatin.* J Biol Chem, 2012. **287**(14): p. 10863-75.
265. Qiu, H., et al., *Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex.* EMBO J, 2012. **31**(16): p. 3494-505.
266. Liu, Y., et al., *Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex.* Mol Cell Biol, 2009. **29**(17): p. 4852-63.
267. Zhou, K., et al., *Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 6956-61.
268. Qiu, H., et al., *The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II.* Mol Cell Biol, 2006. **26**(8): p. 3135-48.
269. Chen, H., et al., *Crystallographic analysis of the conserved C-terminal domain of transcription factor Cdc73 from Saccharomyces cerevisiae reveals a GTPase-like fold.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 8): p. 953-9.
270. Wennerberg, K., K.L. Rossman, and C.J. Der, *The Ras superfamily at a glance.* Journal of Cell Science, 2005. **118**(5): p. 843-846.
271. Warner, M.H., K.L. Roinick, and K.M. Arndt, *Rtf1 is a multifunctional component of the paf1 complex that regulates gene expression by directing cotranscriptional histone modification.* Molecular and Cellular Biology, 2007. **27**(17): p. 6103-6115.
272. Simic, R., et al., *Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes.* EMBO J, 2003. **22**(8): p. 1846-56.
273. Mayekar, M.K., R.G. Gardner, and K.M. Arndt, *The Recruitment of the Saccharomyces cerevisiae Paf1 Complex to Active Genes Requires a Domain of Rtf1 That Directly Interacts with the Spt4-Spt5 Complex.* Mol Cell Biol, 2013. **33**(16): p. 3259-73.
274. Hirtreiter, A., et al., *Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif.* Nucleic Acids Res, 2010. **38**(12): p. 4040-51.
275. Pei, Y. and S. Shuman, *Characterization of the Schizosaccharomyces pombe Cdk9/Pch1 protein kinase: Spt5 phosphorylation, autophosphorylation, and mutational analysis.* J Biol Chem, 2003. **278**(44): p. 43346-56.
276. Ivanov, D., et al., *Domains in the SPT5 protein that modulate its transcriptional regulatory properties.* Mol Cell Biol, 2000. **20**(9): p. 2970-83.
277. Qiu, H.F., C.H. Hu, and A.G. Hinnebusch, *Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters.* Molecular Cell, 2009. **33**(6): p. 752-762.
278. de Jong, R.N., et al., *Structure and DNA binding of the human Rtf1 Plus3 domain.* Structure, 2008. **16**(1): p. 149-59.
279. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Research, 2010. **38**: p. W545-W549.
280. Sobhian, B., et al., *HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP.* Mol Cell, 2010. **38**(3): p. 439-51.

281. Muntean, A.G., et al., *The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis.* Cancer Cell, 2010. **17**(6): p. 609-21.
282. Smith, E., C.Q. Lin, and A. Shilatifard, *The super elongation complex (SEC) and MLL in development and disease.* Genes & Development, 2011. **25**(7): p. 661-672.
283. Krivtsov, A.V. and S.A. Armstrong, *MLL translocations, histone modifications and leukaemia stem-cell development.* Nature Reviews Cancer, 2007. **7**(11): p. 823-833.
284. Marschalek, R., *Mechanisms of leukemogenesis by MLL fusion proteins.* British Journal of Haematology, 2011. **152**(2): p. 141-154.
285. Yokoyama, A., et al., *A Higher-Order Complex Containing AF4 and ENL Family Proteins with P-TEFb Facilitates Oncogenic and Physiologic MLL-Dependent Transcription.* Cancer Cell, 2010. **17**(2): p. 198-212.
286. Krivtsov, A.V., et al., *H3K79 Methylation Profiles Define Murine and Human MLL-AF4 Leukemias.* Cancer Cell, 2008. **14**(5): p. 355-368.
287. Mueller, D., et al., *A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification.* Blood, 2007. **110**(13): p. 4445-4454.
288. Moniaux, N., et al., *The human homologue of the RNA polymerase II-associated factor 1 (hPaf1), localized on the 19q13 amplicon, is associated with tumorigenesis.* Oncogene, 2006. **25**(23): p. 3247-57.
289. Wang, P.F., et al., *HRPT2, a tumor suppressor gene for hyperparathyroidism-jaw tumor syndrome.* Hormone and Metabolic Research, 2005. **37**(6): p. 380-383.
290. Crisucci, E.M. and K.M. Arndt, *The Roles of the Paf1 Complex and Associated Histone Modifications in Regulating Gene Expression.* Genet Res Int, 2011. **2011**.
291. Jaehning, J.A., *The Paf1 complex: platform or player in RNA polymerase II transcription?* Biochim Biophys Acta, 2010. **1799**(5-6): p. 379-88.
292. Chen, Y., et al., *DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation.* Genes Dev, 2009. **23**(23): p. 2765-77.
293. Hartzog, G.A. and J. Fu, *The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation.* Biochim Biophys Acta, 2013. **1829**(1): p. 105-15.
294. Lindstrom, D.L., et al., *Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins.* Mol Cell Biol, 2003. **23**(4): p. 1368-78.
295. Mayer, A., et al., *The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I.* Mol Cell Biol, 2012. **32**(7): p. 1321-31.
296. Wood, A., et al., *The Bur1/Bur2 complex is required for histone H2B monoubiquitination by Rad6/Bre1 and histone methylation by COMPASS.* Mol Cell, 2005. **20**(4): p. 589-99.
297. Holm, L., et al., *Searching protein structure databases with DaliLite v.3.* Bioinformatics, 2008. **24**(23): p. 2780-1.
298. Bian, C., et al., *Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation.* EMBO J, 2011. **30**(14): p. 2829-42.
299. Hauk, G., et al., *The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor.* Mol Cell, 2010. **39**(5): p. 711-23.
300. Tripsianes, K., et al., *Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins.* Nat Struct Mol Biol, 2011. **18**(12): p. 1414-20.

301. Botuyan, M.V., et al., *Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair.* Cell, 2006. **127**(7): p. 1361-73.
302. Law, J.A., et al., *Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1.* Nature, 2013. **498**(7454): p. 385-9.
303. Friberg, A., et al., *Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN.* J Mol Biol, 2009. **387**(4): p. 921-34.
304. Liu, K., et al., *Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain.* Proc Natl Acad Sci U S A, 2010. **107**(43): p. 18398-403.
305. Mathioudakis, N., et al., *The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors.* RNA, 2012. **18**(11): p. 2056-72.
306. Wada, T., et al., *DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs.* Genes Dev, 1998. **12**(3): p. 343-56.
307. Rose, G.D., et al., *Hydrophobicity of amino acid residues in globular proteins.* Science, 1985. **229**(4716): p. 834-8.
308. Cimmperman, P., et al., *A quantitative model of thermal stabilization and destabilization of proteins by ligands.* Biophys J, 2008. **95**(7): p. 3222-31.
309. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.
310. Pantoliano, M.W., et al., *High-density miniaturized thermal shift assays as a general strategy for drug discovery.* J Biomol Screen, 2001. **6**(6): p. 429-40.
311. Wier, A.D., et al., *Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin.* Proc Natl Acad Sci U S A, 2013. **110**(43): p. 17290-5.
312. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
313. Becker, R., B. Loll, and A. Meinhart, *Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II.* J Biol Chem, 2008. **283**(33): p. 22659-69.
314. Ghosh, A., S. Shuman, and C.D. Lima, *Structural insights to how mammalian capping enzyme reads the CTD code.* Mol Cell, 2011. **43**(2): p. 299-310.
315. Kim, D., et al., *Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain.* Nat Struct Mol Biol, 2010. **17**(8): p. 1027-9.
316. Mayer, A., et al., *CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II.* Science, 2012. **336**(6089): p. 1723-5.
317. Kumaki, Y., et al., *Structure of the YSPTSPS repeat containing two SPXX motifs in the CTD of RNA polymerase II: NMR studies of cyclic model peptides reveal that the SPTS turn is more stable than SPSY in water.* Biochim Biophys Acta, 2001. **1548**(1): p. 81-93.
318. Cagas, P.M. and J.L. Corden, *Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II.* Proteins, 1995. **21**(2): p. 149-60.
319. Viktorovskaya, O.V., F.D. Appling, and D.A. Schneider, *Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly.* J Biol Chem, 2011. **286**(21): p. 18825-33.

320. Buchan, D.W.A., et al., *Scalable web services for the PSIPRED Protein Analysis Workbench.* Nucleic Acids Research, 2013. **41**(W1): p. W349-W357.
321. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* Journal of Molecular Biology, 1999. **292**(2): p. 195-202.
322. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.
323. Collaborative Computational Project, N., *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
324. Vagin, A. and A. Teplyakov, *Molecular replacement with MOLREP.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 22-5.
325. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.
326. Bustin, M., *Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins.* Mol Cell Biol, 1999. **19**(8): p. 5237-46.
327. Thomas, J.O. and A.A. Travers, *HMG1 and 2, and related 'architectural' DNA-binding proteins.* Trends Biochem Sci, 2001. **26**(3): p. 167-74.
328. Allain, F.H., et al., *Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding.* EMBO J, 1999. **18**(9): p. 2563-79.
329. Murphy, F.V.t., R.M. Sweet, and M.E. Churchill, *The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition.* EMBO J, 1999. **18**(23): p. 6610-8.
330. Stillman, D.J., *Nhp6: a small but powerful effector of chromatin structure in Saccharomyces cerevisiae.* Biochim Biophys Acta, 2010. **1799**(1-2): p. 175-80.
331. Masse, J.E., et al., *The S. cerevisiae architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding.* J Mol Biol, 2002. **323**(2): p. 263-84.
332. McPherson, A. and B. Cudney, *Searching for silver bullets: an alternative strategy for crystallizing macromolecules.* J Struct Biol, 2006. **156**(3): p. 387-406.
333. Dorigo, B., et al., *Chromatin fiber folding: requirement for the histone H4 N-terminal tail.* J Mol Biol, 2003. **327**(1): p. 85-96.
334. Gordon, F., K. Luger, and J.C. Hansen, *The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays.* J Biol Chem, 2005. **280**(40): p. 33701-6.
335. McBryant, S.J., et al., *Determinants of histone H4 N-terminal domain function during nucleosomal array oligomerization: roles of amino acid sequence, domain length, and charge density.* J Biol Chem, 2009. **284**(25): p. 16716-22.
336. Armache, K.J., et al., *Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 A resolution.* Science, 2011. **334**(6058): p. 977-82.
337. Barbera, A.J., et al., *The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA.* Science, 2006. **311**(5762): p. 856-61.
338. Makde, R.D., et al., *Structure of RCC1 chromatin factor bound to the nucleosome core particle.* Nature, 2010. **467**(7315): p. 562-6.

339. Kato, H., et al., *Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR.* Proc Natl Acad Sci U S A, 2011. **108**(30): p. 12283-8.
340. Roussel, L., et al., *Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A-H2B acidic pocket.* EMBO Rep, 2008. **9**(10): p. 1006-12.
341. Liu, Q., J.C. Greimann, and C.D. Lima, *Reconstitution, activities, and structure of the eukaryotic RNA exosome.* Cell, 2006. **127**(6): p. 1223-37.
342. Scheich, C., et al., *Vectors for co-expression of an unrestricted number of proteins.* Nucleic Acids Res, 2007. **35**(6): p. e43.
343. Mbogning, J., et al., *The PAF complex and Prf1/Rtf1 delineate distinct Cdk9-dependent pathways regulating transcription elongation in fission yeast.* PLoS Genet, 2013. **9**(12): p. e1004029.
344. Godin, S., et al., *The Shu complex interacts with Rad51 through the Rad51 paralogues Rad55-Rad57 to mediate error-free recombination.* Nucleic Acids Res, 2013. **41**(8): p. 4525-34.
345. Tao, J., et al., *Crystal structure of P58(IPK) TPR fragment reveals the mechanism for its molecular chaperone activity in UPR.* J Mol Biol, 2010. **397**(5): p. 1307-15.
346. Han, D., et al., *TPR domain of NrfG mediates complex formation between heme lyase and formate-dependent nitrite reductase in Escherichia coli O157:H7.* Proteins, 2008. **70**(3): p. 900-14.
347. Krenn, V., et al., *Structural analysis reveals features of the spindle checkpoint kinase Bub1-kinetochore subunit Knl1 interaction.* J Cell Biol, 2012. **196**(4): p. 451-67.
348. Morgan, R.M., et al., *Structure of the TPR domain of AIP: lack of client protein interaction with the C-terminal alpha-7 helix of the TPR domain of AIP is sufficient for pituitary adenoma predisposition.* PLoS One, 2012. **7**(12): p. e53339.
349. Bach, H., et al., *Escherichia coli maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies.* J Mol Biol, 2001. **312**(1): p. 79-93.
350. Raran-Kurussi, S. and D.S. Waugh, *The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated.* PLoS One, 2012. **7**(11): p. e49589.
351. Iwata, T., et al., *Parafibromin tumor suppressor enhances cell growth in the cells expressing SV40 large T antigen.* Oncogene, 2007. **26**(42): p. 6176-83.
352. Ali, S.H. and J.A. DeCaprio, *Cellular transformation by SV40 large T antigen: interaction with host proteins.* Semin Cancer Biol, 2001. **11**(1): p. 15-23.
353. Ahuja, D., M.T. Saenz-Robles, and J.M. Pipas, *SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation.* Oncogene, 2005. **24**(52): p. 7729-45.
354. Fanning, E. and R. Knippers, *Structure and function of simian virus 40 large tumor antigen.* Annu Rev Biochem, 1992. **61**: p. 55-85.
355. Li, D., et al., *Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen.* Nature, 2003. **423**(6939): p. 512-8.
356. Gai, D., et al., *Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen.* Cell, 2004. **119**(1): p. 47-60.

357. Lilyestrom, W., et al., *Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor.* Genes Dev, 2006. **20**(17): p. 2373-82.
358. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.
359. Stark, H., *GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM.* Methods Enzymol, 2010. **481**: p. 109-26.
360. Nguyen, V.Q., et al., *Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1.* Cell, 2013. **154**(6): p. 1220-31.
361. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.
362. Amlacher, S., et al., *Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile.* Cell, 2011. **146**(2): p. 277-89.
363. Hakulinen, N., et al., *Three-dimensional structures of thermophilic beta-1,4-xylanases from Chaetomium thermophilum and Nonomuraea flexuosa. Comparison of twelve xylanases in relation to their thermal stability.* Eur J Biochem, 2003. **270**(7): p. 1399-412.
364. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Macromolecular Crystallography, Pt A, 1997. **276**: p. 307-326.
365. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
366. Zucker, F., P.C. Champ, and E.A. Merritt, *Validation of crystallographic models containing TLS or other descriptions of anisotropy.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 8): p. 889-900.
367. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
368. Moriarty, N.W., R.W. Grosse-Kunstleve, and P.D. Adams, *electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1074-80.
369. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W375-83.
370. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 4): p. 849-61.
371. Terwilliger, T.C., et al., *Automated ligand fitting by core-fragment fitting and extension into density.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 8): p. 915-22.
372. Terwilliger, T.C., et al., *Ligand identification using electron-density map correlations.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 1): p. 101-7.
373. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.
374. Brooks, B.R., et al., *CHARMM: the biomolecular simulation program.* J Comput Chem, 2009. **30**(10): p. 1545-614.
375. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.

376. Best, R.B., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles.* J Chem Theory Comput, 2012. **8**(9): p. 3257-3273.
377. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
378. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.
379. Mankouri, H.W., H.P. Ngo, and I.D. Hickson, *Shu proteins promote the formation of homologous recombination intermediates that are processed by Sgs1-Rmi1-Top3.* Mol Biol Cell, 2007. **18**(10): p. 4062-73.
380. Bernstein, K.A., et al., *The Shu complex, which contains Rad51 paralogues, promotes DNA repair through inhibition of the Srs2 anti-recombinase.* Mol Biol Cell, 2011. **22**(9): p. 1599-607.
381. She, Z., et al., *Structural and SAXS analysis of the budding yeast SHU-complex proteins.* FEBS Lett, 2012. **586**(16): p. 2306-12.
382. Tao, Y., et al., *Structural analysis of Shu proteins reveals a DNA binding role essential for resisting damage.* J Biol Chem, 2012. **287**(24): p. 20231-9.