

Development and implementation of in vivo crosslinking for identifying protein
targets of the transcriptional activator Gal4

by

Jody Kristen Lancia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2010

Doctoral Committee:

Professor Anna K. Mapp, Chair
Professor Richard R. Neubig
Professor David H. Sherman
Assistant Professor Jason E. Gestwicki

© Jody Kristen Lancia

2010

Dedicated to my Mom, Dad and Grammie

Acknowledgements

There are so many people that have been instrumental to helping me achieving my goals in graduate school. I would like to thank my research advisor Dr. Anna Mapp who has been an extraordinary mentor. She is an excellent motivator always challenging me to think about my projects from new and interesting perspectives. She also has helped me to think analytically about scientific problems and has played a critical role in my development as a scientist. Thank you Anna.

My first research experience was in Dr. Martin L. Smith's laboratory at Indiana University School of Medicine. He provided my first opportunity to perform biological research. I must thank him for trusting a sophomore chemistry and Spanish major in his molecular biology lab. All of the experience I learned in his lab prepared me for graduate school.

I want to thank members of the Mapp lab both past and present. Specifically, I would like to thank Steve Rowe and Brian Brennan who were excellent mentors especially in the first few years of graduate school. They have also been great friends and always made each day in lab fun and interesting. There have been several lab members Dr. Lori Lee, Adaora Nwokoye, Dr. Chinmay Majmudar, Dr. Malathy Krishnamurthy, Amanda Dugan and Hugo Fung that have been collaborators on various projects and would like to thank as research moves forward when people work together.

I would like to thank my friends who have always been great sources of support and help me to not take life so serious. Emily Wiemer and Dara Leto have helped me to remember to laugh and have been there through it all. Thank you both.

To my family, words cannot express what your love and support has meant to me over the years. I would not have been able to accomplish my goals

without you as graduate school is equally technically and emotionally challenging. To my parents, I have dedicated my thesis to you because I could not have done it without you. You have always been there for me and I thank you from the bottom of my heart. To my brother Jamie and sister-in-law Katie, thank you for always being my cheerleaders. Words can't express my gratitude. To my nephews James and Noah, thank you for being constant reminders that life is short and to have fun and for always putting a smile on my face. To Grammie, you were an extraordinary woman and role model whom I will always admire. I love you all.

To Paul, your endless support, patience, and love have meant more to me than I could ever express. Thank you and I love you.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Figures	viii
List of Tables	x
Abstract	xi
Chapter 1: Introduction	1
A. Overview of transcriptional regulation	1
B. Misregulation and Disease	3
C. Transcriptional activator architecture	5
<i>C.1 DNA binding domains</i>	6
<i>C.2 Transcriptional activation domains</i>	7
<i>C.2.a Structural studies of TADs</i>	8
<i>C.2.b. Transcriptional activation domains interact with masking proteins</i> ...	8
D. Transcriptional activation domains and their direct protein binding partners	10
<i>D.1 General transcriptional machinery as targets of transcriptional activators</i>	11
<i>D.2 The Mediator complex and the proteasome as targets of transcriptional activators</i>	12
<i>D.3 Chromatin-remodeling complexes as targets of transcriptional activators</i>	13
E. The search continues: current methods used to identify activator-targets	15
<i>E.1 Chromatin Immunoprecipitation for identifying targets of transcriptional activators</i>	16
<i>E.2 Förster resonance energy transfer and in vitro photo-crosslinking techniques for detecting targets of TADs</i>	17
F. Thesis Overview	20
G. References	21
Chapter 2: Development of in vivo photo-crosslinking for capturing transcriptional activator Gal4•protein interactions	25
A. Project focus	25
B. Background	26
<i>B.1 Genetically encoding nonnatural amino acids in bacteria</i>	26
<i>B.2 From E.coli to eukaryotes-Using E. coli tRNA/synthetase pairs for nonsense suppression in S. cerevisiae</i>	28

C. Optimizing the incorporation of <i>p</i>-benzoyl-L-phenylalanine into a transcriptional activator	31
<i>C.1 Towards <i>p</i>-benzoyl-L-phenylalanine incorporated into Gal4</i>	32
D. Expanding <i>p</i>-benzoyl-L-phenylalanine incorporation throughout the Gal4 transcriptional activation domain	34
<i>D.1 Selecting residues along Gal4 TAD for <i>p</i>-benzoyl-L-phenylalanine incorporation</i>	35
<i>D.2 <i>p</i>-benzoyl-L-phenylalanine incorporation along Gal4 TAD</i>	36
E. Functional impact of incorporating <i>p</i>-benzoyl-L-phenylalanine into Gal4 transcriptional activation domain	37
F. In vivo photo-crosslinking with Gal4 transcriptional activation domain	38
G. Methods	42
H. References	47

Chapter 3: Identifying the direct binding partners of Gal4 transcriptional activation domain via in vivo crosslinking and MudPIT analysis 50

A. Project Overview	50
B. Background	50
<i>B.1 Mass spectrometric analyses of complex proteomes</i>	53
<i>B.1.a MALDI versus ESI for analyzing proteomes</i>	53
<i>B.1.b. Top-down versus bottom-up proteomics</i>	54
<i>B.2. Methods used for studying proteomic samples</i>	55
<i>B.2.a Two-dimensional polyacrylamide-gel electrophoresis followed by mass spectrometry</i>	56
<i>B.2.b. Two-dimensional liquid chromatography</i>	58
<i>B.2.c Two-dimensional liquid chromatography followed by MS-MS</i>	60
C. Preparing an in vivo photo-crosslinked sample for mass spectrometric analysis	62
<i>C.1. Optimizing yeast lysing conditions and wash buffers</i>	64
<i>C.2. Two affinity columns for purifying crosslinked proteins</i>	66
<i>C.3. Purifying Gal4 with <i>p</i>-benzoyl-L-phenylalanine incorporated</i>	67
<i>C.4. Gel electrophoresis followed by mass spectrometric analysis</i>	69
D. Multidimensional Protein Identification Technology (MudPIT) for identifying protein partners of Gal4	71
E. Using Multidimensional protein identification technology for identifying targets beyond Gal80	73
<i>E.1. Analyzing proteins identified in crosslinked samples</i>	76
<i>E.2. A limitation in the MudPIT analysis</i>	77
<i>E.3. Future MudPIT studies</i>	78
F. Methods	79
G. References	83

Chapter 4: A comparative analysis of photoactivatable crosslinking moieties 86

A. Project Focus	86
B. Background	87

<i>B.1 Photocrosslinking with transcriptional activators</i>	87
<i>B.2 p-benzoyl-L-phenylalanine and p-azido-L-phenylalanine reactive mechanisms and amino acid preferences</i>	90
<i>B.2.a p-benzoyl-L-phenylalanine</i>	90
<i>B.2.b. p-azido-L-phenylalanine</i>	92
<i>B.3 Frequency of preferred amino acids for p-benzoyl-L-phenylalanine and p-azido-L-phenylalanine crosslinking at protein interfaces</i>	94
C. Amino acid propensity within transcription proteins	96
D. In vivo crosslinking with p-benzoyl-L-phenylalanine and p-azido-L-phenylalanine at the Gal4-Gal80 interface	99
E. Methods	104
F. References	106
Chapter 5: Future Directions	108
A. In vivo crosslinking in a histone deacetylase, Rpd3	108
B. Photo-crosslinking with transcriptional activation domains in mammalian cells	111
C. Studying the amino acid preference with photo-crosslinkers	111
D. Additional experiments for verifying direct targets of Gal4 identified by mass spectrometry	112
E. References	113
Appendix: Data from MudPIT analysis	114

List of figures

Chapter 1

Figure 1-1: General scheme of transcription	2
Figure 1-2: A. immuno-staining with anti-REST shows elevated REST protein expression in medulloblastoma cells	4
Figure 1-3: Zinc finger protein transcription factor induces angiogenesis in mice	5
Figure 1-4: Crystal structure of Gal4 DBD (1-100) in complex with DNA	6
Figure 1-5: Amino acid sequences of amphipathic TADs Gal4, Gcn4, VP16	7
Figure 1-6: X-ray crystal structure of p53 peptide in complex with MDM2	9
Figure 1-7: Timeline of proteins and protein complexes discovered as potential targets of TADs	11
Figure 1-8: Structure of PEAS aryl azide photo-crosslinking moiety	18

Chapter 2

Figure 2-1: General nonsense suppression method for site-specifically incorporating nonnatural amino acids in vivo (<i>S. Cerevisiae</i>)	27
Figure 2-2: Some of the nonnatural aminos that have been incorporated into proteins in <i>S. cerevisiae</i> in vivo.....	28
Figure 2-3: tRNA promoter architecture	30
Figure 2-4: Structure of pBpa at the ground (black) and excited (red) states ...	31
Figure 2-5: Structure of Gal4 peptide in complex with Gal80.	32
Figure 2-6: Construct used in this study LexA-Gal4(840-881)-FLAG and the environmental sensitivity of Gal4	33
Figure 2-7: Different tRNA/synthetase pairs (as described above) were used to determine which yielded the highest amount of full-length LexA-Gal4(849 TAG) protein.....	34
Figure 2-8: Site-specific incorporation of pBpa at positions along the Gal4 TAD.....	37
Figure 2-9: β -galactosidase assays showing the functional impact of incorporating pBpa into Gal4 TAD	38
Figure 2-10: In vivo crosslinking of pBpa-containing Gal4 mutants.....	39
Figure 2-11: Gal4 crosslinks to c-Myc-Gal80 in vivo	40
Figure 2-12: In vivo crosslinking with pBpa incorporated at positions near the N-terminal 866, 875, 879	41

Chapter 3

Figure 3-1: General strategy for preparing a photo-crosslinked of Gal4 for mass spectrometric analysis.....	51
Figure 3-2: General schemes for bottom-up and top-down proteomics.....	54
Figure 3-3: General MS-MS strategy for proteomic studies.....	57
Figure 3-4: Example of off-line (top) and on-line (bottom) MS approaches.....	59
Figure 3-5: LexA-Gal4-FLAG-6X-His construct used in the MS experiments ...	63
Figure 3-6: Gel showing the results of different lysis and wash conditions of purifying WT LexA-Gal4-FLAG-6X-His with Ni NTA resin	65
Figure 3-7: Silver stained gel of WT LexA-Gal4 purified with Ni NTA and FLAG affinity resin from 6L of yeast.....	67
Figure 3-8: Silver Stained gel (left) and Western blot (right) of the same samples showing purification of 6L of either WT LexA-Gal4 –pBpa –UV	68
Figure 3-9: Silver stained 2D gel of purified crosslinked LexA-Gal4 Phe849pBpa mutant purified from 12L of yeast.....	70
Figure 3-10: Method for preparing a crosslinked LexA-Gal4 Phe849pBpa sample for MudPIT analysis	71
Figure 3-11: Number of unique and total peptides identified for Gal4 Phe849pBpa and Gal80 with and without UV using MudPIT.....	73
Figure 3-12: Potential targets of Gal4 and the amount of peptides identified from crosslinked Gal4 Phe849pBpa in the presence of galactose	75
Figure 3-13: Crosslinking experiment with LexA-Gal4-FLAG with pBpa incorporated at different positions along Gal4 TAD	78

Chapter 4

Figure 4-1: Scheme of crosslinking with ¹²⁵ I-PEAS	87
Figure 4-2: Sulfo-SBED crosslinking strategy.....	89
Figure 4-3: pAzpa and pBpa reactive mechanisms for crosslinking.....	90
Figure 4-4: Ball and stick model of Calmodulin (left) and the 17 residue α -helical peptide (right).	91
Figure 4-5: Geometric constraints for pBpa hydrogen abstraction of C-H insertion	92
Figure 4-6: Frequency of amino acids at six different interfaces	94
Figure 4-7: Structure of p53 peptide with MDM2	96
Figure 4-8: In vivo photo-crosslinking with pBpa incorporated at 10 positions along the Gal4 TAD.	100
Figure 4-9: Structure of Gal4 peptide in complex with Gal80	101
Figure 4-10: Incorporation and in vivo photo-crosslinking with pAzpa in Gal4.....	103

Chapter 5

Figure 5-1: Incorporation and in vivo photo-crosslinking pBpa in Rpb3	110
--	-----

List of Tables

Chapter 1

Table 1-1: Transcription proteins and complexes identified with various techniques as interacting with amphipathic TADs	15
--	----

Chapter 2

Table 2-1: Residues selected for pBpa incorporation along Gal4 TAD.....	35
--	----

Chapter 3

Table 3-1: Protein copies/cell of potential targets of amphiphatic activators	52
Table 3-2: Strategies for proteomic analysis discussed in this sections	56
Table 3-3: Lysis and wash buffers with different stringency (1<2<3) used in the initial optimization	65
Table 3-4: Proteins identified by MudPIT of crosslinked LexA-Gal4 Phe849pBpa in the presence of galactose.	74
Table 3-5: Samples prepared for MudPIT analysis.....	77

Chapter 4

Table 4-1: Amino acid reactivity preference of pBpa and pAzpa crosslinking ...	92
Table 4-2: Percent of Met, Trp, His and Leu in transcription proteins, arbitrary proteins and potential targets of Gal4 identified by MudPIT	98

Appendix

Table A-1: Samples prepared for MudPIT analysis	114
Table A-2: Data from the MudPIT experiments	115

Abstract

Protein-protein interactions are essential to biological processes; therefore, developing methods to discover and characterize these interactions is of high importance to the scientific community. Transcription, a fundamental cellular function, relies upon protein-protein interactions between a transcriptional activator and the protein complexes that are recruited to DNA in order to initiate the process. Given the relationship between transcription misregulation and disease, protein-protein interactions responsible for transcriptional regulation are attractive targets for therapeutic intervention. Despite considerable effort, the discovery of the direct and functionally relevant protein binding partners of transcriptional activation domains (TADs) of the amphipathic activator class has proven to be challenging, with a number of conflicting models. To create a detailed network map filling this void, an enhanced tRNA/tRNA synthetase pair has been developed to site specifically incorporate the photolabile nonnatural amino acid *p*-benzoyl-L-phenylalanine (pBpa) in vivo (*S. cerevisiae*) into the amphipathic prototypical transcriptional activator Gal4 to capture the interactions through in vivo photo-crosslinking. Applying this methodology, in vivo photo-crosslinking was used to detect a key binding partner of Gal4, the inhibitor protein Gal80. This approach when followed by mass spectrometric analysis of crosslinked protein partners also resulted in identifying Gal80 as a protein in the purified crosslinked reaction mixture. Using this technique other potential targets of Gal4 were identified under conditions where Gal80 is not inhibiting Gal4-activated transcription. Some of these potential targets, Ccr4, Mot2, Tra1, Ssl2, Mlp2, Abf1, Ctr9, Swi1, Hrr25, are consistent with previously suggested targets of amphipathic transcriptional activators while others are novel findings. Further verification of these targets will be critical for elucidating the mechanism of

transcriptional activation, useful information for developing therapeutics designed to regulate aberrant gene expression in diseases.

Chapter 1 Introduction

All cells have a unique transcriptional signature indicative of their overall function, cell cycle status and health.¹ Although strategies are available to characterize these cells by their transcriptional signature(s), many fundamental questions exist surrounding the mechanistic details of activated transcription. One of the unanswered questions is what are the specific protein-protein interactions (PPIs) utilized by transcriptional activators. The focus of my dissertation is aimed at answering this question by developing and implementing in vivo photo-crosslinking for uncovering the direct protein binding partners of the prototypical eukaryotic transcriptional activator Gal4.

A. Overview of transcriptional regulation

In the process of transcription, a cell's genetic material (DNA) is transcribed into mRNA which is then translated into proteins; as outlined above, appropriate regulation of this essential process is vital for the health and viability of the organism. As a result, a variety of extra- and intracellular signal transduction cascades tightly regulate this process by controlling what proteins are localized at a gene promoter.² A more detailed look at the mechanism of transcription initiation paints an exceedingly complicated picture of molecular recognition network that controls this process (Figure 1-1).^{2,3} Playing a central role in the process are transcriptional activators, proteins that are recruited to DNA in order to activate transcription. After the initial signal is relayed to the transcriptional activator, indicating transcription must be initiated, most activators interact with one or more enzymes that confer post-translational modifications. These modifications include phosphorylation, ubiquitylation, methylation, and sumoylation, all of which can alter the conformation, binding properties and

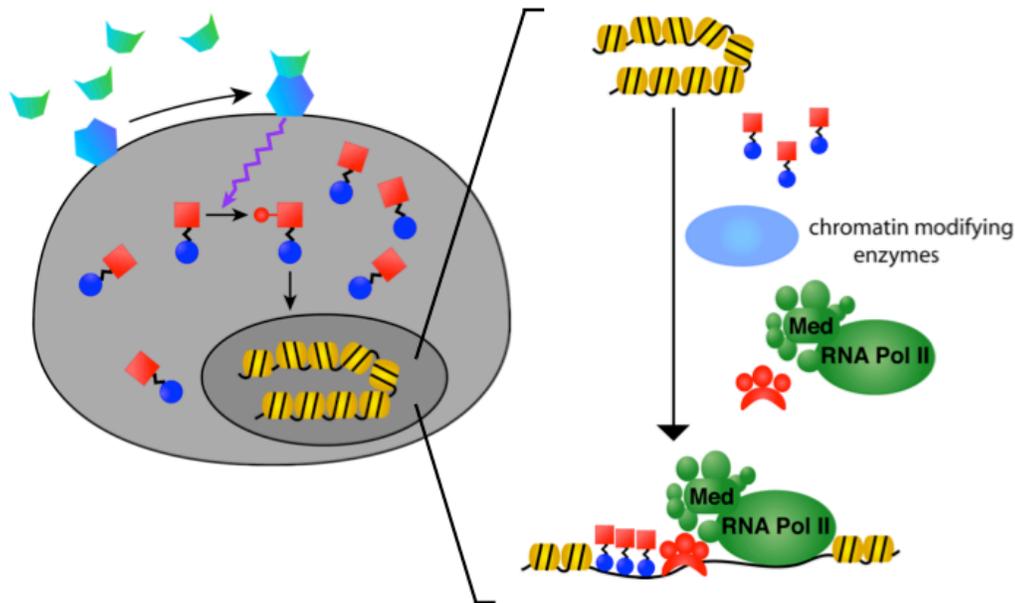


Figure 1-1: General scheme of transcription. Transcriptional activators are represented by blue circles connected to red squares. Upon external signaling a transcriptional activator undergoes a post-translational modification, is shuttled to the nucleus and localized to DNA via the DNA binding domain. Numerous proteins including chromatin modifying enzymes and the general transcriptional machinery including RNA polymerase are recruited to DNA for initiation of activated transcription.⁴ Reproduced with permission.

cellular localization of the activator.^{5,6} If the activators are not already in the nucleus, they must be shuttled there in order to interact with DNA. It is at the stage after the activator locates its cognate binding sequences in DNA that the most complex and poorly understood binding interactions take place, as the activator must assemble the transcriptional machinery (RNA polymerase II and associated protein complexes) at the gene in order for transcription to be initiated through one or more direct binding events.² Transcriptional activators are thought to possess a multipartner binding profile with proteins recruited to DNA. These proteins/complexes include a preinitiation complex (PIC) which is considered to be the general transcriptional machinery and is made up of TFIIA, TFIIB, TFIID (which includes TATA binding protein (TBP)), TFIIE, TFIIF, TFIIF, and RNA polymerase II. Other proteins recruited are TBP-associated factors (Tafs), mediator, and proteins that aid in chromatin remodeling, SAGA and Swi/Snf complexes.³ Protein-protein interactions are essential to the regulation and function of transcriptional activation.

Misregulated gene expression leads to a disruption in the transcriptional signature of the cell and is a characteristic of disease. If misregulation should occur having a detailed map of the functional PPIs a transcriptional activator makes with proteins recruited to DNA for transcription initiation would be useful in designing transcription based-therapeutics designed to bring the transcriptional gene expression signature back to a healthy state. The goal of my thesis research has been to develop tools and applying them to this system in order to capture these interactions in a cellular context.

B. Misregulation and Disease

Gene expression profiling has shown that a hallmark of all diseases is aberrant gene expression.⁷ Due to this, transcription-based therapies intended to correct gene misregulation are being pursued by many researchers.⁸⁻¹² One well-characterized example is medulloblastoma, the most malignant pediatric brain cancer, that is proposed to stem from undifferentiated external granule-layer cells located in the cerebellum. The mechanism of tumorigenesis is not fully characterized; however it is known that one protein, repressor element-1 silencing transcription/neuron-restrictive silencer factor (REST/NRSF), is inappropriately expressed in these cells. REST/NRSF acts as a repressor of a set of neuronal differentiation genes by binding to the promoter region where a transcriptional activator would bind, thereby inhibiting transcription.^{9,10} Conversely, in normal cerebellum tissue samples REST/NRSF is not expressed. To compete with REST/NRSF binding to the promoter region an adenovirus vector for a fusion protein of the REST DNA binding domain and the viral transcriptional activator VP16 (REST-VP16) was used to infect mice with preformed medulloblastoma tumors. After 4 weeks there was a visible decrease in tumor size (Figure 1-2). In addition, gene expression of terminal neuronal differentiation genes and signs of apoptosis were present.¹⁰

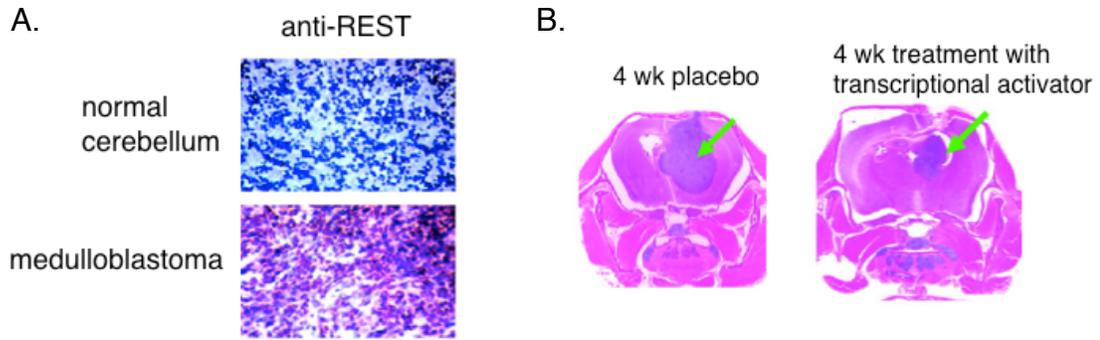


Figure 1-2: A. immuno-staining with anti-REST shows elevated REST protein expression in medulloblastoma cells. B. Mice infected with adenovirus containing REST-VP16 fusion protein results in a reduction in tumor size compared to the placebo after 4 weeks of treatment.¹⁰ Reproduced with permission.

As discussed earlier the zinc finger motif is found in DBDs that bind DNA with high specificity. To regulate gene expression, nonnatural zinc finger proteins (ZFPs) can be engineered to recognize a specific DNA sequence. In one example a ZFP was designed to target a specific site on DNA upstream of the gene coding for vascular endothelial growth factor A (VEGF-A). Aberrant gene expression of VEGF-A is associated with tumorigenesis, diabetic retinopathy, and ischemic heart and limb disease. In this study, a designer ZFP fused to either VP16 or p65 transcriptional activator was able to activate endogenous gene expression of VEGF-A.¹¹ Another study further demonstrated the utility of ZFP based transcription factors by infecting mice with an adenovirus vector containing the ZFP fused to VP16 resulting in induction of angiogenesis (Figure 1-3).¹² Currently, a ZFP is in Phase II clinical trials at Sangamo Biosciences, Inc. It is being assessed for health benefits in diabetic related illnesses.

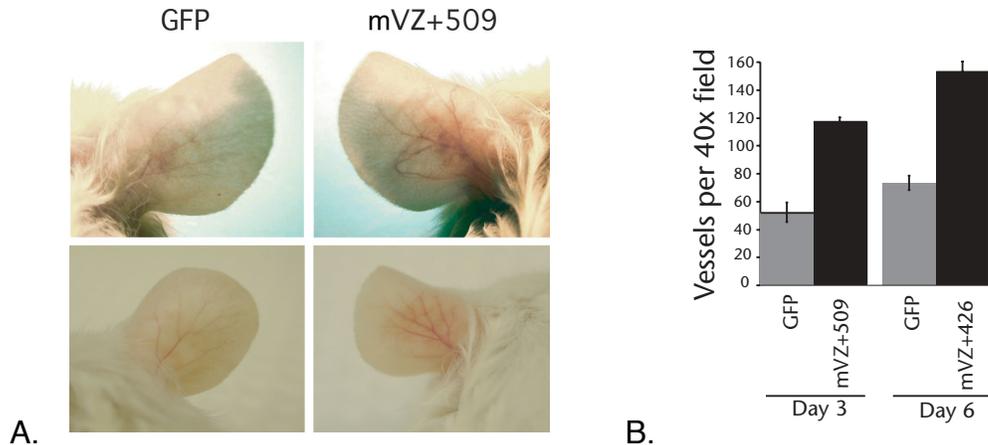


Figure 1-3: Zinc finger protein transcription factor induces angiogenesis in mice. A) Mice injected with adenovirus coding ZFP transcription factor (right) shows visible neovascularization versus mice injected with adenovirus coding GFP on day 3. B) Quantified vessel counts by immunohistochemistry. Day 3 for mVZ+509 and day 6 for mVZ+426 ZFP transcription factors.¹² Reproduced with permission.

C. Transcriptional activator architecture

In order for transcription to be initiated a transcriptional activator must be localized to the DNA.¹³ Transcriptional activators are minimally composed of two modules or domains: a DNA binding domain (DBD) whose job it is to localize the activator to DNA in a site-specific manner and a transcriptional activation domain (TAD) which makes necessary contacts with the general transcriptional machinery recruited to the DNA in order for transcription initiation to occur. These two modules can exist within the same polypeptide or assemble through non-covalent interactions. Interestingly, these two modules may be separated where the TAD of one activator can be fused to the DBD of another and maintain function.^{4,13} One of the first examples of this domain swapping was done by Ptashne and co-workers where they fused the DNA binding domain of the bacterial repressor protein LexA to the TAD of the yeast transcriptional activator Gal4. This fusion protein was able to activate transcription of a reporter gene only when the *lexA* operator was presented near the transcription start site.¹⁴ This seminal experiment revealed the modular architecture of activators and lead

to an extensive effort toward understanding the structure and targets of each domain.²

C.1 DNA binding domains

Of the two key domains, the DBD is better characterized. It is the job of DBDs to bind DNA at a specific site where the TAD is then localized for up-regulating a specific gene.¹⁵ A number of structural studies have revealed key features of their interactions with DNA and one can actually classify these domains structurally.^{16,17} This is in contrast to TADs, which are typically referred to in terms of primary amino acid content because there is so little known with regard to structure which will be discussed in the next section. Common DNA binding structural motifs include helix-turn-helix, zinc finger and leucine zipper.^{16,17} For example, the DBD of the well studied transcriptional activator Gal4 binds DNA as a dimer through a Zn_2Cys_6 binuclear cluster (Figure 1-4).¹⁸ In many cases DBDs, like Gal4, bind DNA as oligomers. They are regulated by post-translational modifications (phosphorylation, ubiquitylation, etc.) which may prevent DNA binding or protect them once on DNA from proteasomal degradation; this is the case with the Gal4 DBD when it is monoubiquitylated.^{5,6}

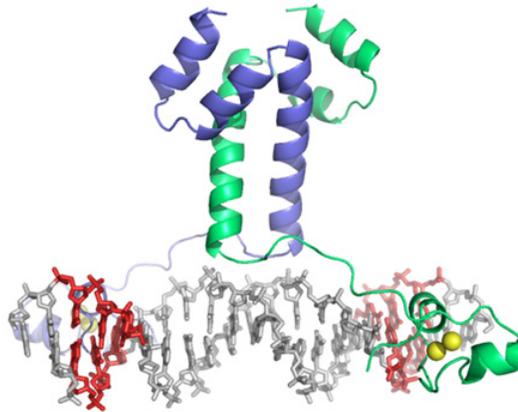


Figure 1-4: Crystal structure of Gal4 DBD (1-100) in complex with DNA. The two subunits of the dimer are colored green and blue. The zinc ion is yellow, DNA is gray and the two CGG DNA sites are in red. 11 base pairs are located in between the two CGG recognition sites.¹⁸ Figure from PDB 3COQ. Reproduced with permission.

C.2 Transcriptional activation domains

TADs typically fall into one of three categories based on their primary amino acid sequence: amphipathic, proline-rich and glutamine-rich. The amphipathic class of activators, distinguished by their interspersion of acidic/polar and hydrophobic residues, is the most studied class. Their function is generally considered to be conserved throughout eukaryotes based on their ability to maintain their function in different eukaryotic species.¹⁹ For example, the yeast activator Gal4 functions in every eukaryote tested, from yeast to mice to human.^{12,20} Therefore, studying this class of activators in one eukaryotic system provides information about how they operate in other systems. Found in this class are the yeast activators Gal4 and GCN4, the herpes virus activator VP16 and the mammalian activator p53, all of which are highly studied and serve as models for how this class of activators operates. Mutagenesis studies of these TADs have led to the identification of the minimal activation domain sequences of the protein that are essential for maintaining functional transcriptional activators (Figure 1-5).²¹⁻²⁴ In the case of

TAD	Sequence
Gal4	⁸⁴⁰ WTDQTAYNAFGITTGMFNTTTMDDVYNYLFDDEDTPPNPKKE ⁸⁸¹
Gcn4	¹⁰⁷ MFHEYENLEDNSKEWTSFLDNDIPVTTDDVSLADKAIES ¹⁴⁴
VP16	⁴⁵⁶ GFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG ⁴⁹⁰

Figure 1-5: Amino acid sequences of amphipathic TADs Gal4, Gcn4, VP16. There is an interspersion of hydrophobic (black) and acidic/polar (green) residues. Despite having little sequence homology all act as TADs.

Gal4, Ptashne and co-workers performed experiments where they created truncated versions of Gal4 and found that minimally, the 881 amino acid protein requires the N-terminal DBD (1-147) fused to the last 30 amino acids (851-881) in order to maintain function.²² Further, fusing Gal4 (1-100) to the last 41 amino acids results in only a 2-fold reduction in activation of transcription when expressed from the ACT1 promoter, remarkable considering ~85% of the protein is deleted.²⁵ This minimal transcriptional activator is fully functional in the sense

that it response to environmental cues. Since Gal4 regulates the genes whose product is responsible for galactose metabolism it is sensitive to sugars in the enviroment. For example, when glucose is present transcription is repressed by Gal4 interacting with an inhibitor a protein, Gal80. Conversely when galactose is present Gal4 is no longer repressed and transcription may occur.¹³

C.2.a Structural studies of TADs

In contrast to DBDs, TADs have proven to be much more challenging to study due to being intrinsically unstructured and having a poorly characterized, multipartner binding profile.² Several studies have suggested that TADs form an α -helix upon protein complex formation with proteins in the transcriptional machinery.²⁶⁻²⁸ One example is that of the TAD of CREB when it interacts with CBP/p300, a member of the transcriptional machinery.²⁶ It is possible that TADs may form other secondary structures upon binding other protein targets.²⁹

C.2.b. Transcriptional activation domains interact with masking proteins

TADs not only contact proteins that are recruited to DNA to activate transcription they also interact with masking proteins that regulate this process by binding to them and inhibiting transcription.³⁰ One such example is the well-characterized p53-MDM2 interaction. p53, a transcriptional activator, becomes inhibited when the MDM2 oncoprotein binds to its activation domain. p53 misregulation occurs in many types of cancer (>50%) and is a result of either a mutation that produces a functionally inert protein or MDM2 is over expressed.³¹⁻
³³ Upon interacting with MDM2, p53 TAD forms an α -helix, burying its hydrophobic residues in MDM2's hydrophobic cleft (Figure 1-5) More recently,

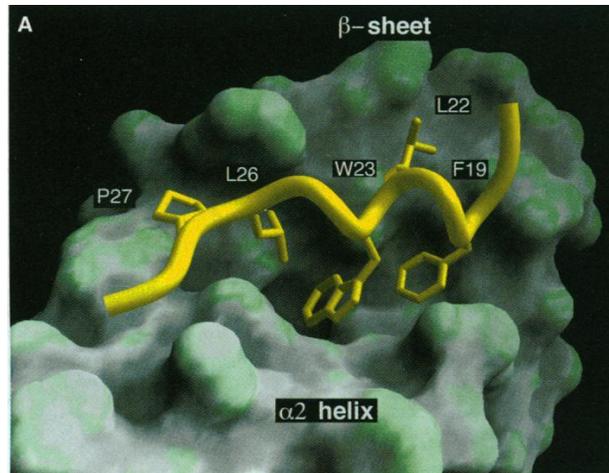


Figure 1-6: X-ray crystal structure of p53 peptide in complex with MDM2. p53 forms an amphipathic α -helix upon binding MDM2, with hydrophobic residues, Phe19, Trp23 and Leu26 on one face of the α -helix buried in the MDM2 binding cleft.³⁰ Reproduced with permission.

an NMR structure of the intrinsically disordered activation domain was solved. In this experiment they used a full-length p53 that contains modifications, not in the TAD, that help to stabilize the protein. This study suggests that p53 forms an α -helix in the absence of MDM2.^{34,35} There is, however, conflicting evidence suggesting that upon binding to its masking protein Gal80, Gal4 TAD is capable of forming a β -hairpin.^{29,36} However, a recent x-ray crystallographic study of Gal4 TAD complexed with Gal80, from *Kluyveromyces lactis* (*K. lactis*), defined an α -helical conformation upon binding Gal80 but this was only a small fragment.³⁷

To conclude, TADs make contacts with the both masking proteins and protein complexes that are recruited to DNA for activated transcription. As discussed earlier, besides the general transcriptional machinery other proteins are localized to DNA for transcription to occur. Many genetic and biochemical approaches, which will be discussed in the next section, have been used to identify the direct protein targets of amphipathic TADs but because these have not clearly defined the targets, conflicting models exist. It is the aim of my dissertation to develop a tool to study these interactions in a cellular environment resolving the complicated TAD-protein target interaction network.

D. Transcriptional activation domains and their direct protein binding partners

As outlined in section B, knowledge about transcription mechanisms can lead to the development of new therapeutic strategies that are dependent on the DNA sequence for delivery of the transcription-based therapy, a work still in progress. Likewise, researchers have been working to uncover the direct protein partners of TADs, information that would be very useful when trying to design new methods for controlling gene misregulation and disease.³⁸⁻⁴⁰ One could use this information to make TAD mimics, that could be used to activate or inhibit transcription depending on the context.

When transcriptional activators are localized to DNA through their DBD they are then poised for making the necessary contacts with proteins that are recruited to the DNA, needed to initiate transcription. Transcriptional activators need to recruit many proteins with different enzymatic functionalities to promoters and they are thought to make contacts with several of these proteins however, which ones they are directly contacting for activating transcription are not clearly defined. Considerable effort has been made toward uncovering these protein-protein interactions, however conflicting models still exist (Figure 1.7). This section outlines a history of the techniques used and the resulting proteins identified as potential targets of amphipathic TADs. It also shows that despite all the efforts thus far, clear targets have not been defined.

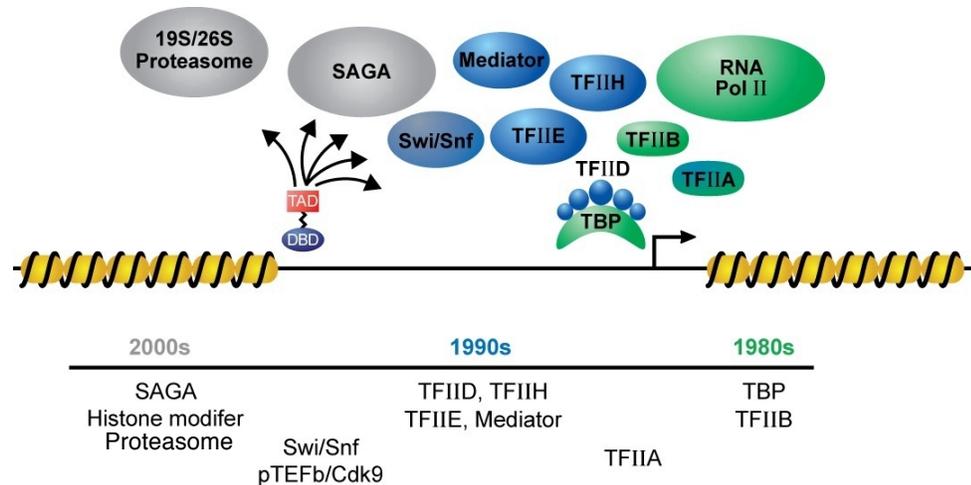


Figure 1-7: Timeline of proteins and protein complexes discovered as potential targets of TADs. The direct and functional targets of TADs are still debatable.² Reproduced with permission.

D.1 General transcriptional machinery as targets of transcriptional activators

Starting in the 1980s, *in vitro* studies were carried out to discover the proteins that interact with amphipathic TADs. In an early study it was found that VP-16 and the mammalian activator ATF interact with the TATA box binding protein (TBP) in TFIID.^{41,42} This was followed by the discovery that TFIIB, a protein that stabilizes the TBP-TATA complex, was a potential target.^{43,44} Next, a mutagenesis study suggested that TFIIA was important for transcription induced by amphipathic activators.⁴⁵ Also, through affinity chromatography of mammalian whole cell extracts interacting with amphipathic activators, p53 and Vp16, TFIIF was isolated, thus adding this protein to the fast-growing list of potential protein targets of TADs.⁴⁶ By the mid-1990's nearly every component of the general transcriptional machinery that makes up the transcription preinitiation complex (PIC) were considered targets of amphipathic TADs.

A detailed study by Pugh and Tijan revealed that proteins that are complexed with TBP and TFIID, known today as TBP-associated factors (Tafs), in fact act as the adaptors between the TADs and the general transcriptional machinery.⁴⁷ This was further confirmed by Poon and Weil who immunopurified TBP in complex with Tafs.⁴⁸ These factors were then thought to be essential for

activator-mediated transcription and thought to be targets of TADs.⁴⁹ In vitro studies have showed that TAFs are not necessary for basal transcription but are needed for activated transcription.^{50 51} However, the idea of Tafs acting as adaptor proteins between TADs and the general transcriptional machinery was questioned based on results from in vivo studies. In these experiments TAF depletion had little affect on transcription initiated by transcriptional activators such as GCN4 and Gal4. These findings indicated that TAFs are not a general requirement for transcriptional activation in yeast.^{50,51} This information drove researchers to focus their attention toward discovering protein complexes other than those associated with TBP and that are localized to DNA during activated transcription for TAD-protein target identification. It was observed in vitro that dosing in excess transcriptional activator caused an inhibition of transcription also known as “squelching” and that adding additional amounts of the general transcriptional factors did not relieve the inhibition.^{21,52,53} This indicated that the transcriptional activators are competing for a target, perhaps a generally unknown factor, and that the general transcription factors are not the clear target.⁵⁴

D.2 The Mediator complex and the proteasome as targets of transcriptional activators

The Mediator is one of the protein complexes that is localized to DNA for transcription to take place. It was identified, then called Srb/Mediator, through purifying RNA polymerase II and functional assays on purified systems.^{55,56} This complex was thought to form stably with RNA polymerase II and in this form it was thought to be recruited to DNA and make contact with transcriptional activators, those tested were of the amphipathic class.^{57,58} Although the discovery of the Srb/Mediator complex was considered a huge step forward toward as this complex was thought to bridge or “mediate the TAD interaction to the RNA polymerase, much speculation arose when one protein, Sug1 an ATPase, originally thought to be a target and component of the Mediator/RNA

polymerase II, was found to actually be a part of the 26S proteasome.^{59,60 61} Indeed, the proteasome is not considered a member of the transcriptional machinery; however, it is thought to play an important role in regulating transcription by controlling the activator stability in the cell or abundance through activator turnover by proteolytically degrading the activator after it initiates transcription.⁶² More recently Sug1 and Sug2, components of the 19S proteasome, have been thought to actually play a role in transcription elongation both in vitro and in vivo, contrasting the original functional role of the proteasome, that it is strictly associated with proteolysis mediated transcriptional regulation, thereby making it a potential target of activators.⁶³⁻⁶⁸

D.3 Chromatin-remodeling complexes as targets of transcriptional activators

As the Mediator complex was being questioned for its validity of being a functional activator target other classes of proteins were being studied for their role in transcription initiation by activators, chromatin remodeling and chromatin modifying complexes. Chromatin remodelers and modifiers play a role in regulating transcription because DNA structural change allows the promoter region to be accessible for activator binding and recruitment of the general transcriptional machinery.^{69,70} It was thus not surprising when Workman et al. found that TADs were able to recruit the SWI/SNF complex, an ATP-dependent nucleosome remodeling enzyme, to promoters.^{39,71} Likewise, chromatin modifying enzymes, like those located in the SAGA complex are histone acetyltransferase (HATs) which modify lysine side chains found in the histone proteins that compose the nucleosomes, have also been thought to be recruited to DNA via TADs.⁷² The SAGA complex participates in at least five functions including: activator binding, making contacts with TBP, recognizing methylated histone H3, HAT activity and deubiquitinating actions therefore this macromolecular complex was not to be dismissed when trying to understand transcriptional activator interactions and mechanisms.⁶⁶ Acetylation by HATs is

reversed by histone deacetylases (HDACs) which is associated with gene repression and have been studied for their role in gene expression.⁷³ Rpd3 the yeast HDAC will be discussed further in Chapter 5.

In summary, this section outlines 25 years of scientific research aimed toward a single goal of determining the transcriptional mechanism surrounding the protein-protein interactions that transcriptional activators make with proteins that participate in transcription regulation. As techniques improved over the years some initial studies were proven false only to uncover more uncertainty in the field with each new target identified as a player in activated transcription. In the recent past new technology has set the stage for successfully painting a clearer picture of what contacts are being made between TADs and the library of potential protein targets. The next section will outline the current stage of the TAD-transcriptional protein target field.

Protein complex	Protein	Methods used
TFIID	TBP Taf12	Affinity chromatography (41) DNase footprinting (42) ChIP(76) ¹²⁵ I-PEAS (38,40,80)
TFIIB		Affinity chromatography(43, 44)
TFIIA		Mutagenesis (45)
TFIIH		Affinity chromatography(46)
Mediator	Med15 Sin4/Med16 Srb9 Med2 Rox3	Affinity chromatography(57) Protein deletion (74) ¹²⁵ I-PEAS crosslinking (38,40, 80)
Proteasome	Sug1, Sug2	Pull-down(59,63)
Swi/Snf	Swi1, Swi2/Snf2, Snf5	Protein deletion(74) Pull-down(39) Sulfo-SBED crosslinking (82)
SAGA	GCN5 Taf5 Taf6 Taf9 Taf10 Ada1 Ada2 Ada3 Ada5 Spt3 Spt7 Spt20 Tra1	ChIP (75,76,77) FRET(78) Protein deletion(74) ¹²⁵ I-PEAS crosslinking (38,40, 80)
	Mbf1	Protein deletion (78)
RSC complex	Rsc1	Protein deletion (78)
Paf complex	Cdc73 Hpr1	Protein deletion (78)
CCR4-NOT complex	Ccr4 Dbf2 Caf1	Protein deletion (78)

Table 1-1: Transcription proteins and complexes identified with various techniques as interacting with amphipathic TADs.

E. The search continues: current methods used to identify activator-targets

Traditional biochemical and genetic techniques have principally answered questions regarding what proteins are present at the time that transcription is occurring and the function of these proteins in relationship to regulating activated transcription. Due to the complex nature of transcription however, these techniques did not answer a more specific question surrounding the direct and functional targets of TADs. Traditional genetic deletions can make it difficult to determine if the effect of the deletion on transcription is directly or indirectly related to loss of TAD contacts, especially if the protein is in a complex. The change in transcription can be a direct result of the TAD not being able to make contact the deleted protein or a result of the complex becoming destabilized as a result of the absent protein. Due to this, the data from genetic deletions of transcription proteins can be difficult to analyze. For example Swanson et. al. performed extensive protein deletions and studied the effect these deletions had on GCN4 transcription. Not surprisingly, they report many proteins found in Swi/Snf, SAGA and mediator complexes as well as several other proteins/complexes as a requirement for GCN4 transcription.⁷⁴ Whether these are relevant targets is yet to be fully understood. Biochemical colocalization technology is difficult to interpret as well because many times the experiments were carried out in vitro and using a purified system. Even chromatin immunoprecipitation (ChIP), which can be performed in vivo shows protein occupancy on DNA, does not determine whether two proteins are direct partners but rather if they are colocalized to DNA. In the last 5-10 years there has been a surge of development and application of new methods for studying these interactions in live systems making the interpretation of the results much more clear and moving the field forward.

E.1 Chromatin Immunoprecipitation for identifying targets of transcriptional activators

As described earlier, the amphipathic class of activators is the most well studied and the three members that are used in recent studies are Gal4, GCN4

and VP16. Towards identifying their direct protein targets, ChIP experiments have been used by researchers. ChIP is an experiment that is designed to study proteins that are localized to DNA. In these experiments formaldehyde is used to crosslink proteins to DNA. Next, the cells are sonicated to shear the DNA. The proteins are immunopurified and then they can be analyzed by Western blots and polymerase chain reaction can be used to determine the DNA that was associated with the proteins bound. Whether or not the proteins are actually contacting one another would require many more experiments. ChIP assays carried out with GCN4 in vivo revealed the HAT GCN5, a member of the SAGA and ADA complex, was targeted to the promoter region by GCN4.⁷⁵ In a similar experiment, a ChIP assay was performed in yeast to identify targets of VP16 resulting in TBP, multiple TAFs and Tra1 being observed.⁷⁶ Likewise, Bhaumik and Green found SAGA to be localized to the Gal4 upstream activation sequence (UAS), a section of the promoter that bears Gal4-binding sites, using this technique.⁷⁷ ChIP experiments do not help determine if the activator is interacting with the proteins directly or if they are simply neighboring each other with another adaptor protein bridging the two.

E.2 Förster resonance energy transfer and in vitro photo-crosslinking techniques for detecting targets of TADs

A few years later more evidence was presented for components of SAGA to potentially interact with TADs through in vivo FRET, a technique used to study proteins that are in close proximity through the use of a pair of fluorophores. Specifically, they found that Tra1, a subunit of the SAGA complex interacts with Gal4 in yeast.⁷⁸ Although FRET is a powerful technique in that one can visualize the interactions in vivo, it too has limitations in capturing protein-protein interactions in this context. One limitation is that false-positives occur by two proteins being brought together by a bridging protein and the two proteins are not directly interacting. Likewise, false-negatives may occur if the fluorophore is not positioned in such a manner to make it capable of interacting with its fluorophore

partner. Further, when proteins are in complexes, as is the case of many transcription proteins, interpreting a positive FRET signal becomes very challenging as it is difficult to say what is bridging the two fluorophores. Additionally, when proteins, like those related to transcription, are found in low abundance in the cell, a FRET signal may be difficult to detect.⁷⁹

Another technique used to study these interactions is *in vitro* photo-crosslinking. The advantage of this technique is that one has the possibility of identifying the direct binding partner and, in some cases, even the actual binding site. In this experiment, a crosslinking capable label is placed on the transcriptional activator and upon UV irradiation covalent bonds are formed between the TAD and any protein(s) that are in close proximity. In the case of transcriptional activators, this had only been accomplished *in vitro* prior to the work outlined in this dissertation. For example, Hahn and coworkers identified direct binding partners of the yeast activator Gcn4 by labeling the protein with ¹²⁵I-PEAS, an aryl azide photo reactive crosslinker at the end of a linker that attaches to proteins through disulfide bond formation with cysteines, discussed in more detail in Chapter 4 (Figure 1-8).^{40,80} They found that in the promoter-bound

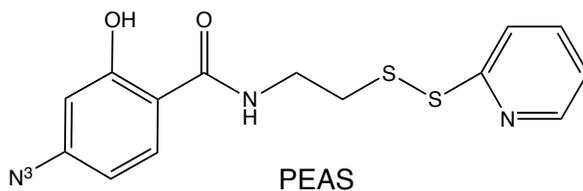


Figure 1-8: Structure of PEAS aryl azide photo-crosslinking moiety.

state in the presence of purified preinitiation complex (PIC), GCN4 formed crosslinks with Tra1, a component of SAGA, Med15(Gal11), part of mediator and Taf12, part of TFIID.^{40,80} Tra1 is part of SAGA complex, Med15 is a component of the Mediator. Based on previous studies, see above, all of these are components of larger complexes that all have been suggested to interact with amphipathic TADs. Also, Sin4/Med16 was identified as a target of only the N-terminal Gcn4 activation domain.⁸⁰ Using ¹²⁵I-PEAS labeling and Gal4 as the TAD, the same group found Tra1, Taf12 and Gal11 to be targets of Gal4 when

photo-crosslinking was carried in the presence of purified PIC and TAD.³⁸ It is interesting that the same binding partners were identified, given the lack of sequence homology of the sequences. This strengthens the idea that identifying the direct targets of an amphipathic TAD, like Gal4 in my thesis research, can be applicable to other amphipathic activators of particular interest in higher eukaryotic organisms. ¹²⁵I-PEAS contains a linker that extends the crosslinking moiety 14 Å from the α carbon of the cysteine to which it is attached so the crosslinker could possibly capture indirect partners that are bridging the TAD to the actual direct partner rather than the direct protein partner itself.⁸¹ Additionally, these experiments were performed in vitro where the crosslinking may not yield true protein targets in the absence of a complex cellular environment.

The above experiments were groundbreaking in that it was the first time that the binding preferences of DNA-bound activators had been defined through crosslinking. There have, however, been a number of studies in which in vitro photocrosslinking was used to investigate the binding behavior of isolated TADs. Sulfo-SBED is a crosslinking reagent that had been used to identify targets of GCN4 and VP16 and components of purified Swi/Snf complex. Sulfo-SBED is conjugated to the lysines on the TAD and contains a photoreactive aryl azide, a disulfide bond and a biotin label. Upon UV irradiation, crosslinks are formed via the aryl azide. DTT is then added to disrupt the disulfide bond, thereby removing the activator and leaving the biotin handle on the target which is used for Western blotting with a streptavidin antibody for detection of crosslinked targets. From these experiments subunits of the Swi/Snf complex, Swi1, Swi2/Snf2 and Swf5, were thought to interact with both GCN4 and VP16.⁸² A disadvantage of using Sulfo-SBED for crosslinking experiments is that it nonspecifically labels all lysines in a protein which can result in many proteins not related to the question at hand being crosslinked. Further nonspecific labeling is not compatible with in vivo studies because all proteins would be susceptible to labeling resulting in everything being crosslinked and little information about direct targets of TADs would be resolved.

Although technically challenging, progress has been made toward capturing the interactions that take place between TADs and proteins recruited to DNA for activated transcription. All of the methods presented here have helped to move the field forward, but they all have limitations in their capabilities to capture the relevant interactions taking place in vivo. Due to the complex nature of transcription and limitations of the current methods used to capture these direct interactions there was a need for a new tool to study these problems.

F. Thesis Overview

Outlined in this Chapter is a description of the general architecture of transcription and the progress that has been made by researchers toward understanding what interactions takes place between the numerous proteins that are recruited to DNA for transcription to occur. Methods intended to control misregulated gene expression however, are limited by the lack of an accurate map of the TAD-protein network involved in transcription regulation. To this end Chapter 2 describes the development and implementation of a new tool, in vivo photo-crosslinking using p-benzoyl-l-phenylalanine (pBpa). After verifying the utility of our chemical biology tool by capturing a known TAD-protein interaction, we extended our study in Chapter 3 to identifying unknown relevant protein targets of Gal4 by analyzing photo-crosslinked samples with mass spectrometry. Finally, in Chapter 4 an analysis of two photo-crosslinking moieties, pBpa and p-azido-l-phenylalanine (pAzpa) for capturing PPIs is studied based on their reaction mechanism and amino acid propensity at hetero-complex interfaces. In summary, this research is aimed at developing and implementing an in vivo method for mapping the protein-protein interactions between Gal4 TAD and its protein partners.

G. References

- (1) Liotta, L.; Petricoin, E. *Nat Rev Genet* **2000**, *1*, 48.
- (2) Mapp, A. K.; Ansari, A. Z. *ACS Chem Biol* **2007**, *2*, 62.
- (3) Thomas, M. C.; Chiang, C. M. *Crit Rev Biochem Mol Biol* **2006**, *41*, 105.
- (4) Ansari, A. Z.; Mapp, A. K. *Curr Opin Chem Biol* **2002**, *6*, 765.
- (5) Archer, C. T.; Delahodde, A.; Gonzalez, F.; Johnston, S. A.; Kodadek, T. *J Biol Chem* **2008**, *283*, 12614.
- (6) Tateishi, Y.; Ariyoshi, M.; Igarashi, R.; Hara, H.; Mizuguchi, K.; Seto, A.; Nakai, A.; Kokubo, T.; Tochio, H.; Shirakawa, M. *J Biol Chem* **2009**, *284*, 2435.
- (7) Chen, X.; Cheung, S. T.; So, S.; Fan, S. T.; Barry, C.; Higgins, J.; Lai, K. M.; Ji, J.; Dudoit, S.; Ng, I. O.; Van De Rijn, M.; Botstein, D.; Brown, P. O. *Mol Biol Cell* **2002**, *13*, 1929.
- (8) Toniatti, C.; Bujard, H.; Cortese, R.; Ciliberto, G. *Gene Ther* **2004**, *11*, 649.
- (9) Lawinger, P.; Venugopal, R.; Guo, Z. S.; Immaneni, A.; Sengupta, D.; Lu, W.; Rastelli, L.; Marin Dias Carneiro, A.; Levin, V.; Fuller, G. N.; Echelard, Y.; Majumder, S. *Nat Med* **2000**, *6*, 826.
- (10) Fuller, G. N.; Su, X.; Price, R. E.; Cohen, Z. R.; Lang, F. F.; Sawaya, R.; Majumder, S. *Mol Cancer Ther* **2005**, *4*, 343.
- (11) Liu, P. Q.; Rebar, E. J.; Zhang, L.; Liu, Q.; Jamieson, A. C.; Liang, Y.; Qi, H.; Li, P. X.; Chen, B.; Mendel, M. C.; Zhong, X.; Lee, Y. L.; Eisenberg, S. P.; Spratt, S. K.; Case, C. C.; Wolffe, A. P. *J Biol Chem* **2001**, *276*, 11323.
- (12) Rebar, E. J.; Huang, Y.; Hickey, R.; Nath, A. K.; Meoli, D.; Nath, S.; Chen, B.; Xu, L.; Liang, Y.; Jamieson, A. C.; Zhang, L.; Spratt, S. K.; Case, C. C.; Wolffe, A.; Giordano, F. J. *Nat Med* **2002**, *8*, 1427.
- (13) Ptashne, M.; Gann, A. *Genes & signals*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2002.
- (14) Brent, R.; Ptashne, M. *Cell* **1985**, *43*, 729.
- (15) Mapp, A. K. *Org Biomol Chem* **2003**, *1*, 2217.
- (16) Brennan, R. G.; Matthews, B. W. *J Biol Chem* **1989**, *264*, 1903.
- (17) Struhl, K. *Trends Biochem Sci* **1989**, *14*, 137.
- (18) Hong, M.; Fitzgerald, M. X.; Harper, S.; Luo, C.; Speicher, D. W.; Marmorstein, R. *Structure* **2008**, *16*, 1019.
- (19) Ponticelli, A. S.; Pardee, T. S.; Struhl, K. *Mol Cell Biol* **1995**, *15*, 983.
- (20) Hulboy, D. L.; Lozano, G. *Cell Growth Differ* **1994**, *5*, 1023.

- (21) Triezenberg, S. J.; Kingsbury, R. C.; McKnight, S. L. *Genes Dev* **1988**, *2*, 718.
- (22) Ma, J.; Ptashne, M. *Cell* **1987**, *48*, 847.
- (23) Hope, I. A.; Struhl, K. *Cell* **1986**, *46*, 885.
- (24) Fields, S.; Jang, S. K. *Science* **1990**, *249*, 1046.
- (25) Wu, Y.; Reece, R. J.; Ptashne, M. *EMBO J* **1996**, *15*, 3951.
- (26) Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741.
- (27) Jonker, H. R.; Wechselberger, R. W.; Boelens, R.; Folkers, G. E.; Kaptein, R. *Biochemistry* **2005**, *44*, 827.
- (28) Asada, S.; Choi, Y.; Yamada, M.; Wang, S. C.; Hung, M. C.; Qin, J.; Uesugi, M. *Proc Natl Acad Sci U S A* **2002**, *99*, 12747.
- (29) Van Hoy, M.; Leuther, K. K.; Kodadek, T.; Johnston, S. A. *Cell* **1993**, *72*, 587.
- (30) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. *Science* **1996**, *274*, 948.
- (31) Hainaut, P.; Hollstein, M. *Adv Cancer Res* **2000**, *77*, 81.
- (32) Hollstein, M.; Sidransky, D.; Vogelstein, B.; Harris, C. C. *Science* **1991**, *253*, 49.
- (33) Lee, L. W.; Mapp, A. K. *J Biol Chem* **2010**, *285*, 11033.
- (34) Joerger, A. C.; Allen, M. D.; Fersht, A. R. *J Biol Chem* **2004**, *279*, 1291.
- (35) Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. *Proc Natl Acad Sci U S A* **2008**, *105*, 5762.
- (36) Leuther, K. K.; Salmeron, J. M.; Johnston, S. A. *Cell* **1993**, *72*, 575.
- (37) Thoden, J. B.; Ryan, L. A.; Reece, R. J.; Holden, H. M. *J Biol Chem* **2008**, *283*, 30266.
- (38) Reeves, W. M.; Hahn, S. *Mol Cell Biol* **2005**, *25*, 9092.
- (39) Neely, K. E.; Hassan, A. H.; Wallberg, A. E.; Steger, D. J.; Cairns, B. R.; Wright, A. P.; Workman, J. L. *Mol Cell* **1999**, *4*, 649.
- (40) Fishburn, J.; Mohibullah, N.; Hahn, S. *Mol Cell* **2005**, *18*, 369.
- (41) Stringer, K. F.; Ingles, C. J.; Greenblatt, J. *Nature* **1990**, *345*, 783.
- (42) Horikoshi, M.; Hai, T.; Lin, Y. S.; Green, M. R.; Roeder, R. G. *Cell* **1988**, *54*, 1033.
- (43) Lin, Y. S.; Ha, I.; Maldonado, E.; Reinberg, D.; Green, M. R. *Nature* **1991**, *353*, 569.
- (44) Choy, B.; Green, M. R. *Nature* **1993**, *366*, 531.
- (45) Ozer, J.; Bolden, A. H.; Lieberman, P. M. *J Biol Chem* **1996**, *271*, 11182.
- (46) Xiao, H.; Pearson, A.; Coulombe, B.; Truant, R.; Zhang, S.; Regier, J. L.; Triezenberg, S. J.; Reinberg, D.; Flores, O.; Ingles, C. J.; et al. *Mol Cell Biol* **1994**, *14*, 7013.
- (47) Pugh, B. F.; Tjian, R. *Genes Dev* **1991**, *5*, 1935.

- (48) Poon, D.; Weil, P. A. *J Biol Chem* **1993**, *268*, 15325.
- (49) Goodrich, J. A.; Tjian, R. *Curr Opin Cell Biol* **1994**, *6*, 403.
- (50) Moqtaderi, Z.; Bai, Y.; Poon, D.; Weil, P. A.; Struhl, K. *Nature* **1996**, *383*, 188.
- (51) Walker, S. S.; Reese, J. C.; Apone, L. M.; Green, M. R. *Nature* **1996**, *383*, 185.
- (52) Gill, G.; Ptashne, M. *Nature* **1988**, *334*, 721.
- (53) Ptashne, M. *A genetic switch : phage [λ] and higher organisms*; 2nd ed.; Cell Press : Blackwell Scientific Publications: Cambridge, Mass., 1992.
- (54) Flanagan, P. M.; Kelleher, R. J., 3rd; Sayre, M. H.; Tschochner, H.; Kornberg, R. D. *Nature* **1991**, *350*, 436.
- (55) Lee, T. I.; Young, R. A. *Annu Rev Genet* **2000**, *34*, 77.
- (56) Bourbon, H. M.; Aguilera, A.; Ansari, A. Z.; Asturias, F. J.; Berk, A. J.; Bjorklund, S.; Blackwell, T. K.; Borggreffe, T.; Carey, M.; Carlson, M.; Conaway, J. W.; Conaway, R. C.; Emmons, S. W.; Fondell, J. D.; Freedman, L. P.; Fukasawa, T.; Gustafsson, C. M.; Han, M.; He, X.; Herman, P. K.; Hinnebusch, A. G.; Holmberg, S.; Holstege, F. C.; Jaehning, J. A.; Kim, Y. J.; Kuras, L.; Leutz, A.; Lis, J. T.; Meisterernest, M.; Naar, A. M.; Nasmyth, K.; Parvin, J. D.; Ptashne, M.; Reinberg, D.; Ronne, H.; Sadowski, I.; Sakurai, H.; Sipiczki, M.; Sternberg, P. W.; Stillman, D. J.; Strich, R.; Struhl, K.; Svejstrup, J. Q.; Tuck, S.; Winston, F.; Roeder, R. G.; Kornberg, R. D. *Mol Cell* **2004**, *14*, 553.
- (57) Kim, Y. J.; Bjorklund, S.; Li, Y.; Sayre, M. H.; Kornberg, R. D. *Cell* **1994**, *77*, 599.
- (58) Koleske, A. J.; Young, R. A. *Nature* **1994**, *368*, 466.
- (59) Swaffield, J. C.; Melcher, K.; Johnston, S. A. *Nature* **1995**, *374*, 88.
- (60) Rubin, D. M.; Coux, O.; Wefes, I.; Hengartner, C.; Young, R. A.; Goldberg, A. L.; Finley, D. *Nature* **1996**, *379*, 655.
- (61) Voges, D.; Zwickl, P.; Baumeister, W. *Annu Rev Biochem* **1999**, *68*, 1015.
- (62) Muratani, M.; Tansey, W. P. *Nat Rev Mol Cell Biol* **2003**, *4*, 192.
- (63) Gonzalez, F.; Delahodde, A.; Kodadek, T.; Johnston, S. A. *Science* **2002**, *296*, 548.
- (64) Ferdous, A.; Gonzalez, F.; Sun, L.; Kodadek, T.; Johnston, S. A. *Mol Cell* **2001**, *7*, 981.
- (65) Sulahian, R.; Sikder, D.; Johnston, S. A.; Kodadek, T. *Nucleic Acids Res* **2006**, *34*, 1351.
- (66) Nalley, K.; Johnston, S. A.; Kodadek, T. *Nature* **2006**, *442*, 1054.
- (67) Reid, G.; Hubner, M. R.; Metivier, R.; Brand, H.; Denger, S.; Manu, D.; Beaudouin, J.; Ellenberg, J.; Gannon, F. *Mol Cell* **2003**, *11*, 695.
- (68) Chang, C.; Gonzalez, F.; Roethermel, B.; Sun, L.; Johnston, S. A.; Kodadek, T. *J Biol Chem* **2001**, *276*, 30956.
- (69) Workman, J. L.; Kingston, R. E. *Annu Rev Biochem* **1998**, *67*, 545.
- (70) Kadonaga, J. T. *Cell* **1998**, *92*, 307.

- (71) Peterson, C. L.; Workman, J. L. *Curr Opin Genet Dev* **2000**, *10*, 187.
- (72) Grant, P. A.; Sterner, D. E.; Duggan, L. J.; Workman, J. L.; Berger, S. L. *Trends Cell Biol* **1998**, *8*, 193.
- (73) Grozinger, C. M.; Schreiber, S. L. *Chem Biol* **2002**, *9*, 3.
- (74) Swanson, M. J.; Qiu, H.; Sumibcay, L.; Krueger, A.; Kim, S. J.; Natarajan, K.; Yoon, S.; Hinnebusch, A. G. *Mol Cell Biol* **2003**, *23*, 2800.
- (75) Kuo, M. H.; vom Baur, E.; Struhl, K.; Allis, C. D. *Mol Cell* **2000**, *6*, 1309.
- (76) Hall, D. B.; Struhl, K. *J Biol Chem* **2002**, *277*, 46043.
- (77) Bhaumik, S. R.; Green, M. R. *Genes Dev* **2001**, *15*, 1935.
- (78) Bhaumik, S. R.; Raha, T.; Aiello, D. P.; Green, M. R. *Genes Dev* **2004**, *18*, 333.
- (79) Evans, S. K.; Aiello, D. P.; Green, M. R. *Biochem Soc Symp* **2006**, 217.
- (80) Herbig, E.; Warfield, L.; Fish, L.; Fishburn, J.; Knutson, B. A.; Moorefield, B.; Pacheco, D.; Hahn, S. *Mol Cell Biol* **2010**, *30*, 2376.
- (81) Ebright, Y. W.; Chen, Y.; Kim, Y.; Ebright, R. H. *Bioconjug Chem* **1996**, *7*, 380.
- (82) Neely, K. E.; Hassan, A. H.; Brown, C. E.; Howe, L.; Workman, J. L. *Mol Cell Biol* **2002**, *22*, 1615.

Chapter 2

Development of in vivo photo-crosslinking for capturing transcriptional activator Gal4•protein interactions*

A. Project focus

Given the central role of protein-protein interactions in biological functions, the development of methods for the discovery and characterization of such interactions has been of paramount importance in the scientific community.¹⁻⁴ Ideally these methods are compatible with live-cell studies so that protein-protein interactions can be investigated in their native environment. Towards this end, recent advances in nonsense suppression technology have enabled the site-specific incorporation of amino acids bearing moieties for photoactivatable cross-linking reactions into proteins in bacteria, yeast and human cells.^{5,6} However, there have been few applications of this technology in eukaryotes, perhaps due to challenges in obtaining efficient incorporation of the nonnatural amino acids as well as limited knowledge about the impact of these changes on structure and function.⁷⁻¹⁰ To study the affect of incorporating nonnatural amino acids into the Gal4 in live yeast and to directly uncover the protein-protein interactions between Gal4 TAD and its protein partners, our goal was to implement this methodology. To this end, p-benzoyl-l-phenylalanine (pBpa), a nonnatural amino acid with photo-crosslinking capabilities, was incorporated in vivo (*S. cerevisiae*) into various positions throughout the amphipathic transcriptional activator Gal4 with photo-crosslinking reactions resulting in a capture of the Gal4 repressor protein, Gal80.¹¹

*portions of this chapter have been published: Majmudar, C.Y., Lee, L.W., Lancia, J.K., Nwokoye, A., Wang, Q., Wands, A.M., Wang, L., Mapp, A.K., *J Am Chem Soc* **2009**, *131*, 14240.

B. Background

With few exceptions, Nature selects from 20 natural amino acids for protein construction.^{6,11} The side chains of the constituent amino acids contribute to the both the structure and function of a protein, participating in intra- and intermolecular interactions as well as functioning directly in chemical reactions. Therefore, perturbations in the amino acid sequence of a protein can have profound effects. Indeed, amino acid mutagenesis is a strategy widely used to characterize the molecular interactions and functions of a protein as well as study the role of individual amino acids in those processes.¹²⁻¹⁴ Both to further expand the utility of mutagenesis for protein study and to potentially enhance the function of proteins, there has been a great deal of effort aimed at expanding the group of amino acids that can be incorporated into proteins beyond the typical twenty.^{6,15} In the following sections, approaches towards accomplishing this are outlined.

B.1 Genetically encoding nonnatural amino acids in bacteria

Incorporating nonnatural amino acids with unique functionality, like photo-caged or photoactivatable moieties, into proteins is useful for studying biological systems. In order to genetically encode a nonnatural amino acid in a living system, several things must be considered. For example, site-specific incorporation of nonnatural amino acids into proteins requires a unique codon for the nonnatural amino acid. In early efforts towards this end, the amber stop codon (TAG) was initially chosen because it possesses two desirable qualities. It is the least used stop codon from bacteria (present at ~7% coding genes) to eukaryotes.^{5,16} During translation, a tRNA that recognizes a certain codon coding for an amino acid is charged with the appropriate amino acid by its cognate tRNA synthetase (Figure 2-1). There is a unique tRNA/synthetase pair in nature for each of the 20 natural amino acids and none that recognize the amber suppressor stop codon. Therefore, a unique tRNA/synthetase pair was engineered, that utilized the endogenous translational

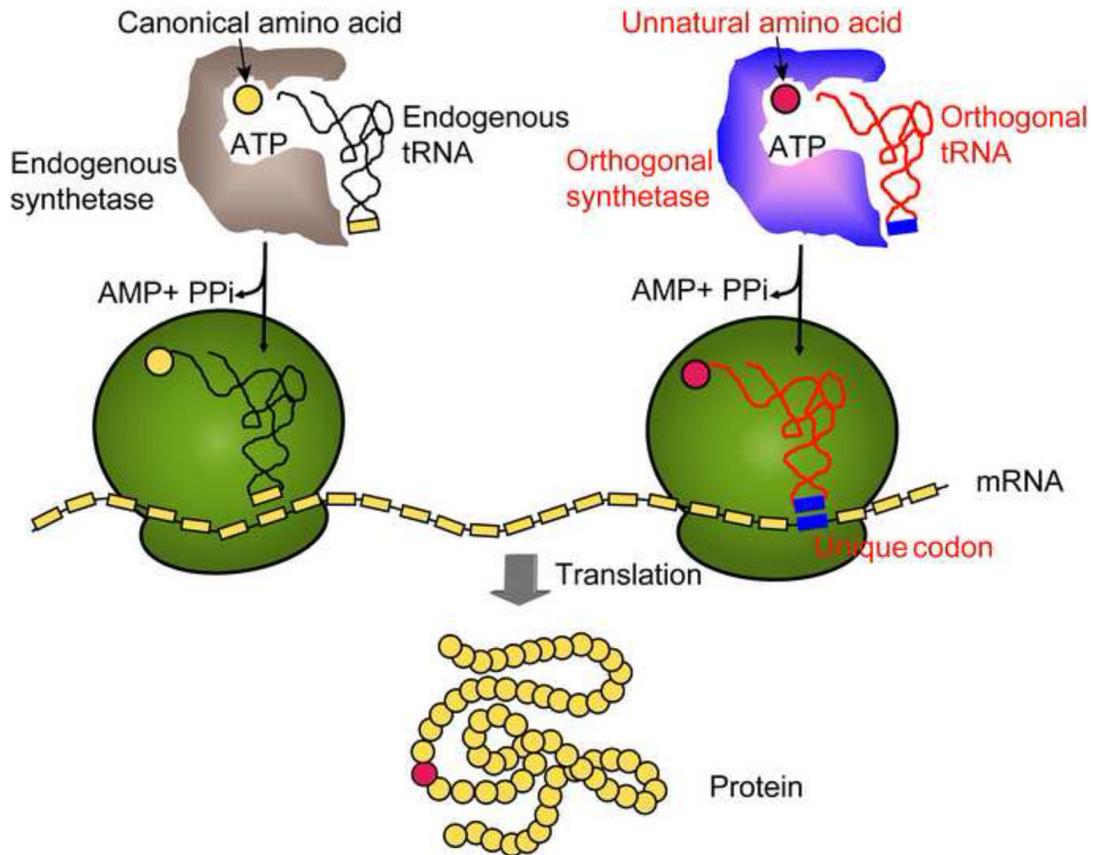


Figure 2-1: General nonsense suppression method for site-specifically incorporating nonnatural amino acids in vivo (*S. cerevisiae*).⁵ Orthogonal tRNA/synthetase pairs are engineered to charge the tRNA, that recognizes the amber stop codon, with a nonnatural amino acid that goes on to work with the endogenous translational machinery for incorporating nonnatural amino acids into proteins. Reproduced with permission.

machinery, to incorporate the nonnatural amino acid at the TAG codon instead of stopping translation or incorporating one of the naturally occurring amino acids into the site.^{6,15}

Designing a new tRNA/synthetase pair de novo is exceedingly challenging because one of the endogenous synthetases may charge the newly designed tRNA with a natural amino acid in the absence of the nonnatural amino acid and cause misincorporation to occur. Schultz et. al. developed the first orthogonal tRNA/synthetase pair in *E. coli* by modifying the tRNA^{Tyr}/tyrosyl-tRNA synthetase ($MjtRNA_{CUA}^{Tyr}/MjTyrRS$) from *Methanococcus jannaschii* (*M. jannaschii*).¹⁵ To do this they generated a library of tRNAs by randomly mutagenizing specific

nucleotides on the tRNA containing the already mutated CUA anticodon that recognizes the amber stop codon, UAG; likewise, a library of tRNA synthetases were created. A series of positive and negative selections were then used to identify an efficient and selective pair. Using the endogenous translational machinery, this new pair was evolved to incorporate the nonnatural amino acid, O-methyl-L-tyrosine, in *E. coli*.¹⁵ Since then, ~30 nonnatural amino acids have been genetically incorporated into proteins in various organisms with varying efficiency (Figure 2-2).^{5,6} Each nonnatural amino acid has an engineered tRNA/synthetase pair designed for their incorporation.

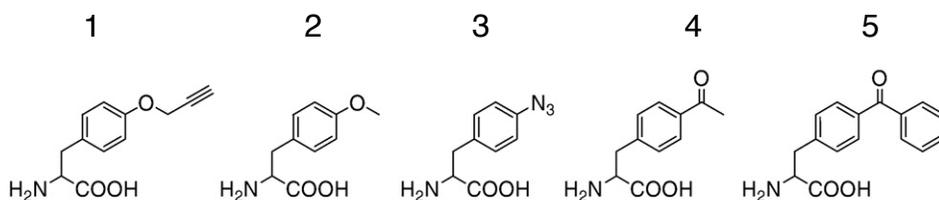


Figure 2-2: Some of the nonnatural amino acids that have been incorporated into proteins in *S. cerevisiae* in vivo. 1. p-propargyloxy-L-phenylalanine (pPpa) 2. p-methoxy-L-phenylalanine (pMpa) 3. p-azido-L-phenylalanine (pAzpa) 4. p-acetyl-L-phenylalanine (pApa) 5. P-benzoyl-L-phenylalanine.

B.2 From E.coli to eukaryotes-Using E. coli tRNA/synthetase pairs for nonsense suppression in S. cerevisiae

Initially *E. coli* tRNA/synthetase pairs (tRNA^{Tyr}_{CUA}/TyrRS) were used to generate functional pairs in yeast by applying a similar selection method described above.¹⁷ Although this was successful, when human superoxide dismutase (hSOD) was expressed in yeast with the nonnatural amino acid p-Acetyl-L-phenylalanine present for incorporating at the amber codon, the yield of full-length protein was approximately five-fold lower than wild-type hSOD expression under the same conditions (~50 ng/mL versus 250 ng/mL)¹⁸ Incorporation of the nonnatural amino acid into hSOD was verified by digesting the protein with trypsin and using mass spectrometry to analyze the peptide fragments. Thus, the next step was to optimize incorporation yields. Strategies used to increase yields include using different promoters and incorporating

elements of tRNA processing elements, described below, that differ in *E. coli* and yeast. One concern associated with increasing incorporation yield is that this may concomitantly increase the amount of read-through. Read-through is when the specialized tRNA/synthetase pair read the amber stop codon in the absence of the nonnatural amino acid, placing a naturally occurring one at this site.

Initially two groups made modifications to the tRNA/synthetase expression plasmid with the goal of increasing incorporation yields. The Hahn laboratory tested strong, medium and weak Pol III promoters driving tRNA production and found that a strong promoter reduced the yeast growth rate; however, the medium-strength N(GTT)PR promoter was reported to produce “normal” levels of proteins with nonnatural amino acid incorporated.¹⁹ The Schultz laboratory modified the tRNA expression plasmid by adding the strong Pol II promoter, pPGKI, that drove transcription of the tRNA and they also included the SUP4 5' and 3'- flanking sequences, an endogenous yeast suppressor tRNA gene, in front of the tRNA. This was included because, unlike *E. coli* tRNAs that are transcribed from a promoter located upstream of the gene, yeast tRNAs are transcribed through promoter elements that are located within the tRNA known as the A- and B-boxes (Figure 2-3A).⁵ SUP4 has a Pol III promoter with internal A and B boxes and flanking regions.²⁰ Additionally they removed the 3' CCA nucleotides from the original *E. coli* tRNA.²⁰ This change was justified by the difference in how *E. coli* and yeast process tRNA. The *E. coli* tRNA gene codes for the entire, fully processed tRNA whereas in yeast the initially transcribed tRNA lacks the 3'-CCA sequence (Figure 2-3A). The modified plasmid resulted in >50 fold increase in tRNA levels compared their original plasmid and a yield ranging 3-10mg/L of hSOD protein when p-propargyloxyphenylalanine was incorporated at three different positions in the protein.²⁰

More recently, Wang and co-workers further improved the tRNA/synthetase expression plasmid. It is known that transcription and processing of tRNAs are different between prokaryotes and eukaryotes.^{17,21} Because generating an A-box and a B-box in *E. coli* tRNA was thought to result

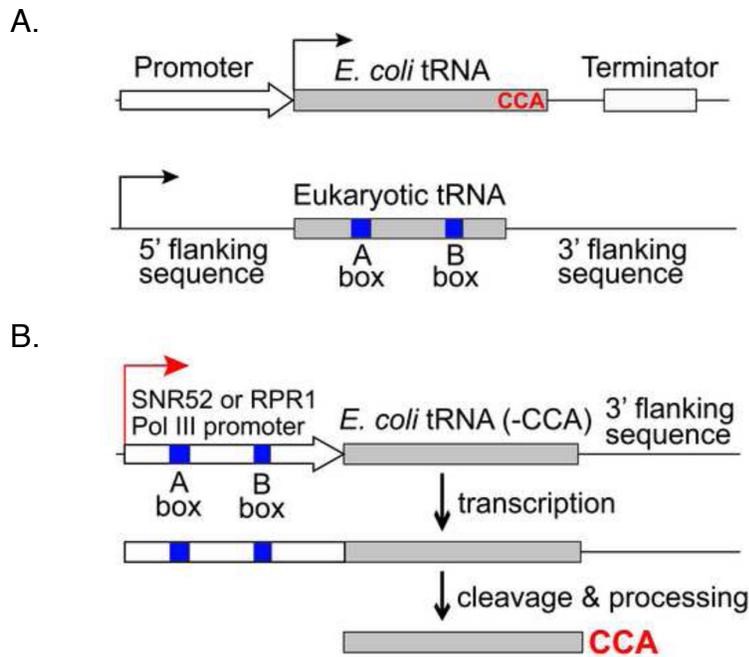


Figure 2-3: tRNA promoter architecture. A. The basic promoter architecture for expressing naturally occurring *E. coli* and eukaryotic tRNAs. B. Wang et. al. construct design for expressing tRNA for nonsense suppression.^{5,22} Reproduced with permission.

in distorted mature tRNA Wang and coworkers added a promoter containing the consensus A- and B-boxes upstream of the *E. coli* tRNA that would be post-transcriptionally cleaved to yield the final tRNA.²² Two yeast genes, SNR52 and RPR1, that are transcribed by Pol III share a promoter arrangement with the A- and B-box promoter elements situated as described above and therefore, it was reasoned that these promoters could be used to enhance production of functional tRNA.²² Similar to Schultz, Wang and co-workers inserted the *E. coli* tRNA lacking the 3'-CCA into the plasmid and the SUP4 3'-flanking sequence was present in their yeast expression plasmid as well (Figure 2-3B).^{20,22} This newly designed tRNA/synthetase expression plasmid yielded tRNA transcripts and the mature tRNA was functional for encoding amber codons in yeast with an increase in protein yield, 6-9 fold compared to the constructs with the SUP4 5'-flanking sequence alone and with little reported read-through (~7%).²² Although having SUP4 5'-flanking sequence alone produced more tRNA than the plasmid with the SNR52 promoter driving tRNA transcription, the latter produced more

incorporated protein product. This result indicates that tRNA processing is an important consideration for tRNA function.

Overall, utilizing this enhanced tRNA/synthetase pair allows one to expand the functionality of amino acid side chains that do not exist in nature to probe eukaryotic biological systems and answer questions yet unresolved. All of the tRNA/synthetase pairs described here were used in the optimization of incorporating a nonnatural amino acid into Gal4 discussed in the next section.

C. Optimizing the incorporation of *p*-benzoyl-L-phenylalanine into a transcriptional activator

The direct protein targets of the transcriptional activators have not been clearly defined.²³ Implementing nonsense suppression technology for incorporating a photo-crosslinking nonnatural amino acid into the prototypical activator Gal4 will be used to identify these targets. Of the two photoactivatable nonnatural amino acids that have been successfully incorporated into proteins in yeast, pBpa and *p*-azido-L-phenylalanine (pAzpa), pBpa was initially selected for incorporation into Gal4 because its excitation wavelength is less damaging to biomolecules than pAzpa excitation wavelength (~365nm versus ~250nm).^{2,5} Briefly, pBpa forms a diradical upon excitation and inserts into C-H bonds within the ~3 Å reactive radius (Figure 2-4).² A detailed analysis of the pBpa and pAzpa crosslinking mechanism is provided in Chapter 4.

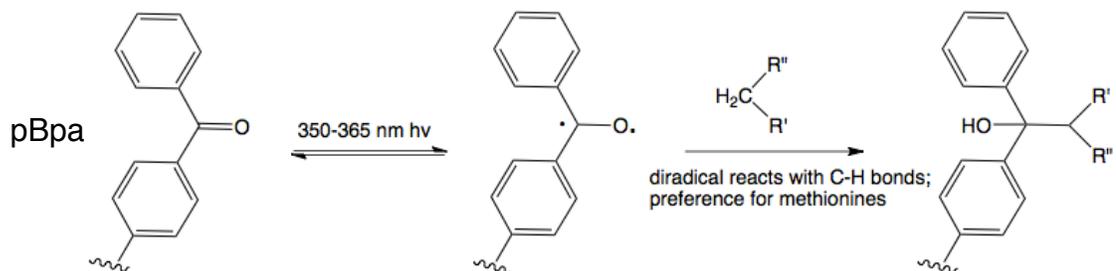


Figure 2-4: Structure of pBpa and reactive mechanism.²

C.1 Towards *p*-benzoyl-L-phenylalanine incorporated into Gal4

Gal4, a 881 residue protein, is one of the most well studied transcriptional activators. Gal4 is responsible for regulating the GAL genes, whose products are responsible for galactose catabolism.²⁴⁻²⁸ As a result, Gal4 is highly sensitive to the nutrient environment, such that only under galactose growth conditions does it function as an activator. The environmental sensitivity of Gal4 is largely regulated via a strong (nanomolar Kd) interaction between the repressor Gal80 and a portion of the C-terminal transcriptional activation domain of Gal4 (Figure 2-5).^{13,26-29} Conventional mutagenesis and structural studies suggest that the

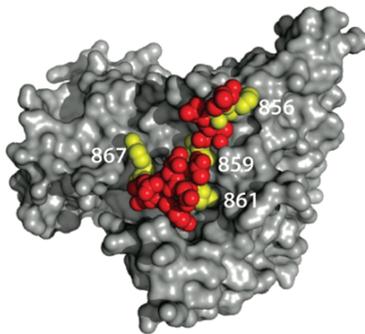


Figure 2-5: Structure of Gal4 peptide in complex with Gal80. Red and yellow highlights the Gal4 peptide (yellow are residues used in this study) and gray represents Gal80 both from *K. lactis*. The residue numbering is from *S. cerevisiae*. PDB ID 3EIK³⁰ Reproduced with permission.

middle region of the transcriptional activation domain (amino acids 851-871) comprises the binding site for Gal80, yet the Gal4-Gal80 complex impacts the function of the entire activation domain (840-881).^{13,30-32} We hypothesized that in vivo crosslinking would provide molecular-level detail regarding this important regulatory interaction. In addition to the interaction with Gal80, the Gal4 TAD engages in a set of interactions with coactivators in the transcriptional machinery to stimulate transcription under restrictive growth conditions (galactose).^{13,23,29-31,33-42} Thus, the impact of pBpa mutations on the binding modes and function of Gal4 can be assessed by examining transcriptional activity under restrictive and permissive growth conditions (Figure 2-6).

Previously it was not known if incorporating pBpa at different positions on the Gal4 TAD was possible and how changes in the amino acid

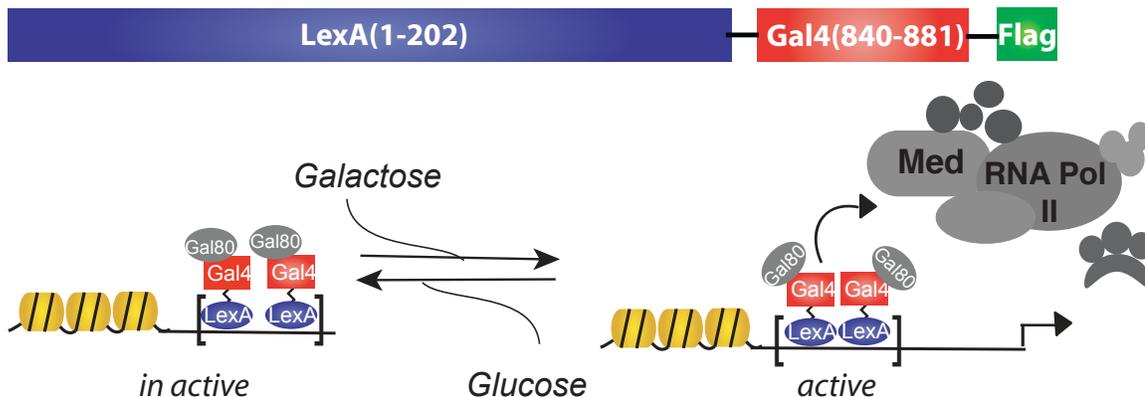


Figure 2-6: Construct used in this study LexA-Gal4(840-881)-FLAG and the environmental sensitivity of Gal4. In the presence of glucose transcription is repressed by Gal80 and under galactose growth conditions Gal4 is no longer repressed and transcription may take place.

sequence would affect the function of said activator, since mutagenesis of proteins can disrupt the natural PPIs required to maintain protein function.¹³ To this end, two colleagues in my laboratory performed a study to first optimize the conditions needed for successful incorporation. Some criteria for incorporating pBpa into Gal4 were that incorporation efficiency needed to be high so that protein could be detected via Western blot and with little read-through and low truncated protein yields. 'Read-through' refers to incorporation of a naturally occurring amino acid at the amber stop codon, a common problem in nonnatural amino acid mutagenesis. Truncated protein occurs when the amber stop codon is not 'read' at all which results in termination of translation. A fusion protein consisting of a LexA DBD, the well studied bacterial repressor protein, and the Gal4 TAD (840-881) was used in this study.¹¹ LexA+Gal4 had previously been used in our laboratory with much success. Because the yeast strain that is used, LS41, contains LexA DNA binding sites upstream of a LacZ reporter, we are able to assess protein function when pBpa is incorporated. A variety of tRNA/synthetase pairs were tested in *S. cerevisiae* for in vivo incorporation efficiency of pBpa into the Gal4 TAD (Figure 2-7). To achieve full-length protein with pBpa incorporated at pre-determined positions, the tRNA/synthetase

plasmid is co-transformed into yeast along with the plasmid coding for the LexA+Gal4 fusion protein. Position 849 of the Gal4 TAD, a phenylalanine, was chosen for the first site for pBpa

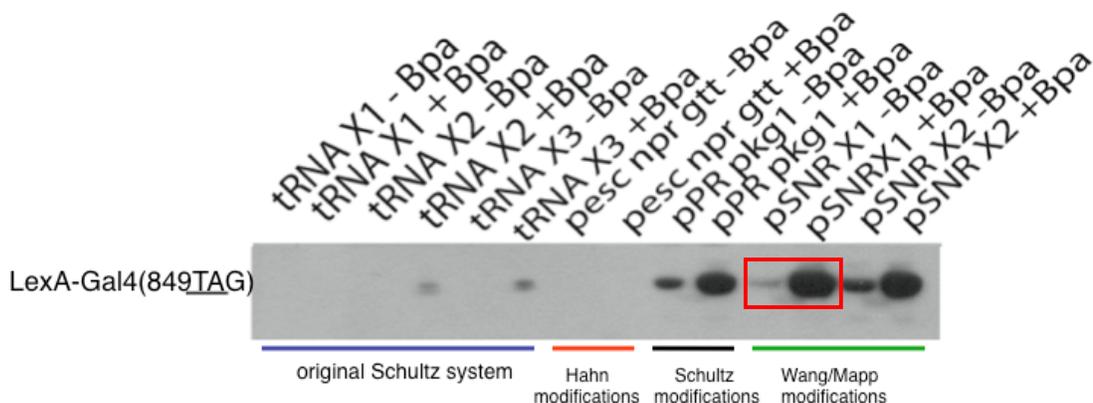


Figure 2-7: Different tRNA/synthetase pairs (as described above) were used to determine which yielded the highest amount of full-length LexA-Gal4(849 TAG) protein and minimized the amount of read-through. The Wang plasmids produced the most protein with least read-through (red box). Experiment by Dr. Chinmay Majmudar and Adaora Nwokoye.

incorporation because its large hydrophobic aromatic character is most reminiscent of the pBpa structure and therefore would minimize the disruption to the primary sequence. Utilizing this construct in live yeast, they were able to incorporate pBpa into LexA+Gal4 with ~20% efficiency compared to wild-type by using the tRNA/synthetase pair driven by the Pol III pSNR52 promoter and with little read-through.¹¹ This tRNA/synthetase construct therefore, was used in the remainder of the study.

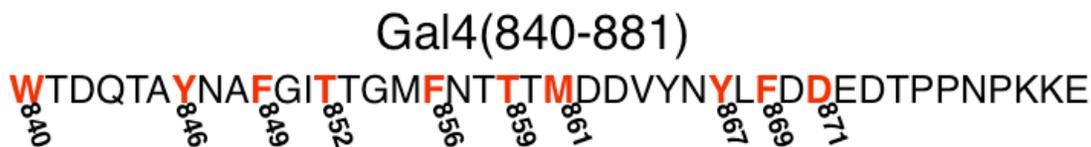
D. Expanding *p*-benzoyl-L-phenylalanine incorporation throughout the Gal4 transcriptional activation domain

We extended this study to test a broader range of mutations throughout the Gal4 TAD. Using the enhanced tRNA/synthetase pair we tested for incorporation at various positions. The goal was to determine (i) does the type of residue (hydrophobic v. polar) affect incorporation efficiency (ii) does incorporating pBpa into Gal4 affect its function and (iii) can we incorporate pBpa

into residues shown by other groups to be important for Gal4 TAD-protein target interactions to capture these interactions via photo-crosslinking.

D.1 Selecting residues along Gal4 TAD for p-benzoyl-L-phenylalanine incorporation

In case mutating the key residue disrupts Gal4 from making key contacts with its protein target(s) we also mutated residues nearby these key residues so as to not disrupt the natural interaction. Table 2-1 outlines the residues where pBpa is incorporated. Thr859 was selected as an incorporation site based on the



Position	Rational	Reference
Trp840, Tyr846, Phe849	Residues thought to be outside the Gal80 binding site, selected to study proteins outside this area	Ptashne and Gann, 2002
Phe856	Reduced Gal80 interaction in gel shift	Ansari et. al. 1998
Thr859	Reduced Gal80 interaction in gel shift	Ansari et. al. 1998
Met861	Adjacent to residue thought to be important for Gal80 binding	Ansari et. al. 1998
Tyr867	Adjacent to residue thought to be important for Gal80 binding	Ansari et. al. 1998
Asp869	Adjacent to residue thought to be important for Gal80 binding, shown to crosslink to Tra1, Gal11, Ste12, Taf12 and Gal80 using ¹²⁵ I-PEAS	Reeves and Hahn, 2005
Asp871	Adjacent to residue shown to crosslink to Tra1, Gal11, Ste12, Taf12 and Gal80 using ¹²⁵ I-PEAS	Reeves and Hahn, 2005

Table 2-1: Residues selected for pBpa incorporation along Gal4 TAD.

work done by Ptashne et. al. that suggested this residue was important for Gal4-Gal80 binding.¹³ They found that mutating Thr859 to cysteine resulted in a decrease in Gal4-Gal80 interaction in a protein gel shift experiment. Similarly, when Phe856 was mutated to cysteine, a reduction in Gal4-Gal80 gel shifting was observed, indicating that this residue likely makes important contacts with Gal80.¹³ In the same study they observed that point mutations at residues Thr860, Val864 and Leu868 to Pro resulted in a decrease in Gal80-mediated transcriptional repression, suggesting that these positions also play an important role in mediating the Gal4-Gal80 complex. We therefore selected residues adjacent to Thr860 or Leu868, those thought to be important for maintaining Gal4-Gal80 interaction (Met861, Tyr867, and Asp869). Further, Reeves and Hahn demonstrated through in vitro crosslinking with ¹²⁵I-PEAS at positions Phe869, Asp870 and Asp873 crosslinking with Tra1, Gal11, Ste12, Taf12 and Gal80, although with varying intensities. Crosslinking to Gal80 was also observed when the crosslinker was attached at Asp871.³⁶ This data supported selecting Phe869 and Asp871 for pBpa crosslinking. Other residues, outside the middle region Gal4(851-871) that makes-up the Gal80 binding site, were chosen to study other proteins targets of Gal4 outside of this region.¹³

D.2 p-benzoyl-L -phenylalanine incorporation along Gal4 TAD

Applying the same procedure used for Phe849pBpa incorporation we observed full-length protein in the presence of pBpa at all positions tested (Figure 2-8). The incorporation efficiency was not equal at all positions tested. It appears that position on the protein (N-terminal versus middle region versus C-terminal) has a greater influence on the incorporation rather than an individual residue being more permissive to pBpa incorporation over another. This is exemplified by Phe869 having significantly less full-length protein than Phe849 despite them both being phenylalanines (Figure 2-8).

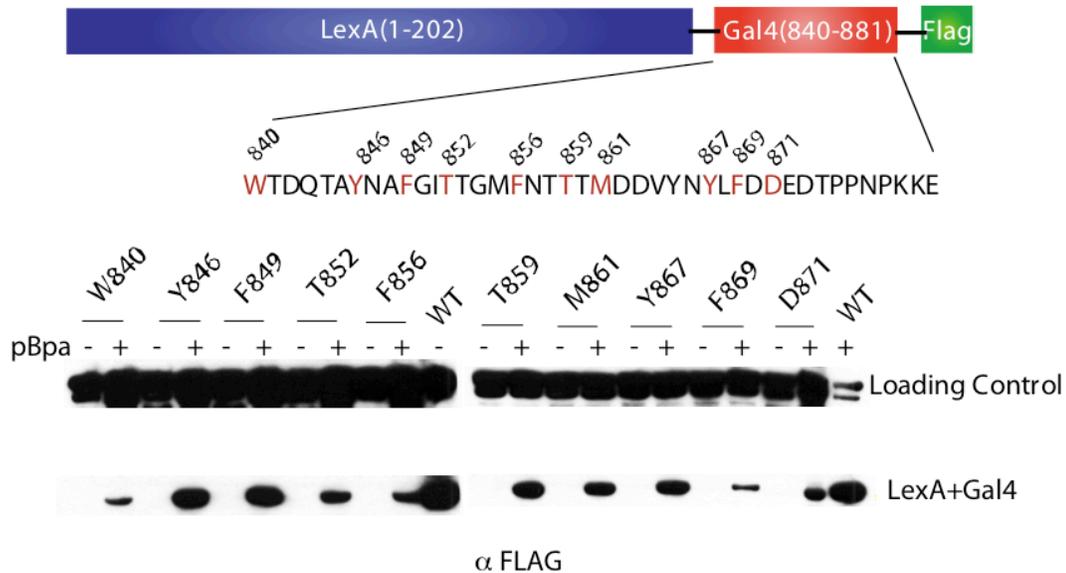


Figure 2-8: Site-specific incorporation of pBpa at positions along the Gal4 TAD.¹¹

E. Functional impact of incorporating *p*-benzoyl-L -phenylalanine into Gal4 Transcriptional activation domain

Next, we examined the functional impact of incorporating pBpa into all positions along LexA+Gal4(840-881) using quantitative β -galactosidase assays. The interaction of the Gal4 TAD with Gal80, an inhibitor protein that interacts with Gal4 when glucose is present, is tighter and more specific than the Gal4-transcriptional machinery interactions. To assess if this interaction was impacted upon pBpa incorporation, the transcriptional activity of each mutant protein was assessed in media containing different sugar sources. When glucose is present Gal4 is repressed by Gal80 and when galactose is present Gal4 is able to activate transcription as Gal80 no longer represses Gal4. What was observed was that under glucose or raffinose growth conditions, little transcriptional activity was observed, indicating that the mutant retained its ability to interact with Gal80 and thus maintain its environmental sensitivity. When galactose was present transcriptional activity increases as it is free to interact with the necessary proteins for up-regulating transcription (Figure 2-9). Based on this data, Gal4 maintains its function when pBpa is incorporated at all positions tested. Also, it was concluded that little read-through occurs since truncation of

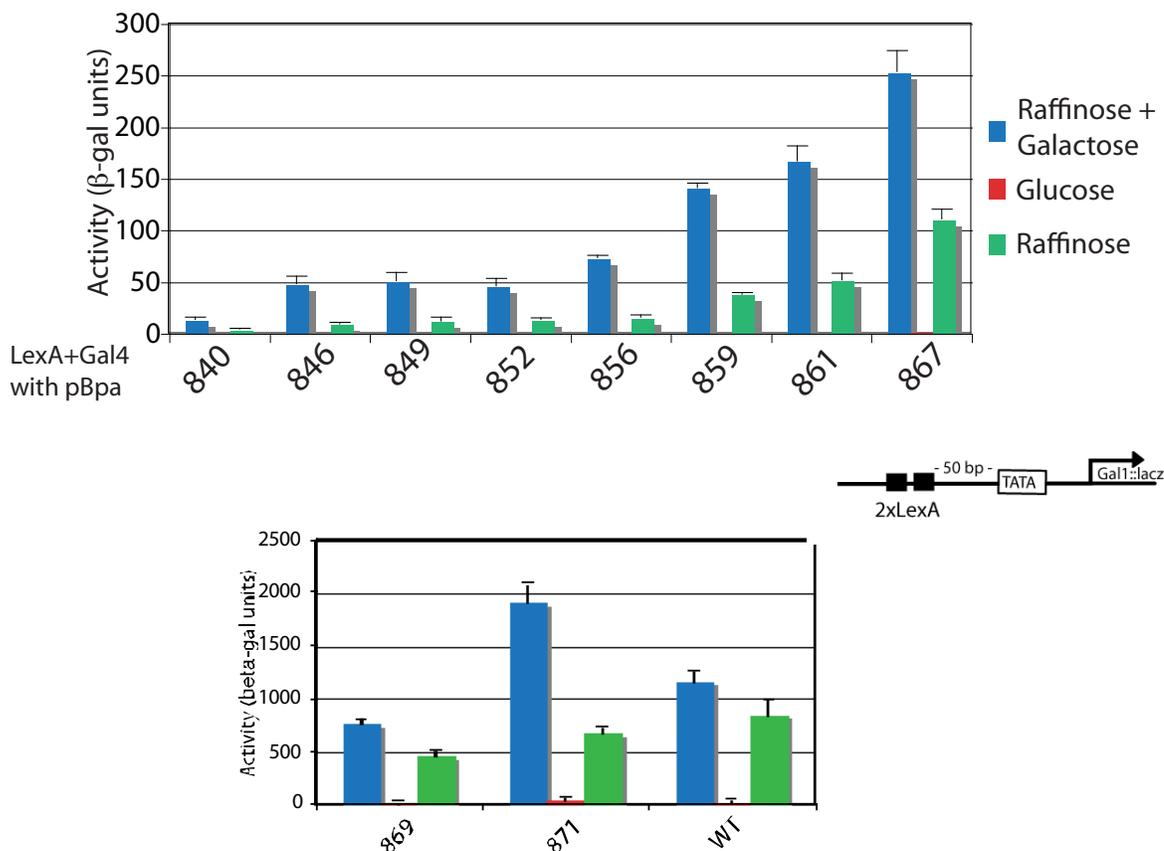


Figure 2-9: β-galactosidase assays showing the functional impact of incorporating pBpa into Gal4 TAD. β-gal assays at all positions tested for pBpa incorporation with different sugar sources.¹¹

protein synthesis at residue 849 would result in a protein with little transcriptional activity;³³ thus activity observed in yeast grown without pBpa would be an indicator of read-through.

F. In vivo photo-crosslinking with Gal4 transcriptional activation domain

The incorporation and functional results encouraged us to move forward in trying to capture the Gal4-Gal80 interaction in vivo by applying photo-crosslinking. Being able to capture a known binding partner of Gal4 would demonstrate the utility of this technique for capturing TAD-protein interactions in vivo. To capture Gal4-Gal80 interactions, live yeast cells expressing the pBpa

mutants were irradiated at 365 nm to create direct Gal4-protein crosslinks (Figure 2-10). Upon lysis and Western blot analysis we observed that all Gal4 mutants

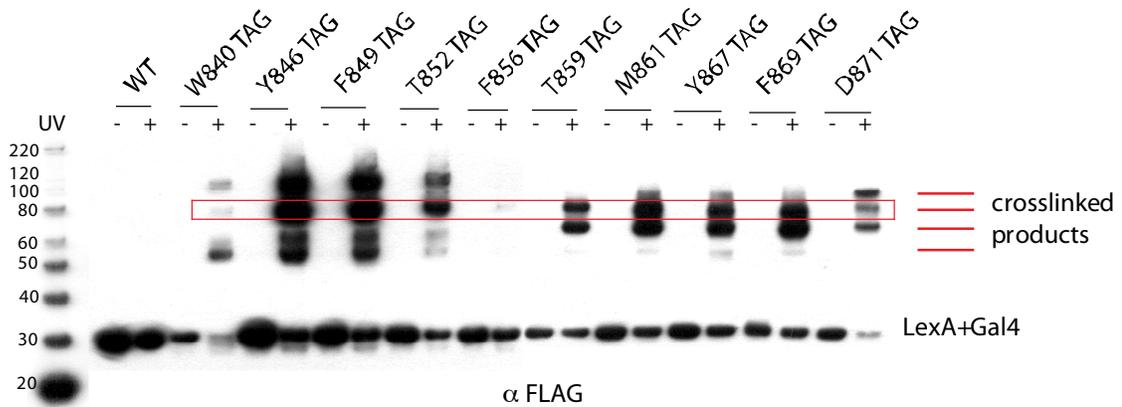


Figure 2-10: In vivo crosslinking of pBpa-containing Gal4 mutants. Only in the presence of UV light are crosslinking products observed. One band is present at all positions at a molecular weight of 80 kDa indicative of Gal4•Gal80 crosslinked product (red box).¹¹

formed several crosslinked products, consistent with the multi-partner binding model proposed for their function.³⁴⁻³⁸ More specifically, all formed a crosslinked product with a molecular weight of ~80 kDa, consistent with a LexA+Gal4•Gal80 product (Figure 2-10, red box). This was confirmed by the introduction of a c-Myc-tagged version of Gal80 into yeast and visualization of the cross-linked products with a c-Myc antibody (Figure 2-11a). Further supporting this, incorporation of pBpa into Gal80 at residue 245 leads to crosslinking with LexA+Gal4 (Figure 2-11b). Although all mutants formed crosslinks with Gal80, the extent varied with

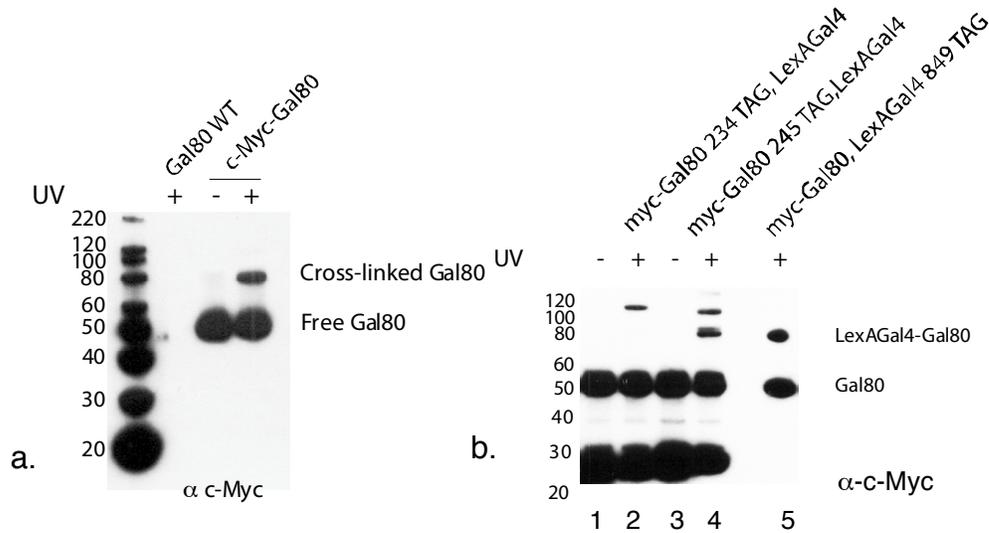


Figure 2-11: Gal4 crosslinks to c-Myc-Gal80 in vivo. a. yeast were cotransformed with a plasmid coding c-Myc-Gal80 and Gal4 with pBpa at position 849. Crosslinking is only observed when both constructs are present and UV light is applied. b. yeast were cotransformed with a c-Myc-Gal80 containing a site for pBpa incorporation and wild-type Gal4 (lanes 1-4) or the same two plasmids used in a (lane 5). Crosslinking was observed in one of the two positions selected for pBpa incorporation on Gal80 (lane 4) at the same molecular weight of the crosslinking experiment in a (lane 5).¹¹

position. Seven of the positions lie within the previously identified binding site for Gal80 (852, 856, 859, 861, 867, 869 and 871).¹³ Most striking is the Phe856pBpa mutant in which little crosslinked product is observed despite considerable evidence that this residue makes key contacts with Gal80.^{13,30} This is not due to attenuated binding affinity, as the pBpa-containing TAD exhibits a K_D for Gal80 nearly identical ($0.7 \pm 0.2 \mu\text{M}$) to that of the native sequence ($1.2 \pm 0.1 \mu\text{M}$) as demonstrated by my coworker Dr. Chinmay Majmudar with fluorescence polarization of Gal4 TAD peptides containing pBpa with interacting with Gal80. In this instance, the benzophenone moiety at position 856 may be poorly positioned for cross-linking, particularly as the efficiency of benzophenone crosslinking is well-known to vary with amino acid identity (discussed in detail in Chapter 4).⁴³ Similarly, the enhanced cross-linking observed with the Thr852pBpa and Tyr867pBpa mutants is not due to an altered affinity for Gal80 (Thr852pBpa: $0.8 \pm 0.3 \mu\text{M}$; Tyr867pBpa: $0.8 \pm 0.1 \mu\text{M}$). Somewhat surprising was that positions beyond the previously defined Gal80 binding sequence (residues 851-871) also

exhibited effective cross-linking. At the amino terminus of the sequence, for example, the Phe849pBpa and Tyr846pBpa mutants both produce strong bands corresponding to the Gal4•Gal80 crosslinked product (Figure 2-10). To probe this further, we also examined pBpa mutations at positions Pro875 and Lys879 and upon irradiation of yeast bearing these mutants, crosslinking with Gal80 was observed (Figure 2-12). Crosslinking efficiency was attenuated with the

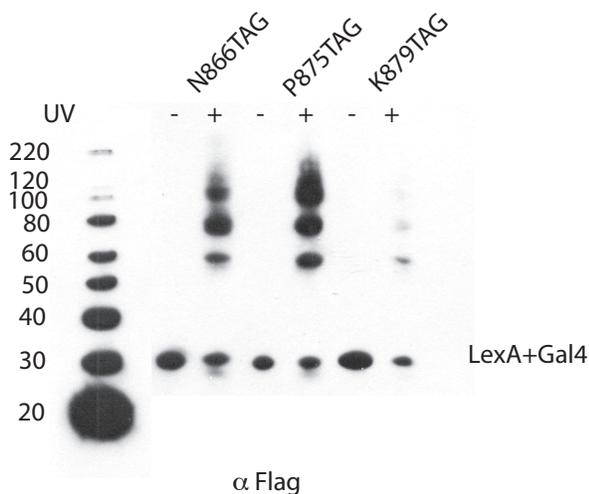


Figure 2-12: In vivo crosslinking with pBpa incorporated at positions near the N-terminal 866, 875, 879. A band at 80 kDa indicated that a more extended surface contacts Gal80 than previously thought (positions (851-871)).^{11,13}

Lys879pBpa mutant, similar to the Phe840pBpa mutant, and this may indicate limited binding interactions at the extreme amino- and carboxy-terminus of the Gal4 TAD. This cross-linking pattern reveals an extending binding interface between Gal80 and Gal4 not previously identified through conventional deletion and mutagenesis experiments, highlighting the higher resolution information accessible by this approach.

In conclusion, we showed the facile incorporation of the nonnatural and photoactivatable amino acid pBpa into a transcription factor (Gal4) in vivo using a modified tRNA/synthetase system. Importantly, the function and binding interactions of Gal4 remain largely intact with this substitution as all mutant proteins retained their ability to activation transcription and to respond to environmental cues. Further, crosslinking in vivo enabled capture of a key

masking protein, Gal80, and characterization of the binding interface with this protein. The improved expression yields observed with this system will be particularly valuable for identifying novel and functionally relevant binding partners of transcription factors that often have high turnover and are present in low abundance inside the cell.

G. Methods

LS41 [JPY9::pZZ41, *Mat α* *his3 Δ 200 leu2 Δ 1 trp1 Δ 63 ura3-52 lys2 Δ 385 gal4 URA::pZZ41] yeast was used for all experiments. pBpa was purchased from Chem-Impex International (Wood Dale, IL). All plasmids described below were constructed using standard molecular biology techniques and the sequences of all the isolated plasmids were verified by sequencing at the University of Michigan Core Facility (Ann Arbor, MI).*

Table of Plasmids used in this study

Plasmid name	Function
pLexAGal4	Expresses LexA(1-202)+Gal4(840-881)+FLAG tag
pLexAGal4 840TAG, pLexAGal4 846TAG, pLexAGal4 849TAG, pLexAGal4 852TAG, pLexAGal4 856TAG, pLexAGal4 859TAG, pLexAGal4 861TAG, pLexAGal4 867TAG, pLexAGal4 869TAG, pLexAGal4 871TAG	Express LexA(1-202)+Gal4(840-881)+FLAG tag with a TAG replacing the codon of the existing amino acid
pSNRtRNA-pBpaRS	Expresses tRNA under the control of the SNR52 promoter and contains synthetase specific for pBpa
ptRNA-pBpaRS	Expresses tRNA with no eukaryotic Pol III promoter and contains synthetase specific for pBpa ¹⁸
pMycGal80	Expresses Gal80 fused to c-Myc tag
pMycGal80 234 TAG, pMycGal80 245 TAG	Expresses Gal80 fused to c-Myc tag with a TAG replacing the codon of the existing amino acid
pmOCR-Gal80	Expresses Gal80 fused to the mOCR solubility tag in <i>E. coli</i>

pLexAGal4

A high copy plasmid expressing LexA(1-202)+Gal4(840-881)+FLAG tag under the control of the ADH1 promoter was created from pNLexA (Origene). First, site-directed mutagenesis was used to mutate the existing EcoRI and BamHI sites at the N-terminus of LexA and subsequently to insert EcoRI and BamHI sites at the C-terminus of LexA, producing the plasmid pCLexA. Primers (5'- TTA CGA ATT CTG GAC GGA CCA AAC TG -3') and (5'- AGT GGA TCC TTA TTT GTC GTC GTC GTC TTT ATA GTC CTC TTT TTT TGG G -3') were used to amplify Gal4(840-881) from EGY48 yeast genomic DNA. The amplified PCR product was digested with EcoRI and BamHI and inserted into pCLexA digested with EcoRI and BamHI and calf intestinal phosphate treated to create pLexAGal4.

pLexAGal4 840TAG, pLexAGal4 846TAG, pLexAGal4 849TAG, pLexAGal4 852TAG, pLexAGal4 856TAG, pLexAGal4 859TAG pLexAGal4 861TAG, pLexAGal4 867TAG, pLexAGal4 869TAG and pLexAGal4 871TAG

Plasmids containing various amber mutants in the Gal4 TAD were derived from pLexAGal4. To create each plasmid, site-directed mutagenesis was used to replace an existing amino acid codon with TAG within the Gal4 TAD. In general, PCR primers were designed to have ~15 bases of homology on either side of the TAG mutation. QuikChange (Stratagene, La Jolla, CA) was used to incorporate the TAG mutants using manufacturer recommended conditions.

pSNRtRNA-pBpaRS

The previously described pSNRtRNA-TyrRS plasmid incorporating tyrosine at the amber position was used to generate a plasmid pSNRtRNA-pBpaRS to incorporate pBpa at the amber position.²² This plasmid was generated by insertion of the pBpa-specific *E. coli* tyrosyl synthetase (amplified from ptRNA-pBpaRS (p-benzoylPheRS-2)¹⁸ obtained from Dr. P. G. Schultz, Scripps Research Institute, La Jolla, CA) using primers 5'- AGT TCA ACT AGT ATG GCA AGC AGT AAC TTG ATT -3' and 5'-TCG ATC TCG AGT TAT TTC CAG CAA

ATC AGA CA-3') into SpeI and XhoI digested and calf intestinal phosphate treated pSNR TyrRS.

pMycGal80, pMycGal80234TAG and pMycGal80245TAG

A plasmid expressing Gal80 fused to the c-Myc tag under the control of the ADH1 promoter was generated by insertion of DNA encoding *S. cerevisiae* Gal80 into the high copy plasmid pADT7 (Clontech, Mountain View, CA). Primers 5'-TGT GAA GCT TAT GGA ACA AAA GTT GAT TTC TGA AGA AGA TTT GGA CTA CAA CAA GAG ATC TTC G -3' and 5' CGT CAA GCT TTT ATA AAC TAT AAT GCG AG -3' were used to amplify Gal80 from LS41 yeast genomic DNA. The amplified PCR product was digested with HindIII and inserted into HindIII digested and calf intestinal phosphate treated pADT7. pMycGal80234TAG and pMycGal80245TAG were constructed by performing site-directed mutagenesis to incorporate a TAG codon to replace the codon of the existing amino acid

Incorporation of pBpa into LexA(1-202)+Gal4(840-881)

LS41 yeast was transformed with various pLexAGal4 TAG mutant plasmids and pSNRtRNA-pBpaRS or ptRNA-pBpaRS. Individual colonies were grown in 5 mL SC media containing 2% glucose but lacking histidine and tryptophan for selection. The cultures were incubated overnight at 30 °C and agitated at 250 rpm. Following incubation, these cultures were used to inoculate 5 mL cultures of SC media containing 2% glucose, with or without 2 mM pBpa (dissolved in 50 µL of 1M NaOH), and 50 mL 1M HCl which were subsequently incubated overnight at 30 °C with agitation to an OD₆₆₀ of 1.5. 3 ODs of cells were isolated, washed with cold, sterile water and stored dry at -20 °C. The samples were lysed in 10 mL 4x NuPAGE LDS Sample buffer (Invitrogen), 15 mL Lysis Buffer (50 mM Tris-Acetate, pH 7.9, 150 mM KOAc, 20% glycerol, 0.2% Tween-20, 2 mM b-mercaptoethanol, 2 mM MgOAc) 5 mL 1M DTT and analyzed using Western blot with the anti-FLAG(M2) antibody (Sigma).

In vivo cross-linking

To perform in vivo cross-linking, individual colonies of each pLexAGal4 TAG mutant were grown in 5 mL SC media containing 2% glucose but lacking histidine

and tryptophan for selection. The cultures were incubated overnight at 30 °C and agitated at 250 rpm. Following incubation, these cultures were used to inoculate 50 mL cultures of SC media containing 2% glucose, with 2 mM pBpa (dissolved in 0.5 mL of 1M NaOH), and 0.5 mL 1M HCl, which were subsequently incubated overnight at 30 °C with agitation to an OD₆₆₀ of 1.5. For each mutant, 50 ODs of cells were isolated, washed with water and either resuspended in 2 mL water and irradiated for 1 h with 365 nm light (Eurosolar 15 W UV lamp) with cooling or kept in the dark at 4 °C. Following irradiation, all the cells were pelleted and stored at -80 °C until lysis. For lysis, cells were resuspended in 600 mL Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche) and lysed using glass beads by vortexing at 4 °C. Subsequently, the lysate was pelleted and the supernatant incubated with 10 mL of LexA antibody (N-19, Santa Cruz Biotechnologies) for 2 h at 4 °C for immunoprecipitation. The protein bound to the antibody was isolated by incubation for 1 h with ~50 mL of prewashed protein G magnetic beads (DynaL Corporation, Invitrogen, Carlsbad, CA) at 4 °C. The beads were washed 3X with 1 mL Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 1% Na-Deoxycholate and 1 mM EDTA) and stored dry at -20 °C. The protein was eluted from the beads by heating at 95 °C for 10 min in NuPAGE LDS Sample buffer (Invitrogen, Carlsbad, CA) containing DTT and probed using Western Blot analysis using anti-FLAG (M2) antibody (Sigma, St. Louis, MO).

β-Galactosidase assays

To evaluate the ability of each LexA+Gal4 TAG mutant in the absence or presence of 2 mM pBpa to activate transcription, saturated cultures (SC media + 2% raffinose) of each mutant were used to inoculate 5 mL SC media lacking histidine and tryptophan supplemented with 2% glucose or 2% raffinose + 2% galactose or 2% raffinose and grown to an OD of 1.5-2.0 before being harvested. The activity of each construct was monitored using β-galactosidase assays as previously described.⁴⁴

Expression of Gal80

Expression of Gal80 fragments fused to the His₆-mOCR solubility tags was carried out in Rosetta2 (DE3) pLysS *E. coli* cells (Novagen). Briefly, cultures (50 mL) from single colonies were grown overnight at 37°C (300 rpm) in Select APS Super Broth (Difco) supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) before addition to 1L of Select APS Super Broth supplemented with ampicillin (100 µg/mL). After an OD₆₀₀ of 0.8 was reached, the cultures were cooled on ice for 30 m, and expression was induced with IPTG (final concentration 0.1 mM) for ~14 h at 20 °C. The cell pellet was resuspended in Buffer A (100 mM Tris pH 7.5 at 4 °C, 10% glycerol (v/v), 150 mM NaCl, 10 mM β-ME and Roche Complete Protease Inhibitor Cocktail), lysed using sonication, and the His-tagged protein was isolated using Ni NTA-Agarose (Qiagen). The Ni-NTA beads were washed 6 times with Wash buffer (100 mM Tris pH 7.5 at 4 °C, 10% glycerol (v/v), 150 mM NaCl, 10 mM β-ME, 30 mM imidazole). The protein was eluted from the beads at 4 °C 3 times using Elution buffer (100 mM Tris pH 7.5 at 4 °C, 10% glycerol (v/v), 150 mM NaCl, 10 mM β-ME, 300 mM imidazole). The resulting mixture was placed in dialysis tubing (Pierce) and the buffer exchanged to a low salt buffer (50 mM Tris pH 7.5 at 4 °C, 10% glycerol (v/v), 50 mM NaCl, 10 mM β-ME) overnight at 4 °C. The solution thus obtained was loaded onto an anion-exchange column (Q sepharose, GE Healthcare) and eluted with a NaCl gradient from 0-1 M. Fractions containing Med15(1-416) were pooled and buffer exchanged to Storage buffer (10 mM PBS pH 7.4, 10% glycerol, 0.01% NP-40, 1 mM DTT) using a PD-10 column (GE Healthcare) and concentrated using a 30K centrifugal filter device. The protein concentration was measured using absorbance at 280 nm. The identity and purity of the protein was verified by reducing SDS-PAGE with appropriate molecular weight standards.

Peptide Synthesis

Gal4(840-870) peptides with pBpa at various positions were synthesized using solid phase peptide synthesis in accordance with standard protocols. Fmoc-

pBpa was purchased from Chem-Impex International. The peptides contained a b-alanine linker at N-terminus and they were labeled using fluorescein isothiocyanate on solid phase and subsequently cleaved using 95% TFA, 2.5% TIS, 2.5% H₂O. The products were purified to homogeneity using reversed-phase HPLC on a C18 column with a gradient solvent system (buffer A: 20 mM Ammonium Acetate, buffer B: MeOH) and stored at -80 °C. The identity was verified using electrospray mass spectrometry (LCT Micromass).

Fluorescence Polarization

Dissociation constants for fluorescein labeled Gal4 TADs and Gal80 were determined using fluorescence polarization as previously described. For each experiment, 50 nM TAD was incubated with varying concentrations of Gal80 in 100 mM PBS pH 7.2, 10% glycerol, 0.01% NP-40, 1 mM DTT and incubated for 10 min in 384 well low volume plate (Corning) before being detected using a TECAN Genios Pro plate reader and the dissociation constant calculated as previously described.

H. References

- (1) Fuentes, G.; Oyarzabal, J.; Rojas, A. M. *Curr Opin Drug Discov Devel* **2009**, *12*, 358.
- (2) Tanaka, Y.; Bond, M. R.; Kohler, J. J. *Mol Biosyst* **2008**, *4*, 473.
- (3) Ruffner, H.; Bauer, A.; Bouwmeester, T. *Drug Discov Today* **2007**, *12*, 709.
- (4) Yin, H.; Hamilton, A. D. *Angew Chem Int Ed Engl* **2005**, *44*, 4130.
- (5) Wang, Q.; Parrish, A. R.; Wang, L. *Chem Biol* **2009**, *16*, 323.
- (6) Wang, L.; Xie, J.; Schultz, P. G. *Annu Rev Biophys Biomol Struct* **2006**, *35*, 225.
- (7) Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. *Nat Methods* **2005**, *2*, 201.
- (8) Ye, S.; Kohrer, C.; Huber, T.; Kazmi, M.; Sachdev, P.; Yan, E. C.; Bhagat, A.; RajBhandary, U. L.; Sakmar, T. P. *J Biol Chem* **2008**, *283*, 1525.
- (9) Huang, L. Y.; Umanah, G.; Hauser, M.; Son, C.; Arshava, B.; Naider, F.; Becker, J. M. *Biochemistry* **2008**, *47*, 5638.
- (10) Mohibullah, N.; Hahn, S. *Genes Dev* **2008**, *22*, 2994.

- (11) Majmudar, C. Y.; Lee, L. W.; Lancia, J. K.; Nwokoye, A.; Wang, Q.; Wands, A. M.; Wang, L.; Mapp, A. K. *J Am Chem Soc* **2009**, *131*, 14240.
- (12) Morrison, K. L.; Weiss, G. A. *Curr Opin Chem Biol* **2001**, *5*, 302.
- (13) Ansari, A. Z.; Reece, R. J.; Ptashne, M. *Proc Natl Acad Sci U S A* **1998**, *95*, 13543.
- (14) Lefevre, F.; Remy, M. H.; Masson, J. M. *Nucleic Acids Res* **1997**, *25*, 447.
- (15) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.
- (16) Chen, D.; Texada, D. E. *Gene Therapy and Molecular Biology* **2006**, *10*, 1.
- (17) Drabkin, H. J.; Park, H. J.; RajBhandary, U. L. *Mol Cell Biol* **1996**, *16*, 907.
- (18) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964.
- (19) Chen, H. T.; Warfield, L.; Hahn, S. *Nat Struct Mol Biol* **2007**, *14*, 696.
- (20) Chen, S.; Schultz, P. G.; Brock, A. *J Mol Biol* **2007**, *371*, 112.
- (21) Kowal, A. K.; Kohrer, C.; RajBhandary, U. L. *Proc Natl Acad Sci U S A* **2001**, *98*, 2268.
- (22) Wang, Q.; Wang, L. *J Am Chem Soc* **2008**, *130*, 6066.
- (23) Mapp, A. K.; Ansari, A. Z. *ACS Chem Biol* **2007**, *2*, 62.
- (24) Klar, A. J.; Halvorson, H. O. *Mol Gen Genet* **1974**, *135*, 203.
- (25) Lohr, D.; Venkov, P.; Zlatanova, J. *FASEB J* **1995**, *9*, 777.
- (26) Wightman, R.; Bell, R.; Reece, R. J. *Eukaryot Cell* **2008**, *7*, 2061.
- (27) Han, Y.; Kodadek, T. *J Biol Chem* **2000**, *275*, 14979.
- (28) Lue, N. F.; Chasman, D. I.; Buchman, A. R.; Kornberg, R. D. *Mol Cell Biol* **1987**, *7*, 3446.
- (29) Ma, J.; Ptashne, M. *Cell* **1987**, *50*, 137.
- (30) Thoden, J. B.; Ryan, L. A.; Reece, R. J.; Holden, H. M. *J Biol Chem* **2008**, *283*, 30266.
- (31) Kumar, P. R.; Yu, Y.; Sternglanz, R.; Johnston, S. A.; Joshua-Tor, L. *Science* **2008**, *319*, 1090.
- (32) Ptashne, M.; Gann, A. *Genes & signals*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2002.
- (33) Ma, J.; Ptashne, M. *Cell* **1987**, *48*, 847.
- (34) Jeong, C. J.; Yang, S. H.; Xie, Y.; Zhang, L.; Johnston, S. A.; Kodadek, T. *Biochemistry* **2001**, *40*, 9421.
- (35) Melcher, K.; Johnston, S. A. *Mol Cell Biol* **1995**, *15*, 2839.
- (36) Reeves, W. M.; Hahn, S. *Mol Cell Biol* **2005**, *25*, 9092.
- (37) Chang, C.; Gonzalez, F.; Rothermel, B.; Sun, L.; Johnston, S. A.; Kodadek, T. *J Biol Chem* **2001**, *276*, 30956.
- (38) Bryant, G. O.; Ptashne, M. *Mol Cell* **2003**, *11*, 1301.
- (39) Wu, Y.; Reece, R. J.; Ptashne, M. *EMBO J* **1996**, *15*, 3951.

- (40) Archer, C. T.; Burdine, L.; Kodadek, T. *Mol Biosyst* **2005**, *1*, 366.
- (41) Bhaumik, S. R.; Green, M. R. *Methods Enzymol* **2003**, *370*, 445.
- (42) Gonzalez, F.; Delahodde, A.; Kodadek, T.; Johnston, S. A. *Science* **2002**, *296*, 548.
- (43) Wittelsberger, A.; Thomas, B. E.; Mierke, D. F.; Rosenblatt, M. *FEBS Lett* **2006**, *580*, 1872.
- (44) Majmudar, C. Y.; Lum, J. K.; Prasov, L.; Mapp, A. K. *Chem Biol* **2005**, *12*, 313.

Chapter 3

Identifying the direct binding partners of the Gal4 transcriptional activation domain via in vivo crosslinking and MudPIT analysis

A. Project Overview

In the previous chapter, it was demonstrated that capturing a known protein-protein interaction, Gal4-Gal80, is possible using an in vivo crosslinking strategy with a genetically encoded crosslinker.¹ Initially, we used Western blot analysis of the crosslinked proteome for identifying the crosslinked products; however, this method is limited since one must have some prior knowledge regarding likely binding partners in order to appropriately select antibodies. In addition, many of the potential protein binding partners of transcriptional activators do not have available antibodies (commercial or academic). Because the goal of our study is to use in vivo crosslinking for identifying the complete suite of direct binding partners of Gal4 in order to resolve the conflicting models surrounding activator-initiated transcription, we needed to analyze our crosslinked sample by a different and more unbiased method. Towards this end, in vivo photo-crosslinking of Gal4 coupled with mass spectrometry (MS) analysis of the resulting crosslinked proteome was undertaken. In this Chapter, we demonstrate that this technique was used to identify nine binding partners of Gal4, including several previously unidentified binding partners.

B. Background

Mass spectrometry was selected for this study over other methods because it is a proven to be a highly sensitive and powerful method to study proteins in complex samples. To identify the direct protein targets of Gal4 using in vivo photo-crosslinking followed by mass spectrometry it was necessary to overcome a number of technical hurdles. Outlined in Figure 3-1 is

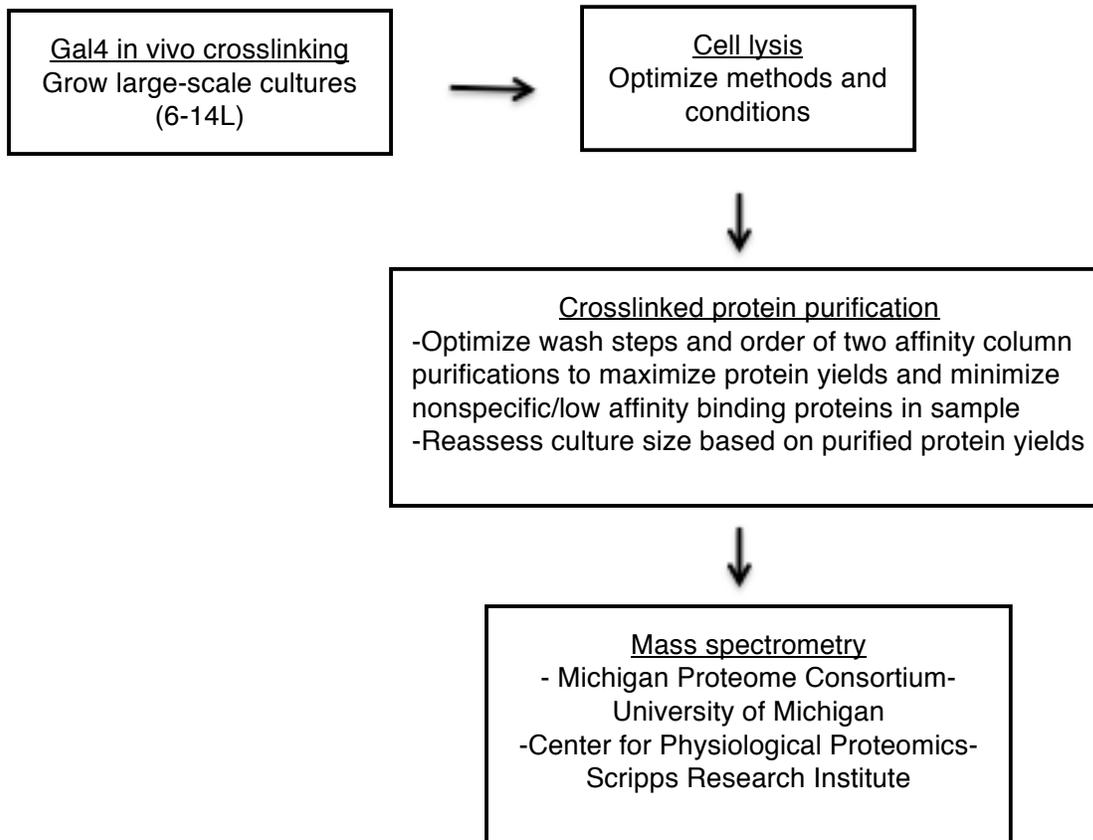


Figure 3-1: General strategy for preparing a photo-crosslinked of Gal4 for mass spectrometric analysis.

a general scheme for preparing a sample for mass spectrometric analysis with crosslinked Gal4. A major challenge in these experiments is that transcription proteins are typically low abundant proteins (Table 3-1), so large culture volumes are required. Even though we typically utilize a high copy plasmid expressing Gal4 from a strong promoter we are incorporating a nonnatural amino acid which adds to the low abundance issue, as these proteins typically express only ~20% of what wild-type protein expressed under the same conditions. In addition, activators such as Gal4 are often rapidly degraded. Developing optimal conditions for lysing procedures with such large volumes (6-12L) is also difficult as yeast are notoriously difficult to lyse.² Since TADs are known to have a multipartner binding profile, one would expect that many proteins are crosslinked. An additional challenge is that effective washing conditions in the

immunoprecipitation steps must be identified in order to minimize the chance of indirect binding partners or proteins that interact with the solid support being in

Protein	Copies/cell
Gal4	166
Swi1	92.3
Snf2	217
Swi2	217
Snf5	217
Rsc1	259
Cdc73	538
Gal11 (Med15)	606
Gal80	784
Taf12	930 ± 45
GCN5	1180
Caf1	1520
Spt3	1710
Spt3	1710
Ada2	1720
Sin4/Med16	1720
Spt7	2360
Ada3	2470
Ccr4	2780
Taf10	3370
Dbf2	3500
Spt20	4150
Ada5	4150
Spt20	4150
Sug1	4700
Rox3	6120
Taf6	6510 ± 1290
Taf9	6580
Ada1	7950
Med2	10,800
Taf5	14800 ± 203
Mbf1	47300
TBP	low signal
Srb9	low signal
Tra1	not visualized
SRB10	not visualized
Hpr1	not visualized
Sug2	not visualized

Table 3-1: Protein copies/cell of potential targets of amphiphatic activators (discussed in Chapter 1) based on the yeast GFP fusion database (<http://yeastgfp.yeastgenome.org/>). Low abundance proteins are <5,000 copies/cell.³ Not visualized means either the tagging was unsuccessful or no signal was detected. Low signal means tagging was successful but the signal was not sufficiently high above background to permit accurate quantitation.

the final sample to be analyzed. Finally, choosing a MS ionization strategy is important since the sample is complex and contains low abundance proteins.

B.1 Mass spectrometric analyses of complex proteomes

To analyze complex protein sample mixtures many analytical techniques involving mass spectrometry have been developed.⁴⁻⁷ These techniques vary with regards to their degree of reproducibility, resolution/sensitivity, throughput, biases and limitations. High resolution and sensitivity with little bias and limitations are desired for analyzing a sample where protein composition is unknown. This will maximize protein coverage so that both low and high abundance proteins are identified. It is also important to consider the reproducibility and throughput of the technique used. Low throughput on complex samples may limit the data collected and ultimately may be unreasonable for such studies.

B.1.a MALDI versus ESI for analyzing proteomes

There are primarily two methods used to ionize biomolecule for MS analysis. Matrix-assisted laser desorption/ionization (MALDI) is a one method for ionization and another used in proteomic analysis is electrospray ionization (ESI). MALDI is a solid-state method where the sample is dried on a metal matrix plate and a laser is used to vaporize the sample.⁸ ESI is flow-based where a sample in solution is aerosolized and droplets become ionized. The liquid in the drops evaporate producing gas-phase ions.⁸ A major advantage of ESI over MALDI is that the flow source can come directly from an attached liquid chromatography column, making it very convenient for analysis of complex mixtures that can be resolved prior to analysis. To achieve a similar type of experiment using MALDI, plate spotters can be coupled to a liquid chromatography system. However, the quality of the separation will be attenuated because several spectra of data will be “summed” in collected fractions rather than a sample coming directly off the separation system to the ionizer where no fractions are collected for analysis. A

disadvantage of ESI is that a sample cannot be reanalyzed, which is not the case with MALDI. Some researchers use both techniques to analyze a similar sample in a complementary approach.⁹⁻¹²

B.1.b. Top-down versus bottom-up proteomics

Two approaches are primarily used for studying biomolecules with MS. The bottom-up approach (Figure 3-2) is a widely used method that involves

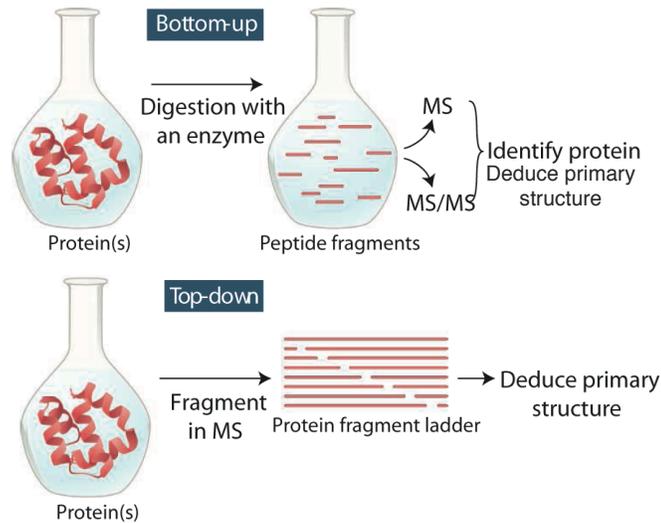


Figure 3-2: General schemes for bottom-up and top-down proteomics.¹³ In bottom-up approach (top panel) proteins are enzymatically digested before MS analysis. In top-down approach (bottom panel) intact proteins are analyzed by MS.

digesting a protein sample mixture to generate peptides that are then subjected to MS analysis. In this analysis, the peptide ions are fragmented, enabling peptide sequences to be determined, ultimately leading to protein identification. The peptide fragments are more easily solubilized and separated compared to the parent proteins which is an advantage of the bottom-up technique for protein identification. A disadvantage of this method is that only a fraction of the peptides are detected, so mapping post-translational modifications (PTMs) found on the parent protein is not straightforward.¹³ Conversely, in the top-down approach intact proteins are subjected to MS analysis where they fragment yielding identification of the protein.¹³ An advantage of this technique over the bottom-up approach is that both the primary structure of the protein and the PTMs can be

understood as long as enough key fragments are observed. A major disadvantage of this approach is that unlike the small peptides, it is difficult to generate significant gas-phase fragmentation ions of full-length protein, particularly of higher molecular weight proteins.¹³ For example it was considered a heroic feat when Han and coworkers broke the ~50 kDa limit upon demonstrating protein fragmentation on a >200 kDa protein.¹⁴ We have chosen the bottom-up approach because we are not initially looking to map protein modifications but simply identify proteins. Additionally, because some transcriptional proteins possess molecular weights near or exceeding the boundaries of top-down (Swi1 ~148 kDa, Snf2 ~194, Tra1 ~433), a bottom-up approach seemed to be a more compatible with our experiments.

B.2. Methods used for studying proteomic samples

Using the various methods of protein separation of complex mixtures coupled with different mass spectrometers the following methods have been used for proteomic studies. Table 3-2 outlines the strategies described in this section.

General strategy	2D separations	ionization technique	advantages/disadvantages	reference
2D PAGE/MS-MS	1. isoelectric point 2. molecular weight	ESI	Decent resolving capabilities/low throughput and little success for identifying low abundant proteins	Gygi et. al. (2000)
2D liquid chromatography/MS	1. isoelectric focusing 2. HPLC	MALDI	Improved sample capacity, more throughput than 2D PAGE/few proteins identified	Wall et. al. (2000)
2D liquid chromatography/MS	1. cation exchange 2. HPLC	ESI	Online method/little success at identifying proteins	Opiteck et. al. (1997)
2D liquid chromatography/MS	1. size exclusion liquid chromatography 2. HPLC	MALDI/ESI	High throughput/lower protein separation resolution than 2D PAGE, offline	Opiteck et. al. (1998)
2D liquid chromatography/MS-MS	1. strong cation exchange 2. HPLC	ESI	Online, high throughput, low abundant proteins identified	Link et. al. (1999) Washburn et. al. (2001)

Table 3-2: Strategies for proteomic analysis discussed in this sections.

B.2.a Two-dimensional polyacrylamide-gel electrophoresis followed by mass spectrometry

One method used to resolve and detect proteins in a complex mixture is two-dimensional polyacrylamide-gel electrophoresis (2D PAGE) followed by mass spectrometry (MS). Proteins are resolved in one dimension by isoelectric point (pI) and then by molecular weight (MW) in a second dimension.¹⁵ Proteins are typically visualized by staining the gel with an MS-compatible stain like Coomassie blue. Once visualized, the stained proteins in the gel are excised and digested with a protease such as trypsin to generate peptide fragments that are then injected into a mass spectrometer for analysis. Although this technique has been effective for the identification of abundant proteins in a sample, when evaluated for its usefulness for analyzing a more complex sample containing protein of varying abundance, it was not sensitive enough for identifying low

abundant proteins (<0.1 codon bias discussed below).¹⁶ Additionally, the manual excision of the protein gel slices lowers the reproducibility and high throughput of 2D PAGE/MS analysis, something that is essential for obtaining scientific data on complex samples. To evaluate the general utility of 2D PAGE/MS for proteomic analysis Gygi and coworkers prepared a 500 μg sample of soluble yeast proteins.¹⁶ They separated the proteins on a 2D PAGE and observed $\sim 1,500$ spots on the gel. They analyzed one 4 cm^2 section where they cut 50 spots from the gel for MS analysis using ESI tandem mass spectrometry (MS-MS).¹⁶ Tandem mass spectrometry is used for protein identification because in the instrument peptides fragment along the backbone in predictable ways, enabling amino acids to be identified based on the mass of the individual fragments and the known mass of amino acids.^{17,18} In MS-MS an initial full survey scan is followed by several product ion scans where peptides are isolated for tandem mass spectrum generation that permits peptide sequencing based on the fragmentation observed in the spectrum (Figure 3-3). Once a peptide sequence

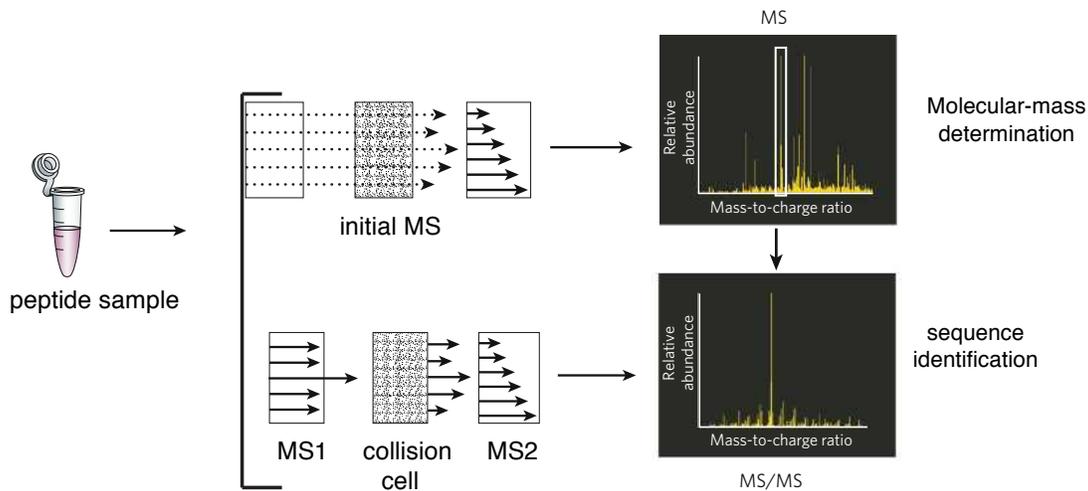


Figure 3-3: General MS-MS strategy for proteomic studies. The peptide ions are first measured by MS (upper panel) to determine the mass of the individual peptides. Each peptide is then isolated in the first mass analyzer (MS1) and moves onto the collision cell where they become fragmented. These fragments are directed to the second mass analyzer (MS2, lower panels).¹⁹ Reproduced with permission.

is obtained a database search is carried out to identify the protein from which the peptide comes from.¹⁷ They evaluated the sensitivity of using 2D PAGE with MS/MS based on the codon bias values. Codon bias values have been established for proteins and is defined as the extent to which a gene uses a subset of optimal codons with a scale of 0.0-1.0.²⁰ Higher expressed proteins have codon bias values of >0.2. In yeast about half of the proteins have a codon bias value of <0.1 and of the proteins identified in their study none of them had a codon bias value <0.1.¹⁶ They concluded that 2D PAGE followed by MS/MS is not a reliable technique for detecting low abundant proteins and in order to do so a much larger protein sample must be prepared to identify these proteins using this method. One limitation of 2D PAGE is sample capacity for loading onto a 2D PAGE gel. If a larger sample is required for low abundant proteins to be detected than is able to be loaded onto a 2D PAGE then sample capacity may be the limiting factor for analyzing these proteins with this method. Also, if an entire analysis of all proteins in the gel would have been carried out (~1,500 spots) it would have been extremely time consuming and very low throughput. Perhaps this is the reason why they only chose a 4cm² area to analyze. A low throughput technique with little success at identifying low abundant proteins is not conducive for large protein mixtures with low abundant proteins and therefore other strategies must be used.

B.2.b. Two-dimensional liquid chromatography

Other researchers explored different ways to resolve protein mixtures before they are analyzed by MS (Table 3-2). In particular, it was desirable to improve the 2D PAGE system by reducing the sample capacity limitations, increase throughput and sensitivity. To this end, 2D-liquid-chromatography followed by MS was developed. One advantage of liquid-phase chromatography over 2D PAGE for resolving proteins is that it has a high loading capacity (can be up to 50 times more than 2D PAGE).^{4-6,21} In an effort to improve throughput and sensitivity Wall et. al. separated proteins from human erythroleukemia cell line

lysate first by isoelectric focusing and then by nonporous reversed-phase high throughput liquid chromatography (HPLC), and analyzed these proteins by matrix-assisted laser desorption/ionization time-of-flight MS (MALDI-TOF MS). They were able to resolve ~700 bands with a rapid gradient and ~1000 bands with a longer gradient but were only able to identify 38 different proteins.⁶ They report an improved resolution of low mass and basic proteins over 2D PAGE protein separation they were not able to identify a significant amount of proteins (38 out of ~700). This method is, however, an improvement in throughput and sample capacity over 2D PAGE.

Another important aspect for improving systems for protein resolution and detection is whether the method used is off-line or on-line. This is an important consideration because it is related to throughput. Off-line is when the sample is handled after the resolving step and before it is injected into the mass spectrometer (MALDI is an example of this) and on-line is when there is no handling of the sample from the resolving step to injecting into the mass spectrometer (Figure 3-4). The latter was used by Opiteck and coworkers who

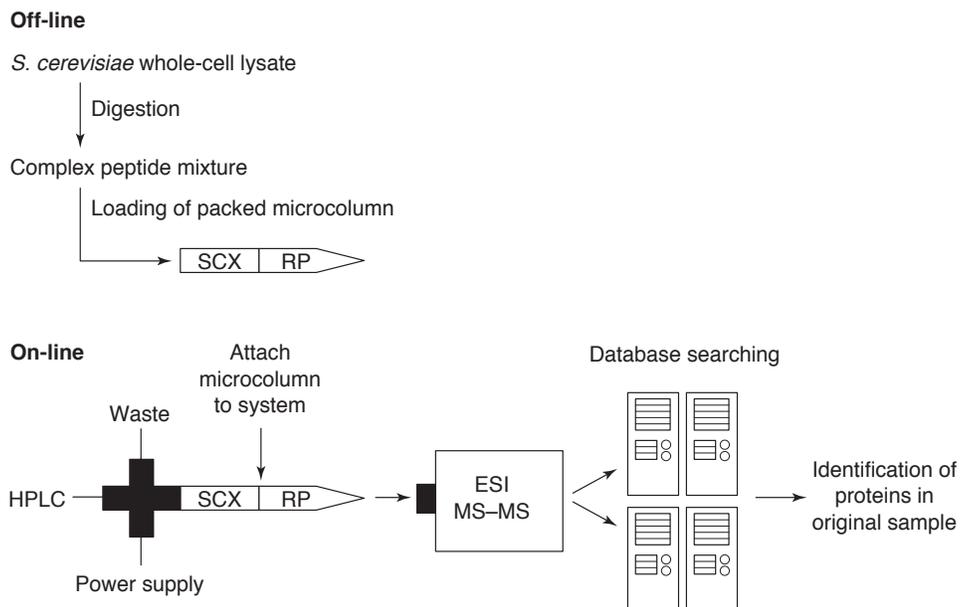


Figure 3-4: Example of off-line (top) and on-line (bottom) MS approaches. Off-line requires sample handling in between protein separation and the mass spectrometer. There is no sample handling once the sample is injected into the protein separation apparatus.²² Reproduced with permission.

developed a fully automated 2D liquid chromatography MS system.⁵ The proteins were separated by cation-exchange chromatography (CE) and reversed-phase HPLC and then detected by an electrospray mass spectrometer (ESI-MS).⁵ On-line systems are higher throughput than off-line systems and they are less prone to sample loss. Despite the advantages of this method, Opiteck and coworkers were able to identify proteins in a protein mixture containing 10 known proteins, but when *E. coli* lysate was analyzed they report one protein with a database search indicating the mass corresponds to 38 proteins within 2% of its molecular weight.⁵ Later the same group modified their method by resolving the protein mixture by size-exclusion liquid chromatography followed by reversed-phase HPLC and analyzed the proteins by either MALDI-TOF MS or ESI-MS.⁴ They identified 14 of the ~450 proteins isolated from *E. coli* lysates by MALDI-TOF MS and Edman sequencing.⁴ Overall, their method was high throughput, but it lacked the resolution that 2D gels possess, a direct interface to the mass spectrometer and automated protein identification technology.⁴

B.2.c Two-dimensional liquid chromatography followed by MS-MS

Perhaps the best strategy for characterizing complex proteomes was achieved by the Yates research group (Figure 3-4 (on-line)). Using a fully integrated method comprised of 2D liquid-chromatography followed by tandem mass spectrometry they were able to identify >100 proteins of a yeast lysate in a single run.⁷ They loaded trypsin-digested proteins onto a biphasic 2D microcapillary column. The peptide sample mixture was separated in the column by strong cation exchange (SCX) and reverse-phase (RP-HPLC). Once on the SCX resin the peptides were displaced through a salt step gradient onto the RP-HPLC resin where they were eluted directly onto an ESI-MS/MS. Before the next salt step in the SCX column the RP-HPLC resin was equilibrated, prepared for a new fraction of peptides.⁷ Peptide sequences were analyzed by the SEQUEST software to identify proteins.

Further optimization by the Yates lab led to increased utility and applicability of their front running technology, termed multidimensional protein identification technology (MudPIT), for detecting proteins in a complex mixture. Applying their method to a large-scale analysis of the yeast proteome they were able to identify 1,484 proteins, the most of anyone at the time.²³ This highly unbiased method yielded proteins identified from all subcellular compartments. They were able to capture proteins with extreme isoelectric points (pI), molecular weight, abundance and hydrophobicity.²³ In particular they evaluated their method in terms being able to identify low abundant proteins. To this end they looked at the codon adaptation index (CAI) of the proteins they were able to identify. CAI, similar to codon bias index, is defined as value that corresponds to cellular protein abundance and is the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes.²⁴ If a protein is coded with the most frequently used codons then it has a higher CAI value. CAI values range from 0.0-1.0 where 0.0-0.2 is related to low cellular protein abundance and values closer to 1.0 are associated with higher cellular protein abundance. It has been reported that 83% of the ~6,000 genes in the yeast genome have CAI levels between 0.0-0.2.²³ When the authors analyzed the proteins identified by their method in terms of their CAI, they identified proteins that have a CAI <0.2.²³ In this large-scale study 53.3% of the proteins had a CAI value <0.2, a major feat as other studies on yeast proteomes have had limited success at identifying proteins with this characteristic.²⁵⁻²⁷ Importantly, they report identifying 45 transcription factors. Amongst them were SNF5, SWI6 and SWI4, proteins that are found in the SWI/SNF complex with CAIs ranging from 0.12-0.15.²³ Based on the CAI analysis of their data, MudPIT may be applied for the study of biological processes involving low abundant proteins such as transcription.

In summary, recent advances in mass spectrometry for analyzing complex protein samples have reduced the limitations of past methods. The Yates laboratory has shown the utility of their method by collaborating with researchers

to help solve biological problems surrounding protein-protein interactions.²⁸⁻³⁰ Due to their proven success, a facility at Scripps Research Institute termed the Center for Physiological Proteomics was opened that uses MudPIT. MudPIT has been used to detect low abundant proteins, a characteristic required for studying transcription proteins therefore is of much interest to reaching our goal of identifying the direct targets of Gal4 TAD. In vivo photo-crosslinking followed by Western blot allowed us to identify a known binding partner of Gal4, Gal80, with little perturbation to the function of the activator proving this method is useful for capturing transcription related proteins. Moving forward, we needed a less biased and limiting method than Western blot for identifying the direct targets of Gal4. Because MudPIT is a technique possessing remarkable sensitivity for identifying low abundant proteins we decided to utilize this method for analyzing our crosslinked protein samples.

C. Preparing an in vivo photo-crosslinked sample for mass spectrometric analysis

Using in vivo crosslinking followed by MS, our initial goal was to observe and confirm the Gal4-Gal80 interaction that was observed on our Western blot. Characterizing Gal4-Gal80 interaction with these techniques would confirm the possibility of using MS for identifying other protein targets of Gal4. Before a sample could be submitted for mass spectrometric analysis an extensive optimization of a protein purification protocol was required for generating a MS sample with minimal nonspecific/low binding proteins that would be carried through the purification process. One challenging aspect of transcription proteins with relation to obtaining MS data, was their low abundance issue meaning that high yeast culture volumes were required for identifying the proteins of interest with MS (Table 3-1).³¹ Additionally, it was improbable that all of the copies of a particular protein target in the cell would be localized to the required proximity of Gal4 TAD for pBpa crosslinking, and, if they were, it was known that pBpa crosslinking was not 100% efficient due to required geometries and amino acid

preference which meant that only a fraction of the already low abundant potential protein targets in the cell would be captured by crosslinking.^{32,33} TADs' intrinsically unstructured character and multipartner binding profile made purifying crosslinked products for MS difficult as potentially many nonspecific proteins could interact with the TAD.³⁴

Many steps were optimized in the protein purification for preparing a MS sample containing low abundance crosslinked proteins (Figure 3-5). The initial

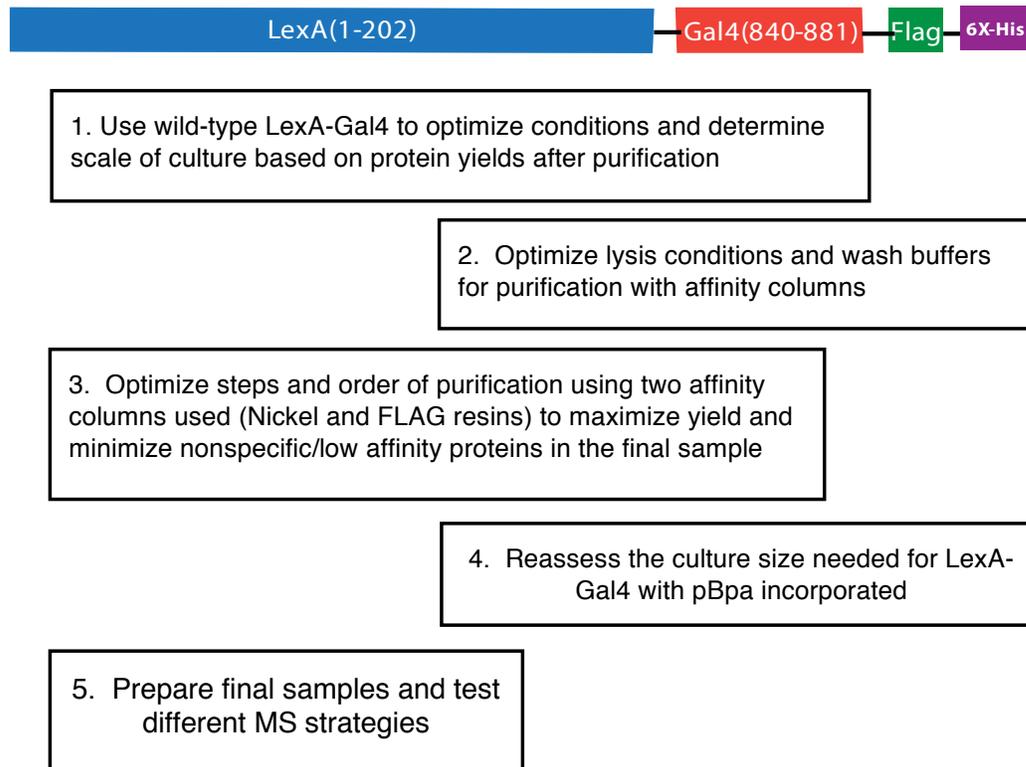


Figure 3-5: LexA-Gal4-FLAG-6X-His construct used in the MS experiments (top). Steps in optimizing a method for preparing a crosslinking protein sample (boxes).

optimization experiments were designed to reduce the nonspecific/low affinity proteins present in the sample. To this end, a sample of full-length wild-type (WT) LexA-Gal4 protein without pBpa and without irradiation was used. Because WT LexA-Gal4 is expressed at higher levels (~5 fold) than Gal4 with pBpa incorporated, this allowed the use of smaller culture sizes. If WT LexA-Gal4 was purified cleanly then it was postulated that the Gal4 mutants containing

pBpa would as well. For purification purposes the LexA-Gal4 protein had both a FLAG and 6x-His epitope tag on the C-terminus.

C.1. Optimizing yeast lysing conditions and wash buffers

Throughout the optimization different yeast lysing methods were tested as yeast require more rigorous lysing methods than other eukaryotic cells or bacteria due to their robust cell wall. In previous experiments with yeast in our lab a mechanical method for lysing was used and included glass beads and a vortex. This protocol was limited to small culture sizes as the vortex only holds 1.5 mL tubes. When preparing samples for MS it was estimated that the culture size would be several orders of magnitude larger than one compatible with the method described. After trying a microfluidizer with little success (meaning whole yeast cells were visualized under a microscope after several times through the instrument) we found that lysing using a combination of incubation with Zymolase (US Biological), a reagent containing an enzyme (β -1,3-glucan laminaripentaohydrolase) used to break down yeast cell walls, followed by treatment with glass beads in either a vortexer or a bead beater resulted in a more complete lysis based on the amount of protein observed on a gel.³⁵ The vortexer can hold 1.5mL microfuge tubes and the bead beater has either a 15 or 50 mL chamber (with a water ice bath chamber that surrounds it) that is placed on top of a rotating blade. Both bead lysing apparati can be controlled in terms of duration and pausing times, both of which were optimized as well. It was important to optimize the amount of time the cells were agitated versus paused especially with the bead beater because it was used on the bench top and the yeast were in direct contact with the blade, thus becoming heated upon use. In the bead beater samples became quite hot if there are not frequent pauses, which was not desirable for protein stability. Ultimately short 15 second agitations with 2 minute pausing in the bead beater was used for lysing large yeast pellets.

To test different combinations of lysis buffer and wash buffer, five 1L cultures of yeast expressing WT LexA-Gal4 protein were grown to test five different combinations (Table 3-3 and Figure 3-6). The protein was purified once

Lysis buffers		
I 100 mM PBS pH 7.0 10% glycerol 10 mM imidazole 0.01% NP-40 10 mM B-me	II 100 mM PBS pH 7.0 150 mM NaCl 10% glycerol 10 mM imidazole 0.5% tween-20 10mM B-me	III 100 mM PBS pH 7.0 350 mM NaCl 10% glycerol 20 mM imidazole 1.0% tween-20 10 mM B-me 1M Guanidine-HCl
Wash buffers		
I 100 mM PBS pH 7.0 4M Guanidine-HCl 20 mM imidazole 0.1% tween-20 10% glycerol 10 mM B-me	II 100 mM PBS pH 7.0 150mM NaCl 4M Guanidine-HCl 30 mM imidazole 1.0% tween-20 10% glycerol 10 mM B-me	III 100 mM PBS pH 7.0 350 mM NaCl 5M Guanidine-HCl 50 mM imidazole 1.0% tween-20 10% glycerol 10 mM B-me

Table 3-3: Lysis and wash buffers with different stringency (1<2<3) used in the initial optimization.

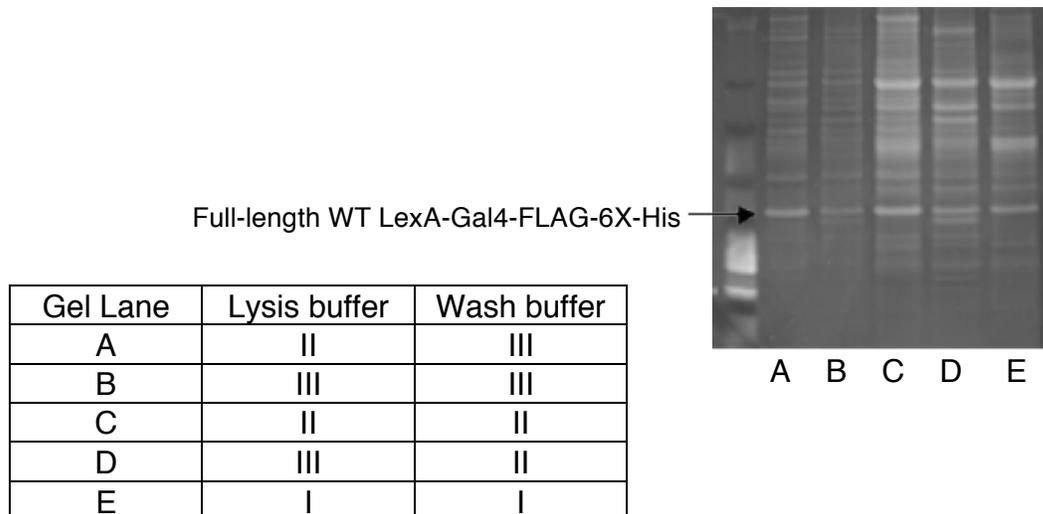


Figure 3-6: Gel showing the results of different lysis and wash conditions of purifying WT LexA-Gal4-FLAG-6X-His with Ni NTA resin. Each lane is purified WT LexA-Gal4-FLAG-6X-His from 1L of yeast expressing this protein. The gel is stained with Sypro Ruby which has a detection limit of .25-1 ng.³⁶

with nickel (Ni NTA) resin to determine what conditions were best for lysing. The wash buffers compatible with Ni NTA resin used were based on what had been used with success in our lab previously and the handbook provided by Sigma that outlined the reagent compatibilities (see Methods for details). Choice of wash buffer was critical because if a wash buffer was incompatible or too stringent, the His epitope tagged protein would be lost over the course of the wash step and too weak a wash buffer could result in high amounts of nonspecific/low affinity proteins stuck to the protein or the Ni NTA resin. The wash buffers used ranged from low stringency to very high stringency (Table 3-3). A lysis buffer was also important because it was in this buffer that the lysate was interacting with the Ni NTA resin therefore minimizing nonspecific binding from the beginning would help in the wash step. It was determined in this study (Figure 3-6) that lane one had the best lysis buffer/wash combination because it had reduced nonspecific bands yet the intensity of the LexA+Gal4 band was not reduced. Further, when one of the most stringent buffers was absent (lane 3) the greatest of nonspecific proteins were observed. Moving forward with this information it was clear that due to the high volume of nonspecific bands with a single purification in the samples a second affinity purification was an order.

C.2 Two affinity columns for purifying crosslinked proteins

Two different affinity columns, Ni NTA and FLAG, were used to purify proteins from a complex protein mixture to reduce the nonspecific proteins in the sample. First the yeast lysate containing WT LexA-Gal4 was bound to FLAG beads which had FLAG antibody conjugated to them, beads were washed, protein was eluted and then bound to the Ni NTA resin, again proteins were washed and eluted off to achieve a final pure protein sample. Figure 3-7 shows the product from the described purification from 6L of yeast culture. In this experiment one band was observed at the correct molecular weight of WT LexA-Gal4 when silver-stained, conditions that have a detection limit of >0.8 ng.³⁶ This

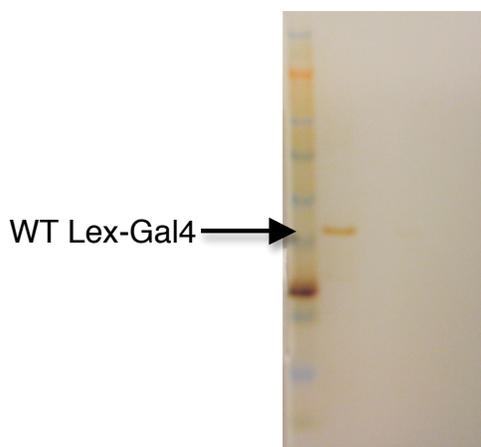


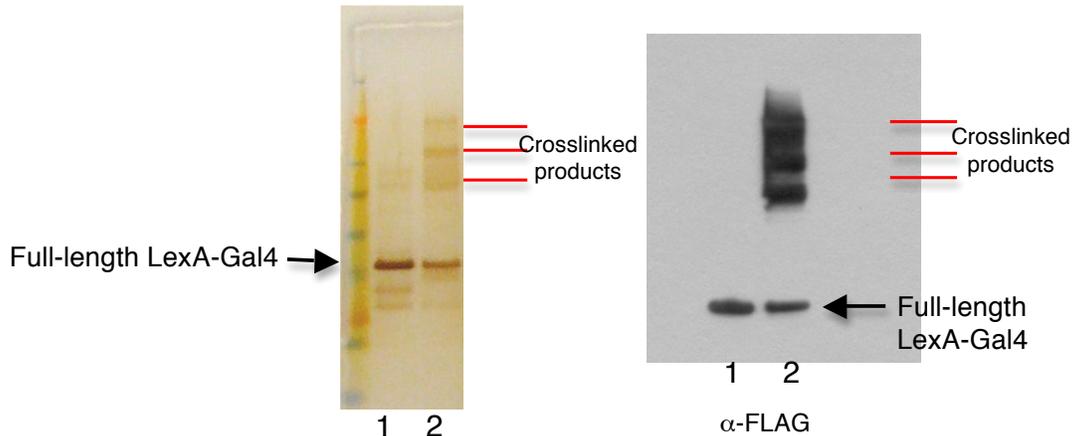
Figure 3-7: Silver stained gel of WT LexA-Gal4 purified with Ni NTA and FLAG affinity resin from 6L of yeast.

result showed two things: (1) the FLAG+Ni NTA dual purification protocol worked, as there were very few, if any, nonspecific bands viewed even with the highly sensitive stain; (2) although 6L of yeast expressing WT LexA-Gal4 protein was visualized well on the sensitive silver stain larger cultures would be required when working with the more poorly expressed protein, LexA-Gal4 with pBpa incorporated.

C.3 Purifying Gal4 with p-benzoyl-L-phenylalanine incorporated

With a purification procedure in place, the next step was to carry out a similar experiment with protein containing pBpa. A major concern moving forward were that the yields seen with WT LexA-Gal4 were not extraordinary considering a 6L sample was prepared. This indicated that when working with the pBpa incorporated LexA-Gal4 protein higher culture sizes would be needed to not only see the full-length protein on a gel but to see the crosslinked products which are at even lower yield. Being able to observe the purified Gal4 protein containing pBpa and crosslinked products in a stained gel was the gauge for obtaining successful MS data. With this in mind, the goal of the next experiment was to understand how much protein was being expressed based on the ability to visualize the proteins with stains that have different detection limits. To this end, 6L yeast culture expressing LexA-Gal4 with pBpa incorporated into position 849

mutant (Phe849pBpa mutant) was crosslinked, purified and run on a gel alongside a 6L sample of not crosslinked WT LexA-Gal4 to compare purity and abundance (Figure 3-8). Position Phe849 was chosen because this site showed the most crosslinked products in the photo-crosslinking experiments, an important aspect to consider when low protein concentration is a concern (see Figure 2-8). The



1. WT LexA-Gal4 (-pBpa -UV)
2. LexA-Gal4 Phe849pBpa (+pBpa +UV)

Figure 3-8: Silver Stained gel (left) and Western blot (right) of the same samples showing purification of 6L of either WT LexA-Gal4 -pBpa -UV (lane 1) and LexA-Gal4 Phe849pBpa +pBpa +UV. This experiment shows that the purification is clean and that the crosslinked products are not very concentrated as they are not very visible considering the sensitivity of the silver stain is >0.8 ng.³⁶ The Western confirms that the higher molecular weight bands in the silver stained gel contain LexA-Gal4 protein.

samples were analyzed by Western blot and three types of staining methods: Coomassie, SYPRO ruby and silver stain with different detection limits (5ng, .25-1ng, and >0.8 ng respectively) and with varying MS compatibility.³⁶ From the data shown in Figure 3-8 it was concluded that the crosslinked bands were seen most clearly in the silver stain which was not MS compatible although Invitrogen does make a MS compatible stain however when tested produced a high background. In both the Coomassie stain and Sypro stain, the full-length protein was visualized in both the WT and Phe849pBpa mutant samples but the crosslinked bands were very light. It was also concluded that the crosslinked bands seen in

the silver stain matched with the Western blot bands nicely indicating that indeed the higher MW bands viewed in the silver stain were crosslinked products and not just nonspecific bands. That did not rule out the possibility that some nonspecific proteins were present and that they could share the molecular weight of a crosslinked target showing up as overlapping bands on the silver stained gel but because the WT LexA-Gal4 showed negligible higher MW nonspecific bands indicated that this was not a clear problem.

C.4 Gel electrophoresis followed by mass spectrometric analysis

As a starting point for MS analysis the University of Michigan's Michigan Proteome Consortium facility was used. A 2D gel was run to resolve protein mixtures and stained protein spots were excised from the gel; these samples were then digested and subjected to MS-MS by MALDI-TOF/TOF. A sample was prepared with a goal of minimally identifying the full-length Gal4 Phe849pBpa protein and the Gal4+Gal80 crosslinked product. Discussions with a staff member at the University of Michigan's Michigan Proteome Consortium indicated that for their MS-MS analysis ideally the proteins needed to be visualized under Coomassie stain. Because 6L culture of irradiated yeast expressing LexA-Gal4-849/pBpa produced crosslinked products that were only minimally observed by Coomassie stain and the more sensitive SYPRO stain, a 12L yeast culture expressing the LexA-Gal4-849/pBpa protein was irradiated and prepared for submission to the Michigan Proteome Consortium for 2D gel resolution followed by MS-MS analysis. Upon staining the 2D gel with Sypro ruby at the facility there were few spots on the gel. It may be that our proteins are just not stained well with SYPRO Ruby, although this would be surprising because it binds to basic amino acids and the polypeptide backbone.³⁷ In our lab we restained the gel with a MS-compatible silver stain (less sensitive than sypro ruby) resulting in an enhancement (Figure 3-9). One concentrated spot in the center appeared near

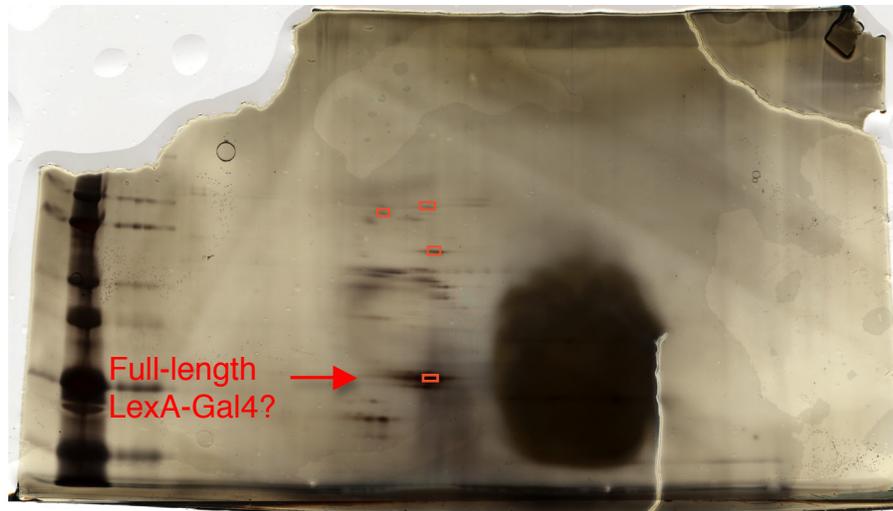


Figure 3-9: Silver stained 2D gel of purified crosslinked LexA-Gal4 Phe849pBpa mutant purified from 12L of yeast. Red boxes indicate spots cut from gel for MS-MS analysis.

the MW of full-length Gal4 Phe849pBpa protein. From this 2D gel four spots were cut out at the facility and analyzed by MS-MS. They were not able to identify proteins in any of the samples tested.

To further define the technical hurdles, a second sample was prepared for submission to the facility. The goal of this experiment was to simply detect the full-length Gal4 Phe849pBpa. If the technique was not sensitive enough to detect the most abundant protein in the protein mixture then it could never be used to detect lower abundant crosslinked proteins. This time 6L of yeast expressing Gal4 Phe849pBpa were irradiated, purified and run on a 1D gel. Only the full-length Gal4 Phe849pBpa (not crosslinked protein) was excised from the gel for MS/MS analysis. From this experiment the facility was able to identify our protein in the sample but only 5 tryptic peptides were identified, certainly a concern.

The lack of promising data obtained at the Michigan Proteome Consortium precipitated the search outside the University of Michigan for facilities that have a more high throughput, sensitive technique. Samples were analyzed at Michigan State University and the Rockefeller University; however, little additional signal was observed. To this end, a collaboration with Center for Physiological

Proteomics at Scripps Research Institute was established where samples are analyzed by MudPIT as described above. Samples from 6-12L yeast cultures were prepared using a modified protocol from above where the major change being the inversion of the order of the affinity purification columns (Figure 3-10). Here the sample was first exposed to Ni NTA resin and then to FLAG beads. Using this purification strategy samples were prepared and sent for analysis.

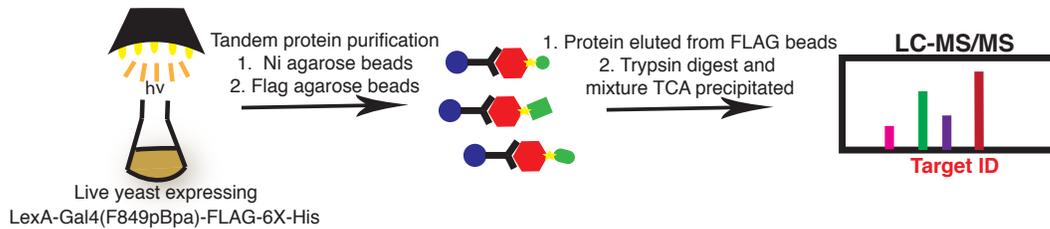


Figure 3-10: Method for preparing a crosslinked LexA-Gal4 Phe849pBpa sample for MudPIT analysis.

D. Multidimensional Protein Identification Technology (MudPIT) for identifying protein partners of Gal4

The first sample prepared for submission to the Center for Physiological Proteomics for MudPIT analysis was crosslinked Gal4 Phe849pBpa. To assess the MudPIT method for detecting our proteins, the yeast were grown in glucose so as to capture the Gal4-Gal80 interaction. A 12L yeast culture expressing Gal4 Phe849pBpa was irradiated, purified and ~1/3 of the purified protein sample was run on a 1D gel and silver stained with silver stain that is not MS compatible because past experiments showed this stain give the best visualization of the proteins. On the same gel another ~1/3 of the samples was run and not stained. The unstained gel was superimposed on the silver-stained gel to guide excision. If LexA-Gal4 Phe849pBpa and a target protein are both identified in an excised gel slice at the correct molecular weight corresponding to a crosslinked product between the two them this will strength the evidence that they are direct protein partners. The remaining 1/3 of the solution sample that was not run on the gel was also submitted. A disadvantage of cutting bands from a gel is that there is always some sample loss relative to simple analysis of the crosslinked solution.

However, one disadvantage surrounding the solution sample is that it is not known if the proteins identified would be present at the correct molecular weight corresponding to a crosslinked product with LexA-Gal4 if run on a gel; therefore further verification of proteins identified is required to confirm that it is a direct target.

One complicating aspect of analyzing the results from these experiments is that it is difficult to identify peptide sequences with the program used that are composed of a part of LexA-Gal4 Phe849pBpa, pBpa and part of the crosslinked protein target. It is challenging because the program used, SEQUEST, does not recognize a mass corresponding pBpa; even if it was incorporated into the program, the peptide fragment would contain parts of two different proteins complicating the database search. We are working the proteomic facility at Scripps to incorporate these aspects into the analysis but, thus far they have not been able to do it in a straightforward manner.

Results from this from the initial study (Figure 3-11) showed that in the solution sample, Gal80 was observed and with significant sequence coverage (54.0%). Additionally, 45 unique peptides with 168 peptides total of LexA-Gal4 Phe849pBpa were identified generating 60.4% sequence coverage, a considerable improvement from the previous method used (only 5 unique peptides identified). Of the other proteins identified, those involved in translational processes made-up ~17% of the all proteins in the irradiated sample. This could be from pBpa crosslinking occurring as the protein is being translated or because these proteins are high abundance proteins that happen to interact with Gal4 Phe849pBpa over the course of irradiation.

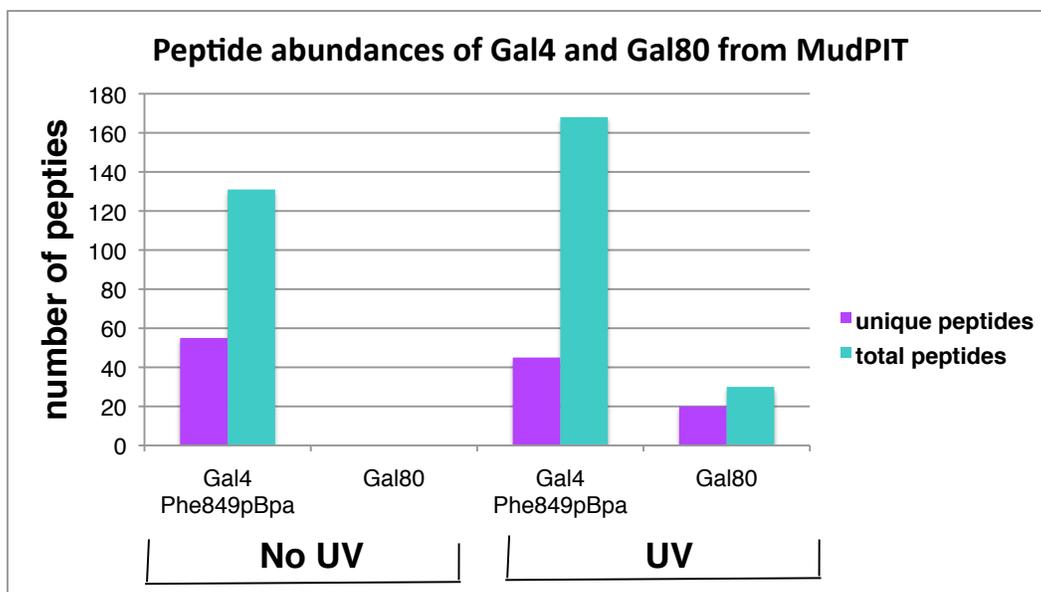


Figure 3-11: Number of unique and total peptides identified for Gal4 Phe849pBpa and Gal80 with and without UV using MudPIT.

Next, a control experiment of 6L yeast expressing Gal4 Phe849pBpa in glucose without irradiation was carried out to investigate whether or not protein binding partners such as Gal80 would be absent. Not surprising, this was the case (Figure 3-11). Additionally, another irradiated control experiment was done where only the tRNA/synthetase plasmid was present in the yeast and no Gal4 Phe849pBpa expression plasmid. This experiment was used for detecting the nonspecific/low binding proteins that are artifacts of the purification process independent from the Gal4 Phe849pBpa protein. From this data about ~62% fewer proteins were identified in this sample than in the Gal4 Phe849pBpa sample that was not irradiated. Indicating that little nonspecific/low affinity proteins were pulled through the process and ones that were can be used to understand the experimental data when LexA-Gal4 Phe849pBpa was present.

E. Using Multidimensional protein identification technology for identifying targets beyond Gal80

Using MudPIT, Gal4 and Gal80 were observed in crosslinked samples, meaning that these techniques have the potential to be used for identifying other

protein partner of Gal4 in vivo. Towards capturing protein targets of Gal4 when transcription is occurring crosslinking experiments were carried out in the presence of galactose and prepared for MudPIT analysis. Two samples, both of Gal4 Phe849pBpa where one was irradiated (12L) and other was (6L) not were prepared and submitted to the MS facility. Figure 3-12 and Table 3-4 outlines the transcription proteins identified that were present only in the sample that was irradiated. None of them were identified in the sample that was not irradiated.

The nine proteins identified are components of six different complexes that are recruited to the DNA during transcription and two of the proteins are not in complexes (Table 3-4).

Complex	Protein(s)
CCR4-NOT	Ccr4 Mot2
SAGA	Tra1
Paf1	Ctr9
Nuclear Pore	Mlp2
TFIIH	Ssl2
Swi/Snf	Swi1
	Abf1
	Hrr25

Table 3-4: Proteins identified by MuDPIT of crosslinked LexA-Gal4 Phe849pBpa in the presence of galactose.

Some of the proteins found were not surprising such as Tra1, a component of SAGA, a complex with histone acetyltransferase function. Through various strategies, discussed in Chapter 1, this complex has been thought of a front runner in the search for direct binding partners.^{38,39} Likewise, Swi1 a component of the SWI/SNF complex that remodels chromatin through their

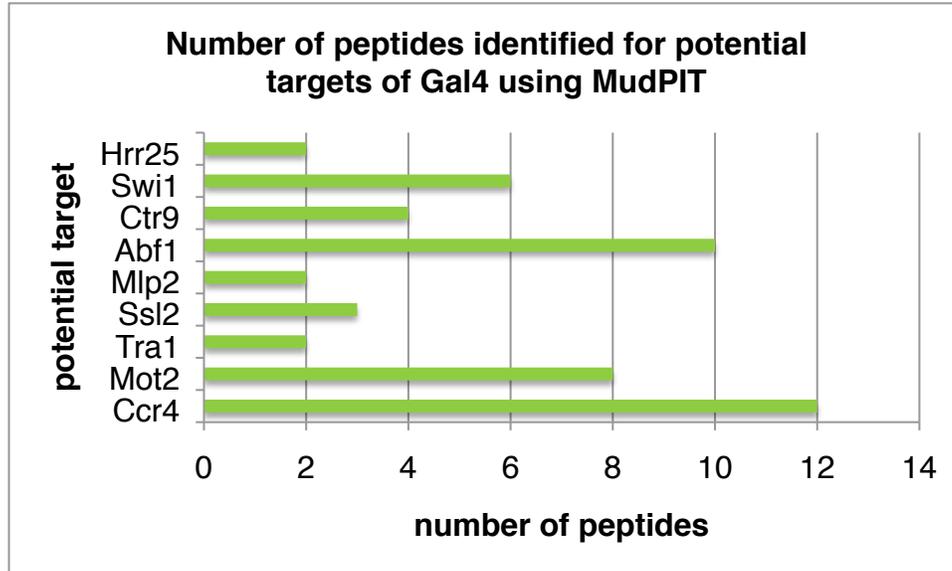


Figure 3-12: Potential targets of Gal4 and the amount of peptides identified from crosslinked Gal4 Phe849pBpa in the presence of galactose.

ATPase capabilities has been suggested as partners of amphipathic activators before (Chapter 1).^{40,41} Both possess enzymatic activity, aiding in chromatin remodeling and are not members of the general transcriptional machinery.

Another protein identified in this study was Mlp2, a component of the inner nuclear basket of the nuclear pore.⁴² A study found that Mlp2 (and Mlp1) interacts with SAGA, the complex mentioned above, and is thought to colocalize to DNA with Gal4 shown through ChIP.⁴³ Interestingly they also found that Mlp2 (and Mlp1) interacts with the GAL upstream activating sequence (UAS) which is the region of the GAL gene that SAGA interacts.⁴² The MudPIT data further strengthens these past studies of nuclear pore proteins being localized to the GAL UAS. Further mechanistic details of how a component of the nuclear pore could be localized to the GAL UAS region and crosslinked with Gal4 could be another area of research to explore in transcription regulation.

Additionally, the proteins Ccr4 and Mot2 were identified. They are part of the CCR4-NOT complex that participates in many transcription processes such as, contacting and regulating TFIID function, contacting SAGA and is thought to contribute to mRNA deadenylation.⁴⁴ Due to the diverse functions of this

complex, involved in regulating transcription both positively and negatively, not all of the components are well characterized.

Ctr9, a protein identified, is a member of the Paf complex which is a five subunit complex that is thought to associate with RNA polymerase II and aid in transcription elongation.⁴⁵ Like the CCR4-NOT complex the Paf complex was not typically on the radar for interacting directly with Gal4. Between these two protein complexes it appears that proteins with diverse functionality are being identified with this approach and ones that were not thought of as leading targets of Gal4 before.

Ssl2, a component of TFIIH, is a DNA helicase that is involved in transcription initiation and nucleotide-excision repair was identified.⁴⁶ Although it is involved in basic transcription perhaps it was specifically targeted to DNA as a result of the UV light that the cells were exposed during the crosslinking process. Considering the function of this protein it may be difficult to study the role that this target has in Gal4 activated transcription under the crosslinking conditions used.

Finally, two proteins that are not associated with a complex were identified Abf1 and Hrr25. Abf1 is thought to carry out several functions including DNA replication, transcription activation and chromatin remodeling.^{47,48} Additionally, it is able to bind DNA in a sequence specific manner and in fact can bind to its own promoter and repress transcription.^{47,48} Hrr25 is a protein kinase and is thought to have a role in the transcriptional processes in response to DNA damage.⁴⁹ It is also thought to interact with the C-terminal domain of RNA polymerase II and phosphorylate Swi6, a component of the Swi/Snf complex.⁴⁹ Neither one of these proteins has historical evidence of interacting with Gal4 and would be a new finding once verified.

E.1 Analyzing proteins identified in crosslinked samples

One of the challenging aspects of analyzing the mass spectrometric data was that many proteins, those related to transcriptional and those that were not, were identified in the experiments (see Appendix). To determine which proteins

identified were relevant over other proteins an analysis of all of the proteins identified in all of the samples submitted for MudPIT analysis was applied (Table 3-4 and Appendix).

Sample	Culture Size	Sugar Source	UV
Gal4 Phe849pBpa	6L	Glucose	No
Gal4 Phe849pBpa	6L	Glucose	Yes
Gal4 Phe849pBpa	6L	Galactose	No
Gal4 Phe849pBpa	6L	Galactose	Yes
Gal4 Phe849pBpa	12L	Galactose	Yes
Only tRNA/synthetase no Gal4 Phe849pBpa present	6L	Glucose	Yes

Table 3-5: Samples prepared for MudPIT analysis.

Only proteins that were functionally related to transcription and only present in irradiated samples grown in either galactose or found in both galactose and glucose were deemed potential targets of Gal4. Other proteins that were not related to transcription could interact with Gal4 and could potentially have unknown functions related to regulating Gal4 function in the cell however, because the goal was to identify direct protein targets that are functionally relevant to Gal4 activated transcription the above proteins were thought to be the most likely candidates.

E.2 A limitation in the MudPIT analysis

One major limitation in the MudPIT analysis was that peptides that contain a portion of Gal4 covalently attached, through pBpa, to a portion of the target protein could not be detected with the SEQUEST software used to analyzed the peptide samples. This would require the software to recognize two different proteins are present in one peptide and that the mass of the nonnatural amino acid pBpa to be understood as a potential mass that is possible in the sample mixture. Due to the complexity of this it is was not possible for the Center for Physiological Proteomics to observe these peptides. This identification would be definitive evidence of the proteins being direct targets of Gal4. Until this type of

analysis is developed other methods, described in the next section must be used to verify proteins identified as direct targets of Gal4.

E.3 Future MudPIT studies

In summary, these results have strengthened the evidence that Gal4 has a multiprotein binding profile and that these proteins come from complexes with diverse functions. Further several of the proteins are novel potential targets of Gal4. When comparing the MS results with a Western blot of crosslinked samples of LexA-Gal4 with pBpa incorporated at various positions in the presence of galactose, many more protein bands were observed than the nine found with MudPIT (Figure 3-13). Of particular interest is a band that is near the molecular weight

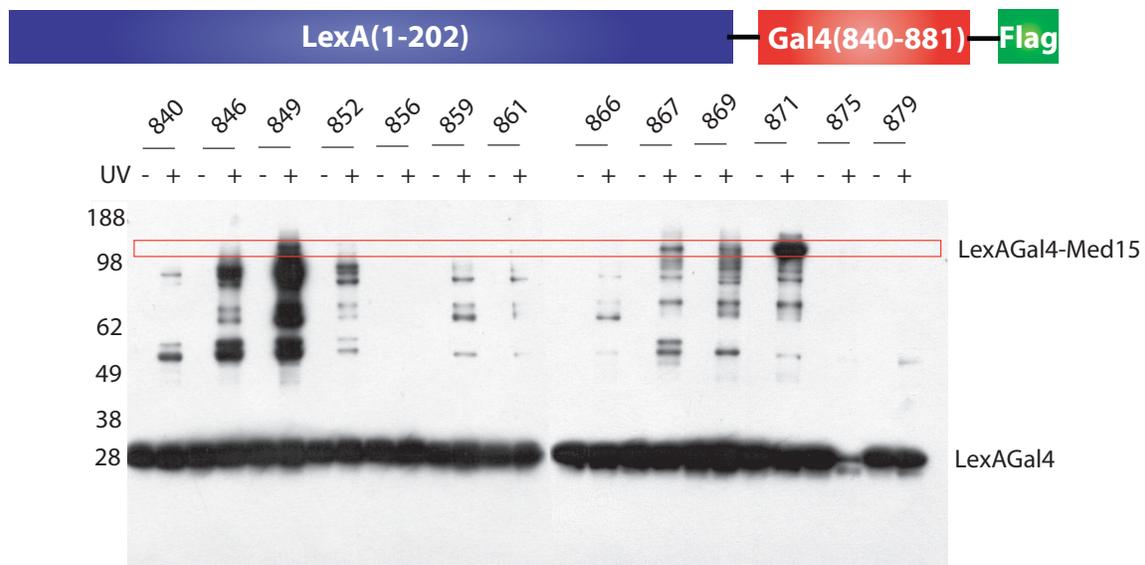


Figure 3-13: Crosslinking experiment with LexA-Gal4-FLAG with pBpa incorporated at different positions along Gal4 TAD. LexA-Gal4 crosslinked to Med15 (~150 kDa) could possibly be bands in the red box.

of Gal4 crosslinked to Med15, a protein of the mediator that has been thought to be a direct target (~150 kDa).^{38,39} Because of the presence of this band and the many other bands the next experiment planned is to scale-up the culture size to 50L so that low abundance proteins that may have been missed in the first experiment could be identified.

Overall, we were able to show that in vivo crosslinking with Gal4 coupled with MudPIT is a useful technique for identifying transcription proteins which have a history of being extremely difficult to study. Additional experiments are planned to further verify and resolve the protein targets of Gal4. Potential targets identified by the MS data will be verified through Western blotting and protein deletions in yeast. Additionally, this method will be used to identify targets of other amphipathic activators, like GCN4 and this data will be compared to Gal4 to understand the shared and unique targets of each activator, information that will be useful for understanding fundamental transcriptional mechanistic questions.

F. Methods

General peptide synthesis for FLAG peptide

Standard Fmoc solid phase peptide synthesis was used to make the FLAG peptide DYKDDDDK. The resin was purchase from Peptides international and all of the Fmoc-protected amino acids were purchased from Peptides International or Chem Impex International. HOBt and HBTU were purchased from CPC Scientific. The synthesis was carried out on an Advanced ChemTech90 synthesizer using establish protocols. N-Methyl-2-pyrrolidone (NMP), Dimethylformamide (DMF), N,N-Diisopropylethylamine (DIEA), piperidine, pyridine, and acetic anhydride were purchase from Fisher Scientific and used without purifying. The peptide was purified to homogeneity using reversed-phase HPLC on a C18 column with a gradient solvent system (buffer A: 0.1% TFA, buffer B: CH₃CN).

Growing cultures for crosslinking experiments

The protocol for growing the yeast and irradiating them is similar to the experiments in Chapter 2. The modifications to the protocol are the following. LS41 yeast were transformed with a plasmid coding WT LexA-Gal4 or LexA-Gal4 Phe849pBpa-FLAG-6X-His mutant plasmid and pSNRtRNA-pBpaRS. A single colony is picked for inoculation of a 5mL starter culture of as described in Chapter 2. Some of the starter culture is used to inoculate a 1L yeast cultures

which were grown in SC media lacking histidine, tryptophan and uracil and containing 2% glucose and 1mM pBpa (dissolved in 10mL of 1M NaOH) and 10mL of 1M HCl. Cultures were grown to an OD₆₆₀ 0.8. Cultures were harvested by centrifugation for 15 minutes at 6000 rpm and at 4°C. Samples that were going to be irradiated were resuspended in SC media lacking Histidine, Tryptophan and Uracil and containing 2% glucose and were placed under the UV lamp (Eurosolar 15 W UV lamp) for 1 hour with a cooling pack underneath the sample. Following irradiation samples were pelleted by centrifugation at 4°C at 4500 rpm for 5 minutes. They were then washed once with 100mM PBS buffer and pelleted again in similar manner. Cultures that were not irradiated were washed once with 100mM PBS buffer and pelleted. All cultures were stored at -80°C until lysed.

Sample lysis and purification for MudPIT

Pelleted yeast cultures are removed from -80°C and thawed on ice or in a water bath on the bench top. ~2X EDTA-free protease inhibitor tablets (Roche) was added to the lysis buffer (100mM PBS, 150 mM NaCl, 10% glycerol, 0.5% NP-40, 10mM B-me). Each pellet containing 3L of yeast culture is resuspended in 10mL of lysis buffer (adjust amount accordingly for different culture sizes). 1000 units of Zymolase 20T (US Biologicals) is dissolved in 1mL of lysis buffer and 0.5mL is added to each 3L resuspended pellet (50mg/3L yeast pelleted culture or ~15mg/1L). Let suspension rotate slowly at 30°C for 30-45 minutes. During this time glass beads and the 15mL or 50mL lysis bead beater chamber (Biospec Products) are pre-chilled in a freezer -80°C. After incubating with Zymolase cool culture on ice for a few minutes. Add yeast to lysis bead beater chamber and add glass beads to fill ~3/4 to top of meniscus. Total time bead beater is turned on for lysis is 3 minutes with 15 seconds turned on with 2 minutes rest in between. Decant yeast lysis bead slurry (add extra lysis buffer to make sure all beads are removed from chamber) into 50mL tubes and centrifuge at 4500rpm for 5 minutes at 4°C. Decant lysate into a clean 50mL tube and pour the beads into 1.5mL microfuge tubes. Poke a hole in the bottom of the 1.5mL tube with a

flamed sterile needle (19G) and place into new 1.5mL microfuge tube and centrifuge both tubes together at 5000rpm for 1 minute at 4°C. Collect the supernatant in the lower tube and add it to the lysate already in the in the 50mL tube. Centrifuge all of the lysate in the 50mL tube at 9000rpm for 5 minutes at 4°C. Decant the supernatant into a new 50mL tube and centrifuge again at 9000rpm for 5 minutes at 4°C. Repeat until there is virtually no pellet observed. Wash 300 μ L of Ni NTA beads (Sigma-Aldrich) 3 times in 100mM PBS buffer at 3000rpm for 3 minutes at 4°C and add to lysate along with 10mM B-me. Rotate and incubate with Ni NTA beads for 1 hour at 4°C. After incubation centrifuge Ni NTA beads 3000rpm for 3 minutes at 4°C, remove the lysate and transfer Ni NTA beads to a 15mL tube. Wash the Ni NTA beads 6 times in 5mL of wash buffer (100mM PBS, 350NaCl, 10% glycerol, 3M Guanidine-HCl, 30mM imidazole, 1.0% tween-20, 10mM B-me) and 2 times in 5mL of 100mM PBS buffer centrifuging at 3000rpm for 3 minutes at 4°C each time and decanting buffer. Transfer Ni NTA beads to a 1.5mL microfuge tube and elute the protein in 500 μ L elution buffer (100mM PBS, 500mM imidazole, 0.5% tween-20) three times with 10 minute incubations each time rotating at 4°C. The 1.5mL eluant is then buffer exchanged in a centrifugal concentrator (Vivaspin 500) with a 10,000 molecular weight cut-off and a polyethersulfone membrane (PES) by first concentrating the eluant to ~50-100 μ L at 10,000 rpm at 4°C and then adding 500 μ L of 100mM PBS buffer and repeat two more times bringing the volume down to 100ul and adding 500ul of 100mM PBS buffer diluting the imidazole for preparation for incubation with anti-FLAG M2 agarose beads (Sigma). After the third time bringing the elution volume down to ~50-100ul bring the final sample volume up to 500ul with PBS. Wash 200 μ l the anti-FLAG M2 agarose beads (Sigma) three times with 100mM PBS in a 1.5mL microfuge tube at 3000 rpm for 3 minutes at 4°C. Add the eluant to the FLAG beads and incubate at 4°C rotating for 2 hours. After incubation centrifuge the FLAG beads at 3000rpm for 3 minutes at 4°C and remove the supernatant. Wash FLAG beads in 1mL wash buffer (100mM PBS, 350mM NaCl, 10% glycerol, 1.0% NP-40) ten times and two times in 100mM

PBS centrifuging at 3000rpm for 3 minutes at 4°C. Make a FLAG elution buffer consisting of a solution of 10mg/mL purified FLAG peptide (DYKDDDDK) in a buffer containing 500µl lysis buffer (see above) and 500ul 100mM PBS. Elute the protein from the FLAG beads in ~300µl of FLAG elution buffer three times (2 30 minute incubations at 4°C rotating and a final one overnight rotating and incubating at 4°C) and collecting the eluant each time after centrifuging at 3000 rpm for 3 minutes at 4°C. Upon collecting the final elution immediately begin concentrating the eluant in a centrifugal concentrator (Vivaspin 500) with a 10,000 molecular weigh cut-off and PES membrane. Buffer exchange the sample into a MS compatible buffer by bringing the volume down to ~50-100µl and adding 500µl of the exchange buffer (50mM Ammonium Acetate or 50mM PBS) three times. The final volume of the sample is ~30ul. Flash freeze the sample and place in the freezer (-20°C) until shipping on dry ice.

Lysis buffers and Wash buffers use for optimization

1L yeast cultures samples were thawed and resuspended in one of the following lysis buffers containing EDTA free protease inhibitor (Roche). Lysis buffer I: 100mM PBS pH 7.0, 0.01% NP-40, 10% glycerol, 10mM imidazole, 10mM B-me
Lysis buffer II: 100mM PBS pH 7.0, 150 mM NaCl, 10mM imidazole, 10% glycerol, 0.5% tween-20, 10mM B-me
Lysis buffer III: 100mM PBS, 350 mM NaCl, 20mM imidazole, 10% glycerol, 1.0% tween-20, 10mM B-me, 1M Guanidine-HCl. Cells were lysed by a incubation with Zymolase enzyme followed by glass beads for 30 minutes, 10 minute pause followed by another 15 minute vortex. Following an incubation with Ni NTA (Sigma) beads the beads were washed 6 times with one of the following wash buffers. Wash buffer I: 100mM PBS pH 7.0, 0.1% tween-20, 10% glycerol, 10 mM B-me, 4M Guanidine-HCl, 20 mM imidazole. Wash buffer II: 100 mM PBS pH 7.0, 150mM NaCl, 10mM B-me, 1.0% tween-20, 4M Guanidine-HCl, 30mM imidazole, 10% glycerol. Wash buffer III: 100mM PBS pH 7.0, 350 mM NaCl, 10mM B-me, 1.0% tween-20, 5M Guanidine-HCl, 50mM imidazole, 10% glycerol. Proteins were eluted

form the Ni NTA resin in 100mM PBS buffer containing 500mM imidazole and 0.5% tween-20, separated on a 1D gel and stained with Sypro ruby.

G. References

- (1) Majmudar, C. Y.; Lee, L. W.; Lancia, J. K.; Nwokoye, A.; Wang, Q.; Wands, A. M.; Wang, L.; Mapp, A. K. *J Am Chem Soc* **2009**, *131*, 14240.
- (2) Muller, F. M.; Werner, K. E.; Kasai, M.; Francesconi, A.; Chanock, S. J.; Walsh, T. J. *J Clin Microbiol* **1998**, *36*, 1625.
- (3) Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. *Nature* **2003**, *425*, 737.
- (4) Opiteck, G. J.; Ramirez, S. M.; Jorgenson, J. W.; Moseley, M. A., 3rd *Anal Biochem* **1998**, *258*, 349.
- (5) Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. *Anal Chem* **1997**, *69*, 1518.
- (6) Wall, D. B.; Kachman, M. T.; Gong, S.; Hinderer, R.; Parus, S.; Misek, D. E.; Hanash, S. M.; Lubman, D. M. *Anal Chem* **2000**, *72*, 1099.
- (7) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd *Nat Biotechnol* **1999**, *17*, 676.
- (8) Siuzdak, G. *Proc Natl Acad Sci U S A* **1994**, *91*, 11290.
- (9) Stapels, M. D.; Barofsky, D. F. *Anal Chem* **2004**, *76*, 5423.
- (10) Andre, M.; Le Caer, J. P.; Greco, C.; Planchon, S.; El Nemer, W.; Boucheix, C.; Rubinstein, E.; Chamot-Rooke, J.; Le Naour, F. *Proteomics* **2006**, *6*, 1437.
- (11) Soskic, V.; Gorlach, M.; Poznanovic, S.; Boehmer, F. D.; Godovac-Zimmermann, J. *Biochemistry* **1999**, *38*, 1757.
- (12) Stapels, M. D.; Cho, J. C.; Giovannoni, S. J.; Barofsky, D. F. *J Biomol Tech* **2004**, *15*, 191.
- (13) Chait, B. T. *Science* **2006**, *314*, 65.
- (14) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. *Science* **2006**, *314*, 109.
- (15) Wright, G. L., Jr.; Farrell, K. B.; Roberts, D. B. *Biochim Biophys Acta* **1973**, *295*, 396.
- (16) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc Natl Acad Sci U S A* **2000**, *97*, 9390.
- (17) Yates, J. R., 3rd *Trends Genet* **2000**, *16*, 5.
- (18) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Anal Chem* **1987**, *59*, 2621.
- (19) Cravatt, B. F.; Simon, G. M.; Yates, J. R., 3rd *Nature* **2007**, *450*, 991.
- (20) Bennetzen, J. L.; Hall, B. D. *J Biol Chem* **1982**, *257*, 3026.

- (21) Raida, M.; Schulz-Knappe, P.; Heine, G.; Forssmann, W. G. *J Am Soc Mass Spectrom* **1999**, *10*, 45.
- (22) Washburn, M. P.; Yates, J. R., 3rd *Trends in Biotechnology* **2000**, *18*, 27.
- (23) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat Biotechnol* **2001**, *19*, 242.
- (24) Sharp, P. M.; Li, W. H. *Nucleic Acids Res* **1987**, *15*, 1281.
- (25) Perrot, M.; Sagliocco, F.; Mini, T.; Monribot, C.; Schneider, U.; Shevchenko, A.; Mann, M.; Jenö, P.; Boucherie, H. *Electrophoresis* **1999**, *20*, 2280.
- (26) Gygi, S. P.; Rochon, Y.; Franz, B. R.; Aebersold, R. *Mol Cell Biol* **1999**, *19*, 1720.
- (27) Futcher, B.; Latter, G. I.; Monardo, P.; McLaughlin, C. S.; Garrels, J. I. *Mol Cell Biol* **1999**, *19*, 7357.
- (28) Wong, C. C.; Xu, T.; Rai, R.; Bailey, A. O.; Yates, J. R., 3rd; Wolf, Y. I.; Zebroski, H.; Kashina, A. *PLoS Biol* **2007**, *5*, e258.
- (29) Xiao, K.; McClatchy, D. B.; Shukla, A. K.; Zhao, Y.; Chen, M.; Shenoy, S. K.; Yates, J. R., 3rd; Lefkowitz, R. J. *Proc Natl Acad Sci U S A* **2007**, *104*, 12011.
- (30) Cass, C. L.; Johnson, J. R.; Califf, L. L.; Xu, T.; Hernandez, H. J.; Stadecker, M. J.; Yates, J. R., 3rd; Williams, D. L. *Mol Biochem Parasitol* **2007**, *155*, 84.
- (31) Peterson, C. L.; Workman, J. L. *Curr Opin Genet Dev* **2000**, *10*, 187.
- (32) Tanaka, Y.; Bond, M. R.; Kohler, J. J. *Mol Biosyst* **2008**, *4*, 473.
- (33) Dorman, G.; Prestwich, G. D. *Biochemistry* **1994**, *33*, 5661.
- (34) Mapp, A. K.; Ansari, A. Z. *ACS Chem Biol* **2007**, *2*, 62.
- (35) Chesters, C. G.; Bull, A. T. *Biochem J* **1963**, *86*, 31.
- (36) Invitrogen *Molecular Probes*.
- (37) Smejkal, G. B.; Robinson, M. H.; Lazarev, A. *Electrophoresis* **2004**, *25*, 2511.
- (38) Reeves, W. M.; Hahn, S. *Mol Cell Biol* **2005**, *25*, 9092.
- (39) Fishburn, J.; Mohibullah, N.; Hahn, S. *Mol Cell* **2005**, *18*, 369.
- (40) Bhaumik, S. R.; Raha, T.; Aiello, D. P.; Green, M. R. *Genes Dev* **2004**, *18*, 333.
- (41) Neely, K. E.; Hassan, A. H.; Wallberg, A. E.; Steger, D. J.; Cairns, B. R.; Wright, A. P.; Workman, J. L. *Mol Cell* **1999**, *4*, 649.
- (42) Luthra, R.; Kerr, S. C.; Harreman, M. T.; Apponi, L. H.; Fasken, M. B.; Ramineni, S.; Chaurasia, S.; Valentini, S. R.; Corbett, A. H. *J Biol Chem* **2007**, *282*, 3042.
- (43) Bhaumik, S. R.; Green, M. R. *Genes Dev* **2001**, *15*, 1935.
- (44) Chen, J.; Rappsilber, J.; Chiang, Y. C.; Russell, P.; Mann, M.; Denis, C. L. *J Mol Biol* **2001**, *314*, 683.

- (45) Rondon, A. G.; Gallardo, M.; Garcia-Rubio, M.; Aguilera, A. *EMBO Rep* **2004**, *5*, 47.
- (46) Furuchi, T.; Takahashi, T.; Tanaka, S.; Nitta, K.; Naganuma, A. *Nucleic Acids Res* **2004**, *32*, 2578.
- (47) Miyake, T.; Reese, J.; Loch, C. M.; Auble, D. T.; Li, R. *J Biol Chem* **2004**, *279*, 34865.
- (48) Beinoraviciute-Kellner, R.; Lipps, G.; Krauss, G. *FEBS Lett* **2005**, *579*, 4535.
- (49) Phatnani, H. P.; Jones, J. C.; Greenleaf, A. L. *Biochemistry* **2004**, *43*, 15702.

Chapter 4

A comparative analysis of photoactivatable crosslinking moieties

A. Project Focus

Protein-protein interactions (PPIs) are essential for carrying out and maintaining cellular processes.¹⁻³ Alteration of PPIs changes the signaling events and the biological processes they are regulating, making them major contributors to the initiation and persistence of diseases like cancer.^{2,4} As a result, there has been enormous effort exerted by the chemical, biological and medical communities aimed at identifying and characterizing PPIs that play a role in disease processes with the longer-term goal of discovering new therapeutic targets.³ As outlined in the previous chapters, PPIs that regulate transcription initiation comprise one such class and thus have been our focus of study.

Photocrosslinking has long been an important tool to capture and characterize PPIs in vitro and in their native context and the ability to genetically incorporate two unnatural amino acids with photoactivatable functionality, p-benzoyl-L-phenylalanine (pBpa) and p-azido-L-phenylalanine (pAzpa) has further enhanced its utility.^{5-7 8,9} However, little attention has been paid to the inherent reactivities and selectivities of these unnatural amino acids in spite of the profound influence these characteristics will likely have on experimental outcome.¹⁰ Here I compare the known reaction pathways for pBpa and pAzpa to predict the utility of these photocrosslinkers in studying PPIs in terms of their likely amino acid preferences. In addition, I analyze these preferences in the context of the known potential for the most preferred amino acids to reside at the surface/interface of proteins. These preferences are illustrated in a direct comparison of pBpa and pAzpa crosslinking in vivo. pBpa and pAzpa

crosslinking is compared at the Gal4-Gal80 interface and shows that placing the crosslinker at different positions results in unique crosslinking profiles. Additionally, placing a different crosslinker in the same position in Gal4 forms crosslinks to Gal80 when pAzpa is present and is attenuated for pBpa. This information will be particularly useful for designing future experiments in capturing the direct protein targets of TADs and is a general consideration for studying PPIs with crosslinkers *in vivo*.

B. Background

B.1 Photocrosslinking with transcriptional activators

Photo-crosslinking to study the targets of transcriptional activators has been used by many groups in an *in vitro* context. Most commonly, the strategy employs an aryl azide as the crosslinking moiety.⁷⁻⁹ One crosslinking strategy that has been used by the Hahn research group involves attaching N-((2-pyridyldithio) ethyl)-4-azidosalicylamide (¹²⁵I-PEAS) through a disulfide linkage to cysteines (Figure 4-1).^{4,8,9} Once attached to a protein and UV light is applied,

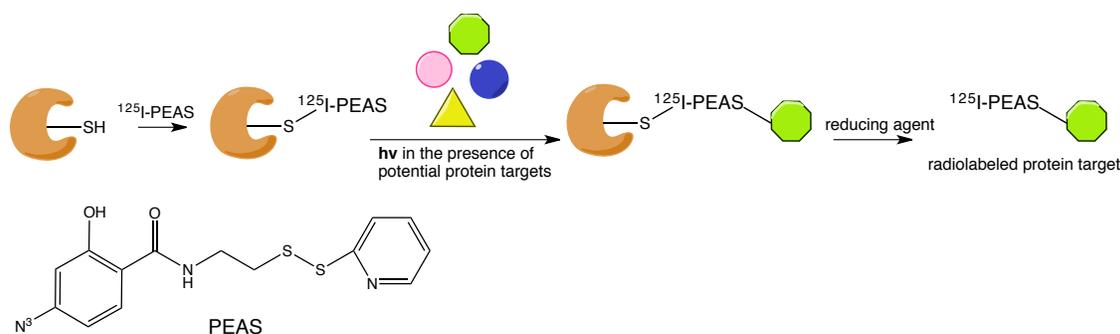


Figure 4-1: Scheme of crosslinking with ¹²⁵I-PEAS.

crosslinks are formed with neighboring biomolecules. After crosslinking, a reducing agent is added, reducing the disulfide bond, releasing the crosslinked products and leaving the targets isotopically labeled for further analysis. Hahn and coworkers applied this crosslinking method to identify the targets of both Gal4 and GCN4 *in vitro*. Both TADs used in their study do not naturally contain any cysteines so first they introduced cysteines at different position along the

TAD. Only one cysteine was mutated at a time so the labeling was site-specific. The labeled proteins were localized on DNA in vitro in the presence of purified preinitiation complex (PICs). This was a significant study that resulted in the identification of multiple direct targets of each activator (Gal11, Tra1, Taf12) in PIC); however, there are limitations to the PEAS crosslinking strategy.^{8,9} One is that the crosslinking moiety on ¹²⁵I-PEAS is 14 Å from the α carbon of the cysteine to which it is attached when fully extended. This means that it is possible that crosslinks to proteins that are not making direct contact with the TAD it is attached could be occurring.¹¹ Another downfall of this crosslinking strategy is that it is not performed in a cellular context but in the presence of a much simpler environment, purified PICs. Further, as discussed later in this Chapter, the chemical reactivity of the phenyl azide moiety may not be optimal for capturing a full complement of binding partners.

Also using a related aryl azide crosslinking moiety, Sulfo-SBED, Workman and coworkers carried out photo-crosslinking label transfer assays to identify direct targets of GCN4 (Figure 4-2).⁷ Sulfo-SBED is a crosslinking moiety that is conjugated to proteins through the lysines on the TAD and contains a photoreactive aryl azide, a disulfide bond and a biotin label. Upon UV irradiation crosslinks are formed via the aryl azide, DTT is added to disrupt the disulfide bond thereby removing the activator and leaving the biotin handle on the target. This is then used for Western blotting with a streptavidin antibody for detection of crosslinked targets (Figure 4-2). Using this method they performed in vitro crosslinking experiments between GCN4 and purified Swi/Snf complex. From these experiments they report that several components of this complex are targets of GCN4 (Swi1, Snf5, Swi2/Snf2).⁷ Like ¹²⁵I-PEAS crosslinking method, Sulfo-SBED has some disadvantages. One aspect is that Sulfo-SBED labeling is not site-specific which means that all nine of the lysines in the GCN4 TAD can be

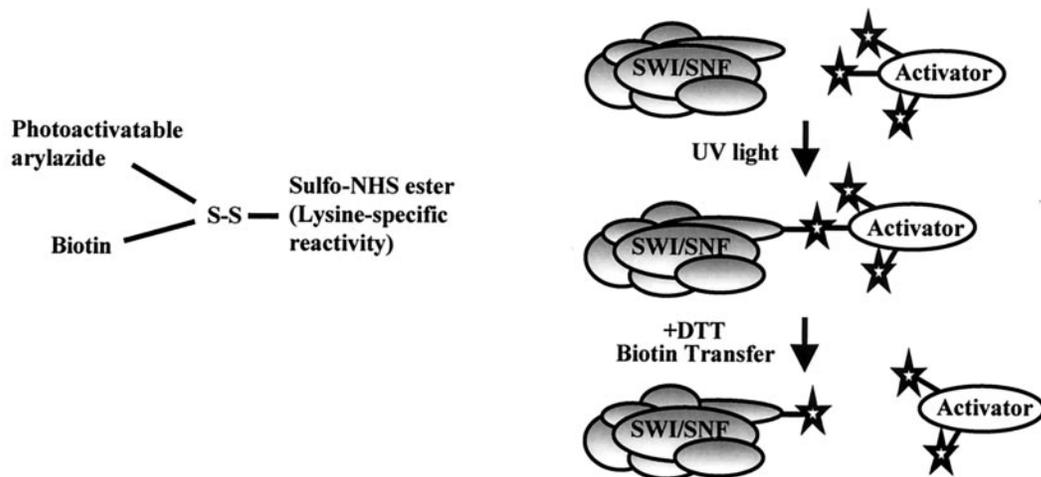


Figure 4-2: Sulfo-SBED crosslinking strategy. Left, organization of aryl azide conjugated to Biotin (for Western blot analysis) and sulfo-NHS (conjugates to lysine) through a disulfide bond. Right, photo-crosslinking label transfer scheme with a transcriptional activator and the Swi/Snf complex.⁷ Reproduced with permission.

labeled. Multiple modifications could have the problem of affecting the interactions and function of Gcn4. Further because they used full-length GCN4, all of the lysines throughout the protein were potentially labeled. In this instance, it is difficult to know which portion of the protein was contributing to the crosslinking.⁷

In summary, previous methods used to identify targets of transcriptional activators using aryl azide crosslinking have been carried out in vitro and a number of activator binding partners have been identified. However, the conclusions that can be drawn are limited due to the positioning and reactivity of the crosslinking moiety employed. The in vivo crosslinking strategy discussed in Chapter 2 has clear advantages over those previously used in that the crosslinker is incorporated site specifically and, further, the crosslinking reactions are carried out in the cellular setting. In addition, as discussed in the subsequent section, the reactivity of the photocrosslinking moiety can have a profound affect on the suite of binding partners identified through crosslinking experiments.

B.2 *p*-benzoyl-L-phenylalanine and *p*-azido-L-phenylalanine reactive mechanisms and amino acid preferences

Two photoactivatable nonnatural amino acids have been successfully incorporated into proteins in vivo (*S. cerevisiae*), pBpa and pAzpa, using nonsense suppression technology.¹² In Chapter 2, I showed that pBpa can be incorporated into positions along the Gal4 TAD and that in vivo crosslinking was feasible. Here I will discuss the reactive mechanisms of both pBpa and pAzpa and with what amino acids they prefer to react, information that affects the experimental outcomes (Figure 4-3).

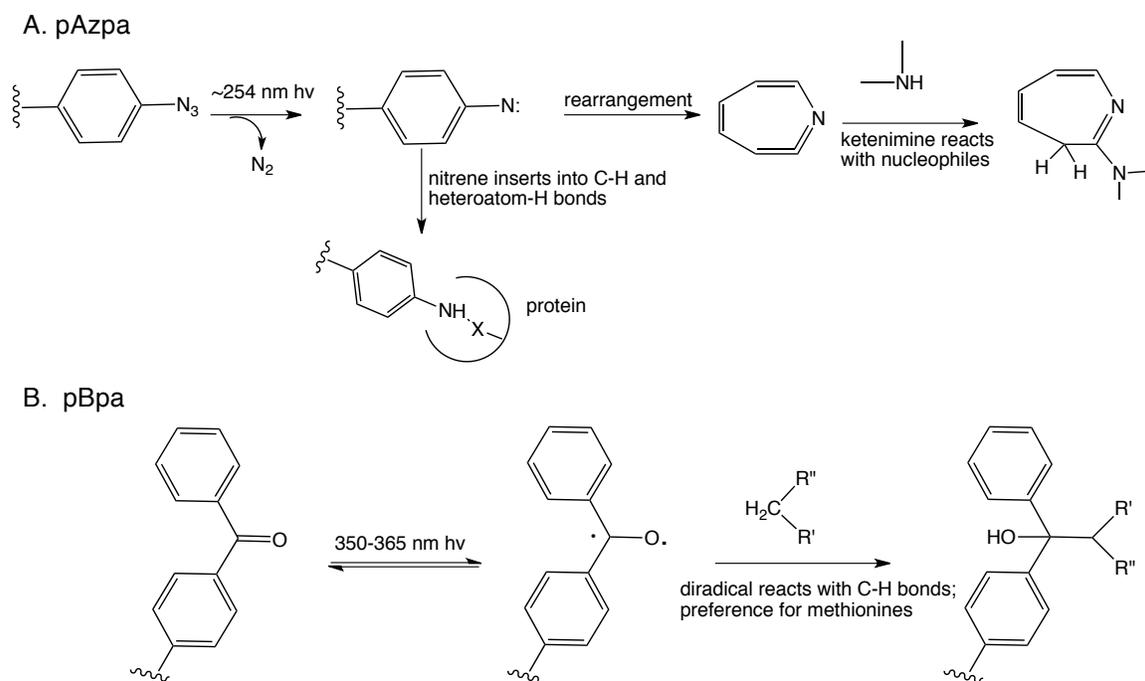


Figure 4-3: pAzpa and pBpa reactive mechanisms for crosslinking. A. pAzpa mechanism. B. pBpa mechanism.^{4,10,13}

B.2.a *p*-benzoyl-L-phenylalanine

pBpa forms a diradical upon UV irradiation at 350-365 nm where it then is able to insert into C-H bonds found in protein backbones and amino acid side chains (Figure 4-3). In the absence of any orientational preferences it reacts with the following carbons in decreasing order $\text{NCH}_x > \text{-SCH}_x > \text{methine} > \text{C=CCH}_x > \text{-CH}_2\text{-} > \text{-CH}_3$.^{10,14} Experimentally pBpa reacts preferentially with Met

where it will react at distances beyond the 3.1 Å reactive radius.¹⁵ This should be considered when analyzing data as the selectivity of Met in pBpa crosslinking may reduce the precision of crosslinking direct targets. In an early study by DeGrado and coworkers using pBpa crosslinking with biomolecules two model α -helical peptides were synthesized containing pBpa at two different positions (Figure 4-4).¹⁶ In the presence of calmodulin both peptides preferentially labeled

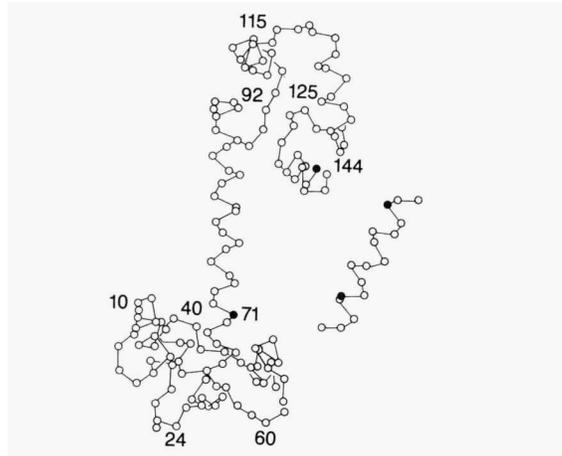


Figure 4-4: Ball and stick model of Calmodulin (left) and the 17 residue α -helical peptide (right). Met-144 and Met-71 that become crosslinked with the peptide are labeled in the picture with filled dots and the filled dots on the peptide indicate where pBpa was placed in the experiments. Only one pBpa at a time was present in each peptide tested.¹⁶ Reproduced with permission.

two different methionines located in the hydrophobic cavity on calmodulin showing the selectivity of pBpa for methionine.¹⁶ It is also important to consider the geometric requirements of pBpa for efficient crosslinking. Even in the presence of amino acids that are preferred by pBpa, crosslinking may not occur due to geometric constraints as it was determined that the ideal angle of pBpa diradical attack is 108.9° and the attack occurs almost in-plane (Figure 4-5).¹⁴ One advantage of pBpa crosslinking is that the diradical can be maintained up to 120 μ s where it either inserts into a C-H bond or relaxes back down to the initial state where it can be excited again aiding in the crosslinking efficiency.¹⁰ For example, if a protein-protein interaction is not captured during the first excitation due to distance or improper geometries from protein conformation perhaps during a second excitation the protein-protein interaction would be captured.

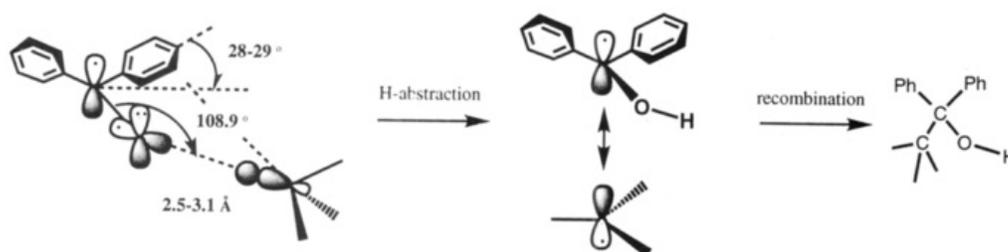


Figure 4-5: Geometric constraints for pBpa hydrogen abstraction for C-H insertion.¹⁴ Reproduced with permission.

B.2.b. *p*-azido-*L*-phenylalanine

The amino acid preference for pAzpa crosslinking is less clear because of its more complex crosslinking mechanism compared to pBpa (Figure 4-3, Table 4-1). During excitation at ~254 nm it forms a nitrene and it is at this state that it

Photo-Crosslinking moiety	Amino Acid preference	Mechanism	reference
pBpa	1. Met, Lys, Arg 2. Leu, Valine	1. prefers CH ₂ groups adjacent to heteroatoms 2. reacts with tertiary centers	Dormán and Prestwich ¹⁴ , 1994
pAzpa	1. Phe, Tyr, Trp, His 2. Lys, His, Trp 3. Ser, Thr, Tyr 4. Cys	1. nitrene inserts into the aryl C-H (also other C-H containing sidechains) 2. N-H insertion possible with the rearranged product and nitrene 3. O-H insertion possible with the rearranged product and nitrene 4. S-H insertion possible with the rearranged product and nitrene	Bayley and Staros ¹⁷ , 1984

Table 4-1: Amino acid reactivity preference of pBpa and pAzpa crosslinking.

inserts into both C-H or heteroatom-H bonds.¹⁰ If however, insertion does not occur during the $\sim 10^{-4}$ excitation (determined for simple nitrenes in a polystyrene matrix) it will rearrange to a more stable ketenimine.^{10,18} Once rearranged it reacts with nucleophiles.¹⁰ To determine which occurs more frequently crosslinks with the nitrene for C-H or heteroatom-H bond insertion or by the ketenimine with nucleophiles Brunner and Richards designed an experiment to answer this question.¹⁹⁻²² In their study they used a synthetic phospholipid bilayer with phenyl azide attached to the fatty acyl chain with the polypeptide antibiotic gramicidin A present.²⁰ The peptide has an established orientation in the membrane with the residues Gly, Ala, Leu and Val in the center of the bilayer where the aryl azide is located and the residues Leu and Trp near the lipid polar heads.^{19,21} From the photo-crosslinking reaction they observed mostly crosslinks with Trp, the only residue with a nucleophilic side chain in gramicidin.²⁰ This experiment suggests that the majority of aryl azide crosslinking takes place between the more stable ketenimine and nucleophilic side chains in biological molecules. One disadvantage of pAzpa as a crosslinker is that because it reacts with nucleophiles, water commonly reacts with the ketenimine intermediate, decreasing the crosslinking efficiency. Another disadvantage of pAzpa compared to pBpa is the unlike pBpa, that can relax back its initial state after excitation state and be excited again, pAzpa has one chance for reactivity at the excited state before it rearranges. Also, because a lower wavelength may be required for crosslinking with pAzpa (~ 254 nm for no substitutions on the phenyl ring as substitutions have been shown to increase the wavelength used for excitation) than pBpa (~ 365 nm) more damage to biomolecules can occur.^{10,23} Overall, pBpa appears to have more advantages over pAzpa in terms of potential efficiency and less damage to biomolecules if a lower wavelength is required.

B.3 Frequency of preferred amino acids for *p*-benzoyl-*L*-phenylalanine and *p*-azido-*L*-phenylalanine crosslinking at protein interfaces

In an effort to understand the role of individual amino acid residues in PPIs, much effort has been focused on characterizing the protein-protein interface by size and shape, complementarity between surfaces, residue propensities, hydrophobicity, hydrogen bonding and electrostatic interactions.²⁴ Of the many factors affecting two proteins interacting one is particularly important when considering the use of, pBpa or pAzpa for photo-crosslinking, the propensity of residues found at the interface of PPIs.^{25,26} To exemplify this Ofran and coworkers performed a comprehensive study on residue composition at six different types of protein-protein interfaces.²⁶ They included 1812 proteins in their study and found that there is a significant difference between each of the protein-protein interfaces that they looked at (Figure 4-6). Here they compared residues

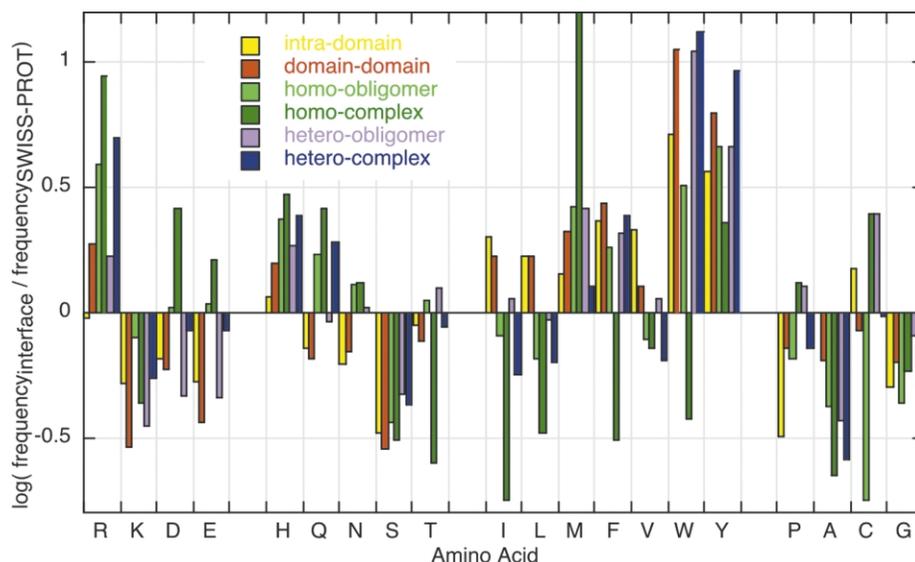


Figure 4-6: Frequency of amino acids at six different interfaces. The SWISS-PROT database was used as background. If a residue is found more frequently at the interface than its frequency in SWISS-PROT this is represented by a positive bar. If a residue is found less frequently this is represented by a negative bar. Equal representation at interface and database results in a zero bar.²⁶ Reproduced with permission.

found at interfaces to the overall amino acid composition of all proteins in SWISS-PROT, a database of protein sequences, as background.²⁶ In this way if an

amino acid is found more frequently at an interface than in the database as a whole then this results in that amino acid having a high propensity for being in that interface. Using the hetero-complex data to assess the amino acid propensities found in the type of protein-protein interfaces of transcription proteins the following amino acids are found more frequently at the interface than compared to all residues found in SWISS-PROT Trp>Tyr>Arg>Phe,His>Gln>Met. The rest of the amino acids were either equal to or less than all residues found in SWISS-PROT.²⁶ Using this information and what is known about pBpa and pAzpa photo-crosslinking we can analyze the utility of each crosslinker in protein-protein interfaces for capturing transcription proteins. Based on this, pBpa crosslinking in hetero-complex interfaces should not be limited by the absence of preferred amino acids for bond formation since methionine is one of the most frequently found amino acids at the interface. When relating the residue preference of pAzpa to those with the highest propensity in the protein-protein interface (discussed above) it is clear that in the context of hetero-complexes photo-crosslinking through the more likely ketenimine is possible. The most frequently found residue at these interfaces, Trp, possesses a nucleophilic side chain.

Overall, both pBpa and pAzpa are useful for probing protein-protein interfaces and capturing interactions whether known or yet to be discovered. For crosslinking transcription proteins, based on the propensity of amino acids found in hetero-complexes, there is evidence that both pBpa and pAzpa have the necessary amino acids present at the protein-protein interface for crosslinking to occur. If a protein-protein interaction is known then having a clear understanding of the amino acids present at the interface can be useful for selection and placement of a photo-crosslinker. If the protein binding partners are not known the best way to probe these systems is by electing to use both crosslinkers as one may capture an interaction that the other may not due to the amino acid preference of the individual crosslinker.

C. Amino acid propensity within transcription proteins

By looking at the residues involved in a well characterized PPI between a transcriptional activator and a known protein partner and by understanding the residue propensity of amino acids, preferred for pBpa or pAzpa crosslinking, in proposed targets of the amphipathic class of TADs (Chapter 1) one could predict if these interactions could be captured through photo-crosslinking. The residues involved in the well-characterized interaction of the transactivation domain of the p53 tumor suppressor and its transcriptional repressor MDM2 are (Phe19, Trp23, Leu26 and Met50, Leu54, Leu57, Gly58, Ile61, Met62, Tyr67, His73, Val75, Phe91, Val93, His96, Ile99, and Tyr100 respectively (Figure 4-7).²⁷ Based on this information an experiment designed with pBpa as a crosslinking moiety placed on p53 would be predicted to result in efficient crosslinking as two methionines are located at the interface. Conversely, when designing a similar experiment with pAzpa as the crosslinker perhaps incorporating in into the MDM2 protein would more likely result in crosslinked products as tryptophan is buried in the MDM2 cleft.²⁷ This is just one example of utilizing what is known about crosslinking moieties and residues found at the interface of a TAD and its direct protein partner. This same rational can be applied to study other PPIs involved in other biological processes.

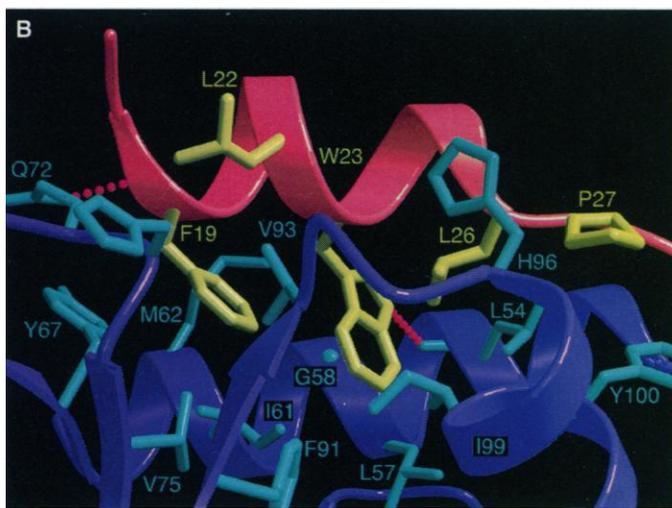


Figure 4-7: Structure of p53 peptide with MDM2. Residues at the interface are labeled p53 (yellow) and MDM2 (blue).²⁷ Reproduced with permission.

An examination of the amino acid propensity in proteins that have been suggested as targets of amphipathic TADs, discussed in Chapter 1, can be useful for determining the probability of crosslinking success. To this end, I have carried out an amino acid analysis on 10 proteins, 9 of which have been suggested as potential targets of amphipathic TADs and one is a known partner of Gal4, Gal80. I looked at the overall propensity of four residues in these proteins Met, Trp, His and Leu (Table 4-2). Met abundance was studied because it is the residue that is preferred for pBpa crosslinking. Trp and His were chosen because of their nucleophilicity, preferred for pAzpa, and because they are amongst the residues thought to be most prevalent at hetero-complex protein interfaces (discussed above). Experimentally, Leu is a residue that is not typically thought of as being one of the highest reactive amino acids with either pBpa or pAzpa so it was used as a comparison. What was observed from this study was that Met, Trp and His were all considerably less abundant than the less reactive Leu. Met was more abundant than Trp and His and in the 10 proteins studied there were ~2.5 times more Mets than Trps. Based on this information, it could be concluded that photo-crosslinking with pBpa would result in more protein crosslinks because there are more Mets. Because the amino acids that are more reactive are not found in excess throughout the protein indicates that during irradiation the reactions are not occurring because there is an overabundant amount of the highly reactive residues but that they are capturing a specific interaction.

Transcription proteins/potential targets of TADs from other studies

	% Met*	%Trp	%His	%Leu
Med15	3.05	0.46	1.2	6.38
Swi1	1.29	0.38	0.99	10.65
Snf2	2.41	0.65	2.17	9.57
Taf12	1.67	0.74	0.93	6.31
Tra1	2.4	1.1	2.22	13.09
Snf5	1.55	0.55	1.99	7.85
GCN5	3.42	0.68	2.05	7.06
TFB1	2.8	0.47	2.18	9.5
ADA2	2.07	0.92	2.76	8.99
GAL80	2.3	0.69	2.3	9.2
Average	2.296	0.664	1.879	8.86
Standard Deviation	0.676694581	0.219959592	0.618105529	2.083165327
*includes n-terminal Met				

Arbitrary Proteins

HSP82	1.83	0.71	0.56	9.31
COQ5	3.26	0.98	1.95	7.82
mps2	0.52	1.29	1.29	8.79
bud1	2.21	0.37	0.74	7.72
Average	1.955	0.8375	1.135	8.41
Standard Deviation	1.13179798	0.391524797	0.625806147	0.666820816

Proteins identified from MudPIT

CCR4	2.62	1.19	3.11	10.99
mlp2	1.61	0.42	1.07	11.26
mot2	1.36	0.34	3.58	6.64
ssl2	2.85	0.71	2.37	7.59
abf1	1.5	0.27	5.34	6.29
ctr9	1.86	0.93	2.14	12.44
hrr25	2.63	0.81	2.83	9.51
Average	2.061428571	0.667142857	2.92	9.245714286
Standard Deviation	0.620307144	0.339249873	1.333516654	2.43723106

Table 4-2: Percent Met, Trp, His and Leu in transcription proteins, arbitrary proteins and potential targets of Gal4 identified by MudPIT.

Next, I looked at the protein composition of the potential protein targets of Gal4 that were identified by MudPIT in Chapter 3 (Table 4-2). Like the analysis of the proteins described above Met, Trp and His were less abundant than Leu. What was different in these proteins was that there was more His than Met. With this information pAzpa may be a better crosslinker for capturing these interactions. Finally, because in vivo crosslinking is useful for studying proteins beyond transcription I also looked at the composition of unrelated arbitrary yeast proteins spanning a heat shock protein to a mitochondrial protein for understanding the utility of this technique for studying other biological processes. Of this small collection of proteins I found a similar trend as with the transcription proteins where Met, Trp and His represented significantly less than Leu and that there are more Met than His and Trp. Overall, it appears that pBpa would be the more efficient crosslinker for capturing PPIs base on the amino acid composition; however pAzpa should not be dismissed as there are some interactions that may be missed by dismissing this crosslinker.

D. In vivo crosslinking with p-benzoyl-L-phenylalanine and p-azido-L-phenylalanine at the Gal4-Gal80 interface

The Gal4-Gal80 interaction has been well characterized through structural studies, protein mutagenesis and by in vivo crosslinking.²⁸⁻³⁰ In Chapter 2 the interaction was further characterized by showing an extending binding surface demonstrated by a crosslinked product corresponding to the correct MW of Gal4-Gal80.³⁰ One observation from this experiment was that crosslinking efficiency varied at the Gal4-Gal80 crosslinked product when pBpa was incorporated at different positions along the Gal4 TAD even at positions important for maintaining the Gal4-Gal80 interaction.²⁹ It was hypothesized that by incorporating pAzpa into the positions on Gal4 that produced little crosslinking with Gal80 the interaction could be captured.

Shown in Figure 4-8, crosslinking at 10 positions produces varying amounts of crosslinked Gal4-Gal80. Particularly perplexing is that although Gal4 (851-871) has been shown to comprise the Gal80 binding site residue Phe856

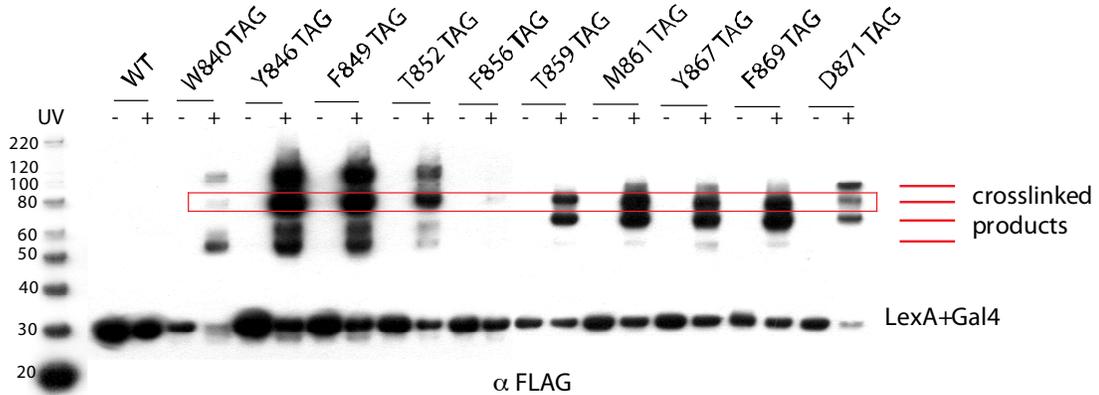


Figure 4-8: In vivo photo-crosslinking with pBpa incorporated at 10 positions along the Gal4 TAD. The red box highlights the a crosslinked band corresponding to the molecular weight of Gal4-Gal80.³⁰

within this region produced little crosslinking to Gal80 and residues outside of this region crosslinked better with Gal80 than Phe856.^{29,30} Interestingly, this residue has been shown to be critical for maintaining Gal4-Gal80 interaction because mutating Phe856 to cysteine resulted in a disruption in the interaction.²⁹ The lack of crosslinking was not due to a disruption in the Gal4-Gal80 interaction from pBpa being incorporated into Phe856 as transcription remained repressed in the presence of glucose, an indicator of Gal80 maintaining contact with Gal4 by repressing transcription.³⁰ Further, the lack of crosslinking at this position is not due to reduced pBpa incorporation at this site as it was show in Chapter 2 that incorporation efficiency at this site was equal to or more than other residues that crosslinked more efficiently with Gal80 (Figure 4-8 compare Phe856 to Phe869). Thoden et. al. solved the structure of *Kluyveromyces lactis* (*K. lactis*) Gal4-Gal80 co-crystallized where the Gal4 peptide of *K. lactis* is thought to mimic the structure of *S. cerevisiae* Gal4.²⁸ They used a 22 residue *K. lactis* Gal4 TAD peptide that has ~68% sequence homology with Gal4 from *S. Cerevisiae* (Figure 4-9). Fortunately, Phe856 (*S. cerevisiae* residue numbering) is shared between the two organisms. From the co-crystallized complex they report that Phe856 lies

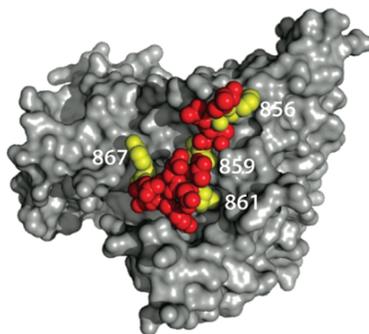


Figure 4-9: Structure of Gal4 peptide in complex with Gal80. Red and yellow highlights the Gal4 peptide (yellow are residues used in this study except 867) and gray represents Gal80 both from *K. lactis*. The residue numbering is from *S. cerevisiae*. PDB ID 3EIK.²⁸ Reproduced with permission.

in a hydrophobic patch consisting of two valines and it is suggested that there Phe856 participates in a stacking interactions with the side chain of a histidine.²⁸ Considering the side chain preference for pBpa crosslinking those particularly reactive residues, like methionine, are not present. Another reason that pBpa crosslinking does not occur at this site could be due to unfavorable geometries for C-H insertion between those present.

We hypothesized that pAzpa could be incorporated at positions along the Gal4 TAD to probe the Gal4-Gal80 interaction and perhaps produce crosslinking of Gal4-Gal80 where pBpa could not. To this end four residues on Gal4 TAD were chosen for comparison with pBpa crosslinking at similar incorporation sites and crosslinking conditions (Phe849, Phe856, Thr859, Met861). Phe849 was one of mutants that formed the most amount of pBpa crosslinks with Gal80. To test a site with the most amount of crosslinks with a different crosslinker would show the difference in crosslinking efficiency between two crosslinkers at a site already known to be permissive to a point mutation. When pBpa was incorporated into Phe856 low crosslinking efficiency was observed. pAzpa may capture this interaction when incorporated into Phe856 especially since it is nearby a histidine. There are two ways the pAzpa may form crosslinking with Gal80 at this position through either the excited nitrene, that inserts into C-H and heteratom-H bonds or the more stable ketenimine may insert into the nucleophilic

side chain. Thr859 was chosen because it is the closest residue mutated in the initial pBpa study to Phe856 to study the changes in crosslinking in nearby position. Met861 was chosen because the structural information suggests that it participates in intramolecular stacking with Tyr865 (Gal4 TAD *S. Cerevisiae* residue numbering) which would suggest that crosslinking to Gal80 could be reduced at this position due to this intramolecular interaction although pBpa was able to form crosslinks at this position.²⁸ At these four residues on Gal4 pAzpa was incorporated and photo-crosslinked in vivo (Figure 4-10). Incorporation was possible at all residues tested with little read-through. An exciting result was that pAzpa crosslinking at position Phe856 produced a band corresponding to the MW of Gal4-Gal80 crosslinked product. Additionally, of the four residues that were irradiated position 856 produced the most prominent crosslinked band at that MW, the opposite of what was observed with pBpa crosslinking at the same positions. Interestingly, Phe849, the position with the most crosslinking to Gal80 by pBpa was exceedingly reduced in pAzpa crosslinking at the same position. It is not certain but the reduced crosslinking to Gal80 at position Met861 could be a result of pAzpa crosslinking at this position to the Tyr865 due to its nucleophilic character.

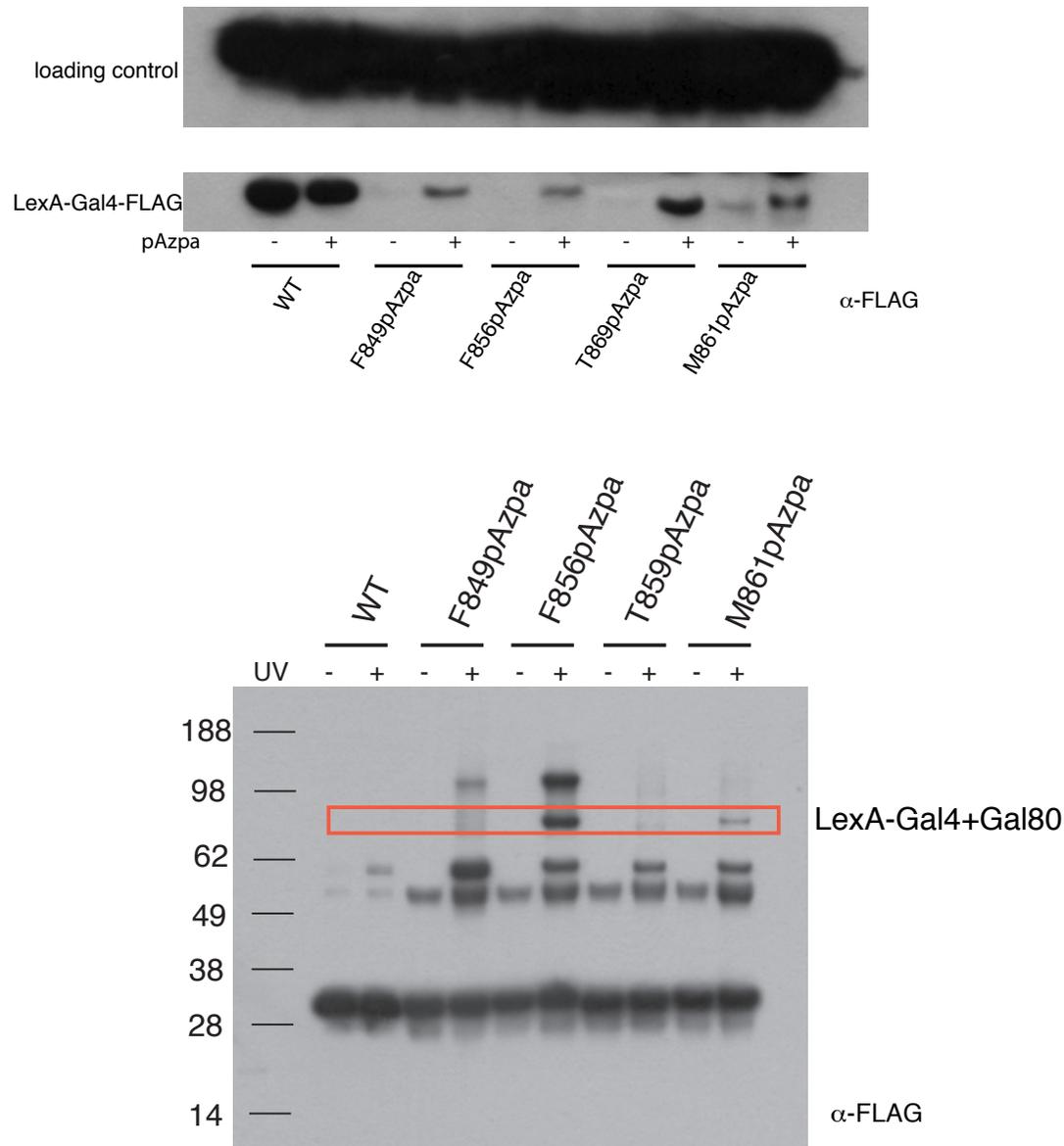


Figure 4-10: Incorporation and in vivo photo-crosslinking with pAzpa in Gal4. Top panel, pAzpa is incorporated into F849, F856, T859, M861. Bottom panel, photo-crosslinking results in varying abundance of crosslinked Gal4-Gal80 corresponding to 80 kDa molecular weight (red box) a different band patterns. Position F856 is shows more crosslinked product with pAzpa than pBpa crosslinking at the same position.

In conclusion, pAzpa was able to capture a Gal4-Gal80 interaction that pBpa was not (at Phe856) when placed under similar conditions. Additionally, crosslinking at position Phe849 to Gal80 was significantly less than what is observed with pBpa crosslinking. This is also true for Thr859 and Met861. Overall, pBpa is a more efficient crosslinker and is able to capture the Gal4-

Gal80 interaction at more of the positions tested than pAzpa. Further, an experimental observation is that pAzpa incorporation is lower and that the crosslinking appears less efficient overall in the system studied than pBpa. When testing for incorporated pAzpa ~0.5 more OD is collected and a 25% larger sample is loaded onto the gel in order to observe incorporated protein. It is not clear if the reduction in crosslinking is from lower yields in pAzpa incorporated protein. It was observed that a more sensitive substrate was needed for developing the Western blots for observing crosslinked product, an indicator of low protein abundance. In summary, careful thought should be placed into experimental design for crosslinking in biological systems from the type of crosslinking moiety to the position of the selected moiety. If for example pBpa was only placed at Phe856 little crosslinking would be observed with Gal80. The best way to capture biomolecular interactions is by placing the crosslinker at various positions. In this way, one can study the binding surface in terms of residues interacting directly with the protein. Incorporating a crosslinker at different positions may capture proteins that are interacting with only a small portion of that protein and it could be missed if a few residues are used in the study. Additionally, a thorough study would incorporate more than one crosslinking moiety for capturing interactions missed by the other moiety.

E. Methods

LS41 [JPY9::pZZ41, *Mat α his3 Δ 200 leu2 Δ 1 trp1 Δ 63 ura3-52 lys2 Δ 385 gal4 URA::pZZ41*] yeast was used for all experiments. pAzpa was purchased from Bachem.

Table of Plasmids used in this study

Plasmid name	Function
pLexAGal4	Expresses LexA(1-202)+Gal4(840-881)+FLAG tag
pLexAGal4 849TAG, pLexAGal4 856TAG, pLexAGal4 859TAG, pLexAGal4 861TAG	Express LexA(1-202)+Gal4(840-881)+FLAG tag with a TAG

	replacing the codon of the existing amino acid
pSNRtRNA-pAzpaRS	Expresses tRNA under the control of the SNR52 promoter and contains synthetase specific for pAzpa

Incorporation of pAzpa into LexA(1-202)+Gal4(840-881)

LS41 yeast were transformed with various pLexAGal4 TAG mutant plasmids and pSNRtRNA-pAzpaRS. Individual colonies were grown in 5 mL SC media containing 2% glucose but lacking uracil, histidine and tryptophan for selection. The cultures were incubated overnight at 30°C and agitated at 250 rpm. Following incubation, these cultures were used to inoculate 5 mL cultures of SC media containing 2% glucose, with or without 2 mM pAzpa (dissolved in 50 µL of 1M NaOH), and 50 mL 1M HCl which were subsequently incubated overnight at 30°C with agitation to an OD₆₆₀ of 0.8-1.2. 3 ODs of cells were isolated, washed with 100mM PBS water, flash freezed and stored at -80 °C. The samples were lysed in 10 µL 4x NuPAGE LDS Sample buffer (Invitrogen), 15 µL Lysis Buffer (50 mM Tris-Acetate, pH 7.9, 150 mM KOAc, 20% glycerol, 0.2% Tween-20, 2 mM b-mercaptoethanol, 2 mM MgOAc) 5 µL 1M DTT and analyzed using Western blot with the anti-FLAG(M2) antibody (Sigma).

In vivo cross-linking

To perform in vivo cross-linking, individual colonies of each pLexAGal4 TAG mutant were grown in 5 mL SC media containing 2% glucose but lacking uracil, histidine and tryptophan for selection. The cultures were incubated overnight at 30 °C and agitated at 250 rpm. Following incubation, these cultures were used to inoculate 50 mL cultures of SC media containing 2% glucose, with 2 mM pAzpa (dissolved in 0.5 mL of 1M NaOH), and 0.5 mL 1M HCl, which were subsequently incubated overnight at 30 °C with agitation to an OD₆₆₀ of 0.8-1.0. For each mutant, 50 ODs of cells were isolated, washed with sterile 100mM PBS and either resuspended in 2 mL of SC media containing 2% glucose but lacking uracil, histidine and tryptophan and irradiated for 1 h with 365 nm light (Eurosolar

15 W UV lamp) with cooling or immediately flash freeze and stored at -80 °C. Following irradiation, all the cells were pelleted and stored at -80 °C until lysis. For lysis, cells were resuspended in 600 µL Lysis buffer (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche) and lysed using glass beads by vortexing at 4 °C. Subsequently, the lysate was pelleted and the supernatant incubated with 10 µL of LexA antibody (N-19, Santa Cruz Biotechnologies) for 2 h at 4 °C for immunoprecipitation. The protein bound to the antibody was isolated by incubation for 1 h with ~40 µL of prewashed protein G magneticbeads (DynaL Corporation, Invitrogen, Carlsbad, CA) at 4 °C. The beads were washed 3X with 1 mL Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 1% Na-Deoxycholate and 1 mM EDTA) and stored dry at -20 °C. The protein was eluted from the beads by heating at 95 °C for 10 min in NuPAGE LDS Sample buffer (Invitrogen, Carlsbad, CA) containing DTT and probed using Western Blot analysis using anti-FLAG (M2) antibody (Sigma, St. Louis, MO).

F. References

- (1) Ruffner, H.; Bauer, A.; Bouwmeester, T. *Drug Discov Today* **2007**, *12*, 709.
- (2) Fuentes, G.; Oyarzabal, J.; Rojas, A. M. *Curr Opin Drug Discov Devel* **2009**, *12*, 358.
- (3) Yin, H.; Hamilton, A. D. *Angew Chem Int Ed Engl* **2005**, *44*, 4130.
- (4) Invitrogen *Molecular Probes The Handbook*.
- (5) Zhou, A. T.; Bessalle, R.; Bisello, A.; Nakamoto, C.; Rosenblatt, M.; Suva, L. J.; Chorev, M. *Proc Natl Acad Sci U S A* **1997**, *94*, 3644.
- (6) Chin, J. W.; Schultz, P. G. *Chembiochem* **2002**, *3*, 1135.
- (7) Neely, K. E.; Hassan, A. H.; Brown, C. E.; Howe, L.; Workman, J. L. *Mol Cell Biol* **2002**, *22*, 1615.
- (8) Reeves, W. M.; Hahn, S. *Mol Cell Biol* **2005**, *25*, 9092.
- (9) Fishburn, J.; Mohibullah, N.; Hahn, S. *Mol Cell* **2005**, *18*, 369.
- (10) Tanaka, Y.; Bond, M. R.; Kohler, J. J. *Mol Biosyst* **2008**, *4*, 473.
- (11) Ebright, Y. W.; Chen, Y.; Kim, Y.; Ebright, R. H. *Bioconjug Chem* **1996**, *7*, 380.
- (12) Chen, S.; Schultz, P. G.; Brock, A. *J Mol Biol* **2007**, *371*, 112.

- (13) Kotzyba-Hibert, F.; Kapfer, I.; Goeldner, M. *Angew Chem Int Ed Engl* **1995**, *34*, 1296.
- (14) Dorman, G.; Prestwich, G. D. *Biochemistry* **1994**, *33*, 5661.
- (15) Wittelsberger, A.; Thomas, B. E.; Mierke, D. F.; Rosenblatt, M. *FEBS Lett* **2006**, *580*, 1872.
- (16) O'Neil, K. T.; Erickson-Viitanen, S.; DeGrado, W. F. *J Biol Chem* **1989**, *264*, 14571.
- (17) Hagan, B.; Staros, J. V. *Azides and nitrenes : reactivity and utility*; Academic Press: Orlando, Fla., 1984.
- (18) Knowles, J. R. *Acc. Chem. Res.* **1972**, *5*, 155.
- (19) Staros, J. V. *Trends in Biochemical Sciences* **1980**, *5*, 320.
- (20) Brunner, J.; Richards, F. M. *J Biol Chem* **1980**, *255*, 3319.
- (21) Weinstein, S.; Wallace, B. A.; Blout, E. R.; Morrow, J. S.; Veatch, W. *Proc Natl Acad Sci U S A* **1979**, *76*, 4230.
- (22) Kelkar, D. A.; Chattopadhyay, A. *Biochim Biophys Acta* **2007**, *1768*, 2011.
- (23) Geiger, M. W.; Elliot, M. M.; Karacostas, V. D.; Moricone, T. J.; Salmon, J. B.; Sideli, V. L.; St. Onge, M. A. *Photochem. Photobio.* **1984**, *40*, 545.
- (24) Jones, S.; Thornton, J. M. *Proc Natl Acad Sci U S A* **1996**, *93*, 13.
- (25) Glaser, F.; Steinberg, D. M.; Vakser, I. A.; Ben-Tal, N. *Proteins* **2001**, *43*, 89.
- (26) Ofran, Y.; Rost, B. *J Mol Biol* **2003**, *325*, 377.
- (27) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. *Science* **1996**, *274*, 948.
- (28) Thoden, J. B.; Ryan, L. A.; Reece, R. J.; Holden, H. M. *J Biol Chem* **2008**, *283*, 30266.
- (29) Ansari, A. Z.; Reece, R. J.; Ptashne, M. *Proc Natl Acad Sci U S A* **1998**, *95*, 13543.
- (30) Majmudar, C. Y.; Lee, L. W.; Lancia, J. K.; Nwokoye, A.; Wang, Q.; Wands, A. M.; Wang, L.; Mapp, A. K. *J Am Chem Soc* **2009**, *131*, 14240.

Chapter 5

Future Directions

This body of work has demonstrated the development and implementation of in vivo photo-crosslinking for studying PPIs between a transcriptional activator Gal4 and its known target Gal80. Additionally, the initial mass spectrometry experiments identified potentially new targets. With additional experimentation this technique could be used to map the entire network of PPIs that take place during transcription. Another interesting aspect of crosslinking revealed was the idea of incorporating different crosslinking moieties, pBpa or pAzpa, into similar protein sites for understanding the potential binding site at the interface of two proteins. Overall these techniques are generally useful for not only identifying unknown partners of Gal4 TAD but for studying and identifying PPIs involved in other biological processes. Outlined here are future ongoing areas of research using photo-crosslinking to study PPIs in cellular environments.

A. In vivo crosslinking in a histone deacetylase, Rpd3

Histone deacetylases (HDACs) are primarily responsible for removing the acetyl group from lysine side chains that has been placed on by histone acetyltransferases (HATs) found in histone proteins.¹ Histones in the octomeric form are called nucleosomes which DNA wraps around and is further compact into chromatin. Removing the acetyl group on lysines of histones generates a positive charge that allows it to interact with the negatively charged phosphate backbone of DNA thereby increasing the histone-DNA interaction and making it more difficult for chromatin to be relaxed, an action associated with activated transcription.² A balance between acetylation and decetylation of histone lysines

is thought to be important for regulating gene expression.³ Because dysregulation is associated with diseases such as cancer inhibitors of HDACs have been developed.³ In fact the HDAC inhibitor, suberoylanilide hydroxamic acid (SAHA), has been approved for the treatment of cutaneous T cell lymphoma.⁴ Although HDACs principally modify lysines on histones there are many other potential non-histone substrates of these proteins, such as NF- κ B and androgen receptor, have been thought to have increased acetylation upon addition of HDAC inhibitors.⁵⁻⁷

In collaboration with Noah Wolfson in the laboratory of Dr. Carol Fierke we will use in vivo photo-crosslinking for characterizing substrates of the yeast homologue, Rpd3, of mammalian Class I HDACs which are primarily localized in the nucleus.² Once these substrates have been identified a more detailed look at the factors that affect substrate specificity (amino acid sequence, structural preference etc.) will be studied.

Rpd3 has been shown in vitro to deacetylated lysines found in H2A, H2B, H3 and H4 with the exception of K16.⁸ To identify other substrates of Rpd3 in vivo three sites were selected for pBpa incorporation (Figure 5-1) based on the structure characterize by Vannini et. al. of the mammalian homologue HDAC8 in complex with a substrate (a p53 derived diacetylated peptide with a fluorogenic coumarin group at its carboxyl terminus or a hydroxamate inhibitor) and on homology maps base on that structure.⁹ HDAC8 shares 43% sequence identity and 65% sequence similarity to Rpd3 and is used for structural studies because unlike other HDACS HDAC8 is found functional as a singular protein instead of a multiprotein complex.^{1,9} Before applying in vivo crosslinking to this system first pBpa was shown to incorporate into Rpd3 with little read-through (Figure 5-1, top panel). In a crosslinking experiment all positions with pBpa incorporated crosslinked products are observed. It appears in this initial study that each position tested has a unique crosslinking product protein pattern which can be indicative of different protein targets captured at different positions. The next step in this project is to incorporate pBpa into more sites along pRpd3 and to

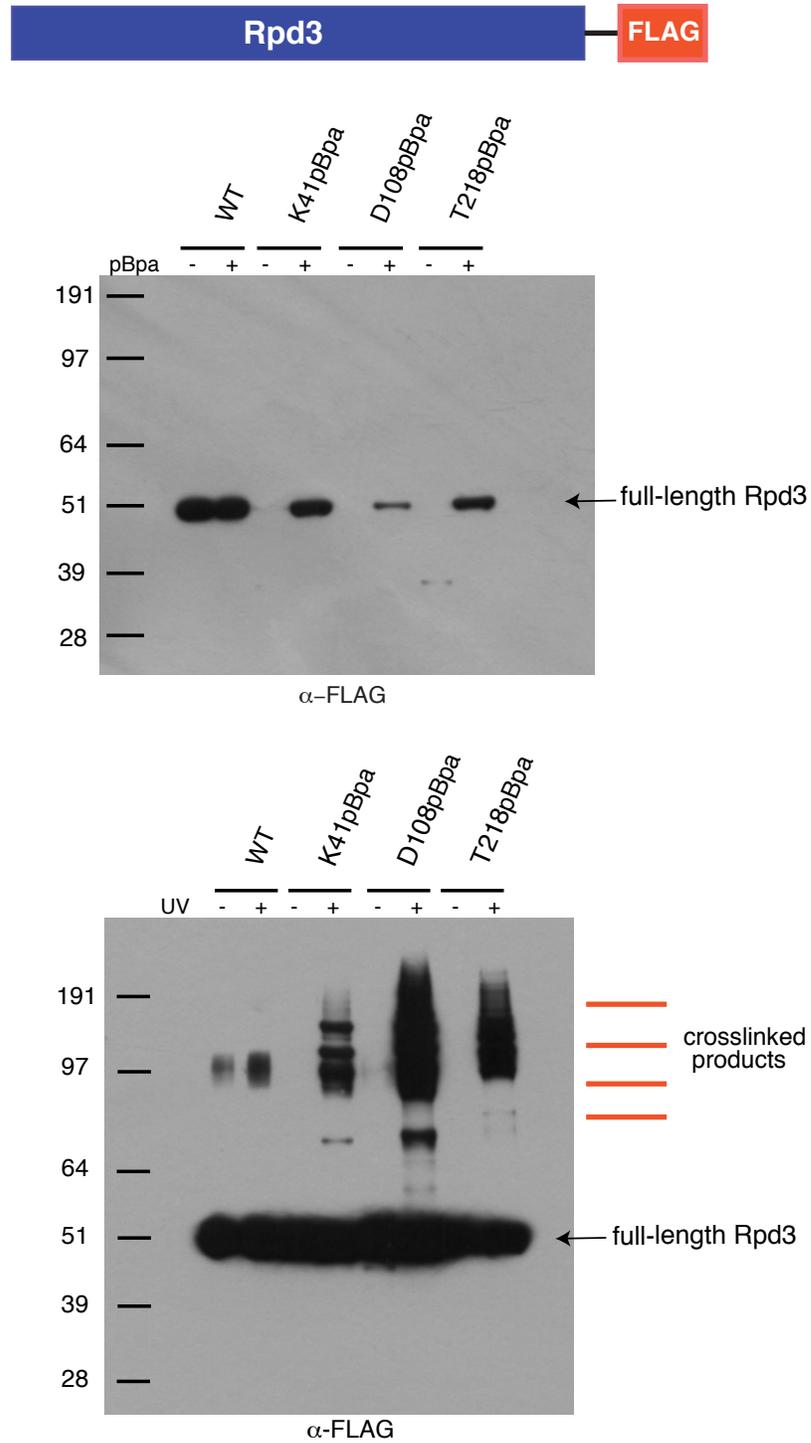


Figure 5-1: Incorporation and in vivo photo-crosslinking pBpa in Rpb3. Top panel, pBpa is incorporated into three position in Rpd3. Bottom panel, crosslinking with pBpa incorporated at three position in Rpd3 show different band pattern at each position tested.

carry out crosslinking in the presence of an HDAC inhibitor to eliminate crosslinked bands that are not actual substrates but are proteins in complex with Rpd3. Another interesting set of crosslinking experiments can be designed to map the substrate binding site once they have been identified by placing pBpa at various distances from the substrate to see if crosslinking is still possible. In this way an extensive binding site can be characterized for individual substrates. The long term goal of this project is to use the yeast model for mapping the substrates of HDACs in human cells.

B. Photo-crosslinking with transcriptional activation domains in mammalian cells

Characterizing the direct protein targets of the prototypical amphipathic transcriptional activator Gal4 in yeast is used as a model for understanding the protein interactions of other transcriptional activators of the same class in other cellular contexts. To this end, a project in our laboratory (carried out by Conor Doss) aims at using the similar crosslinking approach carried out in yeast in mammalian cells for detecting protein partners of Gal4 TAD. After optimizing the conditions required for incorporating a crosslinker into the well studied Gal4 TAD a mammalian transcriptional activator will be used for mapping the transcriptional proteins in a mammalian cells. This information will be useful for understanding fundamental transcriptional mechanistic questions and for designing therapeutics that can inhibit the PPIs between a TAD and its direct protein targets that lead to aberrant gene expression in diseases like diabetes and cancer.

C. Studying the amino acid preference with photo-crosslinkers

As discussed in Chapter 4 the amino acid preference for pBpa is known yet it is not clear what this means in terms of specificity for crosslinking with proteins. In Gal4 TAD (840-881) there is one methionine (Met861), the preferred residue for pBpa crosslinking.¹⁰ At this residue we have incorporated pBpa and have shown that it crosslinks to Gal80 indicating that at this position the two

proteins are at an acceptable distance for pBpa crosslinking. What has not been tested is the importance of methionine in pBpa crosslinking between Gal4-Gal80 when pBpa is incorporated along Gal80. To study this the lone methionine in Gal4 can be mutated to phenylalanine, the closest residue resembling pBpa of the natural amino acids, to ensure that the contact with Gal80 could still be made and the effect of methionine in pBpa crosslinking efficiency could be observed. Although this experiment would initially be carried out in a cell a more simplified in vitro experiment can be designed with peptides of Gal80 and Gal4 coupled with mass spectrometric analysis for quantifying the crosslinking efficiency.

D. Additional experiments for verifying direct targets of Gal4 identified by mass spectrometry

Once all the potential targets have been identified with MudPIT further verification with additional methods will be applied since samples were submitted in solution to maximize proteins identified. Another sensitive and quantitative MS method termed selected reaction monitoring (SRM) can be used to verify direct targets. SRM is applied when a protein is known and therefore the focus of the MS-MS is on detecting the known peptide fragments.¹¹ It is a more biased approach but is useful for focusing on verifying a protein that is known to be present in a sample. To aid in identifying crosslinked peptides in the more focused SRM experiments isotopically labeled pBpa can also be used to aid in identifying crosslinked peptide fragments, one of the challenges with MS analysis since part of the sequence is from one protein and another in a different protein. Beyond MS methods for verifying protein targets, confirmation with Western blots can be straightforward if the protein has a commercially available antibody. If this is not the case a yeast strain with an epitope tag associated with the protein of interest can be used integrated into the genome. This allows for the natural expression level of the protein in the cell and easy probing of a Western blot with an antibody for the epitope tag for confirmation that it is a crosslinked protein partner.

E. References

- (1) Yang, X. J.; Seto, E. *Nat Rev Mol Cell Biol* **2008**, *9*, 206.
- (2) de Ruijter, A. J.; van Gennip, A. H.; Caron, H. N.; Kemp, S.; van Kuilenburg, A. B. *Biochem J* **2003**, *370*, 737.
- (3) Walkinshaw, D. R.; Yang, X. J. *Curr Oncol* **2008**, *15*, 237.
- (4) Duvic, M. *Hematology Meeting Reports* **2008**, *2*, 39.
- (5) Drummond, D. C.; Noble, C. O.; Kirpotin, D. B.; Guo, Z.; Scott, G. K.; Benz, C. C. *Annu Rev Pharmacol Toxicol* **2005**, *45*, 495.
- (6) Chen, L. F.; Greene, W. C. *J Mol Med* **2003**, *81*, 549.
- (7) Fu, M.; Rao, M.; Wang, C.; Sakamaki, T.; Wang, J.; Di Vizio, D.; Zhang, X.; Albanese, C.; Balk, S.; Chang, C.; Fan, S.; Rosen, E.; Palvimo, J. J.; Janne, O. A.; Muratoglu, S.; Avantaggiati, M. L.; Pestell, R. G. *Mol Cell Biol* **2003**, *23*, 8563.
- (8) Suka, N.; Suka, Y.; Carmen, A. A.; Wu, J.; Grunstein, M. *Mol Cell* **2001**, *8*, 473.
- (9) Vannini, A.; Volpari, C.; Gallinari, P.; Jones, P.; Mattu, M.; Carfi, A.; De Francesco, R.; Steinkuhler, C.; Di Marco, S. *EMBO Rep* **2007**, *8*, 879.
- (10) Wittelsberger, A.; Thomas, B. E.; Mierke, D. F.; Rosenblatt, M. *FEBS Lett* **2006**, *580*, 1872.
- (11) Anderson, L.; Hunter, C. L. *Mol Cell Proteomics* **2006**, *5*, 573.

Appendix
Data from MudPIT analysis

This appendix consists of all the MudPIT mass spectrometric data discussed in Chapter 3. Table A-1 outlines the conditions for each sample prepared and Table A-2 shows the data (the proteins identified and the number of peptides in each experiment that were observed for each protein). Those highlighted in yellow represent the proteins that were only present in crosslinked samples in and are related to transcription. These are thought to be potential targets of Gal4.

Sample	Culture Size	Sugar Source	UV
Gal4 Phe849pBpa	6L	Glucose	No
Gal4 Phe849pBpa	6L	Glucose	Yes
Gal4 Phe849pBpa	6L	Galactose	No
Gal4 Phe849pBpa	6L	Galactose	Yes
Gal4 Phe849pBpa	12L	Galactose	Yes
Only tRNA/synthetase no Gal4 Phe849pBpa present	6L	Glucose	Yes

Table A-1: Samples prepared for MudPIT analysis.

Table A-2: Data from the MudPIT experiments.

Protein	Raf/Gal UV 6L	Raf/Gal UV 12L	Glu UV 6L	Raf Gal no UV 6L	Yeast UV 6L	Glu no UV 6L
LEXAGAL4 - no description	255	108	168	88	0	131
YAL003W - EFB1	0	0	0	34	0	0
YAL005C - SSA1	10	8	4	94	0	3
YAL012W - CYS3	0	0	0	6	0	0
YAL016W - TPD3	0	0	0	3	0	0
YAL021C - CCR4	0	12	0	0	0	0
YAL038W - CDC19	2	18	10	48	0	11
YAL042W - ERV46	0	5	0	2	0	0
YAL044C - GCV3	0	0	0	10	0	0
YAL060W - BDH1	17	13	0	8	0	0
YAL061W - BDH2	0	6	0	0	0	0
YBL002W - HTB2	0	0	0	10	0	0
YBL003C - HTA2	0	0	0	4	0	0
YBL005W-B - YBL005W-B	0	45	0	0	0	0
YBL015W - ACH1	0	0	0	4	0	0
YBL024W - NCL1	0	0	0	2	0	0
YBL027W - RPL19B	3	4	0	11	0	0
YBL030C - PET9	3	7	0	9	0	0
YBL039C - URA7	0	2	0	0	0	0
YBL045C - COR1	0	0	0	3	0	0
YBL058W - SHP1	0	0	0	4	0	0
YBL064C - PRX1	0	0	0	3	0	0
YBL066C - SEF1	3	31	0	3	0	8
YBL069W - AST1	0	3	0	0	0	0
YBL072C - RPS8A	0	0	0	6	0	0
YBL076C - ILS1	0	3	3	6	0	2
YBL085W - BOI1	0	0	0	0	0	3
YBL086C - YBL086C	2	0	0	0	0	0
YBL087C - RPL23A	0	0	0	8	0	0
YBL092W - RPL32	0	0	0	2	0	0
YBL099W - ATP1	0	2	0	11	0	0
YBR009C - HHF1	0	0	0	3	0	0
YBR011C - IPP1	0	0	0	13	0	0
YBR012W-B - YBR012W-B	87	45	49	0	0	26
YBR025C - OLA1	0	2	0	3	0	0
YBR031W - RPL4A	0	0	0	23	0	0
YBR035C - PDX3	0	0	0	2	0	0
YBR039W - ATP3	0	0	0	2	0	0
YBR048W - RPS11B	0	0	0	5	0	0
YBR072W - HSP26	0	0	0	9	0	0

YBR079C - RPG1	0	0	0	2	0	0
YBR084C-A - RPL19A	3	4	0	11	0	0
YBR092C - PHO3	0	0	0	4	0	0
YBR106W - PHO88	0	0	0	3	0	0
YBR109C - CMD1	0	0	0	6	0	0
YBR118W - TEF2	91	29	20	107	4	11
YBR121C - GRS1	0	0	0	3	0	0
YBR126C - TPS1	0	4	0	0	0	0
YBR127C - VMA2	0	0	0	21	0	0
YBR129C - OPY1	32	72	16	34	12	66
YBR135W - CKS1	9	17	5	2	4	8
YBR149W - ARA1	0	0	0	7	0	0
YBR181C - RPS6B	0	0	0	6	0	0
YBR189W - RPS9B	0	0	0	4	0	0
YBR191W - RPL21A	0	0	0	2	0	0
YBR196C - PGI1	0	2	0	21	0	0
YBR218C - PYC2	0	13	2	5	0	0
YBR221C - PDB1	0	0	0	2	0	0
YBR238C - YBR238C	52	24	35	15	3	57
YBR248C - HIS7	0	0	0	2	0	0
YBR249C - ARO4	0	2	0	7	0	0
YBR255W - MTC4	13	9	7	5	0	4
YBR263W - SHM1	0	3	0	3	0	0
YBR275C - RIF1	2	2	0	0	0	0
YBR286W - APE3	0	0	0	6	0	0
YCL009C - ILV6	2	20	2	17	0	3
YCL018W - LEU2	0	0	0	31	0	0
YCL019W - YCL019W	0	15	0	0	0	0
YCL030C - HIS4	0	0	0	15	0	0
YCL035C - GRX1	0	0	0	4	0	0
YCL037C - SRO9	167	163	29	139	46	196
YCL039W - GID7	0	2	0	0	0	0
YCL040W - GLK1	0	3	0	11	0	0
YCL043C - PDI1	0	0	0	6	0	0
YCR004C - YCP4	0	0	0	2	0	0
YCR012W - PGK1	0	6	4	97	0	2
YCR031C - RPS14A	0	0	0	7	0	0
YCR053W - THR4	0	0	0	17	0	0
YCR088W - ABP1	0	0	0	3	0	0
YDL004W - ATP16	0	0	0	6	0	0
YDL022W - GPD1	0	9	0	7	0	0
YDL025C - YDL025C	16	9	0	12	0	31
YDL051W - LHP1	0	0	0	3	0	0
YDL055C - PSA1	0	0	0	7	0	0

YDL061C - RPS29B	0	0	0	3	0	0
YDL065C - PEX19	0	0	0	2	0	0
YDL066W - IDP1	0	0	0	4	0	0
YDL075W - RPL31A	0	0	0	6	0	0
YDL081C - RPP1A	0	0	0	7	0	0
YDL082W - RPL13A	0	0	0	6	0	0
YDL083C - RPS16B	0	2	0	7	0	0
YDL084W - SUB2	0	0	0	6	0	0
YDL095W - PMT1	0	3	0	0	0	0
YDL099W - BUG1	0	0	0	2	0	0
YDL124W - YDL124W	0	0	0	4	0	0
YDL126C - CDC48	0	0	0	7	0	0
YDL131W - LYS21	8	0	4	5	0	0
YDL140C - RPO21	0	4	0	0	0	2
YDL145C - COP1	4	10	2	4	0	0
YDL160C - DHH1	0	0	0	0	0	3
YDL171C - GLT1	0	2	0	0	0	0
YDL182W - LYS20	0	8	4	5	0	0
YDL185W - TFP1	0	4	0	12	0	0
YDL192W - ARF1	0	2	0	5	0	0
YDL195W - SEC31	0	0	0	4	0	0
YDL201W - TRM8	0	2	0	0	0	0
YDL223C - HBT1	0	7	0	0	0	2
YDL229W - SSB1	3	0	0	41	0	0
YDL237W - AIM6	0	5	0	0	0	0
YDR002W - YRB1	0	0	0	6	0	0
YDR012W - RPL4B	0	0	0	22	0	0
YDR023W - SES1	0	0	0	8	0	0
YDR025W - RPS11A	0	0	0	5	0	0
YDR034C-D - YDR034C-D	0	15	0	0	0	0
YDR035W - ARO3	0	0	0	3	0	0
YDR037W - KRS1	0	0	0	10	0	0
YDR050C - TPI1	0	0	0	33	0	0
YDR051C - DET1	0	3	0	0	0	0
YDR064W - RPS13	0	0	0	4	0	0
YDR098C-B - YDR098C-B	0	0	0	0	9	0
YDR122W - KIN1	0	2	0	0	0	0
YDR127W - ARO1	5	2	0	4	0	0
YDR148C - KGD2	0	0	0	5	0	0
YDR150W - NUM1	0	0	0	2	0	0
YDR155C - CPR1	0	4	0	22	0	0
YDR164C - SEC1	0	2	0	0	0	0
YDR171W - HSP42	0	3	0	0	0	0
YDR172W - SUP35	3	2	4	0	0	0

YDR210C-D - YDR210C-D	0	0	47	0	9	0
YDR224C - HTB1	0	0	0	10	0	0
YDR225W - HTA1	0	0	0	4	0	0
YDR226W - ADK1	0	0	0	11	0	0
YDR232W - HEM1	10	12	3	6	0	4
YDR258C - HSP78	0	0	0	2	0	0
YDR261C-D - YDR261C-D	0	0	47	0	9	0
YDR261W-B - YDR261W-B	18	14	0	0	0	0
YDR298C - ATP5	0	0	0	5	0	0
YDR300C - PRO1	0	2	0	0	0	0
YDR304C - CPR5	0	0	0	7	0	0
YDR316W-B - YDR316W-B	84	0	47	0	9	0
YDR342C - HXT7	0	5	0	6	0	0
YDR343C - HXT6	0	5	0	0	0	0
YDR353W - TRR1	0	0	0	7	0	0
YDR354W - TRP4	0	3	0	0	0	0
YDR382W - RPP2B	0	0	0	6	0	0
YDR385W - EFT2	0	9	4	55	0	0
YDR394W - RPT3	0	0	0	2	0	0
YDR418W - RPL12B	0	0	0	17	0	0
YDR429C - TIF35	0	0	0	2	0	0
YDR432W - NPL3	35	261	9	42	53	51
YDR447C - RPS17B	0	0	0	7	0	0
YDR450W - RPS18A	0	3	4	6	0	0
YDR454C - GUK1	0	0	0	2	0	0
YDR457W - TOM1	0	4	0	0	0	2
YDR471W - RPL27B	0	0	0	7	0	0
YDR475C - JIP4	6	14	0	3	0	10
YDR477W - SNF1	79	61	19	24	4	70
YDR481C - PHO8	11	69	9	15	6	25
YDR487C - RIB3	0	0	0	3	0	0
YDR502C - SAM2	0	0	0	4	0	0
YDR510W - SMT3	0	0	0	2	0	0
YDR513W - GRX2	0	0	0	4	0	0
YDR529C - QCR7	0	0	0	5	0	0
YDR533C - HSP31	0	0	0	3	0	0
YEL024W - RIP1	0	0	0	2	0	0
YEL030W - ECM10	0	0	0	6	0	0
YEL034W - HYP2	0	0	0	12	0	0
YEL054C - RPL12A	0	0	0	17	0	0
YEL060C - PRB1	0	3	0	0	0	0
YEL071W - DLD3	0	0	2	0	0	0
YER003C - PMI40	0	0	0	8	0	0
YER009W - NTF2	0	0	0	2	0	0

YER018C - SPC25	2	4	0	0	0	0
YER025W - GCD11	0	0	0	3	0	0
YER043C - SAH1	0	5	0	18	0	2
YER062C - HOR2	0	0	0	3	0	0
YER067W - YER067W	30	63	0	19	8	22
YER068W - MOT2	0	8	3	0	0	0
YER074W - RPS24A	0	0	0	8	0	0
YER086W - ILV1	2	11	6	6	3	5
YER088C - DOT6	13	5	0	2	6	5
YER090W - TRP2	0	5	0	3	0	0
YER091C - MET6	0	0	0	27	0	0
YER094C - PUP3	0	0	0	4	0	0
YER101C - AST2	0	4	0	0	0	0
YER102W - RPS8B	0	0	0	6	0	0
YER103W - SSA4	0	0	0	25	0	0
YER113C - TMN3	0	0	0	2	0	0
YER117W - RPL23B	0	0	0	8	0	0
YER120W - SCS2	0	0	0	3	0	0
YER131W - RPS26B	0	0	0	4	0	0
YER132C - PMD1	4	3	4	0	0	5
YER138C - YER138C	0	45	49	57	0	26
YER151C - UBP3	0	3	0	0	0	2
YER155C - BEM2	9	9	0	7	0	6
YER160C - YER160C	87	45	49	0	0	26
YER165W - PAB1	0	3	0	14	0	0
YER177W - BMH1	0	0	0	16	0	0
YER183C - FAU1	2	0	0	0	0	0
YFL002W-A - YFL002W-A	0	15	0	0	0	0
YFL007W - BLM10	3	4	2	0	0	0
YFL014W - HSP12	0	0	0	2	0	0
YFL018C - LPD1	0	0	0	7	0	0
YFL034C-A - RPL22B	0	0	0	2	0	0
YFL039C - ACT1	3	10	0	22	0	9
YFL045C - SEC53	0	0	0	2	0	0
YFR004W - RPN11	0	0	0	2	0	0
YFR015C - GSY1	9	11	0	4	0	0
YFR024C-A - LSB3	0	0	0	3	0	0
YFR028C - CDC14	0	3	0	2	0	0
YFR031C-A - RPL2A	6	8	0	14	0	0
YFR033C - QCR6	0	0	0	2	0	0
YFR044C - DUG1	0	0	0	5	0	0
YFR052W - RPN12	0	0	0	2	0	0
YFR053C - HXK1	0	8	0	45	0	0
YGL008C - PMA1	0	0	0	20	0	0

YGL009C - LEU1	0	0	0	10	0	0
YGL026C - TRP5	6	14	4	16	0	0
YGL030W - RPL30	0	0	0	7	0	0
YGL031C - RPL24A	0	0	0	3	0	0
YGL037C - PNC1	5	39	3	13	7	9
YGL040C - HEM2	0	0	0	2	0	0
YGL043W - DST1	0	0	0	2	0	0
YGL056C - SDS23	0	8	0	2	0	0
YGL062W - PYC1	0	13	2	5	0	0
YGL068W - MNP1	0	0	0	3	0	0
YGL076C - RPL7A	2	5	0	9	0	0
YGL103W - RPL28	6	12	8	5	0	6
YGL105W - ARC1	0	0	0	10	0	0
YGL106W - MLC1	0	0	0	6	0	0
YGL120C - PRP43	0	3	0	0	0	0
YGL123W - RPS2	0	0	0	11	0	0
YGL135W - RPL1B	0	0	0	7	0	0
YGL137W - SEC27	0	0	0	2	0	0
YGL147C - RPL9A	0	5	2	5	0	0
YGL148W - ARO2	2	6	3	8	3	2
YGL157W - YGL157W	0	0	6	0	0	0
YGL173C - KEM1	2	19	4	4	0	3
YGL178W - MPT5	0	6	0	0	0	6
YGL187C - COX4	0	3	0	6	0	0
YGL189C - RPS26A	0	0	0	4	0	0
YGL202W - ARO8	0	5	0	3	0	0
YGL206C - CHC1	0	0	0	2	0	0
YGL234W - ADE5,7	0	0	0	4	0	0
YGL242C - YGL242C	0	0	0	5	0	0
YGL245W - GUS1	0	2	0	11	0	0
YGL253W - HXK2	0	4	0	14	0	0
YGR010W - NMA2	3	0	0	0	0	0
YGR019W - UGA1	0	2	0	0	0	0
YGR034W - RPL26B	0	0	0	3	0	0
YGR037C - ACB1	0	0	0	12	0	0
YGR038C-B - YGR038C-B	0	0	0	0	9	0
YGR061C - ADE6	0	0	0	5	0	0
YGR080W - TWF1	0	0	0	2	0	0
YGR085C - RPL11B	0	0	0	2	0	0
YGR086C - PIL1	0	0	0	12	0	0
YGR094W - VAS1	0	0	0	5	0	0
YGR116W - SPT6	6	23	0	8	0	3
YGR118W - RPS23A	0	0	0	5	0	0
YGR124W - ASN2	301	139	218	103	17	138

YGR148C - RPL24B	0	0	0	3	0	0
YGR155W - CYS4	0	0	0	17	0	0
YGR159C - NSR1	0	0	0	3	0	0
YGR161C-D - YGR161C-D	0	45	0	0	0	0
YGR161W-B - YGR161W-B	0	15	0	0	0	0
YGR185C - TYS1	0	0	0	3	0	0
YGR192C - TDH3	20	21	10	132	5	15
YGR194C - XKS1	0	9	0	0	0	0
YGR204W - ADE3	0	0	0	2	0	0
YGR209C - TRX2	0	0	0	3	0	0
YGR214W - RPS0A	0	2	0	0	0	0
YGR234W - YHB1	0	0	0	7	0	0
YGR240C - PFK1	4	15	4	15	0	2
YGR254W - ENO1	0	0	0	66	0	2
YGR264C - MES1	0	0	0	3	0	0
YGR266W - YGR266W	3	8	0	4	0	0
YGR271W - SLH1	0	0	0	2	0	0
YGR277C - CAB4	0	3	0	0	0	0
YGR279C - SCW4	0	0	0	7	0	0
YGR282C - BGL2	0	0	0	7	0	0
YGR285C - ZUO1	0	0	0	9	0	0
YHL001W - RPL14B	0	0	0	15	0	0
YHL015W - RPS20	0	0	0	3	0	0
YHL018W - YHL018W	12	26	0	4	0	9
YHL023C - RMD11	0	5	0	0	0	6
YHL033C - RPL8A	0	0	0	11	0	0
YHL034C - SBP1	0	0	0	14	0	0
YHR005C-A - TIM10	0	0	0	2	0	0
YHR007C - ERG11	0	2	0	0	0	0
YHR008C - SOD2	0	0	0	4	0	0
YHR010W - RPL27A	0	0	0	7	0	0
YHR018C - ARG4	0	0	0	2	0	0
YHR019C - DED81	0	3	0	5	0	0
YHR033W - YHR033W	0	3	0	0	0	0
YHR051W - COX6	0	0	0	4	0	0
YHR064C - SSZ1	0	0	0	9	0	0
YHR070W - TRM5	2	4	0	0	0	0
YHR073W - OSH3	0	3	0	0	0	0
YHR087W - RTC3	0	0	0	4	0	0
YHR097C - YHR097C	0	14	0	0	3	15
YHR099W - TRA1	0	2	0	0	0	0
YHR106W - TRR2	0	0	0	2	0	0
YHR108W - GGA2	0	0	0	2	0	0
YHR135C - YCK1	0	46	0	0	4	12

YHR146W - CRP1	0	0	0	2	0	0
YHR174W - ENO2	2	3	5	129	0	2
YHR179W - OYE2	0	0	0	6	0	0
YHR183W - GND1	0	8	0	18	0	0
YHR193C - EGD2	0	0	0	12	0	0
YHR203C - RPS4B	0	3	0	18	0	0
YHR205W - SCH9	5	7	4	3	0	12
YHR208W - BAT1	0	0	0	11	0	0
YHR214C-B - YHR214C-B	0	45	0	0	0	0
YHR216W - IMD2	0	0	0	4	0	0
YIL002C - INP51	0	0	0	0	0	3
YIL018W - RPL2B	6	8	0	14	0	0
YIL033C - BCY1	0	0	0	2	0	0
YIL041W - GVP36	0	0	0	2	0	0
YIL051C - MMF1	0	0	0	3	0	0
YIL053W - RHR2	0	0	0	9	0	0
YIL057C - YIL057C	0	2	0	0	0	0
YIL062C - ARC15	0	0	0	3	0	0
YIL069C - RPS24B	0	0	0	8	0	0
YIL070C - MAM33	0	0	0	6	0	0
YIL078W - THS1	0	4	0	0	0	0
YIL109C - SEC24	0	3	0	0	0	0
YIL125W - KGD1	0	10	0	8	0	0
YIL137C - TMA108	0	15	0	3	0	2
YIL138C - TPM2	0	0	0	2	0	0
YIL143C - SSL2	0	3	0	0	0	0
YIL149C - MLP2	0	2	0	0	0	0
YIL162W - SUC2	0	0	0	2	0	0
YIR037W - HYR1	0	0	0	2	0	0
YJL005W - CYR1	0	2	0	0	0	0
YJL020C - BBC1	7	28	3	8	3	8
YJL026W - RNR2	2	2	0	2	0	0
YJL034W - KAR2	0	0	0	6	0	0
YJL041W - NSP1	0	0	0	4	0	0
YJL042W - MHP1	0	0	0	0	0	3
YJL045W - YJL045W	14	5	0	0	0	0
YJL052W - TDH1	0	8	0	25	0	0
YJL056C - ZAP1	0	0	0	0	0	2
YJL080C - SCP160	0	0	0	6	0	0
YJL083W - TAX4	3	4	4	0	0	2
YJL130C - URA2	6	39	15	17	0	12
YJL136C - RPS21B	0	0	0	10	0	0
YJL138C - TIF2	0	2	0	20	0	0
YJL158C - CIS3	0	0	0	2	0	0

YJL159W - HSP150	0	0	0	2	0	0
YJL172W - CPS1	0	7	2	0	0	0
YJL177W - RPL17B	0	0	0	7	0	0
YJL190C - RPS22A	0	0	0	7	0	0
YJL191W - RPS14B	0	0	0	7	0	0
YJR009C - TDH2	0	14	0	44	0	0
YJR016C - ILV3	0	0	0	4	0	0
YJR019C - TES1	113	36	20	26	0	9
YJR033C - RAV1	0	2	0	0	0	0
YJR035W - RAD26	0	6	0	0	0	0
YJR045C - SSC1	0	0	0	29	0	0
YJR048W - CYC1	0	0	0	6	0	0
YJR059W - PTK2	0	2	0	0	0	0
YJR069C - HAM1	0	0	0	3	0	0
YJR077C - MIR1	0	0	0	5	0	0
YJR082C - EAF6	0	4	0	0	0	2
YJR091C - JSN1	0	3	0	0	0	0
YJR104C - SOD1	0	0	0	19	0	0
YJR105W - ADO1	0	0	0	8	0	0
YJR109C - CPA2	0	9	0	13	0	0
YJR121W - ATP2	0	3	0	32	3	0
YJR123W - RPS5	0	0	0	15	0	0
YJR139C - HOM6	0	0	0	12	0	0
YJR145C - RPS4A	0	3	0	18	0	0
YKL006W - RPL14A	0	0	0	15	0	0
YKL016C - ATP7	0	0	0	6	0	0
YKL032C - IXR1	0	4	0	0	0	0
YKL035W - UGP1	5	3	0	3	0	0
YKL060C - FBA1	2	4	0	66	0	0
YKL067W - YNK1	0	0	0	4	0	0
YKL080W - VMA5	0	0	0	6	0	0
YKL081W - TEF4	0	2	0	8	0	0
YKL085W - MDH1	0	0	0	22	0	0
YKL104C - GFA1	0	3	0	2	0	0
YKL112W - ABF1	4	10	2	0	0	0
YKL129C - MYO3	5	10	2	6	0	6
YKL142W - MRP8	0	0	0	5	0	0
YKL148C - SDH1	33	17	0	15	0	0
YKL150W - MCR1	0	0	0	4	0	0
YKL152C - GPM1	0	2	0	41	0	0
YKL164C - PIR1	0	0	0	2	0	0
YKL168C - KKQ8	2	0	0	0	0	0
YKL175W - ZRT3	0	2	0	0	0	0
YKL180W - RPL17A	0	0	0	7	0	0

YKL182W - FAS1	0	10	0	17	0	0
YKL191W - DPH2	0	2	0	0	0	0
YKL210W - UBA1	0	0	0	2	0	0
YKL211C - TRP3	0	2	0	0	0	2
YKL212W - SAC1	0	18	0	3	0	0
YKL213C - DOA1	2	2	0	0	0	0
YKR016W - AIM28	0	0	0	3	0	0
YKR048C - NAP1	0	0	0	4	0	0
YKR057W - RPS21A	0	0	0	11	0	0
YKR059W - TIF1	0	2	0	20	0	0
YKR098C - UBP11	7	9	0	6	0	7
YLL013C - PUF3	0	4	0	0	0	3
YLL018C - DPS1	0	8	0	10	0	0
YLL021W - SPA2	4	3	5	0	0	0
YLL024C - SSA2	9	0	4	84	0	3
YLL026W - HSP104	0	0	0	4	0	0
YLL041C - SDH2	0	2	0	0	0	0
YLL045C - RPL8B	0	0	0	12	0	0
YLL050C - COF1	0	0	0	11	0	0
YLR017W - MEU1	0	3	2	0	0	0
YLR027C - AAT2	0	0	0	4	0	0
YLR028C - ADE16	0	0	0	3	0	0
YLR029C - RPL15A	0	0	0	5	0	0
YLR038C - COX12	0	0	0	2	0	0
YLR043C - TRX1	0	0	0	2	0	0
YLR044C - PDC1	5	23	20	74	0	16
YLR048W - RPS0B	0	2	0	10	0	0
YLR058C - SHM2	8	18	24	17	0	28
YLR061W - RPL22A	0	0	0	7	0	0
YLR075W - RPL10	0	2	0	13	0	0
YLR096W - KIN2	0	7	0	0	0	0
YLR109W - AHP1	0	0	0	12	0	0
YLR134W - PDC5	0	0	0	13	0	0
YLR150W - STM1	0	0	0	7	0	0
YLR153C - ACS2	12	15	8	14	0	13
YLR157C-B - YLR157C-B	0	0	0	0	9	0
YLR167W - RPS31	0	0	0	17	0	0
YLR175W - CBF5	0	0	0	3	0	0
YLR180W - SAM1	0	0	0	6	0	0
YLR192C - HCR1	0	0	0	6	0	0
YLR195C - NMT1	0	4	3	0	0	0
YLR197W - NOP56	0	0	0	5	0	0
YLR200W - YKE2	0	0	0	3	0	0
YLR216C - CPR6	0	0	0	4	0	0

YLR249W - YEF3	8	20	2	20	2	0
YLR250W - SSP120	0	3	0	3	0	0
YLR258W - GSY2	29	34	6	21	0	9
YLR259C - HSP60	0	0	0	26	0	0
YLR262C-A - TMA7	0	0	0	2	0	0
YLR264W - RPS28B	0	0	0	6	0	0
YLR293C - GSP1	0	2	0	2	0	0
YLR295C - ATP14	0	0	0	6	0	0
YLR301W - YLR301W	0	0	0	2	0	0
YLR303W - MET17	2	0	0	3	0	0
YLR304C - ACO1	0	0	0	12	0	0
YLR325C - RPL38	0	0	0	5	0	0
YLR328W - NMA1	22	17	3	14	2	33
YLR340W - RPP0	0	0	0	14	0	0
YLR344W - RPL26A	0	0	0	3	0	0
YLR352W - YLR352W	0	2	0	0	0	0
YLR354C - TAL1	0	0	0	5	0	0
YLR355C - ILV5	0	0	0	22	0	0
YLR367W - RPS22B	0	0	0	7	0	0
YLR371W - ROM2	4	10	5	2	0	3
YLR388W - RPS29A	0	0	0	3	0	0
YLR406C - RPL31B	0	0	0	6	0	0
YLR432W - IMD3	0	2	0	4	0	0
YLR438W - CAR2	26	18	8	5	0	10
YLR441C - RPS1A	0	0	0	10	0	0
YML008C - ERG6	0	0	0	2	0	0
YML017W - PSP2	23	132	9	22	7	35
YML024W - RPS17A	0	0	0	7	0	0
YML026C - RPS18B	0	3	4	6	0	0
YML028W - TSA1	0	5	0	37	0	0
YML035C - AMD1	20	25	5	9	0	0
YML050W - AIM32	3	27	0	2	0	3
YML051W - GAL80	12	13	30	0	0	0
YML059C - NTE1	10	6	5	3	0	5
YML063W - RPS1B	0	0	0	11	0	0
YML070W - DAK1	0	0	0	2	0	0
YML072C - TCB3	0	6	0	2	0	2
YML073C - RPL6A	0	0	0	7	0	0
YML074C - FPR3	0	0	0	4	0	0
YML078W - CPR3	0	0	0	3	0	0
YML081W - YML081W	0	0	0	0	0	3
YML092C - PRE8	0	0	0	2	0	0
YML100W - TSL1	0	4	0	3	0	0
YML126C - ERG13	0	0	0	2	0	0

YMR012W - CLU1	20	27	11	13	3	8
YMR029C - FAR8	0	7	0	0	0	0
YMR062C - ECM40	0	0	0	3	0	0
YMR083W - ADH3	3	3	0	5	0	0
YMR087W - YMR087W	0	6	2	0	0	0
YMR099C - YMR099C	0	2	0	2	0	0
YMR102C - YMR102C	0	0	0	0	0	2
YMR104C - YPK2	10	12	0	9	0	11
YMR108W - ILV2	3	2	0	12	0	0
YMR109W - MYO5	2	0	0	0	0	0
YMR113W - FOL3	3	82	2	10	4	31
YMR116C - ASC1	0	2	0	19	0	0
YMR120C - ADE17	0	0	0	6	0	0
YMR142C - RPL13B	0	0	0	6	0	0
YMR143W - RPS16A	0	2	0	7	0	0
YMR145C - NDE1	0	0	0	3	0	0
YMR146C - TIF34	0	0	0	2	0	0
YMR183C - SSO2	0	0	0	3	0	0
YMR186W - HSC82	6	11	2	29	0	2
YMR194W - RPL36A	0	0	0	3	0	0
YMR203W - TOM40	0	0	0	4	0	0
YMR205C - PFK2	0	5	0	8	0	0
YMR217W - GUA1	5	15	5	12	0	4
YMR226C - YMR226C	0	0	0	3	0	0
YMR230W - RPS10B	0	0	0	11	0	0
YMR237W - BCH1	3	14	3	4	0	2
YMR242C - RPL20A	0	6	0	5	0	0
YMR243C - ZRC1	3	3	0	0	0	0
YMR250W - GAD1	3	9	0	2	0	0
YMR273C - ZDS1	0	0	0	0	0	2
YMR276W - DSK2	0	0	0	4	0	0
YMR304W - UBP15	0	2	0	0	0	0
YMR307W - GAS1	0	0	0	7	0	0
YMR309C - NIP1	0	2	0	2	0	0
YMR315W - YMR315W	0	0	0	3	0	0
YMR318C - ADH6	0	0	0	2	0	0
YNL007C - SIS1	0	2	0	8	0	0
YNL009W - IDP3	0	2	0	0	0	0
YNL015W - PBI2	0	0	0	5	0	0
YNL030W - HHF2	0	0	0	3	0	0
YNL036W - NCE103	0	8	0	2	0	0
YNL037C - IDH1	0	0	0	2	0	0
YNL055C - POR1	0	0	0	12	0	0
YNL064C - YDJ1	4	7	8	3	0	0

YNL067W - RPL9B	0	5	2	5	0	0
YNL071W - LAT1	0	0	0	5	0	0
YNL079C - TPM1	0	0	0	10	0	0
YNL091W - NST1	0	11	0	2	0	0
YNL096C - RPS7B	0	0	0	5	0	0
YNL104C - LEU4	0	0	0	4	0	0
YNL121C - TOM70	0	0	0	3	0	0
YNL131W - TOM22	0	0	0	4	0	0
YNL134C - YNL134C	0	0	0	10	0	0
YNL135C - FPR1	0	0	0	4	0	0
YNL138W - SRV2	0	0	0	2	0	0
YNL154C - YCK2	73	190	25	56	21	120
YNL160W - YGP1	0	0	0	2	0	0
YNL178W - RPS3	0	0	0	13	0	0
YNL183C - NPR1	3	0	0	0	0	0
YNL208W - YNL208W	0	0	0	3	0	0
YNL209W - SSB2	3	7	0	42	0	0
YNL220W - ADE12	0	0	0	2	0	0
YNL244C - SUI1	0	0	0	4	0	0
YNL301C - RPL18B	0	2	0	2	0	0
YNL302C - RPS19B	0	2	0	15	0	0
YNL310C - ZIM17	4	9	3	0	0	0
YNL321W - VNX1	0	0	0	0	0	4
YNR001C - CIT1	0	0	0	8	0	0
YNR016C - ACC1	30	66	10	17	0	7
YNR047W - FPK1	0	3	0	0	0	3
YOL039W - RPP2A	0	0	0	9	0	0
YOL040C - RPS15	0	0	0	5	0	0
YOL044W - PEX15	0	3	0	0	0	0
YOL051W - GAL11	0	0	0	20	0	0
YOL058W - ARG1	0	0	0	5	0	0
YOL059W - GPD2	24	41	15	10	8	36
YOL081W - IRA2	0	4	2	3	0	2
YOL086C - ADH1	7	27	7	40	3	4
YOL087C - YOL087C	36	39	37	25	4	28
YOL109W - ZEO1	0	0	0	2	0	0
YOL120C - RPL18A	0	2	0	2	0	0
YOL121C - RPS19A	0	2	0	15	0	0
YOL123W - HRP1	0	0	0	3	0	0
YOL127W - RPL25	0	0	0	4	0	0
YOL139C - CDC33	0	10	0	6	0	3
YOL145C - CTR9	0	4	0	0	0	0
YOR007C - SGT2	0	0	0	9	0	0
YOR017W - PET127	2	6	0	0	0	0

YOR020C - HSP10	0	0	0	3	0	0
YOR021C - YOR021C	0	0	0	2	0	0
YOR027W - ST11	3	2	0	15	0	0
YOR063W - RPL3	3	4	4	18	0	0
YOR070C - GYP1	0	2	0	0	0	0
YOR096W - RPS7A	0	0	0	10	0	0
YOR117W - RPT5	0	0	0	4	0	0
YOR122C - PFY1	0	0	0	3	0	0
YOR133W - EFT1	0	9	4	55	0	0
YOR136W - IDH2	0	5	0	10	0	0
YOR142W - LSC1	0	0	0	5	0	0
YOR167C - RPS28A	0	0	0	6	0	0
YOR168W - GLN4	0	0	0	2	0	0
YOR184W - SER1	0	0	0	3	0	0
YOR185C - GSP2	0	2	0	2	0	0
YOR187W - TUF1	0	10	0	4	0	0
YOR195W - SLK19	3	0	0	0	0	0
YOR198C - BFR1	0	0	0	6	0	0
YOR202W - HIS3	3	10	2	7	0	4
YOR204W - DED1	0	0	0	4	0	0
YOR230W - WTM1	0	0	0	4	0	0
YOR267C - HRK1	11	19	11	13	5	26
YOR276W - CAF20	19	10	0	14	0	20
YOR285W - YOR285W	0	0	0	8	0	0
YOR286W - AIM42	0	0	0	2	0	0
YOR293W - RPS10A	0	0	0	11	0	0
YOR298C-A - MBF1	0	0	0	3	0	0
YOR310C - NOP58	0	0	0	2	0	0
YOR312C - RPL20B	0	6	0	5	0	0
YOR316C - COT1	0	2	0	0	0	0
YOR317W - FAA1	0	4	0	3	0	0
YOR332W - VMA4	0	0	0	9	0	0
YOR335C - ALA1	0	0	0	4	0	0
YOR369C - RPS12	0	0	0	8	0	0
YOR374W - ALD4	0	8	0	20	0	0
YOR375C - GDH1	0	0	0	9	0	0
YPL004C - LSP1	0	0	0	11	0	0
YPL016W - SW11	7	6	8	0	0	0
YPL037C - EGD1	0	0	0	13	0	0
YPL059W - GRX5	0	0	0	4	0	0
YPL061W - ALD6	0	0	0	12	0	0
YPL063W - TIM50	0	2	0	0	0	0
YPL078C - ATP4	0	0	0	4	0	0
YPL079W - RPL21B	0	0	0	2	0	0

YPL081W - RPS9A	0	0	0	4	0	0
YPL090C - RPS6A	0	0	0	6	0	0
YPL091W - GLR1	0	0	0	3	0	0
YPL106C - SSE1	0	0	0	29	0	0
YPL131W - RPL5	0	0	0	28	0	0
YPL154C - PEP4	0	0	0	8	0	0
YPL160W - CDC60	0	0	0	4	0	0
YPL198W - RPL7B	2	5	0	9	0	0
YPL204W - HRR25	0	2	0	0	0	0
YPL207W - TYW1	4	3	3	0	0	0
YPL218W - SAR1	0	0	0	5	0	0
YPL220W - RPL1A	0	0	0	7	0	0
YPL224C - MMT2	14	19	0	8	0	4
YPL231W - FAS2	0	4	0	17	0	0
YPL237W - SUI3	0	0	0	2	0	0
YPL239W - YAR1	0	0	0	3	0	0
YPL240C - HSP82	0	0	2	27	0	0
YPL248C - GAL4	0	0	0	54	0	0
YPL249C-A - RPL36B	0	0	0	3	0	0
YPL262W - FUM1	0	0	0	3	0	0
YPR016C - TIF6	0	0	0	2	0	0
YPR033C - HTS1	0	0	0	3	0	0
YPR035W - GLN1	0	6	0	0	0	0
YPR036W - VMA13	0	0	0	2	0	0
YPR041W - TIF5	0	0	0	2	0	0
YPR069C - SPE3	0	0	0	2	0	0
YPR072W - NOT5	0	0	0	2	0	0
YPR074C - TKL1	0	0	0	7	0	0
YPR080W - TEF1	91	29	20	107	4	11
YPR102C - RPL11A	0	0	0	2	0	0
YPR129W - SCD6	0	0	0	2	0	0
YPR132W - RPS23B	0	0	0	5	0	0
YPR135W - CTF4	0	6	0	0	0	0
YPR145W - ASN1	110	48	90	27	5	63
YPR163C - TIF3	0	0	0	3	0	0
YPR181C - SEC23	2	3	0	0	0	0
YPR191W - QCR2	0	0	0	7	0	0