UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

hIws1 structural elucidation and protein interactions

Permalink

https://escholarship.org/uc/item/3t89h6bw

Author

Murthy, Nevin

Publication Date

2009

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

hIws1 Structural Elucidation and Protein Interactions

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Nevin Murthy

Committee in charge:

Professor Katherine A. Jones, Chair
Professor James T. Kadonaga, co-Chair
Professor Tracy Johnson

2009

©

Nevin Murthy, 2009

All rights reserved.

The Thesis of Nevin Murthy is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2009

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Sunnie M. Yoh for devoting her valuable time to teach me various laboratory skills and techniques. For that matter, I would like to extend my thanks to all the members of the RBIO-Jones laboratory for their guidance and support the past two years.

I would also like to thank my committee members, Dr. James T. Kadonaga (Co-Chair) and Dr. Tracy Johnson for exhibiting patience and understanding throughout this entire process. They were both kind enough to take the time and allow me to present my findings to them.

Finally, I wish to extend my gratitude to Dr. Katherine A. Jones for giving me the opportunity to work as a member in her laboratory and gain valuable research experience. She was very supportive and provided me with much needed guidance despite her overwhelming schedule.

ABSTRACT OF THE THESIS

hIws1 Structural Elucidation and Protein Interactions

by

Nevin Murthy

Master of Science in Biology

University of California, San Diego, 2009

Professor Katherine A. Jones, Chair
Professor James T. Kadonaga, Co-Chair

There are a number of nuclear factors that play an integral role in transcriptional elongation. Once such factor is hIws1, a nuclear protein first characterized to bind the histone H3 chaperone, Spt6. Our lab previously showed that Spt6 binds directly to the phosphorylated C-terminal tail of RNA polymerase II (RNAPII), and recruits Iws1 to

control specific events that occur during elongation, including mRNA processing, mRNA export, and histone H3K36 methylation. hIws1 interacts directly with the histone H3K36 methyltransferase, Hypb/Set2, as well as the mRNA export adaptor protein, REF1/Aly. Iws1 is reportedly an essential mammalian protein, but little is know about its structure or functional domains.

In the present study, I constructed GST tagged hIws1 protein constructs that were subsequently used in pulldown assays to evaluate its nuclear binding partner proteins. Pulldown assays were performed with either HeLa whole cell extract or HeLa nuclear extract. The N-terminal region of Iws1 (residues 1 through 495) was found to associate with the LEDGF/p75, a protein that has been previously identified to take part in lentiviral genome integration. In addition, I found that the C-terminal region of Iws (residues 522 through 819) binds to the E74-like factor 1 (Elf1). Interestingly, the hIws1 C-terminus contains a novel domain called the LW motif, which is highly conserved in Iws1 proteins from other species, and is also present in two transcription factors, Med26 and TFIIS. To test the role of the LW domain, I created three GST-Iws1 point mutant proteins, which were purified and used in GST-pulldown assays. Two residues (tyrosine-665 and tryptophan-685) contained within the LW motif were found to be crucial for the hIws1-Spt6 interaction. By contrast, Med26 was found to also associate with

Iws1, but this interaction was not mediated by the LW motif, and Iws1 also binds to the LEDGF factor through a region outside of the LW sequence. Thus these findings identify a new domain that is likely to be critical for binding of Iws1 to Spt6 and assembly of the RNAPII elongation complex

INTRODUCTION

Transcription is a vital process to cell viability and results in the production of messenger RNA (mRNA), which is subsequently used as a template, for gene expression and protein biosynthesis. Transcription elongation is the process by which RNA polymerase II (RNAPII) methodically moves along the template DNA strand, coding for mRNA. There are many steps and factors involved in mRNA synthesis that occur concomitantly with transcription elongation, Including pre-mRNA splicing and 3' end formation, release of transcripts from the template DNA, and export of mRNA from the nucleus to the cytoplasm. In addition, several histone modifications are tightly linked to ongoing transcription, including histone H3 Lys4 tri-methylation (H3K4me3), which is a mark of mRNA initiation, and promoter clearance, and H3 Lys36 tri-methylation (H3K36me3), which is found in the transcribed region of active genes and is associated with ongoing elongation.

The RNA CTD

Many of the factors involved in transcription elongation are localized to the C-terminal domain (CTD) of the large subunit of RNAPII. The C-terminal domain essentially works as a scaffold to bind a variety of nuclear factors that interact with nascent mRNAs and with the nucleosomal template. The CTD of RNAPII contains a repeated

heptapeptide consensus sequence and is highly conserved from fungi, plants and animals (Corden, 1990). The sequence itself is of the order YSPTSPS and is repeated 52 times in mammalian RNA polymerase. The CTD heptad repeat is differentially phosphorylated during the process of transcription (Meinhart et al. 2005; Chapman et al. 2007). In particular, phosphorylation of the heptad repeat at the Ser5 position is linked with promoter clearance and early transcription, whereas phosphorylation at the Ser2 position is found only on actively elongating RNA polymerase complexes.

Disrupting the heptad repeat can have lethal consequences at the cellular level. For example, deletion of two-thirds (reduction from 26 repeats to less than 10) or more of the sequence itself resulted in non-viable cells in yeast (Nonet et al. 1987; Zehring et al. 1988). It was also shown that each individual heptad must be consistent in its order of residues. Pairs of heptads can be separated via insertion of a different amino acid, and still exhibit wild-type functionality; however, mutations that alter even a single heptapeptide sequence can be lethal (Stiller and Cook 2004).

Certain genes require that a complete CTD be present for adequate gene transcription. In yeast cells, a removal of approximately 60% of the CTD resulted in lower expression of the *GALIO* gene, but did not

affect transcription of the *HIS4* gene (Scafe et al. 1990). Studies have shown that the CTD is obligatory for transcription elongation *in vivo*; however, *in vitro* studies have elucidated that transcription itself does not require the use of a fully functional CTD. Neither the RNA polymerases I or III nor the prokaryotic polymerase contains a domain resembling that of the RNAPII CTD (Dahmus et al. 1995), indicating that its function is strictly restricted to RNAPII elongation in eukaryotes and is not involved in processes such as rRNA biosynthesis.

General Transcription Factors

RNA polymerase II initiates transcription and does so through the aid of five general transcription factors (TFs): TFIIB, -D, -E, -F and –H (Roeder 1996; Ohkuma 1997). Formation of the pre-initiation complex begins with binding of TFIID to the TATA-containing element of the promoter region (the TATA element is a sequence consisting of repeating thymine and adenine nitrogenous bases) (Roeder 1996; Ohkuma 1997). The binding of TFIID allows TFIIB to associate with regions upstream and downstream of the TATA sequence as well as to RNAPII (Lagrange et al. 1998). Once bound to TFIIB, RNAPII can locate the initiation site and begin constructing RNA transcripts. Even before TFIIB binding, RNAPII is pre-bound to TFIIF, which aids in jump-starting PolII at the initiation site of transcription. TFIIE is involved in binding to all of the general transcription factors in addition to

RNAPII in order to stabilize the initiation complex (Ohkuma 1997). TFIIH is a multi-functional protein that acts as a CTD kinase, an ATPase and DNA helicase: all three of these functions are intricately involved in transcription elongation (Svejstrup et al. 1996).

Phosphorylation of the CTD

The CTD maintains a unique feature of existing in either nonphosphorylated or hyperphosphorylated forms. Analysis of the phosphorylated RNAPII CTD from mouse and HeLa cells has indicated that two serine residues in the heptapeptide repeat are the primary sites of phosphorylation.

After the formation of the initiation complex, the serine-5 residue is targeted for phosphorylation by the kinase TFIIH (Komarnitksy et al. 2000; Hengartner et al. 1998). These results indicate that an RNAPII complex located at initiation sites of genes displays a CTD heavily phosphorylated at Ser5 residues (Komarnitksy et a. 2000).

It was also gleaned that phosphorylation at serine-5 is both necessary and sufficient for recruiting the capping enzyme (CE) involved in adding a 5' cap to the nascent RNA transcript (Cho et al. 1998; Ho and Shuman 1999). Consistent with this finding is that TFIIH (the Ser5 kinase) associates with the capping enzyme and may also be critical to its recruitment (Rodriguez et al. 2000). Such processing mechanisms prevent

mRNA from being degraded upon its exit through nuclear pores. A "7-methyl G5_ppp_N" cap is added as the transcript exits from the RNAPII complex (it is approximately 25 base pairs long once the cap is added and still being actively transcribed) (Jove and Manley 1984; Rasmussen and Lis 1993). Studies were carried out with an RNAPII lacking a CTD and resulted in mRNA transcripts that were poorly capped (McCracken et al. 1997a).

The dimeric elongation factor DSIF (composed of the two human homologs Spt4 and Spt5) associates with NELF to form a complex that arrests transcription at a region slightly downstream of the promoter (this negative effect is eventually relieved through the action of P-TEFb, which catalyzes the phosphorylation of Ser2) (Yamaguchi et al. 1999). This association to the elongation complex is important in that it is also required for recruitment of CE through direct interaction with Spt5. Spt5 subsequently activates the GT domain on CE, which in turn leads to the negative action exerted by NELF (Mandal et al. 2004). This highly coordinated process is crucial in providing time in which capping can be efficiently performed (Sims et al. 2004). The CTD phosphatase Fcp1 is involved in releasing the capping enzyme from the elongation complex by dephosphorylating the Ser5 residue as elongation progresses (Komarnitsky et al.2000; Schroeder et al. 2000).

Recruitment of the histone methyltransferase, Set1, is also dependent upon Ser5 phosphorylation. Set1 functions primarily at initiation sites of actively transcribed genes and works to tri-methylate histone H3 at lysine residue 4 (H3K4me3). As transcription elongation progresses, H3K4me3 levels gradually decrease along with CTD Ser5 phosphorylation.

Modifications also take place at the 3' end of genes. Factors involved in processes such as polyadenylation also bind to a phosphorylated CTD in vitro (McCracken et al. 1997b; Birse et al. 1998). For example, the factor CFIA cross-links at the 3' end of genes and is dependent on Ser2 phosphorylation (Licatalosi et al. 2002; Ahn et al. 2004).

During elongation, serine-2 is the primary target for phosphorylation by the kinase P-TEFb (CTDK-I in yeast) (Marshall et al. 1996; Lee and Greenleaf 1997; Prelich 2002). RNAPII located at regions of actively transcribed genes display significantly more Ser2P than Ser5P. Interestingly, the yeast homolog of P-TEFb, CTDK-I, preferentially phosphorylates Ser2 residues located in repeats that already contain previously Ser5P residues (Jones et al. 2004). Ser2P recruits another methyltransferase, HYPB/Set2, which mediates H3K36me3 in the

transcribed region of active genes, as well as enzymes involved in polyadenylation of the nascent mRNA, such as CFIA.

Set2/HYPB binds to Ser2P residues and is found at internal sites along actively transcribed units (Li et al. 2002, 2003; Krogan et al. 2003; Schaft et al. 2003; Xiao et al. 2003). Yeast and mammalian Set2 enzymes mediate H3K36me3 in a manner that depends upon P-TEFb/CTDK-1 (Krogan et al 2003; Xiao et al. 2003), and the Set2 enzyme binds directly to the Ser2P CTD. However H3K36me3 is not itself required for transcription elongation, and studies in yeast have suggested that it instead serves to reset the normal chromatin structure to ensure that repeated transcription does not disrupt the nucleosome structure within transcribed regions, to prevent cryptic or aberrant transcription from arising at long genes, and to mark regions of the genome that are highly active.

Experiments have also shown that Ser2P plays an important role in 3' mRNA processing (Yamamoto et al. 2001; Ahn et al. 2004). Inactivating the P-TEFb kinase, Cdk9, in Drosophila leads to defective 3' end processing; although transcription elongation is unaffected at some genes. These same studies also showed that cells lacking P-TEFb fail to cross-link 3' end processing factors to the CTD (Ahn et al. 2004). Transcriptional termination is coupled with processing of the 3' end of the nascent mRNA (Hirose and Ohkuma 2007). Addition of the poly(A) tail is a

required signal for termination (Hirose and Manley 2000). This was subsequently verified from studies that showed that mRNAs transcribed by CTD-truncated Pol II were inadequately terminated (McCracken et al. 1997b).

The SPT Proteins

The CTD's main function is to serve as a scaffold upon which a multitude of nuclear factors can bind to and manipulate transcription in different ways. These factors bind to the CTD based upon the phosphorylation patters elucidated. We know that CTD phosphorylation changes as RNAPII moves along a transcriptionally active gene, which means that the factors bound to the CTD might also change as transcription plays out its course. This binding and interaction of these diverse complexes control events such as mRNA 5' and 3' end processing that are coordinated with phosphorylation of the RNAPII CTD.

Other factors involved in the process of transcription elongation include the Spt family of proteins. These factors are conserved across most eukaryotic species and were characterized from studies conducted with yeast (Winston et al. 1984). They include the proteins Spt4, Spt5, and Spt6. While it is apparent that Spt4 and Spt5 are closely associated to form a complex (DSIF), they only weakly associate with Spt6 (Hartzog et al. 1998, Wada et al. 1998). All three of these factors have a role in

transcription elongation in both yeast and humans (Hartzog et al. 1998, Wada et al. 1998). Studies have shown that both Spt6 and Spt5 localize to areas of actively transcribed genes in eukaryotic cells (Kaplan et al. 2000; Andrulis et al. 2000). For example, the heat shock (hsp) gene in *Drosophila melanogaster* exhibits a high recruitment of both Spt5 and Spt6 upon induction, indicating a positive regulatory role by both Spt5 and Spt6 in transcription (Winston 2001). In another instance, only Spt5 was localized to the hsp70 locus in uninduced cells, indicating that it may also exhibit a negative regulatory role (Kaplan et al. 2000; Andrulis et al. 2000). Spt6 has been shown act as an essential transcription factor for many actively transcribed genes, but also, importantly, has been found to function as a histone H3 chaperone. Spt6 has a specific role in nucleosome reassembly during transcription elongation (Adelman et. al. 2006).

A role for Spt6 and Iws1 in the function of the Ser2P RNAPII CTD

Our lab recently discovered that Spt6 associates selectively with the P-TEFb/Ser2P form of RNAPII in extracts, and binds directly to the Ser2P CTD through its SH2 domain (Yoh et al. 2007). Whereas Spt6 is critical for transcription elongation, a point mutation in the SH2 domain (R1358K) that destroys the ability of Spt6 to bind to the Ser2P CTD does not affect transcription elongation, either in vivo or in vitro. Thus Spt6 must be recruited to RNA polymerase and stimulate elongation independently of

it's binding to the Ser2P C-terminal domain of RNAPII. This observation raises the question of what role the binding of Spt6 to the CTD may play in gene expression, and in particular in the activities associated with elongation, such as pre-mRNA processing and histone methylation. Interestingly, the R1358K point mutation in the SH2 domain of Spt6 resulted in nascent mRNAs that contained partial splicing defects (Yoh et al. 2007).

Previous biochemical studies in yeast have shown that Spt6 cross-links with another protein, Iws1 ("interacts-with-Spt6") (Krogan, et al. 2002), an interaction that persists throughout transcription elongation. Mutations in the gene encoding the Iws1 protein results in cells with a phenotype similar to cells with an Spt6 mutation (Lindstrom et al. 2003). Moreover, chromatin immunoprecipitation studies indicate that yeast Iws1 is required for stable binding of Spt6 to the RNA polymerase complex. However, this function of Iws1 is not conserved in mammalian cells, and knockdown of Iws1 by RNAi does not alter the binding of Spt6 with the transcribed region of target genes (Yoh et al. 2008). Additional studies in yeast have shown that Iws1 also associates with RNAPII, similar to the factors Spt4, Spt5, and Spt6 (Krogan et al. 2002) and interacts with the RNAPII elongation factor TFIIS. Iws1 is evolutionary conserved, and a protein alignment analysis has demonstrated that the human and yeast forms share 32% amino acid similarity.

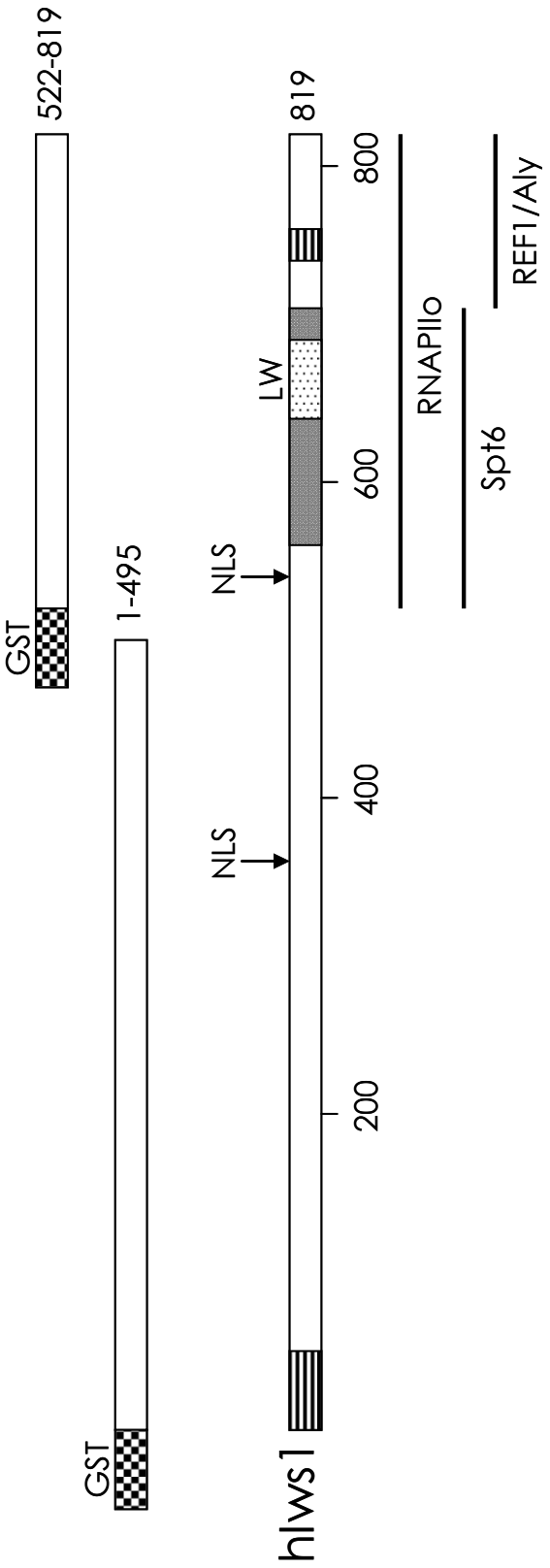
Importantly, knockdown of Iws1 from HeLa cells revealed several consequences for the expression of Spt6 target genes such as the c-Myc gene (Yoh et al. 2007, 2008). Like the R1358K mutant of Spt6, loss of Iws1 is associated with the accumulation of partially spliced mRNAs. In addition, ChIP experiments reveal that H3K36me3 is lost across the coding region of Spt6 target genes in Iws1 RNAi-treated cells. This is associated with a failure to recruit Hypb/Set2, the enzyme that mediates H3K36me3. Because Hypb/Set2 is itself a CTD-binding protein, these findings reveal that stable binding of Set2 to the RNAPII elongation complex depends both upon its intrinsic binding to the RNAPII CTD as well as to its interaction with the Iws1 protein. This interaction is direct and mediated by a region of Iws1 that is largely conserved with the yeast Iws1 homolog. Interestingly, the interaction of Iws1 with Spt6 is strongly enhanced in vitro by including the Ser2P RNAPII in the reaction, which indicates that Iws1 binds tightly to Spt6 only after it has bound to the RNAPII elongation complex. These findings raise the immediate question of how Iws1 associates with Spt6, whether this interaction is regulated, and what other proteins it may bring to the RNAPII elongation complex. In this study, I have analyzed the structure and function of Iws1 and identify a new highly conserved motif that mediates binding of Iws1 to Spt6 and the RNAPII elongation complex.

CHAPTER 1: STRUCTURAL ANALYSIS OF hIWS1

The human Iws (Interacts With Spt6) protein is an 819 amino acid polypeptide that has been identified and characterized as a factor involved in transcription elongation in yeast. The full-length protein is around 90 kilodaltons in size and is known to function primarily in the nucleus. Alignment analysis illustrates that the human Iws1 (hIws1) and the yeast (*Saccharomyces cerevisiae*) Iws1 (yIws1) proteins share approximately 32% amino acid identity across the C-terminal region, whereas the human and *Drosophila* homologs share an overall 48% amino acid similarity. The yIws1 shares over 85% similarity (based solely on primary structure) across other species such as *Mus musculus*, *Rattus norvegicus*, and *Monodelphis domestica*.

The metazoan Iws1 proteins contain a unique N-terminal region of approximately 500 residues that is not found in yeast Iws1. This region is highly acidic and contains a long stretch of predicted phosphorylation sites for the casein kinase protein family, especially CKII. Casein kinases are serine/threonine selective enzymes involved in regulating signal-transduction pathways, including Wnt signaling, cell-cycle regulation, and DNA repair. CKII is tightly associated with several complexes that are known to bind RNA polymerase II, although the exact function of CKII in elongation is unknown. There also exists a well-conserved sequence in

Figure 1.1: Schematic representations of GST-tagged (checkered boxes) polypeptides (numbered according to amino acid residues) and the basic structural features of the hIws1 protein. The location of NLS signals are indicated by arrows. The LW motif is indicated with a dotted-box. The region containing a highly conserved sequence in both yeast and human, including the LW motif, is showed via a shaded box. Two other well-conserved regions that flank a substantial portion of the protein and contain highly acidic residues are indicated by dashed boxes. The regions to which other factors such as Spt6, Aly, and RNAPIIo bind are indicated by solid lines.


522-819
GST
1-495
GST
NLS
NLS
LW
819
800
600
400
200
RNAPIIo
Spt6
REF1/Aly
hlws1

both human and yeast Iws1, consisting of highly acidic residues indicated by a dashed box in figure 1.1. A similar region also exits in the C-terminal portion of the protein, also indicated by a dashed box (figure 1.1), and Iws1 contains two distinct nuclear localization (NLS) domains. The NLS domains are apparent in the central portion of the protein. These signals have been elucidated to be of the order $KK(X)_{11}KKQK$ at position 358 and KRRR at position 542 (figure 1.1) (Ling et al. 2006).

The C-terminal region of approximately 250 amino acids is highly conserved in both yeast and humans (indicated by shaded box in figure 1.1). Recent bioinformatic studies have also indicated that this region contains a novel domain described as the LW motif (figure 1.1). An LW motif exists in the highly conserved N-terminal region of the transcription elongation factor TFIIS; however, this region has not been identified to have any catalytic activity. TFIIS does not possess an NLS and only recombinant proteins containing the LW motif are efficiently targeted to the nucleus (Ling et al. 2006). A similar motif was also identified in the N-terminal region of the two other nuclear factors, elongin A and MED 26 (Booth et al. 2000). The sequences containing the LW motifs in TFIIS, elongin A, MED 26, and hIws1 contain many identical or functionally similar residues (shaded in figure 1.2). However the LW domain does not control the nuclear localization of Iws1, and its function is unknown.

Figure 1.2: Sequence analysis of hlws1. (A) Full-length hlws1 sequence with the LW motif shaded in red. (B) Comparison of LW motif sequence between TFIIS, elonginA, Med26, and hlws1. Amino acids with structural or functional similarities are shaded in grey.

A

Full-Length hIws1

MDSEYYSGDQ SDDGGATPVQ DERDSGSDGE DDVNEQHSGS DTGSVERHSE NETSDREDGL
PKGHHVTDSE NDEPLNLNAS DSESEELHRQ KDSDSESEER AEPPASDSEN EDVNQHGSDS
ESEETRKLPG SDSENEELLN GHASDSENED VGKHPASDSE IEELQKSPAS DSETEDALKP
QISDSESEEP PRHQASDSEN EEPPKPRMSD SESEELPKPQ VSDSESEEPP RHQASDSENE
ELPKPRISDS ESEDPPRHQA SDSENEELPK PRISDSESED PPRNQASDSE NEELPKPRVS
DSESEGPQKG PASDSETEDA SRHKQKPESD DDSDRENKGE DTEMQNDSFH SDSHMDRKKF
HSSDSEEEEH KKQKMDSDED EKEGEEEKVA KRKAAVLSDS EDEEKASAKK SRVVSDADDS
DSDAVSDKSG KREKTIASDS EEEAGKELSD KKNEEKDLFG SDSESGNEEE NLIADIFGES
GDEEEEEFTG FNQEDLEEEK GETQVKEAED SDSDDNIKRG KHMDFLSDFE MMLQRKKSMS
GKRRRNRDGG TFISDADDVV SAMIVKMNEA AEEDRQLNNQ KKPALKKLTL LPAVVMHLKK
QDLKETFIDS GVMSAIKEWL SPLPDRSLPA LKIREELLKI LQELPSVSQE TLKHSGIGRA
VMYLYKHPKE SRSNKDMAGK LINEWSRPIF GLTSNYKGMT REEREQRDLE QMPQRRRMNS
TGGQTPRRDL EKVLTGEEKA LRPGDPGFCA RARVPMPSNK DYVVRPKWNV EMESSRFQAT
SKKGISRLDK QMRKFTDIRK KSRSAHAVKI SIEGNKMPL

B

It is interesting to note that the LW motifs of TFIIS, elongin A, and MED 26 are contained within the N-terminal regions of each protein, whereas in Iws, the LW motif is located within its C-terminal half. Furthermore, nuclear targeting of Iws is dependent only on its nuclear localization signals and not the LW motif. Constructs of tagged Iws containing only an NLS sequence proved to be sufficient for nuclear targeting (Ling et al. 2006). Protein-protein interaction assays also showed that the region of TFIIS containing the LW motif could associate with full-length hIws1 (Ling et al. 2006), which raises the possibility that LWs are self-associating domains that could mediate formation of homo- or heterodimeric complexes.

The conserved C-terminal domain of Iws mediates its binding to a variety of other nuclear factors, including Spt6, REF1/Aly, and the Ser2P RNAPII complex (figure 1.1). REF1/Aly, a factor implicated in RNA splicing and export, associates with a region of Iws that does not contain the LW motif. Spt6, a histone H3 chromatin remodeling protein associates with a region of Iws that does overlap with the LW motif; however it has yet to be determined if the LW region mediates the Iws-Spt6 interaction. Previous biochemical studies have indicated that RNAPII binds to Iws1 indirectly, through Spt6 (Yoh et al. 2007).

To address the nature of the interaction of Iws1 with its various partner proteins, I designed vectors to express different GST-Iws fusion

protein constructs to be used in GST-pulldown assays from HeLa nuclear extracts (figure 1.1). GST-tagged N-terminal construct, containing one of the NLS sequences, was designed and utilized to assess the binding capabilities of the N-terminal region of Iws1. A separate construct spanning the region between residues 522 and 819 was constructed, containing the other NLS sequence and the LW motif to determine the binding capabilities of the C-terminal region of Iws1.

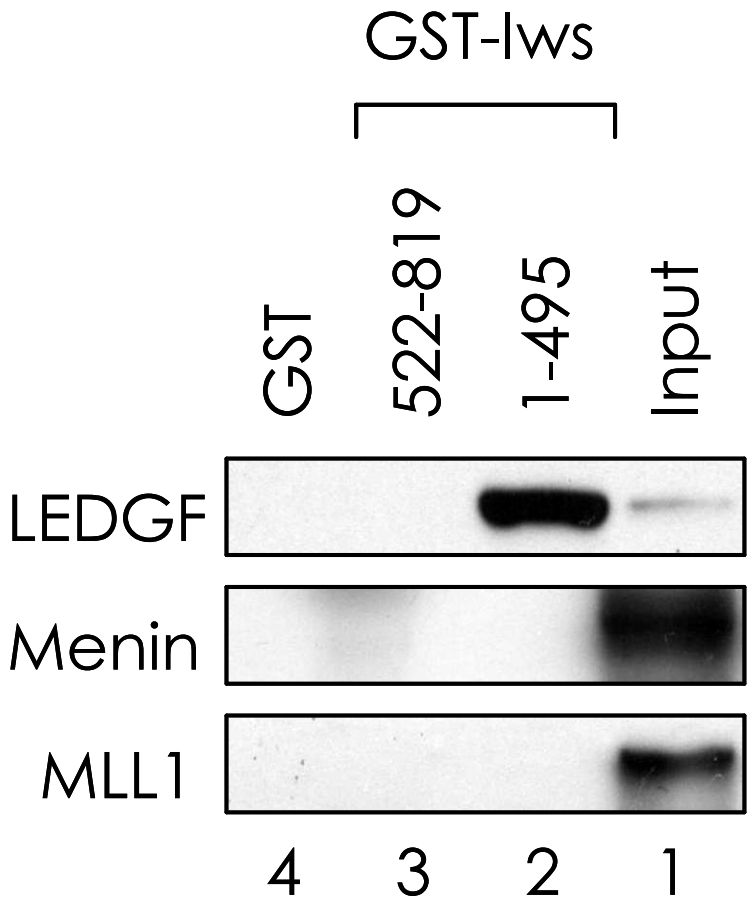
CHAPTER 2: Iws1 PROTEIN-PROTEIN INTERACTIONS

A preliminary in vitro pulldown assay was performed using only the GST-Iws$_{1-495}$ GST-Iws$_{522-819}$ fusion proteins to test for novel protein-protein interactions. The recombinant fusion proteins were purified from E. coli and then immobilized to glutathione beads, which were incubated with pre-cleared whole cell extracts from HeLa cells.

2.1: Iws Interacts with LEDGF/p75

Preliminary data from our lab indicate that Iws1 associates with a nuclear factor designated LEDGF (Yoh et al., unpublished data). LEDGF (Lens epithelium-derived growth factor) was originally isolated and identified as a general transcription activator with identical features to that of a previously characterized protein, p75. A transcriptional co-activator, p75 has been shown to interact with components of the basal transcriptional machinery. LEDGF/p75 has been suggested to increase the survivability of a cell following stress-induced apoptosis events. It binds to promoter elements of heat-shock and other stress-related proteins, altering their expression pattern. Heat shock proteins work antagonistically to factors that are involved in inducing apoptosis and so prolong the cell's lifetime. LEDGF/p75 has also been characterized as an integrase-interacting protein involved in integration of the lentiviral genome, specifically HIV-1. In infected cells lacking endogenous

Figure 2.1: Western blot analysis of the physical association between hlws1 and LEDGF after a GST-pulldown assay using GST-lws bait proteins (and GST control) incubated with HeLa nuclear extracts. Input lane consists of just the nuclear extract. Pulldown assays were separated by a 6% SDS-PAGE, and Western blot assay was performed with anti-LEDGF, anti-Menin, or anti-MLL1 antisera as indicated.


GST-Iws
GST
522-819
1-495
Input
LEDGF
Menin
MLL1
4
3
2
1

LEDGF/p75 protein, HIV-1 genome integration is significantly compromised. Lentiviruses have a tendency to integrate their genome at active transcription units, a process dependent on the presence of LEDGF/p75. Studies have shown that cells devoid of LEDGF/p75 results in integration in non-transcriptional units (Engelman et al. 2008). A nuclear protein, LEDGF/p75 tethers the HIV-1 integrase protein to chromatin and subsequently protects it from degradation. The ability of LEDGF/p75 to direct viral genome integration at active transcription units suggests that it should show some interaction with the transcriptional machinery. Consequently, it is interesting to consider that LEDGF may direct the viral integration complex to the RNAPII elongation complex for integration at active genes. As shown in figure 2.1, the HeLa nuclear LEDGF protein can interact directly with the N-terminal region of GST-Iws1. This interaction suggests that Iws may be involved in either regulating or directing the function of LEDGF in some manner, and raises the possibility that LEDGF is part of the Spt6:Iws1 complex on the RNAPII CTD.

A fraction of nuclear LEDGF associates with a tumor suppressor called Menin, which in turn binds to the H3K4me3-specific methyltransferase protein, MLL1 (Yokoyama et al. 2008). Cells lacking menin have shown a dramatic increase in cell proliferation. Given its interaction with MLL1, there is increasing evidence suggesting that menin

is a crucial regulator of histone modifying proteins by acting as a scaffold to coordinate transcription and proliferation. It has also been shown that Menin is carried along by the transcriptional machinery in regions of actively transcribed genes. To investigate whether MLL1 and Menin might also be part of the Spt6:Iws1 complex on the RNAPII CTD, I tested whether these factors were also present with LEDGF in the GST-Iws1 pulldown fraction. Interestingly, GST-Iws1 does not interact with nuclear Menin (figure 2.1). Similarly MLL1, which is the human homolog of the yeast Set1 methyltransferase, was also not present in western blots of the GST-Iws1 pulldown reaction. Thus, Iws associates with the H3K36 methyltransferase HYPB/Set2, but does not bind to MLL1 and Menin (figure 2.1). MLL1 is involved in histone modification primarily during early elongation near initiation sites, suggesting that Iws may just be involved in histone modification during mid to late phase elongation. Consistent with this possibility, knockdown of Iws1 in HeLa cells affects H3K36me3 at target genes, but does not affect H3K4me3 at the same genes. These findings also indicate that LEDGF may play a role in gene expression or histone modification that is separate from its association with MLL1 and Menin, and further studies are needed to assess whether its association with the Spt6:Iws1:CTD complex might target the HIV-1 integrase protein to actively transcribed cellular genes for proviral integration.

2.2: Iws1 interacts with Spt6 and the Ser2P RNAPII CTD

The key functional interaction of Iws1 is with the transcription elongation factor and histone H3 chaperone, Spt6 (Yoh et al. 2007). Spt6 is a factor that has been characterized in yeast, *Caenorhabditis elegans*, and many mammalian species, indicating its conserved nature (Hartzog et al. 1998). It has been demonstrated that Spt6 acts as a histone H3 chaperone and works to reassemble nucleosomes at promoters after transcription (Bortvin et al. 1996). Interestingly, loss of Spt6 function in yeast so disrupts the local chromatin structure at endogenous promoters that transcription is possible even in the absence of any DNA-bound activators (Adkins and Tyler 2006). A protein roughly twice the size of Iws, Spt6 consists of an N-terminal region with an abundance of acidic amino acids and a C-terminus consisting of an SH2 domain (Yoh et al. 2007). Spt6 associates with the CTD of RNAPII via its SH2 domain (Yoh et al. 2007). In principle, based on genetic studies in yeast, Spt6 might also be expected to interact with components of the DRB-sensitivity-inducing factor (DSIF), which is composed of Spt4 and Spt5. Given the many factors that Spt6 could in principle associate with, I decided to probe for nuclear proteins that associate with Iws1 and assess whether any of these factors might be linked to Spt6 indirectly, via association with Iws1.

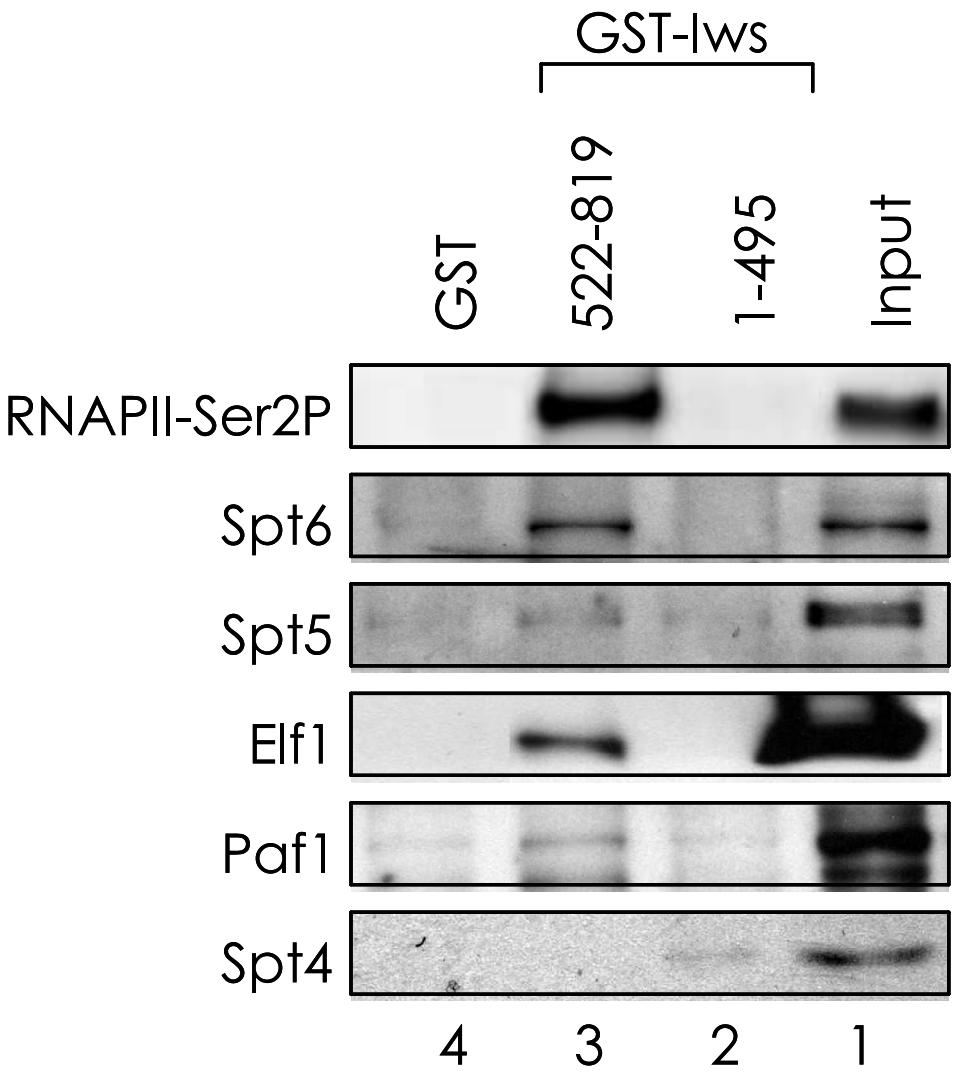
DSIF is a heterodimer composed of the two factors Sp5 and Spt4 and plays roles in both stimulating and repressing transcription. Following transcriptional initiation, RNA polymerase II pauses around 30 nucleotides downstream of the start site (Wada et al. 1998) upon binding of DSIF and another factor, NELF. The CTD kinase and positive transcription elongation factor P-TEFb then phosphorylates the CTD of RNAPII to relieve this arrest in elongation, mediated in part by releasing NELF from the complex (Wada et al. 1998b; Yamaguchi et al. 1999a). P-TEFb also phosphorylates the Spt5 subunit of DSIF, which leads to further activation of elongation (Ivanov et al. 2000). Like Iws, DSIF may also be involved in cellular processing of mRNAs (Sehgal et al. 1976). The repressive activity of DSIF is only apparent in the presence of another protein known as DRB, thus coining the name DRB-sensitivity inducing factor. Importantly, DSIF can also *stimulate* transcription elongation by increasing the ability of PolII to bypass pause sites in the production of long mRNA transcripts (Guo et al. 2000), and thus it plays roles both in pausing and productive elongation. In short, Spt4 and Spt5 are primarily involved in regulating the rate of transcription elongation and therefore can be considered part of the "rate-limiting" step in this overall process.

Iws1 was recently shown to be involved in co-transcriptional processing of mRNAs, in particular splicing, and in facilitating the kinetics

of bulk poly(A)+ mRNA export. Previous studies showed that members of the Paf1 complex are also involved in late elongation events such as pre-mRNA processing (Penheiter et al. 2005; Sheldon et al. 2005). The Paf1 complex is composed of the factors Paf1, Ctr9, Cdc73, Rtf1, and Leo1. The members of the Paf1 complex have been shown to interact with other transcription-related factors including Spt4, Spt5 and FACT (Krogan et al. 2002; Squazzo et al. 2002). In yeast, members of the Paf1 complex are also implicated in regulating phosphorylation of the RNAPII CTD. Deletion of Paf1 or Rtf1 results in a dramatic decrease in Ser2P RNAPII within at coding regions of actively-transcribed genes (Mueller et al. 2004). These studies implicated Paf1 in pre-mRNA processing as well as control of the length of polyadenylated tails of mRNAs.

Using the biochemical GST-pulldown assay, I confirmed that the C-terminal region of Iws1 binds Spt6 (Figure 2.2). This same C-terminal region of Iws1 also pulls out Ser2P RNAPII, whereas the N-terminal region does not. However, my experiments indicated that Iws1 does not interact strongly with either Spt4 or Spt5. This suggests that it may not act as a regulator in pausing/stimulating transcription elongation associated with the DSIF complex, which is consistent with the fact that Iws1 has no effect on transcription initiation or elongation in vitro (Yoh et al. 2007). Moreover,

Figure 2.2: Western blot analysis of the physical association between hIws1 and other nuclear factors after a GST-pulldown assay using GST-Iws bait proteins (and GST control) incubated with HeLa whole cell extracts. Input lane consists of just the cellular extract. Pulldown assays were separated by a 6% SDS-PAGE, and Western blot assay was performed with a-H5, a-Spt6, a-Spt5, a-Elf1, a-Paf1, or a-Spt4 antisera as indicated.


GST-Iws
GST
522-819
1-495
Input
RNAPII-Ser2P
Spt6
Spt5
Elf1
Paf1
Spt4
4
3
2
1

although it shares some similar properties to those factors of the Paf1 complex, my studies further indicate that Iws does not directly interact with Paf1 (Figure 2.2). Iws1 is known to recruit REF/Aly to actively transcribed regions and to facilitate mRNA export from the nucleus; however these studies indicate that this activity is independent of the role of the endogenous Paf1 or Leo1 proteins in these same events.

2.3: Iws associates with Elf1

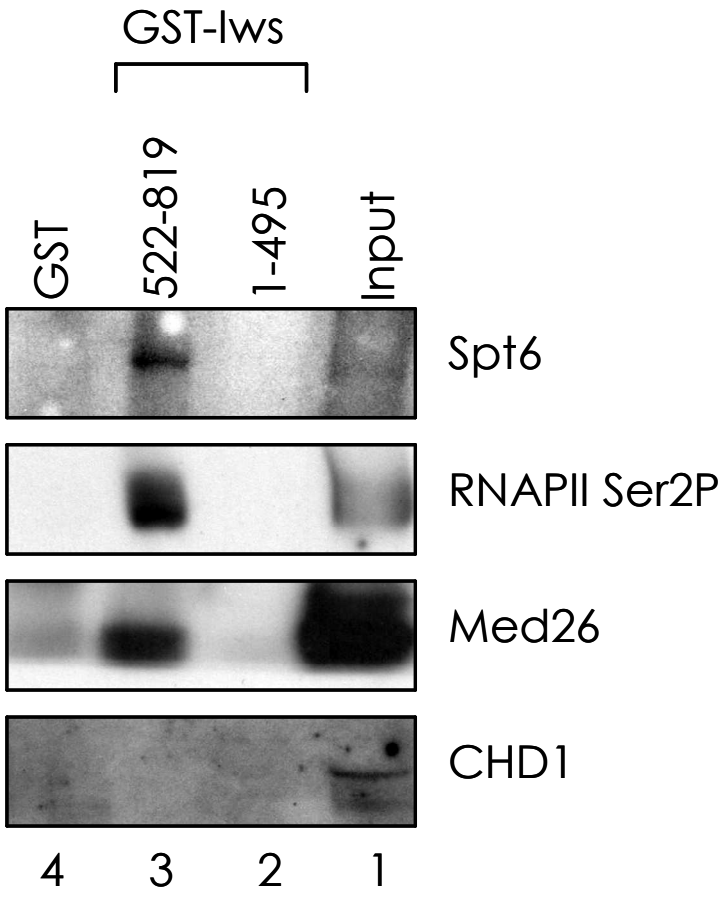
My initial tests also revealed an unexpected interaction between HeLa Iws1 and the factor Elf1 (E74 like factor 1) (Figure 2.2). Elf1 is member of the Ets transcription factor family found in humans, mice, *C. elegans* and *Drosophila*. Most of the members of the ETS transcription factors contain a highly conserved DNA binding domain known as the ETS domain which binds to regions of DNA with a central GGA sequence. More specifically, most Ets proteins bind to specific DNA sequences centered over either GGAA or GGAT motifs. However, Elf1 and another related Ets protein, E74, bind only GGAA motifs. Elf1exists primarily as a 98-kDa form in the nucleus and a 68-kDa form in the cytoplasm. Because Elf1 is a DNA-binding factor, it is not clear why it should associate with Iws1. Perhaps Elf1 recruits Iws1 to target genes prior to its association with Spt6 on the RNAPII CTD, or for functions that are distinct from that of the Spt6:CTD elongation complex. Consequently, I did not further pursue the

mapping of the Elf1 interaction with Iws1, nor is it clear that this is a direct association, and therefore further studies will be needed to assess the relevance of this finding.

2.4: Iws1 associates with Med26, a subunit of the Mediator complex

Studies in yeast have suggested that Iws1 associates with TFIIS, a protein that is an integral part of the elongation complex. TFIIS is a general transcription factor that increases the overall transcription rate of RNA polymerase II by activating transcription elongation complexes that have arrested transcription (Reines, D. 1994). The protein is conserved from yeast to humans while homologs have been identified in Archaea and some viral genomes. TFIIS contains a domain that interacts directly with RNA polymerase II as well as a region that enhances cleavage of the nascent transcript to allow backtracking of the paused RNAPII complex, and both are conserved from yeast to humans. A different, N-terminal domain was recently found to contain a novel LW motif, which is also present in several other nuclear factors, notably Med26 (also called CRSP70) (Ryu, S. et al. 1999) and Elongin A. Med26 is an essential subunit of the CRSP complex, contained in the central core of the mediator complex, which is required for the activity of the enhancer-binding protein Sp1 as well as many other activators(Pan, G. et al. 1997). Whether Med26

Figure 2.3: Western blot analysis of the physical association between hIws1 and CRSP70/Med26 after a GST-pulldown assay using GST-Iws bait proteins (and GST control) incubated with HeLa nuclear cell extracts. Input lane consists of just the nuclear extract (lane 1). Lanes 2 and 3 consist of the pulldown reactions using GST-Iws$_{1-495}$ and GST-Iws$_{522-819}$, respectively. Lane 4 consists of a pulldown reaction using GST alone. Reactions were separated by an 8% SDS-PAGE, and Western blot assay was performed with a-H5, a-Spt6, a-CRSP70, or a-CHD1 antisera as indicated.


GST-Iws
GST
522-819
1-495
Input
Spt6
RNAPII Ser2P
Med26
CHD1
4
3
2
1

is also carried along with the RNAPII elongation complex and might have other roles in gene expression is unknown at present. Although Med26 and TFIIS share homology in their N-terminal region, the function of these regions has yet to be elucidated for either protein. This sequence was later characterized as an LW motif (Ling, Y. et al. 2006) in TFIIS and also shown to share homology with a C-terminal region of hIws1. By introducing point mutations in the LW motif, it was determined that such a region was crucial for nuclear localization in TFIIS (Ling, Y. et. al. 2006) and might possibly maintain a similar function in Med26. Interestingly, the LW motif found in hIws1 does not appear to be involved in nuclear targeting (Ling, Y. et al. 2006). Further experiments also revealed that the TFIIS can associate with hIws1 in nuclear extracts from HeLa cells, and that this association is mediated through the LW domain. Studies are underway in the lab to assess whether or not the human TFIIS protein is carried with the RNAPII elongation complex via association with the Spt6:Iws1 complex.

By means of GST-pulldown assays, I observed that hIws1 associates with a fraction of Med26 present in HeLa extracts. As shown in figure 2.3, I determined that the C-terminal domain of Iws1, GST-Iws$_{522-819}$, interacts with Med26, whereas the N-terminal domain fragment, GST-Iws$_{1-495}$, does not. Relative to the input, the ability of Iws to bind to Med26 is not as strong as its ability to bind Spt6 or RNAPII Ser2P. This may indicate that the

Iws-Med26 interaction might be weak or indirect, i.e., mediated through another factor. Interestingly, Med26 was previously suggested to interact directly with the RNA polymerase II CTD (Naar, A. et al. 2006). Thus, one possibility is that Med26 may be present in the pulldown fraction simply as a part of the Ser2P RNAPII elongation complex. If so, it's binding to the CTD might depend upon Spt6 and Iws1, as was suggested for Hypb/Set2, or it may bind independently of the Spt6:Iws1 complex and be carried along with the RNAPII. Because the GST-Iws$_{522-819}$ construct contains the LW motif, it is alternately possible that Med26 binds directly to the LW motif of Iws1, and might in principle even compete for binding of Iws1 to Spt6 and the RNAPII CTD.

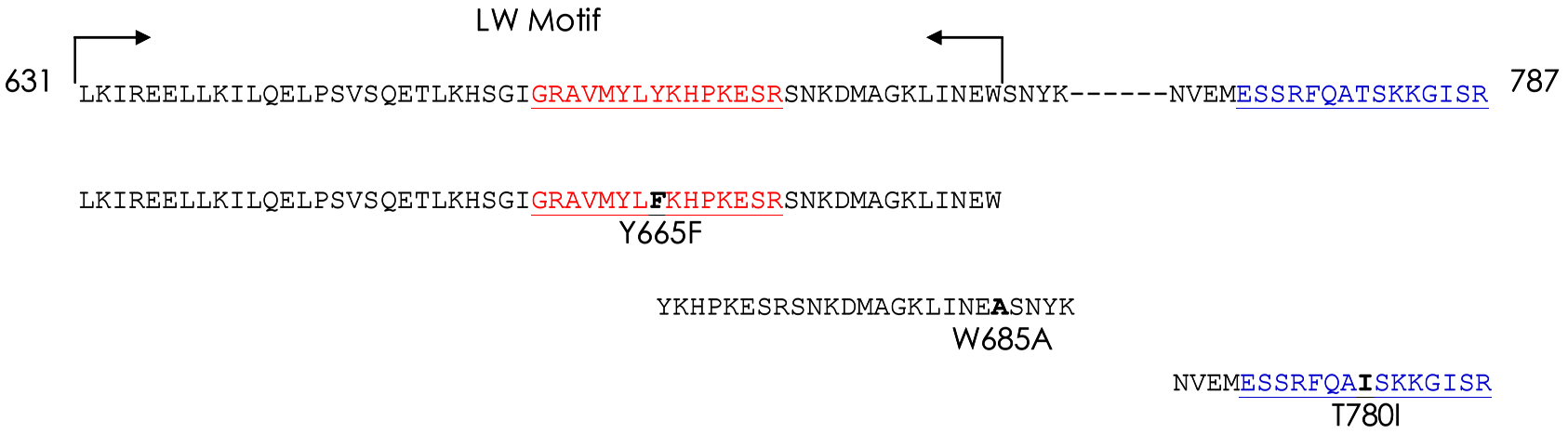
Chapter 3: Iws1 MUTANT PROTEIN-PROTEIN INTERACTIONS

3.1: Iws1 Mutant Proteins

Having determined that Iws1 interacts with Spt6, RNAPII and Med26 through its conserved C-terminal domain, mutant GST-Iws constructs were designed to identify important residues within the LW motif itself that might mediate the binding of Iws1 to one or more of these proteins. Ling et al. already showed that TFIIS interacts with Iws1 through the LW motif. In order to determine whether the hIws1 LW motif directs binding to other factors, I created GST expression vectors to introduce point mutations within the Iws1 LW motif, and tested the activity of these mutant proteins in GST pulldown reactions.

From the motif analysis of Iws1 using the MIT scansite menu, it was gleaned that there exists possible binding sites at tyrosine (Y) residue 665 and threonine (T) residue 780. Tyrosine-665 is located within the LW motif region (figure 3.1a) while threonine-780 is located outside the LW sequence. From figure 3.1, the red-shaded and blue-shaded sequences indicate the regions to which other factors may associate with. A tryptophan (W) residue, which is a signature amino acid within the LW sequence itself, is located at position 685 and may also prove to be critical in protein-protein interactions. Three separate mutations were introduced to each of these sites: the tyrosine was mutated to a

Figure 3.1: Schematic representation of the GST-Iws1 mutant proteins constructed to test for functionality of important residues. The top panel shows the amino acid sequence between residues 631 and 787. The LW motif boundaries are indicated by arrows on either side. Structural analysis using the MIT scansite menu indicates that there exists an SH2 domain binding site (the region shaded in red). A PKC phosphorylation site also exists (shaded in blue). Three mutant proteins were designed. A tyrosine residue (Y) at position 665 was mutated to a phenylalanine (F). A tryptophan residue (W) at position 685 was mutated to an alanine (A). Finally, a threonine (T) residue at position 780 was mutated to an isoleucine (I). Dashed lines indicate omitted regions of the sequence


LW Motif
631
LKIREELLKILQELPSVSQETLKHSGIGRAVMYLYKHPKESRSNKDMAGKLINEWSNYK------NVEMESSRFQATSKKGISR
787
LKIREELLKILQELPSVSQETLKHSGIGRAVMYLFKHPKESRSNKDMAGKLINEW
Y665F
YKHPKESRSNKDMAGKLINEASNYK
W685A
NVEMESSRFQAISKKGISR
T780I

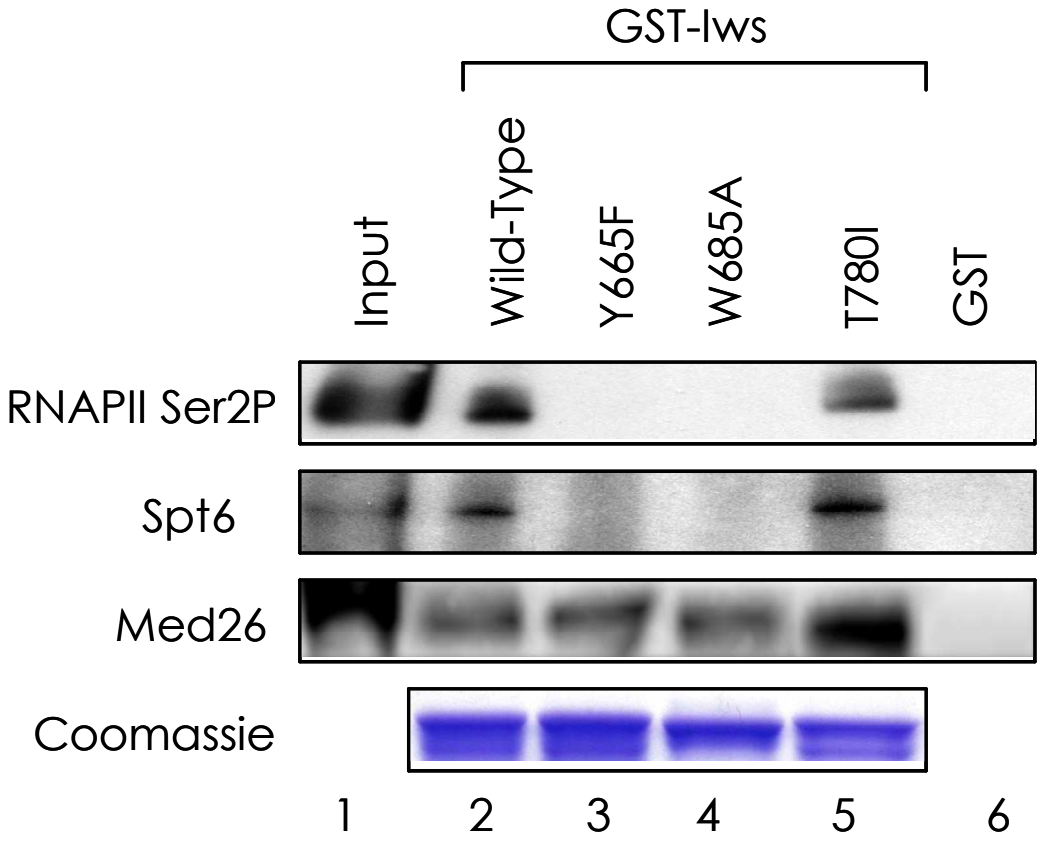
phenylalanine (F); tryptophan-685 was mutated to an alanine (A); finally, threonine-780 was mutated to an alanine (A); finally, threonine-780 was mutated to an isoleucine (I) (figure 3.1). Each mutation was introduced via PCR site-directed mutagenesis. Each protein was over-expressed and purified from E. coli competent cells. The purified proteins were analyzed via SDS-PAGE to ensure proper folding by comparing it to the wild-type protein.

3.2: Tyrosine 665 and tryptophan 685 are required for binding of Iws1 to Spt6 in vitro

GST pulldown assays were performed using the GST-Iws$_{522-819}$ mutant proteins to determine whether the selected mutations affect protein-protein interactions. The GST-Iws purified proteins were first immobilized to glutathione S-transferase beads and washed extensively. Nuclear extract was purified from growing cultures of HeLa cells then pre-cleared with GST alone. The pre-cleared nuclear extract was then incubated with the immobilized bait proteins. The beads were again washed extensively then separated on an 8% SDS-PAGE then transferred to a PVDF membrane.

Figure 3.2 shows the results from the western blot conducted to test for which factors the mutant GST-Iws$_{522-819}$ constructs can bind. As shown in figure 3.2, the wild-type GST-Iws1 protein readily binds to Spt6, RNAPII Ser2P, and can also associate with Med26. Importantly, the Y665F

Figure 3.2: Western blot analysis of the mutant GST-Iws1 pulldown assay. GST-Iws proteins or GST alone was immobilized to glutathione beads then incubated with pre-cleared HeLa nuclear extract. The input lane (lane 1), shows just the nuclear extract. Lanes 2 through 5 show the pulldown reactions using the different GST-Iws mutant proteins. Lane 6 shows the pulldown using just GST. The lanes are labeled with their respective mutations (or wild-type). . Pulldown assays were separated by an 8% SDS-PAGE, and western blot assay was performed with a-H5, a-Spt6, or a-Med26 antisera as indicated.


GST-Iws
Input
Wild-Type
Y665F
W685A
T780I
GST
RNAPII Ser2P
Spt6
Med26
Coomassie
1
2
3
4
5
6

mutation strongly decreased binding of the recombinant GST-Iws1 protein to the HeLa Spt6 and RNAPII Ser2P proteins. A similar decrease in binding was also observed using the Iws1 W685A mutation. By contrast, the T780I mutation did not alter any binding of Iws1 to Spt6 or RNAPII Ser2P. Interestingly, the Y665F mutation and the T685A mutation did not have any apparent effect on the binding of Iws1 to Med26, nor was this association affected by the Iws1 T780I mutation.

The T780I mutation is located outside of the LW motif, while both the Y665F and W685A mutations affect crucial residues in the LW sequence. This indicates that binding of Iws1 to Spt6 is critically dependent upon the LW motif. Loss of either tyrosine-665 or tryptophan-685 is sufficient to destroy the binding of Iws1 to Spt6. Thus these findings reveal that these two residues within the LW core motif are of central importance to this particular protein-protein interaction.

While it is unclear from this experiment whether the Spt6-Iws interaction requires the entire LW motif, it is sufficient to say that Spt6 requires both of these particular residues within the LW motif to be intact for efficient binding to Iws. In addition, the Y665 and W685A bait proteins failed to pull RNAPII Ser2P out of the nuclear extract. This finding is consistent with previous data indicating that Iws1 does not bind directly to RNAPII but rather contacts the elongation complex indirectly, through

binding to the Spt6 protein. The finding that the interaction between Spt6 and Iws1 is greatly strengthened upon binding of the former to RNAPII raises the possibility that the conformation of Spt6 may change upon binding to the CTD in a way that exposes its N-terminus in order to recognize the LW motif present in Iws1.

By contrast, binding of Iws1 to Med26 is not affected by mutation of tyrosine-665 or tryptophan-685 within the LW motif. Therefore, Iws1 does not associate with Med26 through the LW signature, and in addition Med26 does not appear to enter through the RNAPII elongation complex, since binding of endogenous RNAPII to both mutant proteins is greatly disrupted. Unfortunately, I was unable to assess the interaction of Iws1 with TFIIS due to difficulties with the antisera for immunobloting, however studies are underway in the lab to investigate whether TFIIS is recruited to actively transcribed regions of genes in an Iws1-dependent manner in ChIP experiments. Studies with Med26 and other factors associated with TFIIS, such as Med13 and Spt8, will follow if warranted. The present studies reveal the critical role of the LW sequence in governing the binding of Iws1 to the Spt6:RNAPII complex, and further studies analyzing the properties of these mutant Iws1 proteins in vivo should reveal additional new information about the function of the complex in vivo.

DISCUSSION

My studies have highlighted the critical function of a novel domain, termed the LW motif, in the function of the human Iws1 protein. I show that Iws1 contacts the Spt6 elongation factor and associated Ser2P RNAPII elongation complexes, through direct binding to the LW motif. Remarkably, two point mutations within this signature sequence, Y665 and W685, are sufficient to destroy binding of Iws1 to Spt6 and RNAPII in vitro. Another elongation protein, TFIIS, has been observed to bind to Iws1 and to RNAPII elongation complexes through its own LW motif, and further studies are required to assess whether these associations can occur within the same complex, and whether TFIIS might access RNAPII via interaction with Iws1. Alternatively, TFIIS and Spt6 might compete for binding to Iws1 at different stages in transcription, and our lab has previously showed that binding of Iws1 to Spt6 depends upon prior binding of the latter to the RNAPII Ser2P CTD. In addition, further studies are warranted to determine whether the interaction between Iws1 and REF1/Aly, which is direct and is mediated by the conserved C-terminal domain of Iws1, is affected by mutations within the LW motif.

Interestingly, the Med26 protein also possesses an LW motif and has been suggested to bind to the RNAPII CTD directly. My studies suggest an association between Iws1 and Med26, however, this association is not

dependent upon the LW motif of Iws1. Thus one possibility is that these are separate interactions for distinct purposes, and further studies are needed to evaluate whether Med26 is carried with RNAPII and, if so, whether this is mediated through association with Iws1 protein. Alternatively, it would be interesting to see if Med26 association with the CTD does or does not require the LW motif, and, if not, whether the LW motif might recruit Iws1 to distinct Med26:CTD complexes. At present, however, there is no function ascribed to binding of Med26 to the RNAPII elongation complex.

Lastly, my studies also showed that Iws1 associates with the LEDGF cell survival protein through an N-terminal domain that is distinct from the LW motif. This association does not include other know LEDGF-binding proteins, such as Menin or MLL1, and therefore appears to be a distinct function of LEDGF. Because LEDGF plays a unique role in targeting the HIV01 integrase protein to actively transcribed regions of chromatin, it remains possible that the Spt6:Iws1 complex may play a central role in this process, which is crucial for HIV-1 integration and replication. Thus further studies addressing the composition and function of the Spt6:Iws1 complex may continue to provide unexpected new insights into the basic processes that couple mRNA capping, splicing, export and histone modification with the active RNAPII elongation complex at target genes,

and could reveal new intersections between cellular and viral regulatory events.

Recombinant proteins and nuclear extract

GST-fusion protein constructs were designed by Sunnie Yoh. The genes corresponding to each construct were PCR-amplified and ligated into pGEX.kg bacterial vectors containing the ampicillin-resistant and glutathione S-tranferase genes. The vectors were transformed into DH5a competent cells and the plasmids were subsequently purified. The DNA was verified via sequencing (Eton Bioscience). The vectors were then transformed into BL21 competent cells and protein overexpression was achieved through three-hour IPTG-induction. Cells were centrifuged and resuspended in lysis buffer (50 mM Tris-HCL, pH 7.9; 100 mM KCl; 1% Triton X-100). The resulting suspension was subject to three cycles of sonication and incubation on ice then centrifuged at high speeds (30,000 rpm). The cell lysate was incubated for three hours with glutathione S-agarose beads and thoroughly washed. Recombinant GST-lws proteins were eluted from the beads using the appropriate buffer (50 mM Tris-HCl, pH 8.0; 125 mM reduced glutathione) and analyzed via SDS-PAGE. Nuclear and whole cell extracts were isolated from growing cultures of HeLa cells. Cells were seeded on large, 150 cm plates and allowed to grow until the appropriate confluency was reached (80-90%). The cells were harvested using DMEM with 10% FBS media, washed with cold PBS, and

homogenized. The suspension was centrifuged at low speeds (4,000 rpm) and the nuclear pellet was incubated with the appropriate buffer for 30 minutes (50 mM HEPES, pH 7.9; 20% glycerol; 10% sucrose; 0.42 M KCl; 5 mM $MgCl_2$; 0.2 mM EDTA; 1 mM sodium orthovanadate; 1 mM sodium fluoride, 1x protease inhibitor [Roche]; 0.2 mM PMSF).

Protein-protein interaction experiments

For the GST pulldown experiments, 5-10 µg of the GST-fusion proteins were immobilized to glutathione-S-agarose beads. Upon binding, the beads were thoroughly washed to eliminate any non-specifically bound proteins (washing buffer consisted of 1x PBS, 300 mM NaCl, and 1% triton X-100). GST alone was immobilized with glutathione beads as well, then incubated with the nuclear extract for one hour (this step ensures adequate pre-clearing of the extract to eliminate any non-specific binding between GST and nuclear factors). The immobilized GST-fusion proteins were then incubated with 100 µl of the pre-cleared nuclear extract in a final volume of 500 µl of buffer (10 mM HEPES-HCl at pH 8.0, 300 mM KCl, 0.2 mM EDTA, 0.5% NP-40, 10% glycerol, 1mM DTT, 1 mM sodium orthovanadate, 1 mM sodium fluoride, 1x protease inhibitor, and 0.2 mM PMSF). The beads were washed extensively in the same buffer and analyzed via SDS-PAGE and immunoblotting.

PCR Site-Directed Mutagenesis

GST-lws mutant proteins were designed following the protocol of the QuikChange Site-Directed Mutagenesis Kit (Stratagene). The appropriate primers (with desired mutation) were designed through the aid of the PrimerX website. In summation, 125 nanograms of both forward and reverse complementary primers were incubated with 10 nanograms of the pGEX.kg vector containing the gene for $lws_{522\text{-}819}$. Reactions were subject to the following conditions:

Table 1 Reaction conditions of PCR site-directed mutagenesis.

Segment	Cycles	Temperature	Time
1	1	95°C	30 seconds
2	18	95°C	30 seconds
		55°C	1 minute
		68°C	15 minutes

REFERENCES

Adelman, K., Wei, W., Ardehali, M.B., Werner, J., Zhu, B., Reinberg, D., Lis, J.T. "Drosophila Paf1 modulates chromatin structure at actively transcribed genes." Mol. Cell. Biol. (2006): 250-260.

Adkins MW, Tyler JK. "Transcriptional activators are dispensable for >> transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions." Mol Cell 21 (2006):405-416.

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

Andrulis, E.D., Guzman, E., Doring, P., Werner, J., Lis, J.T. "High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter poximal pausing and transcription elongation." Genes & Dev. 14 (2000): 2635-2649.

Birse, C. C., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. 1998. "Coupling termination of transcription to messenger RNA maturation in yeast." Science 280 (1998): 298-301.

Bortvin, A., Winston, F. "Evidence that Spt6 controls chromatin structure by direct interaction with histones." Science. (1996); 272: 1473-1476

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., Eick, D. "Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7." Science (2007): 1780-1782.

Cho, E.J., Rodriguez, C.R., Takagi, T., and Buratowski, S. "Allosteric interactions between capping enzyme subunits and the RNA polymerase II carboxy-terminal domain." Genes & Dev. 12 (1998): 3482-3487.

Corden, J.L. "Tails of RNA polymerase II." Trends Biol. Sci. 15 (1990): 383-387.

Dahmus, M.E. "Phosphorylation of the C-terminal domain of RNA polymerase II." Biochem. Biophys. (1995): 171-182.

Engelman, A., Cherepanov, P. "The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication." PLoS Pathog. (2008).

Guo, S., Yamaguchi, Y., Schilbach, S., Wada, T., Lee, J., Godard, A., French, D., Handa, H., Rosenthal, A. "A regulator of transcriptional elongation controls vertebrate neuronal development." Nature. (2000): 366-369

Hampsey, M., Reinberg, D. "Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation." Cell (2003): 429-432.

Hartzog, G.A., Wada, T., Handa, H., Winston, F. "Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae." Genes & Dev. (1998): 357-369.

Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. "Temporal reguation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases." Mol. Cell 2 (1998): 43-53.

Hirose, Y., Manley, J.L. "RNA polymerase II and the integration of nuclear events." Genes & Dev. (2000): 1415-1429.

Hirose, Y., Ohkuma, Y. "Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression." J. Biochem. (2007): 601-608.

Ho, C.K. and Shuman, S. 1999. "Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme." Mol. Cell. 3 (1999): 405-411.

Jones, J.C., Phatnani, H.P., Haystead, T.A., MacDonald, J.A., Alam, S.M., and Greenleaf, A.L. "C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats." J. Biol. Chem. 279 (2004): 24957-24964.

Jove, R. and Manley, J.L. "In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites." J. Biol. Chem. 259 (1984): 8513-8521.

Kaplan, C.D., Morris, J.R., Wu, C.T., Winston, F. "Spt5 and Spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster." Genes & Dev. (2000): 2623-2634.

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

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

Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., Greenblatt, J.F. "RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach." Mol. Cell. Biol. (2002): 6979-6992.

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

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

Lee, J.M., and Greenleaf, A.L. "Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I." J. Biol. Chem. (1997): 10990-10993.

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

Li, J., Moazed, D., and Gygi, S.P. "Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation." J. Biol. Chem. 277 (2002): 49383-49388.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. "Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II." Mol. Cell. (2002): 1101-1111.

Lindstrom, D.L. Squazzo, S.L., Muster, N., Burkin, T.A., Watcher, K.C., Emigh, C.A., McCleery, J.A., Yates III, J.R., Hartzog, G.A. "Dual roles for Spt5 in pre-

mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins." Mol. Cell. Biol. (2003): 1368-1378.

Ling, Y., Smith, A. J., Morgan, G.T. "A sequence motif conserved in diverse nuclear proteins identifies a protein interaction domain utilized for nuclear targeting by human TFIIS." Nucleic Acids Res. (2006): 2219-2229.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., Reinberg, D. "Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II." Proc. Natl. Acad. Sci. U. S. A. (2004): 7572-7577.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. "Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase." J. Biol. Chem. (1996): 27176-27183.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Program, A.E., Shuman, S., et. al. "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II." Genes & Dev. (1997(a)): 357-361.

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

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. "A structural perspective of CTD function." Genes & Dev. (2005): 1401-1415.

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

Ohkuma, Y. "Multiple functions of general transcription facotrs TFIIE and TFIIH in transcription: possible points of regulation by trans-acting factors." J. Biochem. (1997): 481-489.

Orphanides, G., Reinberg, D. "A unified theory of gene expression." Cell (2002): 439-451.

Phatnani, H.P., and Greenleaf, A.L. "Phosphorylation and functions of the RNA polymerase II CTD." Genes & Dev. (2006): 2922-2936.

Prelich, G. "RNA polymerase II carboxy-terminal domain kinases: Emerging clues to their function." Eukaryot. Cell. (2002): 153-162.

Rasmussen, E.B., and Lis, J.T. "In vivo transcriptinoal pausing and cap formation on three Drosophila heat shock genes." Proc. Natl. Acad. Sci. (1993): 7923-7927.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., and Buratowski, S. "Kin28, the TFIIH-associated CTD kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II." Mol. Cell. Biol. (2000): 104-112.

Roeder, R.G. "The role of general initiation factors in transcription by RNA polymerase II." Trend. Biochem. Sci. (1996): 327-335.

Scafe, C. Chao D., Lopes J., Hirsch J.P., Henry S., Young R.A. "RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals." Nature. (1990): 491-494

Schaft, D., Roguev, A., Kotovic, K.M., Shevchenko, A., Sarov, M., Neugebauer, K.M., and Stewart, A.F. "The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation." Nucleic Acids Res. (2003): 2475-2482.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. "Dynamic association of capping enzymes with transcribing RNA polymerase II." Genes & Dev. (2000): 2435-2440.

Sehgal, P.B., Darnell, J.E., Tamm, I. "The inhibition by DRB of hnRNA and mRNA production in HeLa cells." Cell. (1976): 473-480.

Sims III, R.J., Belotserkovskaya, R., Reinberg, D. "Elongation by RNA polymerase II: the short and long of it." Genes & Dev. (2004): 2437-2468.

Stiller, J.W., and Cook, M.S. "Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs." Eukaryot. Cell (2004): 735-740.

Svejstrup, J.Q., Vichi, P., Egly, J.M. "The multiple roles of transcription/repair factor TFIIH." Trends Biochem. Sci. (1996): 346-350.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., 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): 343-356.

Winston, F. "Control of eukaryotic transcription elongation." Genome Biol. (2001): 2.

Winston, F., Chaleff, D.T., Valent, B., Fink, G.R. "Mutations affected Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae." Genetics (1984): 325-336.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. "Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast." Genes & Dev. (2003): 654-663.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H. "NEFL, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation." Cell (1999): 41-51.

Yamamoto S., Watanabe Y., van der Spek, P.J., Watanabe T., Fujimoto, H., Hanaoka, F., Ohkuma, Y. "Studies of nematode TFIIE function reveal a link between Ser-5 phosphorylation of RNA polymerase II and the transition from transcription initiation to elongation." Mol. Cell. Biol. (2001): 1-15.

Yokoyama, A., Cleary, M.L. "Menin critically links MLL proteins with LEDGF on cancer-associated target genes." Cancer Cell. (2008): 36-46.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., Jones, K.A. "The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export." Genes & Dev. (2007): 160-174.

Yoh, S. M., Lucas, J.S., Jones, K.A. "The Iws1:Spt6:CTD complex controls co-transcriptionaly mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation." Genes & Dev. 24(2008): 3422-34

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