

**Targeted Proteomics and Molecular Mechanisms
of Gene Activation**

by

Christopher Carl Ebmeier

B.S., University of Nebraska-Lincoln, 1994

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Chemistry and Biochemistry

2011

This thesis entitled:
Targeted Proteomics and Molecular Mechanisms of Gene Activation
written by Christopher Carl Ebmeier
has been approved for the Department of Chemistry and Biochemistry

Dylan J. Taatjes, Ph.D.

William Old, Ph.D.

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline

Ebmeier, Christopher Carl (Ph.D., Biochemistry)

Targeted Proteomics and Molecular Mechanisms of Gene Activation

Thesis directed by Dylan J. Taatjes, Ph.D.

Mass spectrometry-based proteomics is a powerful tool when combined with hypothesis driven protein purification. Regulation of protein-protein interactions is a major molecular mechanism for gene activation. Protein purification, functional assays, and antibody-based western blots have traditionally been used to elucidate many of the most critical and perhaps universal protein-protein interactions required for gene expression, such as the assembly of the general transcription machinery. The Mediator complex is an essential part of the general transcription machinery that integrates signals from DNA-binding transcriptional activators through protein-protein interactions to regulate RNA Pol II activity. Gene activation is ultimately determined by incoming stimuli and subsequent inter-cellular signaling. How these signals are integrated in a spatial and temporal fashion for the regulation of distinct genes by activator-Mediator interactions is unclear.

One proposed molecular mechanism of gene activation by the Mediator complex is through a structural shift in the complex. Mediator structural shifts may trigger new protein-protein interactions required for transcription of select genes in response to a specific stimulus. To test this, Mediator complexes were purified with and without transcriptional activators. The activation domains of the

transcriptional activators SREBP-1a and VP16, which generate distinct structures upon binding Mediator, were used to affinity purify activator-bound Mediator complexes. For comparison, antibodies for the Mediator subunits MED1 and CDK8 were used to affinity purify activator-free Mediator complexes. A mass spectrometry-based proteomics platform was established to characterize the protein compositions of each Mediator complex purification. The results showed additional cofactors in the activator-bound Mediator complexes, many of which had known function related to gene expression. Selected cofactors were validated for binding Mediator with an orthogonal purification that combined the activator and antibody purifications and western blotting. Together the proteomics data predicted and the western blotting confirmed new protein-protein interactions relevant for regulation of gene expression that were activator-specific.

If activator-binding triggers new Mediator-cofactor interactions, could distinct activators induce distinct protein complexes that were gene-specific? This was shown upon comparing SREBP-1a and VP16. To further test this idea, other activators were used to purify Mediator complexes, and the associated polypeptides were again identified using MS-based proteomics. We evaluated Mediator complexes bound to one of three isoforms of SREBP: SREBP-1a, SREBP-1c or SREBP2. These three isoforms of SREBP were compared and unique cofactors identified. Another comparison of the activators p53 and p65 was also performed, with unique cofactors identified with each. These data provide further evidence that gene-specific protein complexes can be coordinately assembled upon activator-Mediator binding. Collectively, these

targeted proteomics approaches have generated many new hypotheses and have fundamentally altered our understanding of how gene expression is regulated.

Acknowledgements

I am grateful to many for the ability to pursue this work. My wife, Beth has been a great support from the first days of considering graduate school to the final days finishing this thesis. I would not have been able to do this without her. So I thank Beth from the deepest parts of my heart for enabling me see this through. I also thank my brother Matthew for his support and encouragement.

I have to thank Dylan for his vision and support to pursue the proteomics research I have. Dylan has been a wonderful advisor and mentor, providing a great environment to do science. I must thank Katheryn Resing and Natalie Ahn for offering me the opportunity to use a mass spectrometer in their mass spectrometry facility. I would not have been able to do what I have without it. Will Old has also been a wonderful mentor and collaborator in the mass spectrometry facility. I thank Will for all of the data he processed, his help, time and patience. I am also very grateful for all the help I have received from everyone in the CU-Boulder mass spectrometry facility, especially Stephane Houle and Shuji Kato.

I would like to thank all members of the Taatjes lab past and present, especially Jesse Goossens for his work on the SREBPs, p53 and p65/RelA, Nick Parsonette for his work on p65/RelA and Rhadika Rawat for her work on SREBP-2. I would like to thank Schuyler VanEngelenburg for his wonderful friendship and many discussions about science. There are also a number of collaborators I am very grateful to. I would like to thank Aaron Donner and Joaquin Espinosa for

a really awesome collaboration on CDK8 and P-TEFb. Joaquin has been a source of inspiration and extraordinary to work with. I would also like to thank Mike Kagey, Jamie Newman and Rick Young for a wonderful collaboration on Mediator and Cohesin. Last, but not least, I am grateful to Jacob Schwartz in the Cech lab for his friendship and fascinating collaboration on FUS and the Pol II CTD.

I would like to thank my thesis committee, Xuedong Liu, as well Natalie Ahn, Joaquin Espinosa, Will Old and Dylan Taatjes once again.

Finally, I am grateful for funding from the Creative Training in Molecular Biology Training Grant (MCDB). I have found Mostly Molecular Biology (MMB) to be a very stimulating environment to discuss science and I have very much enjoyed presenting my work in this forum.

Thesis Contents

Chapter 1	Introduction — Background and Significance.....	1
1.1	Cellular Organization and Genome Function.....	2
1.2	Transcription Factories.....	5
1.3	Gene Expression Requires General Transcription Factors...7	
1.4	The Transcription Cycle of RNA Pol II.....	9
1.5	The Mediator Complex is Regulated by Structural Shifts....	14
1.6	Mass Spectrometry-based Proteomics as a Discovery Tool for Biological Sciences.....	18
1.7	Thesis Overview.....	20
Chapter 2	Activator-Mediator Binding Regulates Mediator-Cofactor Interactions.....	22
2.1	Introduction.....	23
2.2	Results.....	24
2.2.1	Activator-Induced Structural Shifts Do Not Dissociate any Mediator Subunits.....	25
2.2.2	Activator Binding Triggers Mediator-Cofactor Interactions.....	27
2.2.3	Distinct Cofactors Associate with SREBP-Mediator vs. VP16-Mediator.....	35
2.2.4	MudPIT Analysis of CDK8-Mediator.....	37
2.3	Discussion.....	39
2.4	Methods.....	44
2.4.1	Mediator Purification.....	44
2.4.2	MudPIT Analysis of Mediator.....	45

Thesis Contents

Chapter 2	
	2.4.3 Antibodies.....47
	2.4.4 Mediator “immunodepletion” experiments.....47
Chapter 3	An Improved Proteomics Platform Allowing Analyses of Very Low Abundance Targeted Protein Samples.....50
3.1	Introduction.....50
	3.1.1 Sample Preparation: Protein Precipitation.....52
	3.1.2 Filter-Aided Sample Preparation (FASP)54
	3.1.3 Orthogonality of Two-Dimensional LC/MS.....58
3.2	Results.....62
	3.2.1 Comparison of Protein Precipitation Protocols.....64
	3.2.2 CDK8-Mediator IP from HeLa Nuclear Extract.....65
	3.2.3 SREBP-1a-Mediator from HeLa Nuclear Extract.....71
	3.2.4 Less than One Microgram VP16-Mediator and RNA Pol II.....73
3.3	Discussion.....76
3.4	Methods.....78
	3.4.1 Mediator Purifications.....78
	3.4.2 Sample preparation and proteomics analysis.....79
Chapter 4	Future Directions — Targeted Proteomics of Activator-Bound Mediator Interactomes.....81
4.1	Introduction.....81

Thesis Contents

Chapter 4

4.1.1	Activator-Specific Mediator Interactome.....	81
4.1.2	SREBPs.....	84
4.1.3	p53 and p65.....	85
4.2	Results.....	86
4.2.1	Mediator-Bound Activators SREBP (-1a, -1c, -2)....	86
4.2.2	Mediator-Bound p53 & NFKb/p65/RelA.....	94
4.3	Discussion.....	100
4.4	Methods.....	101
4.4.1	Activator-Mediator Purifications.....	102
4.4.2	Sample preparation and proteomics analysis.....	102

Chapter 5

	Future Directions — RNA Pol II Activity is regulated by post-translational modifications of the rpb1 C-Terminal Domain.....	105
5.1	Introduction.....	105
5.1.1	The Transcription Cycle.....	106
5.1.2	A CTD Code and Cotranscriptional mRNA processing.....	107
5.2	Results.....	109
5.2.1	CTD Substrates and Kinases.....	109
5.2.2	The RNA Pol II Phospho-CTD Interactome.....	112
5.2.3	A Strategy for the Identification of Site-Specific Phosphorylation on the RNA Pol II CTD.....	119

Thesis Contents

Chapter 5

5.3	Discussion.....	129
5.4	Methods.....	132
5.4.1	GST-CTD Purification.....	132
5.4.2	Purification of CTD Kinases.....	132
5.4.3	Purification of Phospho-CTD Interacting cofactors.....	132
5.4.4	Proteomics Analysis of the Phospho-CTD Interactome.....	133

Bibliography	136
---------------------------	-----

Appendix

1.	Protein Precipitation protocol.....	145
2.	FASP protocol.....	146
3.	2D-high/low pH LC/MS ATRAP protocol.....	147

Tables

Chapter 2

Table 2.1 Spectral Counts for consensus Mediator subunits.....	28
Table 2.2 MudPIT identifies a subset of factors that associate with activator-bound Mediator.....	30,31
Table 2.3 Spectral counts for cofactors observed exclusively in CDK8 IP samples.....	39

Chapter 3

Table 3.1 Proteomics Methodologies at a Glance.....	52
Table 3.2 Comparison of spin filter, short SDS-PAGE and TFE methods.....	57
Table 3.3 Practical Peak Capacity of 2D-LC.....	60
Table 3.4 CDK8 IPs comparing MudPIT and FASP/high-low at 1% FDR.....	66
Table 3.5 Additional P-TEFb-Associated cofactors identified in CDK8 IP using FASP/high-low method.....	67
Table 3.6 GST-SREBP-1a pulldowns from HeLa nuclear extract comparing MudPIT and FASP/high-low at 1% FDR.....	71
Table 3.7 GST-SREBP-1a pulldowns identify Cohesin.....	71
Table 3.8 RNA Pol II (~1ug) and VP16-Mediator (~0.5ug) totals.....	73
Table 3.9 RNA Pol II spectral counts and subunit sequence coverage.....	74
Table 3.10 Mediator spectral counts and subunit sequence coverage.....	75
Table 3.11 Pros or Cons of Proteomics Methodologies.....	76

Chapter 4

Table 4.1 SREBP isoforms Spectral Counts totals for all protein identifications, Mediator and RNA Pol II.....	88
---------------------------------------------------------------------------------------------------------------	----

Tables

Chapter 4

Table 4.2 SREBP Spectral Counts for Mediator by subunit.....	89
Table 4.3 SREBP Spectral Counts for RNA Pol II by subunit.....	90
Table 4.4 SREBP-1a Exclusive Cofactors.....	90
Table 4.5 SREBP-1c Exclusive Cofactors.....	91
Table 4.6 SREBP-2 Exclusive Cofactors.....	93
Table 4.7 p53AD and p65/RelA Spectral Count totals for all protein identifications, Mediator and RNA Pol II.....	95
Table 4.8 p53AD and p65/RelA Mediator Spectral Count totals for each subunit.....	96
Table 4.9 p53AD and p65/RelA RNA Pol II Spectral Count totals for each subunit.....	96
Table 4.10 p53AD exclusive cofactors.....	98
Table 4.11 p65/Rel A exclusive cofactors.....	99

Chapter 5

Table 5.1 Total identifications for CTD/pCTD-interactome.....	115
Table 5.2 Mediator identifications for CTD/pCTD.....	116
Table 5.3 Integrator identifications for CTD/pCTD.....	117
Table 5.4 PCIF1 is an Internal control identified only in phosphorylated CTD pulldowns.....	117

Figures

Chapter 1

Figure 1.1 Architecture of a Chromosome.....	2
Figure 1.2 Visualization of all chromosomes simultaneously in a human fibroblast.....	3
Figure 1.3 Organization of the Mammalian Cell Nucleus.....	5
Figure 1.4 Compartmentalization of Nuclear Processes.....	6
Figure 1.5 The General Transcription Machinery.....	8
Figure 1.6 The RNA Pol II transcription cycle connects each step in a continuous process.....	10
Figure 1.7 The RNA Pol II CTD Coordinates Transcription and Pre-mRNA Processing.....	11
Figure 1.8 The CTD Code and selective binding of protein.....	13
Figure 1.9 The Complexity of the CTD Code.....	13
Figure 1.10 Modular structure of Mediator and interactions with diverse activators.....	15
Figure 1.11 Activator-induced conformational shift in Mediator structure.....	16
Figure 1.12 Activator-specific structural states of the Mediator complex.....	17
Figure 1.13 Consensus Mediator Subunits Identified by MudPIT.....	19

Chapter 2

Figure 2.1 Purification protocols for MudPIT samples.....	25
Figure 2.2 Activator binding triggers new Mediator-cofactor interactions.....	32
Figure 2.3 Activator binding triggers new Mediator-cofactor interactions.....	33

Figures

Chapter 2

Figure 2.4	Orthogonal purification with ADA3L provides additional confirmation of SREBP-Mediator-SAGA interaction.....	34
Figure 2.5	Distinct cofactors associate with SREBP-Mediator vs. VP16-Mediator.....	36
Figure 2.6	Med26 and pol II do not associate with CDK8-Mediator.....	38
Figure 2.7	A model that summarizes the results and implications of this study.....	40
Figure 2.8	Immunodepletion of Mediator experiment does not completely deplete Mediator complexes.....	48

Chapter 3

Figure 3.1	Filter-aided sample preparation (FASP) for mass spectrometry.....	55
Figure 3.2	FASP removes detergents.....	56
Figure 3.3	Normalized retention time plots for two-dimensional orthogonality in LC/MS.....	59
Figure 3.4	Protein Precipitation of glycerol gradient purified Mediator fractions.....	64
Figure 3.5	Filter-Aided Sample Preparation (FASP) samples from CDK8 IPs.....	65
Figure 3.6	P-TEFb associates with CDK8-Mediator.....	68
Figure 3.7	P-TEFb associates with two distinct CDK8 complexes, CDK8-Mediator and the CDK8 four protein subcomplex.....	69
Figure 3.8	Superose 6 column fractions from ammonium sulfate precipitated HeLa nuclear extract input probing CDK8 and P-TEFb.....	70
Figure 3.9	Cohesin (Smc3) and Nipbl co-purify with mediator.....	72

Figures

Chapter 4

Figure 4.1	Activator-Specific Mediator Conformations.....	82
Figure 4.2	Sterol Regulatory Element-Binding Protein (SREBP) family of transcription activators bind Mediator.....	87
Figure 4.3	p53 and p65/Rel A bind Mediator.....	94
Figure 4.4	Venn diagram of all identifications for p53 and p65/RelA compared.....	97

Chapter 5

Figure 5.1	Purification of GST-rpb1-CTD.....	110
Figure 5.2	Purification of TFIIH.....	110
Figure 5.3	Purification of P-TEFb.....	111
Figure 5.4	Kinase reaction timecourse with GST-CTD and TFIIH or P-TEFb.....	113
Figure 5.5	Immobilization of kinase reactions with GSH-Sepharose and beads after Sarkosyl elution to verify fusion protein phosphorylation throughout the nuclear extract pulldowns.....	113
Figure 5.6	Purification of phosphorylated CTD-binding proteins.....	114
Figure 5.7	Western blot probing MED1 and MED23 in CTD/pCTD elutions....	118
Figure 5.8	CTD/pCTD pulldowns from Mediator-enriched fraction, P1.0M/Q1.0M.....	119
Figure 5.9	Beta-elimination and Michael addition of amino-ethyl thiol (AET) converts phospho-serine to lysine analog that is cleavable using Lys-C endoproteinase.....	121
Figure 5.10	Peptide mapping the CTD. Reversed-phase C18 UV 215nm trace of tryptic CTD peptides.....	123
Figure 5.11	Peptide mapping the pCTD.....	123

Figures

Chapter 5

- Figure 5.12 LC/MS total ion count (TIC) chromatograms of Beta-elimination/Michael addition of 2X CTD repeats with a single phosphorylation.....126
- Figure 5.13 Beta-elimination/Michael addition of 2X CTD repeats with a single phosphorylation.....127

Chapter 1

Introduction — Background and Significance

The first draft of the human genome was completed ten years ago (Venter, et al., 2001), providing a comprehensive view of the nearly three billion base pairs (bp) of DNA found in every cell of a human being. However, knowing the sequence of DNA that make up each gene does not unfortunately provide many clues as to how the expression of a gene is regulated, nor how the genome directly controls the function of any given cell. Elusive are the complexities of the development from a single cell zygote to a fully functional adult human.

The central dogma of molecular biology proposed by Watson and Crick dictates the flow of information proceeds very simply from DNA to RNA to protein. DNA must be replicated for cell division, however, once a cell state is defined and the cell goes to work, the expression of many genes (Class II) ultimately proceeds to protein. The processing of DNA into RNA is known as transcription. RNA, specifically messenger RNA (mRNA), is converted into protein through a process known as translation. There are many molecular mechanisms for regulating gene expression of protein coding genes.

1.1 Cellular Organization and Genome Function

All of the DNA that makes up each of our 46 chromosomes must be packaged into each cell. Approximately 146bp of DNA is wrapped around an octamer of histone proteins known as a nucleosome (Figure 1.1.B). The ordered assembly of nucleosomes is referred to as chromatin. Chromatin is organized into higher order structures that can be visualized as dense chromosomes (Figure 1.1.A, bottom). Chromatin presents one barrier for transcription factors that must be negotiated by chromatin remodeling factors (Narlikar, et al., 2002).

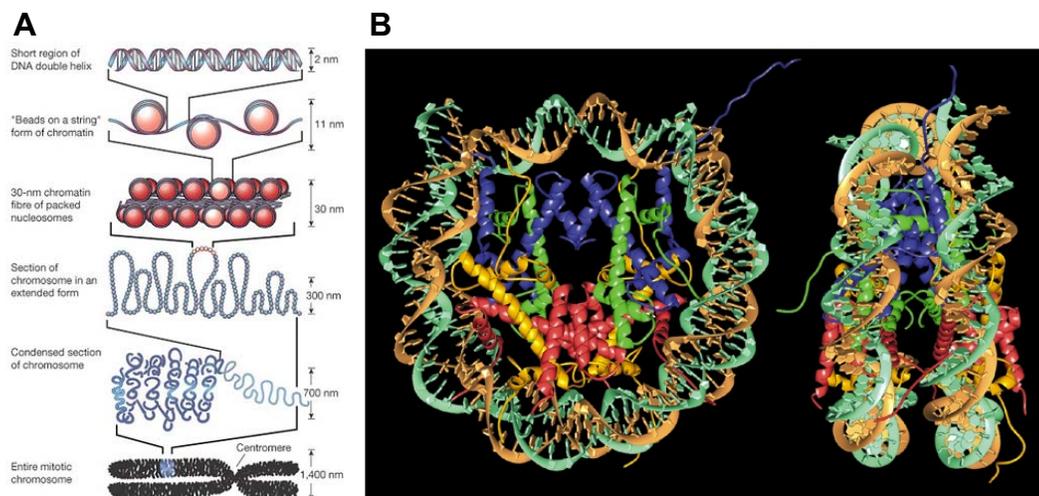


Figure 1.1 Architecture of a Chromosome. (A) DNA is packaged by wrapping 146bp around an octamer of histones to form 'Beads on a string', then further condensed into chromati (Felsenfeld & Groudine, 2003). (B) Crystal structure of a nucleosome (Luger, 2003).

Cells are organized into distinct organelles and compartment with a functional organization. The nucleus is much the same with distinct areas of inactive genes and compartments with high gene expression activity. In these

nuclear domains of high activity are many of the factors required for effective gene expression: transcription factors, chromatin remodeling proteins and mRNA-processing factors (Misteli, 2007). Chromosomes can be visualized using fluorescent probes and assigned positions in three-dimensions (Figure 1.2). Some chromosomes are localized in the nucleus relative to the center, while others localize to the edges, with preferred groups of adjacent chromosomes (Meaburn and Misteli, 2007).

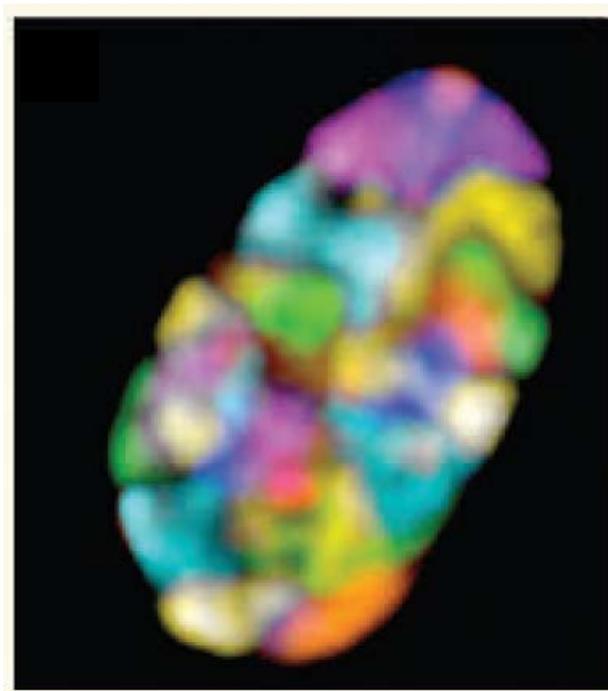


Figure 1.2 Visualization of all chromosomes simultaneously in a human fibroblast (Meaburn and Misteli, 2007).

Why and how some chromosomes localize to the center of the nucleus while other to the periphery is unknown. What is significant is that patterns of chromosome positioning is similar among cell types with common developmental

pathways and that chromosome positions in a given cell type are conserved (Meaburn and Misteli, 2007).

There are two theories of how chromosomes are positioned in the nucleus. One possibility is that a chromosome is positioned by anchoring to a nuclear scaffold. This would require some mechanism for tethering chromosomes with encoded positioning information. It could be envisioned that molecular motors would be involved for translocation of chromosome through the nuclear scaffold. A second possibility is that chromosome position is determined by the pattern of active and silent genes on any given chromosome. The local chromatin structure would be affected by the level of gene activation. Highly active genes have decondensed chromatin (euchromatin) where inactive genes are more condensed (heterochromatin). Therefore, the self-organization of a chromosome could be affected by the amount of euchromatin and heterochromatin (Misteli, 2007; Lanctôt, et al., 2007)

Compartmentalization of the nucleus allows for the concentration of factors necessary for dedicated functions, such as transcription, replication and DNA repair (Figure 1.3). An advantage of such a system would be that a single cofactor could be multifunctional depending on the compartment and the other cofactors present. For example, a population of activated transcription factors could divide the duties, with some targeting transcription and chromatin remodeling proteins while others target the designated chromosome, recruit motor proteins to escort the chromosome to a compartment enriched in

transcription factors ready to activate gene expression. This is an interesting hypothesis.

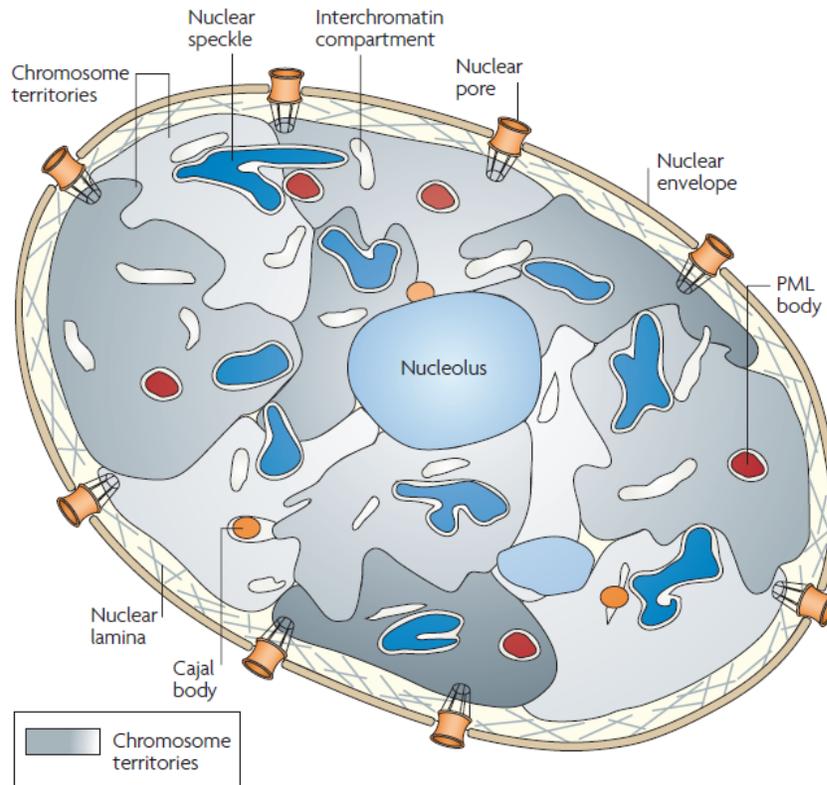


Figure 1.3 Organization of the Mammalian Cell Nucleus (Lanctôt, et al., 2007). The nucleus is compartmentalized for distinct functions. Nuclear pores are found in the nuclear envelope to regulate what goes into and out of the nucleus. The nuclear lamina is a mesh of filaments to maintain shape and structure. Chromatin is organized into chromatin territories.

1.2 Transcription Factories

Visualization of transcription shows thousands of distinct sites distributed throughout the nucleus (Figure 1.4.A) (Wansink, et al., 1993; Misteli, 2007). One view is that these are subnuclear transcription centers, or “transcription factories”

(Cook, 1999; Chakalova, et al., 2005), with a density of polymerase and all other relevant transcription factors present and in sufficient abundance for multiple genes (Cook, 1999). If there is an estimated 65,000 RNA Pol II molecules and 10,000 transcription sites in HeLa cell, then there would be approximate six Pol II per site (Misteli, 2007). This concept is analogous to RNA Pol I transcription in specialized centers (Raska, et al., 2006).

With thousands of compartmental transcription factories (Figure 1.4.A), it seems plausible that the composition of transcription factors could be distinct in different compartments. How cofactors would be delivered to the correct compartment could be complex. What is known is that transcription and

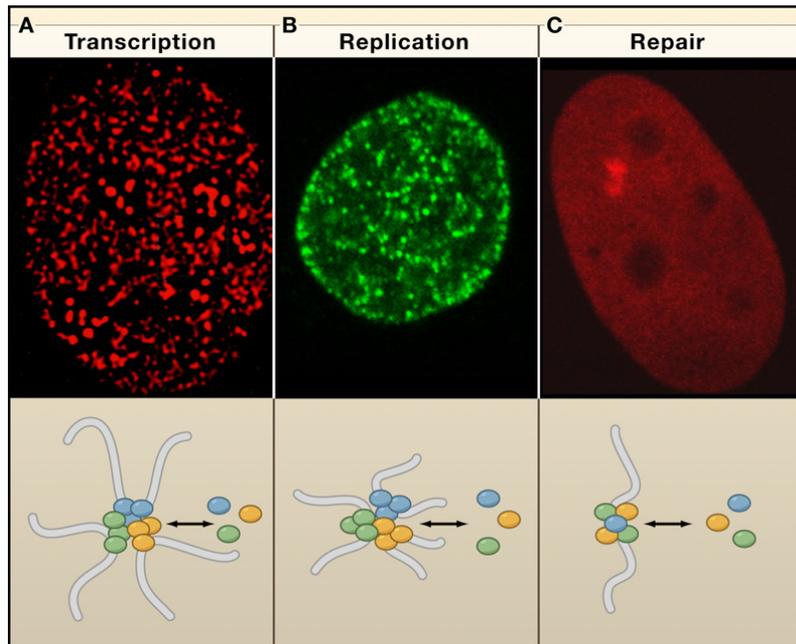
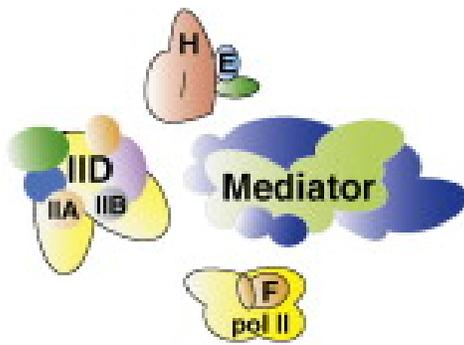


Figure 1.4 Compartmentalization of Nuclear Processes (Misteli, 2007). Transcription sites (A), replication sites (B) and DNA repair sites (C) visualized in the nucleus.

chromatin proteins bind and release in a few seconds (Phair, et al., 2004) allowing the sampling of the nuclear space for high affinity binding sites. Therefore, the stable protein-protein interactions in these sites would likely be highly regulated and protein complexes could be quite large. These complexes would likely include chromatin remodeling factors, the Mediator complex and other coactivators such as SAGA and p300/CBP as well as the general transcription machinery and RNA Pol II.

1.3 Gene Expression Requires General Transcription Factors

Eukaryotic organisms all employ a universal subset of general transcription factors (GTFs) for gene expression of protein-coding genes, which include TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, TFIIH and the Mediator complex (Thomas and Chang, 2006). The complex of GTFs and Mediator has historically been known as the pre-initiation complex (PIC). The PIC including RNA Pol II at ~4MDa is elongation ready, therefore, it is really more of a pre-elongation complex (PEC) (Figure 1.5) (Taatjes, 2010). The role of the PEC is to recognize the promoter of protein-coding genes and bring RNA Pol II to the start site of the gene.



TFIIA: 2 subunits*
TFIIB: 1 protein, 33 kDa
TFIID: 16 subunits (TBP + 15 TAFs)**
TFIIE: 2 subunits
TFIIIF: 2 subunits
TFIIH: 10 subunits
pol II: 12 subunits
Mediator: 26 subunits

T/BS

Figure 1.5 The General Transcription Machinery. (Taatjes, 2010) Each component is shown at a scale representative of each factor. *TFIIA is processed into three subunits. **TFIID composition varies.

There are two popular models for activated transcription. The first model is that a transcription activator targets chromatin remodeling factors to allow access to the gene promoter. The general transcription machinery can then assemble and initiate transcription. The second model is that the transcription activator binds to the DNA to nucleate the Mediator complex, RNA Pol II and the GTFs, forming the PEC to initiate transcription. Both models involve the recruitment of cofactors for activation; either chromatin remodeling factors (Narlikar, et al., 2002) (first model), or transcription proteins (second model). In fact, it is quite reasonable that both models can coexist at some genes at least. A transcription activator can bind chromatin remodeling factors to remove nucleosomes allow access to the enhancer binding site, where the GTFs can then be nucleated. These protein-protein interactions would need to be highly coordinated and tightly regulated for the appropriate amount of gene expression. The Mediator complex is an ideal candidate for a central regulator.

1.4 The Transcription Cycle of RNA Pol II

The traditional view of gene activation, or transcription, is (1) initiation (assembly of PEC), (2) elongation (promoter escape), and (3) termination (transcript cleavage and polyadenylation). The nascent mRNA is then spliced to remove introns, packaged and exported out of the nucleus for translation in the cytoplasm. The prevailing view is that transcription elongation and the processing of mRNA, 5' capping, splicing and cleavage/polyadenylation, are paired together and occur simultaneously (Figure 1.6) (Orphanides and Reinburg, 2002; Perales and Bentley, 2009). The export of mRNA and splicing may also be coupled (Reed and Hurt, 2002). Central to coordinating transcription and RNA processing is the C-terminal domain (CTD) of the largest subunit of RNA Pol II. The CTD consists of repeats of the heptad sequence YSPTSPS, which varies in number from organism to organism depending on genomic complexity. The yeast *Saccharomyces cerevisiae* has 26~27 repeats, *Drosophila* as 45 repeats, and humans and mice have 52 repeats (Bartkowiak and Greenleaf, 2011). The CTD is not required for catalytic activity of the polymerase, but a minimum length is required for viability (Phatnani and Greenleaf, 2006). RNA processing and the CTD were coupled from an experiment where newly synthesized RNA was crosslinked to a hyperphosphorylated form of the CTD. It is now known that the CTD is a large scaffold for binding cofactors throughout the transcription cycle (Phatnani and Greenleaf, 2006). Dynamic reversible post-translational modifications affects the binding specificity of the CTD. The

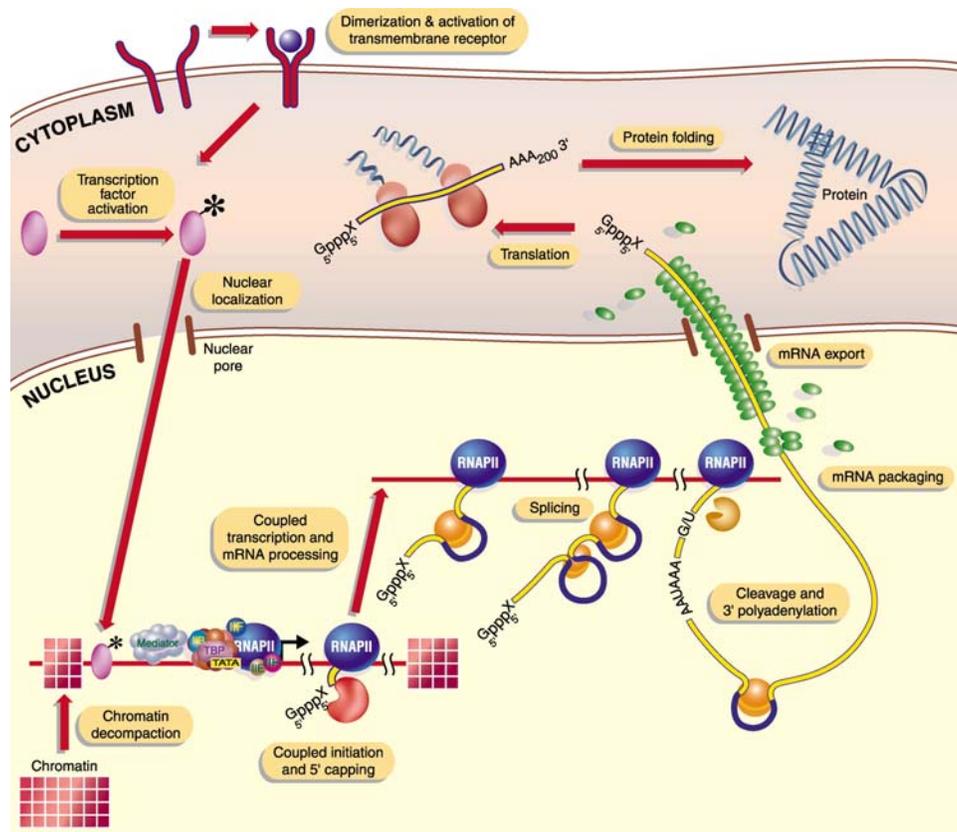


Figure 1.6 The RNA Pol II transcription cycle connects each step in a continuous process (Orphanides and Reinberg, 2002). Initiation and 5' mRNA capping are coupled. Transcription elongation and mRNA processing are coupled. mRNA processing and export may also be coupled.

phosphorylated form of the CTD helps to recruit capping factors to the 5' end of new mRNAs and 3' processing factors to poly(A) sites. The potential for differential modifications suggests a “CTD code” (Buratowski, 2003).

Phosphorylation of tyrosine, threonine and all three serines of the CTD repeats have been detected in vivo (Egloff and Murphy, 2008).

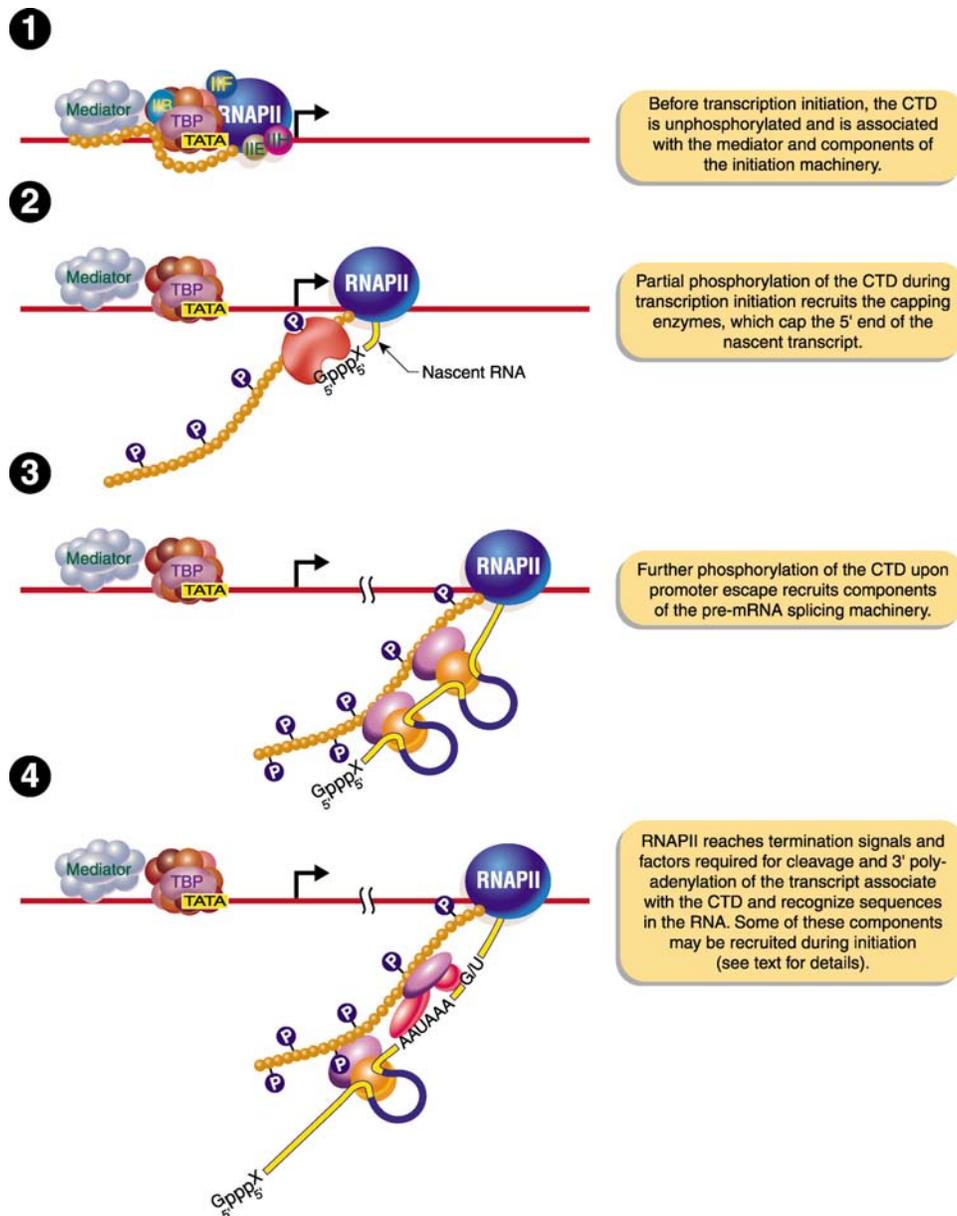


Figure 1.7 The RNA Pol II CTD Coordinates Transcription and Pre-mRNA Processing (Orphanides and Reinberg, 2002). The CTD is 52 repeats of YSPTSPS which functions as a platform for the ordered assembly of pre-mRNA 5' capping, splicing and termination.

The CTD binds the Mediator complex in the PEC in the unphosphorylated form called Pol II (A) (Kim, et al., 1994; Näär, et al., 2002) (Figure 1.7 number 1). The CDK7 kinase subunit of the GTF TFIIF phosphorylates the serine 5 and serine 7 of the repeat. This activity is dependent on the Mediator complex

(Meyer, et al. 2010; Boeing, et al., 2010; Glover-Cutter, et al., 2009) and it is thought that this event disrupts the PEC and induces dissociation of the Mediator complex (Figure 1.7 number 2) (Max, et al., 2007; Svejstrup, et al., 1997). This early phosphorylation mark induces the recruitment of 5' capping enzymes. Further phosphorylation of the CTD by the positive transcription elongation factor b (P-TEFb) phosphorylates serine 2 is consistent with promoter escape and the recruitment of pre-mRNA splicing machinery (Figure 1.7 number 3) (Orphanides and Reignberg, 2003; Bartkowiak and Greenlead, 2011; Egloff and Murphy, 2008). Serine 2 phosphorylation persists through the 3' end of the gene and may recruit cleavage and 3' polyadenylation factors (Figure 1.7 number 4). Some of the phosphorylation-specific cofactors that bind the CTD throughout the transcription cycle are shown in Figure 1.8. The number of sites phosphorylated or any specific pattern is completely unknown. Due to the number of sites in the heptad that can be phosphorylated and the number of repeats, there are a significant number of possible combinations of CTD phosphorylation (Figure 1.9). There are two prolines in each repeat that can be in either *cis* or *trans* orientation giving four possible combination for each repeat (Figure 1.9) (Egloff and Murphy, 2008). Peptidyl-prolyl *cis/trans* isomerases (PPIases), such as Pin1, may catalyze isomerization of CTD prolines in vivo. They target phosphorylated CTD with a high affinity for pSP and pTP (Xu and Manley, 2004, 2007). Phosphatases then remove the phosphorylation marks to reset the Pol II molecule for subsequent rounds of transcription.

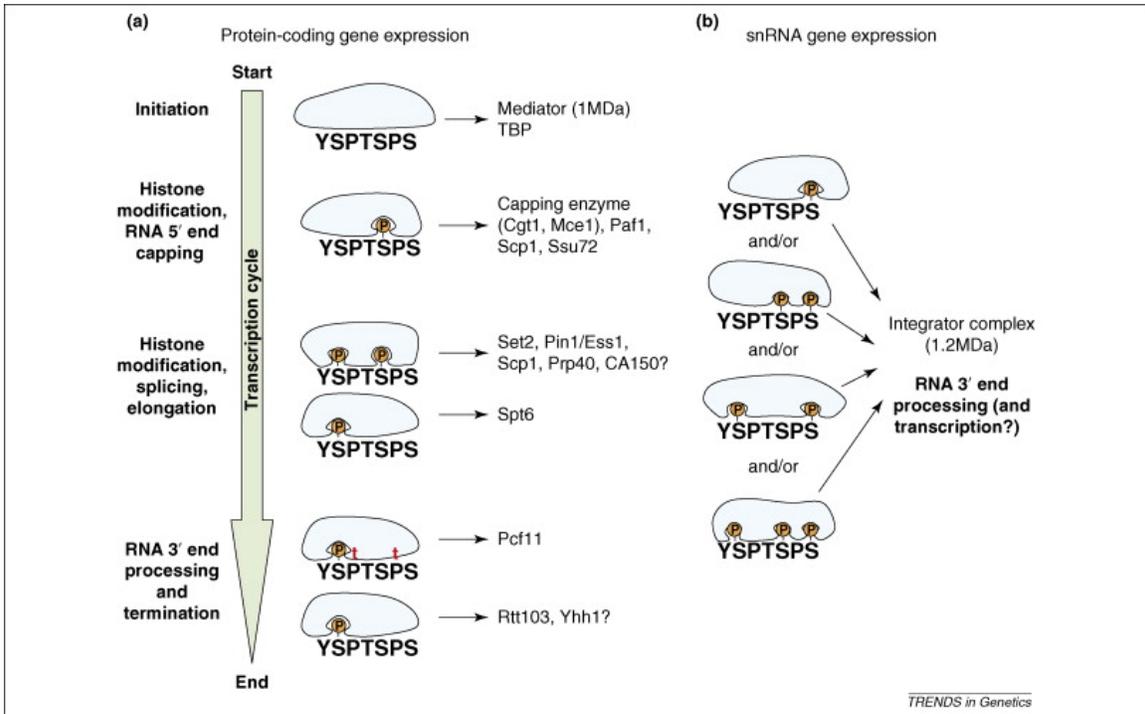


Figure 1.8 The CTD Code and selective binding of protein (Egloff and Murphy, 2008). (a) Factors involved in histone modification, 5' mRNA capping, splicing, and 3' mRNA processing bind the phosphorylated CTD. (b) The integrator complex binds the CTD for small nuclear RNA processing (Egloff, et al., 2007).

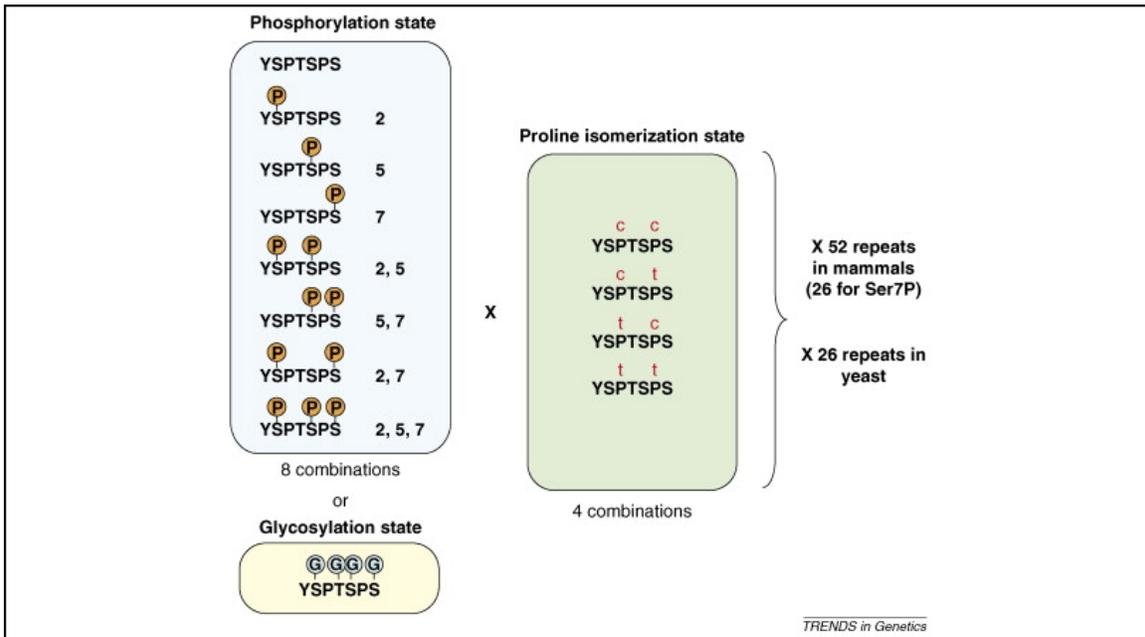


Figure 1.9 The Complexity of the CTD Code (Egloff and Murphy, 2008).

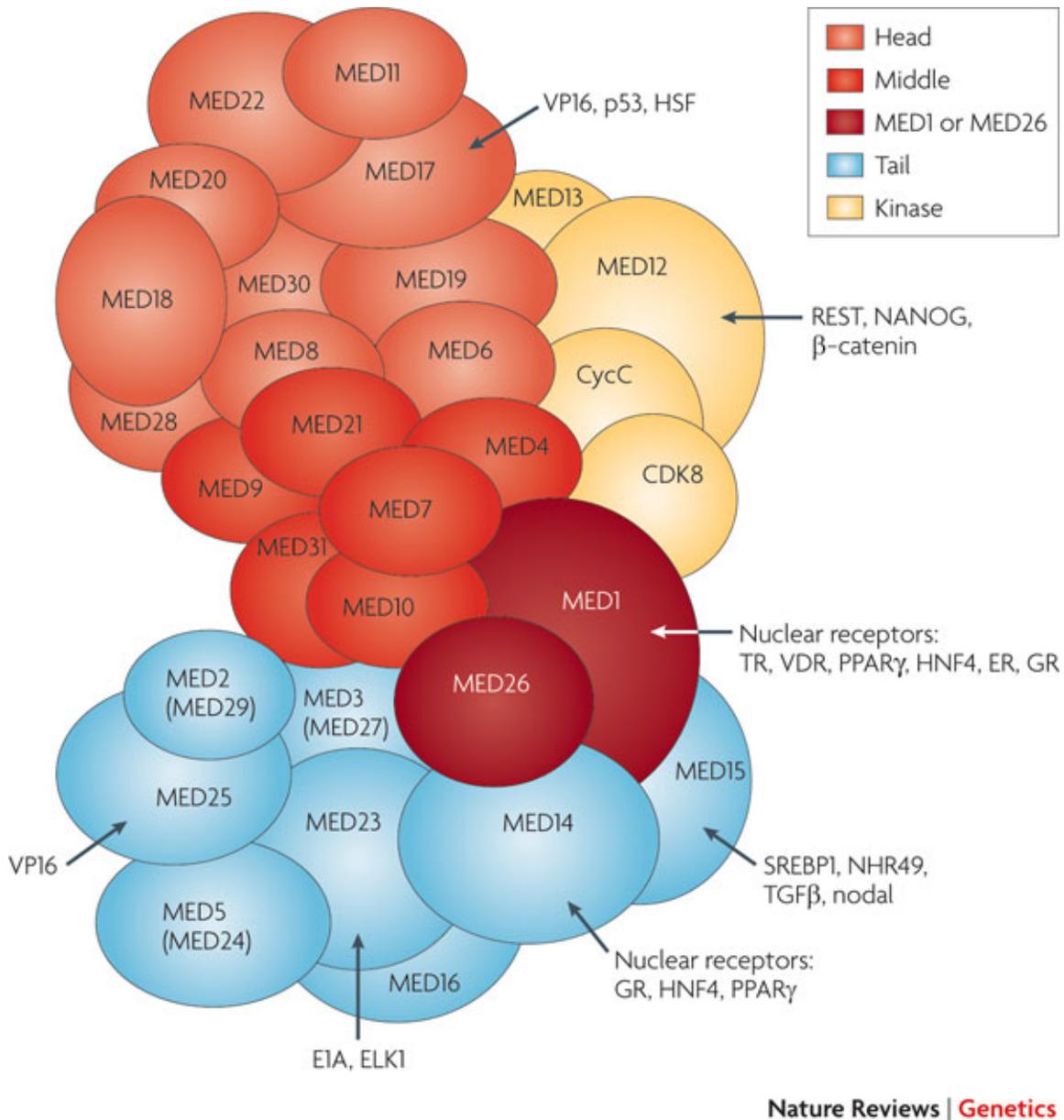
A second large protein complex that directly binds the CTD is the Integrator complex. The Integrator complex is evolutionarily conserved and made up of 12 subunits for ~2MDa. The function of this complex is processing Pol II mediated-small nuclear RNAs (snRNAs) (Baillat, et al., 2005). Furthermore, serine 7 phosphorylation of the CTD repeat is required for snRNA processing (Egloff, et al., 2007).

1.5 The Mediator Complex is Regulated by Structural Shifts

The Mediator complex is a co-activator that regulates gene activation by regulating RNA Pol II. As mentioned earlier, Mediator is really a general transcription factor that is absolutely required for expression of protein-coding genes. Mediator is 26 subunits and 1.2 MDa with an enormous surface area for protein-protein interactions. It is at the heart of the PEC and serves as a molecular bridge connecting DNA-bound activators and the GTFs. Various subunits are targeted by a number of activators (Figure 1.10) (Taatjes, et al., 2004; Malik and Roeder, 2010). The large size of Mediator accommodates a structural plasticity and enormous potential for protein-protein interactions (Taatjes, 2010).

The activator sterol regulator element-binding protein (SREBP) binds the MED15 subunit of the Mediator complex. As a result, MED15 may serve as a master

regulator of lipid homeostasis (Malik and Roeder, 2010). The tumor suppressor p53 has been shown to target MED17 and MED1 subunits, while the viral activator VP16 targets MED25. Nuclear receptors target MED1. The activators



Nature Reviews | Genetics

Figure 1.10 Modular structure of Mediator and interactions with diverse activators (Malik and Roeder, 2010). Transcription activators are shown with their respective Mediator target subunits.

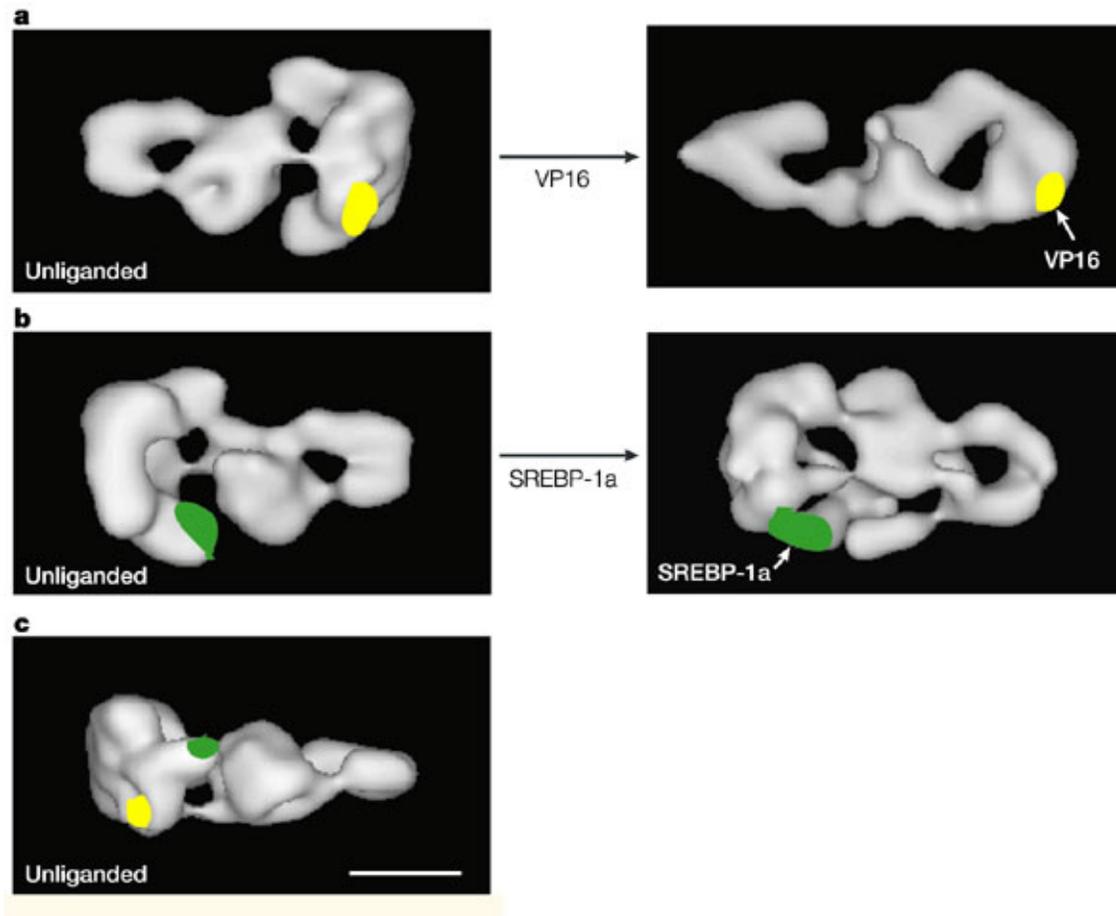


Figure 1.11 Activator-induced conformational shift in Mediator structure (Taatjes, et al., 2004). (a) SREBP-1a and (b) VP16 bind the MED15 and MED25 subunits, respectively. The yellow shows the approximate binding site for VP16 and the green for SREBP-1a. (c) An alternative view of unliganded Mediator rotated 90°.

SREBP and VP16 were used in a study to show that activator-binding induces a global structural shift in the complex that is distinct for each activator (Figure 1.11) (Taatjes, et al., 2002, 2004).

An activator-dependent structure shift would be an elegant molecular mechanism for regulating protein-protein interactions. Unique Mediator

structures are also found with the nuclear receptors vitamin D receptor (VDR) and the thyroid receptor (TR) (Figure 1.12). Despite the fact that many activators bind the Mediator complex, the unique conformation induced by an activator may confer promoter-specific function dependent on Mediator structure (Figure 1.12). The recruitment of enzymatic activities such as chromatin remodeling factors or gene-specific mRNA processing factors could be very tightly regulated with this type of mechanism.

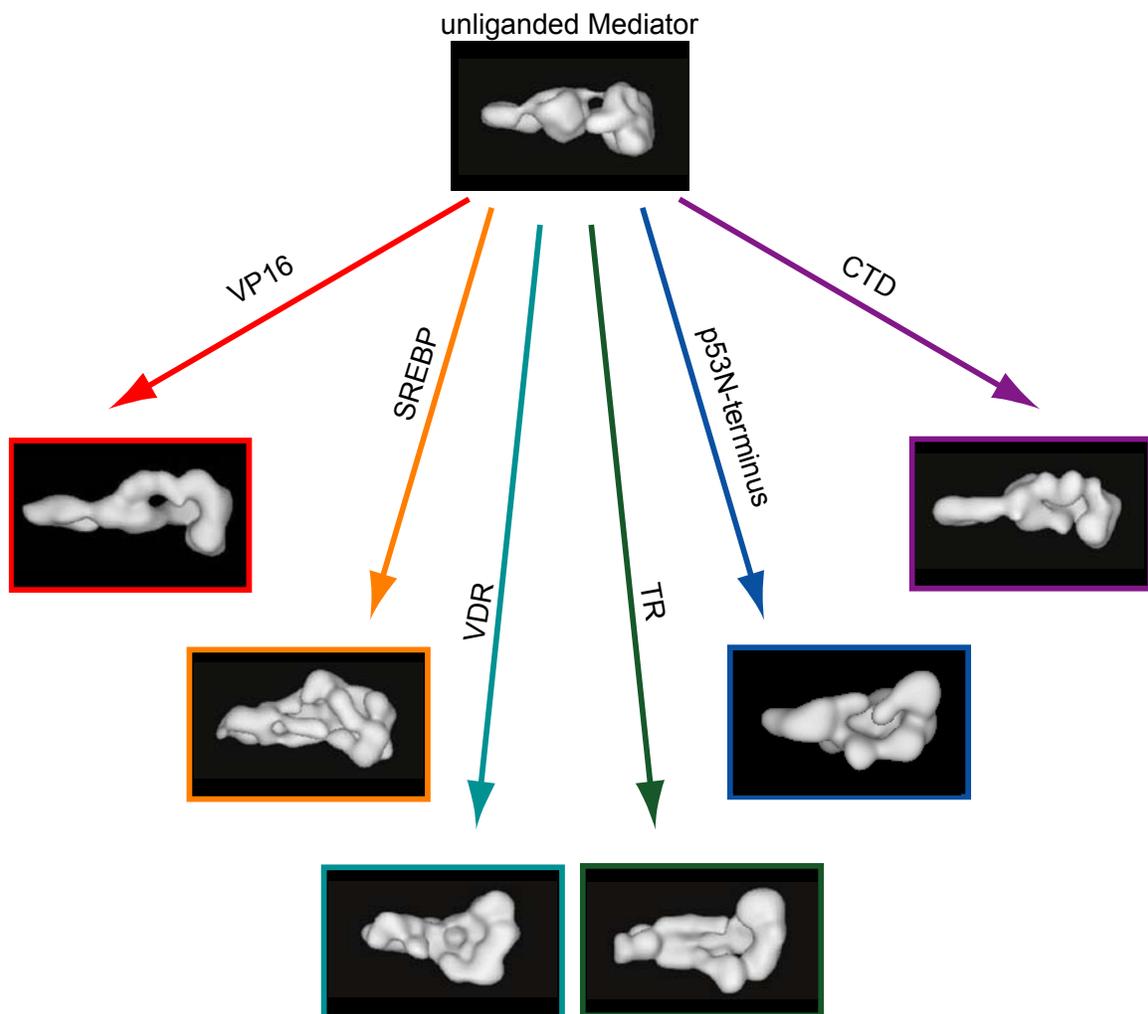


Figure 1.12 Activator-specific structural states of the Mediator complex. A RNA Pol II CTD-bound structure is also included for comparison.

1.6 Mass Spectrometry-based Proteomics Unifies Mediator

Mediator purified in different laboratories were given a different name, creating a variety of names for virtually the same Mediator complex. It was called TRAP/SMCC (thyroid hormone receptor-associated proteins/SRB-Med-containing cofactor), ARC-L (activator-recruited factor-large), DRIP (vitamin D receptor-interacting proteins), mouse Mediator, rat Mediator, PC2 (positive cofactor 2) and CRSP (cofactor required for Sp1 activation) (Sato, et al., 2004). To define the subunit composition of the Mediator complex Sato, et al. employed multidimensional protein identification technology (MudPIT) (Washburn, et al., 2001), which is a mass spectrometry-based method capable of thoroughly characterizing a complex protein sample. Five different subunits of the Mediator complex were stably transfected in HeLa cells with a Flag-tag. Thirty Mediator subunits were identified (Figure 1.13) as a set of consensus mammalian subunits. At about the same time, a unified nomenclature was agreed upon for all Mediator subunits (Bourbon, et al. 2004). The methodology used by Sato, et al., 2004, is very powerful for characterizing protein complexes. This type of mass spectrometry analysis is now more widely used for the study of protein-protein interactions such as molecular machines and protein networks (Kocher and Superti-Furga, 2007), signaling networks (Choudhary and Mann, 2010) as well as affinity purifications (AP-MS) (Gingras, et al., 2007).

	From Literature							MudPIT							MudPIT	
	Large						Small	MudPIT							MudPIT	
	TRAP/SMCC	ARC	DRIP	mMED	ratMED	PC2	CRSP	HeLa	F:Nut2 (MED10)	F:Med25 (MED9)	F:Intersex (MED29)	F:LCMR1 (MED19)	F:MED28	F:CRSP70 (MED26)	P0.5	P1.0
TRAP240L (MED13L)								23	14	18	20	12	18	16		
TRAP240 (MED13)								29	20	26	21	10	2	15		
TRAP230 (MED12)								34	25	23	24	13	19	23	<1*	
TRAP220 (MED1)								47	38	34	33	29	40	43	18	
Rgr1 (MED14)								55	36	50	36	32	41	19	8	
Sur2 (MED23)								39	31	37	29	23	26	12	1	
ARC105 (MED15)								23	15	23	18	17	23	10	5	
TRAP100 (MED24)								59	38	54	35	31	45	32	20	
TRAP95 (MED16)								44	17	41	21	30	19	9	3	
ARC92 (MED25)								20	8	19	20	15	12			
TRAP80 (MED17)								50	42	51	38	56	40	34	20	
CRSP70 (MED26)								26	19	28	23	21	72	23	21	
Cyclin C								6		14	6	4				
Cdk8								33	18	27	23	43	18	22		
Cdk11								22	9	15	9	37	11	10		
TRAP37 (MED27)								44	35	40	39	40	34	16	9	
TRAP36 (MED4)								47	45	48	49	26	50	49	30	
LCMR1 (MED19)								51	13	51	43	32	49	22	26	
MED6								59	39	52	51	56	56	31	19	
MED7								37	22	37	23	22	50	27	19	
MED8								50	54	46	47	47	50	56	35	
p28b (MED18)								53	31	35	44	42	27	32	21	
TRFP (MED20)								48	51	34	47	25	53	50	26	
Med25 (MED9)								31	31	19	31	12	31	45	32	
Intersex (MED29)								46	43	75	34	28	52	27	34	
TRAP25 (MED30)								43	62	43	43	26	61	56	20	
MED28								25	22	18	20	20	35	18	13	
Srb7 (MED21)								58	68	59	46	22	60	49	24	
Surf5 (MED24)								54	59	43	47	26	71	75	69	
HSPC296 (MED11)								45	45	44	48	37	50	34	16	
Soh1 (MED31)								60	36	52	60	46	35	41	34	
Nut2 (MED10)								66	59	59	80	73	70	40	40	

Figure 1.13 Consensus Mediator Subunits Identified by MudPIT (Sato, et a., 2004). Mammalian Mediator subunits identified in different Flag-tagged Mediator subunits. Mediator purified in different labs received different names. (A) Red boxes indicate Mediator subunits present in TRAP/SMCC, ARC, DRIP, CRSP, PC2, mouse and rat Mediator. (B) Sequence coverage for identified subunits (red) for each Flag-tagged Mediator fraction. (C) Phosphocellulose fractions show two distinct Mediator complexes.

1.7 Thesis Overview

The proper regulation of protein-protein interactions in a spatial and temporal fashion is absolutely fundamental for gene expression. The nucleus of a cell is crowded with proteins binding and unbinding. The Mediator complex is a large protein complex with enormous surface area for protein-protein interactions. Distinct activators induce unique structures for the complex, which could be an elegant molecular mechanism for regulating protein-protein interaction. Only when a transcription activator is bound to the enhancer would it be required to recruit transcription cofactors.

In Chapter 2, we provide evidence for the hypothesis that an activator-induced structural shift in the Mediator complex induces new protein-protein interactions. We apply the MudPIT protocol, used to identify the consensus Mediator subunits, to characterize purified Mediator complex without activator ligand and activator-bound Mediator both from HeLa nuclear extract. We identify a subset of cofactors only found in the activator-bound Mediator fractions. Select cofactors were characterized with biochemical assays to assess if they were Mediator-bound or activator-bound. A few were bound only to the activator, while most were Mediator-associated.

Chapter 3 details a method development strategy to improve the mass spectrometry methodology. Many activator-Mediator fractions would be too dilute for the MudPIT analysis used in Chapter 2. Furthermore, technical issues drove the development of an all new method. Significant changes were made to

the protein precipitation and proteolysis strategies as well as the multidimensional chromatography used. Implemented modifications vastly improved sensitivity and ruggedness. A Mediator immunoprecipitation and a activator-Mediator fraction are analyzed with the improved method and directly compared to data generated using the MudPIT strategy. Samples of less than 1 ug highly purified Mediator and 2 ug RNA Pol II are also analyzed to demonstrate the sensitivity of the assay. All subunits of each complex were identified.

In Chapter 4, we put the new improved proteomics platform to work characterizing activator-Mediator fractions from all three isoforms of sterol regulatory element-binding protein (SREBP), p53 and p65/RelA. Unique cofactors as well as Mediator and RNA Pol II are identified with each activator. Many interesting hypotheses can be derived from the datasets.

Chapter 5 focuses more on the RNA Pol II CTD. To probe the interactome of the full length endogenous CTD, a GST-fusion was used to purify CTD-interacting cofactor which were then identified using the proteomics platform. Next, CTD kinases were purified and used to phosphorylate the CTD in vitro to again purify phospho-CTD-interacting cofactors. The kinases TFIIF and P-TEFb were used individually and combined to generate distinct phosphorylation patterns on the CTD. Unique cofactors are again identified with each phospho-form of the CTD. Finally, preliminary data is shown for a mass spectrometry-based strategy to identify exact residues of phosphorylation throughout the length of the CTD. This method would enable the determination of a CTD code, or a unique pattern of modification on the CTD.

Chapter 2

Activator-Mediator Binding Regulates Mediator-Cofactor Interactions

The 26-subunit, 1.2 MDa human Mediator complex is essential for expression of perhaps all protein-coding genes. Activator binding triggers major structural shifts within Mediator, suggesting a straightforward means to spatially and temporally regulate Mediator activity. By using MudPIT mass spectrometry and other techniques, we have compared the subunit composition of Mediator in three different structural states: (1) bound to the activator SREBP-1a, (2) bound to the activator VP16, or (3) an activator-free state. As expected, consensus Mediator subunits were similarly represented in each sample. However, we identify a set of cofactors that interact specifically with activator-bound but not activator-free Mediator, suggesting activator binding triggers new Mediator-cofactor interactions. Furthermore, MudPIT combined with biochemical assays reveals a non-overlapping set of co-regulatory factors associated with SREBP-Mediator vs. VP16-Mediator. These data define an expanded role for activators in regulating gene expression in humans and suggest that distinct, activator-induced structural shifts regulate Mediator function in gene-specific ways.

2.1 Introduction

Transcriptional regulation is driven in large part by transcription factors: DNA-binding proteins that target specific regulatory sites within the genome. Different transcription factors (or activators) recognize different sequence elements via their DNA-binding domains, whereas distinct activation domains within transcription factors interact with one or perhaps several components of the transcriptional machinery. One of the main activator targets within the transcriptional machinery is the Mediator complex (Conaway, et al., 2005; Malik & Roeder, 2005). Direct activator-Mediator interactions are thought to recruit and stabilize Mediator at the promoter; however, EM analyses of Mediator bound to different activation domains have indicated that activators may be playing other regulatory roles (Taatjes, et al., 2004). In particular, activator binding induces significant structural shifts within Mediator (Taatjes, et al., 2002), which imply an additional means to regulate the human Mediator complex. The sheer size (approximately $320 \times 180 \times 160 \text{ \AA}$) and shape of Mediator provides an enormous surface area for protein–protein interactions, and the global structural shifts induced by activator binding likely expose distinct motifs within the Mediator complex. Potentially, such structural shifts may activate Mediator by triggering protein–protein interactions at the promoter.

To test this hypothesis, we purified Mediator in three different structural states. In one instance, Mediator was purified bound to the activator SREBP-1a; in another, Mediator was purified bound to the activator VP16; Mediator was also

purified in an activator-free state (without an activator bound). Relative to the activator-free conformation, Mediator will adopt significantly different structural states when bound to SREBP-1a vs. VP16 (Taatjes, et al., 2002)). To assess Mediator subunit composition and potential associated factors in each structural state, we utilized the multidimensional protein identification technology (MudPIT) methodology, which enables a comprehensive, unbiased assessment of protein composition within even highly complex samples (Chen, et al., 2006). Our results provide clear evidence that activator-induced structural shifts trigger Mediator-cofactor interactions. Moreover, a subset of factors interact specifically with SREBP-Mediator but not VP16-Mediator, suggesting distinct activator-dependent structural shifts within Mediator direct gene-specific regulatory functions. These data indicate that activator binding can dictate subsequent Mediator-cofactor interactions, providing a straightforward means by which Mediator activity (i.e., transcription) can be controlled in a spatial and temporal fashion.

2.2 Results

The MudPIT method uses a two-dimensional liquid chromatography separation method coupled to a tandem mass spectrometer for detection of tryptic peptides. A major advantage of this technique is that protein samples are digested in solution, which minimizes sample loss that is typical for standard in-gel protein sample preparation methods. This feature is especially important for

analysis of Mediator, given the low-abundance of this macromolecular complex in human cells.

2.2.1 Activator-Induced Structural Shifts Do Not Dissociate any Mediator Subunits.

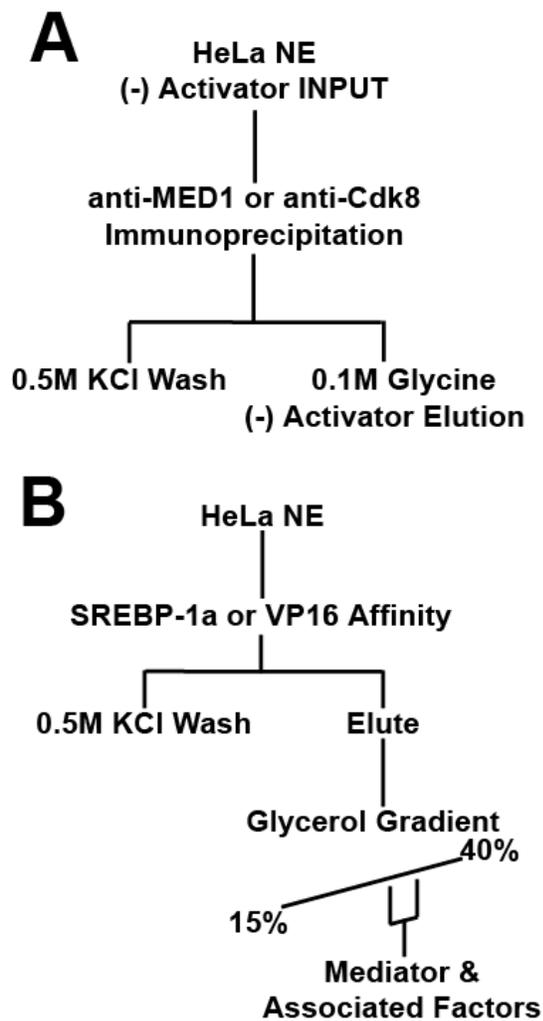


Figure 2.1. Purification protocols for MudPIT samples. (A) Activator-free Mediator immunoprecipitation purification scheme. (B) Activator-bound Mediator purification scheme.

In past studies, we have purified human Mediator by using a combination of ion exchange and affinity chromatography steps, followed by glycerol gradient sedimentation. This rigorous purification protocol yields Mediator complexes that are devoid of additional associated factors. The purpose of this work was to use MudPIT to address whether additional cofactors might stably interact with Mediator and, if so, whether these interactions were dependent upon activator-induced Mediator structural shifts. Consequently, we adopted a simplified purification scheme such that potential Mediator-associated factors could be identified (Figure 2.1). Importantly, the simplified purification protocol still included

a series of high-salt washes to remove weak, nonspecific interactions. To purify Mediator in an activator-free state, we isolated complexes from HeLa nuclear extract by using antibodies against the Med1 or CDK8 subunits. Activator-bound Mediator complexes were purified with affinity chromatography resins containing the activation domain of VP16 (residues 411–490) or SREBP-1a (residues 1–50). After elution, the activator-bound Mediator samples were further purified over a glycerol gradient, and Mediator-containing fractions (>1 MDa in size) were combined for MudPIT analysis.

After analysis of Mediator samples with the MudPIT protocol, identified peptides were subjected to a rigorous 1% false discovery rate threshold; furthermore, peptides that could be assigned to multiple different proteins were assigned with the Isoform Resolver algorithm (Meyer-Arendt, et al., 2011) and quantified by the spectral counting method (Old, et al., 2005). As shown in Table 2.1, all consensus Mediator subunits (Sato, et al., 2004) were identified in each sample (SREBP-Mediator, VP16-Mediator, CDK8 IP, and Med1 IP), with one exception: Med26 was not observed in the CDK8 IP sample. This was expected upon the basis of past work that demonstrated a mutually exclusive association of CDK8 and Med26 within Mediator (Taatjes, et al., 2002). Total spectral counts, defined as an MS/MS event identifying a peptide corresponding to a specific protein, are shown for each of the four different Mediator samples in Table 2.1. With a few exceptions, spectral counts were similar—typically within a 2- to 3-fold range—for each of the 32 consensus Mediator subunits (26 core Mediator subunits, plus the CDK8 submodule). Moreover, when all spectral counts were

summed for activator-bound Mediator and activator-free Mediator (IPs), there was less than a 2-fold difference in total Mediator spectral counts. Importantly, this indicates the total amount of Mediator was similar within each of the four MudPIT samples analyzed. These data also demonstrate that, whereas activator binding alters the conformational state of Mediator, activator induced structural shifts—at least those induced by SREBP or VP16—do not cause dissociation of any Mediator subunit.

2.2.2 Activator Binding Triggers Mediator-Cofactor Interactions.

The data outlined in Table 2.1 represent expected results from MudPIT analysis of Mediator and are consistent with a previous study by the Conaway lab, in which activator-free Mediator samples (i.e., purified with an antibody) were examined by MudPIT (Sato, et al., 2004). The motivation for this study, however, was to determine whether additional cofactors might preferentially associate with activator-bound Mediator complexes, which will adopt distinct structural states compared with activator-free Mediator. As shown in Table 2.2, the MudPIT data provide support for the hypothesis that activator-dependent structural shifts trigger Mediator-cofactor interactions. Spectral counts for identified cofactors were summed for activator-bound Mediator complexes (VP16 or SREBP) and for activator-free Mediator (CDK8 IP or Med1 IP). Cofactors identified that were greater than 4-fold enriched in activator-bound Mediator samples are shown in Table 2.2. Note that for most cofactors, the fold enrichment represents a

Table 2.1 Spectral Counts for consensus Mediator subunits.

calculated MW (KDa)	Mediator subunit	Spectral Counts			
		MED1	CDK8	SREBP	VP16
168.4	MED1	30	68	216	159
29.7	MED4	64	37	117	76
28.4	MED6	21	12	10	38
27.2	MED7	1	1	58	16
32.8	MED8	108	19	61	62
16.4	MED9	21	1	9	27
15.7	MED10	25	27	67	35
13.1	MED11	117	92	155	106
160.7	MED14	54	75	140	106
86.8	MED15	106	127	140	165
96.8	MED16	23	18	114	63
72.9	MED17	80	27	148	48
23.7	MED18	12	22	20	43
26.3	MED19	18	9	30	5
23.2	MED20	36	55	29	66
15.6	MED21	60	31	98	64
16.5	MED22	11	51	30	81
156.2	MED23	38	50	121	190
110.3	MED24	181	63	224	137
78.9	MED25	4	7	17	28
65.5	MED26	19	0	40	25
35.4	MED27	3	17	25	16
19.5	MED28	18	3	8	18
23.5	MED29	13	7	20	8
20.3	MED30	2	10	26	5
15.8	MED31	4	4	24	17
53.3	CDK8	2	8	13	2
	CDK8 or CDK19	10	4	13	27
56.8	CDK19	3	0	2	6
35.6	Cyclin C	15	5	9	13
243.1	MED12	66	134	270	143
239.2	MED13	44	77	202	84
242.6	MED13L	11	15	74	18

minimum value because most proteins listed in Table 2.2 had no detectable peptides in the activator-free Mediator samples (CDK8 IP or Med1 IP). To ensure that potential Mediator-cofactor interactions were not occurring indirectly via any nucleic acid tether, we completed control experiments in which Mediator samples were treated with benzonase, a promiscuous endonuclease that cleaves both single- and double-stranded RNA or DNA. Benzonase treatment did not affect the presence of Mediator-associated factors, as determined by quantitative Western blot analysis. Furthermore, we performed SREBP-Mediator purifications by using a different affinity tag (MBP instead of GST) to confirm that the cofactors identified in Table 2.2 did not result from potential interactions with GST.

It was possible that enrichment of some factors listed in Table 2.2, such as CBP, might result from a direct interaction with SREBP (Oliner, et al., 1996) and not Mediator itself. Our purification protocol (Figure 2.1) includes a glycerol gradient sedimentation step that will separate complexes upon the basis of size; complexes of approximately 1.0 MDa and greater were selected for MudPIT analysis. Thus, a 350 kDa SREBP-CBP binary complex (for example) would be separated during the glycerol gradient step. However, as an additional means to ensure the factors identified in Table 2.2 associate with SREBP-Mediator and not SREBP itself, we completed an orthogonal purification scheme, outlined in Figure 2.2.A. This protocol first involved isolation of SREBP-Mediator complexes with an SREBP affinity resin, as before. Following elution from the resin, complexes were passed over an anti-Med1 or an anti-CDK8 antibody column to ensure that

Table 2.2 MudPIT identifies a subset of factors that associate with activator-bound Mediator.

Transcriptional Regulators		Spectral Counts						MW	description
Fold enrichment (activator bound)	gene	alt. name	SREBP	VP16	CDK8 IP	Med1 IP	MW		
85	TRRAP		79	6	0	0	437600	general co-regulator	
54	SKIL	SnoN	31	23	0	0	77004	co-repressor assoc. with Smads	
22	CREBBP	CBP	22	0	0	0	265351	acetyltransferase	
15	TXLNA	a-taxilin	15	0	0	0	61891	general co-activator	
11	ATM		11	0	0	0	350644	Serine/Threonine-protein kinase	
9	EP300	p300	9	0	0	0	264144	acetyltransferase	
7	VPS39		7	0	0	0	101809	co-regulator; assoc. with Smad complexes	
5	PRKDC	DNA-PK	5	0	0	0	469089	DNA-dependent kinase	
5	ATR		5	0	0	0	301367	Serine/Threonine-protein kinase	
5	MMS19L	MET18	5	0	0	0	113276	interacts with/regulates TFIIF	
5	CXorf15	g-taxilin	5	0	0	0	60606	general co-activator	
4	YEATS4	Gas41	0	4	0	0	26499	Tip60/SRCAP component	
4	GCN5L2	GCN5L	1	3	0	0	93836	acetyltransferase	
152	LRPPRC	LRP130	152	0	0	0	158926	mRNA binding/processing	
50	GCN1L1		50	0	0	0	292930	mRNA processing	
25	C14orf156	SLIRP	24	1	0	1	12349	binds stem-loop RNA	
11	SKIV2L2	DOB1	11	0	0	0	117805	processing/degradation of nuclear RNAs	
22	RUVBL2	reptin/Tip49	14	8	0	1	51157	DNA-dependent ATPase	
20	RUVBL1	pontin/Tip48	17	3	0	0	50228	DNA-dependent ATPase	
17	RECQL5		7	10	0	0	108858	DNA helicase/ATPase	
4	NAV2		4	0	0	1	268133	DNA helicase/ATPase	
8	FAM48A		8	0	0	0	80102	SAGA component	
8	SUPT3H	SPT3	8	0	0	0	35793	SAGA component	
7	TAF5L		7	0	0	0	66155	SAGA component	
7	MGC21874	ADA2B	7	0	0	0	48470	SAGA component	
5	TADA3L	ADA3L	5	0	0	0	48902	SAGA component	
57	POLR2C	Rpb3	45	12	0	1	31441	pol II subunit	
47	POLR2B	Rpb2	33	14	0	1	133897	pol II subunit	
37	CRSP9	Med7	58	16	1	1	27245	Mediator subunit	
11	POLR2E	Rpb5	9	2	0	1	24551	pol II subunit	
9	POLR2H	Rpb8	8	1	0	1	17143	pol II subunit	

Table 2.2. Continued. MudPIT identifies a subset of factors that associate with activator-bound Mediator.

Nuclear Organization/Chromatin			Spectral Counts					MW	description
Fold enrichment (activator bound)	gene	alt. name	SREBP	VP16	CDK8 IP	Med1 IP			
54	KIF21A		26	28	0	0	0	6196	kinesin motor protein
23	TUBB4		23	0	0	0	0	49586	tubulin
16	DYNC1H1		16	0	0	0	0	532408	dynein heavy chain
15	TUBA4A		28	1	1	1	1	50152	tubulin
6	TUBA1B		73	1	11	1	1	50152	tubulin
6	IQGAP1		19	0	2	1	1	189252	actin binding; assoc. with HMMR-CD44
6	SYNE2	nesprin-2	5	1	0	0	0	796442	binds lamin & emerin at nuclear envelope
5	TUBB6		28	0	2	4	4		tubulin
4	DYNC12		4	0	0	0	0	68426	microtubule motor protein
17	HMMR		17	0	0	0	0	84031	chromosome segregation; binds actin
8	ZW10		8	0	0	0	0	88829	chromosome segregation; binds dynein
6	SMC1A		6	0	0	0	0	143233	Cohesin subunit
6	SMC3		6	0	0	0	0	141542	Cohesin subunit
4	NIPBL		4	0	1	0	0	304344	Cohesin loading/unloading factor
22	RUVBL2	reptin/Tip49	14	8	0	1	1	51157	H2A.Z exchange
20	RUVBL1	pointin/Tip48	17	3	0	0	0	50228	H2A.Z exchange
11	EP400	p400	7	4	0	0	0	339981	H2A.Z exchange
Fatty acid metabolism									
97	HADHA		97	0	0	0	0	83000	tri-functional enzyme; b-oxidation of fatty acids
21	AC-ADVL		21	0	0	0	0	75209	acyl CoA dehydrogenase
9	HADHB		9	0	0	0	0	51396	tri-functional enzyme; b-oxidation of fatty acids
5	DECR1		5	0	0	0	0	36068	dienoyl CoA reductase
Others									
10	CCDC131		10	0	0	0	0	226356	PROLINE/SERINE-RICH COILED-COIL 2
9	CCDC87		9	0	0	0	0	96368	COILED-COIL DOMAIN-CONTAINING PROTEIN 87
7	LOH12CR1		7	0	0	0	0	22222	putative tumor suppressor protein

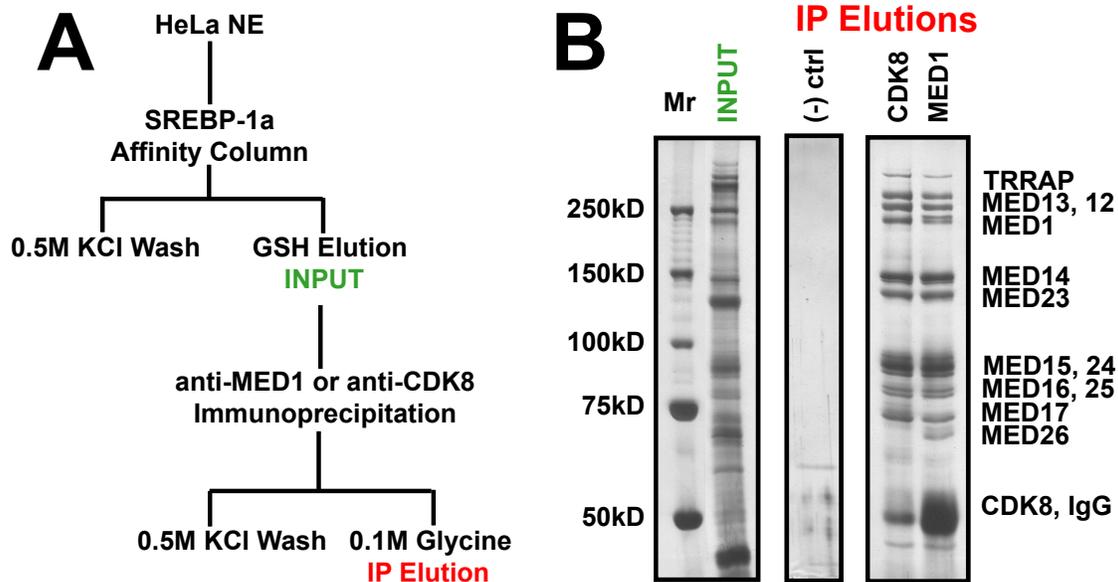


Figure 2.2. Activator binding triggers new Mediator-cofactor interactions.
(A) Orthogonal purification scheme for confirmation of Mediator-associated cofactors.
(B) Silver-stained 7% acrylamide gels representing various stages in the orthogonal purification: IP input, A/G-beads only negative control, and CDK8 or MED1 IP elutions. Subunit identities are listed at the right.

only Mediator complexes would be retained, whereas potential SREBP-CBP complexes (for example) would flow through. After a series of high-salt washes, proteins were eluted from the Mediator antibody resin (Figure 2.2.B). The presence or absence of additional Mediator-associated factors was then examined by western blot analyses.

As shown in Figure 2.3, the results from this orthogonal purification protocol confirms that many cofactors identified in Table 2.2 are in fact Mediator-associated, as they were retained by the second, Mediator-specific antibody affinity resin. Note, however, that the orthogonal purification procedure did identify a few cofactors (LRP130, HADHA, SKIV2L2, SnoN) that do not appear to interact directly with Mediator and likely associate with SREBP itself. These results highlight the effectiveness of the orthogonal purification strategy in

confirming whether cofactors identified in Table 2.2 are truly Mediator-associated. The 12 factors probed following the orthogonal purification Figure 2.3 represent a good cross-section of the cofactors identified in Table 2.2. That is, cofactors representing different activities (e.g. mRNA processing, acetyltransferase, H2A.Z exchange, etc.) were examined. As expected, alternate orthogonal purification protocols further confirmed the results shown in Figure 2.3. For example, orthogonal purification of SREBP-Mediator with an anti-ADA3L antibody resin similarly supported an SREBP-Mediator-SAGA interaction (Figure 2.4). Although past reports have suggested a direct interaction between human

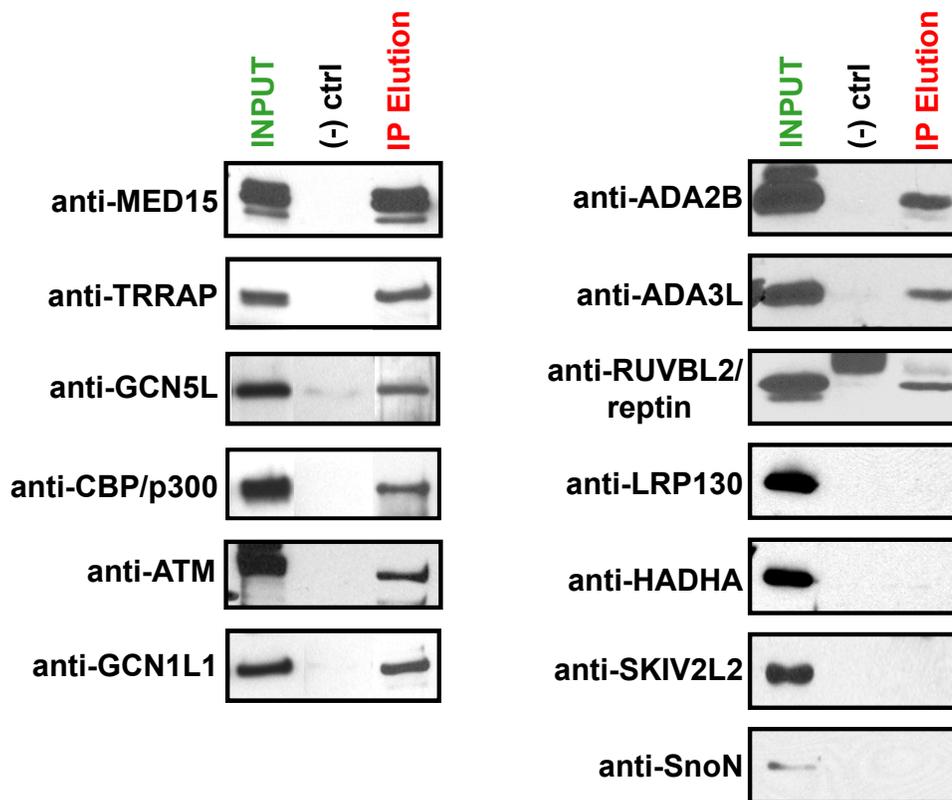


Figure 2.3. Activator binding triggers new Mediator-cofactor interactions. Western blots for various cofactors identified from the MudPIT analysis. Note most, but not all, cofactors were confirmed as Mediator-associated in this assay. LRP130, HADHA, SKIV2L2, and SnoN did not track with Mediator throughout the orthogonal purification, suggesting these factors likely interact directly with the SREBP activation domain and not Mediator.

Mediator and SAGA (Liu, et al., 2008), these data implicate activator-induced structural shifts in promoting and/or stabilizing these interactions. Not every SAGA subunit was identified in Table 2.2, which likely results from its association with only a subset of SREBP-Mediator complexes (i.e., SAGA is substoichiometric relative to Mediator itself). Experiments in which Mediator was immunodepleted from extracts yielded data that also supported the orthogonal purification results; however, such experiments are limited by the fact that Mediator cannot be effectively removed in this way, even following six immunodepletion steps (See Methods and Figure 2.8).

Because eight of twelve factors tested positively through the orthogonal purification protocol (Figure 2.3), it is evident that the majority of factors identified in Table 2.2 are likely Mediator-associated and are not observed because of potential direct interactions with the activation domain itself (VP16 or SREBP-1a).

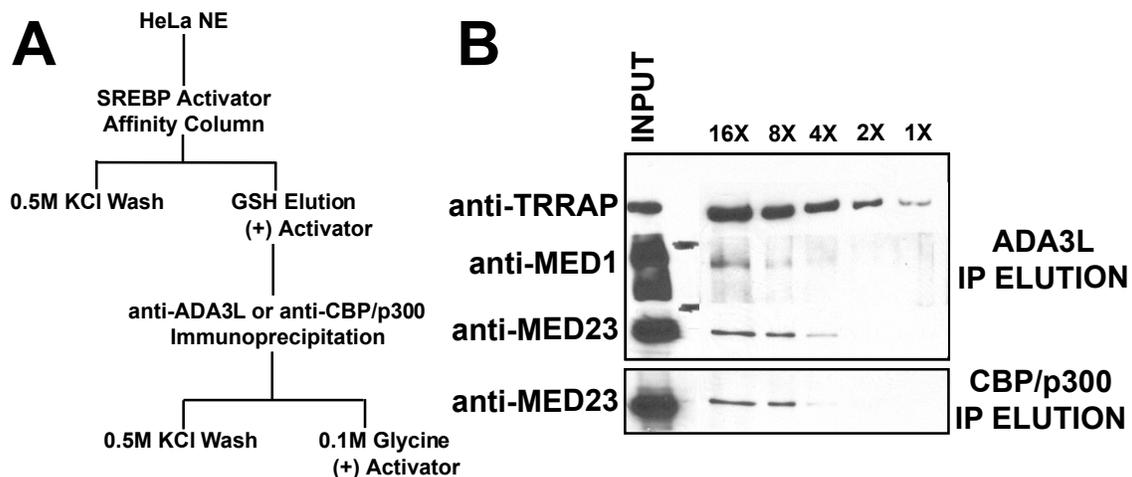


Figure 2.4. Orthogonal purification with ADA3L provides additional confirmation of SREBP-Mediator-SAGA interaction. (A) Purification scheme for the anti-ADA3L orthogonal purification from SREBP-Mediator sample. (B) Western blots probing eluted material from ADA3L orthogonal purification for the presence of Mediator subunits. Note that similar results were observed upon orthogonal purification with CBP/p300 (lower panel).

It remains possible that in some cases a tripartite interaction might occur, in which the cofactor might interact simultaneously with the activation domain and a surface exposed within activator-bound Mediator. In any case, the Mediator-cofactor interactions identified in Table 2.2 and further validated in Figure 2.3 appear to be triggered by activator binding, and, because activator binding causes major structural shifts within Mediator, this observation suggests activator-induced structural shifts regulate subsequent Mediator-cofactor interactions.

2.2.3 Distinct Cofactors Associate with SREBP-Mediator vs. VP16-Mediator.

The sweeping structural shifts induced by activator-Mediator binding likely expose motifs for protein–protein interactions; in agreement with this, additional cofactors were observed to stably associate with Mediator upon activator binding. Because SREBP-Mediator adopts a distinct conformational state relative to VP16-Mediator, it was hypothesized that distinct cofactors might associate with SREBP-Mediator vs. VP16-Mediator. The data in Table 2.2 support this hypothesis, because there are substantial differences between factors associated with SREBP-Mediator vs. VP16-Mediator. To further probe potential activator-selective Mediator-cofactor interactions, we completed a series of quantitative immunoblotting experiments with SREBP- and VP16-Mediator samples (Figure 2.5.A). These samples were purified by using the same protocol

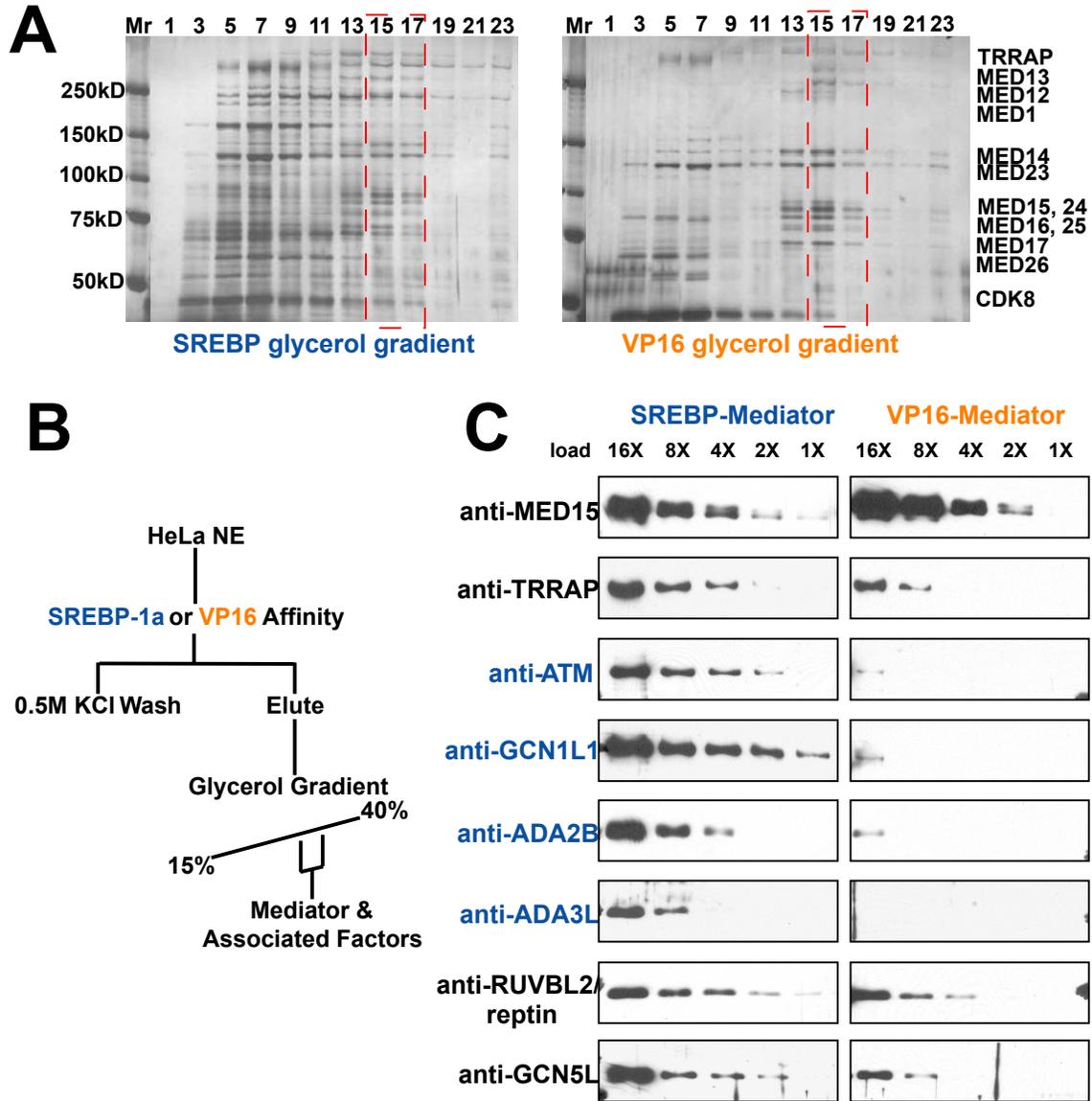


Figure 2.5. Distinct cofactors associate with SREBP-Mediator vs. VP16-Mediator. (A) Silver-stained acrylamide gels showing glycerol gradient fractions from the purification protocols shown in (B). Mediator-containing fractions are denoted by the red boxes. (C) Quantitative western blots confirm MudPIT data. Mediator-associated factors probed in immunoblotting experiments are shown at left. Factors shown in black font were observed to be present in both the SREBP-Mediator and VP16-Mediator samples, whereas factors shown in blue font were observed only in the SREBP-Mediator sample by MudPIT.

used for MudPIT analysis (Figure 2.5.B) and the presence/absence of various polypeptides was examined by western blot analysis. As shown in Figure 2.5.C, the data correlate precisely with the MudPIT results shown in Table 2.2. For

example, TRRAP, GCN5L, and reptin were identified in both SREBP and VP16-Mediator samples by MudPIT (Table 2.2). Each protein was also detected in the SREBP-Mediator and VP16-Mediator samples by Western blot analysis, as shown in Fig. 2.5.C. Similarly, the MudPIT data revealed no peptides corresponding to ATM, GCN1L1, ADA2B, or ADA3L in the VP16-Mediator sample, whereas these proteins were well represented in the SREBP-Mediator sample. In agreement with these data, quantitative western blotting confirmed a significant enrichment of these cofactors in SREBP-Mediator fractions (Fig. 2.5.C), whereas ATM, GCN1L1, ADA2B, and ADA3L were nearly undetectable in the VP16-Mediator sample. Note that equivalent amounts of Mediator were examined in each experiment, as shown by the Med15 immunoblotting experiments (Fig. 2.5.C). Combined with the MudPIT analysis summarized in Table 2.2, the data in Figure 2.5 provide strong evidence that Mediator-cofactor association can be activator-selective and that this selectivity is conferred by distinct activator-bound Mediator structural states.

2.2.4 MudPIT Analysis of CDK8-Mediator.

Mediator exists in at least two major forms in human cells: core Mediator and CDK8-Mediator. Core Mediator is devoid of the CDK8 submodule and contains the Med26 subunit, whereas CDK8-Mediator contains the CDK8 submodule (CDK8, Cyclin C, Med12, and Med13) but lacks Med26 (Taatjes, et al., 2004). Thus, Med26 is specific to core Mediator, whereas the CDK8

submodule is specific for CDK8-Mediator. Note, however, that the CDK8 submodule can also exist as a stable entity on its own (Knuesel, et al., 2009). Med1 represents a Mediator subunit that is shared between core Mediator and CDK8-Mediator; thus, Med1 IP samples will represent a mix of CDK8-Mediator and core Mediator. As expected, Med26 was not detected in the CDK8 IP sample, whereas every other consensus Mediator subunit was identified (Table 2.1). Additionally, no pol II subunits were detected in the CDK8 IP sample, whereas spectral counts for pol II subunits were abundant in each of the other Mediator samples. To verify these results, Med1 and CDK8 IP samples were probed for Med26 and the pol II subunit Rpb1, again by using Med15 to normalize the two samples (Figure 2.6). Quantitative western blotting showed greater than 8-fold enrichment of Med26 and greater than 16-fold enrichment of Rpb1 in the Med1 IP sample, with no detectable Rpb1, nor Med26, in the CDK8 IP sample (Figure 2.6). These results are consistent with past reports that indicated mutually exclusive CDK8/pol II association with Mediator (Knuesel, et al., 2009; Naar, et al., 2002).

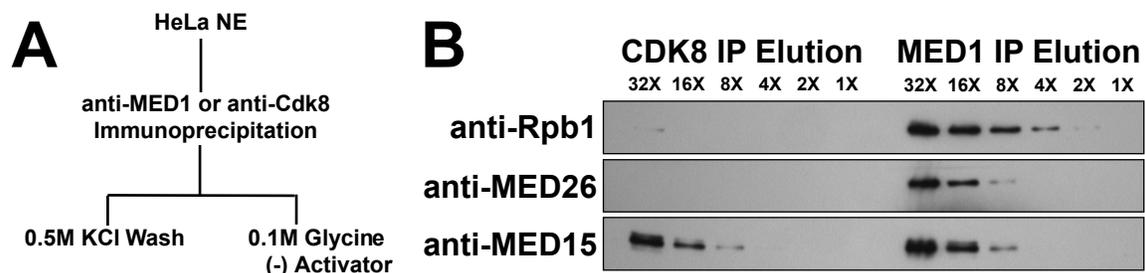


Figure 2.6. Med26 and pol II do not associate with CDK8-Mediator. (A) Western blots probing Med1 or CDK8 IP samples for pol II, Med26, or Med15. Med15 was used as a Mediator loading control.

Table 2.3 Spectral counts for cofactors observed exclusively in CDK8 IP samples

Gene	Alt. name	SREBP	VP16	CDK8	MED1	Protein MW	Description
AFF1	AF4	0	0	7	0	131578	P-TEFb associated factor
AFF4	MCEF	0	0	23	0	127459	P-TEFb associated factor
CCNT1	CycT1	0	0	7	0	80685	P-TEFb subunit
CDK9		0	0	7	0	42778	P-TEFb subunit
ACTBL1		0	0	34	0		Actin, beta-like 1
TLE3		0	0	25	0	82222	Transducin-like enhancer 3
RPLP2		0	0	15	0	11665	Ribosomal protein
FOXP4		0	0	10	0	51011	Transcription factor
YWHAE	14-3-3 epsilon	0	0	7	0		Phosphoserine binding
DDX3X		0	0	5	0	73243	RNA helicase
FLOT1		0	0	5	0	47355	Membrane protein
RPS25		0	0	5	0	13742	Ribosomal protein
SMARCB1	INI1/SNF5	0	0	4	0	44141	SWI/SNF subunit
CD44		0	0	4	0	39416	CD44 isoform 12

Interestingly, MudPIT analysis revealed a set of factors that associate specifically with CDK8-Mediator (Table 2.3); these factors were well-represented in the CDK8 IP sample but not detected in the Med1 IP or activator-bound Mediator samples.

2.3 Discussion

The results described here have broad implications for how gene expression is regulated in human cells. First, it is evident that activators serve roles in controlling gene expression that extend beyond simple recruitment of factors (e.g., Mediator) to the promoter. In fact, activators appear to regulate Mediator function by altering its conformational state, thereby controlling subsequent interactions with other regulatory cofactors. We anticipate that in some cases, cofactor association with Mediator will, in turn, modulate the activity of these factors, upon the basis of previous work that demonstrated the GCN5L

acetyltransferase and the CDK8 kinase alter their activity toward chromatin substrates upon association with Mediator (Knuesel, et al., 2009; Meyer, et al., 2008). Second, our results define a mechanism by which the general transcription machinery, in particular the Mediator complex, might actually adopt different functions in distinct promoter contexts. Different activators help regulate different sets of genes and, intriguingly, different activators induce distinct structural shifts within the human Mediator complex (Taatjes, et al., 2002, 2004;

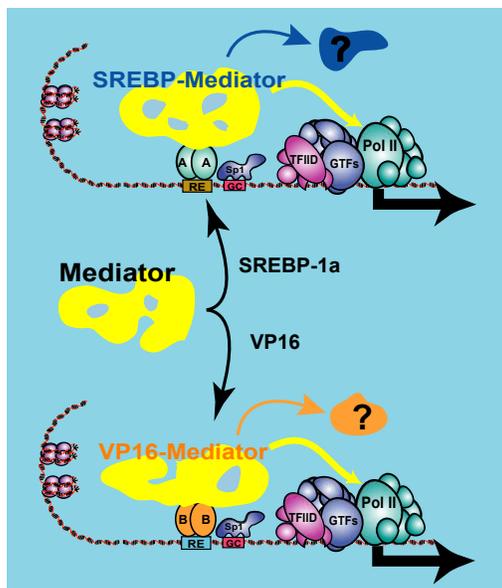


Figure 2.7. A model that summarizes the results and implications of this study. Mediator-cofactor interactions do not occur in the activator-free state; rather, activator binding signals a shift in Mediator structure only when engaged at the promoter, providing a straightforward means by which Mediator activity can be controlled in a spatial and temporal fashion. Note that activator binding not only enables Mediator-cofactor interactions, but different activators, which induce different structural shifts within Mediator, trigger interaction with distinct sets of coregulatory factors, providing a mechanism by which Mediator can adopt activator-specific functionality.

Meyer, et al., 2010). Significantly, we observe these structural differences translate into differences in factors associated with SREBP-Mediator or VP16-Mediator, providing a means for Mediator to adopt gene-specific functions (Figure 2.7). Thus, these data suggest a role for Mediator in orchestrating the recruitment and/or exchange of coregulatory factors at gene promoters and enhancers. Third, the MudPIT data forecast expanded roles for Mediator in the control of human gene expression (see below).

Importantly, numerous Mediator-cofactor interactions identified in the

MudPIT experiments have been functionally validated with in vitro or cell-based experiments. For example, TRRAP and GCN5L associate with activator-bound but not activator-free Mediator complexes. Our laboratory previously determined that TRRAP and GCN5L stably assemble within CDK8-Mediator and that this “T/G-Mediator” complex phosphorylates S10 and acetylates K14 within histone H3, a mark associated with some active genes. In fact, CDK8 and GCN5L function synergistically in this context, providing a mechanistic basis for GCN5L association with CDK8-Mediator (Meyer, et al., 2008). The MudPIT data also identify CBP/p300 and components within the SAGA complex as factors that associate with activator-bound Mediator. A functional cooperativity in gene activation between Mediator-SAGA or Mediator-CBP/p300 has been described in past studies (Liu, et al., 2008; Black, et al., 2006). As another example, the MudPIT data indicate an enrichment in PTEFb (Cyclin-T and CDK9) and AFF4 in the CDK8-Mediator exchange (pontin, reptin, p400), and chromatin architecture (Smc1A, Smc3, IQGAP1, nesprin-2, ZW10, HMMR) are strongly represented in the MudPIT data. Numerous enzymatic activities also associate with Mediator (e.g., GCN5L, CBP/p300, DNA-PK, ATM, RECQL5). These additional Mediator-associated factors are observed almost exclusively with activator-bound Mediator (VP16 or SREBP), whereas activator-free Mediator samples contained only consensus Mediator subunits without associated coregulatory proteins (although see CDK8 IP data, Table 2.3). That additional factors were observed to associate with Mediator upon activator binding provides strong evidence that activator-induced structural shifts trigger Mediator-cofactor interactions. The

biological rationale for this is clear: A requirement for the activator in directing subsequent Mediator-cofactor interactions ensures these interactions are controlled in a spatial and temporal fashion. Indeed, such a strategy prevents Mediator-cofactor interactions from occurring when such interactions might be unproductive, such as when Mediator is not stably bound to the promoter.

Most factors enriched in activator-bound Mediator samples are in fact specific to SREBP-Mediator; beyond the consensus Mediator subunits, few additional factors are observed with VP16-Mediator. This may reflect the fact that VP16 is a viral activator; as such, VP16-Mediator likely evades common regulatory strategies by avoiding association with a host of coregulatory proteins. The factors identified that associate with SREBP-Mediator are almost certainly substoichiometric relative to Mediator itself; that is, these factors likely associate with a fraction of SREBP-bound Mediator complexes and should not be considered consensus Mediator subunits. Human SREBP-1a is important for expression of several dozen genes (Horton, et al., 2003). Factors that associate with SREBP-Mediator (but not VP16-Mediator) might play specialized roles in regulating a subset of SREBP target genes, whereas other Mediator-associated factors (e.g., CBP/p300) clearly serve more general roles at a wide variety of genes. It will be important in future work to further define potential gene-specific functions for cofactors that associate with SREBP-Mediator. These experiments will require comparative analyses at many SREBP target genes that correlate factor recruitment and gene expression with activation and nuclear localization of SREBP-1a. It will also be informative to compare and contrast MudPIT data from

Mediator bound to an array of different transcription factors, such as nuclear receptors or p53, that induce structural shifts within Mediator distinct from SREBP or VP16 (Taatjes, et al., 2004; Meyer, et al., 2010). Potentially, a subset of Mediator-associated factors will be unique to p53-Mediator (for example) and the identity of these factors might provide insight regarding activator-specific regulatory mechanisms.

As anticipated, MudPIT analysis of CDK8-Mediator yielded no spectral counts for Med26, pol II, or the pol II-associated factor Gdown1; however, the MudPIT data also revealed intriguing differences among pol II and Mediator subunits. Several pol II subunits (Rpb2, Rpb3, Rpb5, Rpb8) were vastly over-represented in activator-bound versus activator-free Mediator samples. A distinct role for Med7 in activated transcription was also implicated by the MudPIT data, because Med7 was enriched 37-fold in activator-bound Mediator samples. These results might reflect changes in Mediator-pol II interactions that occur upon activator-Mediator binding; these subunits may also play key roles in activator-dependent transcription. Additional structural and mechanistic studies will be required to confirm this, yet it is interesting to note that activator-induced structural shifts within Mediator have been linked to activation of promoter-bound, stalled pol II complexes (Meyer, et al., 2010).

2.4 Methods

2.4.1 Mediator purification

Activator-bound Mediator was purified from HeLa nuclear extract using GST-SREBP-1a (residues 1–50) or GST-VP16 (residues 411–490) immobilized to Glutathione-Sepharose beads (GE Life Sciences). After binding, the resin was washed five times with 10 column volumes (CV) 0.5 M KCl HEGN (20 mM Hepes, pH 7.6; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40 alternative) and one time with 10 CV 0.15 M KCl HEGN (0.02% NP-40 alternative). Bound proteins were eluted with 30 mM GSH in elution buffer (80 mM Tris, 0.1 mM EDTA, 10% Glycerol, 0.02% NP-40, 100 mM KCl) and applied to a 15% to 40% linear glycerol gradient (in 0.15 M KCl HEG) and centrifuged for 6 h at 55,000 rpm. Mediator-containing fractions (>1.0 MDa) were combined for analysis.

Activator-free Mediator was purified using anti-CDK8 or anti-MED1 antibodies immobilized to Protein A/G Sepharose (Amersham Biosciences). Immobilized antibodies were incubated with HeLa nuclear extract at 4 °C. The antibody resin was then washed three times with 20 CV 0.5 M KCl HEGN and twice with 20 CV 0.15M KCl HEGN. Bound proteins were eluted with 0.1M Glycine, pH 2.7.

Orthogonal purification of activator-bound Mediator samples was completed by initial purification with GST-SREBP-1a, as described above. Eluted material was then applied to an anti-CDK8 or anti-Med1 resin (similar

results were obtained with each), washed four times with 20 CV 0.5 M KCl HEGN, once with 20 CV 0.15 M KCl HEGN, and eluted with 0.1 M glycine, pH 2.7. Wash buffers were supplemented with free GST-SREBP-1a (residues 1–50) to ensure activator-bound Mediator complexes remained in the SREBP-bound conformational state.

2.4.2 MudPIT analysis of Mediator

Mediator subunits and associated proteins were identified using a modified MudPIT procedure (Washburn, et al., 2001) used by the Conaway laboratory to identify the consensus Mediator subunits (Sato, et al., 2004). Mediator-containing fractions from glycerol gradient-purified samples (typically fractions 15 through 17) were combined from four gradients for SREBP and eight gradients for VP16 and TCA precipitated [20% (w/v)]. Proteins were resuspended with 1% RapiGest SF (Waters) in 0.1 M Tris pH 8.5 and processed with a modified method (Florens & Washburn, 2006). Proteins were reduced 30 minutes with 5mMTCEP (Pierce) and alkylated 30 minutes in the dark with 10 mM Iodoacetamide (Sigma). Suspended proteins were then diluted 4-fold to 0.25% RapiGest with 0.1 M Tris pH 8.5 and digested overnight at 37 °C with modified trypsin (Promega) and 2 mM CaCl₂.

Digested peptides were acidified with formic acid to 5% (volume/volume) and pressure loaded and washed on a 250 µm fused silica capillary column packed with 5 cm strong cation exchange resin (Whatman 5 µm Partisphere

SCX) followed by 2 cm rpC18 (Phenomenex Jupiter 5 μm rpC18). Loaded capillary columns were then connected via a 1 μm Nanofilter (Upchurch) to a 75 μm \times 150 mm rpC18 (LCPackings PepMap 3 μm 100A rpC18). The assembled column was attached to an Agilent 1100 HPLC and run at 0.3 $\mu\text{L}/\text{min}$. Peptides were fractionated off of the SCX with six steps of ammonium acetate (75, 150, 200, 250, 350, and 500 mM). Each step was resolved by a 40-minute gradient elution from 3% to 60% acetonitrile with 0.1% formic acid.

Eluted peptides were detected using a Agilent MSD XCT—nano ESI (Picotip, New Objective) ion trap mass spectrometer. Nano-electrospray was achieved with 1.7 kV and 300–1800 m/z was scanned. MS/MS spectra were acquired in a data-dependent fashion from the three most intense parent ions (1.3 V). Data collection was controlled by the Agilent ChemStation software (version A.09.03). A mascot generic file was created using the Agilent Data Analysis software (version 2.2) and Mascot (Matrix Sciences version 2.2) was used to search the human ipi_v3.27 database (Perkins, et al., 1999). Peptides less than seven residues were excluded and a 1% false discovery rate was determined by a separate search of the reversed database. Spectral counts were generated by processing identification data with Isoform Resolver as described (Old, et al., 2005; Meyer-Arendt, et al., 2011; Resing, et al., 2004).

2.4.3 Antibodies

MED1, CDK8, GCN5L, CBP/p300, ADA3L, LRP130, HADHB (Santa Cruz Biotechnology); MED23, TRRAP, GCN1L1, ATM, p400, DNAPK (Bethyl); HADHA, ADA2B, Reptin/TIP49B/RUVB2, SMC3, SLIRP, SnoN, SKIV2L2 (Abcam); Rpb1, MED15, MED26 (lab stocks). Antibodies against IQGAP1 (Cell Signaling), SPT3 (Santa Cruz and Abcam), and MMS19 (Abcam) were also tested but did not yield reliable results in control experiments.

2.4.4 Mediator “immunodepletion” experiments

We completed a series of Mediator immunodepletion trials in an effort to further test Mediator-cofactor interactions. Depletion of Mediator from extracts should similarly deplete the Mediator-associated cofactors from an SREBP affinity purification. These experiments gave the expected results in that factors identified to associate with SREBP-Mediator via MudPIT and orthogonal purification (e.g., ATM, GCN5L, ADA2B) were depleted along with Mediator in SREBP pull-downs from depleted extracts when compared with standard extracts. Over the years, however, we have observed that immunodepletion cannot effectively remove Mediator from a cell extract, thus limiting the usefulness of immunodepletion assays. Why antibodies cannot completely remove Mediator from extracts might derive from the fact that Mediator exists in a variety of structural states, any of which might mask the epitope for a given

antibody. Mediator also interfaces with multiple coregulatory complexes (e.g., the CDK8 submodule) and these Mediator-cofactor associations could also block antibody binding to a subset of Mediator complexes within an extract. Another consideration is that few Mediator antibodies are actually effective in an IP; Med1 and CDK8 work very well, as revealed by silver staining the bound proteins. Upon completing a tandem immunodepletion using first an anti-Med1 affinity resin followed by depletion with an anti-CDK8 antibody resin, it is evident that a significant amount of Mediator remains in the extract. For example, an SREBP-1a affinity resin binds Mediator to an extent similar to an untreated extract, based upon silver stain analysis and quantitative Western blotting. Rerunning the

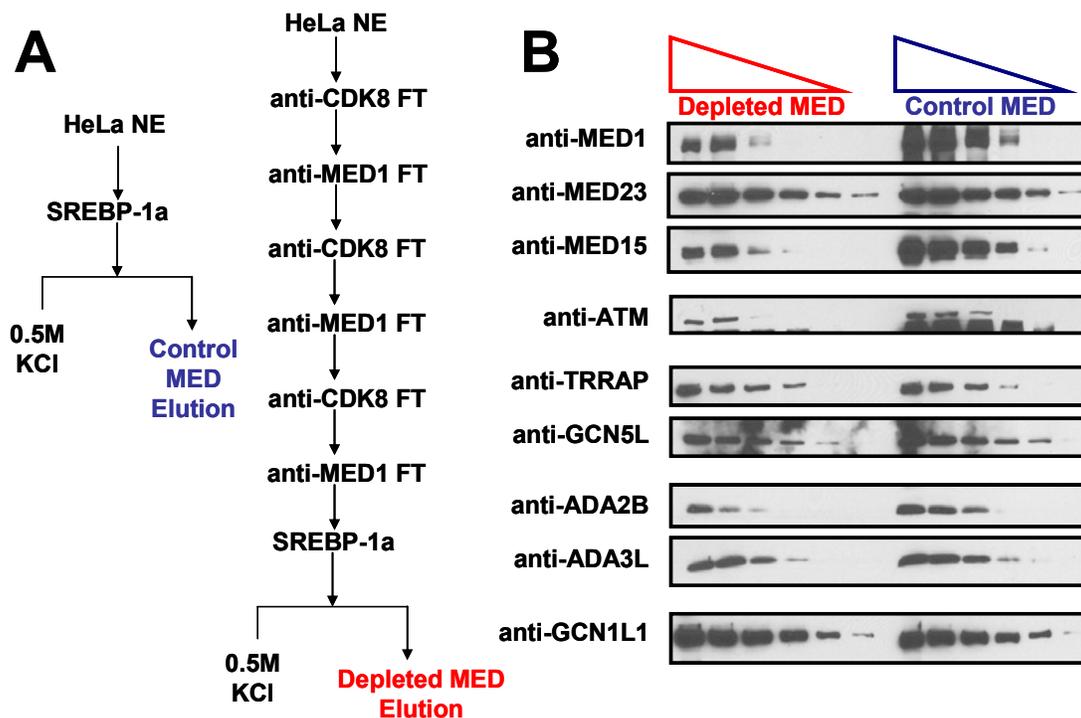


Figure 2.8. Immunodepletion of Mediator experiment does not completely deplete Mediator complexes after a total of six anti-MED1 and anti-CDK8 IPs. (A) Immunodepletion scheme. (B) Western blots of Mediator and cofactors to show degree of immunodepletion.

experiment with 6 successive immunodepletion steps (3X anti-Med1, 3X anti - CDK8) worked to a degree—whereas Mediator was certainly not absent in the final antibody IP flow-through fraction (Figure 2.8), it was at least notably depleted. Comparison of the SREBP affinity purifications from the standard vs. immunodepleted extract (with silver staining and quantitative Western blot analysis) provided the expected results: In fractions depleted of Mediator, we observed a corresponding depletion in Mediator-associated factors.

Chapter 3

An Improved Proteomics Platform for the Analysis of Very Low Abundance Protein Samples

3.1 Introduction

Protein-protein interactions inside the nucleus of a cell are fundamental for the overall function of a cell, not to mention the specific inner workings of regulating the expression of select genes. These protein-protein interactions assemble metastable complexes which are essential for the dynamic associations required for biological functions (Alberts, 1998; Köcher & Superti-Furga, 2007). These dynamic associations must be tightly regulated for efficient gene expression. One way to regulate gene expression is to control critical protein-protein interactions, and in addition, to maintain low copy numbers of critical regulatory cofactors, such as activators, but also entire protein complexes. Some of these low copy cofactors can be very difficult to purify and prepare for analyses with sufficient recovery to achieve interpretable results. The mass spectrometry-based MudPIT protocol worked well to characterize the protein compositions of the purified Mediator complexes in Chapter 2. The samples analyzed were the most concentrated samples that could be purified at the time. Significant limitations of the MudPIT protocol were identified throughout the study. Sensitivity of the method was primarily limited by the recovery from the 20% (w/v)

TCA precipitation which has a lower limit for protein concentration where all or most of the protein is precipitated. When sample protein concentrations are below that limit, sample recovery is dramatically affected in terms of precision and total protein recovery. TCA and acetone washes may also be insufficient to remove residual NP-40 required for the Mediator purifications, which may foul the electrospray (ESI) emitter tip. The first dimension in the MudPIT method is strong cation exchange (SCX) which requires salt to elute. Any salt, even the volatile salts ammonium formate and ammonium acetate dramatically affect the current at the nanospray ESI source and shortens the lifespan of emitter tips. Salt residue also collects on the LC pumps, the interface and inside of the ion trap mass spectrometer requiring more frequent disassembly and cleaning. These are practical points identified running samples to generate the data provided in Chapter 2. The eluent passing through the ESI emitter tip affects the lifetime of that tip which is required for the production of gas phase peptide ions; no ions, no peptides to sequence, no proteins to identify, no data. Therefore, the quality of the sample loaded can have a big impact on the quality of the data collected.

The sample preparation and the LC/MS analysis protocol used in Chapter 2 will be discussed and divided into three parts: (1) precipitation of the purified protein samples, (2) proteolysis of protein samples into peptides suitable for sequencing in an ion trap mass spectrometer, and (3) the LC/MS/MS analysis that separates all the peptides for isolation and sequencing by the ion trap mass spectrometer. This chapter will detail the evaluation of the mass spectrometry-

based MudPIT protocol used in chapter 2, and compare and contrast the protocol developed to replace it (summarized in Table 3.1) allowing for the analyses of protein samples of much lower concentration with considerably less contamination from the protein purification protocol.

Table 3.1 Proteomics Methodologies at a Glance

	MudPIT mass spectrometry	New Proteomics Platform
Protein Precipitation	20% (w/v) TCA, 4°C overnight	Insulin, DOC, 20% (w/v) TCA
Protein Proteolysis	in solution	30k Filter-Aided Sample Prep
Peptides Loaded	Pressure Baume	LC Autosampler
Chromatography	2D-LC(SCX/rpC18)/MS/MS	2D-LC(high/low pH rp)/MS/MS

3.1.1 Removal of Sample Matrix via Protein Precipitation

The preparation of low concentration purified protein samples requires the removal of sample matrix components necessary for the purification of such samples, such as detergents (eg NP-40), glycerol (10~35% (vol./vol.)) and salts; none are compatible with standard liquid chromatography or ESI mass spectrometry. Protein precipitation is an excellent method for concentrating proteins and removing or reducing the sample matrix. There are a variety of common methods for protein precipitation, including the addition of 30~50% (weight/ volume) Ammonium Sulfate to salt out protein that can be recovered/

resuspended with native activities. This method is not generally appropriate for very small samples or for mass spectrometry analyses due to the high salt concentration added, which would need to be removed. Furthermore, maintaining native activity is not required for mass spectrometry analyses, therefore, denaturing protocols such as the addition of 20% (weight/volume) Trichloroacetic acid (TCA) at 4°C or the addition of 80% (v/v) -20°C Acetone (and variants of these two methods) are very useful for removing, or at least reducing salt, glycerol and detergent contamination, as well as concentrating a protein sample prior to proteolysis. Both of these denaturing precipitation protocols have been used successfully for the preparation of sample for mass spectrometry.

Another variant method for protein precipitation utilizing 20% (w/v) TCA includes addition of the detergent deoxycholate (DOC) and the carrier protein bovine pancreatic insulin (MW 5733.66). DOC assists by denaturing the proteins exposing the hydrophobic cores encouraging nucleation and precipitation. Insulin is added as a carrier protein to induce nucleation of precipitating protein allowing for very high precipitation efficiency with the lowest concentrations of protein samples. This method is suitable for in-gel digestion of proteins for mass spectrometry since an acrylamide gel resolves the added insulin. However, this method is not suitable for in solution digestion due to the large amount of insulin added to the sample which become an excess contaminant in the mass spectrometer.

3.1.2 Proteolysis via Filter-Aided Sample Preparation (FASP) for MS

There are two basic strategies for preparation of protein samples for mass spectrometry: (1) in-gel digestion and (2) in solution digestion (as used in the MudPIT protocol in Chapter 2). In-gel digestion is a very robust and clean sample preparation method, but is limited by low peptide recovery, especially for very low concentration samples, and is biased by molecular weight (from the percent acrylamide gel used). Peptide recovery is generally much better with in solution digestion; however, the sample must be free from contamination incompatible with LC/MS (glycerol, detergents, salt, etc.) which can limit the ability to purify certain proteins for mass spectrometry analyses.

A new method described by the Mann group at the Max Plank Institute of Biochemistry, called Filter-Aided Sample Preparation (FASP) (Wiśniewski, et al., 2009) (Figure 3.1), has added another dimension, as a hybrid of an in-gel (using SDS) and in solution protocols, to preparing protein samples for mass spectrometry. The real key to the FASP protocol is that it allows for proteins to be solubilized in SDS and urea, a strong detergent and denaturing chaotrope, respectively. This is possible using an ultrafiltration device designed for concentrating protein samples, such as a Millipore Microcon or a Sartorius Stedim Biotech Vivacon ultra-filtration device, as a proteolysis “reactor” (Wiśniewski, et al., 2011a). When the experiment shown in Figure 3.1.E. was attempted with as 30K MWCO filter using a glycerol gradient purified Mediator, no Mediator was detectable by silver staining prior to the addition of trypsin

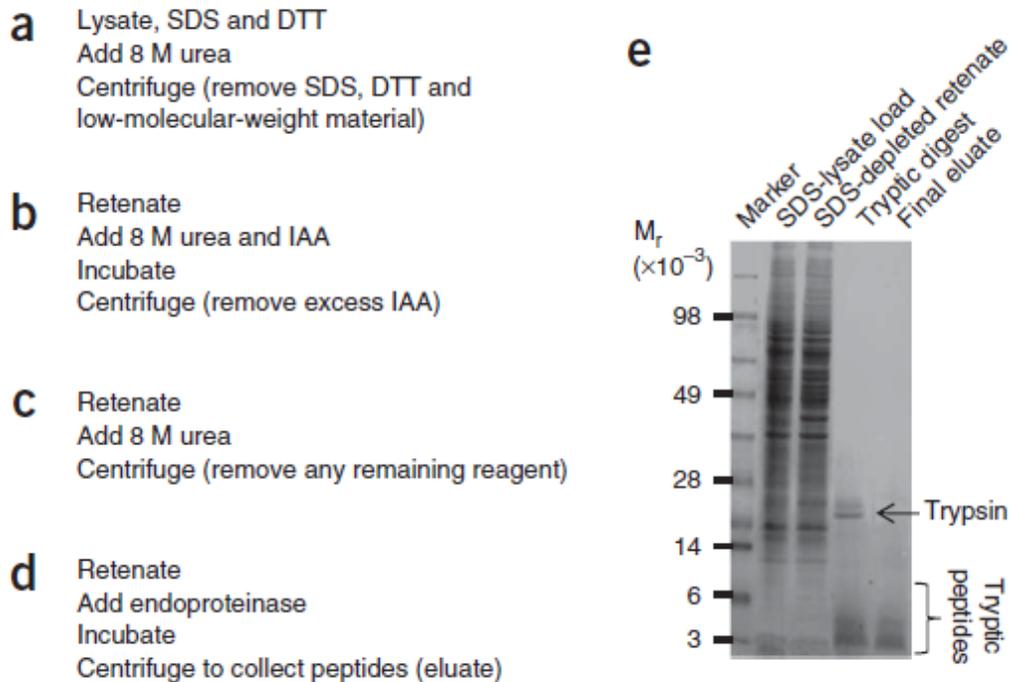


Figure 3.1 Filter-aided sample preparation (FASP) for mass spectrometry (Wiśniewski, et al., 2009). The protocol is (A) solubilize proteins in SDS and reduce with DTT & dilute and wash with 8M Urea. (B) Alkylate and spin out Iodoacetamide. (C) Wash away detergents, etc. with 8M urea. (D) Wash away Urea with ammonium bicarbonate and add protease. Digest protein on the filter and spin out peptides. (E) Acrylamide gel of steps A through D.

(Figure 3.5), suggesting the membrane or plastic of the spin filter bound the protein added (see below).

The purification protocols for Mediator include the detergent NP-40, which is clearly not compatible with LC/MS analysis. Mediator purifications were attempted without NP-40 and with Triton X-100 as an alternative to NP-40, but the yields were very poor compared to the inclusion of NP-40. Therefore, removal or replacement of NP-40 in the purification protocol was not feasible for robust and reproducible Mediator purifications.

This is a major strength of the FASP protocol in that it very effectively removes detergents, such as SDS, NP-40 and Triton X-100 (Figure 3.2) and likely

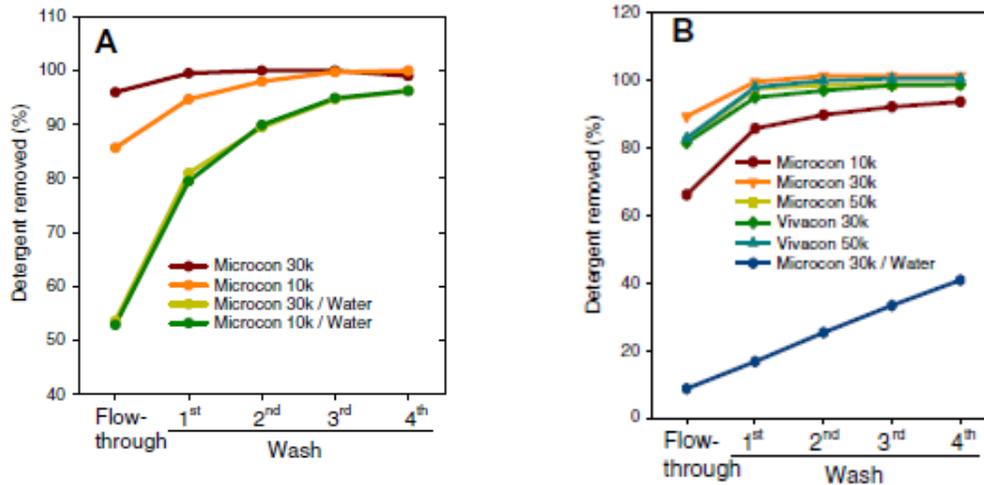


Figure 3.2 FASP removes detergents (Wiśniewski, et al., 2011a).
 (A) Microcon 10/30K filters remove 0.2% (w/v) SDS with four washes of 8M Urea or water. (B) Removal of 0.2% (v/v) Triton X-100, X-114 and NP-40 with four washes of 8M Urea or water determined by absorbance at 275nm.

many other potential contaminants, such as large partially undigested polypeptides that would induce spectral noise. Both the Microcon and Vivcon spin filters were tested with various molecular weight cutoffs for the removal of Triton X-100, Triton X-114 and NP-40 (Figure 3.2.B). The efficient removal of detergents by the filter washed with 8M urea allows for no compromises or modifications to protein purification protocols.

Some controversy (Liebler and Ham, 2009) about the FASP method suggested that it was not a “Universal sample preparation method” as the FASP authors offered it (Wiśniewski, et al., 2009). Liebler and Ham suggested the method “is useful in some applications” but has “considerable limitations” for samples <50ug due to poor peptide recoveries and greater variability. Liebler and Ham offered Table 3.2 which compared the (1) FASP-type spin filter method, (2) an in-gel digest from an acrylamide gel run for a short time and (3)

Table 3.2 Comparison of spin filter, short SDS-PAGE and TFE methods (Liebler and Ham, 2009)

Method	Protein load	Peptide identifications	Protein identifications
Spin filter		5,369	642
Short SDS-PAGE	50 μ g	4,176	593
TFE		4,663	593
Spin filter		86	46
Short SDS-PAGE	150 ng	298	106
TFE		626	150

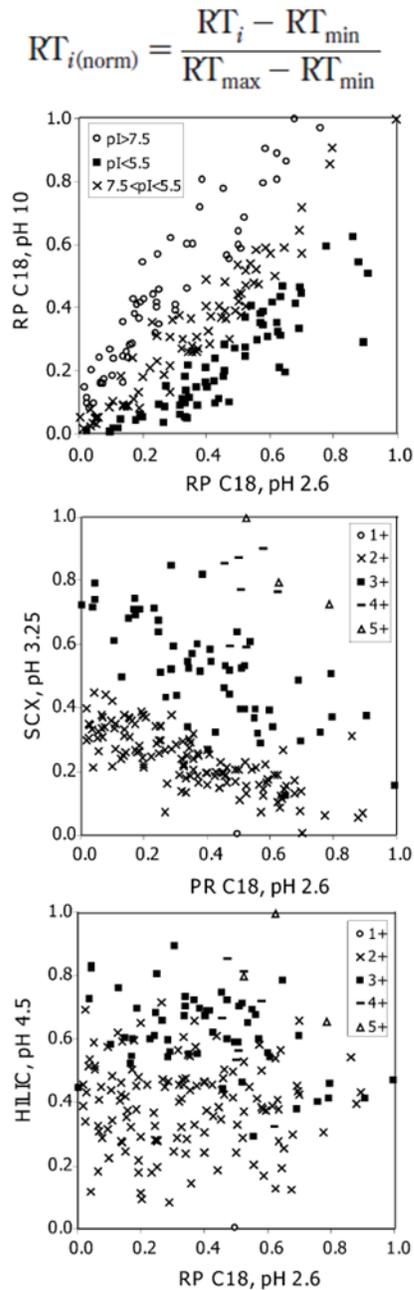
solubilization with trifluoroethanol (TFE), a method developed specifically for micro- and nanoscale proteomics samples (Wang et al., 2005). In fact, Table 3.2 shows clearly when 50ug is processed, the spin filter provides the highest number of peptide and protein identifications. However, when 150ng is processed the spin filter gives the lowest number of peptides and proteins identified. This is readily explained by Liebler and Ham as non-specific protein/peptide binding to the spin filters. This conclusion is consistent with empirical evidence using a CDK8 IP purified Mediator sample with no recovery after the input by silver staining (Figure 3.5), mentioned earlier. It seems likely that virtually any membrane or plastic surface has the capacity to non-specifically bind some protein and/or peptides. This is especially difficult to contend with when protein concentrations are very low. For this reason, siliconized, or low-retention tubes and tips (Fisherbrand) were employed throughout for all of the work presented in this thesis. This strategy was confirmed in a study (Bark and Hook, 2007) evaluating tryptic BSA peptides stored in various tubes.

3.1.3 Orthogonality of Two-Dimensional LC/MS

A practical matter for the analysis of protein samples for mass spectrometry is the depth of sequence coverage. There are many variables that contribute to the number of peptides from a given protein that can be identified. Two major variables are the protease or proteases employed and the resolving power of the chromatography system up-front of the mass spectrometer. The first major variable is the protease or multiple proteases used to digest proteins into peptides suitable for mass spectrometry analysis. Sequence coverage has been shown to improve nearly 3-fold with multiple proteases compared with trypsin only (Swaney, et al., 2010). There is the practical question of how good is good enough? Swaney, et al., identify an additional 595 proteins on top of 3313 proteins identified with trypsin only, mostly low abundance factors. Is it necessary to use a panel of proteases and how many analyses are required? The biological questions and the goals of a well-thought out experiment should lead the way.

So now we have a protein sample digested with trypsin, which cleaves the carboxyl peptide bond of the basic residues arginine and lysine, so long as there is no proline in the carboxyl sequence of the polypeptide. This is ideal for electrospray mass spectrometry which works best for sequencing with positively charged peptide ions where the amino-terminus of the peptide is positively charged and the basic side chain at the cleaved carboxyl terminus is also positively charged, giving at least a plus two charge on the peptide. Depending

on the sample, there could be hundreds to thousands to tens of thousands of peptides that could be identified by the mass spectrometer. The mass spectrometer has a sampling rate that limits the number of scans for peptides and sequencing of each peptide per given amount of time. If the up-front



chromatography system can resolve the peptides such that the capacity of the mass spectrometer to collect data is greater than the number of peptides eluting at any given time, then fewer peptides will be missed and more complete sequence coverage will be the result.

The MudPIT protocol was developed to extend the peak capacity of a 1D-LC/MS analysis with a second dimension SCX which fractionates peptides based on the overall charge of the peptide. Unfortunately, the number of charge states for peptides that can be sequenced by mass spectrometry is limited to 1+, 2+, 3+, 4+ and 5+, though the majority of peptides sequenced are 2+ and 3+, and 1+ are excluded, which does not afford wide variation in separation by the SCX dimension. Today the technology for peptide separations has advanced with Ultra-High

Figure 3.3 Normalized retention time plots for two-dimensional orthogonality in LC/MS. (Top) Calculation of normalized retention time for phosphorylase b tryptic digested peptides. pH 10 vs. pH 2.6 rpC18; SCX vs. pH 2.6 rpC18; HILIC vs. pH 2.6 rpC18.

Table 3.3 Practical Peak Capacity of 2D-LC (Gilar, et al., 2005)

	1st Dim: pH 10 rp 2nd Dim: pH 2.6 rp	1st Dim: SCX 2nd Dim: pH 2.6 rp	1st Dim: HILIC 2nd Dim: pH 2.6 rp
Theoretical 2D Peak Capacity	13291	5880	9050

Performance Liquid Chromatography (UPLC) systems which are capable of very high peak capacities in a 1D-LC format and LTQ-Orbitrap mass spectrometers (Thermo Scientific) which have greater sensitivity and high mass accuracy for greater resolving power. This technology allows an investigator to characterize much more complex samples with greater depth of coverage. But for older technology, such as slower, low resolution 3D-ion traps and traditional nano-flow HPLC systems, enhanced orthogonality in the liquid chromatography separation of peptides can produce impressive results with significantly less complex targeted samples when compared to a UPLC-Orbitrap analysis.

There are only a few chromatographic separations directly compatible with mass spectrometry. Reversed phase is the most well developed for mass spectrometry in that peptides are loaded in an aqueous buffer and eluted with an increasing gradient of the volatile non-polar organic acetonitrile with formic acid added to maintain a low pH (~2.6), ensuring all peptides are positively charged. There are more options in terms of a first dimension fractionation before reversed phase: strong cation exchange (SCX), size exclusion chromatography (SEC), hydrophilic interaction chromatography (HILIC) and pH 10 reversed phase, which deprotonates all acidic residues changing the overall charge state of many

peptides and therefore increases the hydrophilicity of those acidic peptides. An evaluation of the chromatographic modes listed above (Figure 3.4) was performed for one dimension peak capacity as well as two-dimensional peak capacity generating a mathematical characterization for chromatographic orthogonality (See Table 3.3., Gilar, et al., 2005). The orthogonality of each chromatographic phase was determined in an off-line format since the elution from a HILIC column and pH 10 reversed phase would not be compatible with reversed phase retention. Organic modifiers are also frequently added to SCX separations to prevent non-specific interactions of more hydrophobic peptides. This is not a problem for the MudPIT protocol since the acetonitrile elutions for the reversed phase separation also passed through the SCX column in-line. An alternative to off-line separations is to add another pump (if you have it) and a T-connector to dilute the organic phase of the first dimension elution sufficiently to retain on a reversed phase column.

The calculated theoretical peak capacity, or the number of peaks (peptides) that can be separated in a given amount of time, for the three 2D-LC modes are given in Table 3.3 (Gilar, et al., 2005). The theoretical peak capacities calculated for 2D-LC is lowest for SCX with HILIC and pH 10 reversed phases giving considerably higher resolving power. A practical aspect of using HILIC or pH 10 reversed phase for a first dimension fractionation is how the samples are loaded and eluted. The samples for a HILIC column is retained at high organic, sometimes as high as 90% (v/v) acetonitrile. HILIC is also sensitive to salt concentration, therefore samples have to be desalted (additional

processing) and suspended in HILIC equilibration buffer. Elution off the HILIC is performed with a polar solvent, such as water often with salt added. Loading a reversed phase column in aqueous buffer is considerably more convenient in terms of the protease digestion buffers. Elution of a reversed phase column occurs with high organic which must be removed (often evaporated) or diluted for retention on a second dimension reversed phase column.

3.2 Results

The FASP protocol and the pH 10/pH 2.6 reversed phase 2D-LC methodology were adapted at the same time, so any individual contributions from the FASP or 2D-LC/MS were not distinguished. What really made the FASP protocol work so well for very low concentration (<5ug) protein samples is the precipitation protocol including insulin. The carrier protein insulin serves two vital roles in this new proteomics platform: (1) it allows for very good precipitation recovery for virtually any purified protein sample, and (2) in large excess of the protein sample, binds the membrane and plastics of the spin filter effectively blocking it from binding the purified protein sample. The same concept was employed by (Wiśniewski, et al., 2011b), where they used carrier substances polyethylene glycol or dextrans to clinical protein samples to improve peptide yields in the low to submicrogram range.

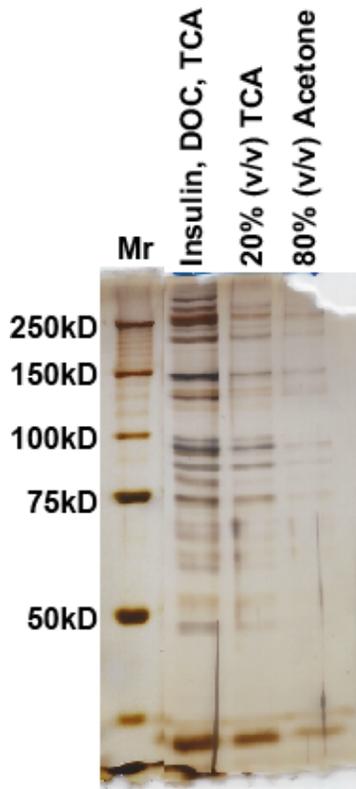
The contribution from the pH 10/pH 2.6 reversed phase 2D-LC/MS is likely to be marginal if it were directly compared to SCX/pH 2.6 reversed phase 2D-

LC/MS due to the much lower complexity of the purified and therefore targeted nature of the samples when compared with a whole cell sample. What really makes the pH 10/pH 2.6 reversed phase 2D better than a SCX/pH 2.6 reversed phase 2D is the entirely volatile acetonitrile elution with 10mM ammonium formate (compared to 500mM ammonium formate to elute the SCX) which is significantly better for the LC pumps with little to no observed effect on the ESI emitter tips. The only drawback and what may have prevented this method from being put in place sooner, is that a third LC pump is required to dilute the first dimension elution with aqueous formic acid buffer through a T-connector. The autosampler loads the sample onto the first dimension pH 10 reversed phase column and then a binary pump step elutes the peptides off which is diluted by the third pump for trapping on a reversed phase C18 trapping column which is washed and placed in-line with a nano-flow reversed phase column in virtually the same configuration as a 1D-LC/MS system.

The results presented in this chapter will show the significant difference in protein recovery with the insulin/TCA precipitation protocol and some of the first samples used to test the new proteomics platform including a CDK8 IP and a glycerol gradient purified SREBP-1a-Mediator, which will both be directly compared with data from Chapter 2 generated using the MudPIT protocol. The absolute sensitivity of the methodology will be demonstrated by Mediator and RNA Pol II samples processed in collaboration with Carry Bernecky in the Taatjes lab.

3.2.1 Comparison of Protein Precipitation Protocols.

A more dilute glycerol gradient Mediator fraction was used to compare the protein precipitation protocols adding: (1) 20%(w/v) TCA at 4°C, (2) 80%(v/v) Acetone at -20°C, and (3) Insulin, DOC and 20%(w/v) TCA at 4°C. All samples were processed identically except for the precipitation. Figure 3.4 clearly shows the enhanced recovery of Mediator polypeptides with the insulin, DOC, TCA method compared to the other two methods. When a more concentrated sample was used with the various precipitation methods, there was no advantage with the insulin, DOC, TCA method. The threshold of protein concentration required for good recoveries with 20% (w/v) TCA at 4°C or 80% (v/v) -20°C Acetone were



not determined, but these are simpler methods that don't require adding anything to the sample if you have a sufficient protein concentration for good sample recovery.

Insulin contamination in the mass spectrometry analysis is not a problem when the spin filters are used for the FASP protocol. Insulin is two polypeptides connected by disulfide bonds. During the processing, those disulfide are reduced and alkylated leaving two peptides (monoisotopic)

Figure 3.4 Protein Precipitation of glycerol gradient purified Mediator fractions. Insulin, DOC, TCA clearly gives superior recovery compared with TCA alone and Acetone.

2338.98 Da and 3398.68 Da (prior to alkylation), of which the excess is removed by the urea washes.

3.2.2 CDK8-IP from HeLa Nuclear Extract

The first sample successfully tested with the FASP and high/low pH 2D-LC methodology was a CDK8 IP from HeLa nuclear extract (NE). This IP was chosen for the fast, reasonably clean good Mediator yield the purification provided and it was a good representative sample. CDK8-bound proteins were

eluted with 0.1M Glycine, pH 2.75 and precipitated with the insulin, DOC, TCA method.

The pellet was suspended with 4% SDS and reduced, then diluted with 8M urea and alkylated. The suspended protein was added

to a Microcon spin filter and washed with 8M urea, then 2M urea, both with 0.1M Tris pH 8.5

(Figure 3.5). This is a deviation from the FASP protocol reported by Wiśniewski, et al., 2009, in

that ammonium bicarbonate is replaced with 2M urea and 0.1M Tris pH 8.5 for the trypsin digestion.

Peptides are spun out of the filter and loaded directly onto the pH 10 first

dimension reversed phase column for “on-line”

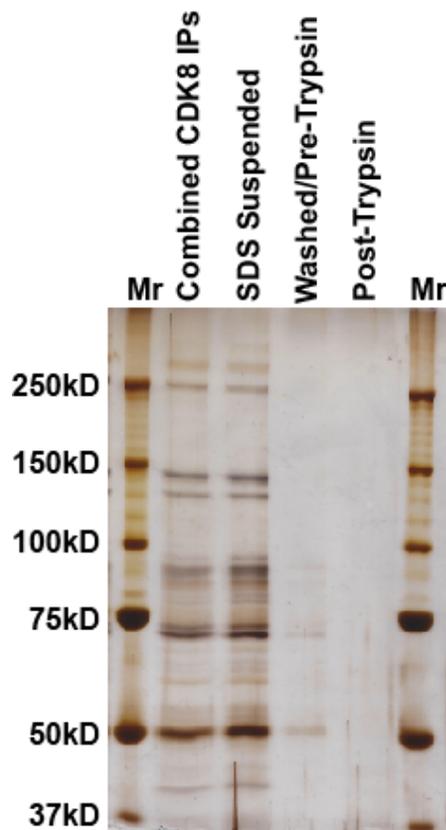


Figure 3.5 Filter-Aided Sample Preparation (FASP) samples from CDK8 IPs.

desalting and fractionation. All data was processed with a 1% false discovery rate threshold.

The sample scale was half for this test of the new proteomics platform FASP/high-low reversed-phase compared with the sample used for MudPIT due to the expected increase in sensitivity from the insulin precipitation protocol primarily. Shown in Table 3.4, as expected, nearly twice the total proteins were identified from over four times the number of peptides. The number of Mediator peptides were nearly double, though two fewer subunits were identified in this n=1 comparison, indicating replicates are absolutely required for best coverage.

Table 3.4 CDK8 IPs comparing MudPIT and FASP/high-low at 1% FDR

CDK8 IPs	IPs /1mL NE	Total		Mediator (33)		% single peptide
		peptides	proteins	peptides	subunits	
MudPIT	12X	1614	160	499	29	55.3%
FASP/high-low	6X	3687	298	785	27	54.5%

There are 138 additional proteins identified from approximately half of the sample with the FASP/high-low method compared with the MudPIT. Most are likely lower abundance factors, some which may be non-specific, but in fact many fit together with other factors in the dataset. In the MudPIT analysis, factors that make up and associate with positive transcription elongation factor b (P-TEFb) were identified only in the CDK8 IPs (Table 3.5). These factors were also identified with the FASP/high-low analysis plus additional P-TEFb associated factors were also identified (Table 3.5). Many of the additional

Table 3.5 Additional P-TEFb-Associated cofactors identified in CDK8 IP using FASP/high-low method.

Spectral Counts		Gene	Prot MW	Description
MudPIT	FASP/high-low			
7	6	CDK9	42778	CELL DIVISION PROTEIN KINASE 9
7	19	CCNT1	80685	CYCLIN-T1
	1	CCNT2	81029	CYCLIN-T2
	6	AFF1	132252	AF4/FMR2 FAMILY MEMBER 1
23	61	AFF4	127459	AF4/FMR2 FAMILY MEMBER 4
	14	ELL2	72324	RNA POLYMERASE II ELONGATION FACTOR ELL2
	5	HEXIM1	40623	PROTEIN HEXIM1
	1	SUPT5H	121000	TRANSCRIPTION ELONGATION FACTOR SPT5
	1	TCEB1	12473	TRANSCRIPTION ELONGATION FACTOR B1
	1	TCERG1	123901	TRANSCRIPTION ELONGATION REGULATOR
	3	RNMT	57725	mRNA CAP GUANINE-N7 METHYLTRANSFERASE
	1	SF3A3	58849	SPLICING FACTOR 3A SUBUNIT 3
	1	SFRS1	27745	SPLICING FACTOR. ARGININE/SERINE-RICH 1
	1	SFRS2IP	128875	SFRS2-INTERACTING PROTEIN
	3	SFRS3	19330	SPLICING FACTOR. ARGININE/SERINE-RICH 3
	1	SFRS6	39587	SPLICING FACTOR. ARGININE/SERINE-RICH 6
	8	KHSRP	73115	KH-TYPE SPLICING REGULATORY PROTEIN
	8	BAT1	50679	SPLICEOSOME RNA HELICASE BAT1
	10	PABPC1	70671	POLYADENYLATE-BINDING PROTEIN 1

cofactors have a single spectral count and by themselves are not very compelling identifications, however, when combined together as they are in Table 3.5, a trend begins to emerge. A hypothesis could be developed that a new function for CDK8-Mediator is a regulator of transcription elongation and mRNA processing. The association of P-TEFb and CDK8 is further explored in Figures 3.6 to Figure 3.8. The association of CDK8 with mRNA processing factors has been further investigated in the Taatjes Lab with additional mass spectrometry of CDK8 IPs from various column fractions, however, this data is beyond the scope of this thesis. The data presented in Table 3.5 goes most directly to show how improvements in the methodology have expanded the hypotheses that can be generated from the same sample, in this case a CDK8 IP.

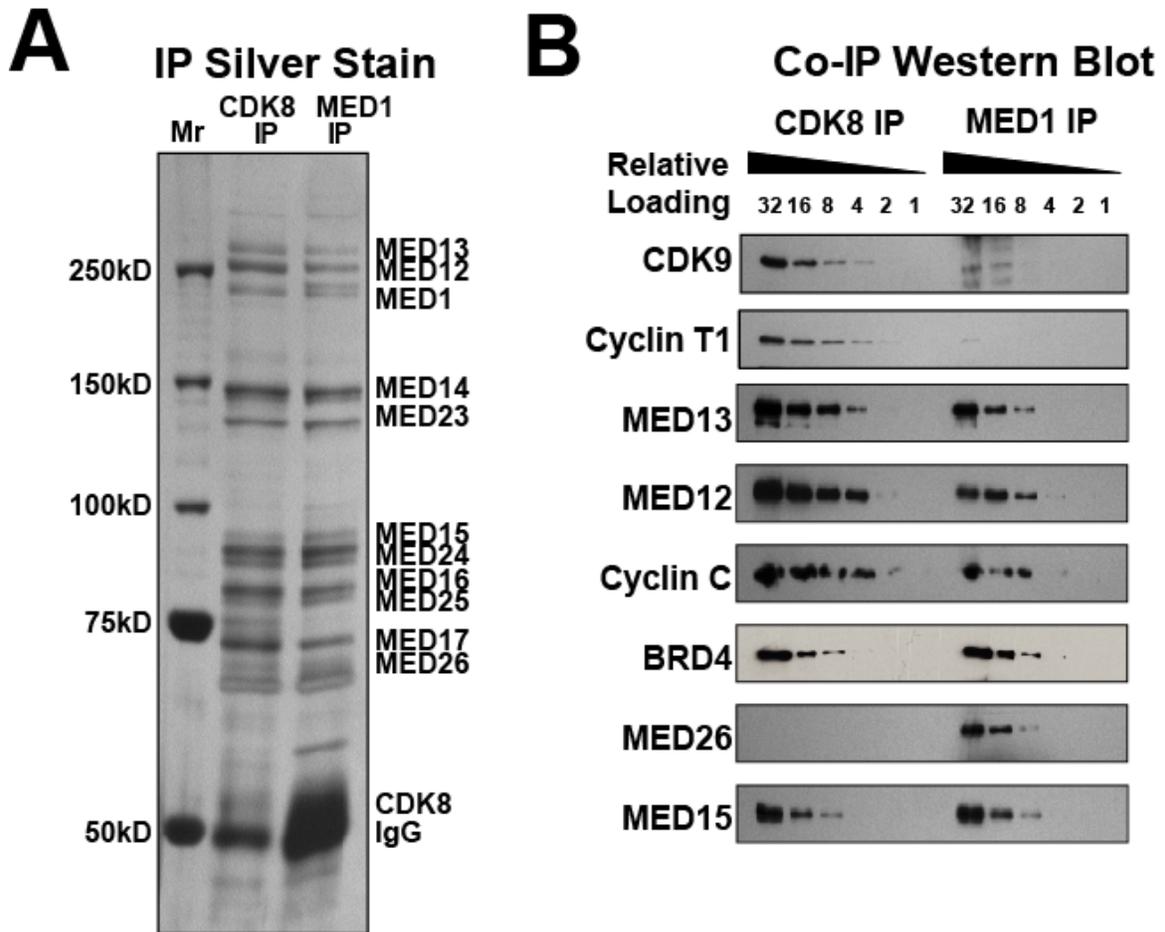


Figure 3.6 P-TEFb associates with CDK8-Mediator. (A) Silver stain of CDK8 and MED1 immunoprecipitates (IPs). Mediator and associated cofactors were immunoprecipitated from HeLa nuclear lysates using antibodies specific for CDK8 (CDK8-Mediator and CDK8 submodule) and MED1 (core Mediator and CDK8-Mediator). **(B)** Quantitative immunoblotting of IP elutions shows CDK9 and cyclin T1 enriched in the CDK8 IP. (Adapted from Donner, et al., 2010)

To further validate the P-TEFb association with CDK8, P-TEFb was probed by quantitative western blotting to compare directly the CDK8 and MED1 IP elutions for P-TEFb subunits CDK9 and Cyclin-T1 along with Mediator subunits (Figure 3.6). The western blot data confirms the mass spectrometry in that P-TEFb is detected most prominently in the CDK8 IP and not in the MED1 IP (Donner, et al., 2010). The cofactor Brd4, a bromodomain-containing protein that binds acetylated chromatin, has been reported to regulate P-TEFb (Yang, et al.,

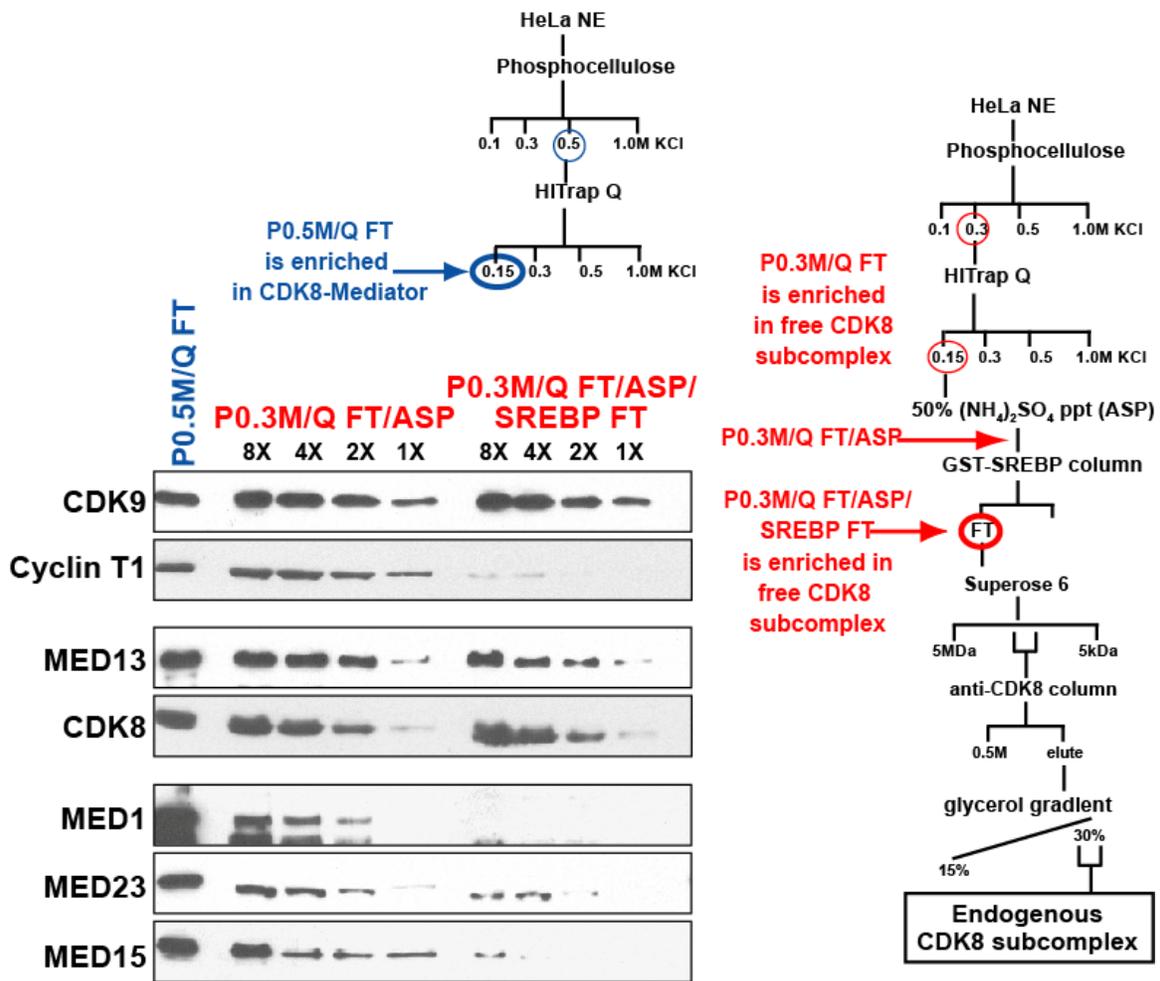


Figure 3.7 P-TEFb associates with two distinct CDK8 complexes, CDK8-Mediator and the CDK8 four protein subcomplex. (Adapted from Donner, et al., 2010)

2005) and associate with the Mediator complex (Jang, et al., 2005), was included in Figure 3.6, though it was not identified in any of the mass spectrometry datasets.

Next, to see how far P-TEFb would track with CDK8 in an extensive endogenous CDK8 purification scheme, various column fractions from a previous study in the Taatjes lab purifying endogenous CDK8 subcomplex and CDK8-Mediator (Knuesel, et al., 2009) were probed for P-TEFb (figures 3.7). Both the CDK8 subcomplex and CDK8-Mediator appears to interact with P-TEFb In this

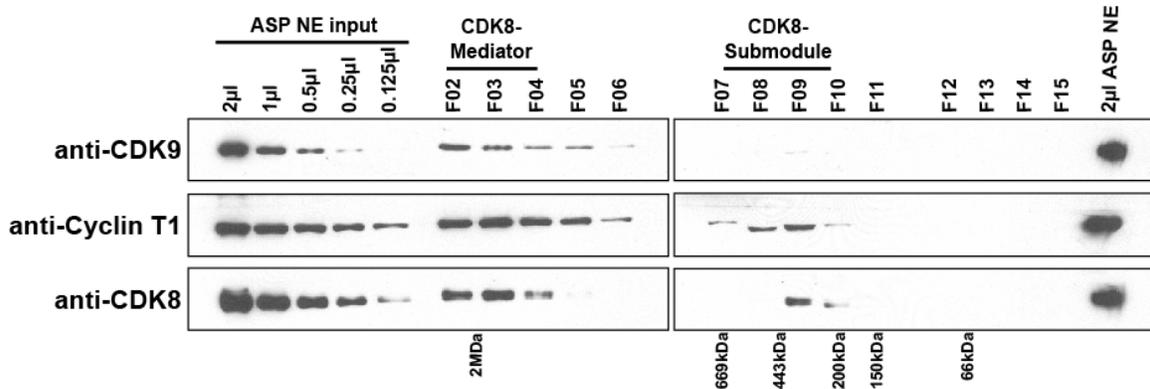


Figure 3.8 Superose 6 column fractions from ammonium sulfate precipitated HeLa nuclear extract input probing CDK8 and P-TEFb. There appears to be more CDK8-Mediator than CDK8 subcomplex, but P-TEFb associates with both complexes.

experiment. A second independent experiment was performed, again with pre-existing fractions, probing for P-TEFb in Superose 6 gel filtration fractions of HeLa nuclear extract (Figure 3.8). The original purpose of this experiment was to quantitate and compare the abundance of CDK8-Mediator and the CDK8 subcomplex. There appears to be more CDK8-Mediator than CDK8 subcomplex in this experiment, but P-TEFb also appears to again interact with both complexes.

The MudPIT data generated a new hypothesis for a functional interaction between CDK8 and P-TEFb. This was further validated (Figures 3.6 through 3.8) with biochemical purification and western blotting showing P-TEFb appears to interact with both CDK8-Mediator and the CDK8 subcomplex. The new FASP/high-low method has expanded the list of transcription elongation-associated cofactors associating with CDK8 to include mRNA capping, splicing and polyadenylation/termination cofactors.

3.2.3 SREBP-1a-Mediator from HeLa Nuclear Extract

Another direct comparison of the MudPIT method with the FASP/high-low method comes from an affinity purification of Mediator using the activation domain of the activator SREBP-1a. Data in Table 3.6 show that with approximately 10% of the sample used for the MudPIT analysis in Chapter 2, more than four times the peptides were identified for an additional 464 protein identifications. Nearly twice the Mediator peptides were identified with the FASP/high-low method for the same number of Mediator subunits identified. Other cofactors identified in the MudPIT analysis included SMC1A and SMC3, components of the Cohesin complex, which holds sister chromatids together during mitosis. NIPBL was also identified, which is known to be a Cohesin

Table 3.6 GST-SREBP-1a pulldowns from HeLa nuclear extract comparing MudPIT and FASP/high-low at 1% FDR

GST-SREBP-1a	GST pulldowns	Total		Mediator (33)		% single peptide
		peptides	proteins	peptides	subunits	
MudPIT	10X	1685	128	1065	32	55.3%
FASP/high-low	1X	8046	592	2025	32	50.5%

Table 3.7 GST-SREBP-1a pulldowns identify Cohesin

Spectral Counts		Gene	Prot MW	Description
MudPIT	FASP/high-low			
6	62	SMC1A	143233	STRUCTURAL MAINTENANCE OF CHROMOSOMES 1A
6	24	SMC3	141542	STRUCTURAL MAINTENANCE OF CHROMOSOMES 3
4	44	NIPBL	304344	NIPPED-B-LIKE PROTEIN
	5	RAD50	138432	DNA REPAIR PROTEIN
	2	STAG2	145751	STROMAL ANTIGEN 2

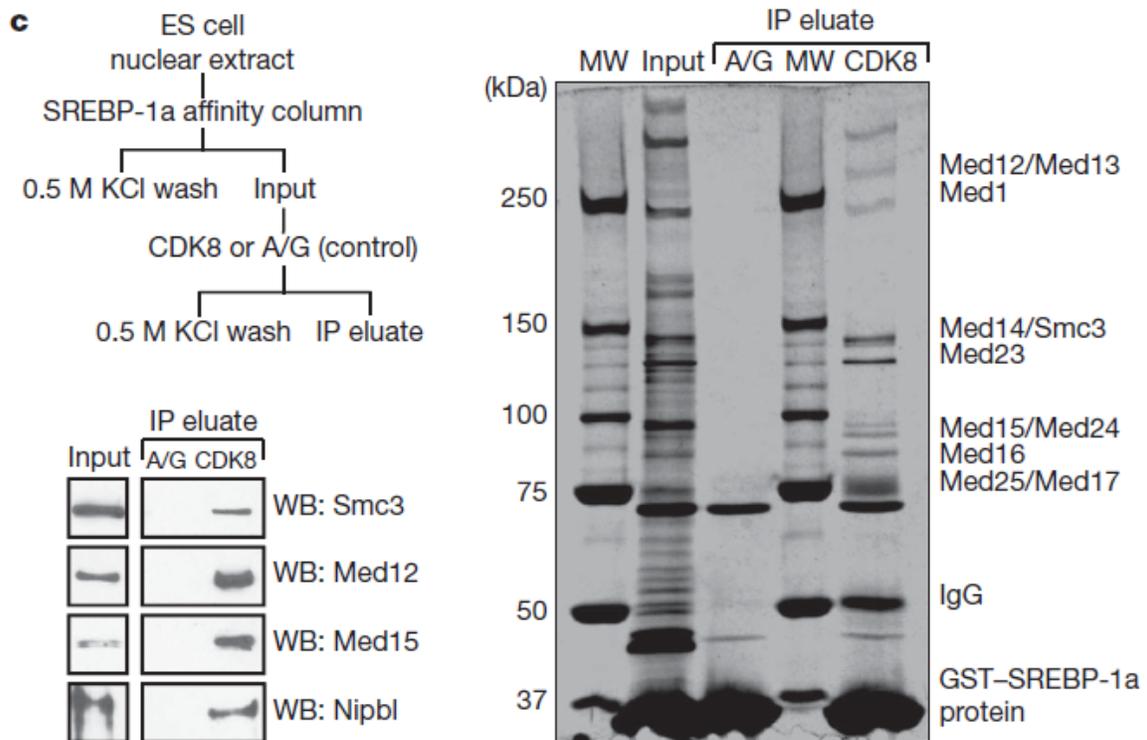


Figure 3.9 Cohesin (Smc3) and Nipbl co-purify with mediator. The input fractions and immunoprecipitated eluate (IP eluate) were examined by western blot and silver staining. (Kagey, et al., 2010)

loading factor. The FASP/high-low method identified 10-times the peptides for SMC1A and NIPBL and 4-times the peptides for SMC3 (Table 3.7). Two additional Cohesin-associated cofactors were also identified, RAD50 and STAG2.

To further validate the association of Cohesin with SREBP-1a-Mediator, the orthogonal purification used in Chapter 2 was employed with HeLa nuclear extract and mouse embryonic stem (mES) cells (Figure 3.9). Once again the MudPIT data predicted a functional protein-protein interaction, in this case between the Mediator complex and the Cohesin complex. This data, along with other data (Kagey, et al., 2010) provide evidence that Mediator and Cohesin protein complexes cooperate in the formation of enhancer-promoter DNA loops.

3.2.4 Sub-Microgram Scale VP16-Mediator and RNA Pol II.

The entire RNA Pol II complex (12 subunits) and highly purified VP16-Mediator were combined with recombinant TFIIF (2 subunits) to generate a highly purified VP16-activator- Mediator- Pol II-TFIIF complex to obtain a structure using electron microscopy (Bernecky, et al., 2011). To thoroughly characterize the compositions of the purified RNA Pol II and the purified VP16-Mediator, the FASP/high-low method was used. Since both the Pol II and Mediator samples are highly purified, there was not much sample to spare, therefore, approximately 1 ug of VP16-Mediator and approximately 2 ug of RNA Pol II were digested into peptides with the FASP protocol. The resulting peptides were split for technical replicates and stored at -80°C until analysis. Each Mediator analysis was approximately 500ng and each Pol II analysis was approximately 1 ug. This sample scale was sufficient for identification of all twelve subunits of RNA Pol II and all expected Mediator subunits from replicate analyses (Table 3.8). Technical replicates were required for all subunits to be identified for both RNA Pol II and Mediator and will be the precedent for further

Table 3.8 RNA Pol II (~1ug) and VP16-Mediator (~0.5ug) totals.

	Total		RNA Pol II		VP16-Mediator	
	peptides	proteins	peptides	subunits	peptides	subunits
RNA Pol II rep1	3471	140	2660	10	6	4
RNA Pol II rep2	3204	157	2022	12	5	4
Mediator rep1	1046	74	53	6	612	26
Mediator rep2	1225	91	59	6	635	28

analyses. The total number of spectral counts for RNA Pol II and Mediator are between half and two-thirds the total number of spectral counts, indicating that they are indeed the major components of those samples. Analytical technical replicates are fairly good for this scale of sample, but sequence coverage and subunit coverage is exceptional, which was the ultimate goal of these experiments (Table 3.9 and Table 3.10).

Table 3.9 RNA Pol II spectral counts and subunit sequence coverage.

RNA Pol II		VP16-Mediator		Gene	Prot MW	Sequence Coverage
rep1	rep2	rep1	rep2			
1191	678	30	37	POLR2A	217206	29.2%
488	509	9	5	POLR2B	133897	32.7%
182	171	3	7	POLR2C	31441	25.5%
262	202	1	1	POLR2D	16311	40.8%
110	78	9	6	POLR2E	24551	29.5%
3	4	0	0	POLR2F	14478	8.7%
243	236	0	0	POLR2G	19294	35.5%
24	44	1	3	POLR2H	17143	46.7%
72	72	0	0	POLR2I	14523	67.2%
0	1	0	0	POLR2J	14131	14.2%
85	16	0	0	POLR2K	7004	14.2%
0	11	0	0	POLR2L	7645	16.4%
2660	2022	53	59	sums		
10	12	6	6	subunits		

Table 3.10 Mediator spectral counts and subunit sequence coverage.

RNA Pol II		VP16-Mediator		Gene	Prot MW	Sequence Coverage
rep1	rep2	rep1	rep2			
2	0	96	88	MED1	168478	14.4%
0	3	50	68	MED14	160607	11.3%
0	0	21	28	MED23	157114	4.3%
0	0	55	55	MED24	110305	7.9%
0	0	49	34	MED16	96793	8.3%
1	0	91	84	MED15	86753	13.6%
0	0	16	15	MED25	84389	5.8%
0	0	19	27	MED17	72876	10.3%
0	0	40	39	MED26	65446	18.3%
3	0	11	6	MED27	35432	3.5%
0	0	10	9	MED8	32819	19.9%
0	0	20	23	MED4	29745	17.0%
0	0	6	9	MED6	29298	24.8%
0	0	11	8	MED7	27245	15.0%
0	0	12	12	MED18	24453	11.1%
0	0	13	11	MED29	23473	26.7%
0	0	23	29	MED20	23222	16.0%
0	0	9	11	MED22	22221	21.4%
0	0	1	1	MED19	20431	4.5%
0	1	5	14	MED30	20277	14.0%
0	0	8	18	MED9	16403	11.6%
0	1	28	29	MED31	15805	26.7%
0	0	13	12	MED10	15688	9.6%
0	0	4	2	MED21	15564	12.5%
0	0	1	3	MED11	13129	23.9%
6	5	612	635	total Mediator peptides		
4	4	26	28	total Mediator subunits		

3.3 Discussion

Mass spectrometry is a powerful technique with enormous discovery potential. The MudPIT method was an elegant way to expand the peak capacity of 1D-LC/MS systems with 3D ion trap mass spectrometers. Furthermore, it has been applied successfully to characterize the consensus Mediator subunits (Sato, et al., 2004) as well as many other experiments. However, developing the MudPIT method on the hardware available, and consequently running many samples (Chapter 2) uncovered many technical challenges (Table 3.11). Some of these challenges could be worked around, and some simply could not.

Table 3.11 Pros or Cons of Proteomics Methodologies.

	MudPIT mass spectrometry	New Proteomics Platform
Protein Precipitation	poor recovery for low/very low concentration protein samples	Very good recovery for all concentrations of samples
Protein Proteolysis	NP-40 residual may exist with in solution digestion	FASP is very good for removal of detergents
LC/MS Analysis	Requires Single-Use Columns Salt Elutions NOT compatible with pumps, ESI emitters, MS	Commercial Columns can be used for extended time Requires Three HPLC pumps

Ultimately the MudPIT method simply was not rugged enough in my hands for reliable and consistent analyses. The high pH-low pH reversed phase 2D-LC separation detailed here was a logical progression of the concept for expanding the peak capacity of a system for greater depth of sample coverage. Many of the technical issues affecting the robustness of the MudPIT method were addressed with implementation of the high-low pH 2D method. It is much more reliable and robust than the MudPIT method ever was. I don't wake up in the middle of the night to drive 30 minutes to check on the mass spectrometer anymore; I don't need to!

The more critical modification to the proteomics platform was clearly the adaptation of the FASP protocol for preparing protein samples for mass spectrometry. In combination with a TCA precipitation protocol using a carrier protein, the method became virtually universal. The poor recovery of low abundance samples was addressed by two points, (1) complete precipitation recovery, and (2) effectively blocking the spin filter with a peptide that can flow through the filter limiting non-specific adsorption of sample protein/peptides. The potential of this method was enormous for a lab that routinely purifies very small amount of very low abundance proteins. Analyzing samples of this nature is not trivial. Many samples that would not have been feasible for analysis can now be characterized in depth. This will be demonstrated in Chapters 4 where activators that give low Mediator yields, p53 and p65/RelA/NFKb, are analyzed and data presented. It simply would not have been possible to analyze these samples with the MudPIT method used in Chapter 2.

3.4 Methods

3.4.1 CDK8 and SREBP-1a Mediator purifications

CDK8-Mediator was purified using anti-CDK8 antibodies (Santa Cruz) immobilized to Protein A/G Sepharose (Amersham Biosciences). Immobilized antibodies were incubated with HeLa nuclear extract at 4 °C. The antibody resin was then washed three times with 20 column volumes (CV) 0.5 M KCl HEGN and twice with 20 CV 0.15M KCl HEGN. Bound proteins were eluted with 0.1M Glycine, pH 2.7.

SREBP-1a-Mediator was purified from HeLa nuclear extract using GST-SREBP-1a (residues 1–50) immobilized to Glutathione-Sepharose beads (GE Life Sciences). After binding, the resin was washed five times with 10 column volumes (CV) 0.5 M KCl HEGN (20 mM Hepes, pH 7.6; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40 alternative) and one time with 10 CV 0.15 M KCl HEGN (0.02% NP-40 alternative). Bound proteins were eluted with 30 mM GSH in elution buffer (80 mM Tris, 0.1 mM EDTA, 10% Glycerol, 0.02% NP-40, 100 mM KCl) and applied to a 15% to 40% linear glycerol gradient (in 0.15 M KCl HEG) and centrifuged for 6 h at 55,000 rpm. Mediator-containing fractions (>1.0 MDa) were combined for analysis.

3.4.2 Sample preparation and mass spectrometry analyses of Mediator and RNA Pol II complexes.

Purified Mediator complex (~1ug) and RNA Pol II complex (~2ug) fractions were precipitated at 4°C using 20%(w/v) TCA, 0.067mg/mL insulin and 0.067%(w/v) deoxycholate. Precipitated protein pellets were washed twice with -20°C Acetone and air dried. Proteins were trypsin digested using a slightly modified Filter-Aided Sample Prep (FASP) protocol (Wiśniewski, et al., 2009). Briefly, protein pellets were suspended with 4%(v/v) SDS, 0.1M Tris pH 8.5, 10mM TCEP and incubated 30m ambient to reduce disulfides. Reduced proteins were diluted with 8M Urea, 0.1M Tris pH8.5 and iodoacetamide was added to 10mM and incubated 30m in total darkness. Reduced and alkylated proteins were then transferred to a Microcon YM-30 spin concentrator and washed twice with 8M Urea, 0.1M Tris pH 8.5 to remove SDS. Three washes with 2M Urea, 0.1M Tris pH8.5 were performed then trypsin and 2mM CaCl₂ were added and incubated approximately 2 hours in a 37°C water bath. Digested peptides were eluted and acidified with 5%(v/v) formic acid. Peptides were desalted online and fractionated with a Phenomenex Jupiter C18 (5µm 300Å; 0.25 x 150mm) column using a two dimensional LC/MS/MS method (Agilent 1100). Seven steps of increasing acetonitrile (3, 6, 9, 12, 16, 20 and 100%B; A: 20mM ammonium formate pH10, 4% acetonitrile and B: 10mM ammonium formate pH10, 65% acetonitrile) at 5µL/minute eluted peptides for a second dimension analysis on a Dionex Acclaim PepMap C18 (3µm 100Å; 0.075 x 150mm) running a gradient at

0.2 μ L/minute from 5 to 25% B in 100 minutes for steps one through six and 10 to 30% B in 100 minutes for step seven (A: 4% acetonitrile & B: 80% acetonitrile, both with 0.1% formic acid pH~2.5). PepMap eluted peptides were detected with an Agilent MSD Trap XCT (3D ion trap) mass spectrometer.

All spectra were searched with Mascot v2.2 (Matrix Sciences) against the International Protein Index (IPI) database version 3.65 with two missed cleavages and mass tolerances of $m/z \pm 2.0$ Da for parent masses and ± 0.8 Da for MS/MS fragment masses. Peptides were accepted above a Mascot ion score corresponding to a 1% false discovery rate (1% FDR) determined by a separate search of a reversed IPI v3.65 database. Peptides were then filtered and protein identifications were assembled using in-house software as described (Meyer-Arendt, et al., 2011; Resing, et al. 2004).

Chapter 4
Future Directions
Activator-Specific Mediator Interactomes

4.1 Introduction

The Mediator complex is a large protein complex that integrates signals from DNA-binding transcription activators. The activator SREBP-1a was shown to induce a structural shift in the Mediator complex compared to an activator-free Mediator complex. A mass spectrometry-based methodology was used to characterize the protein-protein interactions that were associated with each structural state. The results were that activator-bound Mediator contained an additional subset of cofactors compared with activator-free Mediator. The next question to address is whether distinct activators can regulate the recruitment of distinct cofactors. To test this idea the family of three SREBP isoforms, 1a, 1c and 2, are used to purify Mediator complexes for comparison. To further test the idea, additionally activators used to purify Mediator are p53 and p65/RelA (NFκB).

4.1.1 Activator-Specific Mediator Interactome

Distinct activators have been demonstrated to produce unique stable structural conformations of the Mediator complex (Taatjes, et al., 2002 and Figure 4.1). The functional relationship of this structural dynamics may be for the spatial and temporal regulation of protein-protein interactions that are activator and/or gene-specific. Evidence for this concept is provided in Chapter 2 comparing the activators SREBP-1a and VP16. Figure 4.1 shows the unique conformations of the Mediator complex when bound to activators SREBP-1a, VP16, p53 and the RNA Pol II CTD. It is expected that when Activator-bound

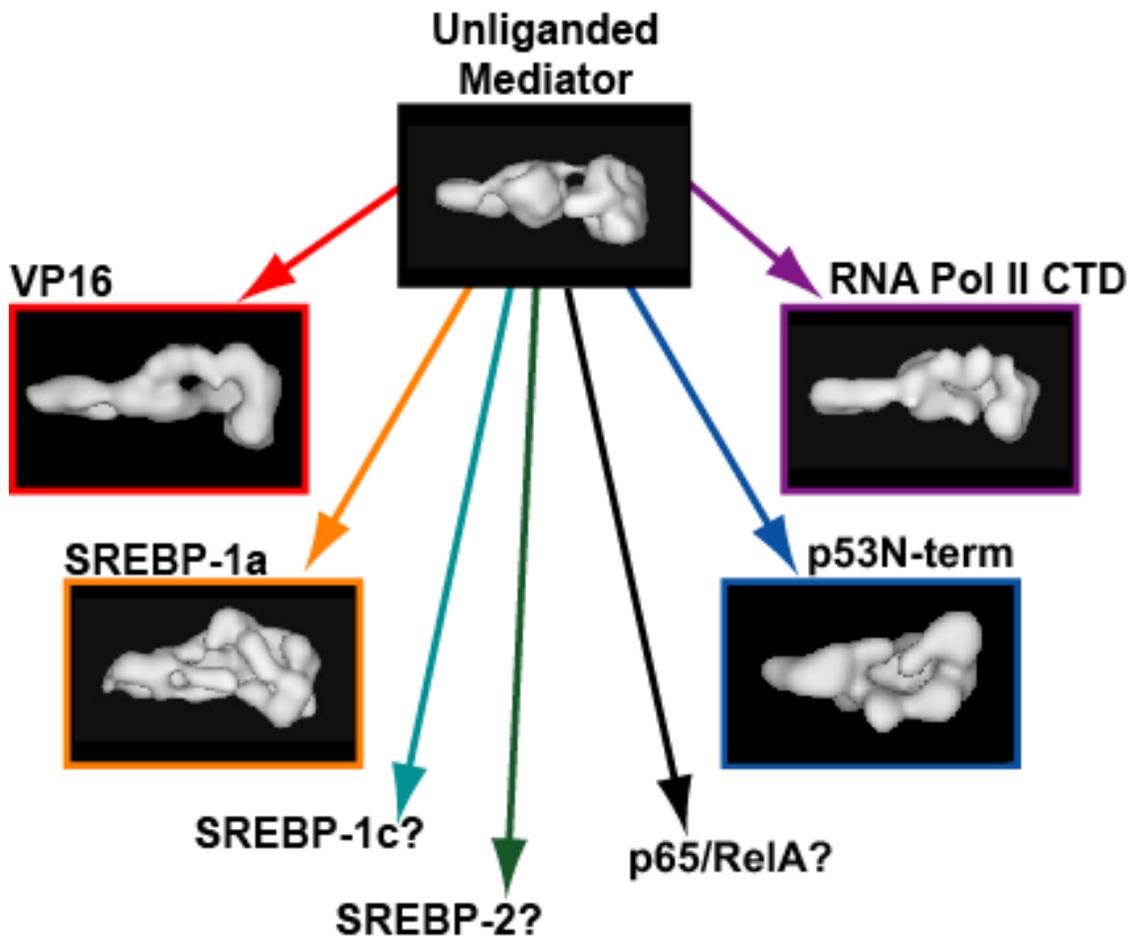


Figure 4.1 Activator-Specific Mediator Conformations. Distinct structures of activator-bound Mediator regulate protein-protein interactions. Activators VP16, SREBP-1a, p53 and the CTD of RNA Pol II. What might the structures of SREBP-1c, SREBP-2 and p65/RelA look like?

Mediator structures are rendered for SREBP-1c, -2 and p65/RelA that they also will be distinct. These unique structural conformations can elegantly control the functional output of a specific gene. By comprehensively characterizing the protein interactome of each activator-Mediator complex, new hypotheses can be generated regarding activator-specific cofactors required for the proper regulation of target genes. Mass spectrometry is an ideal technology for characterizing and comparing protein-protein interaction networks (Köcher T and Superti-Furga, 2007; Gingras, et al., 2007; Choudhary and Mann, 2010). Improvements in the proteomic platform have enabled the analyses of purified protein complex samples such as these performed here. All of the additional activator complexes analyzed in this chapter have reasonably low protein yields and concentrations, which would have been difficult if not impossible to process using the MudPIT methodology from Chapter 2. In fact p53 was previously attempted with insufficient recovery of protein to even attempt to put on the mass spectrometer. The use of insulin as a carrier protein gives very high recovery of purified protein and the adaptation of the FASP sample preparation allows for the removal of the added insulin and efficient purification of target protein tryptic peptides ready for mass spectrometry.

The data presented here is the discovery phase for the generation of new hypotheses for SREBP, p53 and p65/RelA gene activation. Some interesting future directions derived from the data will be offered.

4.1.2 Sterol Regulatory Element-Binding Proteins (SREBP)

Sterol regulatory element-binding proteins (SREBP) are DNA binding transcription activators that regulate fatty acid and cholesterol metabolism (Eberle, et al., 2004; Horton, et al., 2003; Brown and Goldstein, 1997). Two genes encode three SREBP isoforms called SREBP-1a, SREBP-1c and SREBP-2 (Goldstein, et al., 2002). SREBP-1a and SREBP-1c are splice isoforms differing by only the first exon of the gene *srebf1*. SREBP-2 is a product of the *srebf2* gene. SREBPs are unique membrane proteins of the endoplasmic reticulum (ER) in that they are transcription activators. A protein called SREBP cleavage-activating protein (SCAP) acts as a sterol sensor that binds SREBPs to escort them from the ER to the Golgi where the amino-terminal (~500 amino acids) is cleaved and released to translocate to the nucleus to bind the sterol response elements at the promoters of target genes (Goldstein, et al., 2002). All three isoforms target the MED15 subunit of the Mediator complex (Yang, et al., 2006). If overexpressed, all SREBP isoforms activate genes involved in cholesterol and fatty acid synthesis and the low-density lipoprotein (LDL) receptor (Horton, et al., 2003). The different isoforms do have selective roles, however. SREBP-1a is a potent activator for all SREBP-responsive genes, where SREBP-1c is less potent (Horton, et al., 2003), though both are activators for genes that process fatty acids and the assembly of triglycerides and phospholipids. SREBP-2 is an activator for genes that synthesize cholesterol and the LDL receptor (Horton, et al., 2003). Due to the clear and distinct genes

that are activated by SREB-1a/-1c and SREBP-2 (Horton, et al., 2003), it is proposed that there may be subsets of activator-specific cofactors recruited to the Mediator complex that are involved in activator-specific gene activation. Therefore, the activation domains of each SREBP isoform will be used to bind the Mediator complex and any isoform-specific cofactors that may be involved in the regulation of gene activation for a comprehensive protein interactome analysis of each.

4.1.3 p53 and p65/RelA

The transcription activators p53 and p65/Rel A, part of the NFkB complex (Perkins, 2007; Hayden and Ghosh, 2004,2008), are also compared more for convenience, though, it does make for an interesting comparison. These activators regulate many target genes. Both activators bind the Mediator complex (p53: Ito, et al., 2003; Meyers, et al., 2010; and p65/RelA: Näär, et al., 1999) through their activation domains to activate transcription at target gene promoters. The tumor suppressor p53 is often thought of as a sensor for the overall health of a cell directing cell fate to senescence or apoptosis (Kruse and Gu, 2009). NFkB, on the other hand, is often thought of as tumor-promoting and anti-apoptotic transcription factor (O'Shea and Perkins. 2008). Many modes of regulation are known for these transcription factors, however, how these transcription factors activate transcription at the promoter is still less clear. This chapter is not for in-depth molecular mechanisms of gene activation for p53 or

p65/RelA (or SREBPs either), but rather to provide data that supports a hypothesis that unique activators, when bound to Mediator, initiate distinct protein-protein interactions.

4.2 Results

To test the hypothesis that distinct transcription activators can regulate the protein-protein interactions for the activator-bound Mediator complex, a mass spectrometry based approach was applied again to comprehensively characterize activator-specific protein complexes, as in Chapter 2, but with additional activators. Improvements in the proteomics platform, detailed in Chapter 3, which have significantly improved the sensitivity and depth of coverage for protein analyses have made these particular activator-bound Mediator purifications amenable to analyses.

4.2.1 Mediator-Bound SREBP (-1a, -1c, -2)

GST-fusions of the activation domains for SREBP-1a, SREBP-1c and SREBP-2 were used as bait for pulldowns from HeLa nuclear extract (Figure 4.2.A & B). The protein complexes bound were gently eluted with glutathione (GSH) (Figure 4.2.C) and applied to a 2 mL glycerol gradient (Figure 4.2.D), which resolves much of the free fusion-activator and smaller activator-bound cofactors, such as the acetyltransferase coactivators p300/CBP (~300kD glycerol

gradient fractions 5-9, Figure 4.2.D), from the Mediator-containing fractions (~1+MDa, boxed in red, Figure 4.2D). One and two glycerol gradient purifications were used for SREBP-1a, while four to eight glycerol gradient purifications were combined for SREBP-1c and SREBP-2 for a single mass

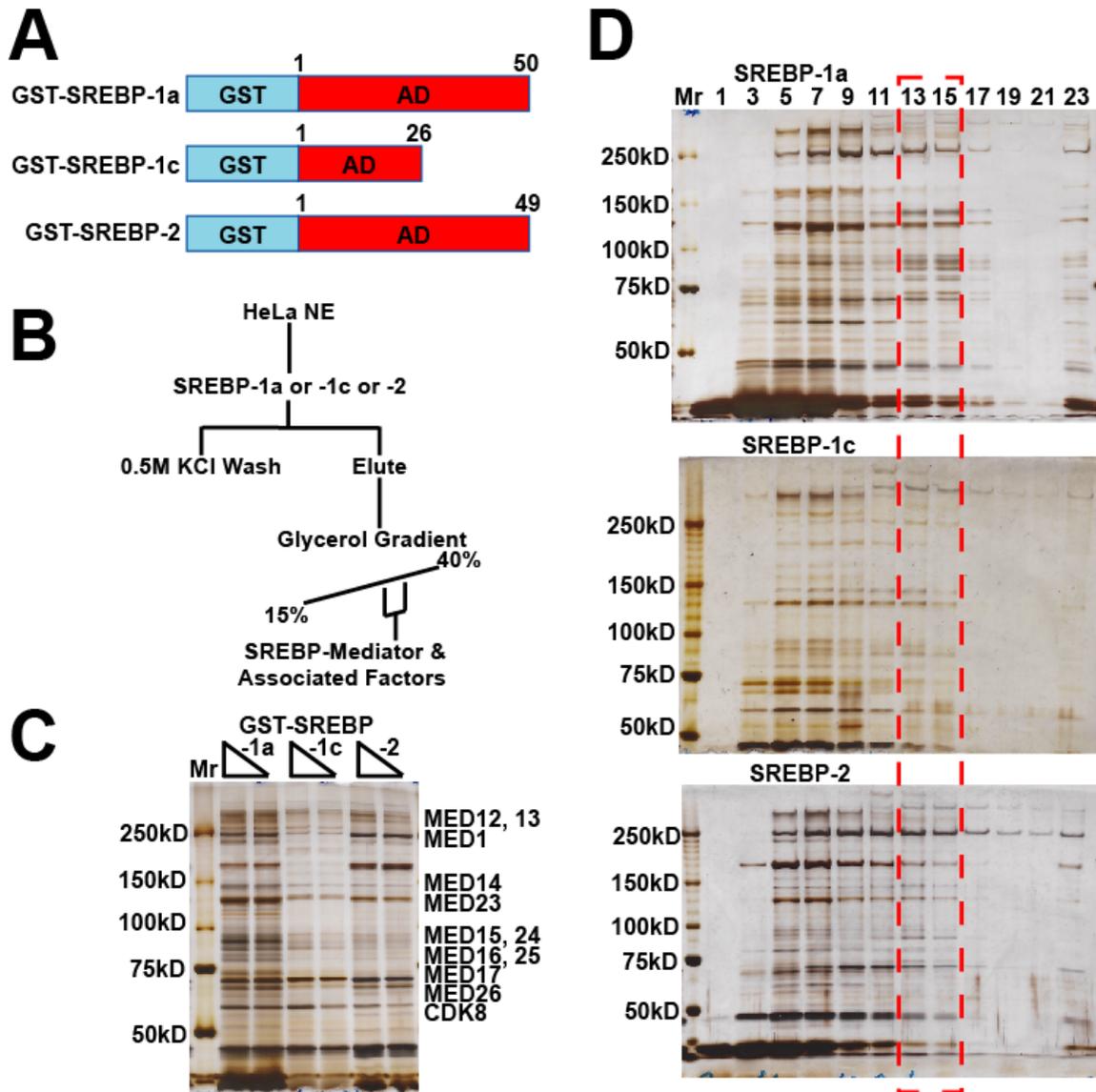


Figure 4.2 Sterol Regulatory Element-Binding Protein (SREBP) family of transcription activators bind Mediator. (A) Diagram of activation domain fusion proteins. (B) Purification scheme. (C) GSH elutions/inputs for SREBP-1a, -1c and -2. (D) Silver stained acrylamide gels of SREBP purifications. What other cofactors are recruited that are activator-specific?

spectrometry analysis. Four replicates were completed for each isoform of SREBP. All 12 analyses were combined for 71,789 total peptides identified; 5465 unique peptides which identified 1535 total proteins with 50.55% single peptide protein identifications. It is common to accept protein identifications with two unique peptides, however, the stringent bioinformatics data processing provides a probability of 1% False Discovery Rate (1%FDR). Many of the single peptide identifications may be noise, or low level contamination, or they can be cofactors that are parts of complexes with more peptides per identification filling in protein sub-complexes that are sub-stoichiometric.

The total peptide & protein identifications, the total number of Mediator peptides and the total number of RNA Pol II peptides identified in this SREBP-Mediator study by replicate are given in Table 4.1. The Mediator complex and RNA Pol II are predominant constituents of each purification. Spectral counts for individual subunits are given in Table 4.2 for the Mediator complex and Table 4.3 for RNA Pol II. Consistent with greater activation potency (Horton, et al., 2003), SREBP-1a pulls down 2 to 4-fold more Mediator and more than two-fold more

Table 4.1 SREBP Isoforms Spectral Counts totals for all protein identifications, Mediator and RNA Pol II.

	SREBP-1a				SREBP-1c				SREBP-2			
total peptides	8046	8335	5087	5211	4639	3411	5160	5755	6988	7377	6014	5766
total proteins	592	602	302	350	317	332	413	522	693	598	432	404
total MEDIATOR peptides	2025	2151	1592	1555	843	883	491	457	598	169	498	483
total MEDIATOR proteins	32	32	32	32	30	29	28	25	30	19	28	27
total RNA Pol II peptides	235	261	266	255	123	143	98	119	133	37	95	67
total RNA Pol II proteins	9	9	9	9	8	6	9	8	8	5	7	7
replicate	1	2	3	4	1	2	3	4	1	2	3	4
	<i>technical replicates</i>	<i>biological replicates</i>										

RNA Pol II by total spectral counts compared with SREBP-1c and SREBP-2, from four to eight-fold fewer gradients. The silver stained acrylamide gels of glycerol gradient fractions confirm at a glance a clear difference in total proteins pulled down as well as total Mediator polypeptides (red box in Figure 4.2.D). Not all Mediator subunits were identified in each replicate, but between four replicates, all Mediator subunits were identified with each SREBP isoform (Table 4.2). Subunits of RNA Pol II are provided in Table 4.3.

Table 4.2 SREBP Spectral Counts for Mediator by subunit.

SREBP-1a	SREBP-1c	SREBP-2	Gene	Prot MW
765	269	219	MED12	243081
88	36	32	MED12 or MED12L	
0	0	1	MED12L	240032
516	247	108	MED13	239318
3	0	0	MED13 or MED13L	
300	91	34	MED13L	242602
709	262	215	MED1	168478
555	127	68	MED14	160607
379	103	33	MED23	156194
480	167	188	MED24	110305
260	64	29	MED16	96793
468	148	121	MED15	86753
89	44	18	MED25	84389
314	67	114	MED17	72876
135	57	35	MED26	65446
31	17	12	CDK19	56802
91	53	49	CDK8 or CDK19	
4	0	0	CDK8	53284
218	71	32	MED27	35432
76	47	17	CCNC	33243
88	48	41	MED8	32819
197	102	57	MED4	29745
25	5	5	MED6	28425
197	101	53	MED7	27245
21	3	1	MED19	26273
26	4	3	MED18	24453
120	24	51	MED29	23473
95	25	23	MED20	23222
106	78	41	MED30	20277
47	11	1	MED28	19520
107	20	11	MED22	16480
138	34	47	MED9	16403
79	21	9	MED31	15805
100	24	14	MED10	15688
374	247	51	MED21	15564
122	57	15	MED11	13129

After comparing the Mediator and RNA Pol II composition of the purifications from each SREBP isoform, a list of exclusive cofactors was also generated for each. We do not propose that each cofactor is truly exclusive, since it could just be below the limit of detection for the other two isoforms, but rather the data presents an opportunity to test new hypotheses.

Factors exclusive to SREBP-1a are given in Table 4.4. Vinculin (VCL) is a cytoskeleton

Table 4.3 SREBP Spectral Counts for RNA Pol II by subunit.

SREBP-1a	SREBP-1c	SREBP-2	Gene	Prot MW
403	218	120	POLR2A	217206
270	93	85	POLR2B	133897
57	25	19	POLR2C	31441
37	15	20	POLR2E	24551
30	13	12	POLR2G	19294
55	37	22	POLR2H	17143
89	64	38	POLR2D	16311
57	15	16	POLR2I	14523
17	3	0	POLR2J	14131
2	0	0	POLR2K	7004

protein and it is completely unknown why it would associate with the activation domain of SREBP-1a. TXLNA and KLRAQ1 are coiled-coil domain-containing proteins, again with unknown function for SREBP-1a

activation. CIAO1 is an adapter protein involved in pre-mRNA processing that associates with FAM96B and MMS19. These cofactors may make up a sub-stoichiometric subcomplex that binds SREBP-1a-Mediator. A number of SREBP-1a exclusive cofactors appear to be involved in vesicle trafficking. Since SREBPs are membrane-bound in the ER prior to cleavage and activation,

Table 4.4 SREBP-1a Exclusive Cofactors.

BP1a	BP1c	BP2	Gene	Prot MW	Protein descriptor
46	0	0	VCL	116722	ISOFORM 1 OF VINCULIN.
37	0	0	TXLNA	61891	ALPHA-TAXILIN.
35	0	0	KLRAQ1	88314	ISOFORM 1 OF KLRAQ MOTIF-CONTAINING PROTEIN 1.
33	0	0	CIAO1	37840	PROBABLE CYTOSOLIC IRON-SULFUR PROTEIN ASSEMBLY PROTEIN CIAO1.
32	0	0	GOLGA4	261140	ISOFORM 1 OF GOLGIN SUBFAMILY A MEMBER 4.
29	0	0	FAM96B	17663	UPF0195 PROTEIN FAM96B.
24	0	0	PIBF1	89805	ISOFORM 1 OF PROGESTERONE-INDUCED-BLOCKING FACTOR 1.
22	0	0	GCC1	87811	GRIP AND COILED-COIL DOMAIN-CONTAINING PROTEIN 1.
21	0	0	STXBP2	67700	CDNA FLJ54775. HIGHLY SIMILAR TO SYNTAXIN-BINDING PROTEIN 2.
18	0	0	KLC4	68640	ISOFORM 1 OF KINESIN LIGHT CHAIN 4.
18	0	0	PDDC1	19539	ISOFORM 3 OF PARKINSON DISEASE 7 DOMAIN-CONTAINING PROTEIN 1.
17	0	0	SEPTIN8	49814	ISOFORM 2 OF SEPTIN-8.
16	0	0	CXorf15	60586	GAMMA-TAXILIN.
16	0	0	NCOA2	159157	NUCLEAR RECEPTOR COACTIVATOR 2.
15	0	0	TRIM11	52774	ISOFORM 1 OF TRIPARTITE MOTIF-CONTAINING PROTEIN 11.
15	0	0	LRRFIP2	82171	ISOFORM 1 OF LEUCINE-RICH REPEAT FLIGHTLESS-INTERACTING PROTEIN 2.
13	0	0	VPS33B	70615	VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN 33B.
12	0	0	PKN2	112035	SERINE/THREONINE-PROTEIN KINASE N2.
12	0	0	RFWD3	85094	RING FINGER AND WD REPEAT DOMAIN-CONTAINING PROTEIN 3.
12	0	0	KLC1	65310	ISOFORM A OF KINESIN LIGHT CHAIN 1.
11	0	0	TSSC4	34285	ISOFORM 1 OF PROTEIN TSSC4 (FRAGMENT).
10	0	0	KLC2	68935	KINESIN LIGHT CHAIN 2.
10	0	0	MMS19	115749	CDNA FLJ55586. HIGHLY SIMILAR TO MMS19-LIKE PROTEIN.
10	0	0	TANC1	202192	ISOFORM 1 OF PROTEIN TANC1.
10	0	0	UBA5	44863	UBIQUITIN-LIKE MODIFIER-ACTIVATING ENZYME 5.

perhaps association of these cofactors is suggestive of a feedback mechanism. Cofactors involved in the ubiquitin proteasome system are also represented in the SREBP-1a exclusive factors list. TRIM11 and RFWD3 are E3 ubiquitin ligases and UBA5 an ubiquitin activating enzyme. In fact, TRIM11 has been reported to interact with MED15 to induce its degradation in the regulation of TFG-beta signaling (Ishikawa, et al., 2006). TRIM11 has also been shown to regulate the neurogenic transcription factor Pax6 also through the ubiquitin proteasome system (Tuoc and Stoykova, 2008). RFWD3 has been reported to bind the MDM2-p53 complex when phosphorylated by ATM/ATR to prevent p53 degradation, which is unique since all other E3 ligases promote p53 degradation (Fu, et al., 2010).

A very interesting subcomplex of cofactors exclusive to SREBP-1c is given in Table 4.5. This subcomplex consists of PIK3C3, BECN1, PIK3R4 and UVRAG which make up a vacuolar protein sorting complex (Behrends, et al., 2010). This vacuolar sorting complex is known to regulate phosphoinositide

Table 4.5 SREBP-1c Exclusive Cofactors.

BP1a	BP1c	BP2	Gene	Prot MW	Protein descriptor
0	57	0	PIK3C3	101549	PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT TYPE 3.
0	28	0	SHCBP1	75660	SHC SH2 DOMAIN-BINDING PROTEIN 1.
0	26	0	DLAT	68997	COMPONENT OF PYRUVATEDEHYDROGENASE COMPLEX. MITOCHONDRIAL.
0	17	0	BECN1	51896	BECLIN-1.
0	15	0	PIK3R4	153103	PHOSPHOINOSITIDE 3-KINASE REGULATORY SUBUNIT 4.
0	15	0	SF3B4	44386	SPLICING FACTOR 3B SUBUNIT 4.
0	10	0	FAM192B	29152	RCNIP30 (FRAGMENT).
0	8	0	UVRAG	78151	UV RADIATION RESISTANCE-ASSOCIATED GENE PROTEIN.
0	8	0	MIS12	24140	PROTEIN MIS12 HOMOLOG.
0	8	0	FEM1B	70264	PROTEIN FEM-1 HOMOLOG B.
0	5	0	PRDX3	27693	THIOREDOXIN-DEPENDENT PEROXIDE REDUCTASE. MITOCHONDRIAL.
0	4	0	C15orf23	35438	PUTATIVE TRAF4-ASSOCIATED FACTOR 1.
0	4	0	MRPS21	10742	28S RIBOSOMAL PROTEIN S21. MITOCHONDRIAL.
0	4	0	C14orf43	114989	UNCHARACTERIZED PROTEIN C14ORF43.
0	4	0	ELF2	62711	ISOFORM 1 OF ETS-RELATED TRANSCRIPTION FACTOR ELF-2.
0	4	0	PYGL	97149	GLYCOGEN PHOSPHORYLASE. LIVER FORM.
0	4	0	TUFM	49875	TU TRANSLATION ELONGATION FACTOR. MITOCHONDRIAL PRECURSOR.

signaling to facilitate the assembly of an autophagosome (Simonsen and Tooze, 2009). Autophagy is the cellular process of recycling defective proteins by collecting them in autophagosomal vesicles for delivery to the lysosome for degradation. In addition to recycling proteins, autophagy catabolizes diverse cellular energy sources (Simonsen and Tooze, 2009). This recently discovered dynamic feedback between degradation and cellular metabolism provides a reasonable premise for a functional interaction of SREBP and a component of the autophagy interaction network. It seems likely the PIK3C3 subcomplex binds directly to the activation domain of SREBP-1c, however, if it did bind Mediator in a SREBP-1c-bound state, it could directly connect transcription regulation and autophagy.

SREBP-2 exclusive cofactors are given in Table 4.6. Many cofactors identified have unknown function and a number are involved in vesicle trafficking, cytosolic enzyme and mitochondrial cofactors. It is possible many of these cofactors have multiple functionalities with their nuclear functions unknown. In this case, there may be many new hypotheses to test. One trend is the number of motor proteins and more so with SREBP-2, the number of proteasome subunits. Ubiquitin-proteasome signaling has been connected to transcription regulation (Hammond-Martel, 2011) but many of the exact molecular mechanism are yet unknown. A cofactor not included in Table 4.6 found significantly enriched in SREBP-2 is KIAA0368, or ECM29. A single spectral count was identified in both the SREBP-1a and SREBP-1c purifications, which is why it did not appear on the exclusive cofactors list (Table 4.6), while 319 spectral counts

were identified in SREBP-2. This is a significant number of spectral counts compared with ~1700 for the 2 MDa Mediator complex. ECM29 is a HEAT repeat protein that interacts with the 26S Proteasome. Genome-wide two hybrid and mass spectrometry identified molecular motors, endosomal components and ubiquitin-proteasomal factors such as ECM29-interacting proteins (Gorbea, et al., 2010). If ECM29 interacted with SREBP-2-Mediator, it would be another example of transcription regulation coupled with ubiquitin-proteasome signaling, but also provide hypotheses for new molecular mechanisms of SREBP gene activation which includes molecular motors and endosomal processing.

Table 4.6 SREBP-2 Exclusive Cofactors.

BP1a	BP1c	BP2	Gene	Prot MW	Protein descriptor
0	0	78	PYCR2	33637	PYRROLINE-5-CARBOXYLATE REDUCTASE 2.
0	0	49	MCM2	101896	DNA REPLICATION LICENSING FACTOR MCM2.
0	0	31	RRP12	143702	ISOFORM 1 OF RRP12-LIKE PROTEIN.
0	0	26	ALDH3A2	54848	ISOFORM 1 OF FATTY ALDEHYDE DEHYDROGENASE.
0	0	25	TFRC	84871	TRANSFERRIN RECEPTOR PROTEIN 1.
0	0	23	ARFGEF1	208767	BREFELDIN A-INHIBITED GUANINE NUCLEOTIDE-EXCHANGE PROTEIN 1.
0	0	23	ENO1	47169	ISOFORM ALPHA-ENOLASE OF ALPHA-ENOLASE.
0	0	19	NCL	65962	HIGHLY SIMILAR TO NUCLEOLIN.
0	0	19	NOC4L	58468	NUCLEOLAR COMPLEX PROTEIN 4 HOMOLOG.
0	0	17	MYO9B	243401	MYOSIN IXB ISOFORM 1.
0	0	17	MYBBP1A	148855	ISOFORM 1 OF MYB-BINDING PROTEIN 1A.
0	0	16	SCFD1	72380	SEC1 FAMILY DOMAIN-CONTAINING PROTEIN 1.
0	0	14	RNF160	205177	ZINC FINGER PROTEIN 294.
0	0	14	ATP5B	56560	ATP SYNTHASE SUBUNIT BETA. MITOCHONDRIAL.
0	0	14	PPIA	18012	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A.
0	0	14	SLC3A2	57945	ISOFORM 2 OF 4F2 CELL-SURFACE ANTIGEN HEAVY CHAIN.
0	0	13	FANCD2	166462	ISOFORM 1 OF FANCONI ANEMIA GROUP D2 PROTEIN.
0	0	13	PSMB5	28480	PROTEASOME SUBUNIT BETA TYPE-5.
0	0	13	P4HA2	60902	ISOFORM IIB OF PROLYL 4-HYDROXYLASE SUBUNIT ALPHA-2.
0	0	13	PSMA2	25899	PROTEASOME SUBUNIT ALPHA TYPE-2.
0	0	12	PAICS	47958	PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE.
0	0	12	POLDIP3	48102	SIMILAR TO POLYMERASE DELTA-INTERACTING PROTEIN3.
0	0	11	TELO2	91747	TELOMERE LENGTH REGULATION PROTEIN TEL2 HOMOLOG.
0	0	11	PFKM	81776	ISOFORM 2 OF 6-PHOSPHOFRUCTOKINASE. MUSCLE TYPE.
0	0	11	PSMA1	29556	ISOFORM SHORT OF PROTEASOME SUBUNIT ALPHA TYPE-1.
0	0	11	INTS6	100390	ISOFORM 1 OF INTEGRATOR COMPLEX SUBUNIT 6.
0	0	11	GALNT2	64733	POLYPEPTIDE N-ACETYL GALACTOSAMINYLTRANSFERASE 2.
0	0	11	CD97	81743	ISOFORM 2 OF CD97 ANTIGEN.
0	0	10	IPO7	119517	IMPORTIN-7.
0	0	10	COPB1	107142	COATOMER SUBUNIT BETA.
0	0	10	SPG20	72833	SPARTIN.
0	0	10	PML	97551	ISOFORM PML-1 OF PROBABLE TRANSCRIPTION FACTOR PML.
0	0	10	HNRNPR	70943	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN R.
0	0	10	HSPH1	92116	ISOFORM BETA OF HEAT SHOCK PROTEIN 105 KDA.
0	0	10	GANAB	109438	ISOFORM 2 OF NEUTRAL ALPHA-GLUCOSIDASE AB.
0	0	10	PFN1	15054	PROFILIN-1.
0	0	10	MRPS34	25650	28S RIBOSOMAL PROTEIN S34. MITOCHONDRIAL.

4.2.2 Mediator-Bound p53 & p65/RelA

GST-fusions of the activation domains for p53 and p65/RelA were used as bait for pulldowns from HeLa nuclear extract (Figure 4.3.A & B). The protein complexes bound were gently eluted with glutathione (GSH) (Figure 4.3.C) and applied to a 2 mL glycerol gradient (Figure 4.3.C), which resolves much of the free fusion-activator and small activator-bound cofactors, such as the acetyltransferase coactivators p300/CBP (~300kD glycerol gradient fractions 5-9, Figure 4.3.C), from the Mediator-containing fractions (~1+MDa, boxed in red,

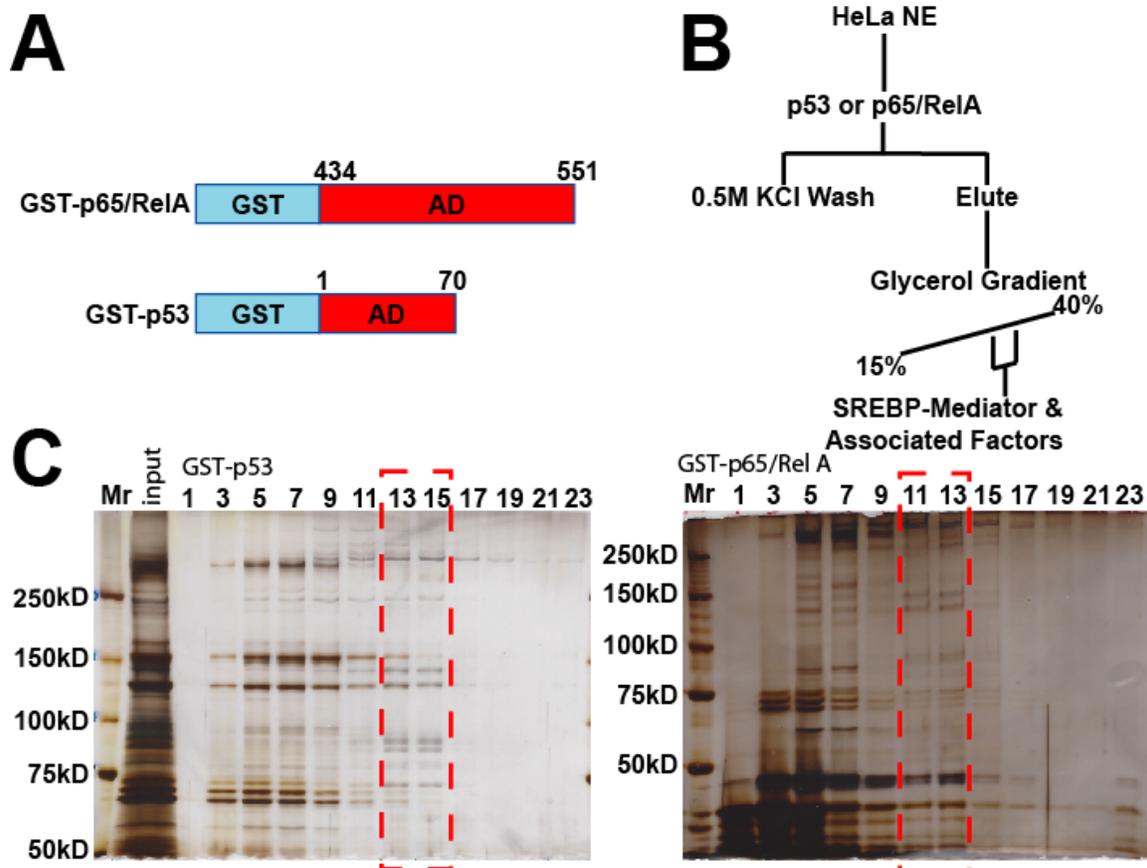


Figure 4.3 p53 and p65/Rel A bind Mediator. (A) diagram activation domain fusion proteins. (B) Purification scheme used. (C) Silver stained acrylamide gels of p53 and p65/RelA-Mediator purifications. What other cofactors are recruited that are activator-specific?

Table 4.7 p53AD and p65/RelA Spectral Count totals for all protein identifications, Mediator and RNA Pol II.

	p53		p65/RelA	
total peptides	4529	4752	2957	3059
total proteins	527	517	321	343
total MEDIATOR peptides	540	605	481	433
total MEDIATOR proteins	29	32	30	29
total RNA Pol II peptides	92	113	50	63
total RNA Pol II proteins	8	9	6	7
replicate	1	2	1	2
	<i>technical replicates</i>		<i>technical replicates</i>	

Figure 4.3.C). Four glycerol gradient purifications were used for p53, while eight glycerol gradient purifications were combined for p65/RelA for a single mass spectrometry analysis. Two replicates were completed for each p53 and p65/RelA. A 1% False Discover Rate (1%FDR) was again applied. Many of the single peptide identifications may be noise, or low level contamination, or they can be cofactors that are parts of complexes with more peptides per identification filling in protein sub-complexes that are sub-stoichiometric.

Total protein and peptide identifications, total Mediator and total RNA Pol II spectral counts are given in Table 4.7. The totals for p53 are very close to those from SREBP-1c and SREBP-2, while p65/RelA total are lower. The same type of experiment with p65/RelA was reported with three cell types, THP1, Jurkat and HeLa (Owen, et al., 2005). A couple of major difference were Owen et al., 2005, (1) did not run GST elutions over a glycerol gradient, and (2) they eluted with samples with loading buffer and prepared samples for mass spectrometry using in-gel digestion. This approach would increase the likelihood

Table 4.8 p53AD and p65/RelA Mediator Spectral Count totals for each subunit.

p53	p65	Gene	Prot MW
134	89	MED12	243081
30	28	MED13L	242602
106	72	MED13	239318
107	74	MED1	168478
112	62	MED14	160607
38	31	MED23	156194
112	95	MED24	110305
21	23	MED16	96793
63	56	MED15	86753
11	14	MED25	84389
36	25	MED17	72876
12	16	MED26	65446
5	9	CDK19	56802
37	13	CDK19 or CDK8	
3	0	CDK8	53284
21	24	MED27	35432
13	31	CCNC	33243
26	29	MED8	32819
53	44	MED4	29745
4	2	MED6	28425
16	22	MED7	27245
6	5	MED19	26273
2	1	MED18	24453
33	27	MED29	23473
9	16	MED20	23222
25	11	MED30	20277
1	12	MED28	19520
17	21	MED22	16480
4	9	MED9	16403
11	13	MED31	15805
8	5	MED10	15688
33	31	MED21	15564
36	4	MED11	13129
1145	914	Total Spectral Counts	

Table 4.9 p53AD and p65/RelA RNA Pol II Spectral Count totals for each subunit.

p53	p65	Gene	Prot MW
100	48	POLR2A	217206
39	15	POLR2B	133897
6	7	POLR2C	31441
16	29	POLR2D	16311
7	8	POLR2E	24551
7	2	POLR2G	19294
18	3	POLR2H	17143
11	1	POLR2I	14523
1	0	POLR2J	14131
205	113	Total Spectral Counts	

of non-specific identifications and also would limit overall sensitivity of the analysis. This is most easily demonstrated by the number of Mediator subunits identified by Owen, et al., 2005, which was nine, while we identified all 31 subunits (Table 4.8). Total protein identifications are not compared due to the completely different methods used for each analysis which cannot account for contaminating cofactors.

All expected Mediator subunits

were identified for both p53 and p65/RelA (Table 4.8). Total spectral counts for Mediator in p53 and p65/RelA were quite close. Nine of twelve RNA Pol II

p53/p65 Protein identifications 1% FDR / >1 spectral counts

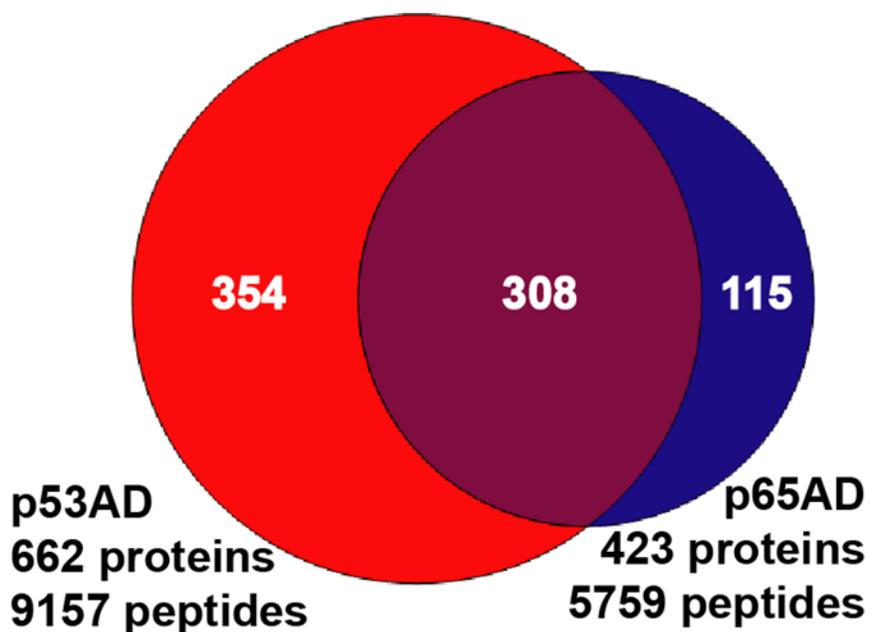


Figure 4.4 Venn diagram of all identifications for p53 and p65/RelA.

subunits were identified with p53 and eight with p65/RelA. Only with a purified RNA Pol II sample have all twelve Pol II subunits been identified.

Total protein identifications were higher for p53 than p65/RelA, despite very similar numbers for Mediator. This may suggest somewhat more promiscuous binding of the p53 activation domain relative to p65/RelA. Figure 4.4 shows a Venn diagram of total protein identifications found in both p53 and p65/RelA as well as unique protein identifications. When p53 exclusive cofactors are assembled, there is a trend in the number of adapter proteins (AP2 and AP1) found. These adapter proteins are found in clathrin-coated vesicles. COPB2 and COPA, also exclusive to p53 are coatamer proteins that regulate the

Table 4.10 p53AD exclusive cofactors.

p53	p65	Gene	Prot MW	Protein descriptor
248	0	RIF1	274466	TELOMERE-ASSOCIATED PROTEIN RIF1
72	0	MRPS25	20116	28S RIBOSOMAL PROTEIN S25. MITOCHONDRIAL
41	0	LANCL1	45283	LANC-LIKE PROTEIN 1
40	0	AP2A1	105370	AP-2 COMPLEX SUBUNIT ALPHA-1
37	0	CALU	38051	CDNA FLJ31776 FIS. HIGHLY SIMILAR TO CALUMENIN
33	0	AP2B1	104553	AP-2 COMPLEX SUBUNIT BETA
30	0	KIF14	186492	KINESIN-LIKE PROTEIN KIF14
24	0	AP2M1	49655	AP-2 COMPLEX SUBUNIT MU
22	0	RPS4X	29598	40S RIBOSOMAL PROTEIN S4. X ISOFORM
19	0	PPP1CB	37187	SER/THR-PROTEIN PHOSPHATASE PP1-BETA CATALYTIC SUBUNIT
19	0	WAPAL	132946	WINGS APART-LIKE PROTEIN HOMOLOG
17	0	REPIN1	63575	REPLICATION INITIATOR 1
15	0	AP2S1	12417	AP-2 COMPLEX SUBUNIT SIGMA
15	0	LACTB	60694	SERINE BETA-LACTAMASE-LIKE PROTEIN. MITOCHONDRIAL
15	0	TAF4	110114	TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 4
14	0	MRPS26	24212	28S RIBOSOMAL PROTEIN S26. MITOCHONDRIAL
14	0	MRPS6	14227	28S RIBOSOMAL PROTEIN S6. MITOCHONDRIAL
14	0	TAF5	86830	TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 5
13	0	TCP1	60344	T-COMPLEX PROTEIN 1 SUBUNIT ALPHA
13	0	YTHDC2	160248	PROBABLE ATP-DEPENDENT RNA HELICASE YTHDC2
12	0	CCT3	60463	CHAPERONIN CONTAINING TCP1. SUBUNIT 3
12	0	MRPS22	41280	28S RIBOSOMAL PROTEIN S22. MITOCHONDRIAL
15	0	AP1B1	20505	AP-1 COMPLEX SUBUNIT BETA-1.
11	0	COPB2	102487	COATOMER SUBUNIT BETA
11	0	CUL4A	87680	CULLIN-4A
11	0	HAUS1	31863	HAUS AUGMIN-LIKE COMPLEX SUBUNIT 1
11	0	HCFC1	208732	HOST CELL FACTOR
11	0	KDM3B	191611	LYSINE-SPECIFIC DEMETHYLASE 3B
11	0	PELP1	124868	PROLINE-. GLUTAMICACID-. LEUCINE-RICH PROTEIN 1
11	0	POLDIP3	48102	HIGHLY SIMILAR TO POLYMERASE DELTA-INTERACTING PROTEIN 3
11	0	PRPF6	106925	PRE-MRNA-PROCESSING FACTOR 6
11	0	RPS6	28681	40S RIBOSOMAL PROTEIN S6
10	0	COPA	138346	COATOMER SUBUNIT ALPHA

transport of proteins between the ER and the Golgi with non-clathrin-coated vesicles. There is a significant representation of cofactors involved in vesicle transport both with and without clathrin. Whether these protein-protein interactions are non-specific contaminants or bona fide functional interactions will require biochemical validation and a determination of specificity for p53-Mediator or p53-alone. The cofactor Rif1, also specific to p53 with 248 spectral counts, is worth noting due to the rather high number of spectral counts. Rif1 is a telomere-binding protein in yeast regulating telomere length. In humans, however, the

Table 4.11 p65/Rel A exclusive cofactors.

p53	p65	Gene	Prot MW	Protein descriptor
0	62	GSTM3	26560	GLUTATHIONE S-TRANSFERASE MU 3
0	22	CTSA	54466	LYSOSOMAL PROTECTIVE PROTEIN
0	14	GLB1	76075	BETA-GALACTOSIDASE
0	14	JUP	62616	HIGHLY SIMILAR TO JUNCTION PLAKOGLOBIN
0	12	GRPEL1	24279	GRPE PROTEIN HOMOLOG 1. MITOCHONDRIAL
0	10	PSMA7	27887	PROTEASOME SUBUNIT ALPHA TYPE-7
0	10	SF3B4	44386	SPLICING FACTOR 3B SUBUNIT 4
0	9	NFS1	50196	CYSTEINE DESULFURASE. MITOCHONDRIAL
0	8	KLHDC2	46099	KELCH DOMAIN-CONTAINING PROTEIN 2
0	8	SETD8	42890	HISTONE-LYSINE N-METHYLTRANSFERASE SETD8
0	8	TADA2A	51496	TRANSCRIPTIONAL ADAPTER 2-ALPHA
0	6	CDK9	42778	CELL DIVISION PROTEIN KINASE 9. CATALYTIC SUBUNIT OF P-TEFb
0	6	HEXIM1	40623	NEGATIVE REGULATOR OF P-TEFb
0	6	PSMD4	40737	26S PROTEASOME NON-ATPASE REGULATORY SUBUNIT 4
0	6	SAPS1	103139	PROTEIN PHOSPHATASE 6 REGULATORY SUBUNIT 1
0	5	KNTC1	250749	KINETOCHORE-ASSOCIATED PROTEIN 1
0	5	PLAUR	36978	UROKINASE PLASMINOGEN ACTIVATOR SURFACE RECEPTOR
0	5	SP1	80693	TRANSCRIPTION FACTOR SP1

ortholog of Rif1 was found to associate only with telomeres with DNA damage (Silverman, et al., 2004). At DNA double-strand breaks, human Rif1 was shown to associate with the DNA damage kinase ATM for the overall protection against DNA damage (Silverman, et al., 2004; Wang, et al., 2009). ATM is a kinase known to target p53 for activation during DNA damage (Shiloh, 2006; Kruse and Gu, 2009), and is also identified in all SREBP, p53 and p65/RelA-Mediator. We have also shown that ATM associates with SREBP-1a-Mediator using a Mediator-specific orthogonal purification with western blotting (Figure 2.5). Future directions should include biochemical validation in terms of an orthogonal purification testing both adapter proteins and Rif1.

Cofactors identified that are exclusive to p65/RelA (Table 4.11) include lysosomal, cytosolic enzymes, and mitochondrial cofactors, along with transcription-related cofactors such as the mRNA splicing factor SF3B4, the histone methyltransferase SETD8, the elongation-associated kinase CDK9 and

its regulator HEXIM1, and the transcription activator SP1. It seems, however, somewhat unlikely that these cofactors generally involved in transcription could be truly specific to p65/RelA, but certainly garner further investigation.

4.3 Discussion

To test the hypothesis that the Mediator complex can regulate protein-protein interactions in an activator-dependent fashion, an improved proteomics platform (Chapter 3) was used to comprehensively characterize the composition of various activator-bound Mediator complexes. Since all activators used in these studies were GST-fusions, each activator can serve as a control for any other activator allowing for a qualitative comparison. Mediator and RNA Pol II specific spectral counts are given and can be compared for each activator in terms of Mediator/Pol II yields. Other Associated cofactors identified exclusively with distinct activators were also assembled from the datasets.

A compelling idea comes from the association of the Beclin3-PIK3C3 subcomplex involved in autophagy (Behrends, et al., 2010) exclusively with SREBP-1c. Interestingly, a recent study comparing the three mammalian SREBPs in a genome-wide ChIP-seq experiment found that SREBP-2 binds preferentially to two different gene-proximal motifs (Seo, et al., 2011). A Gene Ontology (GO) analysis was performed which suggested SREBP-2 targets lipid metabolic processes as expected, but apoptosis and autophagy gene categories were also enriched (Seo, et al., 2011). This is a future direction that could

potentially associate the Mediator complex and transcription regulation with autophagy.

Another worthy future direction is the association of ECM29 (KIAA0368) with SREBP-2. ECM29 has been shown to be an adapter protein that interacts with molecular motor, endosome components and the 26 S Proteasome (Gorbea, et al., 2010). This is interesting considering the fact that molecular motor components, a variety of endosomal components, and many Proteasome components are identified in the mass spectrometry datasets. Other examples of putative protein-protein interaction involving ubiquitin signaling and the Proteasome are also found in the data. There is precedence for ubiquitination of a transcription factors to rapidly turnover the factor after activation to prevent further activation (Salghetti, et al., 2001). The estrogen receptor also requires the 26 S Proteasome for transcriptional activation and subsequent degradation of the receptor (Lonard, et al., 2000).

Finally, a cofactor dramatically enriched in the p53-Mediator complex is Rif1, known to regulate the length of telomeres in yeast. Studies of the human Rif1 ortholog suggest a distinct function in double strand DNA damage with the kinase ATM (Silverman, et al., 2004; Wang, et al., 2009) which is known to target p53. This is also a very interesting potential future direction for this dataset.

4.4 Methods

4.4.1 Activator-Mediator Purifications

SREBP-1c, SREBP-2, and p65/Rel A were a gift from Anders Näär. SREBP-1a(1-50), -1c(1-26), -2(1-49), p53(1-70), and p65/RelA(434-551)-Mediator was purified from HeLa nuclear extract using GST-SREBP-1a (residues 1–50), GST-SREBP-1c (residues 1–26), GST-SREBP-2 (residues 1–49), GST-p53 (residues 1–70) and GST-p65/Rel A (residues 434–551) immobilized to Glutathione-Sepharose beads (GE Life Sciences). After binding, the resin was washed five times with 10 column volumes (CV) 0.5 M KCl HEGN (20 mM Hepes, pH 7.6; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40 alternative) and one time with 10 CV 0.15 M KCl HEGN (0.02% NP-40 alternative). Bound proteins were eluted with 30 mM GSH in elution buffer (80 mM Tris, 0.1 mM EDTA, 10% Glycerol, 0.02% NP-40, 100 mM KCl) and applied to a 15% to 40% linear glycerol gradient (in 0.15 M KCl HEG) and centrifuged for 6 h at 55,000 rpm. Mediator-containing fractions (>1.0 MDa) were combined for analysis.

4.4.2 Sample Preparation and Proteomics Analysis

Purified Mediator complex-containing (~1~10ug) fractions were precipitated at 4°C using 20%(w/v) TCA, 0.067mg/mL insulin and 0.067%(w/v) deoxycholate. Precipitated protein pellets were washed twice with -20°C Acetone and air dried. Proteins were trypsin digested using a slightly modified Filter-Aided Sample Prep (FASP) protocol (Wiśniewski, et al., 2009). Briefly,

protein pellets were suspended with 4%(v/v) SDS, 0.1M Tris pH 8.5, 10mM TCEP and incubated 30m ambient to reduce disulfides. Reduced proteins were diluted with 8M Urea, 0.1M Tris pH8.5 and iodoacetamide was added to 10mM and incubated 30m in total darkness. Reduced and alkylated proteins were then transferred to a Microcon YM-30 spin concentrator and washed twice with 8M Urea, 0.1M Tris pH 8.5 to remove SDS. Three washes with 2M Urea, 0.1M Tris pH8.5 were performed then trypsin and 2mM CaCl₂ were added and incubated approximately 2 hours in a 37°C water bath. Digested peptides were eluted and acidified with 5%(v/v) formic acid. Peptides were desalted online and fractionated with a Phenomenex Jupiter C18 (5µm 300Å; 0.25 x 150mm) column using a two dimensional LC/MS/MS method (Agilent 1100). Seven steps of increasing acetonitrile (3, 6, 9, 12, 16, 20 and 100%B; A: 20mM ammonium formate pH10, 4% acetonitrile and B: 10mM ammonium formate pH10, 65% acetonitrile) at 5µL/minute eluted peptides for a second dimension analysis on a Dionex Acclaim PepMap C18 (3µm 100Å; 0.075 x 150mm) running a gradient at 0.2µL/minute from 5 to 25% B in 100 minutes for steps one through six and 10 to 30% B in 100 minutes for step seven (A: 4% acetonitrile & B: 80% acetonitrile, both with 0.1% formic acid pH~2.5). PepMap eluted peptides were detected with an Agilent MSD Trap XCT (3D ion trap) mass spectrometer.

All spectra were searched with Mascot v2.2 (Matrix Sciences) against the International Protein Index (IPI) database version 3.65 with two missed cleavages and mass tolerances of m/z ±2.0 Da for parent masses and ±0.8 Da for MS/MS fragment masses. Peptides were accepted above a Mascot ion score

corresponding to a 1% false discovery rate (1% FDR) determined by a separate search of a reversed IPI v3.65 database. Peptides were then filtered and protein identifications were assembled using in-house software as described (Meyer-Arendt, et al., 2011; Resing, et al., 2004).

Chapter 5

Future Directions

RNA Polymerase II Activity is regulated by post-translational modifications on the rpb1 C-Terminal Domain

5.1 Introduction

The DNA-dependent RNA Polymerase II (Pol II) is tasked with gene expression of all protein-coding genes as well as some non-coding genes. Pol II is made up of 12 subunits with the largest, rpb1 at 220kD and the smallest at 7kD. It is unique among DNA-dependent polymerases in that the carboxyl-terminal domain (CTD) of rpb1 consists of approximately 40kD of the repeating consensus sequence YSPTSPS. Not all repeats are consensus, however, with a two arginines and seven lysines replacing serine at the consensus position 7 in the last 20kD of the C-terminus. This carboxyl-terminal sequence of rpb1 will be referred to as the CTD. The CTD appears to be a target of signaling events where phosphorylation of the repeating YSPTSPS correlates with transcription initiation and elongation. The CTD is a large domain for protein-protein interactions and it has been shown to bind the Mediator complex (Naar, et al., 2002), the Integrator complex (Baillat, et al., 2005), mRNA capping, splicing and termination cofactors regulating the overall processing of a nascent mRNA (Perales and Bentley, 2009). Many of these cofactors may bind a

phosphorylation-pattern-specific state of the CTD. Antibodies have been used to show the presence of phosphorylation but very little is known regarding how many sites on the CTD are phosphorylated, or whether there is any pattern of phosphorylation on the CTD. Antibodies to any phosphorylation state of the CTD do not convey this information. Many kinases have been shown to phosphorylate the CTD. The two main kinases relevant to the transcription cycle are CDK7 and CDK9. CDK7 is a component of the general transcription factor II H (TFIIH) complex, and CDK9 is part of the positive transcription elongation factor b (P-TEFb). Both of these kinases are purified and used in in vitro kinase reactions to phosphorylate a purified GST-fusion of the ~40kD 52-repeat rpb1 CTD. This phosphorylated CTD (pCTD) is a very useful reagent for identifying new protein-protein interactions specific to the pCTD, as well as for the development of a methodology to identify any site-specific pattern of phosphorylation on the CTD.

5.1.1 The Transcription Cycle

Many protein-protein interactions must be highly coordinated in a spatial and temporal fashion during the transcription cycle. The minimum composition of factors required appears to be the general transcription machinery: TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, TFIIH, The Mediator complex and RNA Pol II (Taatjes, 2010). This is often referred to as the pre-initiation complex (PIC). Transcription initiation is followed by elongation, though the rate can vary widely depending on

the promoter and stimulation. This transition is becoming a new focus for transcription regulation. The CTD is phosphorylated during this transcription cycle at the promoter of genes by TFIIH which is also tasked with promoter melting in preparation for elongation. TFIIH phosphorylates serine five and seven of the consensus repeats YSPT**SPS**. At this point it is thought Pol II is ready for elongation, so this protein assembly is really more accurately a pre-elongation complex (PEC) (Taatjes, 2010). Additional phosphorylation events then occur where P-TEFb phosphorylates the CTD on serine two of the repeats Y**S**PTSPS as well as other elongation factors, such as NELF and DSIF (SPT4 and SPT5) allowing for productive elongation. It is thought that the pCTD electrostatically repels Mediator and the PEC to assist in departure from the promoter (Sogaard and Svejstrup, 2007 and Kim, et al., 1994). The CTD is often thought of as a large binding platform, where phosphorylation can dramatically change the sequence landscape regulating protein-protein interactions. Of course what goes on must come off, so there are phosphatases that must remove the phosphate groups to 'reset' the Pol II for additional transcription events and the cycle can then start over.

5.1.2 Cotranscriptional mRNA Processing and the CTD code

Concurrent with the transcription cycle is the processing of the nascent RNA as well as passage through chromatin. Many cofactors and protein complexes have been demonstrated to be involved. The mature mRNA must be

capped, spliced together and polyadenylated/terminated then prepared for export out of the nucleus. At the same time, chromatin must be cleared by chromatin remodeling factors. The CTD provides a flexible and reasonably large binding platform to coordinate assembly of chromatin remodeling complexes, the spliceosome, and the Integrator complex, which has been shown to process small nuclear RNA (snRNA). The exact CTD or pCTD substrates for these protein-protein interactions are unknown.

Here we undertake an experiment to comprehensively characterize the HeLa nuclear extract interactome of the CTD in distinct phosphorylation states: (1) unphosphorylated, (2) phosphorylated serine five only (TFIIH only), (3) phosphorylated serine two only (P-TEFb only), and (4) phosphorylated serine five and two (TFIIH and P-TEFb both). We identify known CTD-interacting protein complexes: the Integrator and the Mediator complexes. Many other protein cofactors and likely complexes are also identified.

To address the question of a phosphorylation pattern on the CTD, very preliminary data will be presented. The repeating sequence and lack of protease cleavage sites make traditional mass spectrometry incapable of sequencing more than half of the CTD. This strategy is based on chemical biology, “middle-down” proteomics using a supercharging reagent to enhance ionization and charge density along with all available gas phase fragmentation and fractionation methods.

5.2 Results

A proteomics platform was applied to characterize phosphorylation-specific CTD-interacting cofactors. Approximately 35~50 ug GST-CTD was phosphorylated by (1) TFIIH, (2) P-TEFb, and (3) TFIIH and P-TEFb both.

It was expected that Mediator would only interact with the unphosphorylated CTD or at least there would be much less depending on the actual amount of phosphorylation on any given CTD molecule. This, however, was not the case. In fact, there appears to be as much or more Mediator in any of the pCTD as the unphosphorylated control CTD sample, suggesting that perhaps phosphorylation alone is not the dissociating action for Mediator and the CTD.

Another interesting result was identification of the Integrator complex in both the CTD and pCTD samples. It was reported that serine seven must be phosphorylated for recruitment of Integrator (Egloff, et al., 2007), but Integrator was identified in all pCTD regardless of kinase. So this leaves the questions to what regulates the recruitment of either Mediator or Integrator, or is both able to interact simultaneously?

5.2.1 CTD Substrate and Kinases

To obtain a sufficient amount of phosphorylated CTD to perform experiments, an expressed fusion protein GST-rpb1-CTD is purified (Figure 5.1).

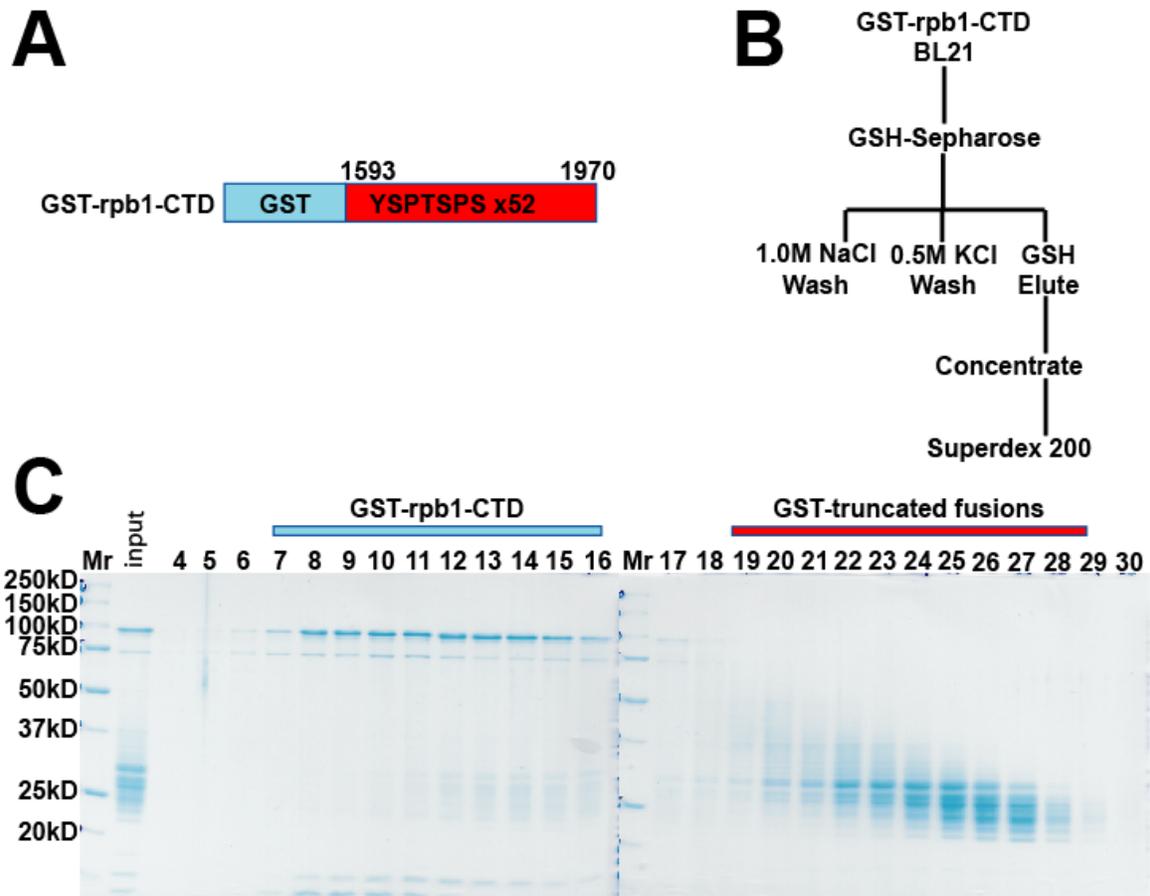


Figure 5.1 Purification of GST-rpb1-CTD. (A) The fusion protein GST-rpb1-CTD used in this study. (B) Purification scheme for GST-rpb1-CTD. (C) Coomassie stain acrylamide gel of Superdex 200 fractions.

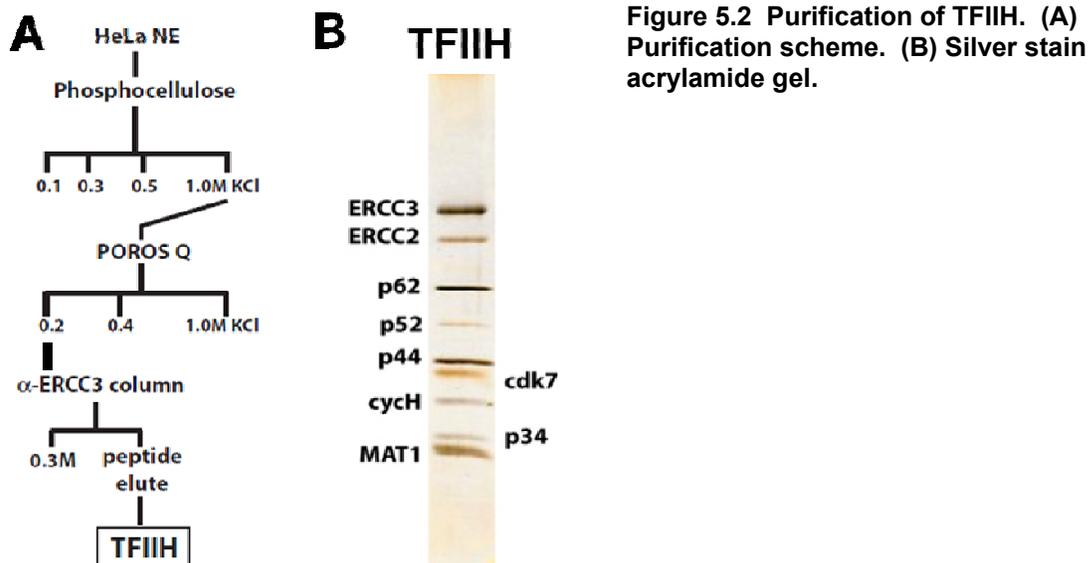


Figure 5.2 Purification of TFIID. (A) Purification scheme. (B) Silver stain acrylamide gel.

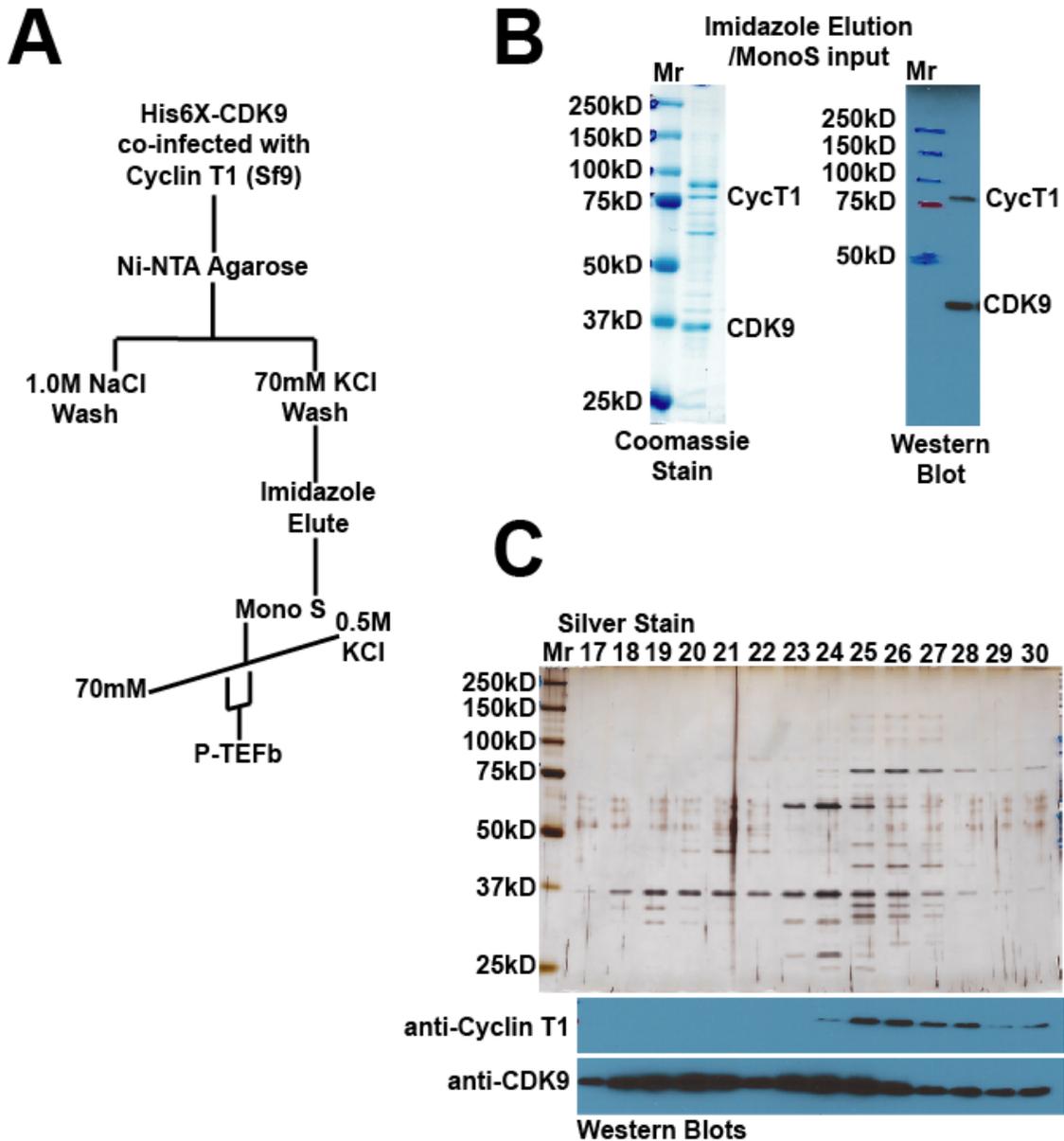


Figure 5.3 Purification of P-TEFb. (A) Purification scheme. (B) Coomassie stained acrylamide of elution from Ni-NTA purification and westerns for cyclin T1 and CDK9. (C) Mono S fractions of P-TEFb.

A two-step purification is employed from BL21 lysates, with a GSH-Sepharose column followed by a Superdex 200 gel filtration column. The gel filtration column effectively removes the excess truncated fusion proteins that are characteristic of GST-CTD expressions. The early gel filtration fractions contain

very pure GST-CTD for in vitro kinase reactions which can then be immobilized again.

The two major CTD kinases were also purified. TFIIH is purified in the lab from P1.0M/Qft/anti-ERCC3 (TFIIH) affinity column (Figure 5.2). P-TEFb is purified from recombinant Sf9 cells co-infected with His6X-CDK9 and Cyclin T1 (Figure 5.3). Sf9 lysates were incubated in batch with Ni-NTA agarose, washed and eluted with imidazole (Figure 5.3. A and B). The imidazole elution was then loaded onto a MonoS for a gradient elution (Figure 5.3.C). P-TEFb-containing fractions were used in kinase reactions with GST-CTD.

5.2.2 The RNA Pol II CTD and Phospho-CTD Interactomes

Kinase reactions were performed with 35~50 ug purified GST-CTD, TFIIH and P-TEFb. Figure 5.4 shows a typical timecourse to an hour at 37°C with both TFIIH and P-TEFb. The amount of phosphorylation can clearly be seen in this silver stained acrylamide gel. The GST-CTD runs at approximately 85kD. As the CTD become more and more phosphorylated the band smears to slower and slower migrating bands with a finite band where apparently no more phosphorylation takes place at approximately 125kD. So neither the purified TFIIH nor the P-TEFb can completely phosphorylate this substrate GST-CTD in solution with purified components only. These kinase reaction were performed at the same scale in 18X replicates for each TFIIH only, P-TEFb only, and TFIIH and P-TEFb both, along with an unphosphorylated CTD control (Figure 5.5).

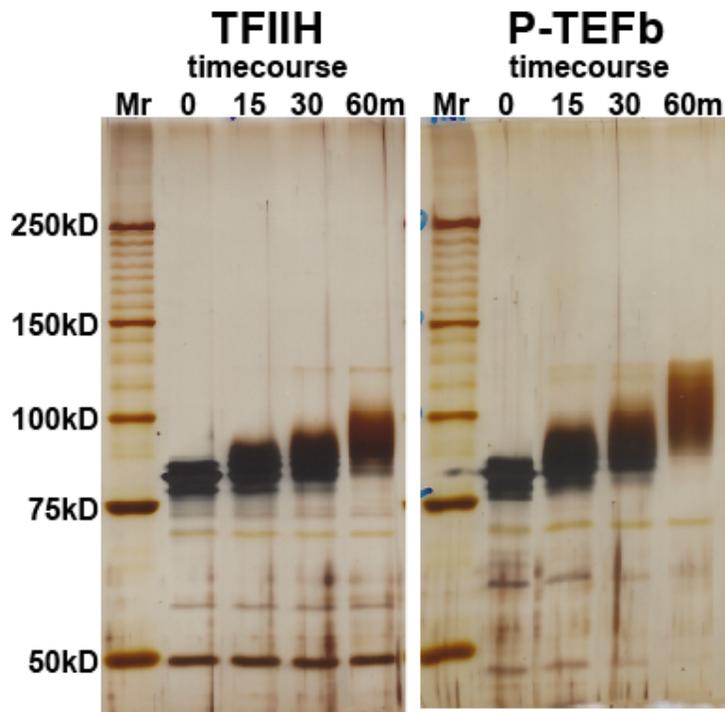


Figure 5.4 Kinase reaction timecourse with GST-CTD and TFIIH or P-TEFb. Silver stained acrylamide gel of T0, T15m, T30m and T60m timecourse.

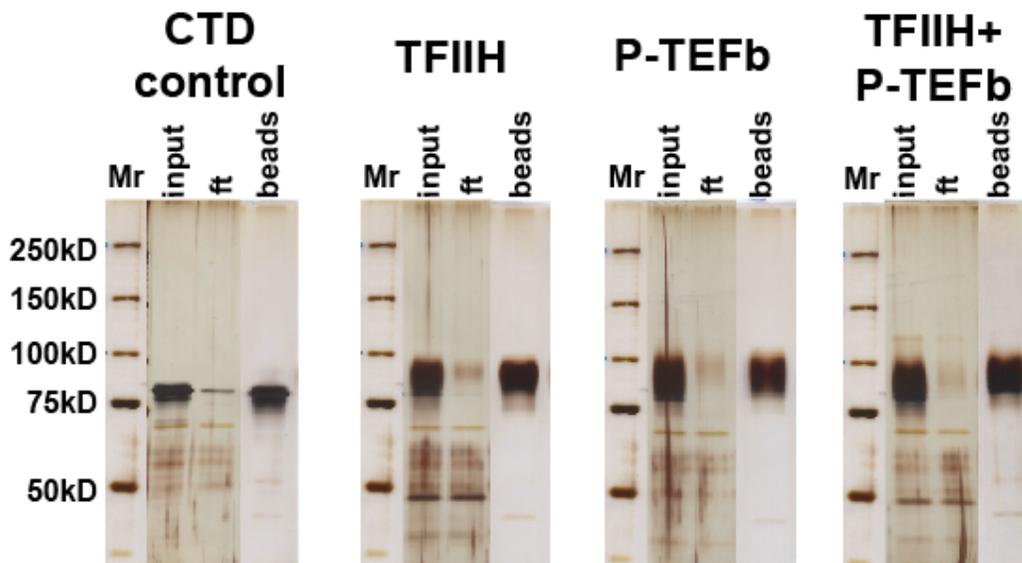


Figure 5.5 Immobilization of kinase reactions with GSH-Sepharose and beads after Sarkosyl elution to verify fusion protein phosphorylation throughout the nuclear extract pulldowns.

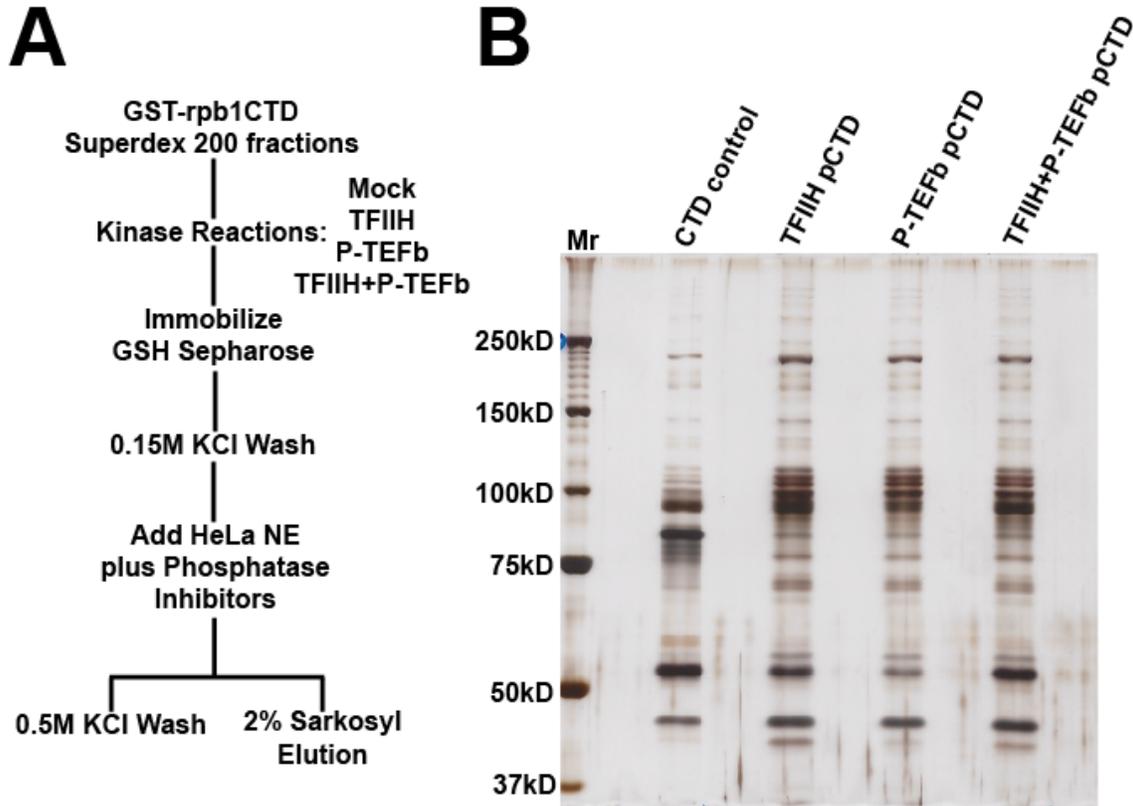


Figure 5.6 Purification of phosphorylated CTD-binding proteins. (A) Purification scheme. (B) Silver stained acrylamide gel of CTD-interacting proteins and phosphorylated CTD-interacting proteins.

The kinase reactions were passed over a GSH-Sepharose column (Figure 5.6.A). The input and flow through of each kinase reaction is shown in Figure 5.5. The immobilized GST-CTD and GST-pCTD were washed with 0.15M KCl buffer to remove the kinase reaction components. The buffer was removed and cleared HeLa nuclear extract supplemented with phosphatase inhibitors was added and mixed at 4°C for 3 hours. The pull-downs were washed with 0.5M KCl buffer and eluted with 2% (w/v) sarkosyl for mass spectrometry analysis. The remaining beads, with most of the fusion protein still immobilized, were then eluted with Laemli sample loading buffer (Figure 5.5 “BEADS”) to show that the amount of phosphorylation had not significantly changed throughout the pull-down.

The sarkosyl elutions (Figure 5.6.B) were prepared for and analyzed by mass spectrometry. The total number of proteins identified in this analysis is given in Table 5.1, along with the totals for two major complexes that are known to bind the CTD, Mediator and Integrator.

Every effort was made match the CTD and pCTD pulldown samples in terms of amount of bait and nuclear extract input. And yet more proteins were identified in each of the pCTD compared to the CTD control. In fact, as much as a third more total proteins were identified in the TFIIH pCTD sample as the CTD control (Table 5.1). This is not unexpected if phosphorylation of the CTD is meant to induce the recruitment of cofactors that do not bind the unphosphorylated CTD. What is unexpected is the number of peptides identified for the Mediator (Tables 5.1 and 5.2) and Integrator (Tables 5.1 and 5.3) complexes in the pCTD samples.

If phosphorylation of the CTD disrupts the interaction with Mediator, then why is there so much Mediator in all of the pCTD samples? Certainly, not all of the available sites are phosphorylated and it may only require two to four repeats

Table 5.1 Total identifications for CTD/pCTD-interactome.

	CTD control	TFIIH pCTD	P-TEFb pCTD	TFIIH+P-TEFb pCTD
total peptides	4073	8013	5628	6346
total proteins	366	757	612	660
total MEDIATOR peptides	214	365	305	300
total MEDIATOR proteins	20	26	24	25
total Integrator peptides	532	1042	1109	1175
total Integrator proteins	12	12	12	12

Table 5.2 Mediator identifications for CTD/pCTD.

Control CTD	PTEFb pCTD	TFIIHpCTD pCTD	TFIIH+PTEFb pCTD	Gene	Prot MW
41	70	56	54	MED1	168478
6	7	17	15	MED4	29745
0	0	1	0	MED6	29298
11	8	8	4	MED7	27245
9	12	10	8	MED8	32819
0	10	0	0	MED9	16403
4	1	2	5	MED10	15688
0	1	4	4	MED11	13129
0	0	0	1	MED12	247334
31	26	42	26	MED14	160607
12	31	37	23	MED15	86753
15	12	12	20	MED16	96793
7	24	35	23	MED17	72876
0	0	0	1	MED18	24453
1	4	5	4	MED19	26273
9	10	20	6	MED20	23222
7	9	9	9	MED21	15564
0	2	7	10	MED22	16480
4	5	5	18	MED23	156194
21	24	27	30	MED24	110305
0	1	16	0	MED25	84389
7	20	9	12	MED26	65446
18	9	20	7	MED27	35432
0	0	3	2	MED28	19520
1	12	6	9	MED29	23473
4	7	11	9	MED30	20277
6	0	3	0	MED31	15805

to bind the Mediator, however, less Mediator would still be expected if the pCTD did not bind Mediator.

Another unexpected result was the amount Integrator complex in the CTD sample as well as all pCTD samples. There appears to be an enrichment of Integrator in the pCTD compare with the CTD control (Table 5.3), however, there are still more than twice the spectral counts for Integrator than the comparably

Table 5.3 Integrator identifications for CTD/pCTD.

Control CTD	PTEFb pCTD	TFIIHpCTD pCTD	TFIIH+PTEFb pCTD	Gene	Prot MW
95	195	235	202	INTS1	244297
47	95	73	105	INTS2	134346
51	152	165	228	INTS3	118013
60	151	120	119	INTS4	108171
22	53	41	52	INTS5	107995
107	150	114	175	INTS6	100390
93	123	132	146	INTS7	106834
33	67	58	44	INTS8	113088
28	48	33	42	INTS9	73814
4	45	48	43	INTS10	82236
11	53	27	42	CPSF3L	67663
10	13	22	28	INTS12	48808

sized Mediator complex in the CTD control. So phosphorylation does not appear to be required for the GST-CTD to bind the Integrator complex in nuclear extract.

An interesting internal control was found with the factor Phosphorylated CTD-Interacting Factor 1 (PCIF1). PCIF1 was identified as a pCTD-interacting factor that did not bind the unphosphorylated CTD (Fan, et al. 2003) and the data presented here (Table 5.4) confirms this. The spectral counts for PCIF1 are fairly well matched for each pCTD sample.

Table 5.4 PCIF1 is an Internal control identified only in phosphorylated CTD pulldowns.

Control CTD	PTEFb pCTD	TFIIH+ PTEFb pCTD	TFIIH+ PTEFb pCTD	Gene	Prot MW	Protein descriptor
0	20	23	26	PCIF1	81351	PHOSPHORYLATED CTD-INTERACTING FACTOR1

To further validate the Mediator association with the pCTD, the samples prepared for mass spectrometry were probed by western blotting for MED1 and MED23 (Figure 5.7). And in fact, the western blot did confirm the mass

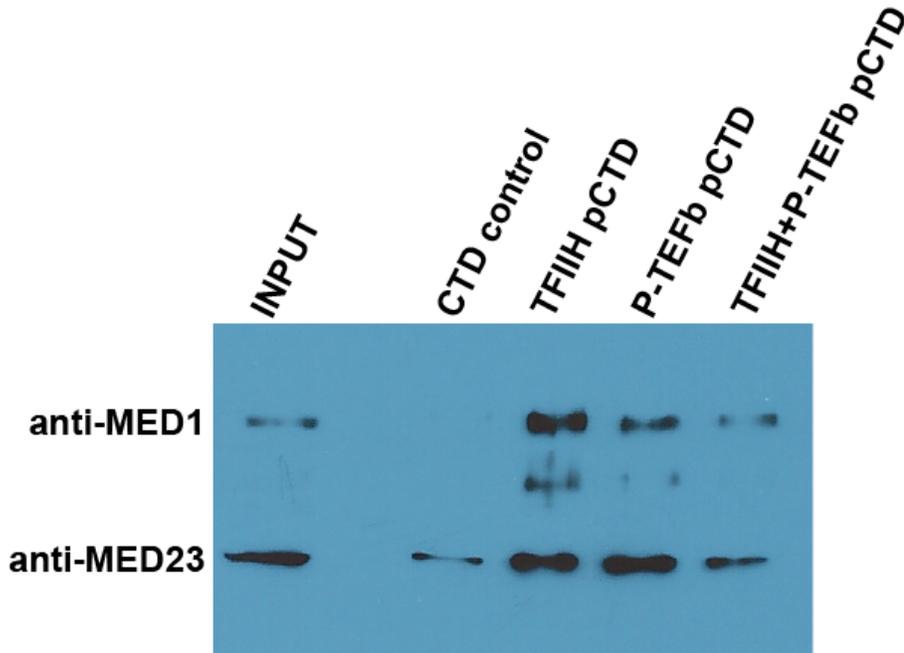


Figure 5.7 Western blot probing MED1 and MED23 in CTD/pCTD elutions.

spectrometry data with more MED1 and MED23 in the pCTD samples relative to the CTD control.

To test whether the pCTD-Mediator interaction was through accessory cofactors and not direct as it was with the unphosphorylated CTD, the input for purifying Mediator was put over two chromatography columns, a phosphocellulose P11 followed by a Poros Q column for an enriched Mediator fraction (Figure 5.8.A). The bait was unphosphorylated CTD and CTD phosphorylated by both TFIIH and P-TEFb. Mediator was bound to both CTD and pCTD from the P1.0M/Q1.0M fraction (Figure 5.8.B). These CTD/pCTD purified Mediator fractions were applied to a Superose 6 gel filtration column to isolate the CTD/pCTD-bound Mediator complexes. However, no Mediator eluted from the Superose 6 column in any experiment. Two possible explanations for

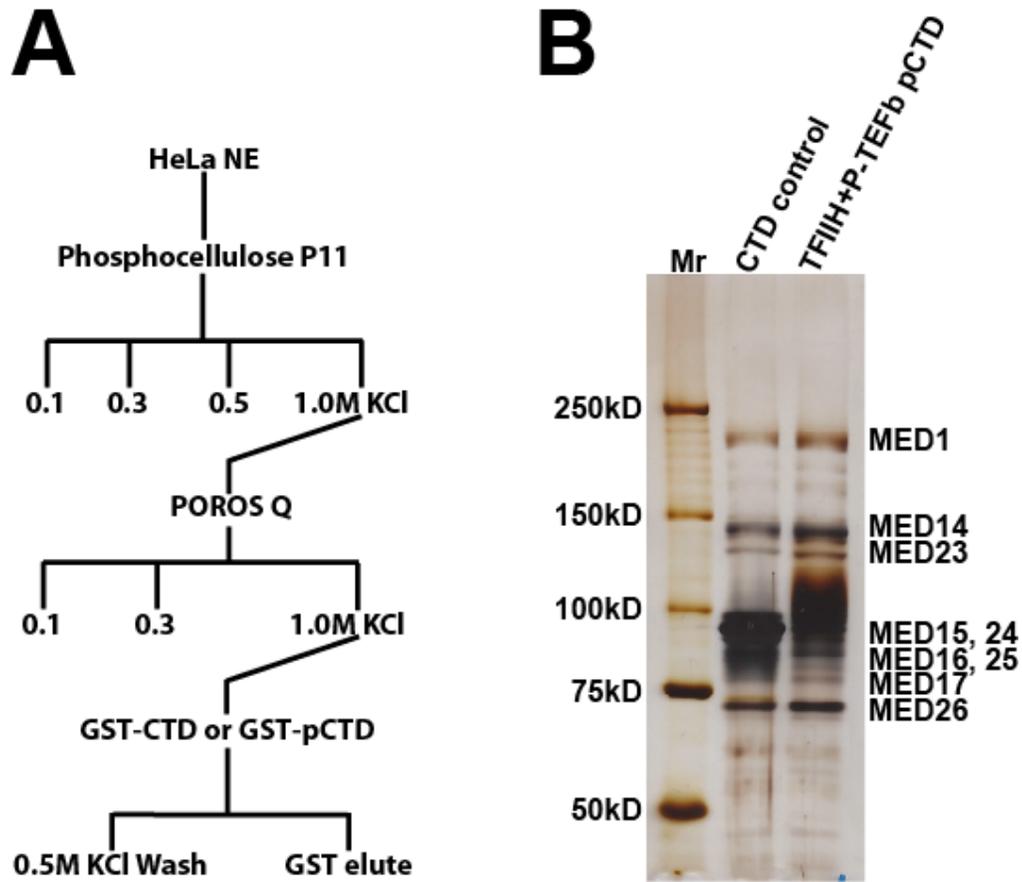


Figure 5.8 CTD/pCTD pulldowns from Mediator-enriched fraction, P1.0M/Q1.0M. (A) Purification scheme. (B) Silver stained acrylamide of CTD/pCTD elutions.

this phenomena are (1) the Mediator complex broke up inside the column (though no subunits were identified in their molecular weight range), or (2) the Mediator complex bound indefinitely to the resin at the salt concentrations used (0.15M KCl).

5.2.3 A Strategy for the Identification of Site-Specific Phosphorylation on the RNA Pol II CTD

A traditional “bottom-up” experiment, such as those offered in Chapters 2, 3 and 4, uses the protease trypsin to reduce target proteins into peptides that can be sequenced by a mass spectrometer. The problem with the CTD is that only half of it has any trypsin-substrate lysines or arginines, the distal half, and the other half, the proximal half, is over 21kD. This is far too large for traditional collision-induced dissociation (CID) gas-phase fragmentation. It is possible, however, with a high resolution mass spectrometer that a 21kD polypeptide could be analyzed in a “middle-down” experiment (Siuti and Kelleher, 2007). Another problem, however, is that a 21kD polypeptide with few basic residues can be a challenge to ionize and get into the gas phase. These are difficult analytical problems that likely have hindered the successful sequencing of the repeating heptad sequence of the CTD using mass spectrometry. Further complicating the issue is that the kinases required for phosphorylating the CTD are difficult reagents to obtain. Unfortunately, purifying kinases and the GST fusion protein and performing kinase reactions to generate phosphorylated CTD is the easiest part of characterizing a pattern of phosphorylation on the CTD.

A strategy for sequencing the repeats of the CTD to identify site-specific modifications and preliminary data will be presented here. The strategy involves:

1. Peptide mapping and hplc purification;
2. Chemical biology (beta-Elimination/Michael Addition);
3. Supercharging reagent to enhance ionization;
4. All available gas-phase fragmentation (ETD, HCD, CID);

All aspects of this strategy work together to provide a solid methodology to tackle this intensive analytical problem that is of great relevance to understanding molecular mechanisms of gene activation. The hplc purification and peptide mapping of CTD polypeptides simplify the downstream analyses and allow for greater flexibility in analyses. A supercharging reagent (Iavarone, et al., 2001) has been tested with great success in increasing the charge states observed for the 21kD polypeptide, though fragmentation efficiency did not benefit as greatly. New technology is available to fragment polypeptides in the gas phase, such as electron transfer dissociation (ETD) and higher-energy collision dissociation (HCD) which may provide high sequencing coverage. However, ETD works better with a higher charge density (Syka, et al., 2004). Phosphorylation adds negative charge neutralizing the positive charge on the CTD. If the negatively charged phosphorylation could be converted to a positively charged group, then the charge density would be increased, likely improving the ETD sequencing efficiency. This chemistry is called beta-elimination and Michael Addition. The phosphate is removed by barium catalyzed beta-elimination leaving a dehydroalanine which can be reacted with the nucleophile amino-ethyl thiol (AET) to leave a lysine analog that can be proteolyzed by trypsin or Lys-C (Figure 5.9) (Knight, et al., 2003; Rusnak, et al., 2004). Trypsin would only

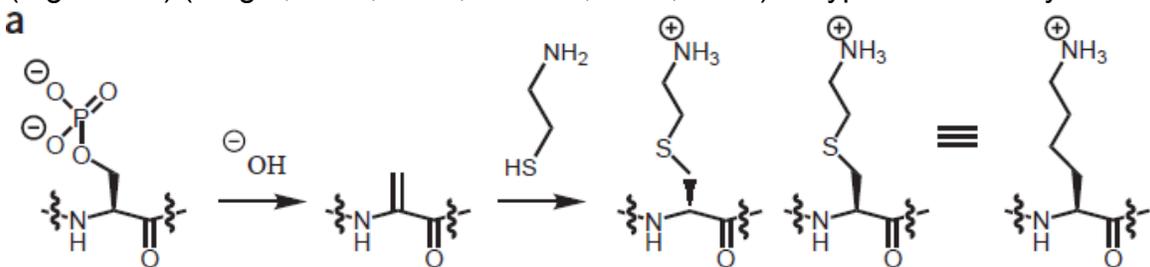


Figure 5.9 (Knight, et al., 2003) Beta-elimination and Michael addition of amino-ethyl thiol (AET) converts phospho-serine to lysine analog that is cleavable using Lys-C endoproteinase.

cleave the position seven with greater efficiency since positions two and five are followed by prolines, which block trypsin access to the active site. Lys-C can be used to cleave through adjacent prolines where a phospho-serine would allow for peptide mapping revealing the position of phosphorylation again with peptide mapping.

Peptide mapping experiments were performed with trypsin, Lys-C and Arg-C. Trypsin gives the most number of CTD peptides, Lys-C gives one fewer and Arg-C gives just two, a proximal and distal CTD polypeptide. The GST-CTD (pGEX-4T3) fusion protein has a thrombin cleavage site, so the kinase reactions are performed in solution, then the reactions were precipitated with 1% formic acid and 80%(v/v) Acetone to remove reaction components. The pellets were suspended with 8M urea, diluted to 2M and thrombin digested for 15 minutes at ambient then loaded onto a reversed-phase C18 hplc column where the ~40kD CTD was collected. The CTD polypeptide was lyophilized then suspended in 0.1M Tris pH 8.5 for trypsin digestion. The tryptic peptides were again loaded onto a reversed-phase C18 hplc column (Figure 5.10). Each CTD peak was manually collected and analyzed using MALDI-TOF-MS to give parent masses for each peptide (Figure 5.10). The first CTD repeat was lost during the thrombin cleavage and CTD, as it appears thrombin also uses that sequence as a substrate. However, all other expected CTD peptides (~49 of 52 repeats) were identified in this fashion. There are two tryptic peptides that are a single repeat that are too hydrophilic and do not retain on the rpC18 column. The asterisks are peptides that belong to trypsin. The next experiment was to perform kinase

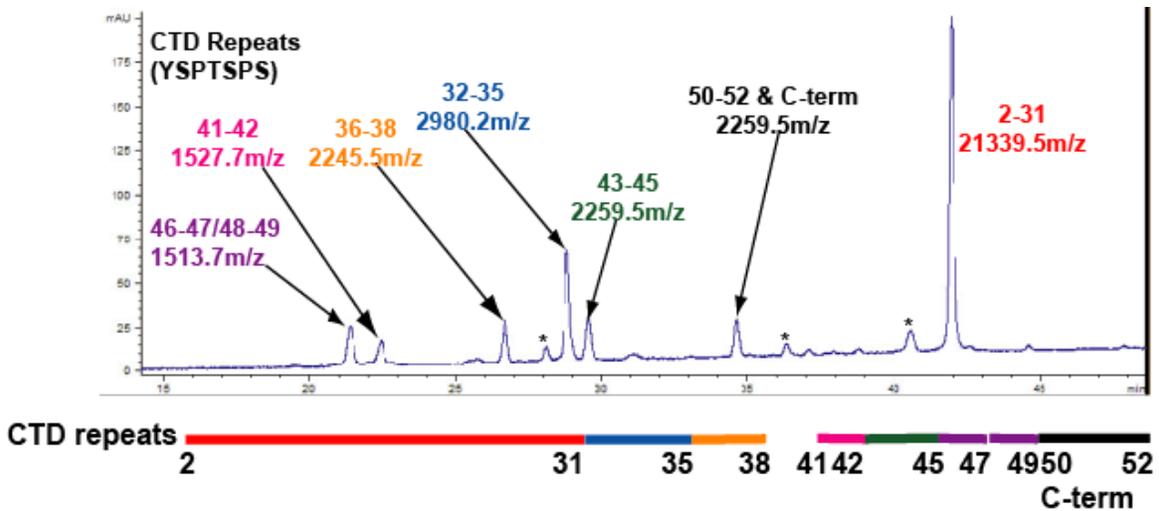


Figure 5.10 Peptide mapping the CTD. Reversed-phase C18 UV 215nm trace of tryptic CTD peptides. Single repeats (two in the CTD) are not retained. (* are peptides of trypsin)

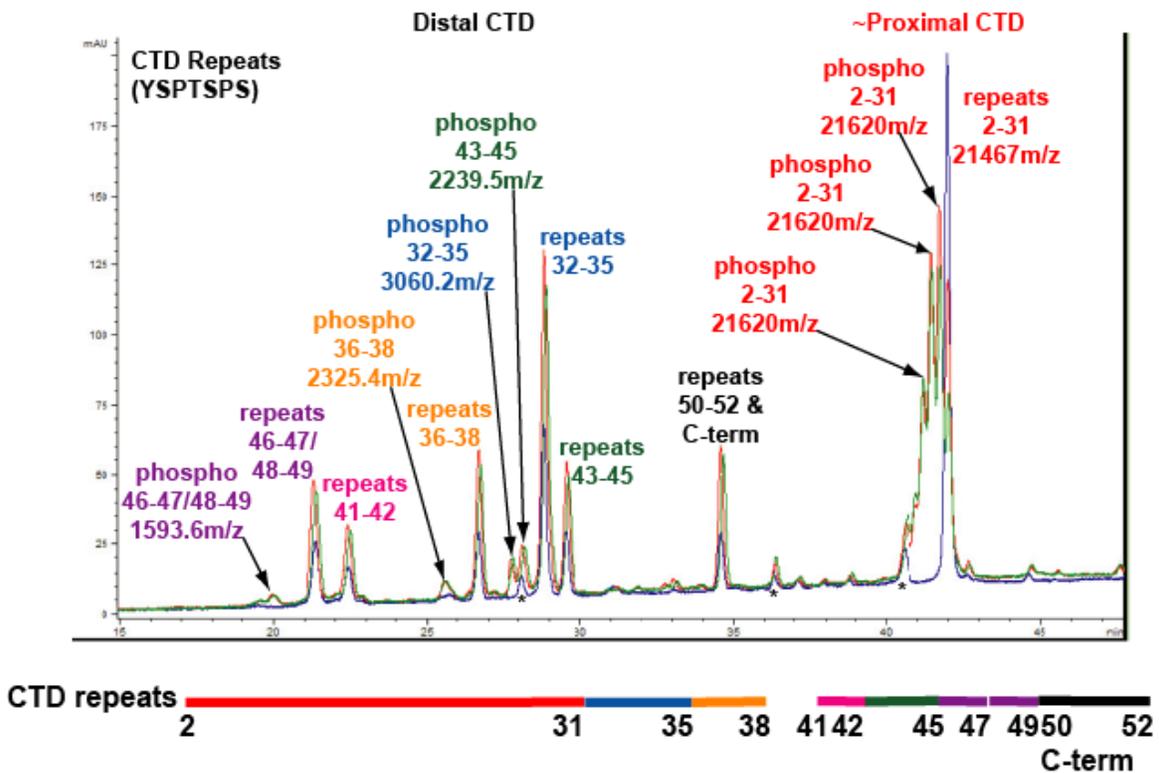


Figure 5.11 Peptide mapping the pCTD. Reversed-phase C18 UV 215nm trace of tryptic CTD and pCTD peptides. Addition of TFIIH kinase prior to trypsin digestion (red) and twice the TFIIH (green). Single repeats (two in the CTD) are not retained. (* are peptides of trypsin)

reactions with the GST-CTD and TFIIH, such as in Figure 5.4. Two titrations (1X and 2X) of TFIIH kinase were added to GST-CTD and incubated at 37°C for maximum kinase activity. The experiment was repeated with Thrombin digestion of the fusion protein, C18 purification of the ~40kD pCTD and trypsin digestion followed by another C18 column (Figure 5.11). Again, all peaks were collected and analyzed using MALDI-TOF-MS for parent masses and peptide mapping. There are only small peaks that rise up in the kinase reaction samples which have masses consistent with phosphorylated CTD peptides, but there is no complete conversion from unphosphorylated to completely phosphorylated in these TFIIH experiments.

The phosphorylated CTD peptides collected in Figure 5.11 were then used for subsequent experiments to develop a beta-elimination/Michael addition protocol (Figure 5.9). There is a fairly large body of literature dealing with beta-elimination of phosphorylated serine and threonine residues, however, two groups (Shokat and Hathaway) have applied this chemistry with a nucleophile for the Michael Addition that is compatible with trypsin and Lys-C active sites (Knight, et al., 2003; Rusnak, et al., 2002 & 2004). Development experiments began with conditions offered as optimal for the model proteins used in these studies. In the methods of Knight, et al., 2003, pTP and pSP sequences (the CTD is Y**SPTSPS**) required 2h at 37°C as opposed to 1h at ambient for all other peptides. The beta-elimination occurs at high pH, which can hydrolyze peptides. Therefore, conditions must be established that optimize beta-elimination/Michael Addition reaction completion and minimize peptide degradation. Critical parameters are

temperature, time, solvent, basicity and the order of addition. These are many parameters to vary with very little pCTD substrate. Therefore, conditions were tested as published (Knight, et al., 2003) with limited success. Several experiments were performed with no products at all. Upon further investigation in the literature, intramolecular linkage are common in beta-elimination reactions and that the hydrophobicity of the solvent can be a critical parameter for pTP and pSP sequences (Tinette, et al., 2007). Therefore, the hydrophobicity of the reactions was increased with the addition of 50% (v/v) acetonitrile, and since no products had been identified, a much shorter timepoint was tested. Product peaks were discovered at 5m that were gone at 10m, so other reactions were clearly taking place leaving no desired products. Finally, an experiment testing 5m beta-elimination and 10m sequential Michael Addition provided a product peak with a mass consistent with a modified CTD peptide (Figure 5.12). This experiment was run on an LC/MS instead of the UV-hplc, which allowed for partial MS/MS sequencing. The sequencing data is also consistent with the correct Michael Addition product for the two repeat pCTD peptide used for this experiment (Figure 5.13). This reaction did not go to completion, however, it was the first experiment to successfully provide the correct product or any product for that matter. Further development experiments will be required to optimize the beta-elimination/Michael addition for the unique repeating, serine/proline-containing sequence of the CTD.

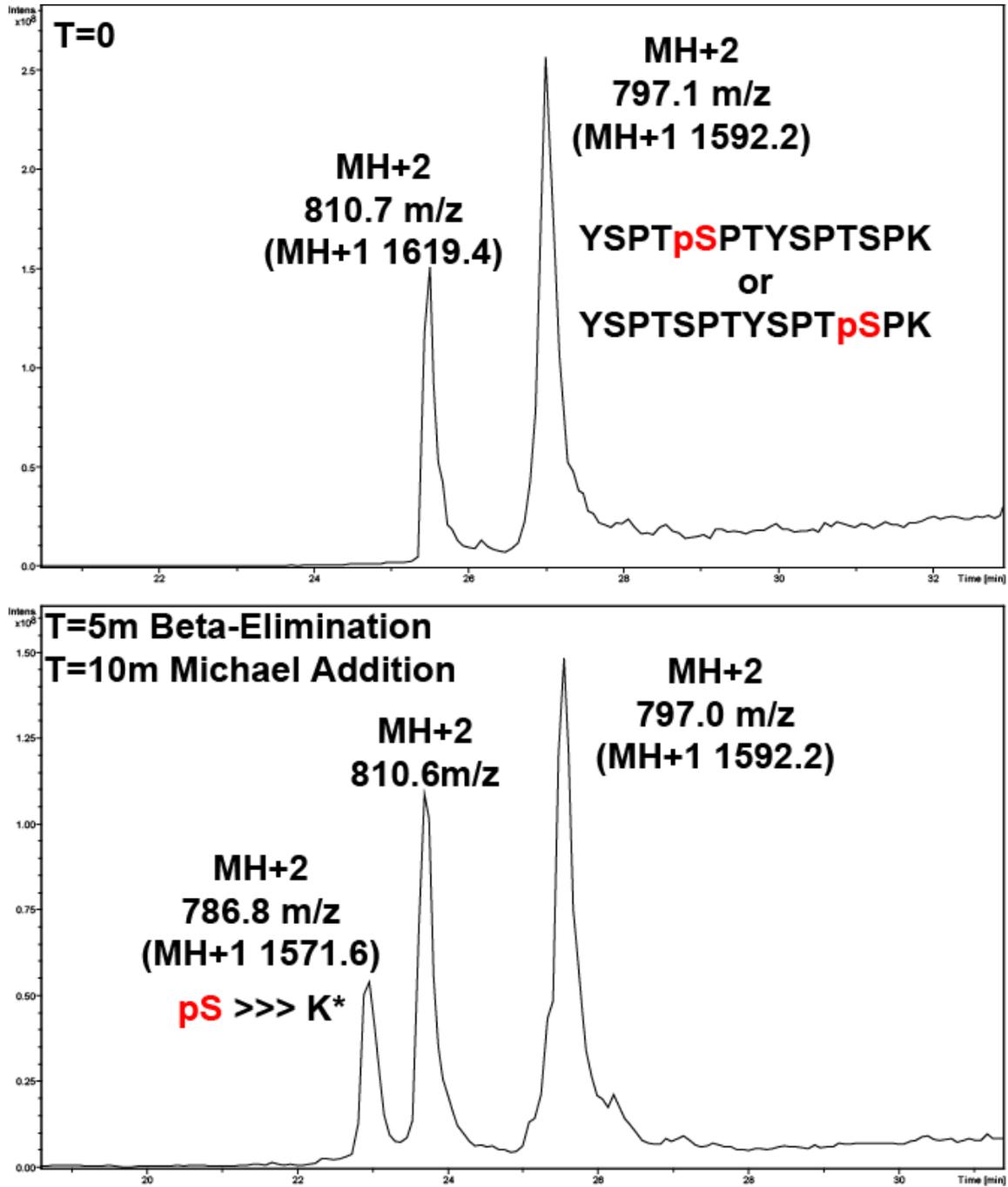


Figure 5.12 LC/MS total ion count (TIC) chromatograms of Beta-elimination/Michael addition of 2X CTD repeats with a single phosphorylation. The 810m/z peak is thought to be a formylation product not resolved on the first column (Figure 5.10).

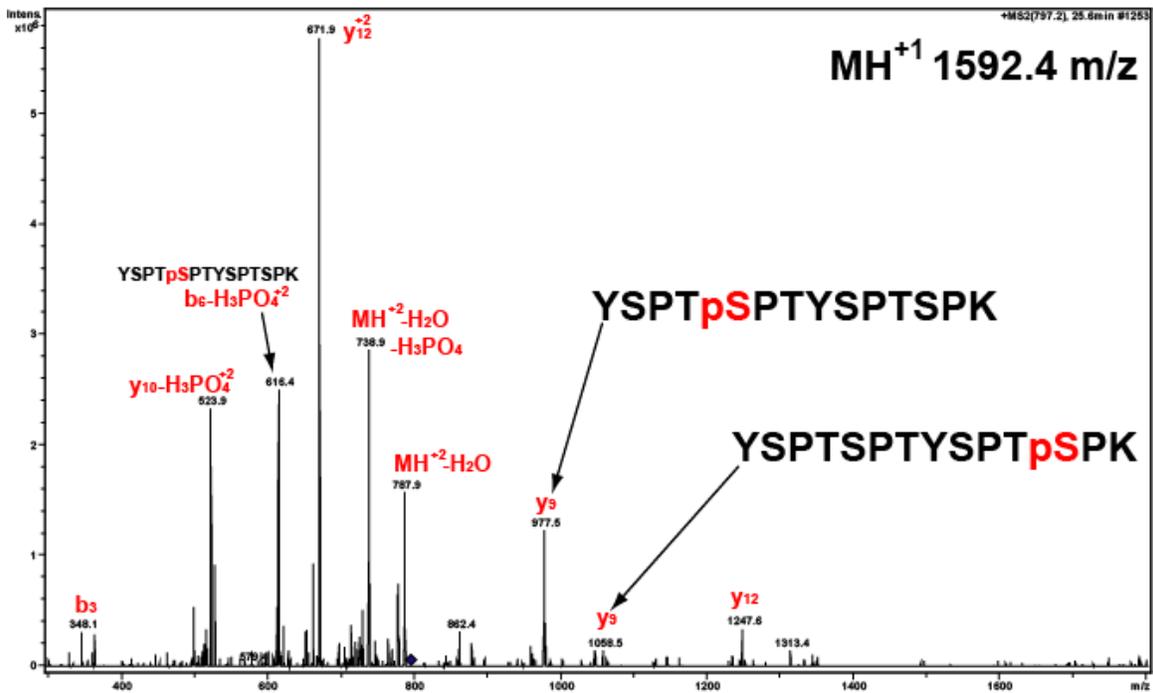
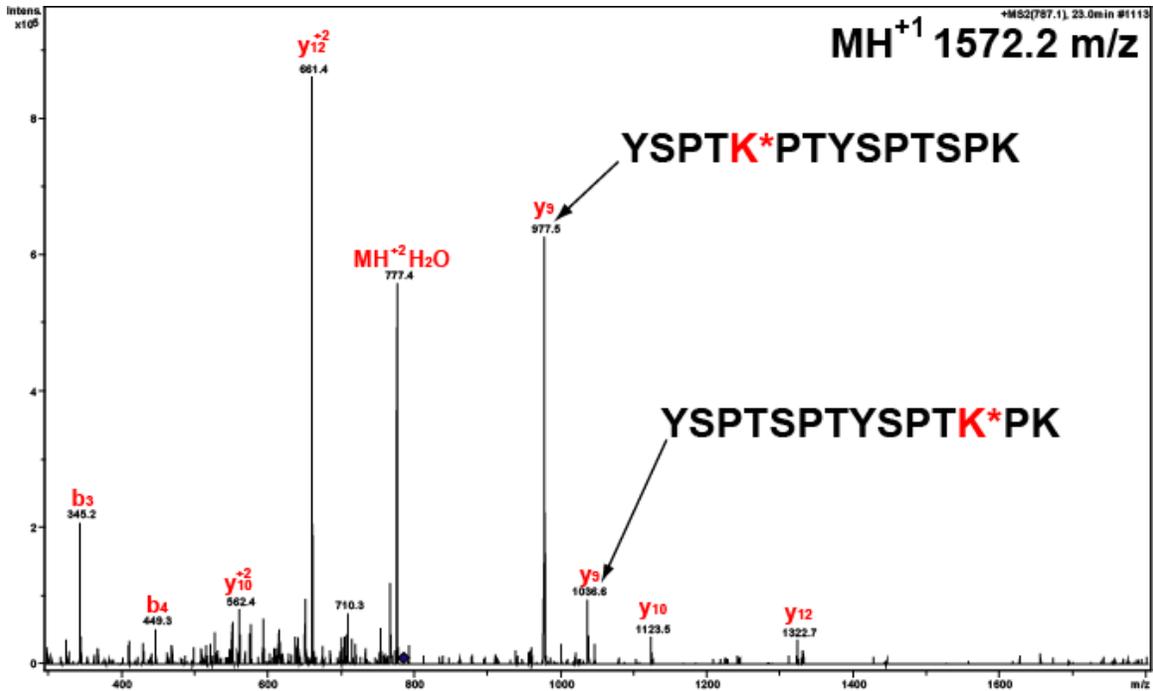


Figure 5.13 Beta-elimination/Michael addition of 2X CTD repeats with a single phosphorylation. The 810m/z peak is thought to be a formylation product since it does not match any expected masses.

The advantage of purifying kinase reaction products is the variety of actual CTD substrate peptides, as opposed to ordering a synthetic pCTD peptide, which would be much more expensive and may not be representative of all pCTD peptides. Future development experiments will employ the CTD kinase P-TEFb which is recombinantly expressed in baculovirus and exhibits somewhat greater activity with the CTD than the endogenous purifications of TFIIH. More phosphorylated CTD products will allow for more thorough development experiment to optimize the beta-elimination/Michael Addition chemistry. Successful application of this chemistry with the ~21kD pCTD fragment would convert negatively charged phosphates to positively charged lysine analogs, increasing the charge density improving ionization, along with a supercharging reagent, also improving sequencing efficiency with gas-phase ETD fragmentation (Syka, et al., 2004). ETD has been tested (without the beta-elimination/Michael Addition chemistry) with an Orbitrap Velos using a supercharging reagent. Only a small portion of the ~21kD CTD fragment was fragmented, but theoretical c-ions were identified, so it can be done. Increasing sequencing efficiency will be required for a successful method to characterize a pattern of modification on the CTD. Although, for the Orbitrap, which has a fairly narrow mass window compared with other types of mass analyzers (only up to ~3kD), only multiple charge states are shown, which greatly increases the complexity of the mass spectrum. A different style of mass analyzer may provide more readily interpretable results. A new quadrupole-time of flight (Q-TOF) mass spectrometer with ETD, HCD and CID will allow the interrogation of spectra up to

100kD. The full option of gas-phase fragmentation and the “top-down/middle-down” capability of the Q-TOF may provide the data to identify site-specific modification of a highly purified pCTD.

5.3 Discussion

RNA Pol II is the protein really at the center of it all in terms of gene activation. The responsibility of this protein is great in terms of being highly regulated for activation, but once activated, it must negotiate through a chromatin environment for productive transcription and elongation, and then it must ensure the processing of nascent mRNA transcripts. This involves coordinating the timely arrival of mRNA Capping enzymes at the initiation of transcription and then coordinating the Spliceosome to remove intronic sequences, and finally cofactors that polyadenylated and terminate the 3'-end of the mRNA transcript. The ~40kD CTD is the business portion for these critical protein-protein interactions. Phosphorylation of the CTD appears to be a molecular mechanism for regulating the spatial and temporal protein-protein interactions crucial for transcription-coupled mRNA processing. The modification of such a large repeating sequence begs the question of a “CTD Code” or a pattern of modification that is specific for distinct binding domains. This would allow for substrate specificity and affinity to be regulated by associating cofactors, such as the Mediator complex. These experiments could be performed with a method to accurately identify site-specific

phosphorylation or any modification on the CTD. Furthermore, the benefits of these types of experiments are:

1. True substrate recognized by RNA Processing Factors
2. Define unique biologically relevant epitopes for production of antibodies
3. Biochemical purification of factors interacting with Phospho-CTD motifs
4. Establish CTD phospho-patterns: mRNA processing and transcription
5. Mechanistic insight into CTD kinase specificity and processivity.

To our knowledge the entire ~40kD CTD has never been used to probe the protein-protein interactions, or the CTD interactome, in a cell. Small synthetic peptides have been used, but never the actual endogenous substrate that is the full length CTD. It is somewhat challenging to express and purify even the GST-CTD fusion protein to do experiment such as this, let alone the kinases that modify it. Having derived highly purified reagents for these experiments, the CTD was phosphorylated and used to pulldown HeLa nuclear extract. Both of the known major kinases were used to phosphorylate the CTD and a combination of both with the addition of a non-phosphorylated control CTD. These four samples were analyzed by mass spectrometry to characterize the nuclear interactome of each 'phosphorylation state' of the GST-CTD fusion. Every effort was made to match the GST-CTD pulldowns. The most striking results are the interactions of the Integrator and Mediator complexes with the both the unphosphorylated and phosphorylated CTD. It has been thought for

many years that the phosphorylation of the CTD is what dissociates the PIC, Mediator and RNA Pol II, via electrostatic repulsion of the negatively charged phosphate groups. Not all of the CTD repeats are phosphorylated in any of the samples used for pulldowns, however, it would be expected that there would be less Mediator in the pCTD pulldowns. Western blotting has confirmed the association of Mediator with a phosphorylated CTD. To test whether additional cofactors were likely linking the Mediator complex to the pCTD and that it is not a direct interaction, a column fraction enriched in Mediator (P1.0M/Q1.0M) was used for an input of both the unphosphorylated and a TFIIH/P-TEFb phosphorylated CTD. Mediator was pulled-down by both the CTD and pCTD. This data fits nicely with chromatin immunoprecipitation-sequencing (ChIP-Seq) data showing Mediator colocalizing with serine 5 phosphorylated RNA Pol II (Kagey, et al., 2010; Takahashi, et al., 2011). A future direction for these experiments is to repeat them with purified endogenous RNA Pol II complex. If Mediator binds the phosphorylated CTD, then that is not the cue for releasing Pol II from the promoter, and there is some other mechanism coinciding with phosphorylation of the CTD.

5.4 Methods

5.4.1 GST-CTD Purification

The 52-repeat CTD of rpb1 (residues 1593-1970) was cloned into a pGEX-4T-3 vector for bacterial expression. The fusion protein was transformed into Codon Plus RIPL BL21 strain (Stratagene) which contains rare codons in bacteria. The BL21s were grown at 37°C to OD 0.6 and then induced for 10 minutes. IPTG was then added to 0.5 mM and the induction was overnight at 13°C. Cell pellets were lysed and purified using a standard GSH-Sepharose purification. The GST-CTD fusion protein was eluted with 30 mM GSH and concentrated with a Microcon spin filter. The concentrated GST-CTD was loaded onto a Superdex 200 gel filtration column and 0.3 mL fractions were collected and analyzed by SDS-PAGE and Coomassie staining. Fractions were stored at -80°C.

5.4.2 Purification of CTD Kinases

TFIIH was purified from HeLa nuclear extract as reported (Kneusel, et al., 2009) and P-TEFb was purified from Sf9 baculovirus cell pellets as reported (Tahirov, et al., 2010).

5.4.3 Purification of Phospho-CTD Interacting cofactors

To generate the phosphorylated CTD baits for pulldowns, kinase reactions were performed with purified GST-CTD, TFIIH and P-TEFb as reported (Knuesel, et al., 2009). The kinase reactions were added to washed GSH-Sepharose beads and incubated 1h at 4°C. Two 0.15M HEGN washes removed the kinase reactions and all buffer was taken off the beads. Cleared HeLa nuclear extract supplemented with Microcystin LR, sodium orthovanadate, sodium fluoride, sodium pyrophosphate and beta-glycerophosphate was added to the CTD/pCTD immobilized beads. Pulldowns were incubated 3h at 4°C. After binding, the resin was washed five times with 10 column volumes (CV) 0.5 M KCl HEGN (20 mM Hepes, pH 7.6; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40 alternative) and one time with 10 CV 0.15 M KCl HEGN (0.02% NP-40 alternative). Bound proteins were eluted with 2% (w/v) Sarkosyl in elution buffer (80 mM Tris, 0.1 mM EDTA, 10% Glycerol, 0.02% NP-40, 100 mM KCl).

5.4.4 Proteomics Analysis of the Phospho-CTD Interactome

Sarkosyl elutions from GST-CTD and GST-pCTD (TFIIH, P-TEFb and both TFIIH & P-TEFb) pulldowns from HeLa NE were precipitated at 4°C using 20%(w/v) TCA, 0.067mg/mL insulin and 0.067%(w/v) deoxycholate. Precipitated protein pellets were washed twice with -20°C Acetone and air dried. Proteins were trypsin digested using a slightly modified Filter-Aided Sample Prep (FASP) protocol (Wiśniewski, et al., 2009). Briefly, protein pellets were suspended with 4%(v/v) SDS, 0.1M Tris pH 8.5, 10mM TCEP and incubated 30m ambient to

reduce disulfides. Reduced proteins were diluted with 8M Urea, 0.1M Tris pH8.5 and iodoacetamide was added to 10mM and incubated 30m in total darkness. Reduced and alkylated proteins were then transferred to a Microcon YM-30 spin concentrator and washed twice with 8M Urea, 0.1M Tris pH 8.5 to remove SDS. Three washes with 2M Urea, 0.1M Tris pH8.5 were performed then Lys-C was added and incubated overnight at ambient on a nutator. Trypsin and 2mM CaCl₂ were added and incubated approximately 4 hours in a 37°C dry incubator on a nutator. Digested peptides were eluted and acidified with ~5% (v/v) formic acid.

Peptides were desalted online and fractionated with a Phenomenex Jupiter C18 (5µm 300Å; 0.25 x 150mm) column using a two dimensional LC/MS/MS method (Agilent 1100). Thirteen steps of increasing acetonitrile (2, 4, 6, 8, 10, 12, 15, 18, 21, 25, 30, 50 and 100%B; A: 20mM ammonium formate pH10, 4% acetonitrile and B: 10mM ammonium formate pH10, 65% acetonitrile) at 5µL/minute eluted peptides for a second dimension analysis on a Phenomenex Jupiter Proteo C12 (4µm 90Å; 0.075 x ~300mm) running a gradient at 0.2µL/minute from 5 to 25% B in 100 minutes for steps one through six and 10 to 30% B in 100 minutes for step seven (A: 4% acetonitrile & B: 80% acetonitrile, both with 0.1% formic acid pH~2.5). Eluted peptides were detected with an Agilent MSD Trap XCT (3D ion trap) mass spectrometer.

All spectra were searched with Mascot v2.2 (Matrix Sciences) against the International Protein Index (IPI) database version 3.65 with two missed cleavages and mass tolerances of m/z ±2.0 Da for parent masses and ±0.8 Da for MS/MS fragment masses. Peptides were accepted above a Mascot ion score

corresponding to a 1% false discovery rate (1% FDR) determined by a separate search of a reversed IPI v3.65 database. Peptides were then filtered and protein identifications were assembled using in-house software as described (Meyer-Arendt, et al., 2011; Resing, et al., 2004).

Bibliography

Alberts B. 1998. The cell as a collection of protein machines: Preparing the next generation of molecular biologist. *Cell* **92**: 291-294.

Baillat D, Hakimi MA, Näär AM, Shilatifard A, Cooch N, Shiekhattar R. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*. 123(2):265-276.

Bark SJ, Hook V. 2007. Differential recovery of peptides from sample tubes and the reproducibility of quantitative proteomic data. *Journal of Proteome Research*. **6**: 4511-4516.

Behrends C, Sowa ME, Gygi SP, Harper JW. 2010. Network organization of the human autophagy system. *Nature*. **466**: 68-76.

Bernecky C, Grob P, Ebmeier CC, Nogales E, Taatjes DJ. 2011. Molecular Architecture of the Human Mediator–RNA Polymerase II–TFIIF Assembly. *PLoS Biology*. **9**: e1000603. doi:10.1371/journal.pbio.1000603.

Black JC, Choi JE, Lombardo SR, Carey M. 2006. A mechanism for coordinating chromatin modification and preinitiation complex assembly. *Molecular Cell*. **23**: 809–818.

Boeing S, Rigault C, Heidemann M, Eick D, Meisterernst M. 2010. RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. *Journal of Biological Chemistry*. **285**: 188-196.

Bourbon HM, Aguilera A, Ansari AZ, Asturias FJ, Berk AJ, Bjorklund S, Blackwell TK, Borggreffe T, Carey M, Carlson M, Conaway JW, Conaway RC, Emmons SW, Fondell JD, Freedman LP, Fukasawa T, Gustafsson CM, Han M, He X, Herman PK, Hinnebusch AG, Holmberg S, Holstege FC, Jaehning JA, Kim YJ, Kuras L, Leutz A, Lis JT, Meisterernst M, Naar AM, Nasmyth K, Parvin JD, Ptashne M, Reinberg D, Ronne H, Sadowski I, Sakurai H, Sipiczki M, Sternberg PW, Stillman DJ, Strich R, Struhl K, Svejstrup JQ, Tuck S, Winston F, Roeder RG, Kornberg RD. 2004. A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Molecular Cell*. **14**: 553-557.

Brown MS, Goldstein JL. 1997. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell*. **89**: 331-340.

Chakalova L, Debrand E, Mitchell JA, Osborne CS, Fraser P. 2005. Replication and transcription: shaping the landscape of the genome. *Nature Reviews Genetics*. **6**: 669-677

Bibliography

- Chen EI, Hewel J, Felding-Habermann B, Yates JR.** 2006. Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). *Molecular and Cellular Proteomics*. **5**: 53–56.
- Choudhary C, Mann M.** 2010. Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews Molecular and Cellular Biology*. **1(6)**: 427-439.
- Conaway RC, Sato S, Tomomori-Sato C, Yao T, Conaway JW.** 2005. The mammalian Mediator complex and its role in transcriptional regulation. *Trends in Biochemical Sciences*. **30**: 250–255.
- Cook PR.** 1999. The organization of replication and transcription. *Science* **284**: 1790-1795.
- Cravatt BF, Simon GM, Yates JR.** 2007. The biological impact of mass-spectrometry based proteomics. *Nature*. **450**: 991–1000.
- Donner AJ, Ebmeier CC, Taatjes DJ, Espinosa JM.** 2010. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature Structural and Molecular Biology*. **17**: 194-201.
- Eberlé D, Hegarty B, Bossard P, Ferré P, Foufelle F.** 2004. SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie*. **86**: 839-848.
- Egloff S, Murphy S.** 2008. Cracking the RNA polymerase II CTD code. *Trends in Genetics*. **24**: 280-288.
- Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S.** Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*. **318**: 1777-1779.
- Fan H, Sakuraba K, Komuro A, Kato S, Harada F, Hirose Y.** 2003. PCIF1, a novel human WW domain-containing protein, interacts with the phosphorylated RNA polymerase II. *Biochemical and Biophysical Research Communications*. **301**: 378-385.
- Felsenfeld G, Groudine M.** 2003. Controlling the double helix. *Nature* **421**: 448-453.
- Florens L, Washburn MP.** 2006. Proteomic analysis by multidimensional protein identification technology. *Methods in Molecular Biology* **328**: 159–175.

Bibliography

Fu X, Yucer N, Liu S, Li M, Yi P, Mu JJ, Yang T, Chu J, Jung SY, O'Malley BW, Gu W, Qin J, Wang Y. 2010. RFW3-Mdm2 ubiquitin ligase complex positively regulates p53 stability in response to DNA damage. *Proceedings of the National Academy of Sciences U S A.* **107**: 4579-4584.

Gilar M, Olivova P, Daly AE, Gebler JC. 2005. Orthogonality of separation in two-dimensional liquid chromatography. *Analytical Chemistry.* **77**: 6426-6434.

Gingras AC, Gstaiger M, Raught B, Aebersold R. 2007. Analysis of protein complexes using mass spectrometry. *Nature Reviews Molecular and Cellular Biology.* **8**: 645-654.

Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL. 2009. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Molecular and Cellular Biology.* **29**: 5455-5464.

Hammond-Martel I, Yu H, Affar EB. 2011. Roles of ubiquitin signaling in transcription regulation. *Cell Signaling.* 2011 Oct 17. [Epub ahead of print]

Hayden MS, Ghosh S. 2004 Signaling to NF-kappaB. *Genes and Development.* **18**: 2195-2224.

Hayden MS, Ghosh S. 2008. Shared principles in NF-kappaB signaling. *Cell.* **132**: 344-362. Review.

Horton JD, Shah NA, Warrington JA, Anderson NN, Park SW, Brown MS, Goldstein JL. 2003. Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. *Proceedings of the National Academy of Sciences USA.* **100**: 12027-12032.

Iavarone AT, Jurchen JC, Williams ER. 2001. Supercharged protein and peptide ions formed by electrospray ionization. *Analytical Chemistry.* **73**: 1455-1460.

Ishikawa H, Tachikawa H, Miura Y, Takahashi N. 2006 TRIM11 binds to and destabilizes a key component of the activator-mediated cofactor complex (ARC105) through the ubiquitin-proteasome system. *FEBS Letters.* **580**: 4784-4792.

Ito M, Yuan CX, Malik S, Gu W, Fondell JD, Yamamura S, Fu ZY, Zhang X, Qin J, Roeder RG. 1999. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Molecular Cell.* **3**: 361-370.

Bibliography

- Jang MK, Mochizuki K, Zhou M, Jeong HS, Brady JN, Ozato K.** 2005. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular Cell*. **19**: 523-534.
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA.** 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. **467**: 430-435.
- Knight ZA, Schilling B, Row RH, Kenski DM, Gibson BW, Shokat KM.** 2003. Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nature Biotechnology*. **21**: 1047-1054.
- Knuesel MT, Meyer KD, Bernecky C, Taatjes DJ.** 2009. The human CDK8 subcomplex is a molecular switch that controls Mediator co-activator function. *Genes and Development*. **23**: 439-451.
- Knuesel MT, Meyer KD, Donner AJ, Espinosa JM, Taatjes DJ.** 2009. The human CDK8 subcomplex is a histone kinase that requires Med12 for activity and can function independently of mediator. *Molecular and Cellular Biology*. **29**: 650-661.
- Köcher T and Superti-Furga G.** 2007. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nature Methods* **4**: 807-815.
- Kruse, JP and Gu W.** 2009. Modes of p53 Regulation. *Cell*. **137**: 609-622.
- Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T.** 2007. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Reviews Genetics*. **8**: 104-115.
- Liebler DC & Ham A-JL.** 2009. Spin filter-based sample preparation for shotgun proteomics. *Nature Methods*. **6**: 785; author reply 785-6.
- Liu X, Vorontchikhina M, Wang Y, Faiola F, Martinez E.** 2008. STAGA recruits Mediator to the MYC oncoprotein to stimulate transcription and cell proliferation. *Molecular and Cellular Biology*. **28**: 108-121.
- Lonard DM, Nawaz Z, Smith CL, O'Malley BW.** 2000. The 26S proteasome is required for estrogen receptor-alpha and coactivator turnover and for efficient estrogen receptor-alpha transactivation. *Molecular Cell*. **5**: 939-948.

Bibliography

- Luger, K.** 2003. Structure and dynamic behavior of nucleosomes. *Current Opinion in Genetics & Development* **13**: 127-135.
- Malik S, Roeder RG.** 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends in Biochemical Sciences*. **30**: 256–263.
- Max T, Søgaard M, Svejstrup JQ.** 2007. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *Journal of Biological Chemistry*. **282**: 14113-14120.
- Meaburn KJ, Misteli T.** 2007. Cell biology: chromosome territories. *Nature*. **445**: 379-781.
- Meyer KD, Donner AJ, Knuesel MT, York AG, Espinosa JM, Taatjes DJ.** 2008. Cooperative activity of CDK8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *EMBO Journal*. **27**: 1447–1457.
- Meyer KD, Lin S, Bernecky C, Gao Y, Taatjes DJ.** 2010. p53 activates transcription by directing structural shifts in Mediator. *Nature Structural and Molecular Biology*. **17**: 753-760.
- Meyer-Arendt K, Old WM, Houel S, Renganathan K, Eichelberger B, Resing KA, Ahn NG.** 2011. IsoformResolver: A peptide-centric algorithm for protein inference. *Journal of Proteome Research*. **10**: 3060-3075.
- Misteli T.** 2007. Beyond the sequence: cellular organization of genome function. *Cell*. **128**: 787-800.
- Näär AM, Beaurang PA, Zhou S, Abraham S, Solomon W, Tjian R.** 1999. Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature*. **398**: 828-832.
- Näär AM, Taatjes DJ, Zhai W, Nogales E, Tjian R.** Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. 2002. *Genes and Development*. **16**: 1339-1344.
- Narlikar GJ, Fan HY, Kingston RE.** 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell*. **108**: 475-487.
- Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG.** 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular and Cellular Proteomics*. **4**:1487–1502.

Bibliography

- Oliner JD, Andresen JM, Hansen SK, Zhou S, Tjian R.** 1996. SREBP transcriptional activity is mediated through an interaction with the CREB-binding protein. *Genes and Development*. **10**: 2903–2911.
- Orphanides G, Reinberg D.** 2002. A unified theory of gene expression. *Cell*. **108**:439-451.
- O'Shea JM, Perkins ND.** 2008. Regulation of the RelA (p65) transactivation domain. *Biochem Soc Trans*. **36(Pt 4)**: 603-608.
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS.** 1999. Probability-based protein identification by searching sequence databases using mass spectrometry. *Electrophoresis* **20**: 3551–3567.
- Perkins ND.** 2007. Integrating cell-signalling pathways with NF-kappaB and IKK function. *Nature Reviews Molecular and Cellular Biology*. **8**: 49-62.
- Phair RD, Scaffidi P, Elbi C, Vecerová J, Dey A, Ozato K, Brown DT, Hager G, Bustin M, Misteli T.** 2004. Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Molecular and Cellular Biology*. **24**: 6393-6402.
- Raska I, Shaw PJ, Cmarko D.** 2006. New insights into nucleolar architecture and activity. *International Review of Cytology*. **255**: 177-235.
- Reed R, Hurt E.** 2002. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell*. **108**: 523-531.
- Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, Old WM, Cheung HT, Russell S, Wattawa JL, Goehle GR, Knight RD, Ahn NG..** 2004. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Analytical Chemistry* **76**: 3556–3568.
- Rusnak F, Zhou J, Hathaway GM.** 2002. Identification of phosphorylated and glycosylated sites in peptides by chemically targeted proteolysis. *Journal of Biomolecular Techniques*. **13**: 228-237.
- Rusnak F, Zhou J, Hathaway GM.** 2004. Reaction of phosphorylated and O-glycosylated peptides by chemically targeted identification at ambient temperature. *Journal of Biomolecular Techniques*. **15**: 296-304.
- Salghetti SE, Caudy AA, Chenoweth JG, Tansey WP.** 2001. Regulation of transcriptional activation domain function by ubiquitin. *Science*. **293**: 1651-1653.

Bibliography

- Sato S, Tomomori-Sato C, Parmely TJ, Florens L, Zybaylov B, Swanson SK, Banks CA, Jin J, Cai Y, Washburn MP, Conaway JW, Conaway RC.** 2004. A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Molecular Cell* **14**: 685–691.
- Seo YK, Jeon TI, Chong HK, Biesinger J, Xie X, Osborne TF.** 2011. Genome-wide localization of SREBP-2 in hepatic chromatin predicts a role in autophagy. *Cell Metabolism*. **13**: 367-375.
- Shiloh, Y.** 2006. The ATM-mediated DNA-damage response: taking shape. *TRENDS in Biochemical Sciences*. **31**: 402-410.
- Silverman J, Takai H, Buonomo SB, Eisenhaber F, de Lange T.** 2004. Human Rif1, ortholog of a yeast telomeric protein, is regulated by ATM and 53BP1 and functions in the S-phase checkpoint. *Genes and Development*. **18**: 2108-2119.
- Simonsen A, Tooze SA.** 2009. Coordination of membrane events during autophagy by multiple class III PI3-kinase complexes. *Journal of Cellular Biology*. **186**: 773-782.
- Singh R, Cuervo AM.** 2011. Autophagy in the cellular energetic balance. *Cell Metabolism*. **13**: 495-504.
- Siuti N, Kelleher NL.** 2007. Decoding protein modifications using top-down mass spectrometry. *Nature Methods*. **4**: 817-821.
- Svejstrup JQ, Li Y, Fellows J, Gnatt A, Bjorklund S, Kornberg RD.** 1997. Evidence for a mediator cycle at the initiation of transcription. *Proceedings of the National Academy of Sciences U S A*. **94**: 6075-6078.
- Swaney DL, Wenger CD, Coon JJ.** 2010. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *Journal of Proteome Research*. **9**: 1323-1329.
- Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF.** 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences U S A*. **101**: 9528-9533.
- Taatjes DJ, Marr MT, Tjian R.** 2004. Regulatory diversity among metazoan co-activator complexes. *Nature Reviews Molecular Cell Biology*. **5**: 403–410.
- Taatjes DJ, Naar AM, Andel F, Nogales E, Tjian R.** 2002. Structure, function, and activator-induced conformations of the CRSP coactivator. *Science*. **295**: 1058–1062.

Bibliography

Taatjes DJ, Schneider-Poetsch T, Tjian R. 2004. Distinct conformational states of nuclear receptor-bound CRSP-Med complexes. *Nature Structural and Molecular Biology*. **11**: 664–671.

Taatjes DJ. 2010. The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in Biochemical Sciences*. **35**: 315-322.

Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. 2010. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature*. **465**: 747-751.

Takahashi H, Parmely TJ, Sato S, Tomomori-Sato C, Banks CA, Kong SE, Szutorisz H, Swanson SK, Martin-Brown S, Washburn MP, Florens L, Seidel CW, Lin C, Smith ER, Shilatifard A, Conaway RC, Conaway JW. 2011. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell*. **146**: 92-104.

Thomas and Chang. 2006. The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology*. **41**: 105-178.

Tinette S, Feyereisen R, Robichon A. 2007. Approach to systematic analysis of serine/threonine phosphoproteome using Beta elimination and subsequent side effects: intramolecular linkage and/or racemisation. *Journal of Cellular Biochemistry*. **100**: 875-882.

Tuoc TC, Stoykova A. 2008. Trim11 modulates the function of neurogenic transcription factor Pax6 through ubiquitin-proteasome system. *Genes and Development*. **22**: 1972-1986.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al. 2001. The sequence of the human genome. *Science*. 2001. **291**:1304-1351.

Vousden, KH and Prives C. 2009. Blinded by the Light: The Growing Complexity of p53. *Cell*. **137**: 413-431.

Wang H, Qian WJ, Mottaz HM, Clauss TR, Anderson DJ, Moore RJ, Camp DG 2nd, Khan AH, Sforza DM, Pallavicini M, Smith DJ, Smith RD. 2005. Development and evaluation of a micro- and nanoscale proteomic sample preparation method. *Journal of Proteome Research*. **4**: 2397-403.

Wang H, Zhao A, Chen L, Zhong X, Liao J, Gao M, Cai M, Lee DH, Li J, Chowdhury D, Yang YG, Pfeifer GP, Yen Y, Xu X. 2009. Human RIF1 encodes an anti-apoptotic factor required for DNA repair. *Carcinogenesis*. **30**: 1314-1319.

Bibliography

Wansink DG, Schul W, van der Kraan I, van Steensel B, van Driel R, de Jong L. 1993. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *Journal of Cell Biology*. **122**: 283-293.

Washburn MP, Wolters D, Yates JR. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*. **19**: 242–247.

Wiśniewski JR, Ostasiewicz P, Mann M. 2011b. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *Journal of Proteome Research*. **10**: 3040-3049.

Wiśniewski JR, Zielinska DF, Mann M. 2011a. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Analytical Biochemistry*. **410**: 307-309.

Wiśniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nature Methods*. **6**: 359-362.

Xu YX, Manley JL. 2004. Pinning down transcription: regulation of RNA polymerase II activity during the cell cycle. *Cell Cycle*. **3**: 432-435.

Xu YX, Manley JL. 2007. Pin1 modulates RNA polymerase II activity during the transcription cycle. *Genes and Development*. **21**: 2950-2962.

Yang F, Vought BW, Satterlee JS, Walker AK, Jim Sun ZY, Watts JL, DeBeaumont R, Saito RM, Hyberts SG, Yang S, Macol C, Iyer L, Tjian R, van den Heuvel S, Hart AC, Wagner G, Näär AM. 2006. An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature*. **442**: 700-704.

Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, Zhou Q. 2005. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular Cell*. **19**: 535-545.

Appendix 1. Protein Precipitation protocol

- 1. Combined Sample Fractions _____, Total Volume: _____ μ L.
- 2. Prepare 1.0mg/mL Insulin with 0.15M HEGN (for a saturated insulin solution).
_____ mg Insulin / ~ _____ mL 0.15M HEGN.
- 3. Divide Total Sample Volume: _____ μ L / 100 μ L ~ _____ -fold.
- 4. For every 100 μ L sample add in order immediately on ice then vortex:
10 μ L 1.0mg/mL Insulin X _____ ~ _____ μ L
10 μ L 1% DOC X _____ ~ _____ μ L
30 μ L TCA X _____ ~ _____ μ L
- 5. Incubate on ice 20m then Centrifuge 14,000rpm, 4C, 30m.
- 6. Aspirate Supernatant & Add 0.5mL Acetone(-20C) wash then vortex.
- 7. Incubate on ice 10m then Centrifuge 14,000rpm, 4C, 15m.
- 8. Repeat steps 6. & 7.: 0.5mL Acetone(-20C) wash, vortex, incubate, centrifuge.
- 9. Aspirate Supernatant & Air Dry Pellet.
- 10. Store Dry Pellet at -80C.

Appendix 2. Filter-Aided Sample Preparation (FASP) for mass spectrometry

- 1. Prepare 4% SDS, 0.1M Tris pH 8.5, 10mM TCEP... (100 μ L) [4% SDS]
_____ μ L 20% SDS(20 μ L) _____ μ L 0.1M TCEP(10 μ L)
_____ μ L 2M Tris pH8.5(5 μ L) _____ μ L HPLC water(65 μ L)
- 2. Prepare FRESH 8M Urea, 0.1M Tris pH 8.5 (weigh 0.6~0.7g in 1.5mL tube)
_____ mg Urea/FW60.06*8M ~ _____ mL total [UA]
_____ μ L 2M Tris pH8.5 _____ μ L HPLC water
- 3. Prepare 2M Urea, 0.1M Tris pH 8.5... (500 μ L) [UB]
_____ μ L 2M Tris pH8.5(25 μ L) _____ μ L UA (125 μ L)
_____ μ L HPLC water (350 μ L)
- 4. Suspend ~ _____ μ g protein pellet(s) with 30 μ L 4% SDS... Did it go into solution? If YES, then & Incubate 20~30m ambient to Reduce all proteins... go on to step 5. If NO, then Add 200 μ L UA and 20 μ L 0.1M TCEP & Incubate 20~30m ambient to Reduce all proteins...
- 5. Add 200 μ L UA to 30 μ L SDS suspended pellet... if it hasn't been already... & Add 13 μ L 0.5M Iodoacetamide for ~25mM... Incubate 20~30m wrapped in foil (dark)...
- 6. Add 200 μ L UA to YM-30 (Microcon) & centrifuge 14,000rpm ~5min to wash... Discard FT... Flip the filter & spin out any residual UA...
- 7. Add reduced/alkylated sample to washed YM-30 & centrifuge 14,000rpm ~10min... Discard FT...
- 8. Add 250 μ L UA... & centrifuge 14,000rpm ~10min... Repeat 2X's... ..
- 9. Add 100 μ L UB... & centrifuge 14,000rpm ~10min... Repeat 2X's... ..
- 10. Add 50 μ L UB & _____ μ L 1 μ g/ μ L Lysyl Endopeptidase (MS grade)... then gently mix ~1m... Transfers filter to a Fresh Collection tube & Wrap with parafilm... Incubate digest ambient o/n... ~ _____ hours...
- 11. Add 0.2 μ L 0.5M CaCl₂ & _____ μ L 1 μ g/ μ L MS grade Trypsin... then gently mix ~1m... Incubate Trypsin digestion in 37°C waterbath for _____...
- 12. Centrifuge 14,000rpm ~15min... Transfer _____ μ L FT to a 0.5mL low-retention tube ...
- 13. Add 50 μ L 0.5M NaCl in HPLC Water & Centrifuge 14,000rpm ~15min... Combine _____ μ L FT in 0.5mL low-retention tube ...
- 14. Freeze & store -80C...

Appendix 3. 2D-high/low pH LC/MS protocol for ATRAP

- 1. Prepare 1st Dimension rpC18 column: Cut _____ mm X 250µm fused silica capillary tubing. Wash with methanol & assemble 1µm SS nanofilter. Wash with methanol again. Pack 5µ Jupiter C18 resin >150mm: ~ _____ mm.
- 2. Purge Capillary hplc pumps 10min each with high pH buffers & wash autosampler at 100µL/m 100% CapB (65% ACN). Backpressure: _____ bar.
- 3. Attach 75µm fused silica connect with 2µm SS nanofilter (head of 1st Dim. C18 column) to packed 5µ Jupiter C18 & wash at 20µL/min (micro mode) 100% Cap B for >1h. Backpressure: _____ bar.
- 4. Equilibrate 1st Dimension 5µ Jupiter C18 at 20µL/min 100% Cap A (micro mode) for >1h. Shut off flow & allow pressure to drop slowly to avoid disturbing the column packing. Trim the head & measure: ~ _____ mm.
- 5. Assemble 1st Dimension 5µ Jupiter C18 to autosampler 6-port & load method "LOAD.m" to start flow 10µL/min 100% CapA (micro/split flow mode). Backpressure: _____ bar.
- 6. Start Accessory Capillary pump (dilution) flow at 50µL/min (3.5% ACN, 0.2% FA) & connect 3-way T-connect.
- 7. Attach 1st Dimension 5µ Jupiter C18 flowing 5µL/min & 75µm connect (with 1µm nanofilter) to 3-way T-connect. Attach 75µm connect to 10-port switching valve. Record Accessory Cap (dilution) pump backpressure: _____ bar & 1st Dimension 5µ Jupiter C18 backpressure: _____ bar.
- 8. Set 10-port switching valve to position 1 "load trap" to pressure test. Record Accessory Cap (dilution) pump backpressure: _____ bar & 1st Dimension 5µ Jupiter C18 backpressure: _____ bar.
- 9. Load method "LOAD.m" to equilibrate system for sample load (10-port switching valve to position 2 "inject trap"). Allow ~15min to equilibrate pressures & Record Accessory Cap (dilution) pump backpressure: _____ bar & 1st Dimension 5µ Jupiter C18 backpressure: _____ bar.
- 10. Write hplc methods sequence _____ to load & fractionate peptides. Print to file & double check sequence...
- 11. Thaw peptide sample & acidify... Centrifuge 14,000rpm, ~4C, 30min. Inspect for pellet & carefully transfer _____ µL supernatant to a low-retention hplc sample insert.
- 12. Start sequence to run 2D-LC method...