Characterization of the Association of mRNA Export Factor Yra1 with the C-terminal Domain of RNA Polymerase II \textit{in vivo} and \textit{in vitro}

by

April Lee MacKellar

Department of Biochemistry
Duke University

Date:_______________________

Approved:

___________________________
Arno Greenleaf, Supervisor

___________________________
Kenneth Kreuzer

___________________________
Michael Been

___________________________
Laura Rusche

___________________________
Mariano Garcia-Blanco

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2011
ABSTRACT

Characterization of the Association of mRNA Export Factor Yra1 with the C-terminal Domain of RNA Polymerase II in vivo and in vitro

by

April Lee MacKellar

Department of Biochemistry
Duke University

Date:_______________________

Approved:

___________________________
Arno Greenleaf, Supervisor

___________________________
Kenneth Kreuzer

___________________________
Michael Been

___________________________
Laura Rusche

___________________________
Mariano Garcia-Blanco

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2011
Abstract

The unique C-terminal domain (CTD) of RNA polymerase II (RNAPII), composed of tandem heptad repeats of the consensus sequence YSPTSPS, is subject to differential phosphorylation throughout the transcription cycle. Several RNA processing factors have been shown to bind the appropriately phosphorylated CTD, and this facilitates their localization to nascent pre-mRNA during transcription. In *Saccharomyces cerevisiae*, the mRNA export protein Yra1 (ALY/REF in metazoa) has been shown to cotranscriptionally associate with mRNA and is thought to deliver it to the nuclear pore complex for export to the cytoplasm. Based on a previous proteomics screen, I hypothesized that Yra1 is a *bona fide* phosphoCTD associated protein (PCAP) and that this interaction is responsible for the pattern of Yra1 cotranscriptional association observed *in vivo*. Using *in vitro* binding assays, I show that Yra1 directly binds the hyperphosphorylated form of the CTD characteristic of elongating RNAPII. Using truncations of Yra1, I determined that its phosphoCTD-interacting domain (PCID) resides in the segment comprising amino acids 18-184, which, interestingly, also contains the RNA Recognition Motif (RRM) (residues 77-184). Using UV crosslinking, I found that the RRM alone can bind RNA, although a
larger protein segment, extending to the C-terminus (aa 77-226), displays
stronger RNA binding activity. Even though the RRM is implicated in both RNA
and CTD binding, certain RRM point mutations separate these two functions:
thus, mutations that produce defects in RNA binding do not affect CTD binding.
Both functions are important in vivo, in that RNA binding-defective or CTD
binding-defective versions of Yra1 engender growth and mRNA export defects.
I also report the construction and characterization of a useful new temperature
sensitive YRA1 allele (R107AF126A). Finally, using chromatin
immunoprecipitation, I demonstrate that removing the N-terminal 76 amino
acids of Yra1 (all of the PCID up to the RRM) results in a 10-fold decrease in Yra1
recruitment to genes during elongation. These results indicate that the PCTD is
likely involved directly in cotranscriptional recruitment of Yra1 to active genes.
Based on the results of this study, I propose that the hyperphosphorylated CTD
of RNAPII, rather than or in addition to the TREX complex, recruits Yra1 to its
nascent mRNA cargo during transcription elongation in Saccharomyces cerevisiae.
# Table of Contents

Abstract .......................................................................................................................... iv

List of Tables .................................................................................................................. ix

List of Figures .................................................................................................................. x

Abbreviations .................................................................................................................. xiv

Acknowledgements .......................................................................................................... xvi

1. Introduction .................................................................................................................. 1

  1.1. RNA polymerases ................................................................................................. 1

  1.2. RNA polymerase II and the CTD ........................................................................ 1

  1.3. Modifications of the CTD .................................................................................... 5

      1.3.1. Phosphorylation ............................................................................................. 5

      1.3.2. Other CTD modifications ............................................................................... 9

  1.4. CTD binding proteins .......................................................................................... 10

      1.4.1. Early/initiation ............................................................................................... 10

      1.4.2. Elongation ..................................................................................................... 12

      1.4.3. Termination ................................................................................................... 15

  1.5. Nucleocytoplasmic transport of mRNA ................................................................. 16

  1.6. Yra1 and mRNA export ....................................................................................... 17

2. Characterization of Yra1 binding interactions in vitro .............................................. 25

  2.1. Introduction ........................................................................................................... 25

  2.2. Materials and methods ......................................................................................... 28
2.2.1. Reagents .......................................................................................................................... 28
2.2.2. Cloning ............................................................................................................................ 28
2.2.3. Protein expression and purification ................................................................................. 29
2.2.4. Far Western assay .......................................................................................................... 31
2.2.5. Immobilized CTD peptide binding assay ...................................................................... 31
2.2.6. UV crosslinking ............................................................................................................. 32
2.3. Results .................................................................................................................................. 33
2.3.1. Expression and purification of recombinant Yra1 ......................................................... 33
2.3.2. Yra1 interacts directly with the hyperphosphorylated yeast CTD .............................. 34
2.3.3. The phosphoCTD interacting domain of Yra1 spans the N-variable and RRM regions ............................................................... 36
2.3.4. Mutations in the PCID of Yra1 did not affect the CTD binding interaction ................ 40
2.3.5. The RNA and CTD binding domains contain the RRM .............................................. 42
2.3.6. The RNA and CTD binding domains involve different residues ............................. 46
2.3.7. The interaction between Iws1 and Yra1/ALY is conserved ........................................ 49
2.4. Discussion .......................................................................................................................... 50
3. Role of CTD binding in mRNA export by Yra1 in vivo .................................................... 54
3.1. Introduction ....................................................................................................................... 54
3.2. Materials and methods ..................................................................................................... 55
3.2.1. Yeast strains and growth ............................................................................................. 55
3.2.2. Cloning ......................................................................................................................... 56
3.2.3. Fluorescence in situ hybridization (FISH) and immunostaining .............................. 56
3.2.4. Chromatin immunoprecipitation .......................................................... 57

3.3. Results ........................................................................................................ 59

3.3.1. V5-tagged Yra1 supports viability in S. cerevisiae ................................ 59

3.3.2. Yra1 CTD and RNA binding mutants are expressed and display growth phenotypes ......................................................................................... 61

3.3.3. Yra1 PCTD and RNA binding mutants are sensitive to hydroxyurea .... 62

3.3.4. RRM mutations do not alter Yra1 subcellular localization .................. 63

3.3.5. Yra1 CTD and RNA binding mutants display mRNA export defects .... 64

3.3.6. PCID truncation leads to a loss of Yra1 at the PMA1 and PYK1 genes .... 66

3.4. Discussion ................................................................................................... 72

4. Future Directions .......................................................................................... 76

4.1. Yra1, the PCTD, and cotranscriptional recruitment .................................. 76

4.2. Yra1 RNA binding and regulation ............................................................... 78

4.3. Interdependence of the recruitment of export factors ............................... 79

References ......................................................................................................... 81

Biography .......................................................................................................... 94
List of Tables

Table 3.1: Primers used for quantitative real-time PCR analysis. ............................................ 68
**List of Figures**

Figure 1.1: Amino acid sequences of the mammalian, *Drosophila melanogaster*, and *Saccharomyces cerevisiae* C-terminal domains. Residues that conform to the consensus sequence are in black; residues that deviate from the consensus are in red. The residues at the C-terminus of the domain that do not conform to the consensus are in blue. ........ 3

Figure 2.1: Coomassie-stained 4-20% gradient SDS-PAGE gel of fractions from the two-step purification of recombinant His-tagged Yra1. After cell cracking, the soluble protein lysate was applied in batch to IMAC resin (IMAC OP and FT). As the IMAC elutions contained low molecular weight degradation products or contaminating proteins (indicated by asterisks), these fractions were pooled and loaded onto a Mono S ion exchange column (Mono S OP). The peak of full length Yra1 from the Mono S salt elutions is shown in fractions 32-36. ............................................................................................................. 34

Figure 2.2: Far Western Assay of Yra1 and the PCTD. The Coomassie-stained gel (left) shows purified recombinant His-tagged Yra1 (approximately 28 kDa). The purified protein was subjected to a Far Western assay probed with a yeast CTD fusion protein hyperphosphorylated ([32P]-labeled) with CTDK-I. The autoradiograph (right) shows that the PCTD probe was retained at the same mobility as Yra1. ................................................................. 35

Figure 2.3: Immobilized CTD peptide binding assays. A) Synthetic biotinylated three-repeat consensus CTD peptides that were used for the subsequent binding assays. Serines in blue indicate phosphoserines. B) The CTD peptides shown in A were immobilized on avidin resin. Yra1 was applied to the columns and onput (OP), flowthrough (FT), wash (W1 and W2), and salt elution fractions (0.3 M and 1 M NaCl) were analyzed on 4-20% SDS PAGE gels stained with Coomassie Blue. ........................................ 36

Figure 2.4: Far Western assay of Yra1 fragments with the PCTD. A) Yra1 architecture based on homology among the REF family of proteins (Zenklusen et al, 2001). The blue regions, including the Nbox, RNA Recognition Motif (RRM), and Cbox, display higher conservation, while the red regions, including the N-variable (Nvr) and C-variable (Cvr), have variable sequences and lengths among the REF proteins. B) Top, Coomassie-stained gel of Yra1 truncation proteins subjected to Far Western assay. Bottom, Autoradiograph of Far Western assay of Yra1 truncation proteins probed with [32P]-labeled hyperphosphorylated CTD fusion protein. ................................................................. 38

Figure 2.5: Immobilized CTD peptide binding assay of Yra1 fragments. Coomassie-stained gels of fractions from the immobilized peptide column binding assays of Yra1
truncations. Yra1 aa 1-226 and 1-184 bound to the immobilized 2,5P$_3$ peptide and eluted with 0.3 M NaCl, but Yra1 aa 77-226 and 70-190 only predominantly in the flowthrough (FT) and wash (W1 and W2) fractions of the column. 

Figure 2.6: Binding of Yra1 aa 18-184 to the PCTD. A) The N-terminal and RRM regions (residues 18-184), which lacks the Nbox and the entire C-terminus (top), was and subjected to a Far Western assay probed with [$^{32}$P]-labeled hyperphosphorylated CTD fusion protein (stained gel and autoradiograph below the protein diagrams). B) Coomassie stained gels of fractions (onput-OP, flowthrough-FT, washes-W1 and W2, and elutions in two salt steps) from an immobilized peptide binding assay (2,5P$_3$ peptide) of full-length Yra1 1-226 (top) and the truncated variant 18-184 (bottom).

Figure 2.7: Alignment of the PCID region of Yra1 (S.c.Yra1; NP_010669.1) with the Drosophila (D.m.ALY; NP_609574.1) and mouse (M.m.ALY; NP_062357.3) homologues. The residues indicated with dots were mutated to alanines. The RNP1 and RNP2 motifs are labeled. ClustalW and BoxShade were used to make the alignment.

Figure 2.8: Far Western assay of alanine mutations with the PCID of Yra1. Top, Coomassie-stained gel of Yra1 mutants subjected to Far Western assay. Bottom, Autoradiograph of Far Western assay of Yra1 mutants probed with [$^{32}$P]-labeled hyperphosphorylated CTD fusion protein.

Figure 2.9: UV crosslinking of full-length Yra1 to a [$^{32}$P]-labeled 15-mer RNA probe (5'-CAGUCGCAUAGUGCA-3') with increasing exposure times to UV light. Top, phosphorimage of crosslinked RNA. Bottom, Coomassie-stained gel of proteins.

Figure 2.10: UV crosslinking of full-length (1-226) and truncations (1-184, 77-226, and 77-184) of Yra1 to the 15-mer RNA probe. Top, phosphorimage of crosslinked RNA. Bottom, Coomassie-stained gel of proteins.

Figure 2.11: UV crosslinking of 15-mer RNA to full-length Yra1 with no competitor (lanes 1, 2, 5) or 25-fold (lanes 3, 6) and 100-fold (lanes 4, 7) molar excess of cold RNA or 2,5P$_3$ CTD peptide competitors, respectively. Top, phosphorimage of crosslinked RNA. Bottom, Coomassie-stain of proteins.

Figure 2.12: A) UV crosslinking of increasing amounts of full-length Yra1 expressed in E. coli (lanes 1-3) and baculovirus-infected Sf9 cells (lanes 4-6) to the 15-mer RNA probe. Top, phosphorimage of crosslinked RNA. Bottom, Coomassie-stained gel of proteins. B) Coomassie-stained gels of fractions from the immobilized peptide column binding assays of Yra1 expressed in E. coli (top) and baculovirus-infected Sf9 cells (bottom). ...
Figure 2.13: Amino acid alignment of the RRM sequences of Yra1 and homologous proteins. The RNP1 and RNP2 motifs are labeled. Three conserved residues indicated with dots were identified (R107, Y112, & F126 in Yra1) and mutated to alanines. The star below the alignment (N123 in Yra1) indicates the position of a conserved phenylalanine in most (non-Yra1) RRM sequences. BLAST was used to identify homologues of Yra1, and ClustalW and BoxShade were used to make the alignment. S.c. – *Saccharomyces cerevisiae* Yra1 (NP_010669.1); S.p. – *Schizosaccharomyces pombe* Mlo3 (NP_595715.1) and THO subunit (NP_595161.1); N.c. – *Neurospora crassa* hypothetical protein NCU01793 (XP_956471.1); C.e. – *Caenorhabditis elegans* ALY1 (NP_501588.1) and ALY2 (NP_501594.1); D.r. – *Danio rerio* THO subunit 4 (NP_001098578.1); G.g. – *Gallus gallus* hypothetical protein (XP_001232393.1); M.m. – *Mus musculus* ALY (NP_062357.3).

Figure 2.14: A) UV crosslinking of wild-type and alanine mutants of Yra1 to a $^{32}$P-labeled 15-mer RNA probe. Top, phosphorimage of crosslinked RNA. Bottom, Coomassie-stained gel of proteins. B) Coomassie-stained gels of fractions from the immobilized 2,5P$_3$ peptide column binding assays of Yra1 R107A, Y112A, F126A and R107AY112AF126A (Triple) mutants. All mutants display binding to the immobilized 2,5P$_3$ peptide and elute with 0.3 M NaCl.

Figure 2.15: Far Western assay of Yra1 and Iws1. Left, Coomassie stained gel of Yra1 truncation proteins subjected to Far Western assay. Right, Anti-His tag visualization of Iws1 Far Western assay of Yra1 truncation proteins.

Figure 3.1: Characterization of the Yra1V5-expressing strain. A-C) Western blots of total cell extracts from the WT YRA1 strain, WT YRA1 strain expressing plasmid-encoded yra1V5, and yra1::HIS3 strain expressing plasmid-encoded yra1V5.

Figure 3.2: Western blot of V5-tagged protein expression in *S. cerevisiae*. Top, anti-V5 Western blot of total yeast extracts. Bottom, anti-G6PDH Western blot of total yeast extracts.

Figure 3.3: Growth of the RRM point mutant (R107A, Y112A, F126A, R107AF126A, and Y112AF126A) and N-terminal truncation (77-226) strains. Serial 10-fold dilutions of cells were spotted on rich medium (YPD plates) and plates were incubated at 30°C (left), 37°C (middle), or 16°C (right).

Figure 3.4: Effects of DNA damaging agents on growth of the RRM (R107A, Y112A, F126A, R107AF126A, and Y112AF126A) and PCID (77-226) mutants. Serial 10-fold dilutions of cells were spotted on rich medium in the absence (YPD) or presence of DNA.
damaging agents [2 mM methyl methanesulfonate (MMS), 200 mM hydroxyurea (HU), and 50 µg/mL doxorubicin (DOX)].

Figure 3.5: In situ localization of Yra1 proteins in yeast. V5 epitope-tagged Yra1 constructs were detected by immunofluorescence using a FITC-labeled anti-V5 antibody (green); DAPI (blue) was used to stain the nuclei.

Figure 3.6: Yra1 PCID truncation strain displays an export defect at 37°C. Fluorescence in situ hybridization (FISH) of polyA+ RNA is shown in green. DAPI (blue) was used to stain the nuclei. Nonheat shock (30°C; left) and heat shock (37°C; right) conditions were used.

Figure 3.7: Yra1 RRM point mutations (F126A, R107AF126A, and Y112AF126A) produce export defects at 30°C. FISH of polyA+ RNA is shown in green, and DAPI (blue) was used to stain the nuclei.

Figure 3.8: Schematic showing the primer set locations on the PMA1 and PYK1 genes.

Figure 3.9: Chromatin immunoprecipitation of wild-type (WT) and truncated (-PCID) Yra1 at the PMA1 gene.

Figure 3.10: Chromatin immunoprecipitation of wild-type (WT) and truncated (-PCID) Yra1 at the PYK1 gene.

Figure 3.11: Anti-V5 Western of the WT and –PCID (77-226) extracts used for the ChIP experiments. A and B for each stain indicate biological replicates. Crosslinks were reversed by incubation at 95°C for 1 hour, and proteins were separated on a 4-20% gel.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>ChIP</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain of RNA polymerase II</td>
</tr>
<tr>
<td>CTDK-I</td>
<td>C-terminal domain kinase I</td>
</tr>
<tr>
<td>DOX</td>
<td>doxorubicin</td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescent <em>in situ</em> hybridization</td>
</tr>
<tr>
<td>GSTyCTD</td>
<td>glutathione S-transferase fused to the yeast CTD</td>
</tr>
<tr>
<td>HU</td>
<td>hydroxyurea</td>
</tr>
<tr>
<td>HS</td>
<td>heat shock</td>
</tr>
<tr>
<td>IMAC</td>
<td>immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>MMS</td>
<td>methyl methanesulfonate</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>mRNP</td>
<td>messenger ribonucleoprotein</td>
</tr>
<tr>
<td>NHS</td>
<td>non-heat shock</td>
</tr>
<tr>
<td>NPC</td>
<td>nuclear pore complex</td>
</tr>
<tr>
<td>PCAP</td>
<td>phosphoCTD associated protein</td>
</tr>
<tr>
<td>PCID</td>
<td>phosphoCTD interaction domain</td>
</tr>
<tr>
<td>PCTD</td>
<td>phosphorylated C-terminal domain of RNA polymerase II</td>
</tr>
<tr>
<td>polyA⁺</td>
<td>polyadenylated RNA</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>pre-mRNA</td>
<td>precursor messenger RNA</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>Ser2P</td>
<td>phosphorylated serine 2 of the CTD</td>
</tr>
<tr>
<td>Ser5P</td>
<td>phosphorylated serine 5 of the CTD</td>
</tr>
<tr>
<td>Ser7P</td>
<td>phosphorylated serine 7 of the CTD</td>
</tr>
<tr>
<td>Ser2/5P</td>
<td>phosphorylated serines 2 and 5 of the CTD</td>
</tr>
<tr>
<td>Ser2/5/7P</td>
<td>phosphorylated serines 2, 5, and 7 of the CTD</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>SRI domain</td>
<td>Set2 Rpb1 interaction domain</td>
</tr>
<tr>
<td>TREX</td>
<td>transcription export complex</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>ts</td>
<td>temperature sensitive</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet light</td>
</tr>
<tr>
<td>WT</td>
<td>wild type</td>
</tr>
</tbody>
</table>
Acknowledgements

There are many individuals that contributed, directly or indirectly, to the completion of this dissertation. First and foremost, the guidance and expertise of Arno Greenleaf was vital. In addition to providing a home for me in his lab, he has been eager to assist me when needed and has allowed me grow as an independent scientist. Craig Bennett has also been a valued colleague and a teacher of all things yeast. Past and present lab members, including Hemali Phatnani, Jan Jones, Mary Beth Danforth, Tiffany Winsor, Peter Decker, Pengda Liu, and Bart Bartkowiak, have provided valuable discussions and laughs throughout the years. I have been privileged to share the lab with people who are both intelligent and genuinely friendly. Pei Zhou and Jon Werner-Allen have also aided on my research by helping me run the NMR and troubleshoot problems related to the structural characterization of the Yra1-PCTD complex. While this work did not pan out, I appreciate the time and effort they put into these experiments.

I deeply appreciate the efforts of my thesis committee, who have supported and helped me develop my project. Each meeting I had with them improved my research and helped me to overcome many experimental obstacles. Mariano Garcia-Blanco and Mike Been provided their expertise in RNA and RNA binding; Laura Rusche taught me how to perform ChIP; and Ken Kreuzer always had valuable suggestions, comments, and words of encouragement. I would also like to thank the Department of
Biochemistry, who provided the funding that allowed me to finish my dissertation. I am indebted to all of these individuals who have helped me professionally over the years.

Outside of the lab, my family has been an immense source of emotional and psychological support. My parents, John and Barbara, taught me the value of hard work and the importance of education. They have been understanding, patient, and positive, which I greatly appreciate. I would also like to thank my sister and her family, Lisa DeCarlo and Stephanie and Damien Horrace, for their support and love. I love you all.

Finally, I have so many friends to thank. You know who you are, what you’ve done, and how much I appreciate it. I have met the most wonderful people at Duke over the years, and they have enriched this experience.
1. Introduction

1.1. RNA polymerases

The DNA-dependent RNA polymerases are complex, multisubunit enzymes that catalyze the polymerization of ribonucleic acids de novo based on a DNA template sequence, a process termed Transcription. Bacteria encode one RNA polymerase that is responsible for transcribing all RNA species. This enzyme is composed of four subunits: \( \alpha \) (rpoA), \( \beta \) (rpoB), \( \beta' \) (rpoC), \( \omega \) (rpoZ), and \( \sigma \) (rpoD). The core of RNA polymerase, \( \alpha_2 \beta \beta' \omega \), is responsible for DNA-dependent RNA synthesis, while the \( \sigma \) subunit recognizes and melts the promoter sequence, which facilitates initiation.

Due to the greater genetic complexity, one RNA polymerase is not sufficient for transcription in eukaryotic cells. There are three RNA polymerases (I, II, and III) in eukaryotes that synthesize the different classes of RNAs. RNA polymerase I transcribes all of the ribosomal RNAs (rRNAs), except 5S rRNA. RNA polymerase II transcribes all messenger RNAs (mRNAs) and some small nuclear RNAs (snRNAs). RNA polymerase III transcribes transfer RNAs (tRNAs), 5S rRNA, and some snRNAs. These enzymes are highly conserved with the prokaryotic RNA polymerase; they consist of 12-15 subunits (three of which, Rpb5, 6, and 8, are shared among the three enzymes) (Kolodziej et al, 1990; Woychik et al, 1990), and the two large subunits Rpb1 and Rpb2 are homologous to \( \beta' \) and \( \beta \), respectively (Allison et al, 1985; Sweetser et al, 1987).

1.2. RNA polymerase II and the CTD

Unlike RNA polymerases I and III, RNA polymerase II (RNAPII) was found to exist in three forms designated IIO, IIA, and IIB, which had large subunits (IIo, IIa, and IIb,
respectively) with distinct apparent molecular masses/electrophoretic mobilities on a
denaturing gel. Christmann and Dahmus (1981) generated a monoclonal antibody to calf
thymus RNA polymerase II and found that it cross-reacted with the IIo and IIa species
but not the IIb form, indicating that the antigenic determinant in the IIo/a forms was
lost in IIb (Christmann & Dahmus, 1981). Corden et al. (1985) determined the amino acid
compositions of IIa and IIIb and found that IIa contained more serine, proline, threonine,
and tyrosine residues in a ratio of 3:2:1:1 compared with IIb. In addition, this sequence
was unique to RNAPII, as the largest subunit of RNA polymerase III did not have a
homologous sequence (Allison et al, 1985). Using the Drosophila RNAPIIa/o subunit
sequence as a probe, they then identified the mouse RPB1 gene and found that the
region carboxyl-terminal to the b’ homologous region consisted of highly conserved
repeats of the consensus sequence Y₁S₂P₃T₄S₅P₆S₇ (Corden et al, 1985). These data
indicated that the IIa and IIo subunits contained the full β’-like subunit, while the IIb
variant lacked the C-terminal extension.

While this unique carboxyl-terminal extension (called the C-terminal domain – CTD)
is well conserved among eukaryotes, the length of the CTD varies among organisms,
with 26-27 repeats in Saccharomyces cerevisiae, 42 repeats in Drosophila melanogaster, and
52 repeats in mammals (Allison et al, 1988). In addition, the level of conservation is also
variable. In S. cerevisiae, the sequence of the CTD predominantly consists of consensus
heptad repeats, but the D. melanogaster CTD contains few fully consensus repeats (Fig.
1.1). The CTD is thought to be a largely unstructured domain. The heptad repeats
contain SPXX motifs (S₂P₃T₄S₅ and S₆P₇Y₁), which are known to form beta-turn
structures in solution. Indeed, synthetic CTD repeat peptides displayed the tendency to
form beta-turns by circular dichroism and nuclear magnetic resonance analyses (Cagas
Figure 1.1: Amino acid sequences of the mammalian, *Drosophila melanogaster*, and *Saccharomyces cerevisiae* C-terminal domains. Residues that conform to the consensus sequence are in black; residues that deviate from the consensus are in red. The residues at the C-terminus of the domain that do not conform to the consensus are in blue.

The role of the CTD is clearly important, as truncation or deletion of the CTD in *Drosophila* or yeast is lethal. In yeast, cells expressing Rpb1 with at least 10 heptad
repeats grow as well as wild type, but truncation below 8 repeats results in inviability (Nonet et al, 1987; West & Corden, 1995). In *Drosophila*, deletion of approximately half of the CTD produces homozygous lethality; however, this variant is catalytically functional *in vivo* and *in vitro*. Moreover, treating the purified enzyme with protease to completely remove the CTD also does not affect *in vitro* transcription, indicating that the essential function of the CTD was likely unrelated to the polymerization of RNA (Zehring et al, 1988). In a mouse cell line, truncation of the CTD to 36 repeats does not affect the ability of the polymerase to confer amanitin resistance; however, 23 CTD repeats are not sufficient to confer resistance (Bartolomei et al, 1988). Taken together, these studies indicated that while the CTD is required for viability, it is not catalytically important for the polymerase and the heptads display functional redundancy.

The relationship between the two CTD-containing forms of RNAPII (IIa and IIo) was first demonstrated by Cadena and Dahmus (1987); in this study, treatment of the IIo form with alkaline phosphatase converted it to the electrophoretic mobility of the IIa form (Cadena & Dahmus, 1987). Likewise, the IIa form could be treated with a kinase and shifted to the mobility of the IIo form (Payne et al, 1989; Zhang & Corden, 1991b). The roles of these forms *in vitro* seemed to be distinct: IIA displayed higher activity with the adenovirus-2 major late promoter (which required initiation) (Chesnut et al, 1992; Kim & Dahmus, 1989), while IIo seemed to be important for transcription elongation based on photoaffinity labeling experiments (Bartholomew et al, 1986; Cadena & Dahmus, 1987; Payne et al, 1989). In addition, an antibody that has high affinity for mammalian IIa blocked transcription initiation *in vitro*, but addition of this antibody after initiation could not block elongation (Laybourn & Dahmus, 1989; Moyle et al, 1989). Together, these results indicated that IIa and IIo were the same polypeptide that
was reversibly post-transcriptionally modified, predominately by phosphorylation, and that the dynamics of modifications were important for mRNA transcription.

1.3. **Modifications of the CTD**

1.3.1. **Phosphorylation**

The first identified and best-characterized post-translational modification of the CTD is phosphorylation. The sequence of the CTD contains multiple residues that could be phosphorylated, including three serines, one threonine, and one tyrosine. Using phosphoamino acid analysis, the mammalian CTD was first shown to be phosphorylated in vivo and in vitro on serine and threonine residues, but phosphotyrosines were not found (Zhang & Corden, 1991a). Weak reactivity of RNAPII immunoprecipitated from HeLa cells with an anti-pTyr antibody was later found, indicating that a low level of tyrosine phosphorylation occurs in vivo (Baskaran et al, 1993); however, the function of this modification is not known.

The importance of CTD modifications in vivo has been investigated by a mutational analysis in budding yeast. Alleles in which serines 2 or 5 or both serines 2 and 5 in the heptad repeats were substituted with either alanines or glutamates could not support viability. Inversion of the Ser2 with Pro3 (i.e., Y₁₃P₂S₃T₄S₅P₆S₇) and substitution of Tyr1 with a phenylalanine were also lethal. However, if only half of the repeats contained substitutions, most combinations were viable, except for partial substitution of serines 2 and 5 with glutamates (West & Corden, 1995). These results indicated that modification of these residues in vivo is important for CTD function and mRNA transcription. More recently, phosphorylation of Ser7 of the CTD was also found in budding yeast and mammals (Akhtar et al, 2009; Chapman et al, 2007). While this mark is also found on
protein-coding genes, the main role of Ser7P has been attributed to snRNA transcription (Egloff et al, 2007).

The cyclin-dependent kinases responsible for CTD phosphorylation have been well characterized. Phosphorylation of Ser5 and Ser7 during the formation of the preinitiation complex is mediated by Kin28/Ccl1 in yeast (CDK7/Cyclin H in metazoa), which is a subunit of the TFIIH complex (Akhtar et al, 2009; Glover-Cutter et al, 2009; Tietjen et al, 2010). In opposition to Kin28/CDK7, Srb10/Srb11 (CDK8/Cyclin C in metazoa) inhibits initiation, possibly by phosphorylating the CTD prior to preinitiation complex formation, by phosphorylating and inhibiting Kin28/CDK7, or both mechanisms (Akoulitchev et al, 2000; Hengartner et al, 1998). Productive elongation is promoted by the actions of two enzymes: CTDK-I (CDK12/13 in metazoa) (Bartkowiak et al, 2010; Lee & Greenleaf, 1989; Lee & Greenleaf, 1991; Sterner et al, 1995) and Bur1/Bur2 (P-TEFb aka CDK9/Cyclin T in metazoa) (Marshall et al, 1996; Murray et al, 2001; Yao et al, 2000). CTDK-I is thought to provide the bulk of Ser2 CTD phosphorylation during elongation, especially during diauxic growth (Cho et al, 2001; Patturajan et al, 1999); however, it also possesses the ability to phosphorylate Ser5 and efficiently generates doubly phosphorylated repeats if presented with peptides containing Ser2P or Ser5P in vitro (Jones et al, 2004). In addition to CTD phosphorylation, CTDK-I also seems to play a role in removing basal transcription factors after initiation, but this role is independent of its kinase activity (Ahn et al, 2009). Bur1 (P-TEFb) also phosphorylates Ser2 as well as Ser7 of the CTD (Murray et al, 2001; Qiu et al, 2009; Tietjen et al, 2010) and the elongation factor Spt4/5 (DSIF in metazoa) (Ivanov et al, 2000).

CTD phosphorylation is also modified by CTD phosphatases. Fcp1, which is evolutionarily conserved and essential in yeast, was the first identified CTD
phosphatase and specifically removes the Ser2P mark (Chambers & Dahmus, 1994; Cho et al, 2001). Ssu72, a component of the cleavage/polyadenylation factor complex, is responsible for removing Ser5 phosphorylation (Krishnamurthy et al, 2004). Recently, a third phosphatase, Rtr1, was also shown to dephosphorylate the Ser5 of the CTD during elongation (Mosley et al, 2009). Interestingly, the phosphatase that is responsible for Ser7P dephosphorylation has not been identified.

These kinases and phosphatases collaborate to produce a variable and dynamic pattern of CTD phosphorylation. The use of phospho-specific CTD antibodies has improved our understanding of the pattern of CTD phosphorylation and its importance during transcription. Using the H5 and H14 antibodies for immunoprecipitation of Ser2P and Ser5P CTD, respectively, Ser5 phosphorylation was found at promoter regions and Ser2 phosphorylation was found within coding regions (Komarnitsky et al, 2000); however, the H5 antibody was subsequently shown to react with Ser2/5P CTD peptides more strongly than Ser2P, and H14 reacted equally well with Ser5P and Ser2/5P repeats (Jones et al, 2004). Improvements in the Ser5P-, Ser2P-, and Ser7P-specific antibodies and the implementation of genome-wide chromatin immunoprecipitation have improved our understanding of the dynamics of CTD phosphorylation. The consensus of such studies (represented in Fig. 1.2) is that there is a peak of Ser5 phosphorylation towards the 5’ end of genes that decreases as the polymerase elongates. Ser2P levels increase during elongation and remain until termination. On protein-coding genes, Ser7P levels increase with Ser5P levels but remain high throughout transcription (Fig. 1.2A); however, on non-coding genes, Ser2P levels seem to be overall lower, and Ser7P levels decrease as the polymerase elongates, similarly to Ser5P (Kim et al, 2010; Mayer et al, 2010; Tietjen et al, 2010) (Fig. 1.2B). While these studies give us a general overview of the phosphorylation of the CTD, there are
inherent problems with these techniques, such as epitope masking due to protein binding and the influence of other CTD modifications on the recognition properties of the phospho-specific antibodies. Moreover, there are clearly differences among the modification of the CTD of RNAPII as it transcribes different genes, which we do not yet fully understand.

A. Protein-coding Genes

Figure 1.2: A) For protein-coding genes, RNAPII (blue) is recruited to the promoter with an unphosphorylated CTD (IIA form). Upon preinitiation complex formation, the CTD is phosphorylated at the Ser5 (black) and Ser7 (red) positions. During elongation, an increase in Ser2 phosphorylation (green) produces the hyperphosphorylated form of the CTD, which is probably an ensemble of singly, doubly, and triply phosphorylated heptads.
Figure 1.2 (continued): As the polymerase elongates towards the 3' end of the gene, the activity of Ser5-specific phosphatases decreases the Ser5 phosphorylation levels, while the Ser2 and Ser7 phosphate levels remain largely unchanged. B) For non-coding genes, RNAPII is also recruited to the promoter with an unphosphorylated CTD (IIA form). Upon preinitiation complex formation, the CTD phosphorylation of Ser5 (black) and Ser7 (red) increases. During elongation, Ser2 phosphorylation (green) increases, while Ser7 phosphorylation begins to decline, presumably due to the activity of a yet unidentified Ser7 phosphatase. At the 3' end of the gene, the activity of Ser5-specific phosphatases decreases the Ser5 phosphorylation levels, Ser7 phosphorylation levels continue to decrease, and Ser2 phosphate levels remain largely unchanged.

1.3.2. Other CTD modifications

While phosphorylation is the most abundant and best-characterized modification, other CTD modifications have been observed. The CTD has been shown to be glycosylated but only if it is unphosphorylated, indicating that these modifications are mutually exclusive (Kelly et al, 1993); however, the extent of glycosylation and its function during transcription is currently unknown. Isomerization of the CTD peptidyl-proline bonds by Ess1 (Pin1 in humans) has been shown to play an important role in mediating protein binding. Although multiple studies have suggested that Ess1 in yeast and Pin1 in humans are involved in transcription and CTD phosphorylation (Krishnamurthy et al, 2009; Wilcox et al, 2004; Xu & Manley, 2007), most structures of CTD peptide-protein complexes revealed that the CTD proline residues are exclusively in the more energetically stable trans state. However, the Ser5-specific CTD phosphatase Ssu72 was found to bind the cis conformation of a Ser5-Pro6 motif within the heptad repeat (Werner-Allen et al, 2011; Xiang et al, 2010). In addition, the activity of the proline isomerase Ess1 facilitated the rapid dephosphorylation of the CTD by Ssu72 in vitro, suggesting that this cis/trans interconversion plays a role in regulating the phosphorylation state of the CTD (Werner-Allen et al, 2011). These findings indicate that proline isomerization increases the number of distinct CTD states, increasing the
possible number of CTD conformations. It is still unclear whether proline isomerization is a general property of RNAPII transcription or is gene specific (Werner-Allen et al, 2011).

1.4. CTD binding proteins

The essential role of the CTD (and by extension its modifications) is to provide a plastic binding surface for proteins involved in transcription and various transcription-related processes, including pre-mRNA capping, processing, and 3'-end formation and chromatin modification. These proteins display phospho-specific binding; therefore, as the phosphorylation state of the CTD changes during transcription, these factors are able to bind in a temporal manner. Accordingly, these proteins can be classified according to the stage in transcription at which they bind and the corresponding state of the CTD. It is clear that the CTD plays an important role in coupling mRNA-specific processes to RNAPII.

1.4.1. Early/initiation

While the phosphorylated forms of the CTD play an important role in protein recruitment (discussed more below), the nonphosphorylated form of the CTD is necessary for the association of general transcription factors and regulators at promoters. In particular, Mediator is a versatile and key regulatory complex that orchestrates transcription activation and pre-initiation complex formation at protein-coding genes (Baek et al, 2006; Kim et al, 1994). Moreover, it directly affects CTD phosphorylation by stimulating the activity of Kin28/CDK7 and associating with Srb10/CDK8 (Kim et al, 1994; Knuesel et al, 2009). Mediator specifically binds a CTD fusion protein that is unphosphorylated, and removal of the CTD by proteolysis abrogated the effect of Mediator on RNAPII basal transcription in an in vitro system.
(Myers et al, 1998). In addition to playing a role in recruiting Mediator, CTD phosphorylation has been shown to promote the release of Mediator from RNAPII, which then promotes productive elongation (Max et al, 2007). At snRNA genes, Mediator is required for Ser7 phosphorylation by CDK7, which then promotes the PCTD-mediated recruitment of the Integrator complex (Boeing et al, 2010; Egloff et al, 2007; Egloff et al, 2010). Mediator therefore uses the CTD as a binding scaffold to facilitate preinitiation complex formation, and then it promotes productive elongation and its subsequent release from RNAPII by stimulating CTD phosphorylation via Kin28/CDK7.

Pre-mRNA capping has also been found to be associated with the CTD and Ser5 phosphorylation. The 5’ 7-methylguanosine cap is specific to eukaryotic mRNAs and is thought to stabilize transcripts and play a role in translation (Furuichi et al, 1977; Gallie, 1991). The process of capping occurs cotranscriptionally when the transcript reaches about 20-30 nucleotides in length (Coppola et al, 1983; Rasmussen & Lis, 1993). The enzyme responsible for adding the 5’ cap consists of an RNA triphosphatase, which removes the terminal phosphate on the 5’ nucleotide; an RNA guanylyltransferase, which uses GTP to conjugate guanosine monophosphate onto the remaining diphosphate of the nascent pre-mRNA; and an RNA guanine-7-methyltransferase, which then methylates the N7 position of the guanine. In yeast, the RNA triphosphatase and guanylyltransferase activities are encoded within two distinct polypeptides (Cet1 and Ceg1) (Ho et al, 1998; Itoh et al, 1987), while in mammals a single multifunctional polypeptide performs both functions (Mce1) (Yagi et al, 1983). In mammals, truncation of the CTD decreases the ratio of capped vs. uncapped mRNAs; moreover, Mce1 and Ceg1 were shown to interact with the phosphorylated form of the CTD (Cho et al, 1997; McCracken et al, 1997a). Interestingly, the guanylyltransferase domain of Mce1 (aa 211-
binds to both Ser2P- and Ser5P-containing peptides; the addition of Ser5P CTD peptide stimulated the activity of the guanylyltransferase, but the Ser2P peptide did not (Ho & Shuman, 1999). ChIP experiments have also shown that association of capping enzyme with RNAPII on an active gene coincides with the peak of Ser5 phosphorylation (Komarnitsky et al., 2000; Mayer et al., 2010). The requirement of the CTD in capping enzyme recruitment and activation explains why only RNAPII transcripts are capped in vivo. In addition, these results indicate that the CTD contributes to both localization and allosteric stimulation of capping enzyme.

Phosphorylation of the CTD early in transcription also promotes subsequent kinase activity. The elongation kinase Bur1 contains a C-terminal phosphoCTD interaction domain (PCID) and binds to Ser5P heptad repeats in vitro. In addition, the use of a temperature sensitive allele of Kin28 demonstrated that disruption of Kin28 activity decreased recruitment of Bur1 by ChIP (Qiu et al., 2009). Thus, Ser5 phosphorylation of the CTD potentially signals the completion of initiation and promotes further CTD modification and the transition into elongation by binding Bur1.

1.4.2. Elongation

The number of proteins that bind to the CTD during elongation is likely much higher than those present during initiation and termination, as elongation is the longest phase of transcription. A proteomics screen has shown that at least a hundred proteins could bind to the CTD of elongating RNAPII based on the Ser2/5P mark; these potential PCTD associating proteins (PCAPs) are responsible for a wide range of processes, including transcription, RNA processing, chromatin modification, DNA metabolism, protein synthesis, and others (Phatnani et al., 2004). While there are likely many uncharacterized and unvalidated PCAPs, a few elongation-associated proteins have been demonstrated to bind directly and specifically to Ser2/5P CTD repeats.
The connection between mRNA processing and the PCTD was first demonstrated by the colocalization of RNAPII0 and U1 snRNP 70K protein on Drosophila polytene chromosomes (Weeks et al, 1993). In mammals, truncation of the CTD produces defects in mRNA processing (McCracken et al, 1997b). Multiple yeast two-hybrid screens have demonstrated that Serine-Arginine (SR) domain-containing proteins, which are known to be involved in mRNA processing, interact with the PCTD (Bourquin et al, 1997; Yuryev et al, 1996). Moreover, the addition of an anti-CTD antibody or CTD peptides to in vitro splicing reactions is sufficient to block the mRNA splicing (Yuryev et al, 1996). In another study, deletion of the catalytic subunit of CTDK-I (ctk1) in yeast produced a splicing defect following return from starvation in vivo (Phatnani et al, 2004).

One protein that has been shown to connect the CTD to the splicing apparatus is Prp40, a bridging factor that binds the U1 snRNP and branch point binding protein. Prp40 contains a WW domain, which is one of the few known CTD binding domains, that specifically interacts with the hyperphosphorylated form of the CTD and with peptides containing Ser2/5P CTD heptads (Morris & Greenleaf, 2000; Phatnani et al, 2004). Likewise, a related mammalian putative splicing factor, CA150 (aka TCERG1), has also been found to bind the hyperphosphorylated form of the CTD (Carty et al, 2000); while the function of CA150 is still not clear, knockdown experiments showed that it is required for mRNA processing and may influence alternative splice site choice (Pearson et al, 2008).

Recently, a connection was also found between the CTD and U2AF, which recognizes the 3’ splice site of pre-mRNA. U2AF consists of U2AF65, which binds polypyrimidine tracts proximal to the 3’ end of introns, and U2AF35, which binds the conserved AG dinucleotide at the 3’ end of the intron (Wu et al, 1999; Zamore & Green, 1989). Fusion of the CTD to the SR protein ASF/SF2 and addition of cofactors within the
20-40% ammonium sulfate fraction of mammalian nuclear extract (NF20-40) stimulated splicing of pre-mRNA substrates with weak polypyrimidine tracts, and CTD phosphorylation was required for this effect (Millhouse & Manley, 2005). Through further separation of NF20-40, Prp19 and U2AF65 were found to be responsible for the CTD-dependent splicing of the weak polypyrimidine tract-containing substrate; moreover, U2AF65 was able to bind to the PCTD in vitro (David et al, 2011).

Based on these interactions, an interesting model has been proposed in which SR proteins tether the 5’ splice site to the CTD as the intron is transcribed. After transcription of the 3’ splice site, it will also be tethered to the CTD by U2AF; branch point binding protein will then bind to both the branch point and the tethered 5’ splice site-containing complex, thereby promoting lariat formation and efficient splicing of the 5’ and 3’ ends of the neighboring exons in association with the PCTD (Greenleaf, 1993; Morris & Greenleaf, 2000). In this way, the PCTD mechanistically contributes to splicing and possibly splice site choice; however, much of this model remains to be tested.

During elongation, the structure of the chromatin is also altered, and particular modifications mark the coding regions of active genes (Bernstein et al, 2002; Santos-Rosa et al, 2002). These modifications are thought to create a “memory” of recent transcriptional events (Ng et al, 2003). Chromatin methylation during transcription has been linked to the CTD of RNAPII through the histone methyltransferase Set2. Set2 methylates histone H3 lysine 36 (Strahl et al, 2002), and both Set2 and H3 K36 methylation are found throughout the coding regions of genes (Krogan et al, 2003; Xiao et al, 2003). Set2 copurifies with the phosphorylated form of RNAPII in both yeast and mammals (Krogan et al, 2003; Schaft et al, 2003; Xiao et al, 2003). Truncation of the CTD or deletion of the catalytic subunit of CTDK-I abrogates H3 K36 methylation (Xiao et al, 2003). Set2 has been shown to directly bind the Ser2/5P CTD through its C-terminal Set2
Rpb1 interaction (SRI) domain, and deletion of the SRI domain leads to a loss of H3 K36 methylation within active genes (Kizer et al, 2005). In this way, H3 K36 methylation is coupled to active transcription through the interaction of Set2 with the PCTD.

Finally, a handful of elongation-related proteins have been shown to interact with the hyperphosphorylated CTD \textit{in vitro}; however, the biological importance of these interactions is still not known. Hrr25, a kinase involved in DNA damage responses, and Ssd1, a member of the ribonuclease II family, have been shown by BIAcore analysis to bind to peptides containing Ser2/5P and Ser2/5P or Ser2P, respectively (Phatnani et al, 2004). The stress-related Hog1 kinase interacts with Rpb1 in a CTD phosphorylation-dependent manner and contranscriptionally associates with target genes during osmotic stress (Proft et al, 2006); however, it is unclear whether it directly interacts with the PCTD. In collaboration with the Janscak lab, we have shown that human RecQ5β, which is a putative “anti-recombinase,” is associated with active genes and binds directly to the hyperphosphorylated CTD (Kanagaraj et al, 2010). As more putative PCAPs are identified and characterized, the list of transcription elongation-associated processes will surely continue to expand.

\textbf{1.4.3. Termination}

During the final stage of transcription, the pre-mRNA must be cleaved and polyadenylated. In addition to releasing the mRNA from RNAPII, cleavage and addition of the poly(A) tail increases the stability of the transcript and plays a role in translation (Gallie, 1991). Truncation or deletion of the CTD results in a decrease in the efficiency of polyadenylation and affects the selection of polyadenylation site \textit{in vivo} (Hirose & Manley, 1998; Licatalosi et al, 2002), indicating that the factors responsible for 3' end processing are CTD associated. In yeast, these processes are mediated by the cleavage
factor complexes CFI (which can be subdivided into CFIA and CFIB) and CFII,
polyadenylation factor PFI, and poly(A) polymerase, Pap1 (Chen & Moore, 1992).
Deletion of the catalytic subunit of CTDK-I in yeast alters polyadenylation site selection
(Skaar & Greenleaf, 2002) and decreases the association of the 3’ end processing factors
Rna14 and Rna15 in CFIA and Cft1 and Cft2 in CFII at active genes (Ahn et al, 2004). In
the CFIA complex, two proteins, Rna15 and Pcf11, have been shown to directly bind the
PCTD of RNAPII (Barilla et al, 2001). Pcf11 cotranscriptionally associates at the 3’ end of
the coding regions of genes and binds to the PCTD predominately phosphorylated on
Ser2 (Licatalosi et al, 2002). Moreover, in mammals, the CstF p50 subunit of the CstF
complex (which is analogous to the CFIA complex in yeast) has been shown to bind to
the PCTD through a domain at its N-terminus (Fong & Bentley, 2001; McCracken et al,
1997b). In addition to cleavage and polyadenylation, processing of non-polyadenylated
messages (such as snoRNAs) is coupled to the PCTD through the hnRNP-like protein
Nrd1, which specifically binds Ser5P-containing peptides (Conrad et al, 2000; Steinmetz
et al, 2001; Vasiljeva et al, 2008; Yuryev et al, 1996). Based on these studies, the PCTD of
RNAPII plays a clear role in both the recruitment and function of RNA 3’ end processing
factors.

1.5. **Nucleocytoplasmic transport of mRNA**

The compartmentalization of eukaryotic cells requires pathways that direct and
mediate the transport of macromolecules across intercellular membranes. For gene
expression to successfully occur, mRNAs, which are transcribed in the nucleus, must be
translocated to the cytoplasm to be translated into polypeptides. Interestingly, there are
specific pathways that are responsible for the nuclear import and export of proteins and
nucleic acids. While nuclear export of most RNA species relies on the karyopherin-β Ran pathway, mRNAs are exported through a Ran-GTP independent pathway that involves a specific set of conserved export receptor and adaptor proteins. There appears to be one universal receptor for mRNA export, Mex67/Mtr2 in yeast and TAP/p15 in mammals, which interacts with the mature mRNP and the Nuclear Pore Complex (NPC) to facilitate export. The export receptor functions in conjunction with the export adaptor proteins, which cotranscriptionally associate with the nascent mRNA. There are two main mRNA export adaptor proteins, Yra1/ALY in the transcription and mRNA export (TREX) complex and Sac3 in the TREX-2 complex. It is not clear whether these pathways have distinct cargo or functions. Export mediated by TREX-2 and Sac3 has been shown to be coupled to chromatin modification through Sus1, a common factor in both TREX-2 and the Spt–Ada–Gcn5 acetyltransferase (SAGA) complex (Lei et al, 2003; Rodriguez-Navarro et al, 2004). The Yra1/ALY-mediated pathway is described in detail below.

1.6. Yra1 and mRNA export

YRA1 was first identified as a gene that induces a cell cycle arrest when overexpressed in yeast (Espinet et al, 1995). It was first purified based on its ability to anneal complementary single stranded RNAs (Portman et al, 1997), which is a characteristic of heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) (Kumar & Wilson, 1990). Yra1 is an essential gene that contains the second longest intron of all S. cerevisiae genes (766 nt). It is predominantly localized to the nucleus; however, the identity of its nuclear localization signal (NLS) and whether an NLS is required for its localization is unknown (Portman et al, 1997; Strasser & Hurt, 2000; Zenklusen et al, 2001).
The RNA binding activity of Yra1 was attributed to the observation that it contained an RNA Recognition Motif (RRM; also known as the RNA-binding domain and the ribonucleoprotein domain), which is the best-characterized and most abundant RNA binding structure in proteins. RRMs are characterized by two consensus sequences called RNP1 (Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe/Tyr-Val/Ile/Leu-X-Phe/Tyr) and RNP2 (Ile/Val/Leu-Phe/Tyr-Ile/Val/Leu-X-Asn-Leu) (Maris et al, 2005), which form the face of a four-stranded beta sheet (Clery et al, 2008). While the RNP2 sequence of Yra1 matches the consensus, the RNP1 motif is more divergent. Little is known about the three-dimensional structure of Yra1. Partial structures of the metazoan REF proteins have been determined, but these structures consisted of little more than the RRM regions (Golovanov et al, 2006; Perez-Alvarado et al, 2003).

The role of Yra1 in mRNA export was found through its genetic interaction with Mex67, which was known to associate with nuclear pores and polyadenylated RNA. As YRA1 is essential in yeast, a GAL1::GFP-YRA1 allele was used to deplete Yra1 from growing yeast. In addition to stopping growth, depletion of Yra1 also induced nuclear accumulation of polyadenylated (polyA⁺) RNA based on fluorescence in situ hybridization, which is a hallmark of an export defect (Strasser & Hurt, 2000). In addition, a temperature sensitive allele, yra1-1 (which contained five point mutations – S42G, V80A, G137E, K210E and F223S), also displayed nuclear accumulation of polyA⁺ RNA at 37°C. Using both pull downs and purified proteins, Yra1 was found to physically interact with Mex67 (Strasser & Hurt, 2000; Stutz et al, 2000). Based on the effects of Yra1 conditional mutants and its physical and genetic interactions with Mex67, it was defined as an mRNA export factor.

The expression of YRA1 has been extensively studied. Increasing the copy number of the gene through the use of a 2µ plasmid does not increase the protein level of Yra1;
however, deletion of the 766-nt intron ($\Delta IVS$ or $yra1-\Delta i$) increases the expression level of Yra1 protein from a low copy plasmid. The increase in Yra1 expression is detrimental, as the $\Delta IVS/yra1-\Delta i$ strain grows slower and displays nuclear accumulation of polyA$^+$ RNA compared with WT (Preker et al., 2002; Rodriguez-Navarro et al., 2002). This negative feedback regulatory mechanism has been shown to be related to splicing of the transcript, and three factors that affect splicing, and therefore Yra1 expression regulation, have been identified: the length of the intron, the noncanonical branchpoint, and the C-terminus of the Yra1 protein (Preker & Guthrie, 2006). The current model for the autoregulation of YRA1 expression involves export of the excess unspliced pre-mRNA and degradation in the cytoplasm by the 5’ to 3’ pathway after decapping by the Edc3 decapping activator and the Dcp1/Dcp2 decapping enzyme complex (Dong et al., 2007).

Interestingly, Zenklusen et al. (2001) found that a second YRA1-like gene, which they designated YRA2, is able to rescue the YRA1 deletion if overexpressed, although YRA2 itself is nonessential. It did appear to be present in RNPs pulled down with either Mex67 or Yra1. It has been proposed that Yra2 is a redundant protein, but due to a lack of study, this has not been demonstrated, and Yra2 remains a protein of unknown function.

To identify important and potentially functional regions of Yra1, regions of Yra1 have been isolated and evaluated based on three aspects of Yra1 function: ability to support viability, ability to promote RNA export, and ability to bind RNA. Surprisingly, the entire RRM could be deleted with only minor effects on growth and mRNA export at 37°C. Deletion of both termini (using constructs encompassing aa 14-210 and 14-167) resulted in a loss of viability. Deletion of only the N-terminus before the RRM or the C-terminus after the RRM (aa 77-227 and 1-167, respectively) causes only slow growth and
slight export defects at 37°C. Finally, all regions of Yra1 seemed to have affinity for an RNA probe, except for the isolated RRM (Zenklusen et al, 2001). These results indicated that the RRM of Yra1 might not be involved in RNA binding or be functionally important. However, this interaction was not analyzed in the context of the full-length Yra1 protein.

Using chromatin immunoprecipitation (ChIP), many groups have found Yra1 associates with actively transcribed genes. In general, Yra1 is present at low levels near the 5′-ends of genes. During transcription elongation (i.e., in the middle of the gene), the level of Yra1 peaks; this level stays relatively constant until the end of the gene (Abruzzi et al, 2004; Johnson et al, 2009; Lei et al, 2001; Zenklusen et al, 2002). While this pattern is consistent with an indirect interaction between Yra1 and the transcription apparatus via the nascent RNA, treatment of extracts with RNase prior to ChIP did not significantly affect the Yra1 signal, indicating that Yra1 is likely tethered to the polymerase during elongation by a protein-protein interaction (Abruzzi et al, 2004).

Yra1 has been found to be associated both genetically and physically with a set of proteins collectively known as the Transcription Export (TREX) complex, which includes the THO complex of transcription elongation factors (Hpr1, Tho2, Thp2, and Mft1), the Sub2 RNA helicase, and the uncharacterized protein Tex1 (Strasser et al, 2002; Zenklusen et al, 2001). Deletion of the nonessential THO genes HPR1, THO2, MFT1, and THP2 resulted in slight mRNA export defects (Strasser & Hurt, 2000). In addition, some studies have shown that Δhpr1 displays a decrease in the levels of both Sub2 and Yra1 occupancy on certain genes by ChIP, indicating that the absence of Hpr1 (and TREX) results in a loss of Yra1 recruitment (Zenklusen et al, 2002); however, as Δhpr1 has been shown to affect transcription elongation, this decrease in Yra1 and Sub2 may be due to
the elongation defect rather than a direct effect of Hpr1 on the export proteins (Rondon et al, 2003). As a complex, TREX is thought to facilitate the coupling of mRNA transcription and export through its association with Yra1. In addition, ChIP of Yra1 in conditional mutants of sub2 showed that Yra1 recruitment was not decreased in the absence of Sub2 (Johnson et al, 2009). In the currently accepted model of Yra1-mediated mRNA export, the TREX complex is thought to recruit Yra1 to nascent mRNAs cotranscriptionally. Yra1 then binds to and packages the mRNA (along with many other mRNA binding proteins (Hogan et al, 2008; Oeffinger et al, 2007)) for export via Mex67/Mtr2 and the Nuclear Pore Complex (NPC) (Fig. 1.3).
Figure 1.3: Model of Yra1-mediated mRNA export in yeast. Yra1 (purple) cotranscriptionally associates with RNAPII on active genes through interaction with the THO complex (green) and Sub2 (orange). Yra1 is then poised to dissociate from TREX and associate with the nascent mRNA (pink), which is packaged by many mRNA binding proteins (peach) in addition to Yra1 to create the messenger ribonucleoprotein (mRNP). When the mRNP reaches the Nuclear Pore Complex (NPC), Yra1 interacts with the Mex67/Mtr2 receptor complex (grey) to facilitate nucleocytoplasmic transport of the mRNP.

Yra1 is thought to be a general or bulk mRNA export adaptor protein because it is essential for viability and conditional alleles display severe export defects. However, studied have shown that Yra1 does not associate to all genes or bind to all RNAs. As the mammalian homologue ALY has been shown to be associated with the Exon Junction Complex (Le Hir et al, 2000; Lejeune et al, 2002; Tange et al, 2005), Abruzzi et al. (2004) hypothesized that Yra1 may be important for export of spliced transcripts. By comparing the occupancy of Yra1 on a larger set of genes using ChIP, Yra1 seemed to be associated with some but not all intronless and intron-containing genes, indicating that splicing did not define the mRNAs exported by Yra1 (Abruzzi et al, 2004). Moreover, depletion of Yra1 protein does not affect pre-mRNA splicing in *S. cerevisiae* (Wang &
Rymond, 2003); therefore, the recruitment of Yra1 to a nascent mRNA does not appear to require splicing in yeast. Yra1 has been shown to be dispensable for export of heat shock mRNAs (Rollenhagen et al, 2007). RNA immunoprecipitation (RIP) with Yra1 was used to characterize the RNAs that Yra1 binds (and therefore exports). Yra1 bound and co-immunoprecipitated 1002 transcripts, which is 16.7% of the S. cerevisiae genome (Hieronymus & Silver, 2003). While this is consistent with the idea that Yra1 is not responsible for exporting all transcripts, a consensus Yra1 binding sequence was not identified from these data; therefore, it is still unclear what directs Yra1 to certain mRNAs for export.

In addition to the TREX complex, Yra1 has been shown to interact with other transcription-related proteins. Yra1 interacts genetically and physically with Rrp45 and genetically with Rrp6, subunits of the exosome complex; this interaction indicates that Yra1 plays a role in coupling mRNA export to turnover (Zenklusen et al, 2002). In addition, Yra1 physically interacts with 3’-end processing factor Pcf11 in vitro; using a conditional mutant of pcf11, Yra1 cotranscriptional association decreased at the nonpermissive temperature (Johnson et al, 2009). These results are consistent with observations that defects in 3’-end formation inhibit mRNA export (Lei & Silver, 2002).

yra1-1 genetically interacts with the transcription factors spt4D, spt6-140, paf1Δ, and spt16-197 (Burckin et al, 2005); its mammalian homologue, ALY, has also been shown to interact with Spt6 through a physical interaction with the scaffold factor Iws1 (Yoh et al, 2007).

Finally, a recent study on the Prp19 complex, which plays an essential role in splicing, has shown that this splicing factor might affect Yra1 recruitment. Deletion of the Syf1 subunit of Prp19 decreases Yra1 recruitment to intron-less and intron-containing genes by up to 20% during elongation. However, deletion of syf1 also
seemed to affect transcription processivity of RNAPII, so it is unclear whether the effect on Yra1 is due to a direct interaction with Prp19 or an indirect effect based on the processivity defect (Chanarat et al, 2011).

Because the mRNA export pathway is distinct from the other RNA export pathways and requires the cotranscriptional recruitment of adaptor proteins, the proteins involved in mRNA export are candidate PCAPs. In previous work from our lab, Yra1 was identified as a potential phosphoCTD associating protein (PCAP) using a systematic proteomics-based screen (Phatnani et al, 2004). Based on this observation, the aims of this dissertation were to determine whether Yra1 directly interacts with the CTD, characterize the interaction, identify the PCTD interaction domain in Yra1, isolate mutations and truncations that affect PCTD binding, and explore the biological role of the PCTD in Yra1-mediated mRNA export.
2. Characterization of Yra1 binding interactions in vitro

2.1. Introduction

The carboxyl-terminal domain (CTD) of RNA polymerase II (RNAPII) is a conserved unstructured repeat domain that extends from the largest subunit of RNAPII, Rpb1. The consensus heptapeptide sequence of the CTD, YSPTSPS, is conserved in eukaryotes, although the number of heptapeptide repeats varies (from 26 repeats in budding yeast up to 52 repeats in mammals). The CTD is essential for viability; however, the essential function has been shown to be independent of the catalytic activity of RNAPII (Bartolomei et al, 1988; Nonet et al, 1987; Zehring et al, 1988). The CTD acts as a binding platform for factors involved in mRNA transcription-associated processes, including, but not limited to, chromatin remodeling and mRNA capping, splicing, and 3′-end processing; for review, see (Buratowski, 2009; de Almeida & Carmo-Fonseca, 2008; Meinhart et al, 2005; Pandit et al, 2008; Phatnani & Greenleaf, 2006).

The CTD is subject to reversible post-translational modification, namely phosphorylation, and the extent and pattern of phosphorylation have been shown to change during the course of transcription of protein-coding genes. The CTD is hyper-phosphorylated (referred to as the RNAPIIO form) during the elongation phase of transcription, with serines 2 (Ser2) and 5 (Ser5) of each repeat the primary phoshoacceptors. In addition to Ser2 and Ser5 phosphorylation, recent studies have also demonstrated Ser7 phosphorylation is present on transcribing RNAPII in yeast and mammals (Chapman et al, 2007; Kim et al, 2009). ChIP and ChIP-chip experiments with CTD phospho-specific antibodies have shown that CTD phosphorylation varies across different genes: Ser5 phosphorylation peaks at the 5′-end of genes and the peak of Ser2
phosphorylation follows and persists to the 3’-end of genes (Kim et al, 2010; Komarnitsky et al, 2000; Mayer et al, 2010; Tietjen et al, 2010).

Four kinases and three phosphatases collaborate to modulate CTD phosphorylation in yeast. Ser5 of the CTD is phosphorylated by Kin28 (CDK7) of TFIIH and Srb10 (CDK8) of Mediator during transcription initiation (Hengartner et al, 1998). More recently, Kin28 (CDK7) was also shown to be responsible for serine 7 phosphorylation in yeast and mammals (Akhtar et al, 2009; Kim et al, 2009). The other two kinases, Ctk1 (CDK12) of CTD Kinase I (CTDK-I) and Bur1 (CDK9), have been genetically and functionally linked to transcription elongation. Ctk1 genetically interacts with multiple elongation factors (Jona et al, 2001; Lindstrom & Hartzog, 2001) and increases transcription elongation efficiency \textit{in vitro} (Lee & Greenleaf, 1997). In addition, aberrant CTD phosphorylation has been found in \textit{ckt1} mutants (Lee & Greenleaf, 1991). CTDK-I can phosphorylate Ser2 and Ser5 \textit{in vitro} and prefers a substrate that is already phosphorylated on one serine (Jones et al, 2004). Bur1/2 was recently found to be a PCTD binding protein as well as a CTD kinase. Bur1/2 binds to Ser5P CTD, as produced by Kin28, and phosphorylates Ser2 near promoters (Qiu et al, 2009). In addition, Bur1/2 has been shown to replace Ser7 phosphates during elongation (Tietjen et al, 2010). Finally, the yeast phosphatases Fcp1, Ssu72, and Rtr1 have been shown to remove phosphates from CTD substrates (Mosley et al, 2009).

The synergistic action of these enzymes produces distinct binding surfaces on the CTD that recruit factors during the stage of transcription at which they are required. While interactions between the PCTD and many mRNA processing factors have been characterized, PCTD-mediated association of mRNA packaging and export factors has not been demonstrated. It is likely that packaging and export are also coupled to transcription through CTD interactions, as defects in transcription elongation, splicing,
and 3'-end processing have been shown to affect export (Brodsky & Silver, 2000). In a screen of yeast proteins that potentially bind to a CTDK-I–hyperphosphorylated CTD fusion protein, only one known mRNA export factor was identified, Yra1 (Phatnani et al, 2004).

Yra1 (REF/ALY in mammals) has been described as a general mRNA export factor because it genetically and physically interacts with the mRNA export receptor Mex67/Mtr2 (TAP/NXF1 in mammals) and depletion of Yra1 protein or the use of a temperature sensitive allele induces nuclear retention of polyadenylated RNA (Strasser & Hurt, 2000; Stutz et al, 2000; Zenklusen et al, 2001). While only a few targets of Yra1-mediated mRNA export are known, it is clear that Yra1 is not recruited to all loci and is not required for export of all mRNAs (Abruzzi et al, 2004; Hieronymus & Silver, 2003; Rollenhagen et al, 2007). However, there is evidence that distinct subpopulations of functionally similar transcripts are exported by a given export adaptor (Hieronymus & Silver, 2003; Hogan et al, 2008) and that a given mRNA likely has binding sites for many export adaptors and packaging factors (Hogan et al, 2008).

Yra1 is thought to exist as part of a conserved complex referred to as the TRanscription EXport complex (TREX), which is made up of the THO elongation complex, the Sub2 RNA helicase, Yra1, and Tex1 (Strasser et al, 2002). The TREX complex members cotranscriptionally associate with actively transcribed genes during elongation, but the mechanism that directs recruitment is unclear (Abruzzi et al, 2004; Zenklusen et al, 2002). Yra1 and ALY have also been shown to interact with other transcription factors, including Iws1/Spn1 (Yoh et al, 2007) and Pcf11 (Johnson et al, 2009), and these proteins have also been hypothesized to recruit the export factor.

In previous work from our lab, Yra1 was identified as a potential phosphoCTD associating protein using a systematic proteomics-based screen (Phatnani et al, 2004). To
test the idea that Yra1 binds the phosphoCTD (PCTD) directly and to examine its poorly characterized RNA-binding properties, I purified and analyzed recombinant constructs of Yra1 that represent either the complete protein or defined segments. I identified both a minimal PCTD-binding segment and an overlapping but distinct RNA-binding segment. Mutations in the overlapping region affected its RNA binding activity but did not affect PCTD binding. In addition, a hyperphosphorylated CTD peptide was unable to compete with RNA for binding to Yra1, indicating that the RNA and PCTD interactions are distinct. These results indicate that the PCTD could play an important role in recruiting Yra1 to nascent mRNAs.

2.2. Materials and methods

2.2.1. Reagents

The biotinylated CTD peptides were synthesized by Anaspec. All enzymes used for cloning were purchased from New England Biolabs. All Gateway cloning reagents were purchased from Invitrogen. Precast gels and nitrocellulose were from Bio-Rad. Chemicals, including protease inhibitor cocktails, were purchased from Sigma-Aldrich.

2.2.2. Cloning

For expression of full-length Yra1 and truncations used in the Far Western assay, the YRA1 cDNA was expressed in bacteria as Maltose Binding Protein (MBP) fusions using the pMal vector (NEB). All plasmids were sequenced before transformation into BL21(DE3)RIL cells (Promega) for expression. For expression in baculovirus-infected Spodoptera frugiperda Sf9 cells, the YRA1 cDNA was inserted into the pFastBac-HTb vector with the Bac-to-Bac system (Invitrogen). To increase expression in bacteria, the amino acid sequence of Saccharomyces cerevisiae YRA1 was obtained from the
Saccharomyces Genome Database (www.yeastgenome.org) and used to design a synthetically codon optimized gene for expression in *E. coli* (Integrated DNA Technologies). The synthetic gene was cloned into the pET15b vector (Novagen), and full-length Yra1, truncations, and point mutants were expressed as His tag fusion proteins in BL21(DE3) cells (Promega). Point mutations were created using the QuikChange Site-Directed Mutagenesis system (Stratagene) according to the manufacturer’s instructions. Mutants were verified by sequencing and transformed into BL21(DE3) cells (Promega) for protein expression.

### 2.2.3. Protein expression and purification

Recombinant GST-tagged yeast CTD (GSTyCTD) was expressed in *E. coli* and purified as described previously (Morris et al, 1997). Yra1 MBP fusions were expressed in BL21(DE3)RIL cells and purified using Amylose resin (New England Biolabs) according to the manufacturer’s protocol.

His-tagged full-length Yra1, truncations and mutants were expressed in BL21(DE3) cells grown in Luria-Bertani (LB) medium with antibiotics. Protein expression was induced with a final IPTG concentration of 0.5 mM for 4 hours. Cells were lysed in 50 mM Tris-HCl, pH 8.0 and 500 mM NaCl buffer with Protease Inhibitor Cocktail for General Use (Sigma) by two passages though a cell cracker (Microfluidics), and the cell debris was pellet by centrifugation. Imidazole (20 mM) was added to the soluble extract to decrease background binding, and the extract was incubated in batch with Ni Sepharose Fast Flow Resin (GE Healthcare) for 30 min at 4°C with rotation. The resin was collected and washed with >10 column volumes of high salt wash buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 20 mM imidazole). The protein was eluted with 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 250 mM imidazole, and 8% glycerol. Elutions from
the Ni Sepharose were pooled and applied to an Akta FPLC system equipped with a Mono S 5/5 HR column (GE Healthcare) at a flow rate of 1 ml/min with 50 mM Tris-HCl, pH 8.0, 8% glycerol, and 0.1 mM EDTA as Buffer A and 50 mM Tris-HCl, pH 8.0, 8% glycerol, 0.1 mM EDTA, and 1 M NaCl as Buffer B. A 40 column volume gradient from 0.35 to 1.0 M NaCl was used to elute proteins, and fractions containing Yra1 were identified by SDS-PAGE and pooled. The purified protein was exchanged into storage buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 8% glycerol) and concentrated to approximately 1-2 µg/mL with microcentrifuge concentrators; aliquots of the purified protein were stored at -80°C until use.

CTDK-I and Yra1 were expressed in baculovirus-infected *S. frugiperda* (Sf9) cells at a multiplicity of infection of 1 for 72 hours at 27°C. Cells were resuspended in cytoplasmic lysis buffer (10 mM HEPES, pH 8.0, 320 mM sucrose, 3 mM calcium chloride, 2 mM magnesium acetate, and 0.5% NP-40) with Protease Inhibitor Cocktail for Tissue Extracts (PIC; P8340, Sigma), and the solution was centrifuged 1500 g for 15 min to pellet the nuclei. Because CTDK-I and Yra1 are predominantly nuclearly localized, the pellet was resuspended in nuclear resuspension buffer (20 mM HEPES, pH 8.0, 1.5 mM MgCl₂, 20 mM KCl, 25% glycerol, and PIC). Nuclei were lysed by the addition of 12 mL nuclear extraction buffer (80 mM HEPES, pH 7.6, 1.2 M NaCl, 20% glycerol, 0.06% Triton X-100 and PIC). One volume of PEG solution (18% PEG 8000, 1 M NaCl, and protease inhibitor cocktail) was then added, and the solution was incubated at 4°C for 45 min with rotation. Chromatin and debris were removed by centrifugation at 20,000 g for 30 min. The soluble lysate containing the protein of interest was then diluted to adjust the salt concentration to around 300 mM, and imidazole was added to a final concentration of 10 mM. His-tagged rCTDK-I and Yra1 were purified by Ni affinity and ion exchange
chromatography as described above, except lower salt conditions were used for CTDK-I (300 mM NaCl in lysis and wash buffers and 100 mM NaCl in elution buffer). CTDK-I was eluted from the Mono S column with 60% B, which was confirmed by kinase assays. Fractions with kinase activity were stored at -80°C until use.

2.2.4. Far Western assay

Far Western assays were performed as described previously (Phatnani & Greenleaf, 2004). Briefly, duplicate samples of purified proteins were separated on a 4-20% precast SDS gel. One half of the gel was stained with Coomassie Brilliant Blue and the second half was transferred to nitrocellulose membrane with 0.75 Amps for 2 hours at 4°C. The membrane was incubated overnight at 4°C in blocking/renaturation buffer, containing 1X PBS, 3% nonfat dry milk, 0.2% Tween 20, 0.1% PMSF, and 5 mM NaF, with 2 mM dithiothreitol. The nitrocellulose was then probed with ≥300,000 cpm [32P]-labeled GSTyCTD fusion protein that had been hyperphosphorylated with rCTDK-I for 4 hours at 4°C. After extensive washing, the nitrocellulose was air-dried and exposed to X-ray film (Kodak) for 1-2 days and developed. The stained gel was scanned with a LiCor Odyssey Scanner (LiCor), and the film was scanned with a document scanner.

2.2.5. Immobilized CTD peptide binding assay

Synthetic biotinylated peptides were dissolved in PBS and incubated with approximately 300 µL of TetraLink Tetrameric Avidin Resin (Invitrogen) for 45 min at room temperature. The amounts of peptide in the column output, flowthrough and washes were monitored by absorbance at 280 nm, and these values were used to approximate the amount of peptide on the column. In general, 30-50 µg of biotinylated
peptide was conjugated to the 300-µL column. The peptide columns were stored in PBS with sodium azide at 4°C and were stable for at least 2 months.

About 20 µg of purified Yra1 and 5 µg of bovine serum albumin (BSA) were dissolved in PBS for a final volume of 500 µl as the output; 450 µl of the output was applied to the peptide resin and incubated for 20 min at room temperature with mixing every 5 min. The flowthrough and two washes with half-column volumes of PBS were collected. The resin was then extensively washed with 5 ml of buffer containing 25 mM HEPES, pH 7.6, 8% glycerol, 150 mM NaCl, and 0.1 mM EDTA (HGNE150). The bound protein was then eluted with four half-column volumes of HGNE300 and four half-column volumes of HGNE1000. The resin was regenerated with 5 ml HGNE1000 and 5 ml PBS.

2.2.6. UV crosslinking

The UV crosslinking assay was similar to that described previously (Hung et al, 2010). A synthetic oligonucleotide RNA (CAGUCGCAUAGUGCA; IDT) was end-labeled with γ-32P ATP and used for binding experiments. The labeled RNA was mixed with 1 µg of protein in crosslinking buffer (15 mM HEPES, pH 7.8, 75 mM NaCl, 0.1 mM KCl, 0.2 mM EDTA, 5 mM MgCl₂, 0.05% Tween 20, and 10% glycerol). Due to the salt in the protein solution (300 mM), the final NaCl concentration in the binding reaction was 90 mM. The reaction mixture was exposed to 254 nm light for 5 min on ice in a Stratalinker. The samples were run on a 4-20% precast gel, which was stained with Coomassie blue and dried. The dried gel was exposed to a Phosphorimager (GE Healthcare) and scanned.
2.3. **Results**

2.3.1. **Expression and purification of recombinant Yra1**

To verify that the previously observed interaction of Yra1 with the hyperphosphorylated CTD was direct (Phatnani et al., 2004), I expressed the *YRA1* cDNA or a codon-optimized synthetic construct as a recombinant six histidine-tagged protein in *E. coli* and purified it under high salt conditions over two columns (Fig. 2.1). The use of high salt facilitated the gentle removal of the majority of copurifying RNA species. The total protein lysate was first incubated in batch with an immobilized metal affinity chromatography (IMAC) resin to purify the His-tagged Yra1 protein by affinity. While the majority of the protein that eluted from the IMAC step was full length Yra1 (indicated with an arrow), there were many low molecular weight degradation products or contaminating proteins (indicated with asterisks in Fig. 2.1). Based on the positive charge of Yra1 (pI = 11.9), the protein eluted from the IMAC resin was further purified by ion exchange chromatography on a Mono S column. After salt elution from the Mono S, many of the small proteins from the IMAC were separated from full-length Yra1. The peak fractions of full-length Yra1 from the Mono S were pooled, concentrated, desalted, and used for the subsequent *in vitro* assays.
2.3.2. Yra1 interacts directly with the hyperphosphorylated yeast CTD

The purified recombinant Yra1 was tested for its ability to bind to the PCTD by a “Far Western” assay. Specifically, the purified Yra1 protein was subjected to SDS-PAGE, transferred to nitrocellulose, and probed with a hyperphosphorylated $[^{32}\text{P}]-\text{CTD}$ fusion protein generated by CTDK-I (Materials and Methods). The radioactive PCTD fusion protein (retained on the nitrocellulose by virtue of binding to the immobilized protein) was visualized by autoradiography, and a discrete band at the same mobility as the purified Yra1 protein was observed (Fig. 2.2).

Because phosphorylation by CTDK-I likely produces a combination of CTD phosphoepitopes on the fusion protein, synthetic biotinylated three-repeat consensus CTD peptides were used to characterize the specificity of the Yra1 interaction (Fig. 2.3A).
The synthetic peptides were immobilized on Tetrameric Avidin resin (Phatnani et al., 2004), and purified Yra1 was bound and eluted with two salt steps (Fig. 2.3B). Under these conditions (150 mM NaCl), Yra1 bound to the hyperphosphorylated (2,5P₃) peptide, did not wash off the peptide with excess buffer (5 mL PBS) and eluted with 300 mM NaCl. Yra1 did not bind well to a CTD charge control, in which both Ser2 and 5 were substituted with glutamates (E2E₅), because Yra1 protein was observed only in the flowthrough and washes. Likewise, peptides that contained phosphoserines only at position 2 or 5 of the heptad (2P₃ or 5P₃, respectively) were unable to bind Yra1. These results demonstrate that Yra1 binds directly to a hyperphosphorylated CTD fusion protein and specifically to synthetic heptad repeats with phosphates on both Ser2 and 5.

Figure 2.2: Far Western Assay of Yra1 and the PCTD. The Coomassie-stained gel (left) shows purified recombinant His-tagged Yra1 (approximately 28 kDa). The purified protein was subjected to a Far Western assay probed with a yeast CTD fusion protein hyperphosphorylated ([³²P]-labeled) with CTDK-I. The autoradiograph (right) shows that the PCTD probe was retained at the same mobility as Yra1.
Figure 2.3: Immobilized CTD peptide binding assays. A) Synthetic biotinylated three-repeat consensus CTD peptides that were used for the subsequent binding assays. Serines in blue indicate phosphoserines. B) The CTD peptides shown in A were immobilized on avidin resin. Yra1 was applied to the columns and output (OP), flowthrough (FT), wash (W1 and W2), and salt elution fractions (0.3 M and 1 M NaCl) were analyzed on 4-20% SDS PAGE gels stained with Coomassie Blue.

2.3.3. The phosphoCTD interacting domain of Yra1 spans the N-variable and RRM regions

Based on homology with other RNA Export Factor (REF) family members, the architecture of Yra1 has been defined as in Figure 2.4A (Strasser & Hurt, 2000; Zenklusen et al, 2001). Amino acids 77-184 contain an RNA Recognition Motif (RRM) in the center of Yra1, and two regions that vary in sequence and length among the REF proteins, called the N-variable and C-variable regions, flank the RRM. The termini are more conserved within the family and are designated the N-box and C-box regions.
Based on this primary structure model, I generated truncated versions of Yra1 and used them to identify the PCTD interaction domain (PCID) (Fig. 2A). Using PCR, Yra1 segments representing amino acids 1-226, 1-184, 77-226, 77-184, 1-76, and 184-226 were sub-cloned, and the resulting proteins were expressed and purified as MBP fusions (Materials & Methods). As for full-length Yra1 (Fig. 2.2), the purified proteins were tested for binding to the \[^{32}\text{P}\]-CTD fusion protein generated with CTDK-I (Fig. 2.4B). Both full-length Yra1 (1-226) and the C-terminal truncation (1-184) bound the PCTD probe; the latter interaction was verified using the immobilized hyperphosphorylated (2,5P,3) peptide column (Fig. 2.5). While there was some signal observed for the 77-226 construct, which lacks the region N-terminal to the RRM, in the Far Western assay (Fig. 2.4B), this protein did not bind well to the immobilized 2,5P,3 peptide column (Fig. 2.5). The majority of the protein appeared in the column flowthrough and washes. Finally, the RRM region alone (aa 77-184) did not interact with the PCTD probe in the Far Western assay (Fig. 2.4B); in addition, a larger RRM construct (70-190) that included more flanking sequence, which would presumably improve the folding of the RRM, did not display binding to the immobilized 2,5P,3 peptide column (Fig. 2.5). These data indication that the PCTD interaction involves Yra1 amino acids 1-184 and the C-terminus of Yra1 is not required for PCTD binding.
Figure 2.4: Far Western assay of Yra1 fragments with the PCTD. A) Yra1 architecture based on homology among the REF family of proteins (Zenklusen et al, 2001). The blue regions, including the Nbox, RNA Recognition Motif (RRM), and Cbox, display higher conservation, while the red regions, including the N-variable (Nvr) and C-variable (Cvr), have variable sequences and lengths among the REF proteins. B) Top, Coomassie-stained gel of Yra1 truncation proteins subjected to Far Western assay. Bottom, Autoradiograph of Far Western assay of Yra1 truncation proteins probed with [$^{32}$P]-labeled hyperphosphorylated CTD fusion protein.
Figure 2.5: Immobilized CTD peptide binding assay of Yra1 fragments. Coomassie-stained gels of fractions from the immobilized peptide column binding assays of Yra1 truncations. Yra1 aa 1-226 and 1-184 bound to the immobilized 2,5P₃ peptide and eluted with 0.3 M NaCl, but Yra1 aa 77-226 and 70-190 only predominantly in the flowthrough (FT) and wash (W1 and W2) fractions of the column.

To subdivide the 1-184 amino acid fragment, I truncated the N-terminus to eliminate the N-box region, yielding a protein fragment composed of residues 18-184. The 18-184 fragment retained PCTD binding activity as assessed by both Far Western and immobilized peptide binding assays (Fig. 2.6). These results indicate that the N-box region (aa 1-17) is not required for interaction with the PCTD. Expression and stability problems precluded the construction and analysis of further truncations of this region of Yra1. Amino acids 18-184 were therefore designated the minimum stable PCID-containing fragment of Yra1.
Figure 2.6: Binding of Yra1 aa 18-184 to the PCTD. A) The N-terminal and RRM regions (residues 18-184), which lacks the Nbox and the entire C-terminus (top), was and subjected to a Far Western assay probed with \[^{32}P\]-labeled hyperphosphorylated CTD fusion protein (stained gel and autoradiograph below the protein diagrams). B) Coomassie stained gels of fractions (onput-OP, flowthrough-FT, washes-W1 and W2, and elutions in two salt steps) from an immobilized peptide binding assay (2,5P peptide) of full-length Yra1 1-226 (top) and the truncated variant 18-184 (bottom).

2.3.4. Mutations in the PCID of Yra1 did not affect the CTD binding interaction

To further characterize the residues in the PCID involved in the interaction between Yra1 and the PCTD, conserved positively charged residues were identified based on conservation within this region. In Figure 2.7, the amino acid sequences for Yra1, Drosophila ALY, and mouse ALY are aligned. While the metazoan sequences are much more similar to each other than to Yra1, there are some residues that are conserved in all
three species. Six of these residues, including R55, R59, K79, R107, R134, and R162, were selected as candidate residues for interaction with the PCTD. Alanine substitution mutations were introduced at each of these positions to determine whether removal of the positively charged residues would affect CTD binding. The R55A, R59A, K79A, R107A, R134A, and R162A mutants were purified, and the Far Western assay was used to determine whether these mutants displayed any differences in binding (Fig. 2.8). Interestingly, all of the alanine mutants bound similarly to the hyperphosphorylated CTD fusion protein compared to wild type in the Far Western assay. The variations in signal in the Far Western are similar to the variations in protein amounts on the Coomassie-stained gel, indicating that they are not indicative of binding differences. In addition, the three residues in the N-variable region (R55, R59, and K79) were also mutated simultaneous to alanines (R55A R59A K79A) or glutamates (R55E R59E K79E). The triple N-variable mutants also displayed WT-like binding to the 2,5P3 CTD peptide (data not shown). Thus, mutations of these candidate residues did not affect PCTD binding determined by Far Western assay and the immobilized peptide binding assay.
Figure 2.7: Alignment of the PCID region of Yra1 (S.c.Yra1; NP_010669.1) with the Drosophila (D.m.ALY; NP_609574.1) and mouse (M.m.ALY; NP_062357.3) homologues. The residues indicated with dots were mutated to alanines. The RNP1 and RNP2 motifs are labeled. ClustalW and BoxShade were used to make the alignment.

<table>
<thead>
<tr>
<th></th>
<th>S.c.Yra1</th>
<th>D.m.ALY</th>
<th>M.m.ALY</th>
<th>consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCID</td>
<td>AGSNRARVGCTGRPGPV---GSQRRSLPVRK---DIRKN---TRAPPH NAA---</td>
<td>GKKPQAARGGPGKRZQGQFRFGGARQRPQ---NAGSR---KPQSGVLRPEQYEA CA</td>
<td>GQR---GQRGQ---QRGQHSAQGRRGQVQAARVQGQGQRNPARAARQAAAG---CR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>RNP1</td>
<td>RPYA</td>
<td>RPYA</td>
<td>RPYA</td>
<td>RPYA</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>75</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>RNP2</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

2.3.5. The RNA and CTD binding domains contain the RRM

It has been previously reported that the RRM of Yra1 is unable to bind RNA, while the N- and C-terminal regions of the protein can bind RNA (Zenklusen et al, 2001). To test RNA binding with a different method and to determine whether the PCID was also involved in RNA binding, I used a UV crosslinking assay with purified Yra1 and a [P<sup>32</sup>] end-labeled synthetic 15-base RNA oligonucleotide (5’-CAGUCGCAUAUGUGCA-3’).
This probe was utilized because the RNA sequence specificity of Yra1 is currently not known and this sequence has been shown to bind to the metazoan homolog of Yra1, ALY (Hung et al, 2010). Similarly to ALY, full-length Yra1 crosslinked to the $[^{32}P]$-RNA probe after exposure to UV radiation for increasing lengths of time (Fig. 2.9). The level of crosslinking increased until 5 min of exposure, and little increase in crosslinking was found at 10 min (Fig. 2.9); therefore, 5 min of UV exposure was used for the subsequent assays.

![Figure 2.9: UV crosslinking of full-length Yra1 to a $[^{32}P]$-labeled 15-mer RNA probe (5'-CAGUCGCAUAGUGCA-3') with increasing exposure times to UV light. Top, phosphorimage of crosslinked RNA. Bottom, Coomassie-stained gel of proteins.](image)

UV crosslinking was then used to characterize the RNA binding activity of the truncated Yra1 proteins (1-184, 77-226, and 77-184). I found that all of the truncated proteins, including the RRM (77-184), crosslinked to the RNA probe (Fig. 2.10). The full length (1-226) and N-terminal deletion (77-226) bound better than the C-terminal deletion (1-184) and the RRM (77-184). These data indicate that the RRM (77-184), while not sufficient for maximal RNA binding, is indeed able to bind RNA and that there is potential overlap between the PCID and the RNA binding region.
Because the minimal PCID fragment contains the RRM, which binds RNA, the ability to block RNA binding of full-length Yra1 with excess 2,5P, CTD peptide was analyzed. When a 100-fold molar excess of unlabeled RNA was preincubated with Yra1 before exposure to UV, crosslinking to the \[^{32}\text{P}\]-RNA probe was almost completely blocked. In contrast, preincubation with a 100-fold molar excess of 2,5P CTD peptide had no effect on crosslinking to the RNA probe (Fig. 2.11), indicating that the CTD does not compete with RNA for binding. The amount of competitor required for this assay is likely related to the inefficiency of UV crosslinking, as it has been estimated that crosslinking occurs in only 0.1-10\% of the complexes present in solution (Chodosh, 2001).

Interestingly, I found that Yra1 protein expressed in baculovirus-infected Spodoptera frugiperda (Sf9) cells would not bind RNA in the UV crosslinking assay (Fig. 2.12A); using the same conditions and protein concentrations, Yra1 expressed in E. coli showed robust UV crosslinking (as in Fig. 2.9; see Fig. 2.12A). On the other hand, both versions of Yra1 bound to the doubly phosphorylated CTD peptide (Fig. 2.12B). While the cause
of the difference between the two versions of Yra1 is unclear, I hypothesize that, like ALY (Hung et al, 2010), post-translational modifications on Yra1 expressed in Sf9 cells block RNA binding but not CTD binding. In any case, these results further demonstrate that CTD binding and RNA binding are distinct events.

![Image](image.png)

**Figure 2.11**: UV crosslinking of 15-mer RNA to full-length Yra1 with no competitor (lanes 1, 2, 5) or 25-fold (lanes 3, 6) and 100-fold (lanes 4, 7) molar excess of cold RNA or 2,5P, CTD peptide competitors, respectively. Top, phosphorimage of crosslinked RNA. Bottom, Coomassie-stain of proteins.
To further probe the involvement of the RRM in RNA and CTD binding, I first identified conserved amino acids in the RRM sequences of budding yeast Yra1 and eight Yra1 homologues from other eukaryotes (Fig. 2.13); these include one positively charged residue (R107, leftmost dot) and two aromatic residues (Y112 and F126, middle and right dots). These two aromatic residues are extremely well conserved in evolution, except in...
**Caenorhabditis elegans.** In most RRM s of proteins that bind RNA (other than Yra1 homologues), a well-conserved aromatic residue is found within the middle of the RNP1 motif (Maris et al, 2005), which is indicated with a star in Fig. 2.13. In Yra1 and its homologues, this residue is an aspartic acid or asparagine, the presence of which has been proposed to account for the inability of these RRMs to bind RNA in the Yra1-like proteins; however, there is a nearby aromatic amino acid, F126, that I hypothesized might play a role in RNA binding in the absence of the normally conserved residue in the RRM.

![Amino acid alignment of the RRM sequences of Yra1 and homologous proteins. The RNP1 and RNP2 motifs are labeled. Three conserved residues indicated with dots were identified (R107, Y112, & F126 in Yra1) and mutated to alanines. The star below the alignment (N123 in Yra1) indicates the position of a conserved phenylalanine in most (non-Yra1) RRM sequences. BLAST was used to identify homologues of Yra1, and ClustalW and BoxShade were used to make the alignment.]

S.c. – *Saccharomyces cerevisiae* Yra1 (NP_010669.1); S.p. – *Schizosaccharomyces pombe* Mlo3 (NP_595715.1) and THO subunit (NP_595161.1); N.c. – *Neurospora crassa* hypothetical protein NCU01793 (XP_956471.1); C.e. – *Caenorhabditis elegans* ALY1 (NP_501588.1) and ALY2 (NP_501594.1); D.r. – *Danio rerio* THO subunit 4 (NP_001098578.1); G.g. – *Gallus gallus* hypothetical protein (XP_001232393.1); M.m. – *Mus musculus* ALY (NP_062357.3).

To test whether the conserved residues R107, Y112, and F126 are involved in RNA binding, I performed an alanine scan and analyzed the purified mutant proteins by UV crosslinking (Fig. 2.14A). The R107A and Y112A mutations did not seem to significantly
affect RNA binding \textit{in vitro}; however, the F126A mutant protein showed a decrease in RNA binding. If all of these mutations were made (R107AY112AF126A; Triple mutant in Fig. 2.14A), the resulting mutant displayed little RNA binding. In contrast, all of these mutant proteins, including the R107AY112AF126A triple mutant, retained the ability to bind to the immobilized CTD 2,5P$_3$ peptide (Fig. 2.14B) (slight apparent differences in column elution behavior are within the range of experimental variation; e.g., see Fig. 2.6B and Fig. 2.12B). These results demonstrate that the RRM as an isolated protein segment binds RNA very weakly, whereas the same RRM in the context of the full-length protein is required for optimal RNA binding. Together, the RNA binding and PCTD binding results indicate that while the RNA and PCTD binding activities share the same protein segment, the two activities involve distinct amino acids and possibly distinct binding surfaces.
2.3.7. The interaction between Iws1 and Yra1/ALY is conserved

The mammalian homologue of Yra1, ALY, has also been shown to interact with Spt6 through a physical interaction with the scaffold factor Iws1 (Yoh et al, 2007). As Iws1 and ALY both have homologues in *S. cerevisiae*, I hypothesized that this interaction
would also be conserved. Using the Far Western assay, the ability of the Yra1 truncations to interact with purified Iws1 was examined (Fig. 2.15). Similarly to the RNA and PCTD binding interactions, Isw1 interacted with full-length Yra1 and the C-terminal and N-terminal truncations (1-184 and 77-226, respectively). Although these results have not been independently verified in solution, it does indicate that the Iws1-Yra1/ALY interaction is conserved from yeast to mammals. In addition, taken together with the PCTD and RNA binding results, it seems that the RRM domain is an important interaction surface on Yra1.

Figure 2.15: Far Western assay of Yra1 and Iws1. Left, Coomassie stained gel of Yra1 truncation proteins subjected to Far Western assay. Right, Anti-His tag visualization of Iws1 Far Western assay of Yra1 truncation proteins.

2.4. Discussion

The phosphoCTD of RNAPII plays an important role in the cotranscriptional recruitment of elongation, chromatin remodeling, and mRNA processing factors; in this study, the PCTD binding of an mRNA export factor, Yra1, was characterized. Previously, using ChIP approaches, Yra1 has been shown to be cotranscriptionally
recruited to active genes during elongation (Abruzzi et al., 2004; Johnson et al., 2009; Lei et al., 2001; Zenklusen et al., 2002). Although it was not clear whether Yra1 associated with the transcription machinery or the nascent mRNA during elongation, the ChIP association was shown to be significantly resistant to nuclease treatment (Abruzzi et al., 2004), which is consistent with a protein-protein interaction. I showed here that Yra1 directly binds to the hypersphosphorylated form of the CTD and to CTD peptides phosphorylated on Ser2 and 5, the phosphorylation pattern thought to predominate during transcription elongation. These in vitro data are consistent with the previous ChIP studies, and I propose that Yra1 may be recruited to genes through interaction with the PCTD.

Because Yra1 is a uniformly positively charged RNA binding protein, I sought to demonstrate that the PCTD interaction is a bona fide binding event. To this end, I showed that Yra1 binds to the CTD phosphorylated in an elongation-specific pattern (Ser2,5P, but not Ser5P or Ser2P). The interaction is specific for phosphoserines in the context of the CTD, as a peptide with substitution of Ser2 and 5 with the phosphomimetic amino acid glutamate was not bound by Yra1. Using truncations of Yra1, I further showed that CTD binding is confined to amino acids 18-184, while optimal RNA binding (out of the truncations analyzed) was obtained with amino acids 77-226. While these regions both contain the RRM, preincubation of Yra1 with the Ser2,5P CTD peptide did not interfere with RNA binding. Moreover, while Yra1 expressed in baculovirus-infected Sf9s is unable to bind RNA potentially due to posttranslational modifications, CTD peptide binding is not blocked. These results therefore demonstrate that Yra1 binds specifically to the PCTD in a manner that is distinct from RNA binding.

Previous studies indicated that all regions of Yra1 except the RRM are responsible for RNA binding (Zenklusen et al., 2001). Using UV crosslinking, I was able to detect
binding of the isolated RRM to a synthetic RNA probe; however, a construct that included the sequence C-terminal to the RRM displayed RNA binding that was the closest to full length. Moreover, point mutations in the RRM of full-length Yra1 affected RNA binding but not CTD binding. This is the first demonstration that the RRM of Yra1 is directly involved in RNA binding. It is possible that, when expressed alone, the RRM does not fold well enough to bind RNA strongly and that increasing the amount of sequence around the RRM improves folding. In addition, the C-terminus of Yra1 may contribute to RNA binding. In ALY/REF2-I, the RRM and N-terminus of the export factor were shown to bind RNA in chemical shift perturbation studies (Golovanov et al, 2006). Similar studies on Yra1 would clarify the roles of the N-terminal and C-terminal regions in RNA binding.

It is not surprising that the RRM in Yra1 may play a role in a protein-protein interaction, as many RRMs have been shown to be important for protein but not RNA interaction. For example, the splicosomal components U2AF35 and U2AF65 both contain non-canonical RRMs that mediate interactions with U2AF65 and SF1, respectively (Kielkopf et al, 2004). Moreover, a recent study indicated that the non-canonical RRM of U2AF65 may also mediate an interaction between the splicing factor and the CTD (David et al, 2011). The RRM thus may be another domain that proteins utilize to interact with the CTD of RNA polymerase II.

Taken together, these in vitro studies indicate that Yra1 binds specifically to the Ser2/5 phosphorylated form of the CTD of RNAPII, and its PCID is located within amino acids 18-184. In addition, Yra1 binds RNA principally with amino acids 77-226, and mutation of the phenylalanine within RNPI (F126) affects RNA binding, indicating that the RRM contributes to RNA binding. Although the regions involved in CTD and RNA binding overlap, they are in fact distinct, as RRM mutations that diminish RNA
binding do not affect CTD binding and truncation of the PCID does not affect RNA binding.
3. Role of CTD binding in mRNA export by Yra1 in vivo

3.1. Introduction

In the previous in vitro analyses, I found that Yra1 interacted with the Ser2/5P form of the CTD of RNAPII. These data are consistent with previous Yra1 ChIP data, which show that the level of Yra1 association with active genes peaks during transcription elongation (i.e., in the middle of the gene) (Abruzzi et al, 2004; Johnson et al, 2009; Lei et al, 2001; Zenklusen et al, 2002), which is also when the Ser2/5P form or hyperphosphorylated form of the CTD of RNAPII is observed (Kim et al, 2010; Mayer et al, 2010; Tietjen et al, 2010). Moreover, treatment of extracts with RNase prior to ChIP did not significantly affect the Yra1 signal, indicating that Yra1 is likely tethered to the polymerase during elongation by a protein-protein interaction (Abruzzi et al, 2004), which also agrees with a PCTD interaction. However, I sought to demonstrate that this PCTD interaction is important for Yra1 function in vivo.

Because YRA1 is an essential gene in S. cerevisiae (Portman et al, 1997), genetic and functional analyses require the use of conditional alleles, and a handful have been developed and utilized. The yra1-1 allele, which contains five point mutations (S42G, V80A, E137G, K210E, and F223S), has been used as a temperature sensitive (ts) allele. While it displays no growth and nuclear accumulation of poly(A)+ RNA at 37°C, this strain also grows much slower than wild type at 30°C (Strasser & Hurt, 2000). The yra1-2 allele contains a point mutation (V164D) and a G-to-C change in the 5′ acceptor splice site of the intron, which is thought to disrupt normal splicing and require the use of an upstream alternative splice site, creating a two amino acid deletion (Kashyap et al, 2005). The yra1-2 allele grows poorly at the permissive temperature and only a little slower at 34°C; however, it was more sensitive than wild type to hydroxyurea, which depletes
nucleotide pools. The ts allele GFP-\textit{yra1}-8 contains two amino acid substitutions (D10K and D11K) that are innocuous on their own but create temperature sensitive growth and mRNA export phenotypes when introduced in conjunction with fusion to green fluorescent protein (Zenklusen et al, 2002). A \textit{yra1} allele in which the sequence encoding the RRM has been deleted (\textit{yra1ΔRBD} or \textit{yra1ΔRRM}) has been used in many studies to identify synthetic lethal interactions (Strasser & Hurt, 2001; Strasser et al, 2002), but removal of this domain results in only minor ts growth and export phenotypes (Zenklusen et al, 2001). Due to the limitations of these alleles, they have mainly been used to identify genetic interactions but not mechanistic aspects of Yra1 function.

To study the role of the PCTD in Yra1 function, I created a series of yeast strains based on the results of the \textit{in vitro} analyses in the previous chapter. While a point mutation or series of point mutations that affect PCTD binding were not identified, truncation of the N-terminus (\textit{yra1} 77-226) abrogated PCTD binding but did not affect RNA binding; therefore, a strain expressing aa 77-226 was constructed. In addition, because point mutations that decreased RNA binding but not PCTD binding (F126A and R107AY112AF126A) were identified, these point mutations were introduced. Protein expression, growth, sensitivity to DNA damaging agents, protein localization, mRNA export, and protein cotranscriptional association of these strains were characterized.

3.2. \textit{Materials and methods}

3.2.1. Yeast strains and growth

The FY84 strain of \textit{Saccharomyces cerevisiae} [\textit{MATa his3Δ200 leu2Δ1 ura3-52 lys2-128Δ}] was used to produce the \textit{yra1} mutant strains. To create a \textit{YRA1} plasmid shuffle strain, the chromosomal copy of \textit{YRA1} was deleted by homologous recombination with the
*kanMX* cassette (Wach et al, 1994) or the *HIS3* gene (Brachmann et al, 1998) with a covering plasmid, which consisted of a V5-tagged full-length *YRA1* fusion in the pAG416GPD Gateway vector (Alberti et al, 2007).

To compare growth of the mutants to WT, overnight yeast cultures were diluted to an OD$_{600}$ of 1 and used to make four 10-fold serial dilutions. Each dilution was spotted onto a YPD plate (3 µL/spot), and the plates were incubated at 16°C for 7 days, 30°C for 2 days or 37°C for 3 days. Expression of Yra1 mutant proteins was verified in whole cell extracts made in ethanol as described previously (Lee & Greenleaf, 1991) by Western blot with an anti-V5 epitope antibody (MCA1360, AbD Serotec).

### 3.2.2. Cloning

For expression of WT and mutant proteins in yeast, the full-length *YRA1* gene was PCR amplified from FY84 genomic DNA, tagged with the V5 epitope and inserted into the pAG416GPD and pAG415GPD plasmids using the Gateway method (Alberti et al, 2007) (Invitrogen). The *F126A*, *R107AF126A*, and *Y112AF126A* point mutants and the 77-226 truncation (-*PCID*) mutant were produced by PCR and inserted into the pAG415GPD plasmid. After sequencing, the plasmids were transformed into the *YRA1* plasmid shuffle strain by the lithium acetate method (Gietz & Woods, 2002). The *YRA1* covering plasmid was counterselected using 5-fluoroorotic acid-containing plates.

### 3.2.3. Fluorescence *in situ* hybridization (FISH) and immunostaining

The method of Tokunaga and Tani (2008) was used for FISH with some modifications (Tokunaga & Tani, 2008). Briefly, the *S. cerevisiae* strains were grown in YPD to mid-log phase at 30°C. The cultures were then split to produce two identical cultures; one culture was incubated at the non-heat shock temperature (NHS, 30°C) and the other at the heat shock temperature (HS, 37°C) for 30 min. After heat shock, the cells
from 1.5 mL of each culture were harvested by centrifugation (3000 g for 2 min at 4°C) and fixed for 1 hour at room temperature in 4% formaldehyde. The fixed cells were permeabilized with 40 μg of Zymolyase for 20 min at 37°C. The spheroplasts were resuspended in 1.2 M sorbitol in 0.1 M sodium phosphate buffer, pH 6.0; the suspension allowed to settle onto poly-L-lysine coated glass slides for 30 min at room temperature. The slides were then washed with 70% ethanol, dehydrated with 70%, 90%, and 100% ethanol, and dried. After incubating the slides for 30 min in prehybridization solution (4X SSC, 5X Denhardt solution and 1 mg/mL baker’s yeast tRNA), the slides were hybridized overnight at 42°C with hybridization solution (prehybridization solution containing 1 ng/mL digoxigenin-labeled dT₅₀ probe). The hybridized slides were stained with FITC-labeled anti-digoxigenin antibody (1:200 dilution; Roche) and DAPI (0.1 μg/mL; Pierce). A coverslip was mounted using glycerol containing 1 mg/mL p-phenylenediamine, and the slides were imaged on a Zeiss Axio Observer microscope with the 100X oil objective.

For immunostaining of V5-tagged Yra1, cells were grown, fixed, and permeabilized as described above. After allowing the cells to settle on poly-L-lysine coated glass slides, the samples were blocked with 5% BSA in PBST. The slides were then incubated with a 1:200 dilution of anti-V5 FITC antibody (MCA1360, AbD Serotec) in 5% BSA in PBST overnight at 4°C. After washing with 5% BSA in PBST, the cells were counterstained with 0.1 μg/mL DAPI, mounted, and imaged as described above.

3.2.4. Chromatin immunoprecipitation

The method described in Rusche and Rine (2001) was used with some modifications (Rusche & Rine, 2001). Biological duplicate yeast cultures were grown in YPD at 30°C to an OD₆₀₀ of 1.0. Approximately 50 OD₆₀₀s of cells were fixed with 1% formaldehyde for
30 min at room temperature. The cells were harvested by centrifugation at 3000 rpm for 5 min, resuspended in 400 µL of lysis buffer [50 mM HEPES, pH 7.6, 1 mM EDTA, 140 mM NaCl, 1% Triton X-100, 0.1% Na deoxycholate, 1 mM PMSF, 5 µg/mL chymostatin, 2 µg/mL pepstatin A, 1 mM benzamidine, and 1:100 dilution of protease inhibitor cocktail for fungal extracts (P8215, Sigma-Aldrich)] and lysed by bead beading with 0.5 mm Zirconia/Silica beads (BioSpec) for three pulses of 1 min each with cooling in an ethanol ice bath between cycles. The lysate was then collected by pipetting and pooled with 400 µL of lysis buffer used to wash the beads. The pooled lysate was sonicated four times for 10 s to shear the DNA. The cell debris was removed by centrifugation for 10 min at 4°C, and 160 µL of the soluble lysate was diluted to 400 µL with lysis buffer. After preclearing with salmon sperm DNA and BSA-blocked Protein G Dynabeads (Invitrogen), the lysate was incubated overnight with 2.5 µg of mouse monoclonal antibody (Yra1: anti-V5 antibody, MCA1360, AbD Serotec; Rpb3: anti-Rpb3 antibody, WP012, Neoclone; Mock: anti-β-galactosidase antibody, Z378A, Promega). The bound complexes were precipitated with blocked Protein G Dynabeads, washed, and eluted with 1% SDS in 0.1 M NaHCO₃. The crosslinks were reversed and the DNA was isolated using standard techniques (Moore & Dowhan, 2002). The samples were used for quantitative real-time PCR analysis with SYBR Green (Bio-Rad) and an Eppendorf Mastercycler realplex system (Eppendorf). Data were quantified relative to the standard curve of input DNA and are presented as the mean and standard deviation of independent immunoprecipitations from biological replicates. The error bars represent the standard deviation. The primers used for quantitative PCR are detailed in Table 1.
3.3. **Results**

3.3.1. V5-tagged Yra1 supports viability in *S. cerevisiae*

Before testing the effect of the mutants of interest in *S. cerevisiae*, the genomic copy of *YRA1* needed to be knocked out and covered with *YRA1* on a *URA3*-containing plasmid to set up a plasmid shuffle system. A PCR product containing a selectable marker [the *kanMX* cassette (Wach et al, 1994) or the *HIS3* gene (Brachmann et al, 1998)] with sequence flanking the *YRA1* ORF at the 5’ and 3’ ends was generated. This PCR product was transformed into a *YRA1* strain (FY84) with a plasmid encoding *YRA1* and *URA3*. In addition, the plasmid-encoded version of *YRA1* also contained a 14 aa V5 epitope tag (GKIPNPLLGLDST). After selection, the resulting strain will be Δ*yra1* with a *yra1V5* plasmid that can be counterselected on plates containing 5-fluoroorotic acid.

As shown in Figure 3.1A-C, the FY84 strain without the covering plasmid had one protein band of approximately 28 kDa that was reactive to an anti-Yra1 antibody and no bands that are reactive to an anti-V5 antibody (Fig. 3.1 B vs. C). After transformation with the covering plasmid, the FY84 strain expresses two proteins with slightly different molecular weights that are reactive to the anti-Yra1 antibody, and the upper band is also reactive to the anti-V5 antibody. After transformation and selection for the disruption cassette (*yra1Δ::HIS3* in this case), the lower anti-Yra1 reactive band is no longer expressed, indicating that the chromosomal *YRA1* gene has been deleted (Fig. 3.1A-C). While it appears that the V5-tagged protein is expressed at a lower level relative to the untagged endogenous protein (Fig. 3.1B), this is probably an artifact of epitope blocking by the V5 antibody, as this Western was simultaneously probed with both
antibodies. Because the intron of YRA1 has been shown to determine the expression level of Yra1 independently of the promoter used or plasmid copy number (Preker & Guthrie, 2006; Preker et al, 2002), the plasmid-encoded WT Yra1V5 and all mutants discussed later should be expressed at similar levels due to the presence of the 766-nt intronic sequence.

Based on the growth of these strains, plasmid-encoded yra1V5 is able to support viability similarly to the FY84 strain (YRA1), indicating that the V5 tag does not interfere with the function of Yra1 (Fig. 3.1D). Therefore, the V5 tag was used to tag all mutants of Yra1, and the yra1V5-expressing strain will be referred as the “wild type” (WT) strain in the rest of this chapter.

**Figure 3.1:** Characterization of the Yra1V5-expressing strain. A-C) Western blots of total cell extracts from the WT YRA1 strain, WT YRA1 strain expressing plasmid-encoded yra1V5, and yra1Δ::HIS3 strain expressing plasmid-encoded yra1V5.
Figure 3.1 (continued): Note that the V5-tagged protein shows an increase in apparent molecular weight (displayed as a doublet in anti-Yra1 western in B). D. Growth of the WT YRA1 strain, WT YRA1 strain expressing plasmid-encoded yra1V5, and yra1Δ::HIS3 strain expressing plasmid-encoded yra1V5. Serial 10-fold dilutions of cells were spotted on rich medium (YPD) and plates were incubated at 30°C.

3.3.2. Yra1 CTD and RNA binding mutants are expressed and display growth phenotypes

Based on the in vitro binding experiments, I used yra1 mutants that affect CTD binding (77-226 truncation) or RNA binding (F126A) to probe the requirements for these interactions in vivo; I also constructed double mutants of the RRM residues (Y112AF126A and R107AF126A). Western blots for the V5-tagged proteins showed that all of the mutant Yra1 proteins were expressed at similar levels relative to the loading control, glycerol-6-phosphate dehydrogenase (G6PDH) (Fig. 3.2).

In dilution plating experiments at 30°C, the R107A and Y112A yra1 mutants displayed wild-type growth, whereas F126A, R107AF126A and 77-226 appeared to display a slight slow growth phenotype. In contrast, the Y112AF126A strain was very slow growing at 30°C (Fig. 3.3, left panel). When the strains were tested for growth at high (37°C; Fig. 3.3, middle panel) and low (16°C; Fig. 3.3, right panel) temperatures, the
R107A and Y112A strains behaved like WT, while the F126A and 77-226 strains showed slight growth defects at high but not at low temperature. In contrast, the double mutants showed marked but distinct temperature sensitive phenotypes. The R107AF126A strain was inviable at 37°C but grew like wild type at 16°C. The Y112AF126A strain, which was slow growing at 30°C, displayed slower growth at both higher and lower temperatures. Thus, while these Yra1 mutant proteins are functional enough to support viability, the F126A, R107AF126A, Y112AF126A, and 77-226 strains manifest growth defects at one or more temperature.

![Figure 3.3](image)

**Figure 3.3:** Growth of the RRM point mutant (R107A, Y112A, F126A, R107AF126A, and Y112AF126A) and N-terminal truncation (77-226) strains. Serial 10-fold dilutions of cells were spotted on rich medium (YPD plates) and plates were incubated at 30°C (left), 37°C (middle), or 16°C (right).

### 3.3.3. Yra1 PCTD and RNA binding mutants are sensitive to hydroxyurea

In addition to thermal stress, I asked whether the yra1 mutant strains were sensitive to DNA damaging agents. The yra1-2 allele has been shown to be sensitive to hydroxyurea (HU) (Kashyap et al, 2005), and our lab has found a connection between sensitivity to DNA damage via doxorubicin (DOX) and PCAPs (Westmoreland et al,
Moreover, the TREX complex has been connected to hyperrecombination and genomic instability (Jimeno et al, 2002). To test the sensitivity of these strains to DNA damage, dilutions of the *yra1* mutant strains were incubated on plates containing no drug (YPD), 2 mM methyl methanesulfonate (MMS), 200 mM HU, or 50 µg/mL DOX (Fig. 3.4). MMS is an alkylating agent, HU inhibits ribonuclease reductase, and DOX inhibits topoisomerases. The *yra1* mutants do not display marked sensitivity to MMS or DOX. In contrast, the F126A and Y112A single mutants are slightly more sensitive than WT or R107A to HU, and the R107AF126A and Y112AF126A double mutants and the 77-226 truncation are very sensitive to HU. The Y112AF126A double mutant also seems to be sensitive to MMS and DOX. These data indicate that disruption of RNA binding or CTD binding of Yra1 might make yeast more sensitive to HU treatment.

![Figure 3.4: Effects of DNA damaging agents on growth of the RRM (R107A, Y112A, F126A, R107AF126A, and Y112AF126A) and PCID (77-226) mutants. Serial 10-fold dilutions of cells were spotted on rich medium in the absence (YPD) or presence of DNA damaging agents [2 mM methyl methanesulfonate (MMS), 200 mM hydroxyurea (HU), and 50 µg/mL doxorubicin (DOX)].](image)

**3.3.4. RRM mutations do not alter Yra1 subcellular localization**

As deletion of the N-variable region has been shown to change the subcellular localization of Yra1 (Zenklusen et al, 2001), I used immunostaining to verify that the point mutations did not alter nuclear localization. As shown in Figure 3.5, the full-length
WT protein and the RRM mutant proteins F126A, R107AF126A, and Y112AF126A displayed nuclear and perinuclear staining. Thus, the RRM point mutations do not alter Yra1 subcellular localization. In contrast, the 77-226 protein was diffusely localized in the nucleus and cytoplasm, which is consistent with previous studies. Therefore, the growth defect found in the RRM double mutant is not due to mislocalization of the protein.

![Image of localization of Yra1 proteins in yeast](image)

3.3.5. Yra1 CTD and RNA binding mutants display mRNA export defects

To determine whether the growth defects of the mutant strains resulted from defects in mRNA export, the distribution of polyA⁺ RNA was analyzed by FISH. In wild-type cells, polyA⁺ RNA, which was detected with a FITC-labeled oligo dT₅₀ probe, was uniformly distributed throughout cells (Fig. 3.6), and the polyA⁺ RNA distribution did not change at high temperature. In contrast, the CTD-binding deficient mutant (77-226) displayed significant nuclear accumulation of polyA⁺ RNA at 37°C (Fig. 3.6). This
result is consistent with the growth of the 77-226 strain, which was nearly WT at 30°C but slow growing at 37°C.

Figure 3.6: Yra1 PCID truncation strain displays an export defect at 37°C. Fluorescence in situ hybridization (FISH) of polyA⁺ RNA is shown in green. DAPI (blue) was used to stain the nuclei. Nonheat shock (30°C; left) and heat shock (37°C; right) conditions were used.

In addition, at the permissive temperature, the F126A, R107AF126A, and Y112AF126A strains showed greater nuclear retention of polyA⁺ RNA compared with wild-type cells (Fig. 3.7). Nuclear retention was particularly severe in the Y112AF126A strain, which also displays the most severe growth defect. These results are consistent with the growth assays and the hypothesis that the PCID of Yra1 is required for optimal cotranscriptional recruitment of Yra1 and its mRNA export activity. Moreover, as the RRM point mutations decreased RNA binding in vitro, the export defects in the RRM mutant strains are consistent with the proposition that the RRM of Yra1 is indeed important for RNA binding in vivo.
3.3.6. PCID truncation leads to a loss of Yra1 at the \textit{PMA1} and \textit{PYK1} genes

I hypothesized that the role of PCTD binding is to increase the efficiency of Yra1 recruitment to its nascent mRNA cargo. One measure of cotranscriptional recruitment can be obtained by chromatin immunoprecipitation (ChIP), which provides an indication of protein occupancy along a gene of interest. Using primer sets specific to the 5’-end, middle, and 3’-end of the \textit{PMA1} and \textit{PYK1} genes (Table 1, Fig. 3.8), I used ChIP assays to measure the occupancy levels of Yra1 and Rpb3 on these two genes (Fig. 3.9-3.10). Consistent with previous studies (Abruzzi et al, 2004; Zenklusen et al, 2002), wild-type Yra1 displayed low occupancy at the 5’-ends of both genes, and the level of Yra1 occupancy increased along the length of the genes. An internal control, RNAPII (as indicated by the Rpb3 subunit), showed essentially constant occupancy levels across the genes, as seen previously (Kim et al, 2010; Mayer et al, 2010; Tietjen et al, 2010).

Interestingly, the level of the PCID truncated protein (aa 77-226, “–PCID” in Fig. 3.9-3.10), at either \textit{PMA1} or \textit{PYK1}, was lower than the WT protein by up to 10-fold (mean values of 9.60 ± 0.18% input for WT vs. 0.97 ± 0.46% input for -PCID with the 2263-2421.
primer set of *PMA1*). Note that the PCID truncation of Yra1 did not affect levels of RNAPII at either gene, as shown by the Rpb3 occupancy values. Moreover, mock IPs with an anti-β-galactosidase antibody showed that the background was consistently less than 0.1% input. A region on the left telomere of chromosome I (TELIL) was also used as a transcriptionally inactive control segment for ChIP. The decrease in protein association with the two genes was not due to the lack of expression of the 77-226 (-PCID) mutant because the protein was clearly detected by Western in the ChIP extracts after reversal of the crosslinks (Fig. 3.11). Thus, a version of Yra1 lacking part of the PCID is present on transcription units at much lower levels than the full-length protein.

**Figure 3.8:** Schematic showing the primer set locations on the *PMA1* and *PYK1* genes.
Table 3.1: Primers used for quantitative real-time PCR analysis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer Set</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5’- AAAGCCTGCTAAGACTTACGATGACG -3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’- TTCTTCTGGAACTGTCTAGCTTCAC -3’</td>
</tr>
<tr>
<td></td>
<td>+72</td>
<td>5’- GTTTGCCAGCTGTTACCACCAC -3’</td>
</tr>
<tr>
<td></td>
<td>+235</td>
<td>5’- TTCTTCTTCTGGAAGCAGCCAAC -3’</td>
</tr>
<tr>
<td></td>
<td>+1010</td>
<td>5’- TTATGGGTATGTCTATTATTTTG -3’</td>
</tr>
<tr>
<td></td>
<td>+1250</td>
<td>5’- AATCAACCAGTTTTCAGTCAAGG -3’</td>
</tr>
<tr>
<td></td>
<td>+2263</td>
<td>5’- CGTTGTTGCTGTTCTGACCTTG -3’</td>
</tr>
<tr>
<td></td>
<td>+2421</td>
<td>5’- CAATGACAGACTTGTGATTTC -3’</td>
</tr>
<tr>
<td></td>
<td>+32</td>
<td>5’- CGCCAACGATGTTTGGACC -3’</td>
</tr>
<tr>
<td></td>
<td>+193</td>
<td>5’- CAAGTCACCTCTGGACCA -3’</td>
</tr>
<tr>
<td></td>
<td>+651</td>
<td>5’- CTTGGTTACCAGATGCCCAAGA -3’</td>
</tr>
<tr>
<td></td>
<td>+801</td>
<td>5’- AAGATACGAAATTCTCCAGCC -3’</td>
</tr>
<tr>
<td></td>
<td>+1263</td>
<td>5’- TTTGTAACCCGACCTTACC -3’</td>
</tr>
<tr>
<td></td>
<td>+1416</td>
<td>5’- GGCTGGGAATGTTGATGC -3’</td>
</tr>
</tbody>
</table>
Figure 3.9: Chromatin immunoprecipitation of wild-type (WT) and truncated (-PCID) Yra1 at the PMA1 gene.
Figure 3.9 (continued): The TELIL primer set recognizes a region in the telomere of chromosome I that should be transcriptionally inactive based on annotation of the yeast genome. The mock IP was performed with anti-β-galactosidase mouse monoclonal antibody. Note that the Yra1 and Rpb3 graphs are plotted on the same scale for comparison, but the mock IP is plotted on a different scale. The data are shown as the percentage of DNA immunoprecipitated relative to the input DNA for each strain. For each sample, the mean of two independent immunoprecipitations from biological replicates is shown, and the error bars represent the standard deviation of the biological replicates.
Figure 3.10: Chromatin immunoprecipitation of wild-type (WT) and truncated (-PCID) Yra1 at the PYK1 gene.
Figure 3.10 (continued): The TELIL primer set recognizes a region in the telomere of chromosome I that should be transcriptionally inactive based on annotation of the yeast genome. The mock IP was performed with anti-β-galactosidase mouse monoclonal antibody. Note that the Yra1 and Rpb3 graphs are plotted on the same scale for comparison, but the mock IP is plotted on a different scale. The data are shown as the percentage of DNA immunoprecipitated relative to the input DNA for each strain. For each sample, the mean of two independent immunoprecipitations from biological replicates is shown, and the error bars represent the standard deviation of the biological replicates.

![Image](image.png)

Figure 3.11: Anti-V5 Western of the WT and -PCID (77-226) extracts used for the ChIP experiments. A and B for each stain indicate biological replicates. Crosslinks were reversed by incubation at 95°C for 1 hour, and proteins were separated on a 4-20% gel.

It is important to note that the localization defect caused by truncation of the N-terminus of Yra1 (Fig. 3.5) probably also contributes to the ChIP results, as a lower amount of Yra1 in the nucleus would likely decrease the amount of protein associated with a given gene. However, altogether, these data indicate that Yra1 specifically binds to the hyperphosphorylated CTD of RNAPII in vivo, and disruption of this interaction results in a defect in Yra1 recruitment.

3.4. **Discussion**

Based on the in vitro analyses in Chapter 2, I hypothesized that the PCTD would play an important role in recruiting Yra1 to mRNAs at active genes. While I was unable to
identify point mutations that affect PCTD binding, I found that deletion of the N-terminus of the protein decreased PCTD binding but had little effect on RNA binding. Therefore, I used the N-terminally truncated protein to investigate the role of PCTD binding in Yra1 function in S. cerevisiae. The 77-226 strain showed a slight slow growth phenotype and only a minor mRNA export phenotype during heat stress. This is consistent with the in vitro results, as the ability to bind RNA is likely the essential activity of Yra1 and deletion of the N-terminus did not affect RNA binding. Yet, the ChIP assay indicated that even at 30°C, recruitment of Yra1 to the PMA1 and PYK1 genes was severely compromised. The lack of severe growth or export phenotypes in the 77-226 truncation strain indicates that PCTD binding is not essential for export at the normal growth temperature. It is possible that at the higher temperature the co-transcriptional recruitment defect of the 77-226 protein is more severe or that the mutant protein is recruited to a different subset of genes.

While the 77-226 strain displayed a drastic decrease in protein that associated to two active genes (PMA1 and PYK1) during elongation, truncation of the N-terminus also created a defect in Yra1 nuclear localization. The altered localization of Yra1 in this stain would also decrease the steady-state amount of Yra1 in the nucleus, which would, by extension, decrease the amount of Yra1 associated with a given gene. At this time, the extent at which the mislocalization of Yra1 in the 77-226 strain contributed to the ChIP results is not clear. While this subcellular localization defect is thought to be due to the deletion of a nuclear localization signal (NLS) in the N-variable region (Zenklusen et al, 2001), Yra1 may not require an NLS for nuclear localization based on the following: the sequence of an NLS has not been identified; Yra1 contains a large proportion of basic amino acids (22.6% of the sequence is R or K), which could function as an NLS; and the small size of WT and the N-terminal truncation (26 and 18 kDa, respectively) may allow
these proteins to diffuse through the NPC without the requirement for an NLS. Based on the results presented in this study, it is possible that truncation of the PCID actually causes the protein to spend more time in the cytoplasm due to the lack of efficient recruitment to RNAPII.

It is also currently unclear how the Yra1-PCTD interaction is involved in or affects the co-transcriptional association of other Yra1-interacting proteins, such as the TREX complex (which contains the THO proteins, Hpr1, Tho2, Mft1, and Thp2, along with Yra1 and Sub2), Iws1, or Mex67/Mtr2. While the proteins in TREX usually coimmunoprecipitate, Yra1 may associate weakly with the complex, as the levels of Yra1 and Sub2 are often substoichiometric; in addition, a high salt wash has been shown to remove Yra1 from the complex completely (Strasser et al, 2002). An implied interaction between Yra1 and TREX, supported by the observation that the ChIP enrichment of Yra1 at PMA1 is decreased approximately 50% in an hpr1Δ strain (Zenklusen et al, 2002), appears not to be universal because ChIP analysis of TREX complex members on multiple genes showed that the recruitment patterns of Yra1 and Hpr1 are not always the same (Abruzzi et al, 2004). Moreover, deletion of HPR1 has been shown to affect transcription elongation, so the decrease in Yra1 occupancy may be due to an elongation defect rather than a direct effect of Hpr1 on the export protein (Rondon et al, 2003). Thus the TREX complex may not be important for Yra1 cotranscriptional association at all genes. In this study, the decrease in Yra1 association at PMA1 and PYK1 is unlikely to be due to global elongation defects because the levels of Rpb3 (and therefore RNAPII) were not altered across either gene in the 77-226 strain. Interestingly, the PCID potentially overlaps with part of the Mex67 binding site on Yra1, which has been shown to include the N-variable region (Strasser & Hurt, 2001), while the Yra1-Sub2 interaction involves a different region. It will be informative to investigate the recruitment behavior of other
export related factors in a strain in which Yra1 recruitment *via* the CTD is defective. The mutants identified here may contribute to our understanding of the order of recruitment of multiple mRNA packaging and export factors.

In the course of these investigations, I generated a double point mutant version of Yra1 with novel and potentially useful properties. *R107AF126A* displays wild-type growth at 30°C but is inviable at 37°C. In contrast, the previously identified Yra1 temperature sensitive mutants (Kashyap et al, 2005; Strasser & Hurt, 2000; Zenklusen et al, 2002) displayed defective growth at 30°C. The normal growth of the *R107AF126A* mutant at the permissive temperature makes it potentially more useful for future examination of Yra1 function and mRNA export in yeast.
4. Future Directions

The aim of this research project was to characterize the interaction between Yra1 and the PCTD and determine whether this interaction is important for Yra1 cotranscriptional association and function in vivo. In the course of conducting this research and from the results obtained, many new questions and directions for inquiry have emerged.

4.1. Yra1, the PCTD, and cotranscriptional recruitment

The in vitro analysis of the Yra1-PCTD interaction centered on particular phosphorylation patterns, namely Ser2P, Ser5P, and Ser2/5P. However, after the conclusion of these experiments, it became clear that Ser7P is also present at protein-coding genes in S. cerevisiae (Kim et al, 2010; Mayer et al, 2010; Tietjen et al, 2010). It would be interesting to analyze how Ser7P affects Yra1 binding. Like the Ser2/5P pattern, Ser5/7P and Ser7/2P may be sufficient for Yra1 binding. As the Ser2/5/7P pattern is probably the form present during elongation in vivo, it is likely that Yra1 will bind to this triply phosphorylated pattern as well as or better than the Ser2/5P pattern used in this study.

Based on the in vitro PCTD binding assays, I hypothesized that truncation of the N-terminus of Yra1 should create a PCTD binding-deficient form of the protein. Indeed, a yeast strain that only expressed aa 77-226 of Yra1 was viable (as it was still able to bind RNA) and displayed export defects at 37°C. This strain displayed a drastic (as much as 10-fold) decrease in protein that associated to two active genes (PMA1 and PYK1) during elongation; however, truncation of the N-terminus also created a defect in Yra1 nuclear localization. Unfortunately, this complicates the interpretation of the ChIP results because a lower steady-state level of the 77-226 truncation in the nucleus would likely
also decrease the amount of protein associated with a given gene. This subcellular localization defect is thought to be due to the deletion of a nuclear localization signal (NLS) in the N-variable region (Zenklusen et al, 2001); if this is true, addition of a strong NLS (such as KRRK) to the 77-226 truncation of Yra1 should allow us to uncouple the localization defect from the PCTD binding defect. It is possible that Yra1 may not require a classical NLS for nuclear localization based on the following: the sequence of an NLS has not been identified in Yra1; Yra1 contains a large proportion of uniformly distributed basic amino acids (22.6% of the sequence is R or K), which could function as multiple NLSs; and the small size of WT and the N-terminal truncation (26 and 18 kDa, respectively) may allow these proteins to passively diffuse through the NPC without the requirement for an NLS. Based on the results presented in this study, it is possible that truncation of the PCID actually causes the protein to spend more time in the cytoplasm due to the lack of efficient recruitment to RNAPII. Based on this hypothesis, the addition of a strong NLS to Yra1 might not completely restore nuclear localization of the 77-226 truncation. Inhibition of transcription, for instance by treatment of cells with a small molecule inhibitor, may also increase the lifetime of full-length Yra1 in the cytoplasm. In addition, altering the phosphorylation pattern of the CTD (discussed further below), which would decrease Yra1 binding, could also induce Yra1 cytoplasmic localization.

Because the in vitro experiments indicated that Yra1 specifically bound the Ser2/5P form of the CTD, the cotranscriptional association of Yra1 with genes should be sensitive to the phosphorylation state of the CTD. I therefore hypothesize that alterations in CTD phosphorylation patterns, such as those created by inhibiting CTD kinase activities in vivo, will produce abnormalities in Yra1 recruitment. I predict that Yra1 cotranscriptional chromatin association, analyzed by ChIP, will be decreased in strains carrying a deletion of the nonessential catalytic subunit of CTDK-I (ΔcTKI) or a ts allele of
the Bur1 kinase (Keogh et al, 2003); however, it may be difficult to separate CTD phosphorylation-specific effects from effects caused by inefficient elongation in these strains. Therefore, it would be beneficial to use yeast strains with analogue-sensitive alleles of these kinases, such as the bur1-as strain described previously (Qiu et al, 2009), to specifically inhibit kinase activity without affecting other elongation-related, non-enzymatic functions of these enzymes. After short-term exposure of the yeast to the small molecule ATP analogue, the association of Yra1 to particular genes could be analyzed by ChIP. Similarly to the Yra1 PCID truncation experiment, I expect that alterations of the pattern of CTD phosphorylation will decrease the amount of Yra1 that immunoprecipitates with genes. This result would provide additional evidence that Yra1 indeed requires the PCTD to associate to genes during transcription.

4.2. Yra1 RNA binding and regulation

While this study was primarily focused on CTD binding, I found that mutations of residues within the RRM of Yra1 decreased its ability to bind RNA. These mutations (F126A, Y112AF126A, and R107AF126A) were introduced into yeast, and I found that these mutant strains were viable but had defects in mRNA export. It would be interesting to examine chromatin association of these yra1 alleles to determine the role of RNA binding in Yra1 recruitment. We may find that RNA binding also plays an important role in cotranscriptional recruitment. In addition, the level of Yra1 at the gene might be affected specifically at the 3' ends of protein-coding genes in the F126A, Y112AF126A, and R107AF126A strains.

Comparing recombinant Yra1 expressed in bacteria and insect (Sf9) cells revealed that post-translational modifications probably modulate its ability to bind RNA, but not the PCTD. ALY/REF, the mammalian homolog of Yra1, has been shown to be
extensively mono- and dimethylated on arginine residues, and the modified form of the protein does not bind RNA as well as the unmodified form (Hung et al, 2010). Therefore, it is likely that the difference between the protein expressed in bacteria and insect (Sf9) cells involves arginine methylation. Yra1 purified from WT and arginine methyltransferase deficient (Δhmt1) *S. cerevisiae* should be analyzed for the ability to bind RNA. Tandem mass spectrometry techniques could be used to map the locations of these modifications on Yra1. As these modifications do not affect PCTD binding, the pattern of arginine methylation may give us more information concerning the differences between the RNA and PCTD interactions and indicate unmodified residues that are likely involved in PCTD binding. Moreover, as post-translational modifications did not affect the PCTD binding of Yra1 *in vitro*, the results of ChIP experiments in Δhmt1 would likely be the same as or similar to WT cells. These experiments would begin to unravel the regulation of Yra1 and mRNA export in *S. cerevisiae*.

### 4.3. Interdependence of the recruitment of export factors

One of the remaining, and potentially very interesting, puzzles concerning mRNA export and the proteins involved is whether the association of export and mRNA packaging proteins is coordinated or whether the proteins bind the mRNA independently. The prevailing model has been that Yra1 is recruited to genes via interaction with the other members of the TREX complex (composed of Hpr1, Tho2, Thp2, Mft1, Sub2, Yra1, and Tex1), but this has not been directly demonstrated. In this study, I created a Yra1 mutant that decreased the association of Yra1 at a gene of interest by up to 10-fold relative to WT; this mutant could be used to address the relationship between Yra1 and the rest of TREX. Using ChIP, the effect of Yra1 PCID truncation (and
its inability to associate with genes) on TREX and other mRNA packaging/export proteins could be studied. It is possible that Yra1, by binding to the PCTD, is responsible for the association of the other TREX members; therefore, ChIP of Hpr1 or Sub2 in the 77-226 strain could show that these proteins (and by extension the TREX complex) display a defect in chromatin association similar to that observed for Yra1. If Yra1 association with the transcription elongation machinery is largely independent of the other TREX members, there should be no change in the levels of Hpr1 or Sub2 that immunoprecipitate. In addition, genome-wide ChIP studies might be particularly suitable for investigating the proposal that Yra1 and TREX are differentially recruited to subsets of genes.
References


Chambers RS, Dahmus ME (1994) Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. *J Biol Chem* **269**: 26243-26248


Fong N, Bentley DL (2001) Capping, splicing, and 3’ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* **15**: 1783-1795


Harding MM (1992) NMR studies on YSPTSPSY: implications for the design of DNA bisintercalators. *J Med Chem* **35**: 4658-4664


Lee JM, Greenleaf AL (1989) A protein kinase that phosphorylates the C-terminal repeat domain of the largest subunit of RNA polymerase II. Proc Natl Acad Sci U S A 86: 3624-3628

Lee JM, Greenleaf AL (1991) CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of Saccharomyces cerevisiae. Gene Expr 1: 149-167


Lei EP, Silver PA (2002) Intron status and 3′-end formation control cotranscriptional export of mRNA. Genes Dev 16: 2761-2766


McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL (1997a) 5’-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* **11**: 3306-3318


Preker PJ, Guthrie C (2006) Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. *RNA* **12**: 994-1006


Strasser K, Hurt E (2000) Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. EMBO J 19: 410-420


Wilcox CB, Rossettini A, Hanes SD (2004) Genetic interactions with C-terminal domain (CTD) kinases and the CTD of RNA Pol II suggest a role for ESS1 in transcription initiation and elongation in *Saccharomyces cerevisiae*. *Genetics* 167: 93-105


Biography

April Lee MacKellar, daughter of John David MacKellar and Barbara Ann MacKellar, was born November 4, 1982 in Lewiston, New York. She earned a Bachelor of Science degree from Bradley University in Peoria, Illinois in 2004. April has been working on her doctoral research in the laboratory of Arno Greenleaf at Duke University since 2004. She received a Graduate/Postdoctoral Travel Fellowship Award from the American Society of Biochemistry and Molecular Biology in 2009 and a Best Thematic Poster Award from the American Society of Biochemistry and Molecular Biology in 2011. Her publications include the following:

Kanagaraj, Radhakrishnan; Huehn, Daniela; MacKellar, April; Menigatti, Mirco; Zheng, Lu; Urban, Vaclav; Shevelev, Igor; Greenleaf, Arno; and Janscak, Pavel. “RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription” (2010) Nucleic Acids Research. Dec 1; 38 (22): 8131-40.


* These authors contributed equally.