

GLOBAL ANALYSIS OF SUMO E3 LIGASE SPECIFICITY UNCOVERS
CROSSTALK-MEDIATED KINASE ACTIVATION

by
Ijeoma K. Uzoma

A dissertation submitted to The Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March 2014

© Ijeoma K. Uzoma 2014
All rights reserved

ABSTRACT

Protein microarrays are highly flexible, ranging from large-scale proteome microarrays to smaller customizable microarrays, making the technology amenable for detection of a broad spectrum of biochemical properties of proteins. Functional protein microarrays have greatly contributed to advances in proteomics and are becoming an important platform for network reconstruction and posttranslational modification (PTM) substrate discovery. Currently, conjugation of Small Ubiquitin-related Modifier (SUMO) to many known substrates is sufficiently catalyzed by the E1 and E2 enzymes *in vitro*, however, emerging *in vivo* evidence suggests that E3 ligases may play an important role in numerous SUMO-dependent signaling pathways. While several studies have attempted to identify the full set of SUMO-modified proteins in the human proteome, a global approach examining the role of SUMO E3 ligases and their substrate specificity has not been reported. This thesis describes efforts to identify a comprehensive set of SUMO modified proteins, in order to illuminate the E3 ligase-substrate network, and identify the common and distinct substrates of several E3 ligases. To this end, we have developed a protein microarray-based approach to globally construct SUMOylation networks, utilizing the Human Proteome Microarray containing approximately 17,000 individually purified recombinant human proteins. In our study over 2,700 SUMO E3 substrates were identified. Many were unique SUMOylation targets for particular E3 ligases, suggesting E3 ligases play a large role in determining substrate selection *in vitro*. Gene ontology analysis recovered known biological functions for many E3 ligases and more importantly, revealed many novel functions that could

potentially link SUMOylation to signaling pathways involving protein kinases. Our studies in chapter 3 reveal enrichment of SUMOylation within the MAP kinase family.

Over 100 kinases were recovered as SUMO substrates in our assays which points to a global phenomenon of kinase regulation by SUMO modification. The studies presented in chapter 4 describe our efforts to validate and characterize SUMO modification of focal adhesion, tyrosine kinase, Pyk2. Here, we demonstrate that SUMOylation of Pyk2 is a novel PTM that serves to amplify intrinsic kinase activity by enhancing autophosphorylation. Further biochemical studies reveal SUMOylation of Pyk2 enhances its association with putative binding protein Src kinase, promotes phosphorylation of downstream focal adhesion protein paxillin, and mediates ERK activation. These findings led us to examine the role of SUMOylation of Pyk2 in cell migration. Our results suggest SUMOylation of Pyk2 amplifies its promigratory function in MDA-MB-231 breast cancer cells. These studies have revealed a novel mechanism for Pyk2 activation, regulated by crosstalk between phosphorylation and SUMOylation. Collectively, we have illuminated the connections between SUMOylated kinases along a particular signaling axis, promoting enzyme activity, protein interactions, and activation of numerous nodes in a pathway. We believe that the work presented in this thesis will have a profound impact on the studies of SUMOylation and its crosstalk with other PTMs in a broad range of biological processes.

Thesis Advisor:

Dr. Heng Zhu

Thesis Reader:

Dr. Jiang Qian

ACKNOWLEDGEMENTS

I would like to take this opportunity to acknowledge and thank all the people who have helped me accomplish my goals as a scientist, and who have contributed to my life successes thus far. First and foremost, I would like to thank my thesis advisor, Dr. Heng Zhu, for his careful guidance, patience, and unwavering support throughout this journey. Heng has always given me the freedom to pursue my own scientific avenues, resulting in an exciting project that we were able to create together. I appreciate his trust in me which has inspired the confidence I have come to feel towards the end of my PhD candidacy. I have learned many important life lessons from him. His own path exemplifies persistence and discipline which I admire. Heng has always encouraged me to aim high whether it is applying for a fellowship, designing a challenging experiment, or submitting to a top tier journal. He believes I am capable of achieving my goals which has been an unspoken, yet important source of support. I am so thankful that I was able to join his lab, which has truly felt like home for the past six years.

I want to thank my thesis committee members Dr. Pamela Meluh, Dr. Michael Matunis, and Dr. Cynthia Wolberger for taking the time to assist the development and evolution of my project. I appreciate all their insightful contributions, valuable ideas, and suggestions over the years. I would also like to thank Dr. Seth Blackshaw and Dr. Jiang Qian for advising me as if I was their own student at times. Dr. Kate Guzzo has also been an invaluable mentor and friend, who has always taken the time to look over my work, provide great experimental suggestions, and discuss results. I must also thank my collaborators Dr. Jianfei Hu, Dr. Zhi Xie, Dr. Jianying Zhou, who have all provided major contributions to this project.

I would also like to thank all of the members of the High Throughput (HiT) Biology Center for creating a uniquely collaborative and inspiring atmosphere to train in as a graduate student. I feel very lucky to have all of the resources and support that the HiT center provides to its members. I was also involved in the multidisciplinary Technology Center for Networks and Pathways (TCNP) initiative funded by the NIH. Through TCNP, I was given the opportunity to participate in highly collaborative projects, present at numerous meetings at the NIH, and attend many high quality TCNP seminars at Johns Hopkins.

I must also thank the members of the Zhu lab, both past and present. Though some of the members have changed over the years, the familial aspect remains present. I can honestly say that there has never been a day where I have not wanted to wake up and come into the lab, even when none of my experiments showed any hope. The positive and friendly atmosphere in our lab, and in the HiT Center, has made my graduate career both enjoyable and successful. In particular, I would like to thank Eric Cox. I am certain that this experience would not be the same without Eric. We started graduate school the same year and spent countless hours together studying biophysics, bioorganic, and pathways. Ultimately, we both joined the Zhu lab and were offered the opportunity to work on the same project. I cannot think of a better person to have worked alongside every day for nearly seven years. We have been through some significant scientific disappointments and successes, but we have always supported each other's growth as scientists. He has helped me troubleshoot through hundreds of chip assays, failed transfections, western blots and IPs, and managed to make me laugh at my own mistakes.

I am so thankful for his friendship and support in the lab. Without him this would not have been possible.

I extend my deepest thanks to my family for their unending love and support. I thank my parents for instilling in me a passion for education at an early age. Having them in my corner has made my time in graduate school not only possible but more pleasant and comfortable. I thank my father, Val Uzoma, for continually reminding me that education is truly a privilege with extraordinary transformative power. My mother, Laurie Uzoma, has always encouraged me to be well-rounded academically and has been a major influence on my pursuit of higher education. My step-mother, Alison Otis-Uzoma, has been an amazing listening ear and is always there to offer her superior editing skills. I am so thankful that I was able to share a year in Baltimore with my sister, Ngozika Uzoma. I will always cherish the time we had living together at 222 Saratoga while we both struggled to find our footing in our careers. I would also like to thank Dwayne Ferguson for his love, faith, and encouragement. He has been a calming force, especially during the seemingly volatile periods of time towards the end of graduate school.

I would like to thank the Department of Pharmacology for offering me the chance to earn a PhD from Johns Hopkins. This has been a lifelong dream and I am grateful for the opportunity. Finally, I would like to extend a special thanks to Mimi Guercio, Paula Mattingly, Amy Loveland, and Amy Forcier for all of their behind the scenes work that has made this possible. Thank you!

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	vii
List of Figures	x
List of Tables	xiii
Chapter 1: Introduction to Protein Microarray Technology.....	1
Abstract	2
Introduction	3
1.1 Network Construction	7
1.2 Pathogen-host Interactions	12
1.3 Biomarker Identification	16
Perspectives	20
References	24
Chapter 2: SUMOylation: Features, Conjugation, and Function.....	30
2.1. Features	31
2.2. SUMO Conjugation	33
2.3. SUMO E3 Ligases	35

2.4. Deconjugation	37
2.5. SUMO Substrate Specificity	37
2.6. Functional Importance of SUMOylation	39
2.7. PTM Crosstalk	41
References	43
Chapter 3: Constructing Global SUMO E3 ligase Substrate Networks Utilizing Protein	
Microarray Technology.....	48
Abstract	49
Introduction	51
3.1. Experimental Strategy	55
3.2. Analysis of Raw Data	64
3.3. Results of Global SUMOylation Data Analysis	65
3.4. Validation of MAP Kinases as Targets of SUMOylation	76
Discussion	79
Methods	84
References	90

Chapter 4: Characterization of Pyk2 SUMOylation in Cell Migration Pathways.....	94
Abstract	95
Introduction	96
4.1. Results	103
4.1.1. Validation of Pyk2 SUMOylation in HeLa Cells.....	103
4.1.2. Mapping SUMOylation Sites in Pyk2.....	106
4.1.3. SUMOylation of Pyk2 Enhances Autophosphorylation at Y402..	111
4.1.4. SUMOylation Stimulates Src Association with Pyk2.....	114
4.1.5. Role of SUMO modification in Pyk2-mediated Paxillin Phosphorylation.....	117
4.1.6. Pyk2 SUMOylation Mutant, 4KR, Shows reduced Signaling through ERK1/2 Axis.....	120
4.1.7. SUMOylation of Pyk2 affects paxillin-mediated cell migration.	120
Discussion	125
Methods	130
References	136
Curriculum Vitae	142

List of Figures

Chapter 1: Introduction to Protein Microarray Technology

Figure 1-1. Reconstituted interaction networks in cellular systems generated through protein microarray studies.....22

Figure 1-2. The human proteome microarray.....23

Chapter 2: SUMOylation: Features, Conjugation, and Function

Figure 2-1. The SUMO conjugation cycle.....34

Figure 2-2. Phosphorylation dependent SUMOylation.....39

Chapter 3: Constructing Global SUMO E3 Ligase Substrate Networks Utilizing Protein Microarray Technology

Figure 3-1. Siz1 E3 ligase-dependent SUMOylation of PCNA Lysine 164.....52

Figure 3-2. PIAS α specific SUMOylation of STAT1 under limiting concentrations of E1 and E2.....56

Figure 3-3. SUMO pilot chip.....57

Figure 3-4. E1 and E2 Enzyme titration on pilot chip.....58

Figure 3-5. PIAS3 enhances SUMOylation of NR2E3 *in vitro* and requires intact RING domain.....60

Figure 3-6. E3 ligase SUMOylation activity and specificity on pilot array.....61

Figure 3-7. Study design and assay schematic.....64

Figure 3-8. E3 ligase SUMOylation substrate specificity network.....69

Figure 3-9. Clustering of E3 ligase substrates based on similarity of E3 ligase profiles.....	70
Figure 3-10. Gene Ontology analysis of significant biological processes and molecular functions.....	71
Figure 3-11. Phylogenetic kinase tree overlaid with SUMOylation enrichment..	74
Figure 3-12. Kinase SUMOylation substrates in PPI and KSR network.....	75
Figure 3-13. E3 ligase substrate specificity in the MAP kinase family.....	76
Figure 3-14. PIAS directed SUMOylation of the MAP kinase family.....	78
Figure 3-15. E3 ligase specificity for MAPK11/P38 β	79

Chapter 4: Characterization of Pyk2 SUMOylation in cell migration pathways

Figure 4-1. Bioactive peptide induced signaling pathway is enriched for SUMOylation substrates.....	96
Figure 4-2. Architecture of focal contacts.....	99
Figure 4-3. Domains, motifs, and phosphorylation sites in FAK family tyrosine kinases FAK and Pyk2.....	100
Figure 4-4. Pyk2 signaling within focal contacts.....	102
Figure 4-5. Pyk2 SUMOylation on Human Proteome chip.....	104
Figure 4-6. Validation of Pyk2 SUMOylation and E3 ligase specificity.....	105

Figure 4-7. PIAS1 directed modified 3KR mutant shows impaired SUMOylation compared to wild type.....	107
Figure 4-8. Pyk2-SUMO1 immunoprecipitate visualized by silver staining.....	109
Figure 4-9. K581 SUMO1 modification spectra.....	109
Figure 4-10. Pyk2 4KR mutant exhibits significantly reduced SUMOylation compared to wild type.....	110
Figure 4-11. Autophosphorylation at Tyr402 is impaired in 4KR mutant.....	112
Figure 4-12. SUMOylation stimulates autophosphorylation of Pyk2 <i>in vitro</i>	113
Figure 4-13. SUMOylation of WT Pyk2 but not 4KR promotes association with endogenous Src.....	116
Figure 4-14. SUMOylation of Pyk2 promotes phosphorylation of paxillin.....	118
Figure 4-15. Pyk2 4KR mutant inhibits activation of ERK1/2 but not P38/MAPK.....	120
Figure 4-16. SUMOylation of Pyk2 stimulates cell migration in MDA-MB-231 breast cancer cells.....	122
Figure 4-17. Crosstalk-mediated activation of Pyk2 model.....	124

List of Tables

Table 1. Applications of protein microarrays in diverse biological network construction.....	7
Table 3-1. Summary of Global E3 ligase SUMOylation utilizing the Human Proteome microarray.....	68

CHAPTER 1

INTRODUCTION PROTEIN MICROARRAY TECHNOLOGY

This review was originally published in the Journal Genomics Proteomics Bioinformatics. Uzoma I, Zhu H. Interactome mapping: using protein microarray technology to reconstruct diverse protein networks. Genomics Proteomics Bioinformatics. 2013;11:18-28. Copyright © 2013 Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

Abstract

A major focus of systems biology is to characterize interactions between cellular components, in order to develop an accurate picture of the intricate networks within biological systems. Over the past decade, protein microarrays have greatly contributed to advances in proteomics and are becoming an important platform for systems biology. Protein microarrays are highly flexible, ranging from large-scale proteome microarrays to smaller customizable microarrays, making the technology amenable for detection of a broad spectrum of biochemical properties of proteins. In this chapter, we will focus on the numerous studies that have utilized protein microarrays to reconstruct biological networks including posttranslational protein modifications (PTMs), lectin-glycan recognition, pathogen-host interactions and hierarchical signaling cascades. The diversity in applications allows for integration of interaction data from numerous molecular classes and cellular states, providing insight into the structure of complex biological systems. We will also discuss emerging applications and future directions of protein microarray technology in the global frontier.

Introduction

Since the completion of major whole genome sequencing efforts, the scientific community has been faced with the challenge of identifying and characterizing the expressed gene products of given organisms [1]. The post-genomics era gave birth to the field of proteomics that aimed to systematically chart the biochemical properties and functions of all expressed proteins [2]. With a global view in mind, we now strive to integrate complex “omics”-data from all molecular ranks. The scope of proteomics is not limited to identifying protein-protein interactions, but also includes identification of protein posttranslational modifications (PTMs) and of interactions with DNA and RNA sequences, lipids and glycans. Weaving these layers together will allow us to construct the carefully tuned network that exists within live cells. Improvements in high throughput proteomic technologies coupled with advances in genomics and bioinformatics have laid a framework to enable this level of research.

Two of the most powerful platforms for proteomic studies are mass spectrometry and protein microarray technologies. Although mass spectrometry is well suited for high throughput protein identification, quantification and PTM site mapping [3], it still has its disadvantages such as bias against low abundance proteins and modifications, as well as undersampling of complex proteomes [4]. On the contrary, the protein microarray platform avoids these limitations and is particularly suited for unbiased global profiling [5].

A protein microarray, also termed a protein chip, is created by immobilization of thousands of different proteins (*e.g.*, antigens, antibodies, enzymes and substrates, *etc.*) in discrete spatial locations at high density on a solid surface [6]. Depending on their

applications, protein microarrays can be categorized into two varieties: analytical and functional protein microarrays. Analytical protein microarrays are usually composed of well-characterized biomolecules with specific binding activities, such as antibodies, to analyze the components of complex biological samples (*e.g.*, serum and cell lysates) or to determine whether a sample contains a specific protein of interest [7]. They have been used for protein activity profiling, biomarker identification, cell surface marker/glycosylation profiling, clinical diagnosis and environmental/food safety analysis [8-10]. Alternatively, functional protein microarrays are constructed by printing a large number of individually purified proteins and are mainly used to comprehensively query biochemical properties and activities of those immobilized proteins. In principle, it is feasible to print arrays composed of virtually all annotated proteins of a given organism, effectively comprising a whole-proteome microarray [11].

In 2001 the Snyder group reported the fabrication of the first proteome microarray in the budding yeast, representing a major advance for the field [12]. In order to construct this array, approximately 5,800 full-length yeast ORFs were individually expressed in yeast and their protein products purified as N-terminal GST-fusion proteins. Each purified protein was then robotically spotted on a single glass slide in duplicate at high-density to form the first “proteome” microarray, covering more than 75% of the yeast proteome. High throughput protein production:

To overcome these hurdles, the Snyder group created a high-throughput protein purification protocol in the budding yeast [13]. Using a homologous recombination-based strategy, more than 5800 full-length yeast open reading frames (ORFs) were cloned into a yeast expression vector that, upon galactose induction, produces glutathione-*S*-transferase

(GST)-tagged N-terminal fusion proteins. The purification protocol took advantage of both a 96-well format and immobilized affinity chromatography. This strategy allowed parallel purification of unprecedented numbers of proteins—up to 1152 per day. The success of this approach is built upon several unique aspects: First, it utilizes a eukaryotic expression system that both generates high levels of recombinant proteins and tends to produce a high fraction of soluble proteins. Compared with bacterial expression systems, in which a large fraction of recombinant proteins end up in inclusion bodies, this is a huge advantage when a large number of eukaryotic proteins are being generated. Second, the expression of recombinant proteins is only induced over about two total cell cycles, which greatly reduces toxicity and cell death. Third, a foreign eukaryotic protein purified from yeast is more likely to be active because post-translational modifications (PTMs) necessary for function are more likely to occur correctly than in either bacteria or a cell-free system. Fourth, the use of an N-terminal GST tag helps protein fold correctly and therefore, improve its stability and solubility. Other commonly used tags include the so-called tandem affinity purification (TAP-tag) and 6xHis, to name a few.

More recently, proteome microarrays have been fabricated from the proteomes of viruses, bacteria, plants, and humans. Functional protein microarrays have been successfully applied to identify protein–protein, protein–lipid, protein–antibody, protein–small molecule, protein–DNA, protein–RNA, protein–lectin, and lectin–cell interactions [8, 9, 12, 14-18], and to identify substrates or enzymes for phosphorylation, ubiquitylation, acetylation, and nitrosylation, as well as to profile immune response [11, 19-23]. There are a myriad of inventive applications for protein microarrays that have resulted in

significant findings that contribute to understanding the complex interactions within cells (Table 1).

Table 1. Applications of protein microarrays in diverse biological network construction

Assay type	Array content	Type of probe	Application	Ref
	Network construction			
Protein-DNA interaction	4191 human proteins	DNA motif	Protein-DNA interaction network	[16]
Kinase assay	2158 <i>Arabidopsis</i> proteins	Protein kinase	Signaling network	[23]
Ubiquitylation assay	Yeast proteome	Ubiquitylation enzymes	PTM network	[20]
	~9000 human proteins	Concentrated cell extract	PTM network	[19]
Acetylation assay	Human Protoarray, Invitrogen			
	Yeast proteome	Acetyltransferase	PTM network	[21]
	Pathogen-host interactions			
Viral kinase assay	4191 human proteins	Conserved viral kinases	Viral PTM target network	[31]
Protein-protein interaction	60 EBV viral proteins	Human protein	Protein-protein interaction network	[13]
	4191 human proteins	Viral protein	Protein-protein interaction network	[44]
Protein-RNA interaction	Yeast proteome	BMV SLD RNA loop	Protein-RNA interaction network	[17]
	Biomarker identification			
Antigen-antibody interaction	5011 human proteins	AIH patient sera	Biomarker identification	[47]
	<i>82 corona virus proteins</i>	SARS patient sera	Antibody profiling	[48]
	<i>E. coli</i> K12 proteome	IBD patient sera	Biomarker identification	[8]
Lectin-glycan interaction	Yeast proteome	Lectins	Protein glycosylation profiling	[53]
	94 lectins	Live mammalian cells	Cell surface biomarker identification	[9]

1.1. Network construction

A solid understanding of the molecular mechanisms of biological functions requires systematic profiling of dynamic interactions between biomolecules. Processes such as transcriptional regulation, viral infection, numerous PTMs and protein-protein interactions account for a small fraction of the potential molecular interactions within a cell but highlight how fundamental these networks are for essential functions. High throughput technologies strive to provide an unbiased platform for charting these relationships at the proteome and genome scale. In this section we will review several studies that demonstrate the utility of protein microarrays in reconstructing interaction networks.

MAP kinase substrate phosphorylation network

The mitogen-activated protein kinase (MAPK) signaling cascade involves a hierarchy of kinases that activate one another through consecutive phosphorylation events in response to extracellular or intracellular signals [24]. Standard methods have only been able to establish a few combinatorial connections from upstream MKK-activating kinases (MKKKs) to downstream MPK-activating kinases (MKKs), MAPKs and their cytoplasmic and nuclear substrates [25, 26]. Constructing this complicated interconnected network necessitates a systematic unbiased high-throughput approach to avoid confounding issues of redundancy and functional pleiotropy [24]. Akin to the protein microarray based kinase assays developed by Ptacek et al. [19], Popescu et al. employed high-density protein microarrays to identify novel MPK substrates. The authors first determined which *Arabidopsis thaliana* MKKs preferentially activate 10 different MPKs *in vivo* and used the activated MPKs to probe *Arabidopsis* protein microarrays containing

2158 unique proteins to reveal their phosphorylation substrates [24]. The initial screen identified 570 nonredundant MPK phosphorylation substrates with an average of 128 targets per activated MPK. With this data the authors were able to reconstruct a complex signaling cascade involving nine MKKs, 10 MPKs and 570 substrates [24]. Moreover, the resulting nodes and edges highlighted the specificity conserved within these interactions: 290 (51%) of MPK phosphorylation targets were hit by only one MPK and only 94 (16%) were phosphorylated by two or more MPKs [24]. Gene ontology (GO) analysis of effector substrates showed enrichment in TFs involved in the regulation of development, defense and stress responses [24]. The network that emerged from this study suggests the MAPK signaling cascade regulates transcription through combinatorial enzyme specificity and discrete phosphorylation events.

Ubiquitin E3 ligase substrate discovery

Ubiquitylation is one of the most widespread PTMs and mediates a huge range of cellular events and processes in eukaryotes [27]. Understanding ubiquitin substrate specificity is a complex combinatorial question, as it is conferred by unique permutations of E1, E2 and E3 enzymes. Lu et al. developed an assay to determine substrates of a HECT domain E3 ligase, Rsp5, using yeast proteome microarrays [21]. Over 90 novel proteins were found to be readily ubiquitylated by Rsp5, eight of which were validated as *in vivo* targets. Deeper *in vivo* characterization of two substrates, Sla1 and Rnr2, revealed that Rsp5-dependent ubiquitylation affects either the posttranslational process of the substrate or subcellular localization [21]. This design offers the ability to dissect the molecular

mechanisms of a complex enzymatic cascade and gives the field a tool to understand how the system is organized globally.

Identification of non-histone substrates of protein acetyltransferases in yeast

Acetylation is a major epigenetic PTM widely known for its role in regulating chromatin state. However, it is suspected to regulate nonnuclear functions as well [28]. In yeast, no non-histone proteins were reported as substrates of histone acetyltransferases (HATs) and histone deacetylases (HDACs). The catalytic enzyme, Esa1, of the essential nucleosome acetyltransferase of the complex, NuA4, is the only essential HAT in yeast [29], strongly suggesting that it may mediate acetylation of non-histone proteins critical for cell survival. Another intriguing question was whether HATs could regulate activity of cytosolic proteins or even enzymes like protein kinases. To comprehensively discover the non-chromatin substrates of the NuA4 HAT complex in the yeast proteome, Lin et al. developed *in vitro* acetylation reactions on the yeast proteome microarrays, containing 5800 yeast proteins, using NuA4 and [¹⁴C]-Acetyl-CoA [22]. Over 90 non-histone proteins were readily acetylated by the NuA4 complex. Although it was expected that the majority of the substrates would be involved in nucleosome assembly and histone binding categories, a significant number of the identified substrates were cytoplasmic proteins and metabolic enzymes [22]. Twenty proteins involved in a variety of cellular functions such as metabolism, transcription, cell cycle progression, RNA processing and stress response were selected for further validation. Standard double-immunoprecipitation techniques were used to validate 13 of the 20 substrates, including phosphoenolpyruvate carboxykinase (Pck1p). To understand the physiological relevance of non-chromatin

acetylation, the authors focused on the cytosolic enzyme Pck1p to explore a connection between acetylation and metabolism. Tandem mass spectrometry (MS/MS) identified lysine 19 (K19) and K514 as the acetylation sites of Pck1p and site-directed mutagenesis revealed that acetylation of K514 is critical for its enzymatic activity and promotes extension of life span in yeast growing under starvation conditions. These findings demonstrate a functional role for non-chromatin acetylation in yeast metabolism and longevity.

Based on GO analysis, acetylation may regulate several other cellular processes as well. In a follow up study, Lu et al. investigated the impact of acetylation on another NuA4 substrate, Sip2, a regulatory subunit of the SNF1 kinase complex (yeast AMPK). Based on the MS/MS analysis and site-directed mutagenesis studies, the authors found that Sip2 acetylation enhances its interaction with the catalytic subunit Snf1 and inhibits Snf1's kinase activity [30]. As a result, phosphorylation of one of Snf1's downstream targets, Sch9 (homolog of Akt/S6K), is decreased, ultimately leading to slower growth but extended replicative life span. Finally, the authors demonstrated that the anti-aging effect of Sip2 acetylation is independent of extrinsic nutrient availability and TORC1 activity. These studies are now echoed by recent discoveries of many mitochondrial and cytosolic enzymes as substrates of acetyltransferases in higher eukaryotes via MS-based PTM profiling [31-33].

Global ubiquitylation substrate discovery from cell extracts

Readily generating a snapshot of global protein PTM profiles under various cellular conditions could be considered the Holy Grail for those researching PTMs. General PTM

substrate identification strategies require enrichment from a cell extract sample followed by MS or *in vitro* assays using purified components. While both approaches have their strengths and weaknesses, a hybrid of the two is possible. The use of concentrated mammalian cell extracts in combination with protein microarrays can serve to identify PTM targets in a semi-*in vivo* setting while alleviating the challenge of analyzing a complex mixture. Merbl and Kirschner generated cell extracts that replicate the mitotic checkpoint and anaphase release to identify differentially regulated polyubiquitylation substrates [20]. The synchronized cell extracts were incubated with Invitrogen's Human ProtoArray composed of 8000 proteins and the resulting polyubiquitylated proteins were detected with antibodies directed to ubiquitin chains [20]. The authors expected to recover substrates of the anaphase promoting complex (APC), the major ubiquitin ligase in mitosis and G1. To differentiate polyubiquitylation substrates of the APC from other ligases, Merbl and Kirschner designed three experimental set ups. All cell extracts were arrested with nocodazole as the control which inhibits the APC, in the second condition the sample was released from checkpoint arrest with the addition of UbcH10, an E2 ligase, and the final condition was supplemented with both UbcH10 and a specific inhibitor of APC. Approximately 132 proteins were differentially polyubiquitylated, 11 of which were known APC substrates, confirming the validity on the experimental design. Validation studies performed in rabbit reticulocyte lysate confirmed the degradation/ubiquitylation of 7 novel APC substrates [20]. This study demonstrates the efficacy of using protein microarrays in combination with cell extracts to recapitulate the global PTM signature in a specific cellular state.

1.2. Pathogen-host Interactions

Protein microarrays allow for exploration of hypotheses that cannot be addressed by standard methods. Investigating the interactions between viral encoded proteins and the proteins within the infected host has been an important yet cumbersome task. Protein microarrays composed of either the host or the viral proteome can be fabricated and subsequently used to examine the relationships between the viral machinery and the host. This *in vitro* approach recapitulates viral infection in that the viral genome/proteome are allowed to physically interact with the host. The Hayward and Zhu groups have recently developed this new paradigm to examine direct interactions between viral and host proteins [14, 34, 35], leading to a deeper understanding of the mechanisms by which the viral proteins hijack the host as well as uncovering the direct targets of major viral enzymes.

Herpesvirus kinase-phosphorylome

The human α , β , and γ herpesviruses cause diseases distinct from one another, ranging from mild cold sores to pneumonitis, birth defects and cancers [34]. Although the viruses are different, once they enter the host cells they all must reprogram cellular gene expression, sense cell-cycle phase, modify cell-cycle progression and reactivate the lytic life cycle to produce new virions to spread infection [36]. Many lytic cycle genes involved in replication of the viral genomes are highly conserved across the herpesvirus family. For example, each herpesvirus encodes for an orthologous serine/threonine kinase [37] that shares structural similarity with human cyclin-dependent kinases (CDKs) [38] and phosphorylates the substrates of CDKs [37]. The ability of viral kinase to mimic host

CDKs results in hijacking of key pathways to potentiate their own replication. Particular cellular phosphorylation events are observed during herpes infection and specific phosphorylation of antiviral drugs in infected cells are mediated by the conserved viral kinases [39]. Identifying the collective host targets of the viral kinases would reveal the commonly shared mechanisms and signaling pathways among different herpesviruses to promote their lytic replication. This knowledge will increase the therapeutic target options necessary for developing pan-antivirals.

To test this idea, Li et al. utilized the human transcription factor (TF) proteome array containing 4191 human proteins to identify commonly shared substrates of herpesvirus-encoded kinases [34]. Parallel kinases assays were performed using the four viral kinases, UL31, UL97, BGLF4 and ORF36, which is encoded by herpes simplex type 1 (HSV1), human cytomegalovirus (HCMV), Epstein-Barr virus (EBV) and Kaposi Sarcoma associated-virus (KSHV), respectively [37]. In total, 643 nonredundant substrates were identified across the four kinases and 110 substrates were targets of at least three kinases. GO analysis of the 110 shared substrates indicates that DNA damage functional class was significantly enriched. Among the DNA damage proteins, TIP60 was selected as a lead candidate for regulation of viral replication, due to its roles in DNA damage as well as transcriptional regulation through its HAT activity. Phosphorylation of TIP60 by BGLF4 in EBV-infected B cells was validated during further analysis. BGLF4 is known to phosphorylate multiple EBV proteins and only a small number of host proteins [37, 40]. The functions of its previously-characterized targets are varied, implying that the kinase plays multiple roles to promote viral replication [40]. It is expressed in the early phase of the lytic infection cycle and is localized mainly in the nuclei of EBV-infected cells [41].

BGLF4 knockdown revealed that it is critical for release of infectious virus during viral lytic reactivation [40]. Subsequent experiments demonstrated that BGLF4-mediated phosphorylation enhanced TIP60 HAT activity by ten-fold, linking the phosphorylation event to viral replication. They also demonstrated the importance of phosphorylation of host DNA damage proteins for viral replication. More specifically, phosphorylation and activation of TIP60 by BGLF4 triggers EBV-induced DNA damage response (DDR) and promotes positive transcriptional regulation of critical lytic genes involved in viral replication. Lastly, the study confirmed that TIP60 was also required for efficient lytic replication in HCMV, KSHV and HSV-1. Taken together, this unbiased approach provides a novel paradigm for discovery of conserved targets of viral enzymes. While herpes kinases have been credible therapeutic candidates, knowing their targets and the signaling pathways they exploit will better enable the development of widely effective antiviral drugs.

BGLF4-SUMO2

In a follow up study, Li et al. took the inverse approach that employed a herpesvirus EBV protein microarray to assess human-host protein binding events [14]. Small ubiquitin-related modifier (SUMO) is covalently attached to proteins via an enzymatic cascade analogous to the ubiquitin pathway. SUMO is involved in a broad range of cellular processes including signal transduction, regulation of transcription, DNA damage response and mediation of protein-protein interactions [42, 43]. Both latent and lytic EBV proteins interact with components of the SUMO machinery [14, 43]. While covalent modification by SUMO is more commonly understood, noncovalent interactions with

SUMO also contribute to SUMO effector signaling [42, 43]. Noncovalent binding to SUMO is often mediated through SUMO-interaction motif (SIM) domains on target proteins [42, 43]. To comprehensively identify the EBV proteins that bind to the SUMO moiety, the authors fabricated a protein microarray of full length proteins from EBV and KSHV individually purified from yeast. The array was used to perform a protein-protein binding assay using the SUMO2 paralog. They identified 11 EBV proteins as potential SUMO partners, including BGLF4, a conserved kinase [14]. As BGLF4 is known to play a multitude of roles in EBV, the authors pursued the importance of the cellular PTM in BGLF4 function. The BGLF4 SIM domains were mapped and when mutated at both the N- and C-terminal SIMs, the intracellular localization of the kinase shifted from nuclear to cytoplasmic. A mutation in the N-terminal SIM showed largely nuclear localization, whereas the C-terminal SIM mutation generated an intermediate phenotype with nuclear and cytoplasmic expression. The authors found that BGLF4 inhibits SUMOylation of lytic cycle transactivator ZTA and demonstrated that the SIM domains as well as kinase activity are required for inhibition [14]. SIM domains of BLGF4 were also shown to be necessary for suppressing global SUMOylation, inducing cellular DDR and promoting EBV lytic replication.

The virus takes advantage of the SUMOylation system by encoding proteins that are SUMO modified and those that bind to SUMO [14]. As previously mentioned SUMO is involved in DDR, which is further supported by the finding that BGLF4 appears to interact with sites of DNA damage via SUMO binding, revealing an additional mechanism promoting EBV-mediated DDR and lytic replication. SUMO interaction is as important as the kinase activity for the function of BGLF4.

1.3. Biomarker identification

Biomarker identification represents a major effort in modern biomedical and clinical research, as it allows for better screening methods, diagnosis criteria, prognosis predictions and ultimately superior treatment for a broad range of diseases. Traditionally, biomarker discovery has utilized popular methods such as MS, ELISA, gene expression and antibody arrays to profile serum samples [44]. In recent years, protein microarray technology has extended into clinical proteomics and is becoming a powerful tool for biomarker discovery. Proteins on functional protein microarrays were originally viewed as substrates and binding partners, but when applied to immunology, the proteins on the array could be potential antigens associated with certain diseases. By comparison, protein microarray based-serum profiling is much more sensitive and can be performed at higher throughput while requiring less amount of sample. Here we will review a variety of clinically-relevant applications for protein microarrays in biomarker identification.

SARS-CoV diagnosis

Protein microarrays can also be used as a diagnostic tool for infectious diseases. Severe acute respiratory syndrome (SARS) is an infectious disease, caused by a novel coronavirus (CoV), which appeared in Guangdong, China in November 2002. As of March 2003, the virus had spread globally and by July over 8000 SARS cases and approximately 800 deaths were reported worldwide [45]. At the time of the outbreak, no effective treatment of SARS was available, thus isolation and infection control were the best way to limit the spread of the virus. Therefore, rapid and reliable, early diagnosis is

critical to control such an epidemic. Zhu et al developed the first virus protein microarray, which included all the SARS-CoV proteins as well as proteins from five additional coronaviruses that can infect human (HCoV-229E and HCoV-OC43), cow (BCoV), cat (FIPV) and mouse (MHV-1) [46]. The SARS microarray was used to screen sera from infected and noninfected individuals in a double-blind format. The samples were quickly distinguished as SARS positive or SARS negative based on the presence of human IgG and IgM antibodies against SARS-CoV proteins, with a 94% accuracy rate compared to a standard ELISA diagnostic test. The SARS microarray improved the sensitivity of the assay 50 fold over the ELISA and dramatically reduced the amount of sample required. This method may be suitable for diagnosis for many viral infections.

Novel serological biomarkers for inflammatory bowel disease

The two most common subtypes of inflammatory bowel disease (IBD) are Crohn's disease (CD) and ulcerative colitis (UC). They are idiopathic in nature and are both characterized by an abnormal immunological response in the gut [47]. IBD is clinically thought to have autoimmune etiology, although, anti-microbial antibodies to normal bacteria are present in the sera of patients, leading to the pathogenesis of the disease [8]. The known serological antibodies are currently used as partial diagnostic criteria as they are not robust enough to stand alone [48]. Chen et al. elected to use an *E. coli* proteome microarray to characterize the differential immune response (serum anti-*E. coli* antibodies) in patients with CD and UC compared to healthy controls (HC). The microarray included 4256 *E. coli* proteins, encompassing the vast majority of the

proteome of *E. coli* K12 strain. The sera from HC (n=29), CD (n=66) and UC (n=39) were profiled using this array and the reactive anti-*E. coli* antibodies were detected with anti-human IgG antibodies. Data analysis revealed differential immunogenic response to 417 proteins between these three groups: 169, 186 and 19 were highly immunogenic in HC, CD and UC, respectively. Two robust sets of novel serological biomarkers were identified that can discriminate CD from HC or UC with >80% overall accuracy and sensitivity [8]. This is the first study to identify serological biomarkers in human immunological diseases with respect to the entire proteome of a microbial species. The underlying molecular pathology of other immune system related diseases can also be examined with this proteome microarray approach.

Lectin study: protein-glycan interaction

Cell surface glycosylation is a complex and highly-varied PTM that in turn is not amenable to standard high-throughput techniques. Glycosylation is present on the surface of all vertebrate cells, and it serves to distinguish cell types through very delicate differences [9]. It is also shown to be associated with cell differentiation, malignant transformation and subcellular localization [49-53]. Glycan binding proteins, known as lectins, are used to characterize glycosylation marks due to their ability to discriminate sugar isoforms [54]. Lectin microarrays have already been employed to characterize glycoproteins and lysates [55, 56], however, they have not been used to systematically profile cell surface glycosylation signatures of mammalian cell types. Such studies have the potential to provide a tool for distinguishing normal versus abnormal cell surface profiles based on glycan-lectin interactions. Tao et al. fabricated a lectin microarray

composed of 94 non-redundant lectins selected for defining cell surface glycan signatures [5]. Using 23 well-studied mammalian cell lines, the authors developed a systematic binary analysis of binding interactions of the selected lectins and cell types. They observed a broad range of binding potential and specificity across cell types, implying a high level of variation in cell surface glycans within mammalian cell types. For example, less than 20 lectins could capture the hESC, Caco-2, D407 and U937 cells, while more than 50 lectins captured the HEK293, K1106 and MCF7 cells [9]. Interestingly, similar cell types such as various breast cancer cell lines did not reveal overlapping lectin binding profiles, indicating lectins can discern subtle differences between physiologically-related cells.

To further test the utility of the lectin microarray for biomarker discovery, Tao et al. analyzed lectin binding in a model cancer stem-like system by comparing cell surface glycan signatures of all 24 cell types [9]. Focusing on MCF7, a breast cancer cell line that adopts cancer stem-like phenotypes when grown under specific conditions, the authors demonstrated that different growth conditions give rise to distinct lectin binding profiles that can distinguish these cancer cell subpopulations [9]. The lectin LEL was identified as a biomarker that can discriminate between MCF7 subpopulations. The authors propose that combined with other stem cell enrichment methods, lectin microarray technology is a potential tool for identifying cell surface markers in tumors, enabling the discovery of cancer stem cell-like targeted therapies.

Perspectives

Over the past decade, protein microarrays have evolved into a powerful and versatile tool for systems biology. They capitalize on femtomolar sensitivity, profiling full proteomes and high-throughput yet straightforward assays. We have described their utility for a myriad of applications that have resulted in impactful scientific findings including pathogen-host interactions, biomarker identification, unconventional transcription factors and PTM substrates (Figure 1-1). While protein microarrays leverage the advantage of uniform protein expression, for proteomics, their impact is limited by the extent of coverage. A remarkable advance was put forth by the Zhu laboratory with the construction of the first human proteome microarray containing over 17,000 full length proteins [15], the largest available to date (Figure 1-2). The discovery potential for this technology is dramatically increased by expanded proteome coverage. Multiple large-scale studies intended to link PTM substrates with their upstream enzymes, such as kinases, SUMO E3 ligases and ubiquitin ligases, are ongoing with the human proteome microarrays. As the number of *bona fide* PTMs increase and more substrates are found to acquire numerous modifications, we cannot ignore coregulation of PTMs. Directed studies to recapitulate crosstalk between enzymes, PTMs and their common substrates are possible with protein microarrays and may uncover key nodes of regulation and critical points where pathways converge. While MS is an ideal technology for the discovery of novel PTMs, such as the crotonylation PTM [57], it is not well suited to identify the enzymes responsible for novel modification. The richness of 17,000 natively-purified proteins on a single surface provides an ideal platform for discovery of novel enzyme

function. The human proteome array can also be harnessed as a tool for high-throughput characterization of monoclonal antibody (mAb) specificity from hybridomas [15].

The capabilities of microarray technology are further expanding with the development of label-free optical techniques that monitor the real-time dynamics of biomolecular interactions. Oblique-incidence reflectivity difference (OIRD) is an emerging technique that measures the changes in reflectivity of polarized light [58, 59]. OIRD has recently been applied to DNA and protein microarrays and has successfully determine association and dissociation rates of biomolecular interactions in a high-throughput format [60, 61].

Constructing complex interaction networks involving the full range of cellular components is critical for deciphering how organisms are organized and is essential for understanding the aberrant changes that result in diseases. We have explored the vast applications of protein microarrays for global characterization of interactomes and the significance of their findings for creating a comprehensive view of biological systems. Protein microarray technology is no longer in its infancy and will undoubtedly serve as an invaluable tool for proteomics and systems biology.

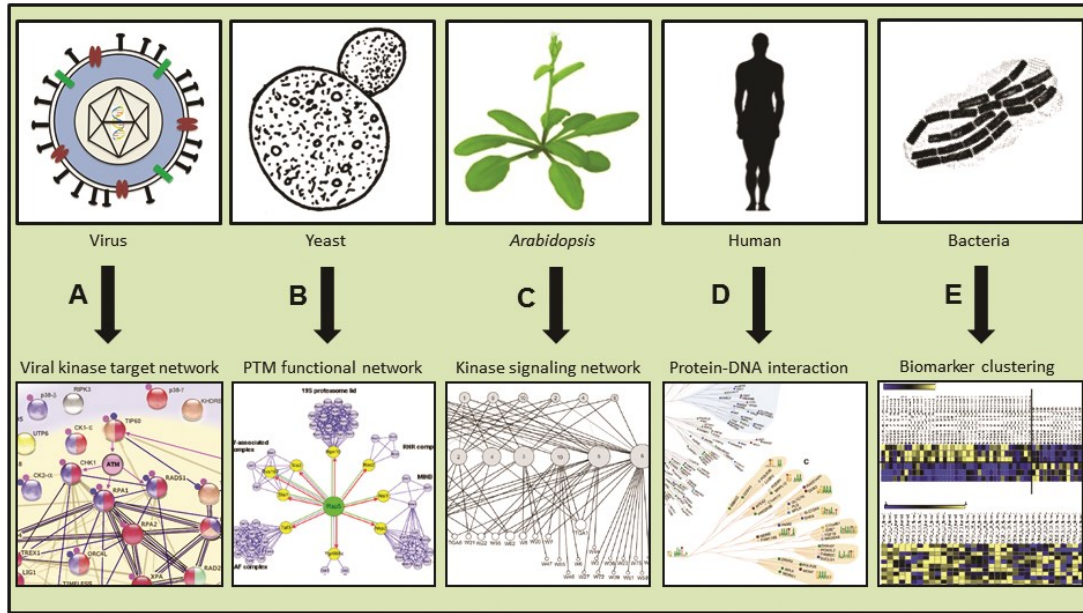


Figure 1-1. Reconstituted interaction networks in cellular systems generated through protein microarray studies

Interaction mapping with protein microarrays has been applied to numerous organisms to achieve diverse representations of molecular networks. **A.** Li et al. probed a human transcription factor (TF) microarray with four conserved kinases encoded by herpesviruses to reveal the host targets of the viral kinases [34]. **B.** Using a yeast proteome microarray, Lu et al. identified the substrates of the HECT E3 ligase Rsp5 [21]. Through gene ontology analysis Rsp5 was linked to subgroups of substrates based on function. **C.** The *Arabidopsis thaliana* MAP kinase signaling network was reconstructed using an *Arabidopsis* protein microarray [24]. The hierarchical phosphorylation network depicts the MKKs (upper nodes), MPKs (middle nodes), and substrates (bottom nodes). (Adapted with author's permission) **D.** DNA binding specificity of unconventional DNA-binding proteins (uDBPs) was characterized using the TF microarray [17]. The uDBPs are clustered based on target sequence similarity and proteins of different functional classes are color-coded. "C" denotes consensus sequences for each sub-branch are shown. **E.** The *E. coli* proteome microarray was used to identify differentially immunogenic proteins between HC and CD patient samples depicted in the heat map [8]. The yellow and blue colors indicate high and low immunogenic responses, respectively. HC, healthy controls; CD, Crohn's disease.

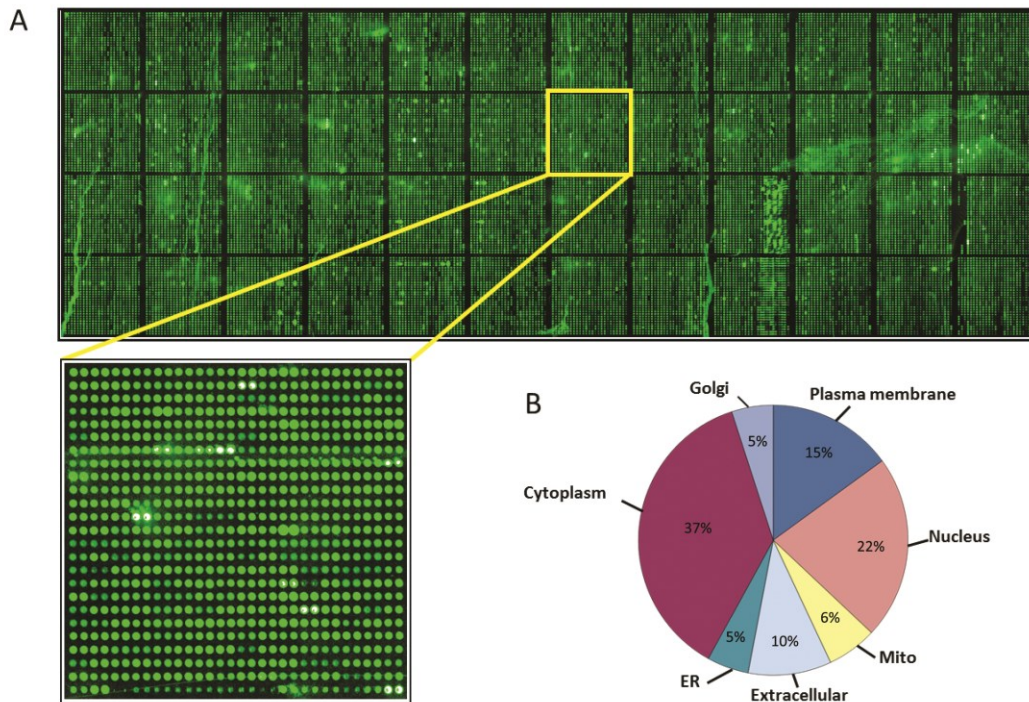


Figure 1-2. The human proteome microarray

A. The human proteome microarray composed of 16,368 unique full-length recombinant proteins printed in duplicate on Full Moon glass slides. To monitor the quality, the microarray was probed with anti-GST monoclonal antibody, followed by Alexa-555 secondary antibody to visualize the signals. The proteins positively detected by the anti-GST antibody are represented in green. **B.** Cellular distribution of the proteins included in the human proteome microarray.

ER, endoplasmic reticulum; Mito, mitochondria.

References

- [1] Kenyon GL, DeMarini DM, Fuchs E, Galas DJ, Kirsch JF, Leyh TS, et al. Defining the mandate of proteomics in the post-genomics era: workshop report. *Mol Cell Proteomics*. 2002;1:763-80.
- [2] Neet KE, Lee JC. Biophysical characterization of proteins in the post-genomic era of proteomics. *Mol Cell Proteomics*. 2002;1:415-20.
- [3] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422:198-207.
- [4] Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol*. 2005;6.
- [5] Yang L, Guo S, Li Y, Zhou S, Tao S. Protein microarrays for systems biology. *Acta Biochim Biophys Sin (Shanghai)*. 2011;43:161-71.
- [6] Smith MG, Jona G, Ptacek J, Devgan G, Zhu H, Zhu X, et al. Global analysis of protein function using protein microarrays. *Mech Ageing Dev*. 2005;126:171-5.
- [7] Zhu H, Qian J. Applications of functional protein microarrays in basic and clinical research. *Adv Genet*. 2012;79:123-55.
- [8] Chen CS, Sullivan S, Anderson T, Tan AC, Alex PJ, Brant SR, et al. Identification of novel serological biomarkers for inflammatory bowel disease using *Escherichia coli* proteome chip. *Mol Cell Proteomics*. 2009;8:1765-76.
- [9] Tao SC, Li Y, Zhou J, Qian J, Schnaar RL, Zhang Y, et al. Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology*. 2008;18:761-9.
- [10] Kumble KD. Protein microarrays: new tools for pharmaceutical development. *Anal Bioanal Chem*. 2003;377:812-9.
- [11] Hu S, Xie Z, Qian J, Blackshaw S, Zhu H. Functional protein microarray technology. *Wiley Interdiscip Rev Syst Biol Med*. 2011;3:255-68.

- [12] Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, et al. Global analysis of protein activities using proteome chips. *Science*. 2001;293:2101-5.
- [13] Zhu H, Klemic J, Chang S, Bertone P, Casamayor A, Klemic K, et al. Analysis of yeast protein kinases using protein chips. *Nat Genet*. 2000;26:283-9.
- [14] Li R, Wang L, Liao G, Guzzo CM, Matunis MJ, Zhu H, et al. SUMO binding by the Epstein-Barr virus protein kinase BGLF4 is crucial for BGLF4 function. *J Virol*. 2012;86:5412-21.
- [15] Jeong JS, Jiang L, Albino E, Marrero J, Rho HS, Hu J, et al. Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. *Mol Cell Proteomics*. 2012;11:1-10.
- [16] Huang J, Zhu H, Haggarty SJ, Spring DR, Hwang H, Jin F, et al. Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc Natl Acad Sci U S A*. 2004;101:16594-9.
- [17] Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J, et al. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell*. 2009;139:610-22.
- [18] Zhu J, Gopinath K, Murali A, Yi G, Hayward SD, Zhu H, et al. RNA-binding proteins that inhibit RNA virus infection. *Proc Natl Acad Sci U S A*. 2007;104:3129-34.
- [19] Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, et al. Global analysis of protein phosphorylation in yeast. *Nature*. 2005;438:679-84.
- [20] Merbl Y, Kirschner MW. Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. *Proc Natl Acad Sci U S A*. 2009;106:2543-8.
- [21] Lu JY, Lin YY, Qian J, Tao SC, Zhu J, Pickart C, et al. Functional dissection of a HECT ubiquitin E3 ligase. *Mol Cell Proteomics*. 2008;7:35-45.
- [22] Lin YY, Lu JY, Zhang J, Walter W, Dang W, Wan J, et al. Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell*. 2009;136:1073-84.

- [23] Foster MW, Forrester MT, Stamler JS. A protein microarray-based analysis of S-nitrosylation. *Proc Natl Acad Sci U S A*. 2009;106:18948-53.
- [24] Popescu SC, Popescu GV, Bachan S, Zhang Z, Gerstein M, Snyder M, et al. MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Dev*. 2009;23:80-92.
- [25] Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, et al. MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature*. 2002;415:977-83.
- [26] Ichimura K, Mizoguchi T, Yoshida R, Yuasa T, Shinozaki K. Various abiotic stresses rapidly activate *Arabidopsis* MAP kinases ATMPK4 and ATMPK6. *Plant J*. 2000;24:655-65.
- [27] Hill APVaCP. Structural basis of ubiquitylation. *Curr Opin Struct Biol*. 2002;12:822-30.
- [28] Close P, Creppe C, Gillard M, Ladang A, Chapelle JP, Nguyen L, et al. The emerging role of lysine acetylation of non-nuclear proteins. *Cell Mol Life Sci*. 2010;67:1255-64.
- [29] Allard S, Utlely RT, Savard J, Clarke A, Grant P, Brandl CJ, et al. NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *EMBO J*. 1999;18:5108-19.
- [30] Lu JY, Lin YY, Sheu JC, Wu JT, Lee FJ, Chen Y, et al. Acetylation of yeast AMPK controls intrinsic aging independently of caloric restriction. *Cell*. 2011;146:969-79.
- [31] Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 2009;325:834-40.
- [32] Chen Y, Zhao W, Yang JS, Cheng Z, Luo H, Lu Z, et al. Quantitative acetylome analysis reveals the roles of SIRT1 in regulating diverse substrates and cellular pathways. *Mol Cell Proteomics*. 2012;11:1048-62.

- [33] Zhang T, Wang S, Lin Y, Xu W, Ye D, Xiong Y, et al. Acetylation negatively regulates glycogen phosphorylase by recruiting protein phosphatase 1. *Cell Metab.* 2012;15:75-87.
- [34] Li R, Zhu J, Xie Z, Liao G, Liu J, Chen MR, et al. Conserved herpesvirus kinases target the DNA damage response pathway and TIP60 histone acetyltransferase to promote virus replication. *Cell Host Microbe.* 2011;10:390-400.
- [35] Shamay M, Liu J, Li R, Liao G, Shen L, Greenway M, et al. A protein array screen for Kaposi's sarcoma-associated herpesvirus LANA interactors links LANA to TIP60, PP2A activity, and telomere shortening. *J Virol.* 2012;86:5179-91.
- [36] Arvin A C-FG, Mocarski E, et al., . *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Cambridge: Cambridge University Press; 2007.
- [37] Gershburg E, Pagano JS. Conserved herpesvirus protein kinases. *Biochim Biophys Acta.* 2008;1784:203-12.
- [38] Romaker D, Schregel V, Maurer K, Auerochs S, Marzi A, Sticht H, et al. Analysis of the Structure–Activity Relationship of Four Herpesviral UL97 Subfamily Protein Kinases Reveals Partial but not Full Functional Conservation†. *J Med Chem.* 2006;49:7044-53.
- [39] Meng Q, Hagemeyer SR, Fingerroth JD, Gershburg E, Pagano JS, Kenney SC. The Epstein-Barr virus (EBV)-encoded protein kinase, EBV-PK, but not the thymidine kinase (EBV-TK), is required for ganciclovir and acyclovir inhibition of lytic viral production. *J Virol.* 2010;84:4534-42.
- [40] Gershburg E, Raffa S, Torrisi MR, Pagano JS. Epstein-Barr virus-encoded protein kinase (BGLF4) is involved in production of infectious virus. *J Virol.* 2007;81:5407-12.
- [41] Gershburg E, Pagano JS. Phosphorylation of the Epstein-Barr Virus (EBV) DNA Polymerase Processivity Factor EA-D by the EBV-Encoded Protein Kinase and Effects of the L-Riboside Benzimidazole 1263W94. *J Virol.* 2002;76:998-1003.
- [42] Gareau JR, Lima CD. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol.* 2010;11:861-71.
- [43] Johnson ES. Protein modification by SUMO. *Annu Rev Biochem.* 2004;73:355-82.

- [44] Mattoon D, Michaud G, Merkel J, Schweitzer B. Biomarker discovery using protein microarray technology platforms: antibody-antigen complex profiling. *Expert Rev Proteomics*. 2005;2:879-89.
- [45] Satija N, Lal SK. The molecular biology of SARS coronavirus. *Ann N Y Acad Sci*. 2007;1102:26-38.
- [46] Zhu H, Hu S, Jona G, Zhu X, Kreiswirth N, Willey BM, et al. Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. *Proc Natl Acad Sci U S A*. 2006;103:4011-6.
- [47] Thoreson R, Cullen JJ. Pathophysiology of inflammatory bowel disease: an overview. *Surg Clin North Am*. 2007;87:575-85.
- [48] Bossuyt X. Serologic markers in inflammatory bowel disease. *Clin Chem*. 2006;52:171-81.
- [49] Hakomori S. Aberrant glycosylation in cancer cell membranes as focused on glycolipids: overview and perspectives. *Cancer Res*. 1985;45:2405-14.
- [50] Hakomori S. Glycosylation defining cancer malignancy: new wine in an old bottle. *Proc Natl Acad Sci U S A*. 2002;99:10231-3.
- [51] Dennis JW, Granovsky M, Warren CE. Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta*. 1999;1473:21-34.
- [52] Dwek MV, Ross HA, Leatham AJ. Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. *Proteomics*. 2001;1:756-62.
- [53] Kung LA, Tao SC, Qian J, Smith MG, Snyder M, Zhu H. Global analysis of the glycoproteome in *Saccharomyces cerevisiae* reveals new roles for protein glycosylation in eukaryotes. *Mol Syst Biol*. 2009;5:308.
- [54] Hirabayashi J. Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconj J*. 2004;21:35-40.

- [55] Ebe Y, Kuno A, Uchiyama N, Koseki-Kuno S, Yamada M, Sato T, et al. Application of lectin microarray to crude samples: differential glycan profiling of lec mutants. *J Biochem.* 2006;139:323-7.
- [56] Kuno A, Uchiyama N, Koseki-Kuno S, Ebe Y, Takashima S, Yamada M, et al. Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nat Methods.* 2005;2:851-6.
- [57] Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell.* 2011;146:1016-28.
- [58] Fan Chen HL, Zhenghao Chen, Tong Zhao, and Guozhen Yang. Optical real-time monitoring of the laser molecular-beam epitaxial growth of perovskite oxide thin films by an oblique-incidence reflectance-difference technique. *J Opt Soc Am B.* 2000;18:1031-5.
- [59] Landry JP, Zhu XD, Gregg JP. Label-free detection of microarrays of biomolecules by oblique-incidence reflectivity difference microscopy. *Opt Lett.* 2004;29:581-3.
- [60] Lu H, Wen J, Wang X, Yuan K, Li W, Lu H, et al. Detection of the specific binding on protein microarrays by oblique-incidence reflectivity difference method. *Journal of Optics.* 2010;12:1-5.
- [61] Xu W, Heng L, Juan W, Kun Y, Hui-Bin L, Kui-Juan J, et al. Label-Free and High-Throughput Detection of Protein Microarrays by Oblique-Incidence Reflectivity Difference Method. *Chinese Physics Letters.* 2010;27:1-4

CHAPTER 2

SUMOYLATION: FEATURES, CONJUGATION, AND FUNCTION

2.1. Features

Posttranslational modifications (PTMs) are critical regulatory events that alter the properties of a protein by proteolytic processing or addition of charged chemical groups, small proteins, sugars, or lipids. Ubiquitin-like (Ubl) proteins are reversible PTMs catalyzed by dynamic coordination of transferase enzymes to regulate diverse cellular processes. The SUMO proteins (Small Ubiquitin-related MOdifiers) are some of the best characterized members of the Ubl family and are known to be conjugated to over >800 putative targets based on large scale screens and empirical studies [1-3]. Attachment of the SUMO moiety to a substrate protein is important for regulation of membrane receptors down to cytosolic proteins and nuclear transcription factors. The mechanisms of attachment for SUMOylation and ubiquitylation are analogous and the three-dimensional structures are virtually superimposable, however, SUMO only shares 18% sequence identity with ubiquitin [4]. The surface charge topology is distinct between the two molecules creating distinct negative and positively charged patches [5]. Differences in sequence and charge allow for recognition by diverse effector molecules and unique downstream signaling outputs. SUMO protein is approximately 11 kDa and composed of around 100 amino acids, containing a flexible N-terminal region which is absent in other Ubl proteins. Within this unstructured stretch of amino acids are several lysines that serve as sites for polymeric chain formation.

SUMO proteins are highly conserved throughout eukaryotes. In yeast and other invertebrates, a single gene encodes for SUMO, known as SMT3 in *Saccharomyces*

cerevisiae. Vertebrates encode for four distinct SUMO paralogs on separate genes, designated as SUMO1 through SUMO4. SUMO1 is an 11 kDa protein first described as a covalent modification on nuclear pore protein RanGAP1 [6]. In vertebrates, SUMO2 and SUMO3 are commonly referred to as SUMO2/3 because they share 97% sequence identity, only differing by three amino acids. SUMO1 shares roughly 50% sequence identity with SUMO2/3 and is distinguishable by immune-based detection methods. Similar to ubiquitin, SUMO2/3 is capable of forming polymeric SUMO2/3 chains through the SUMOylation consensus motif, ΨKXE, present in its N-terminal extension. SUMO1 lacks this attachment site and therefore is not known to form chains *in vivo*, however, it is proposed to function as a polymeric chain terminator [7]. SUMO4 was initially thought to be a pseudogene because it lacks introns but recent studies have suggested that its protein product contains an unconserved proline residue that interferes with proteolytic processing, thus preventing maturation and conjugation [8, 9]. There are some gross as well as subtle differences between the SUMO paralogs. Certain substrates are preferentially modified by a particular SUMO while others are modified equally with either SUMO. For instance, RanGAP1 is efficiently modified by SUMO1 but modified weakly by SUMO2 [10]. In addition, in the cell there is a large pool of unconjugated or “free” SUMO2/3 while there is limited SUMO1 because the majority of SUMO1 is conjugated to cellular proteins [10]. Free SUMO2/3 is rapidly conjugated to substrates when the cell is exposed to stress stimuli where SUMO1 is not [11].

2.2. SUMO Conjugation

The SUMO conjugation cascade covalently attaches SUMOs to lysine residues in a reversible mechanism. The cascade resembles ubiquitin conjugation, however, the enzymes involved are unique. SUMO is synthesized as an immature precursor protein with a C-terminal region of amino acids that block its reactivity. The family of SUMO specific proteases (knowns as sentrin proteases or SENPs) cleave several C-terminal residues revealing a di-glycine motif rendering the polypeptides competent for conjugation. SAE1/UBA2, the heterodimeric SUMO E1 enzyme, catalyzes the first step in the three-part reaction. To begin the reaction, the C-terminal glycine carboxyl group of SUMO attacks ATP at the gamma phosphate, producing a high energy adenylated-SUMO, thereby releasing pyrophosphate [12]. The thiol group of the active site cysteine of the E1 attacks the SUMO-adenylate, forming a high energy thiolester bond with the C-terminus of SUMO and the E1 cysteine. In the following step, Ubc9, the sole conjugating enzyme (E2) transfers the SUMO to its active cysteine residue, forming a second thiolester intermediate. In the final step of the reaction SUMO is transferred from the E2 to the amino group of a lysine residue in the substrate protein, forming an isopeptide bond [12-14] (Figure 2-1). A significant percentage of SUMO lysine attachment sites lie within the Ubc9 binding motif Ψ KXE/D, where Ψ is a large hydrophobic residue, K is the acceptor lysine, X is any amino acid. The E2 can guide SUMOylation of the substrate through its ability to bind to the Ψ KXD/E motif [15]. The chemistry of this mechanism is virtually identical to ubiquitylation, however, there are drastic organizational differences. The number of enzymes and transferases for SUMOylation is far more modest than in the ubiquitylation system. Ubiquitylation uses two E1s and over

30 E2s enzymes as compared SUMOylation's single E1 and E2. Ubiquitylation is catalyzed by over 600 E3 ligases whereas number of SUMO E3 ligases is growing but remains less than 15.

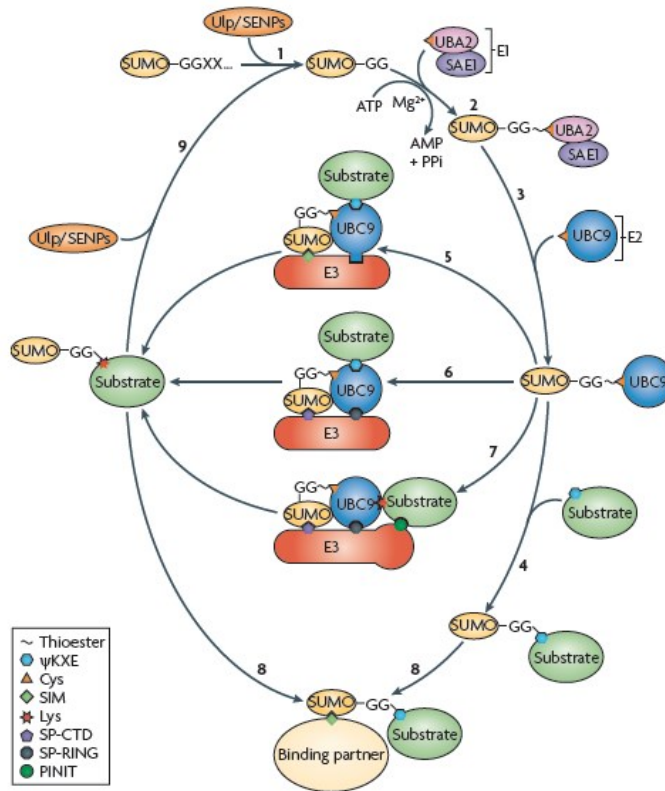


Figure 2-1. The SUMO conjugation cycle. Adapted from Gareau and Lima 2010, [16].

SUMO undergoes processing by SENPs to its mature form (step 1). SUMO is then adenylated by the E1 complex (SAE1/UBA2) in an ATP-dependent reaction. Following activation, SUMO is transferred to the catalytic cysteine of the E2 (UBC9) (step 3). It can then catalyse conjugation to a substrate in an E3 ligase-independent manner through recognition of SUMO consensus motifs (Ψ KXD/E) (step 4). Substrate specificity imparted by E3–substrate interactions are thought to be particularly important for directing conjugation to non-consensus lysine residues. Substrates modified by SUMO can interact with SUMO-binding proteins through their SUMO-interacting motifs (SIMs) (step 8). SUMO cleavage is catalyzed by SENP proteases to release free SUMO for future conjugation (step 9) [16].

2.3. SUMO E3 ligases

Since the discovery of SUMO, the role of SUMO E3 ligases has been mysterious. At first, it was even unclear whether E3s participated in this system at all because they are not required for SUMOylation of many substrates *in vitro* [17, 18]. In addition, they do not form covalent intermediates with SUMO substrates during conjugation, making them difficult to identify. The basic definition of a SUMO E3 is that it binds the E2, binds to the substrate, and assists the transfer of SUMO from the E2 to the substrate *in vitro* [19, 20]. Yeast demonstrate E3-dependent SUMOylation *in vivo* and *in vitro* [21, 22]. The contribution of E3 ligases to SUMOylation in vertebrates remains unclear and the majority of *in vivo* SUMO substrates are not linked to an upstream E3 ligase. There are

estimated to be between 600 and 1000 ubiquitin E3 ligases while to date only 15 SUMO E3 ligases have been identified. Among those discovered, there are three distinct types of SUMO E3 ligases: the PIAS (Protein inhibitor of activated STAT) family, that also function as STAT transcription factors; the nuclear pore protein RanBP2; and the polycomb group protein Pc2.

The mammalian PIAS proteins are encoded by four genes, PIAS1, PIAS2(xb), PIAS3, and PIAS4(y). They share a high degree of sequence homology and range in size from 507 (PIASy) and 650 (PIAS1) amino acids. Notably, the PIAS proteins contain SP-RING domains that resemble RING domains of ubiquitin E3 ligases [23, 24]. Thus their ligase activity is thought to be analogous to ubiquitin RING domain E3 ligases through binding the E2~SUMO and substrate, bringing them in closer proximity with the substrate [19, 23, 25-27]. PIAS proteins also have the ability to interact with SUMO through their SUMO interaction motifs [28], however mutation of the SIMs show minimal effect on E3 activity [26]. In Onishi et al 2007, two point mutations in the RING domain of PIAS3 ablated its E3 ligase activity [29].

RanBP2 does not contain a ring domain and functions in a different manner than the PIAS E3 ligases. Its E3 domain contains multiple regions that have ligase activity, that bind either SUMO or the E2, coordinating an optimal orientation of the E2~SUMO thioester for catalysis and substrate interaction [16]. In addition to RanGAP1, RanBP2 enhances SUMOylation of Sp100, HDAC, hnRNP C1 and PML proteins *in vitro* [30-33]. There is little *in vivo* evidence to corroborate this, but it is presumed that RanBP2 enhances SUMOylation of these and other proteins during nuclear import and export [31]. Several other SUMO E3 ligases have been reported in the literature namely Pc2, G-

protein Rhes, DNA topoisomerase binding protein TOPORS, and HDAC4 [34-37], though the mechanisms by which they enhance SUMOylation are not clear. How exactly E3 ligases achieve substrate specificity remains enigmatic. Some E3 ligases show specificity for substrates while others demonstrate redundancy. In other cases, E3 ligases show paralog-preference for SUMOylation of targets. Chapter 3 of this thesis will discuss our efforts to identify the substrates and paralog preference of several E3 ligases in a global screen. Our studies have been able to establish connections between six E3 ligases and hundreds of their substrates to reveal the specificity and redundancy within SUMO E3 ligase substrate network.

2.4. Deconjugation

SUMOylation is a reversible process where SUMO is rapidly deconjugated from substrates by SENPs, which are the same family of proteases involved in SUMO maturation. In yeast there are two proteases, Ulp1 and Ulp2, with six homologs in vertebrates known as SENP1-3 and SENP5-7. The SENPs have distinct roles in executing maturation and deconjugation for the SUMO paralogs. SENP1 and SENP2 are involved in deconjugation for SUMO1 and SUMO2/3, while SENP6 and SENP7 are not involved in maturation or deconjugation but serve as polymeric chain editors.

2.5. SUMO Substrate Specificity

Of the SUMOylation sites that have been successfully mapped, many contain the Ψ KXD/E motif, but those sites do not account for all target lysines. Thus, this motif cannot universally explain all of the determinants of SUMOylation. Additionally, this motif is very common and occurs very frequently in protein sequences, though in most

cases is not SUMOylated [19]. For example, in yeast, out of 5884 open reading frames (ORFs) the (IVL)KXE occurs 2799 times in 1913 ORFs [19]. This points to the concept that there must be other criteria required for a target lysine or consensus motif to be a bona fide SUMOylation site. Several SUMO attachment sites that do not occur within a consensus motif have been identified and other proteins in which all consensus motifs mutated are still modified [19]. A common presumption is that E3 ligases assist in mediating target selection, especially for lysines that do not lie in a consensus motif. E3 ligases can discriminate between SUMO isoforms and in certain cases demonstrate preference for a particular SUMO. E3 ligases add to the complexity of SUMOylation with their ability to show preference and redundancy for substrates and SUMO isoforms. The work presented in chapter 3 will also shed light on E3 ligase SUMO paralog preference *in vitro* and in cell-based systems.

A Phosphorylation dependent SUMOylation motif (PDSM) has been described, Ψ KX(D/E)XXSP, where in the phosphorylation of the downstream serine residue enhances SUMOylation at the target lysine [38] (Figure 2-2). In the case of the PDSM containing substrates PPAR γ and MEF2A, phosphorylation and subsequent SUMOylation switches their activity from transcriptional activation to transcriptional repression [38] [39]. Negatively Charged Amino Acid Dependent SUMOylation Motif (NDSM) is a variation on the PDSM motif where the negatively charged C-terminal residues mimic the negatively charge phospho-serine in the PDSM [40]. The negatively charged extension increases the interaction surface with Ubc9 [40].

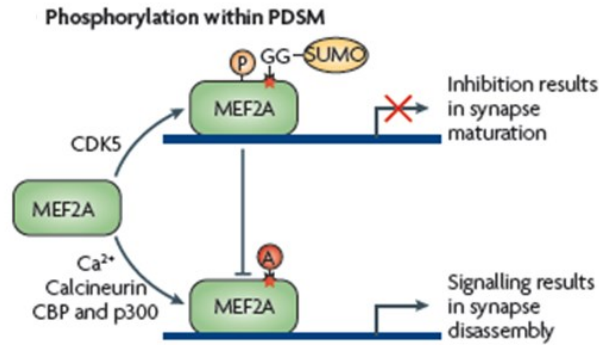


Figure 2-2. Phosphorylation dependent SUMOylation. Adapted from Gareau and Lima, 2010 [16].

Phosphorylation of the transcription factor MEF2A by CDK5 within a PDSM enhances its modification by SUMO, inhibiting transcription and resulting in synapse maturation. Figure adapted from [16].

SUMO Interaction Motifs (SIMs) are typically found in SUMO binding proteins. They mediate non-covalent recognition of SUMOylated proteins and serve as readers/effectors or the small protein PTM. SIM domains within a target protein also contribute to substrate selection by recognizing the E2~SUMO species and promoting the transfer of SUMO onto itself [41]. The one known SUMO consensus motif V/I-X-V/I-V/I or V/I-V/I-X-V/I is flanked by acidic residues and/or serines. Through phosphorylation the serines acquire a negative charge and may mimic the acidic residues to strengthen binding [42, 43].

2.6. Functional Importance of SUMOylation

SUMO has been implicated in a wide variety of cellular functions such as transcription activation and repression, protein-protein interactions, enzyme activity, chromosome

segregation, DNA repair, protein localization, development, and nuclear transport [4, 29, 31, 38, 39, 44, 45]. Although SUMO is involved in a wide variety of cellular processes, the mechanism by which its effect is produced is substrate specific.

SUMOylation of RanGAP1 is a hallmark example of regulation of protein localization. Once SUMOylated, RanGAP1 binds to RanBP2, a component of the nuclear pore and SUMO E3 ligase, and is tethered to the nuclear envelope where it functions in nuclear protein import [6]. Prior to the discovery of SUMO modification on RanGAP1, it was known to be localized to both the cytoplasm and the NPC, though the molecular switch was unknown [4].

SALL4 is a transcription factor that functions in maintenance and self-renewal of embryonic and hematopoietic stem cells. SUMOylation of SALL4 positively regulates its stability by increasing the half-life of the protein [45]. SUMOylation deficient SALL4 is enriched at chromatin but not in the cytoplasm suggesting that SUMOylation drives cytoplasmic localization and may prevent chromatin association [45]. NR2E3 is a transcription factor that is enriched in developing photoreceptor cells. Paradoxically, it functions as both a transcriptional activator of rod genes and a transcriptional repressor of cone genes. PIAS3 mediates SUMOylation of NR2E3, conferring its opposing transcriptional functions. SUMOylation of NR23E is required for transcriptional repression of cone-specific genes [29].

An example of SUMO-mediated enzyme activation is SUMO substrate and proto-oncogene, Akt. Systematic analysis of lysines in Akt revealed K276 is essential for its activity [46]. This critical lysine also occurs in a SUMO consensus motif. PIAS1

mediates SUMOylation of Akt and over expression of PIAS1 enhances Akt activity, whereas expression of SENP1 reduces its activity [46].

Bloom's Syndrome Protein, BLM is critical for repair of damaged DNA replication forks and has been shown to have both pro and anti-recombinogenic functions in homologous recombination (HR) [47]. SUMOylation mediates the shift between the opposing roles in HR. When BLM is the non-SUMOylated form, it inhibits DNA repair at damaged replication forks whereas SUMOylated BLM promotes DNA replication for repair via HR, by promoting the function of the critical recombinase, RAD51 [47]. Here, SUMO is implicated as a critical signal in the DNA damage system.

2.7. PTM Crosstalk

These examples highlight the importance SUMO modification in cell signaling and maintenance of overall homeostasis, however SUMOylation may confer different signals in the context of other PTMs. Emerging evidence implicates SUMOylation as a player in PTM crosstalk with ubiquitin and phosphorylation. One residue may have the potential to be modified by various PTMs with opposing functions. For example, the ubiquitylation of K21 of I κ B leads to its degradation, whereas K21 SUMOylation stabilizes I κ B [48] (Figure 2-2). Moreover, numerous PTMs may be present on a protein at one time. A specific combination of PTMs can determine the localization, activation state, interaction partners, stability, or turnover of the substrate. Positive crosstalk between phosphorylation and SUMOylation is exemplified by PDSM substrates MEF2A, GATA-1, and ERR γ . On the contrary, phosphorylation of AIB1 inhibited its SUMOylation [38, 49]. The evidence in such empirical studies suggests SUMOylation and phosphorylation

may be connected and serve to co-regulate protein functions, but whether this is a systems level regulatory mechanism remains unanswered. In chapter 4, I will present work on characterization of the function of SUMOylation of in the dynamics tyrosine kinase, Pyk2. We reveal a bivalent mechanism of activation and interplay between the SUMO and phosphorylation systems.

References

- [1] Vertegaal AC, Andersen JS, Ogg SC, Hay RT, Mann M, Lamond AI. Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol Cell Proteomics*. 2006;5:2298-310.
- [2] Golebiowski F, Matic I, Tatham MH, Cole C, Yin Y, Nakamura A, et al. System-wide changes to SUMO modifications in response to heat shock. *Sci Signal*. 2009;2:ra24.
- [3] Yao Q, Li H, Liu BQ, Huang XY, Guo L. SUMOylation-regulated protein phosphorylation, evidence from quantitative phosphoproteomics analyses. *J Biol Chem*. 2011;286:27342-9.
- [4] Matunis MJ, Wu J, Blobel G. SUMO-1 Modification and Its Role in Targeting the Ran GTPase-activating Protein, RanGAP1, to the Nuclear Pore Complex. *The Journal of Cell Biology*. 1998;140:499-509.
- [5] Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F, Jaenicke R, et al. Structure Determination of the Small Ubiquitin-related Modifier SUMO-1. *J Mol Biol*. 1998;280:275-86.
- [6] Matunis MJ, Coutavas E, Blobel G. A Novel Ubiquitin-like Modification Modulates the Partitioning of the Ran-GTPase-activating Protein RanGAP1 between the Cytosol and the Nuclear Pore Complex. *The Journal of Cell Biology*. 1996;135:1457-70.
- [7] Matic I, van Hagen M, Schimmel J, Macek B, Ogg SC, Tatham MH, et al. In Vivo Identification of Human Small Ubiquitin-like Modifier Polymerization Sites by High Accuracy Mass Spectrometry and an in Vitro to in Vivo Strategy. *Mol Cell Proteomics*. 2008;7.1:132-44.
- [8] Bohren KM, Nadkarni V, Song JH, Gabbay KH, Owerbach D. A M55V polymorphism in a novel SUMO gene (SUMO-4) differentially activates heat shock transcription factors and is associated with susceptibility to type I diabetes mellitus. *J Biol Chem*. 2004;279:27233-8.
- [9] Owerbach D, McKay EM, Yeh ET, Gabbay KH, Bohren KM. A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem Biophys Res Commun*. 2005;337:517-20.

- [10] Saitoh H. Functional Heterogeneity of Small Ubiquitin-related Protein Modifiers SUMO-1 versus SUMO-2/3. *J Biol Chem.* 2000;275:6252-8.
- [11] Tempe D, Piechaczyk M, Bossis G. SUMO under stress. *Biochem Soc Trans.* 2008;36:874-8.
- [12] Johnson ES. Ubc9p Is the Conjugating Enzyme for the Ubiquitin-like Protein Smt3p. *J Biol Chem.* 1997;272:26799-802.
- [13] Schwarz SE, Matuschewski K, Liakopoulos D, Scheffner M. The ubiquitin-like proteins SMT3 and SUMO-1 are conjugated by the UBC9 E2 enzyme. *Proceedings of the National Academy of Sciences.* 1998;95:560-4.
- [14] Desterro JMP, Thomson J, Hay RT. Ubch9 conjugates SUMO but not ubiquitin. *FEBS Lett.* 1997;417:297-300.
- [15] Sampson DA, Wang M, Matunis MJ. The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *J Biol Chem.* 2001;276:21664-9.
- [16] Gareau JR, Lima CD. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol.* 2010;11:861-71.
- [17] Desterro JMP. Identification of the Enzyme Required for Activation of the Small Ubiquitin-like Protein SUMO-1. *J Biol Chem.* 1999;274:10618-24.
- [18] Okuma T, Honda R, Ichikawa G, Tsumagari N, Yasuda H. In Vitro SUMO-1 Modification Requires Two Enzymatic Steps, E1 and E2. *Biochem Biophys Res Commun.* 1999;254:693-8.
- [19] Johnson ES. Protein modification by SUMO. *Annu Rev Biochem.* 2004;73:355-82.
- [20] Hershko A, Ciechanover A. The Ubiquitin System. *Annu Rev Biochem.* 1998;67:425-79.
- [21] Johnson ES, Gupta AA. An E3-like Factor that Promotes SUMO Conjugation to the Yeast Septins. *Cell.* 2001;106:735-44.

- [22] Takahashi Y, Toh EA, Kikuchi Y. A novel factor required for the SUMO1/Smt3 conjugation of yeast septins. *Gene*. 2001;275:223-31.
- [23] Hochstrasser M. SP-RING for SUMO: New Functions Bloom for a Ubiquitin-like Protein. *Cell*. 2001;107:5-8.
- [24] Jackson PK. A new RING for SUMO: wrestling transcriptional responses into nuclear bodies with PIAS family E3 SUMO ligases. *Genes Dev*. 2001;15:3053-8.
- [25] Kahyo T, Nishida T, Yasuda H. Involvement of PIAS1 in the SUMOylation of Tumor Suppressor p53. *Mol Cell*. 2001;8:713-8.
- [26] Sachdev S, Bruhn L, Sieber H, Pichler A, Melchior F, Grosschedl R. PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes Dev*. 2001;15:3088-103.
- [27] Takahashi Y, Kahyo T, Toh EA, Yasuda H, Kikuchi Y. Yeast Ull1/Siz1 is a novel SUMO1/Smt3 ligase for septin components and functions as an adaptor between conjugating enzyme and substrates. *J Biol Chem*. 2001;276:48973-7.
- [28] Minty A, Dumont X, Kaghad M, Caput D. Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif. *J Biol Chem*. 2000;275:36316-23.
- [29] Onishi A, Peng GH, Hsu C, Alexis U, Chen S, Blackshaw S. Pias3-dependent SUMOylation directs rod photoreceptor development. *Neuron*. 2009;61:234-46.
- [30] Pichler A, Gast A, Seeler JS, Dejean A, Melchior F. The Nucleoporin RanBP2 Has SUMO1 E3 Ligase Activity. *Cell*. 2002;108:109-20.
- [31] Vassileva MT, Matunis MJ. SUMO Modification of Heterogeneous Nuclear Ribonucleoproteins. *Mol Cell Biol*. 2004;24:3623-32.
- [32] Tatham MH, Kim S, Jaffray E, Song J, Chen Y, Hay RT. Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection. *Nat Struct Mol Biol*. 2005;12:67-74.

- [33] Kirsh O, Seeler JS, Pichler A, Gast A, Muller S, Miska E, et al. SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase. *The EMBO Journal*. 2002;21:2682-91.
- [34] D. Wotton JCM. Pc2 and SUMOylation. *Biochem Soc Trans*. 2007;35:1401-4.
- [35] Subramaniam S, Sixt KM, Barrow R, Snyder SH. Rhes, a striatal specific protein, mediates mutant-huntingtin cytotoxicity. *Science*. 2009;324:1327-30.
- [36] Weger S, Hammer E, Heilbronn R. Topors acts as a SUMO-1 E3 ligase for p53 in vitro and in vivo. *FEBS Lett*. 2005;579:5007-12.
- [37] Yang Y, Tse AK, Li P, Ma Q, Xiang S, Nicosia SV, et al. Inhibition of androgen receptor activity by histone deacetylase 4 through receptor SUMOylation. *Oncogene*. 2011;30:2207-18.
- [38] Hietakangas V, Anckar J, Blomster HA, Fujimoto M, Palvimo JJ, Nakai A, et al. PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A*. 2006;103:45-50.
- [39] Yamashita D, Yamaguchi T, Shimizu M, Nakata N, Hirose F, Osumi T. The transactivating function of peroxisome proliferator-activated receptor gamma is negatively regulated by SUMO conjugation in the amino-terminal domain. *Genes Cells*. 2004;9:1017-29.
- [40] Yang SH, Galanis A, Witty J, Sharrocks AD. An extended consensus motif enhances the specificity of substrate modification by SUMO. *The EMBO Journal*. 2006;25:5083-93.
- [41] Zhu J, Zhu S, Guzzo CM, Ellis NA, Sung KS, Choi CY, et al. Small ubiquitin-related modifier (SUMO) binding determines substrate recognition and paralog-selective SUMO modification. *J Biol Chem*. 2008;283:29405-15.
- [42] Hecker CM, Rabiller M, Haglund K, Bayer P, Dikic I. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem*. 2006;281:16117-27.
- [43] Song J, Durrin LK, Wilkinson TA, Krontiris TG, Chen Y. Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc Natl Acad Sci U S A*. 2004;101:14373-8.

- [44] Guzzo CM, Berndsen CE, Zhu J, Gupta V, Datta A, Greenberg RA, et al. RNF4-dependent hybrid SUMO-ubiquitin chains are signals for RAP80 and thereby mediate the recruitment of BRCA1 to sites of DNA damage. *Sci Signal*. 2012;5:ra88.
- [45] Yang F, Yao Y, Jiang Y, Lu L, Ma Y, Dai W. Sumoylation is important for stability, subcellular localization, and transcriptional activity of SALL4, an essential stem cell transcription factor. *J Biol Chem*. 2012;287:38600-8.
- [46] Li R, Wei J, Jiang C, Liu D, Deng L, Zhang K, et al. Akt SUMOylation regulates cell proliferation and tumorigenesis. *Cancer Res*. 2013;73:5742-53.
- [47] Ouyang KJ, Woo LL, Zhu J, Huo D, Matunis MJ, Ellis NA. SUMO modification regulates BLM and RAD51 interaction at damaged replication forks. *PLoS Biol*. 2009;7:e1000252.
- [48] Desterro JMP, Rodriguez MS, Hay RT. SUMO-1 Modification of I κ B α Inhibits NF- κ B Activation. *Mol Cell*. 1998;2:233-9.
- [49] Wu H, Sun L, Zhang Y, Chen Y, Shi B, Li R, et al. Coordinated regulation of AIB1 transcriptional activity by sumoylation and phosphorylation. *J Biol Chem*. 2006;281:21848-56.

CHAPTER 3

CONSTRUCTING GLOBAL SUMO E3 LIGASE SUBSTRATE NETWORKS UTILIZING PROTEIN MICROARRAY TECHNOLOGY

Abstract

Currently, SUMOylation of many known substrates is sufficiently catalyzed by the E1 and E2 enzymes *in vitro*, however, emerging *in vivo* evidence suggests that the E3 ligases may play an important role in many SUMO-dependent signaling pathways. Though it is widely accepted that E3s enhance the degree of SUMOylation of their substrates, there are few cases where SUMO modification is E3-dependent. While numerous studies have attempted to identify the complement of SUMO-modified proteins in the human proteome, a global approach examining the role of SUMO E3 ligases and their substrate specificity has not been reported. Our goal is to identify a comprehensive set of SUMO modified proteins, illuminating the E3 ligase-substrate network, and identifying the common and distinct substrates of several E3 ligases. To this end, we have developed a protein microarray-based approach to globally construct SUMOylation networks, utilizing the Human Proteome Microarray containing approximately 17,000 individually purified recombinant human proteins. More specifically, we successfully purified six active SUMO E3 ligases and performed SUMOylation reactions on the protein microarrays to identify E3-dependent substrates, using high and low E1/E2 reaction mixtures as comparisons. Over 2700 SUMO E3 substrates were identified under stringent cutoff values. Many were unique SUMOylation targets for a particular E3 ligase, suggesting E3 ligases play a large role in determining substrate selection *in vitro*. Gene ontology analysis recovered known biological functions for many E3 ligases and more importantly, revealed many novel functions that could potentially link SUMOylation to signaling pathways involving protein kinases. We have pursued carefully selected candidates for validation to confirm physiological relevance of

E3 ligase-dependent SUMOylation in HeLa cells. We believe that the work presented in this chapter will have a profound impact on the studies of SUMOylation and its crosstalk with other PTMs.

Introduction

Complexity in the SUMOylation pathway

The SUMOylation pathway is fraught with complexity at each level. To begin with, unlike most PTMs (e.g. ubiquitin or phosphorylation), there are four unique SUMO proteins with redundant and non-redundant features and functions. The E2 enzyme has the ability to direct SUMOylation of many targets *in vitro* without the need for E3 ligases [1]. There is one widely accepted SUMOylation motif however, it is neither necessary nor sufficient for SUMOylation, as SUMOylation also occurs at non-consensus motif lysines, and not all consensus motifs are SUMOylated [2, 3]. In some cases, *in vivo*, substrates are preferentially modified by SUMO1 over SUMO2 (e.g. RanGAP1) [2, 4], and *visa versa*, although the true reasons are unknown. E3 ligases do not form covalent intermediates with SUMO in assisting transfer and thus are difficult to identify. To date, many of the E3 ligases that have been characterized also have other critical cellular functions (e.g. PIAS are transcriptional co-regulators [5, 6]) making it difficult to assess their E3 ligase function with knockdown experiments. E3 ligases enhance SUMOylation of targets however, only a few cases describe true E3 ligase dependent SUMOylation [7]. PCNA is SUMOylated at two lysines, K127 which lies within a consensus motif, and K164 which is a non-consensus lysine. Siz1, a yeast PIAS homolog, is able to enhance SUMO modification at K127 although E1 and E2 are sufficient [7]. On the other hand, E1 and E2 cannot SUMOylate K164 and it is only modified in the presence of Siz1 (Figure 3-1) [7].

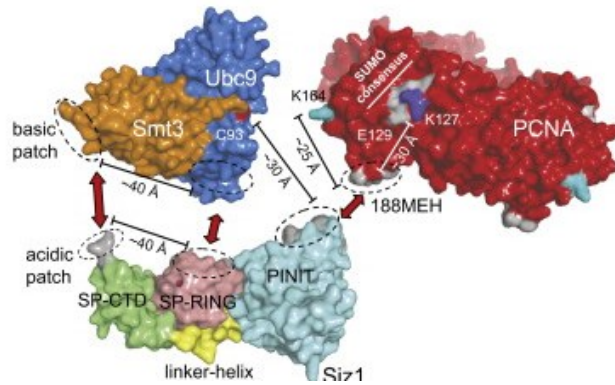


Figure 3-1. Siz1 E3 ligase-dependent SUMOylation of PCNA Lysine 164. Adapted from Yanus et al 2009, [7].

The model for activated E3:E2~SUMO-PCNA complex with the Siz1 E3 ligase, Ubc9 (blue), Smt3 (orange), and PCNA (red). Lys164 of PCNA is colored cyan, the SUMO consensus motif is colored gray with the target Lys127 colored blue. Siz1 activity is not required for modification at Lys127 although it enhances the SUMOylation. On the other hand, Lys164 SUMOylation requires Siz1 activity [7].

As alluded to, there are many layers to specificity in SUMOylation: the substrate, the SUMO paralog, and the E3 ligase. In order to develop a clear understanding of the organization and regulation of this complex system, global analysis of substrates, SUMO isoform preference, and E3 ligase specificity is needed. By identifying all of the SUMO substrates in the proteome, establishing connections between E3 ligases and their targets, and designation of the correct SUMO paralog, the intricate SUMO network will begin to emerge.

A study published in 2006 by Vertegaal et al attempted systematic identification of targets of SUMO1 and SUMO2 using the unbiased quantitative proteomic technique SILAC [3]. SUMO conjugates were purified from HeLa cells expressing His6-SUMO1 or His6-SUMO2. When dealing with whole cell extracts it is important to appreciate technical challenges due to the high activity of the SENPs that remove SUMO from the substrates, compounded with the low abundance of SUMO modified proteins. The nuclei were isolated and the high molecular weight proteins were separated by SDS PAGE and analyzed by tandem mass spectrometry. In total, 53 SUMOylation substrates were identified, including 9 that were previously discovered (e.g. RanGAP1 and PML) [3]. Twenty five substrates were preferential substrates of SUMO1, 19 were preferential substrates of SUMO2, and 9 exhibited little to no preference for SUMO1 or SUMO2 [3]. This is further indication that SUMO1 and SUMO2 likely have both overlapping and non-overlapping functions. Analysis of the SUMOylation consensus motif in the sequences of the 53 proteins revealed 112 consensus sites, which comprised 39 of the 53 targets (62%) [3]. This is consistent with theme that SUMO substrate selection is not entirely mediated through the consensus motif.

One critical difference between the functions of SUMO1 and SUMO2 lies in the response to cellular stress. There is a large pool of free unconjugated SUMO2/3 as compared to SUMO1, and after exposure to stress like heat shock, conjugation of SUMO2/3 but not SUMO1 occurs [8, 9]. In an effort to understand the role of SUMO in response to heat shock, Golebiowski et al performed a system wide proteomic analysis of targets of SUMO2. SILAC and tandem mass spectrometry identified 574 novel and 192 previously

known SUMO2 substrates. The findings further implicate SUMO2 as a key signaling feature of the cellular response to heat shock [10].

These studies provide us with a wealth of potential SUMO substrates, though there are likely many more. Additional proteomic studies are needed to identify the remaining SUMO substrates and alternative approaches are required to establish the substrate specificity of SUMO E3 ligases. Mass spectrometry is a powerful tool for proteomics as it allows for *in vivo* identification of substrates as well as the attachment sites. There are limitations to consider when using mass spectrometry for comprehensive SUMO proteomics 1. The inherently low relative abundance of SUMO substrates makes their recovery difficult 2. E3 ligase substrate specificity is not captured with these methods. Protein microarrays offer an alternative approach that mitigates these challenges. More specifically, protein microarrays provide an unbiased platform for screening the activity and specificity of the SUMOylation enzymes *in vitro*. Thousands of potential substrates can be queried at once using protein microarray technology, providing a versatile way to evaluate E3 ligase specificity as well as SUMO isoform preference.

Protein microarrays have been an important proteomics tool for substrate identification of PTMs such as phosphorylation, acetylation, and ubiquitylation [11-13]. The protein microarray method was previously employed in a proof of principal study for SUMOylation by Oh et al [9]. A fragment of RanGAP1 that is known to be SUMOylated was purified and spotted on a glass slide [14]. A K524A lysine mutant and bovine serum albumin were spotted as negative controls. The chip with the immobilized substrate and negative controls was incubated with a mixture of E1, E2, SUMO1, and ATP in reaction buffer. SUMOylation was detected with anti-SUMO1 antibody, followed by Cy3-labeled

anti-IgG secondary antibody for fluorescent detection. The SUMOylation reaction proved to be specific only for the WT RanGAP1 fragment and not the negative controls. This study was the first example of On-Chip SUMOylation. Blending this approach with breadth of substrates printed on the human proteome microarray provides an unbiased global substrate discovery tool (Figure 1-2).

3.1. Experimental Strategy

To include SUMO E3 ligase specificity in a protein microarray substrate discovery assay, the proportions of E1:E2:E3 enzymes must be titrated such that the true activity of the E3 can be assessed. As stated earlier, the E1 and E2 are often sufficient to SUMOylate many targets *in vitro* and thus simply adding an E3 ligase may be futile as the activity of the E1 and E2 may occlude the activity of the E3 or render it unnecessary. Reducing the concentrations of E1 and E2 to the extent that little to no SUMOylation is observed is the ideal background condition. By introducing the E3 ligase to the limiting concentration E1 and E2 mixture detection of the true E3 ligase-dependent or -specific substrates is possible. Rogers et al 2006 used this strategy to verify that SUMOylation of STAT1 with SUMO1 is enhanced with PIASx- α but not PIAS1 [1] (Figure 3-2).

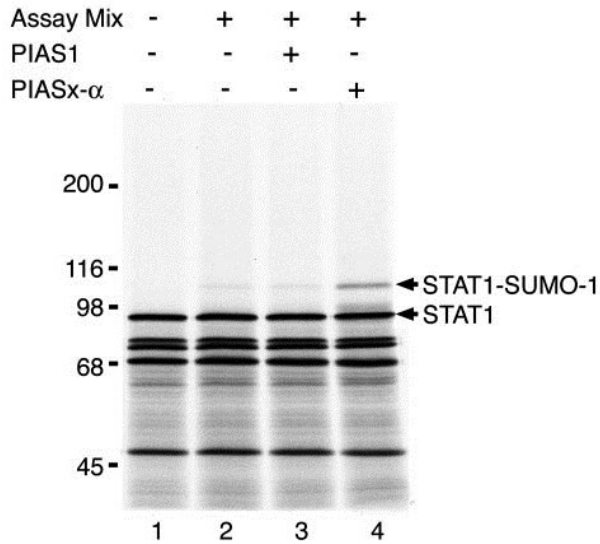


Figure 3-2. PIASx- α specific SUMOylation of STAT1 under limiting concentrations of E1 and E2. Adapted from Rogers et al 2003, [1].

[³⁵S]methionine-labeled STAT1 was transcribed and translated in rabbit reticulocyte lysate (lane 1). The substrate STAT1 was SUMO modified under limiting concentrations of E1 and E2 (lane 2). PIAS1 and PIASx- α were added to assays E3 ligase activity and specificity (lanes 3 and 4). PIASx- α , but not PIAS1, specifically enhanced modification of STAT1 [1].

Construction of SUMOylation Pilot Chip and Assay Optimization

To optimize reaction conditions for On-Chip SUMOylation a pilot chip was constructed. We selected 66 previously known SUMO substrates, 12 proteins involved in SUMO conjugation, and the four SUMO isoforms. These proteins were purified as GST fusions from yeast and printed on the FullMoon glass surface (Figure 3-3).

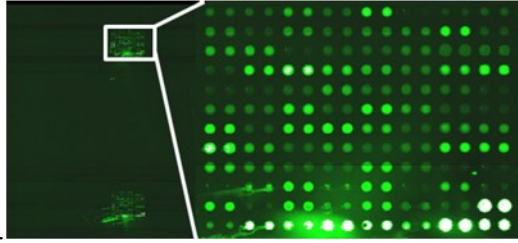


Figure 3-3. SUMO pilot chip

Anti-GST image showing 82 SUMO substrates and conjugation enzymes printed in duplicate.

Because many targets are SUMOylated *in vitro* with E1/E2, we intended to use the high concentrations of E1 and E2 as a positive control for baseline SUMOylation. To achieve this, we titrated the E1 and E2 enzymes with the standard solution-based concentrations as our “1X” condition [1]. Concentrations ranging from 5X to 0.001X E1 and E2 showed a dynamic range of SUMOylation on our pilot array. Subsequent batches of purified E1 and E2 yielded different activity and thus slightly different amounts of enzymes to achieve the same results in our assays. Irrespective of the activity of the enzymes, a 50 fold difference in concentrations of E1 and E2 mixture was found to be optimal for high and low E1 and E2 controls. At the 1X condition well characterized substrates and SUMO enzymes like RanGAP1, NME1, RanBP2 and Uba2 are modified by SUMO1. At 5X additional targets such as BLM, H3, H4, H2B, and RUVBL2 appear. At the highest concentration, 10X, AKT, PCNA, SENP2, MGMT become modified (Figure 3-4). At the lowest concentration, 0.001X E1/E2 RanGAP1 is no longer modified. A common concern with protein microarray assays is whether the amount of protein printed on the

chip correlates with the targets modification (not shown). We performed a regression analysis comparing the abundance of the protein printed on the chip, using anti-GST signal as a proxy, to the amount of SUMOylation at 10X with anti-SUMO antibody. The resulting R^2 value is 0.3827 indicating there is no relationship between the amount of protein printed and the SUMOylation of a target. This provided confidence that our assay is sensitive and specific.

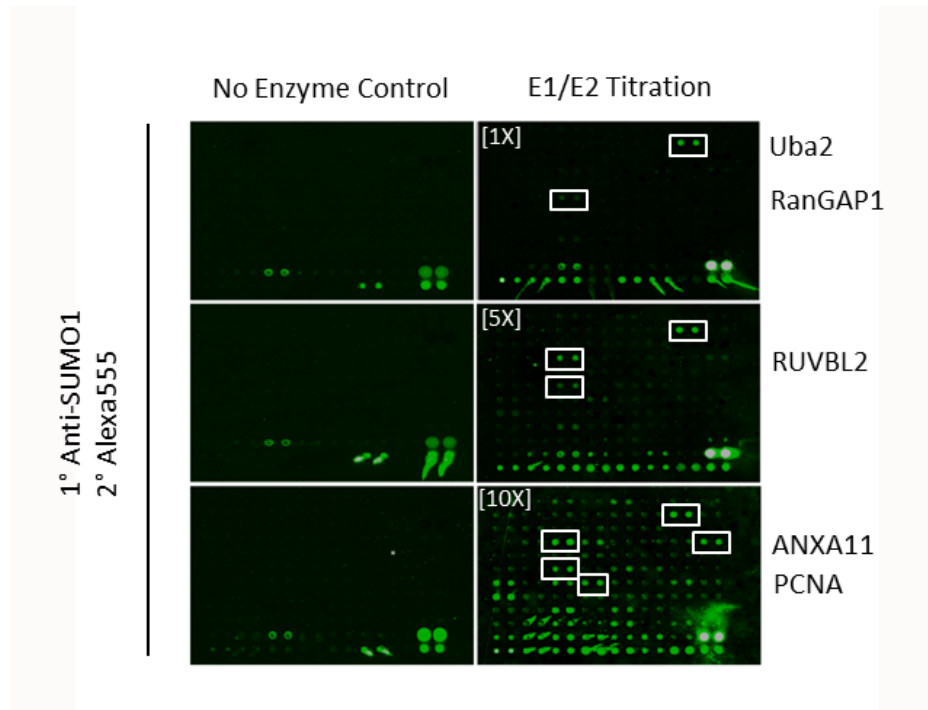


Figure 3-4. E1 and E2 Enzyme titration on pilot chip

The pilot chip was incubated with increasing concentrations of E1 and E2 enzymes to determine the appropriate concentration for high and low concentration controls for the human proteome global screening assays. At 1X only Uba2 and RanGAP1 are SUMOylated, at 5X RUVBL2, SUMO4, H2B, H3, H4 become modified, at 10X many

additional targets are modified including SAE1, NAP1L1, GTF2IRD2, p53, PCNA, and ANXA11.

Antibody based detection of SUMOylation results in high background on the glass surface. Directly labeling SUMO with either Alexa555 or 647 allows for detection of directly labeled substrates, eliminates the need for numerous antibody amplification steps, reduces the duration of the experiment, and dramatically improves the foreground/background signal which boosts the quality of our analysis. Alexa-555 labeled SUMO, gave the better fluorescent signal than Alexa-647, thus we employed Alexa-555 SUMO for all future chip-based experiments.

Purification of Active E3 ligases

At the outset of this study, 11 proteins had been reported to show SUMO E3 ligase activity: PIAS1-4, RanBP2, TOPORS, Pc2, MUL1, Rhes, HSP27, MMS21 [15-19]. The quality of the evidence and the number of reported substrates varied. In an attempt to carry out the most comprehensive study possible, we purified all of the reported E3s from bacteria. In addition to wild type E3 ligases, we purified a mutant form of PIAS3 with two point mutations in the RING domain. Using a rabbit reticulocyte *in vitro* transcription and translation system, we expressed the reported substrates of the E3s with [³⁵S] labeling. To determine if the proteins showed true E3 ligase activity against the substrate we carried about the gold standard *in vitro* assay as described in Rogers et al 2003 [1]. We were able to confirm E3 ligase activity for all of the PIAS proteins, RanBP2, and TOPORS. For those whose activity was not confirmed, numerous attempts

at purification and activity testing were made. The PIAS3 Δ SUMO RING domain mutant was deficient at modifying the WT PIAS3 target NR2E3 (Figure 3-5).

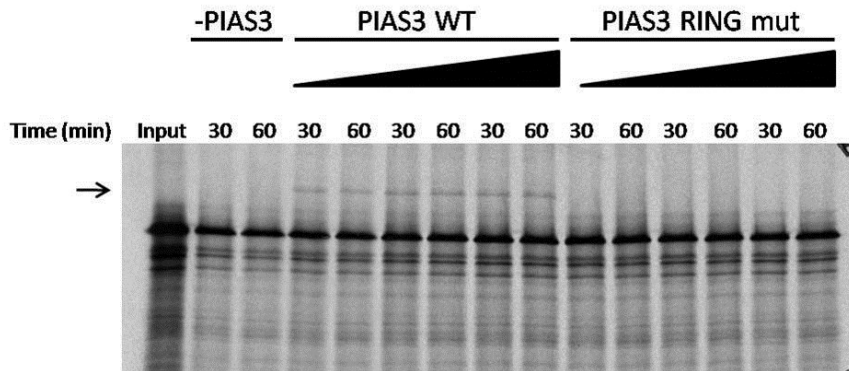


Figure 3-5. PIAS3 enhances SUMOylation of NR2E3 *in vitro* and requires intact RING domain

[³⁵S] methionine-labeled NR2E3 was transcribed and translated in rabbit reticulocyte lysate. The substrate NR2E3 was SUMO modified under limiting concentrations of E1 and E2. PIAS 3 and PIAS3 RING mutant were added to assay E3 ligase activity and specificity. PIAS3, but not PIAS3 RING mutant, specifically enhanced modification of NR23E.

E3 Ligase Substrate Discovery on Pilot Array

Next, the E3 ligases that passed the solution based activity test were applied to the pilot chip in a reaction mixture containing low concentrations of E1 and E2. On the pilot chip, the low concentration of E1 and E2 SUMOylated only few a targets including Uba2,

RanGAP1, PIAS1, RanBP2, and Bcl-2. By increasing the concentration of E1 and E2 50-fold, several more targets were SUMO modified: BLM, GFT2IRD2, SUMO1, SUMO2, PPID, H3, and PTEN. With the addition of each E3 ligase additional targets were SUMOylated. In many cases the E3 ligases were able to rescue the SUMOylation of substrates that were modified with high concentrations of E1 and E2 enzymes. Of particular interest are the targets that were not modified by high concentrations of E1 and E2 but were efficiently SUMOylated with the presence of an E3 ligase. These substrates are what we consider E3 ligase dependent SUMOylation substrates (e.g AKT1 and SMAD4) (Figure 3-6).

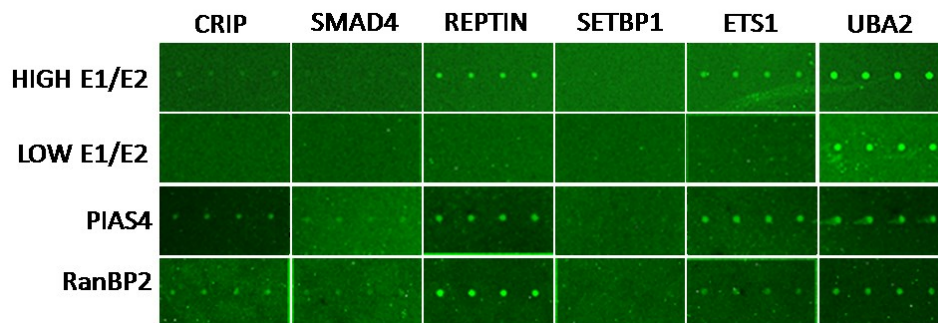


Figure 3-6. E3 ligase SUMOylation activity and specificity on pilot array

Substrate proteins were spotted on the array in quadruplet. In parallel assays, SUMOylation of these targets was assayed by incubation with high concentrations of E1 and E2 and low (limiting) concentrations of E1 and E2. E3 ligases were added to the low concentration E1 and E2 mixture to determine ligase activity and specificity. PIAS4 and RanBP2 activity is pictured. E3 ligases were able to rescue SUMOylation of targets modified under high E1 and E2 conditions as well as modified E3-dependent targets that were not modified by high concentrations of E1 and E2 alone.

E3 Ligase Substrate Discovery utilizing Human Proteome Microarray

E3 ligases with robust, reproducible activity are essential for pursuing an *in vitro* global SUMO substrate screen. Based on the results from the pilot chip this assay has the power to identify the E3 ligase-dependent SUMO substrates for PIAS1-4, RanBP2, and TOPORS. We will have the ability to readily distinguish unique substrates from shared, those preferentially modified by SUMO1 or SUMO2, and the overall SUMO paralog preference of each E3 ligase. The substrate specificity of each E3 ligase will finally be queried in a comprehensive systematic manner.

The Human Proteome Microarray, developed in our lab contains over 17,000 human proteins. The ORFs in this collection are derived from the Invitrogen Ultimate ORF collection. The inserts from ORF collection entry clones were shuttled into pEGHA yeast expression vector using LR Clonase recombination. Successful subcloning was verified by restriction digest. The plasmids were then transformed into Y258 yeast cells for subsequent expression and purification. The proteins were purified as GST fusions and printed onto the Fullmoon glass surface (Figure 1-2). The relative abundance, spot morphology, and overall quality of the array was monitored by anti-GST antibody followed by a Cy3-secondary antibody. The signal intensity of each spot was analyzed with GenePix Pro software (Figure 1-2).

Using recombinant purified E1, E2, E3 enzymes, and Cy3-labeled SUMO, we set up a systematic global E3 ligase substrate screen using the Human Proteome chip. Chips incubated with high concentrations of E1 and E2, and SUMO1 or SUMO2, were used as a positive control for SUMOylation. To compare the activities of the six active E3 ligases, low concentrations of E1 and E2, and SUMO1 or SUMO2, were spiked with each E3 ligase. In a parallel reaction, E3s were omitted as a control for ligase activity. Chips incubated with reaction buffer and SUMO1 or SUMO2 served as negative controls. Including all experimental permutations, a total of 18 reactions conditions were carried out. All reactions were performed in triplicate to ensure a high quality, reproducible dataset. The Human Proteome chips were blocked overnight at 4°C, the appropriate reaction mixture was added to the surface of the chip for 90 minutes at 37°C. Following the reaction, the chips were washed under denaturing conditions to remove residual enzymes as well as SUMO that was not covalently incorporated during the reaction (Figure 3-8). Again GenePix Pro 6.0 was used to scan and analyze the signal intensity of the spots. All experiments, with the exception of PIAS2/SUMO2 were successful. The signal produced by the PIAS2/SUMO2 assays, in two of three attempts, resulted in a fully saturated image that was uninterpretable by eye and could not be analyzed by GenePix Pro.

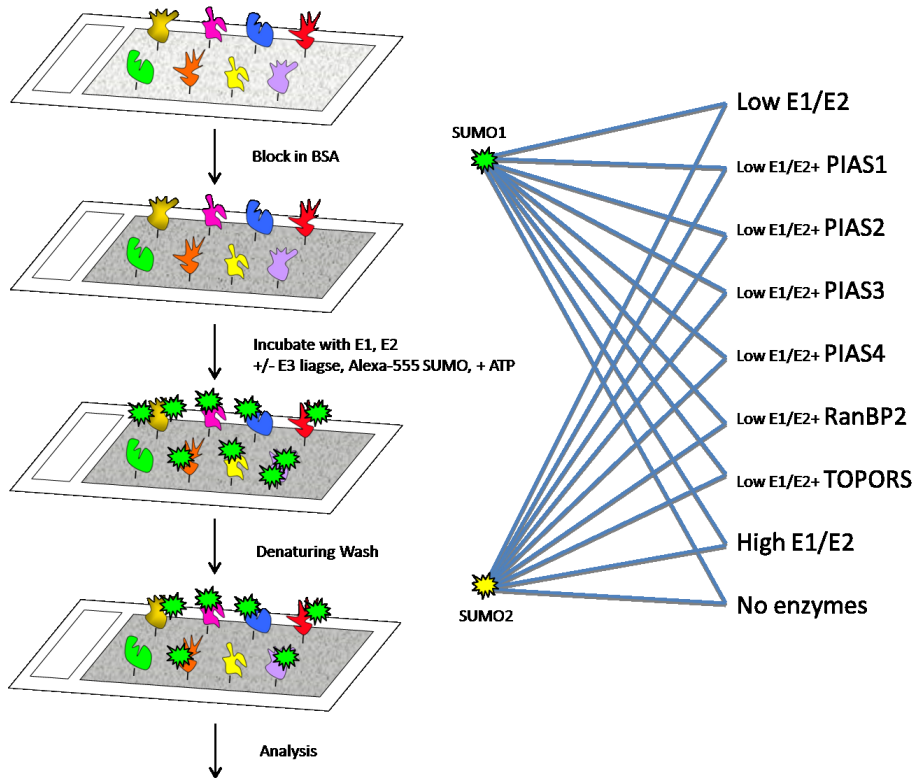


Figure 3-7. Study design and assay schematic

Protein microarrays were blocked overnight in 2% BSA, then incubated with low concentrations of E1 and E2, plus PIAS1-4, RanBP2, TOPORS, high concentrations of E1 and E2, and no enzyme controls. Assays were performed with alexa-555 SUMO1 and SUMO2. Following the reaction chips were scanned and analyzed with the GenePix Pro system.

3.2. Analysis of Raw Data

The images were analyzed with GenePix Pro software. The foreground to background intensities ($S=F/B$) was used to generate a score for each spot. Given that the majority of

the proteins on the array are not substrates, the distribution of the scores resembles a normal distribution centered around the value of 1. Standard deviation of 5 was assigned as the stringent cut-off value. Spots are printed on the chip in the duplicate, and the experiments were performed in triplicate. If a duplicate spot, showed signal intensity greater than SD5 and was reproduced on 2 of the 3 chips, it is considered a positive hit in our dataset. Because the data analysis criteria requires that a spot is reproduced in at least 2 of 3 experiments, by definition the PIAS2/SUMO2 assay did not result in any positive hits. No spots were scored as positive in the negative control and only two proteins appeared in the low concentration E1 and E2 control, SAE2 and RanGAP1. This is not surprising as RanGAP1 remains the most efficiently SUMO modified protein in human cells. Furthermore, SAE2 is one protein in the E1 heterodimer that covalently attaches to SUMO. This indicates that any signal present in the reactions with an E3 is attributable to the ligase activity of the protein.

3.3. Results of Global SUMOylation Data Analysis

The Human Proteome chip contains over 17,000 unique proteins that served as potential SUMO substrates in our assays. The high concentration of E1 and E2 resulted in 2346 hits with SUMO1 and 1933 hits with SUMO2, indicating about 13% of proteins on our chip have the potential to be SUMOylated *in vitro*. The addition of PIAS1 to the low concentration of E1 and E2 showed strong activity as 767 proteins were SUMOylated with SUMO1 and 853 proteins with SUMO2. As mentioned above the PIAS2/SUMO2 assays were not included in our final data set due to experimental challenges, however, PIAS2 with SUMO1 SUMOylated 329 proteins. In parallel assays, PIAS3 SUMOylated only 70 proteins with SUMO1, yet modified 1092 with SUMO2. Similarly, PIAS4 only

SUMOylated 3 proteins with SUMO1 while 249 were hit with SUMO2. The IR fragment of RanBP2 SUMOylated 181 proteins with SUMO1 and 99 with SUMO2. Finally, TOPORS stimulated SUMOylation of 197 proteins with SUMO1 and 73 with SUMO2 (Table 3-1). Initial observations indicate E3 ligases do show preference for one SUMO isoform over another (*e.g.* PIAS3, PIAS4, TOPORS). The number of proteins modified per assay spans a large range from only three proteins modified by PIAS4 SUMO1 up to 1092 modified with PIAS3 SUMO2.

By subtracting the hits in the high concentration E1 and E2 experiment from the E3 ligase assays, we can reveal the E3 ligase-specific or -dependent targets. These targets could not be identified *in vitro* without the experimental design used in our assays. It would also be difficult to assign an E3 ligase for the same substrates if they were recovered from mass spectrometry pulldown experiments. Again, there is a large range of specific hits for the E3 ligases in our assays. PIAS3 SUMO2 revealed 478 unique targets whereas TOPORS SUMO2 only modified two proteins specifically (Table 3-1). This is the first time that E3 ligase specific substrates have been identified in a systematic manner. We looked for the SUMOylation consensus motif in the subsets of proteins described above and discovered a low occurrence of consensus lysines in our substrates. This is expected for E3 ligase specific SUMOylation although the abundance of the motif is significantly lower than in the SUMO-2 heat shock study described in the introduction (Table 3-1) [11].

The level of redundancy among the substrates of SUMO E3 ligases has been consistently been asked in the literature however, the answer has been elusive until now. In our dataset there is some degree of overlap among the substrates of virtually all of the

pairwise E3 ligase comparisons. For example, PIAS4 SUMO2 and PIAS2 SUMO1 share 45.2% of their substrates, PIAS1 SUMO1 and PIAS2 SUMO1 share 13.4% of their substrates, and RanBP2 SUMO2 and PIAS3 SUMO2 share only 3.7 % of their substrates. These numbers suggest that while there is redundancy there is also a great deal of specificity for SUMOylation of substrates. Over 1000 of the substrates recovered in our assays only occurred in one unique enzymatic reaction out of the fourteen that were carried out. To visually capture the substrate specificity in the SUMOylation network, a figure was generated where the nodes representing the SUMO substrate proteins are connected to the E3 ligases that they are modified by (Figure 3-8). With this analysis one can visualize how E3 ligase specificity is organized in the SUMOylation network.

Table 3-1. Summary of Global E3 ligase SUMOylation utilizing the Human Proteome microarray

	Total	Unique	ΨKXD/E motifs	SIMs	Phospho-protein	Kinases	ZINC fingers
High E1/E2 S1	2346	---	194	240	1286	81	80
High E1/E2 S2	1933	---	162	205	1073	72	62
PIAS1 SUMO1	767	184	62	105	445	32	39
PIAS1 SUMO2	853	366	74	99	478	34	43
PIAS2 SUMO1	329	48	37	44	224	12	15
PIAS3 SUMO1	70	0	2	9	47	4	4
PIAS3 SUMO2	1092	470	96	136	601	32	55
PIAS4 SUMO1	3	0	0	1	3	0	0
PIAS4 SUMO2	249	26	38	37	181	10	11
RANBP2 SUMO1	181	33	8	17	116	7	2
RANBP2 SUMO2	99	7	4	7	51	4	1
TOPORS SUMO1	197	46	13	15	111	6	10
TOPORS SUMO2	73	2	5	8	50	7	3

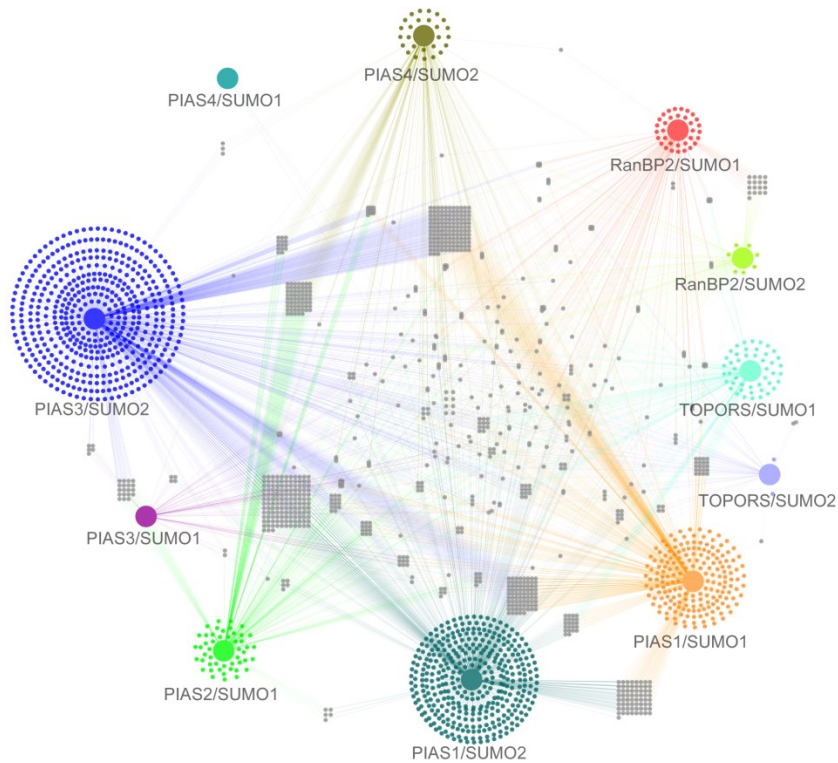


Figure 3-8. E3 ligase SUMOylation substrate specificity network

A network showing the connections between each E3 ligase/SUMO paralog pairing and the modified substrates was generated using Cytoscape. The colored edges depict the connection to upstream E3 ligase. Many substrates are connected to more than more one E3 ligase/SUMO pairing, thus revealing the overlap and redundancy between E3 ligases.

As seen in Figure 3-8, each substrate displays its own unique profile of E3 ligase activity. Some substrates cluster together because they are modified by a similar set of E3 ligases while others lie far apart because of their divergent E3 ligase profiles. By computing a score to represent the degree of similarity between each substrate to all

others, a sophisticated, multidimensional cluster analysis was generated revealing the substrates that share the most similar and dissimilar E3 ligase profiles (Figure 3-9).

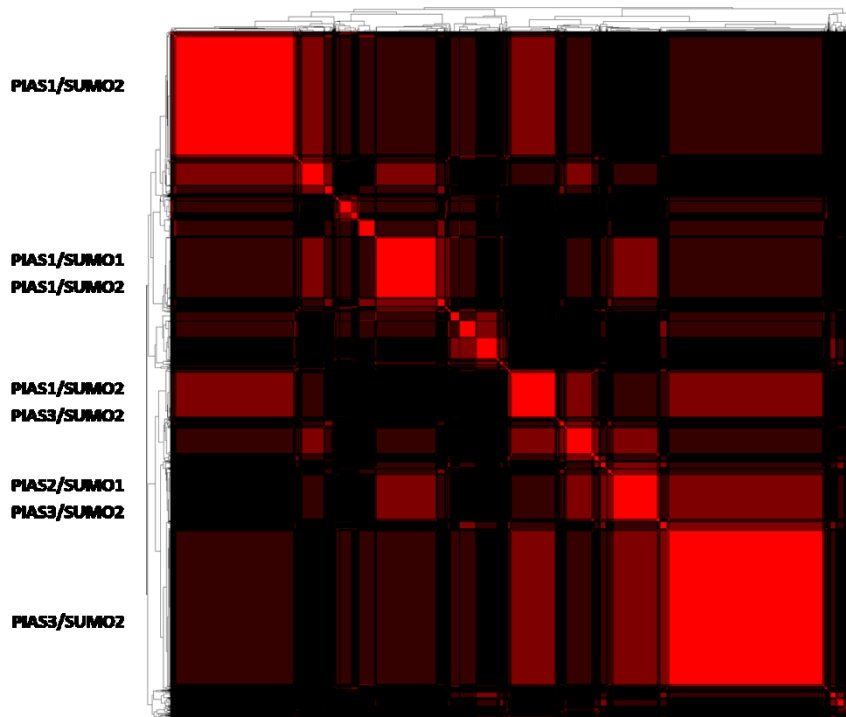


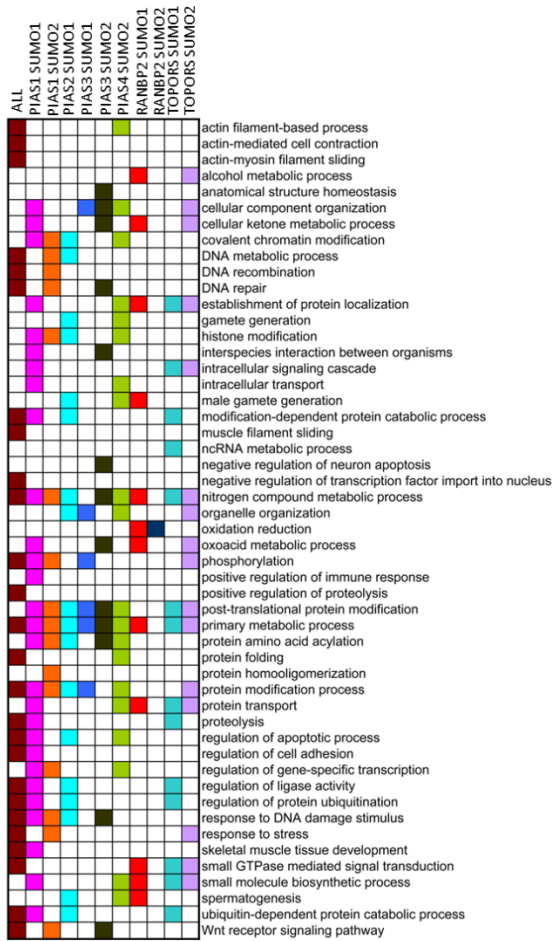
Figure 3-9. Emap clustering of E3 ligase substrates based on similarity of E3 ligase profiles

Multidimensional cluster analysis was performed for each substrate based on the E3 ligase specificity profiles. Target proteins that cluster together are closely connected based on the E3 ligases that target them.

To extract the biologically relevant information from our data set, we performed gene ontology analysis (GO) for enrichment in biological processes and molecular functions,

to provide deeper insight into the functions SUMOylation may serve to regulate. GO analysis was run on the collection of all SUMO substrates, as well as the individual substrate sets for each E3 ligase SUMO pairing. The full set of proteins printed on the chip served as the background for statistical analysis. Several of the significant categories returned were those that SUMO has been determined to play a role in such as response to stress, DNA damage, DNA recombination, protein localization, and protein transport. Of note, many of the significant molecular function terms were related to enzymatic function *e.g.* kinase activity, MAP kinase activity, GTP binding, transferase activity, hydrolase activity (Figure 3-10). Presently SUMOylation is not known to regulate families of enzymes thus we were particularly interested in the role of SUMOylation in kinase activity. In the collective data set, 129 proteins were annotated under the molecular function kinase activity, with a p value of 3.8E-04. Substrates modified by PIAS1 with SUMO1 and SUMO2, RanBP2 SUMO1, and TOPORS with SUMO1 and SUMO2 were also significantly enriched for kinase activity. Furthermore, MAP kinase activity was also significantly enriched in the collective data set with a resultant *p*-value of 1.5E-04.

Biological Process



Molecular Function

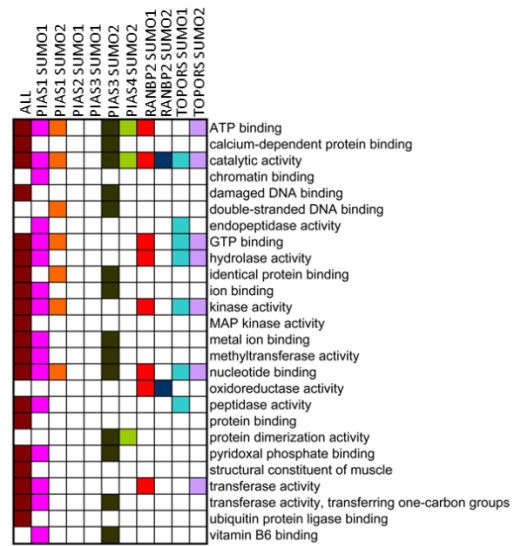


Figure 3-10. Gene Ontology analysis of significant biological processes and molecular functions

Overrepresented biological process and molecular function GO categories were identified. The composite dataset and each E3 ligase/SUMO pairing were analyzed separately.

Phosphorylation is also a GO enriched biological process within SUMOylation targets, pointing to potential interplay between the two PTMs on a systems level. Our lab has also

pursued extensive kinome studies in which the substrates of each kinase were identified using protein microarrays [20]. Combining the kinase substrate relationship (KSR) dataset with the SUMOylation data, there are over 1200 mutual substrates of the two PTMs. Examples of crosstalk between phosphorylation and SUMOylation have been described and we would like to explore this phenomenon on a systems level using protein microarray technology. Briefly, pathway enrichment analysis was performed using large curated databases such as KEGG, Biocarta, and Reactome as sources. Wnt signaling, cell cycle checkpoints, and the proteasome were among the significantly enriched pathways.

As kinases are a well characterized and extremely important, diverse family of proteins, we were interested to see if SUMOylation was enriched among kinase subfamilies. We constructed a phylogenetic tree consisting of the kinases printed on the Human Proteome chip and calculated the significance of the enrichment of SUMOylation of the kinases in each of the seven major groups (Figure 3-11). Though kinases in all groups were targets of SUMOylation in our screen, the CMGC group showed the strongest enrichment with 19 of 49 members (p -value 4.62E-08). The MAP kinase family is included within the CMGC family which provides an additional layer of confidence for the likely biological relevance of SUMOylation in the MAP kinase family.

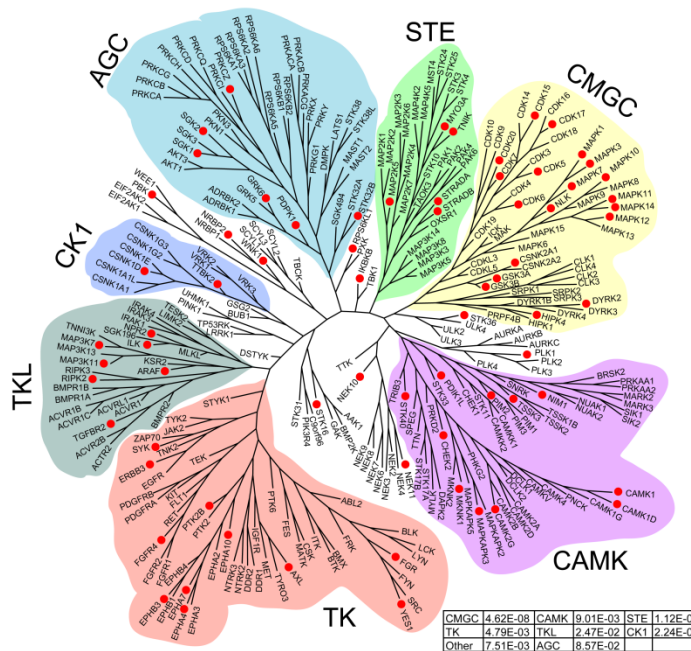


Figure 3-11. Phylogenetic kinase tree overlaid with SUMOylation enrichment

The amino acid sequences of kinase domain of all human kinase proteins have been annotated by Manning et al with Hidden Markov Model [21]. We collected these sequences from kinase.com and built the phylogenetic tree by Mega 5 [22]. We marked different kinase family by distinct color shadow and marked the SUMOylated kinases with a red circle. The enrichment analysis by cumulative hypergeometric method indicated that SUMOylation kinases are very significantly enriched in CMGC family (p -value 4.64E-8).

To establish further connections with SUMOylation and kinase function we used a protein-protein interaction (PPI) database and KSR information to generate an interaction and directional phosphorylation web of kinases that are SUMOylated in our dataset (Figure 3-12). This type of integrative analysis may reveal kinases that are regulated by

SUMOylation or those that are interaction hubs. ERK2 showed the most edges for both PPIs and KSRs totaling 14.

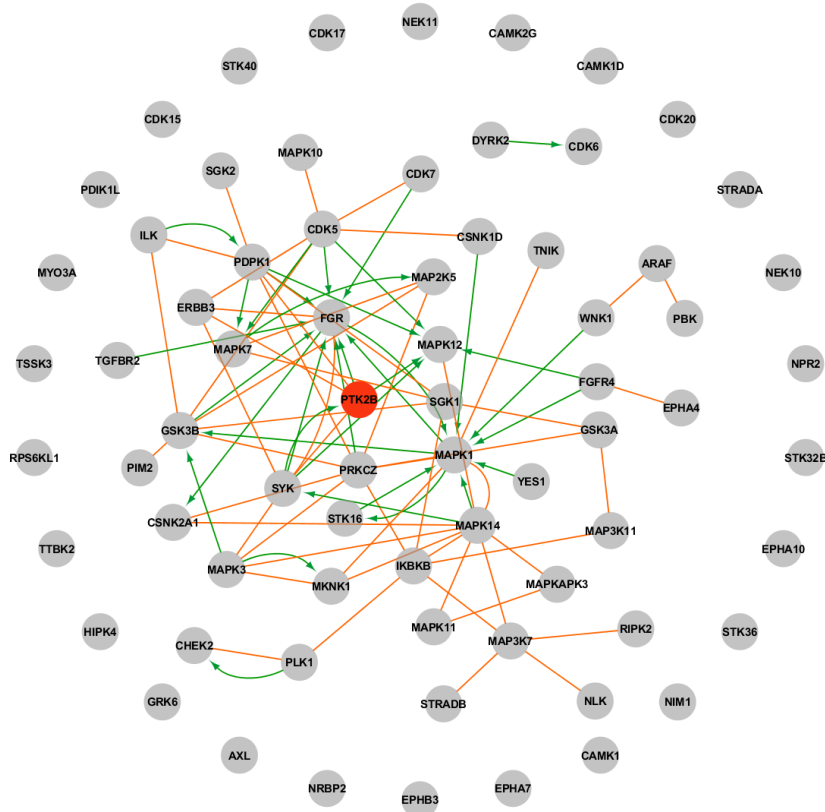


Figure 3-12. Kinase SUMOylation substrates in PPI and KSR network

Protein-protein interaction analysis was performed on the subset of kinases identified as SUMO substrates in our array-based assays (orange edges). Kinase substrate relationships (KSRs) are represented by green directional arrows.

3.4. Validation of MAP kinases as targets of SUMOylation

To investigate the fidelity of the Human Proteome microarray SUMOylation reactions we selected several SUMO substrates for validation in HeLa cells. Based on GO analysis and phylogenetic enrichment the MAP kinase family appears heavily implicated in SUMOylation. We intended to validate E3 ligase specificity for the substrates MAPK3/ERK1, MAPK10/JNK3, MAPK11/P38 β , MAPK12/P38 γ , and MAPK14/P38 α . The MAP kinase family is an ideal validation set due the varied E3 ligase profiles for the substrates (Figure 3-13). Because all six E3s show activity with at least one kinase yet each substrate's E3 ligase profile is unique, we will be able to address accuracy of our assays for individual substrates and the kinase family.

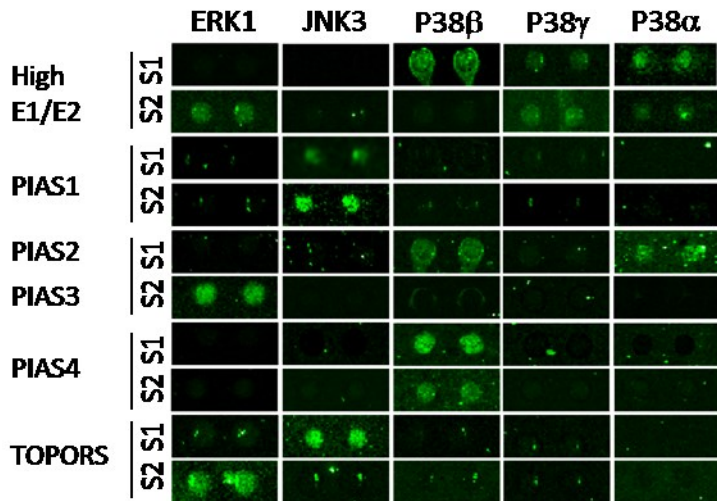


Figure 3-13. E3 ligase substrate specificity in the MAP kinase family

Images from each SUMOylation experiment on the Human Proteome chip were assembled to depict the enrichment and E3 specificity of MAP kinases as SUMO substrates.

Candidate kinase substrates were subcloned into a V5 mammalian expression vector and co-transfected with MYC-tagged SUMO, plus and minus the appropriate E3 ligase. SUMOylation status was assessed by V5 immunoprecipitation followed by western blotting for MYC-tagged SUMO. PIAS directed SUMOylation was examined with the following pairings: JNK3 modified by PIAS1 with SUMO1 and SUMO2, P38 α modified by PIAS2 with SUMO1 and SUMO2, and ERK1 modified by PIAS3 with SUMO1 and SUMO2. Modification of ARAF kinase by PIAS1 was also examined in our validation studies. ARAF is a member of the MAP kinase family however it is a MAPKKK, two tiers above the MAPKs. All of the MAPK validation experiments returned positive results that confirmed the data collected in the protein microarray assays (Figure 3-14). To evaluate true E3 ligase specificity, P38 β was co-transfected with PIAS1-4 and TOPORS, and SUMO1 or SUMO2 (Figure 3-15).

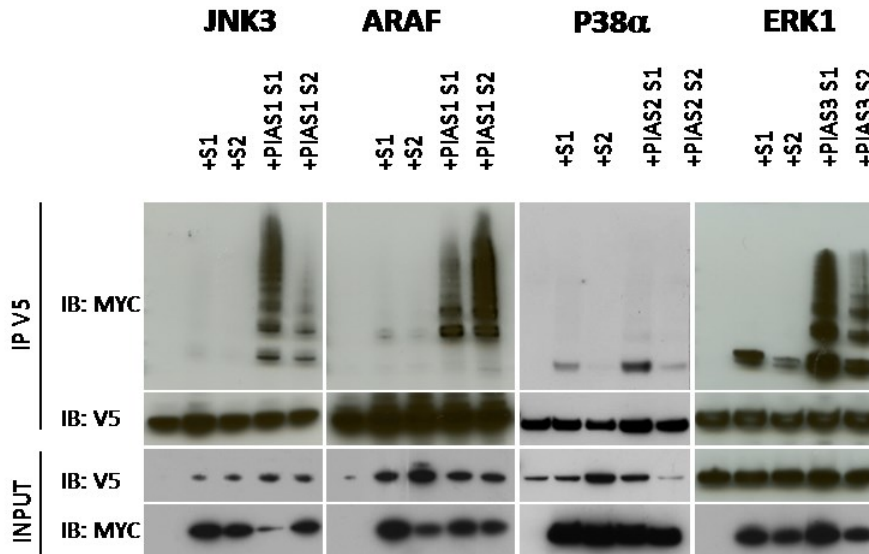


Figure 3-14. PIAS directed SUMOylation of the MAP kinase family

Human protein array E3 ligase specificity is reproduced in HeLa cell-based validation studies with several MAP kinase candidate substrates. HeLa cells were transfected with V5-tagged constructs of JNK3, ARAF, P38, ERK1, along with MYC-SUMO1 or SUMO2, and FLAG-tagged E3 ligases as indicated. V5-tagged MAP kinases were immunoprecipitated and SUMOylation was assessed by immunoblotting with an anti-MYC antibody.

These results also confirmed our chip data with only minor variations. *In vitro* evidence showed P38 β can be modified by SUMO1 with E1 and E2 alone, which was upheld in the cell based assays. The protein microarray assays with PIAS2 and SUMO2 was not included in our screen thus we did not have a prediction for the cell based validation assays. In HeLa cells PIAS2 modified P38 β with both SUMO1 and SUMO2, though stronger modification was observed with SUMO1. *In vitro* PIAS4 modified P38 β with both SUMO1 and SUMO2 although only modification with SUMO1 was detected in

HeLa cells (Figure 3-15). The activity of SENPs cannot be predicted based on our *in vitro* experiments and thus we cannot rule the possibility that there is differential protease activity for each SUMO isoform. The opposing and activity of E3 ligases and SENPs may result in different outcomes in cell-based assays versus *in vitro* protein microarray assays.

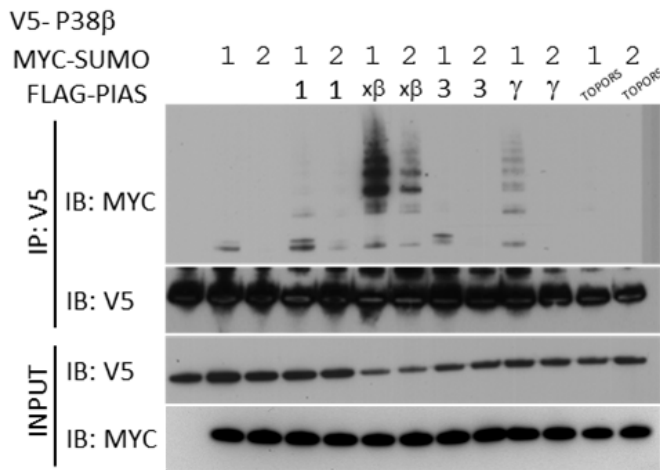


Figure 3-15. E3 ligase specificity for MAPK11/P38β

MAPK11/ P38β E3 ligase SUMOylation specificity is reproduced in HeLa cell-based validation using PIAS1-4 and TOPORS.

Discussion

The majority of systematic SUMOylation studies employ affinity pull downs coupled to mass spectrometry. This is an ideal method to identify substrates and attachment sites at a specific biological time point, however such studies are clearly lacking information indicating how and why the modifications occurred. Although previous studies have

identified large numbers of SUMOylated proteins, only a small number of substrates have been linked to ligases. Using recombinant E3 ligases in conjunction with the richness of the Human Proteome microarray our assays have the power to reveal a global SUMO substrate network and enriched downstream processes it may regulate. Our study yielded a comprehensive data set that integrates both E3 ligase and SUMO paralog specificity, enabling us to define the role of E3 ligases in substrate selection and SUMO paralog preference at a proteomics level.

While much of the evidence for E3 activity seems merely anecdotal many scientists in the field have used hypothetical E3 function to explain how specificity is coordinated in the SUMO system. Here, we were able to generate data to explain the numerous roles E3 ligases play in regulating SUMO substrate selection. The assay design creates a straightforward method for revealing the substrates that are shared or redundant (modified in more than one condition), the E3 ligase specific substrates (only modified in the presence of one E3), and those that do not show E3 specificity (only modified by high E1 and E2). We successfully assigned thousands of SUMO substrates to E3s and developed high-confidence, detailed substrate profiles for each ligase. Even on the surface, we can note vivid distinctions in the E3 ligase properties. Perhaps the most obvious is great variation in the number of substrates that E3s modified. A unifying feature is that they all modify a subset of substrates specifically. Surprisingly, PIAS3, PIAS4, and TOPORS demonstrated dramatic SUMO isoform preference for conjugation. This also gave us confidence that our assay was not biased towards SUMO1 or SUMO2.

The SUMO isoforms have been shown to have roles in different biological processes and thus we expected that the global level of redundancy would be moderate. In most cases

the percent of substrates modified by SUMO1 and SUMO2 for a particular E3 ligase was below 10%. It has long been speculated that E3 ligases are responsible for directing SUMOylation specificity two-fold: by selecting the substrate and discriminating between SUMO paralogs in conjugation. We observed that substrates which were only modified by one SUMO paralog in the high E1 and E2 condition were modified by SUMO1 and SUMO2 in the reactions with an E3 ligase. For example, IKBKB, is only modified by SUMO2 under high E1 and E2 reaction conditions whereas PIAS1 and PIAS3 are able to stimulate SUMOylation with SUMO1 and PIAS1 and TOPOR2 cause SUMOylation with SUMO2. In a separate case, the substrate RNF5 is only targeted by SUMO1 under high E1 and E2 conditions while it can be modified by SUMO2 in the presence of PIAS1 and PIAS3. E3 ligases have the ability to mediate differential conjugation of the SUMO isoforms for hundreds of proteins in our screen.

MAP kinase validation studies in HeLa cells demonstrated the validity of the E3 ligase substrate specificity on the human proteome array as it was reproduced in cells. We are not surprised that SUMO paralog specificity was not consistently upheld from *in vitro* to cell based assay because paralog preference specificity cannot be faithfully reproduced *in vitro* due to many confounding factors such as SENPs [2].

Because we do not know the sites of attachment for our SUMO substrates we can only make correlative connections to SUMO consensus motif conjugation. Between 2.8-15.26 percent of proteins recovered in our assays contained the Ψ KX(D/E) motif. This is significantly lower than the predictions in an earlier large scale study where 74.3 percent (564 of 759) of targets contained the motif [10]. There are major differences between our study design and those previously published, which may explain the drastic differences.

E3 ligases are often said to direct SUMOylation of non-consensus site lysines which is in accordance with the low abundance of motifs. Our study identified more than three times the number of hits as the Golebiowski and can thus shift the percentage of predicted sites because the sample size is so much greater. Additionally, their study was executed with a SILAC/mass spectrometry and exclusively focused on SUMO2 targets. Overall our study suggests that E3 ligases largely direct non-consensus lysine SUMOylation.

E3 ligases coordinate modification of specific substrates likely for explicit biological purposes. Our gene ontology analysis revealed enrichment in many novel biological processes and molecular functions for E3 ligase directed SUMOylation. Processes such as small GTPase signaling, phosphorylation, ligase activity, Wnt receptor signaling, and protein folding are new and exciting areas for SUMOylation function. The molecular function analysis uncovered enriched categories related to ATP binding, catalytic activity, kinase activity, and transferase activity. In combination with the gene ontology results from other SUMO proteomics studies we are now building an extensive global function map of processes where SUMOylation is enriched. This marks the first E3 ligase directed SUMOylation study where substrate specificity is globally analyzed for discrete functions. It cannot be ignored that our *in vitro* studies cannot provide context for the function of SUMOylation. Candidates must be selected for in depth characterization to confirm the relationship between SUMOylation of substrates in enriched GO categories. We have been able to validate several SUMO substrates in phosphorylation and MAP kinase activity categories in cell based overexpression systems with a 100 percent success rate (6 of 6). Protein microarray based ubiquitylation screens from our lab also yielded successful validation rates [13]. We speculate that is due to the inherent

reduction of false positives because three enzymes are required function together in a cascade to achieve modification, thereby increasing the fidelity of the reaction. The low levels of E1 and E2 in our assays are insufficient to modify proteins on the proteome array and depend on the E3 activity for robust modification. We suspect the mechanism of this reaction and the design of our assay reduces the potential promiscuity of the enzymes. The E3 ligase specificity was confirmed in cell based assays with minor variations in the efficiency of SUMO paralog attachment or discrimination. Validation of SUMO modification in cells is merely the first step in evaluating function of SUMOylation for a substrate. Ideally, lysine attachment site(s) would be identified with SUMO site prediction programs or tandem mass spectrometry. Lysine to arginine mutants serve to delineate the function of the SUMO modified versus unmodified forms of the kinase to understand the physiological importance of the modification.

The enriched phosphorylation and kinase activity categories from the gene ontology analysis suggest possible systems level connections between phosphorylation and SUMOylation. Individual studies of several substrates suggest crosstalk between SUMOylation and phosphorylation may co-regulate protein function. This phenomenon is further supported by a large scale study of SUMO-regulated phosphorylation wherein the authors report that expression of SUMO2/3 in HEK293 cells stimulates an increase in global tyrosine phosphorylation [23]. They identified several proteins with enhanced phosphorylation and linked them together as substrates of one kinase, CAMKII, which upon SUMOylation experiences increased activity. Chapter 4 further explores the theme of SUMO regulation of kinase activity for the tyrosine kinase Pyk2.

Material and Methods

Protein Purification and Pilot chip fabrication

SUMO substrate ORFs were expressed as GST fusion proteins in yeast. Cultures (6 ml) were grown at 30°C to an optical density at 600 nm of 0.7 to 0.9 and induced with 2% galactose for 4 to 6 h. Harvested cells were lysed with glass beads in lysis buffer (100 mM Tris-HCl (pH 7.4), 100 mM NaCl, 1 mM EGTA, 0.1% 2-mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 0.1% Triton X-100 plus protease inhibitor cocktail (Roche). GST fusion proteins were bound to glutathione beads (GE Healthcare) for 1 h at 4°C and washed three times with wash buffer I (50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM EGTA, 0.1% Triton X-100, 0.1% β -mercaptoethanol, and 0.5 mM PMSF) and three times with wash buffer II (50 mM HEPES, pH 7.4, 100 mM NaCl, 1 mM EGTA, 10% glycerol, 0.1% β -mercaptoethanol, and 0.5 mM PMSF) and eluted by glutathione competition elution buffer (100 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM MgCl₂, 40 mM glutathione, and 30% glycerol). The eluate was collected through a filter unit and stored in a 384-well plate. Sixty-six SUMO substrate proteins were successfully purified based on anti-GST antibody probing. The purified substrates along with 16 recombinant control proteins, acquired through generous contributions, were printed in duplicate on modified glass (Full Moon Biosystems) microscope slides using a 48-pin contact printer (Bio-Rad) employing four pins.

Protein Purification

E3 ligase purification: Full length PIAS1, PIAS3 and PIAS3 Δ SUMO were expressed in bacterial as GST fusions in the pDEST15 bacterial expression vector. PIASx β and PIAS γ

were expressed in bacteria in the pQLink 6XHis plasmid. Fragments of TOPORS (268-644) and the RanBP2 IR region were subcloned and expressed in pDEST15. All constructs were expressed in *Escherichia coli* and purified with Glutathione Sepharose 4B (GE Healthcare) or Nickel NTA agarose (Qiagen).

Solution-based in vitro E3 ligase SUMOylation assays

Substrates were produced by *in vitro* transcription and translation in rabbit reticulocyte lysate in the presence of [³⁵S] methionine according to the manufacturer's instructions (Promega). 2 ul of translation product was added to a 40 ul reaction. *In vitro* SUMO modification of reported E3 ligase substrates was performed using two individual assay conditions similar to that described in Rogers et al 2003. Assays performed using high concentrations of SAE1/UBA2 (E1) and Ubc9 (E2) contained 200 nM E1 enzyme, 600 nM E2 enzyme, 1.0 uM SUMO1 or SUMO2, 1 mM ATP, 20 units/ml creatine phosphokinase, 5 mM phosphocreatine, 0.6 mg/ml inorganic pyrophosphatase, 20 mM KEPES-KOH, pH 7.3, 110 mM potassium acetate, 2 mM magnesium acetate, and 1 mM dithiothreitol. Assays performed using low concentrations of E1 and E2 contained 35 nM E1 and 14 nM E2. E3s were supplemented at 5-20 nM. The reactions were incubated at 37°C for 1 hour then quenched by addition of SDS-PAGE sample buffer and analyzed by SDS-PAGE and autoradiography.

Protein Microarray SUMOylation

Protein chips were incubated overnight in blocking buffer containing 0.2% bovine serum albumin (BSA) and 0.05% Tween-20 at 4°C. SUMOylation on the protein chips was performed under two assays conditions. Assays performed using high concentrations of

SAE1/UBA2 (E1) and Ubc9 (E2) contained 2.3 μ M E1, 6.25 μ M E2, 0.7 μ M mature SUMO-Alexa555, 5 mM ATP in 20 mM HEPES, 100 mM NaCl, pH 7.4, 10 mM MgCl₂, 0.1 mM dithiothreitol. E3 ligases SUMOylation assays performed using limiting concentrations of E1 and E2 contained 45 nM E1, 125 nM E2, 0.7 μ M mature SUMO-Alexa555, 5-20 nM E3 enzyme, 5 mM ATP in 20 mM HEPES, pH 7.4, 100 mM NaCl, 10 mM MgCl₂, 0.1 mM dithiothreitol. Assays carried out for each E3 ligase enzyme were performed in triplicate on the same day. The arrays were incubated with approximately 200 μ l of the reaction mixture and covered with LifterSlips (Thermo Scientific) then incubated in a humidity chamber for 90 minutes at 37°C. Arrays were subjected to three 10 minute washes with TBST wash buffer (25 mM Tris-HCl, pH 7.4, 3 mM KCl, 150 mM NaCl, 0.05% Tween-20), three 10 minute washes with 1% SDS solution shaking at 55°C, rinsed briefly with doubly distilled water and spun dry. Control slides were incubated in parallel with SUMO reaction mixture without enzymes.

Protein Microarray data acquisition and Bioinformatic Analysis of E3 ligase Substrates

Protein Microarray Data Analysis

To identify positive hits for each chip first bad spots were removed, such as spots in stained region or spiked with Coefficient of Variation (CV) >1.5. A background correction was applied defined as the Signal Intensity (SI) of each spot as the odds ratio of the foreground median divided by the background median. If a spot does not have signal, its foreground median and background median will be close, thus its SI will be around one. To normalize the signal we suppose the real reaction/signal are rare and

almost evenly dispersed in each block, thus we force each block on a chip to have a median SI of one. To be considered a positive hit, the duplicate spots of each gene must both have signal intensities (Foreground/Background ratio) five standard deviations above the mean. Additionally, positive hits from the triplicates of each enzymatic reaction must be identified by at least two of the three replicates. Positive hits also identified by Negative Control (NC) chips were removed. The names of all genes were checked and non-official gene name were replaced by official gene symbol name.

Prediction of SUMOylation consensus motif

We collected the known sites from small-scale *in vivo* or *in vitro* experiments and used them as foreground dataset [1, 2, 8]. For the known sites shorter than 15 amino acids, we mapped them back to the corresponding protein amino acid sequences and extracted the 15 mers centered on the functional sites. We randomly extracted 100,000 15 mers from the human proteome and took them as background dataset. We then used our M3 algorithm to predict the functional sites on all substrate proteins of each condition [20].

SUMOylation Substrate Network

We used Cytoscape to create the SUMOylation substrate network. The E3 ligase/SUMO pairing and substrates are represented by large filled circles and corresponding small filled circles, respectively. The E3 ligase and ligase-specific substrates are marked by same ligase/SUMO-specific color with substrates circled around the E3 ligase. For example, PIAS3/SUMO2 and its specific substrates are marked by blue. The shared substrates are marked by gray color and connected to corresponding case by case-specific color lines. For example, the substrate sets shared by PIAS3/SUMO2 and PIAS1/SUMO1

are connected to PIAS3/SUMO2 by green colored lines, and connected to PIAS1/SUMO1 by gold colored lines.

Phylogenetic kinase tree overlaid with SUMOylation enrichment

The amino acid sequences of kinase domain of all human kinase proteins have been annotated by Manning et al with Hidden Markov Model [21]. We collected these sequences from kinase.com and built the phylogenetic tree by Mega 5 [22]. We marked different kinase family by distinct color shadow and marked the SUMOylated kinases with a red circle.

Kinase protein-protein and KSR interaction network

71 of the 2150 SUMOylation substrates (not including the substrates of high concentration E1 and E2) are kinases. To analyze the relationship of these 71 SUMOylation kinases, we built their functional relationship network. In this network, two kinases are connected by an un-directed orange line if there is protein-protein interaction (PPI) between them. Two kinases are connected by a directed green line with arrow pointing to the substrate if there is phosphorylation relationship between them. The kinases that do not have phosphorylation or PPI relationships are represented by orphan node.

Validation of candidate substrates in HeLa cells

Candidate substrates were subcloned into the PCAGIG-V5 mammalian expression vector, SUMO1 and SUMO2 were subcloned into PCAGIG-MYC mammalian expression vector, and E3 ligases were subcloned into PSG5-FLAG mammalian

expression vector. V5-substrate and MYC-SUMO were transfected together with and without FLAG-E3 ligases constructs with Fugene 6 transfection reagent (Promega) into HeLa cells seeded at 2×10^5 cells per well. After 48 hours, the cells were washed with phosphate buffered saline and lysed in RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 0.1 mM dithiothreitol, 1% Triton-X-100) containing 0.5 mM PMSF, Roche Protease Inhibitor Cocktail, 20mM *N*-ethylmaleimide and 1% SDS, to inhibit deSUMOylation and dissociate non-covalent protein complexes. After initial lysis in 1% SDS, the lysate was diluted to achieve a 0.2% SDS concentration to accommodate immunoprecipitation. The lysate was incubated with anti-V5- agarose (Sigma-Aldrich) for two hours to immunoprecipitate the substrate. The beads were washed three times with RIPA buffer containing 0.1% SDS then incubated with 2X LDS sample buffer containing β -mercaptoethanol then heated at 100°C for 10 minutes. Samples were resolved in parallel on 4-12% NuPage Bis-Tris gels (Life Technologies) and subject to western blotting using anti-V5-HRP and anti-MYC-HRP antibodies (Life Technologies).

References

- [1] Rogers RS, Horvath CM, Matunis MJ. SUMO modification of STAT1 and its role in PIAS-mediated inhibition of gene activation. *J Biol Chem*. 2003;278:30091-7.
- [2] Johnson ES. Protein modification by SUMO. *Annu Rev Biochem*. 2004;73:355-82.
- [3] Vertegaal AC, Andersen JS, Ogg SC, Hay RT, Mann M, Lamond AI. Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol Cell Proteomics*. 2006;5:2298-310.
- [4] Matunis MJ, Wu J, Blobel G. SUMO-1 Modification and Its Role in Targeting the Ran GTPase-activating Protein, RanGAP1, to the Nuclear Pore Complex. *The Journal of Cell Biology*. 1998;140:499-509.
- [5] Jackson PK. A new RING for SUMO: wrestling transcriptional responses into nuclear bodies with PIAS family E3 SUMO ligases. *Genes Dev*. 2001;15:3053-8.
- [6] Onishi A, Peng GH, Hsu C, Alexis U, Chen S, Blackshaw S. Pias3-dependent SUMOylation directs rod photoreceptor development. *Neuron*. 2009;61:234-46.
- [7] Yunus AA, Lima CD. Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. *Mol Cell*. 2009;35:669-82.
- [8] Saitoh H. Functional Heterogeneity of Small Ubiquitin-related Protein Modifiers SUMO-1 versus SUMO-2/3. *J Biol Chem*. 2000;275:6252-8.
- [9] Oh YH, Hong MY, Jin Z, Lee T, Han MK, Park S, et al. Chip-based analysis of SUMO (small ubiquitin-like modifier) conjugation to a target protein. *Biosens Bioelectron*. 2007;22:1260-7.
- [10] Golebiowski F, Matic I, Tatham MH, Cole C, Yin Y, Nakamura A, et al. System-wide changes to SUMO modifications in response to heat shock. *Sci Signal*. 2009;2:ra24.
- [11] Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, et al. Global analysis of protein phosphorylation in yeast. *Nature*. 2005;438:679-84.

- [12] Lin YY, Lu JY, Zhang J, Walter W, Dang W, Wan J, et al. Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell*. 2009;136:1073-84.
- [13] Lu JY, Lin YY, Qian J, Tao SC, Zhu J, Pickart C, et al. Functional dissection of a HECT ubiquitin E3 ligase. *Mol Cell Proteomics*. 2008;7:35-45.
- [14] V. B-V, D.A. S, M.J. M, C.D. L. Structural Basis for E2-Mediated SUMO Conjugation Revealed by a Complex between Ubiquitin-Conjugating Enzyme Ubc9 and RanGAP1. *Cell*. 2002;108:345-56.
- [15] Weger S, Hammer E, Heilbronn R. Topors acts as a SUMO-1 E3 ligase for p53 in vitro and in vivo. *FEBS Lett*. 2005;579:5007-12.
- [16] D. Wotton JCM. Pc2 and SUMOylation. *Biochem Soc Trans*. 2007;35:1401-4.
- [17] Braschi E, Zunino R, McBride HM. MAPL is a new mitochondrial SUMO E3 ligase that regulates mitochondrial fission. *EMBO Rep*. 2009;10:748-54.
- [18] Subramaniam S, Sixt KM, Barrow R, Snyder SH. Rhes, a striatal specific protein, mediates mutant-huntingtin cytotoxicity. *Science*. 2009;324:1327-30.
- [19] X. Z, G. B. A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proceedings of the National Academy of Sciences* 2005;102:4777-82.
- [20] Newman RH, Hu J, Rho HS, Xie Z, Woodard C, Neiswinger J, et al. Construction of human activity-based phosphorylation networks. *Mol Syst Biol*. 2013;9:655.
- [21] G M, G.D. P, T. H, S. S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci*. 2002;27:514-20.
- [22] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731-9.

[23] Yao Q, Li H, Liu BQ, Huang XY, Guo L. SUMOylation-regulated protein phosphorylation, evidence from quantitative phosphoproteomics analyses. *J Biol Chem.* 2011;286:27342-9

CHAPTER 4

CHARACTERIZATION OF PYK2 SUMOYLATION IN CELL MIGRATION PATHWAYS

Abstract

Global analysis of SUMO E3 ligase substrates revealed kinase activity is an enriched molecular function within the set of SUMO modified proteins identified in our screen. Over 100 kinases were recovered as SUMO substrates in our assays which points to a global phenomenon of kinase regulation by SUMO modification. The studies presented in chapter 4 describe our efforts to validate and characterize SUMO modification of non-receptor tyrosine kinase, Pyk2. Here, we demonstrate that SUMOylation of Pyk2 is a novel PTM that serves to augment intrinsic kinase activity by enhancing autophosphorylation. Further biochemical studies reveal SUMOylation of Pyk2 enhances its association with putative binding protein Src, even in the absence of autophosphorylation. Stimulation of Pyk2 activity by SUMOylation also promotes phosphorylation of downstream focal adhesion protein paxillin. These findings led us to examine the role of SUMOylation of Pyk2 in cell migration. Our results suggest SUMOylation of Pyk2 amplifies its promigratory function in MDA-MB-231 breast cancer cells. Our study describes a novel mechanism for Pyk2 activation regulated by crosstalk between phosphorylation and SUMOylation.

Introduction

Protein microarray SUMOylation assays identified kinases as an enriched protein family that is potentially regulated by the modification. Several kinases have been verified as SUMO targets however, in most cases the modification is not connected to a phenotype. Intracellular signaling was also an enriched biological process. Analysis of canonical signaling pathways highlighted many that were significantly enriched with mediators that are targets of SUMOylation. The MAP kinase pathway and the bioactive peptide induced signaling pathway are both heavily decorated by SUMOylation pointing to a role for SUMOylation in major signaling cascades (Figure 4-1). As we have already validated SUMOylation of many MAP kinases, we desired a kinase further upstream with many substrates, interaction partners, and connections to other SUMOylated kinases.

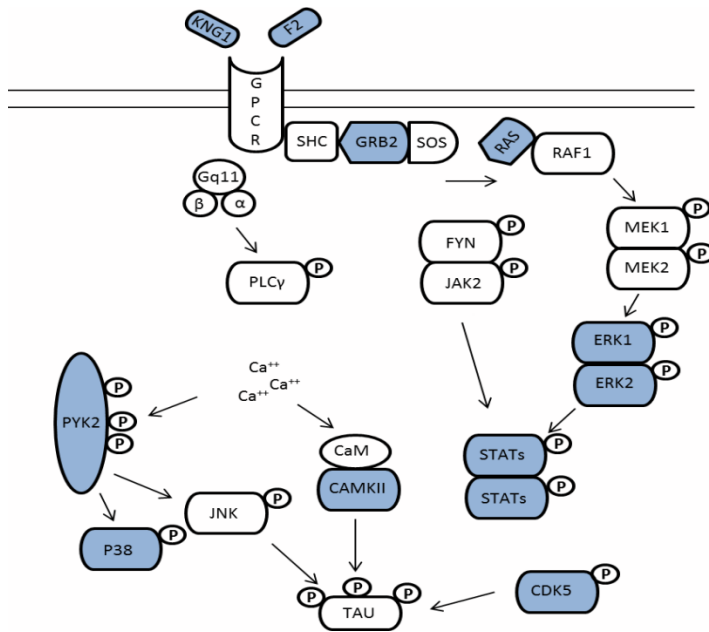


Figure 4-1. Bioactive peptide induced signaling pathway is enriched for SUMOylation substrates.

GPCR pathways exhibit crosstalk with other signaling pathways including tyrosine kinase growth factor receptors and MAP kinase pathways. Transactivation of either receptor tyrosine kinases like the EGF receptor or focal adhesion complexes can stimulate Ras activation through the adaptor proteins Shc, Grb2 and Sos, and downstream MAP kinases activating Erk1 and Erk2. Src kinases may also play an essential intermediary role in the activation of Ras and MAP kinase pathways by GPCRs [1]. Enriched SUMO substrates highlighted in blue; p-value 0.001637.

With this criteria in mind we selected, Pyk2, a member of the focal adhesion kinase (FAK) family. This family consists of FAK (Focal Adhesion Kinase) and Pyk2 (also known as RAFTK, FAK2, CAK- β , or CADTK). Both FAK and Pyk2 are large, 110-125 kDa, non-receptor tyrosine kinases that are similar in their domain structure and amino acid sequences as they share 48% identity and 65% similarity [2]. FAK and Pyk2 are distinct from one another yet they function in many of the same processes, such as integrin-mediated signal transduction pathways, and both kinases localize to focal adhesion-like structures [3]. Integrin mediated adhesion to the extracellular matrix is essential for the regulation of a number of cellular activities including survival, proliferation, and migration. By linking the extracellular matrix to the actin cytoskeleton, integrins provide mechanical basis for transducing productive force required for cell motility. In addition, by sensing the adhesive interaction, biophysical information is

translated into the initiation of cellular signaling pathways through the formation of the focal adhesion complex [4]. Focal adhesions are large multi-protein scaffolds composed of effector proteins that mediate signaling inputs (Figure 4-2). Focal adhesions also serve as a point of convergence for signaling initiated by growth factors and G-protein coupled receptors to integrate information to regulate cell growth and migration. The focal adhesion kinases FAK and Pyk2 function as critical mediators for the activation of signaling pathways that regulate cell migration, proliferation, and survival [5].

The discovery of FAK precedes Pyk2 and thus Pyk2 is frequently compared to FAK as their similarities and difference are characterized. Pyk2 and FAK execute different roles within the same signaling pathways, partially due to unique interaction partners. Further, the two kinases possess different tissue specificity profiles. FAK is quite ubiquitously expressed whereas Pyk2 is more commonly localized to the endothelium, central nervous system, and cells of hematopoietic lineage [6]. As compared to FAK, Pyk2 is only weakly activated in response to fibronectin-integrin binding [7]. Furthermore, in the context of phosphotyrosine activation, Pyk2 is demonstrated to have greater sensitivity to cell depolarization, osmolarity stress, or calcium ionophore than FAK [8-10]. Some of these differences may be owing to distinct subcellular localization. FAK is restricted to sites of focal contact and while Pyk2 is found at focal contacts but also maintains cytoplasmic and perinuclear localization [6, 7, 10].

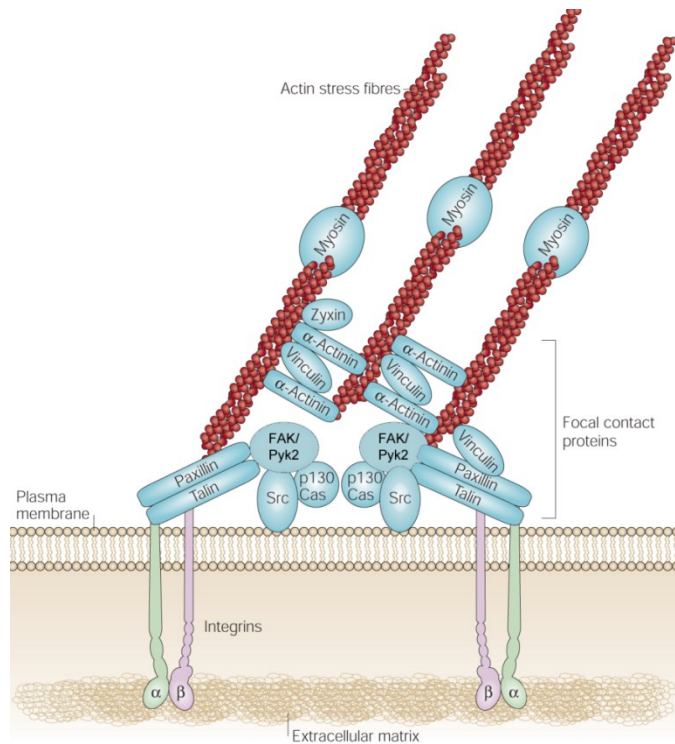


Figure 4-2. Architecture of focal contacts. Adapted from Mitra et al 2005, [6].

The extracellular matrix, integrins, and the cell cytoskeleton interact at sites called focal contacts. Focal contacts are dynamic groups of structural and regulatory proteins that transduce external signals to the cell interior and can also relay intracellular signals to generate an activated integrin state at the cell surface. The integrin-binding proteins paxillin and talin recruit focal adhesion kinases and vinculin to focal contacts. Protein tyrosine kinase Src and the adaptor protein p130Cas associate with focal contacts following integrin clustering. Integrin-mediated FAK activation is mediated in part by matrix binding or by force-dependent changes in cytoskeletal linkages. Several other proteins such as ERK2 and calpain are known to be transiently present at focal contacts (not shown). The composition of a focal contact is therefore constantly varying depending on external cues and cellular responses [6].

The functional domains of Pyk2 include an N-terminal FERM domain, a central catalytic domain, a number of proline rich sequences, and the C-terminal focal adhesion targeting domain or FAT. The FAT domain of Pyk2 also interacts with the N-terminus of paxillin and the C-terminus of gelsolin and regulates its activity linking Pyk2 to the regulation of actin cytoskeleton organization [6] (Figure 4-3).

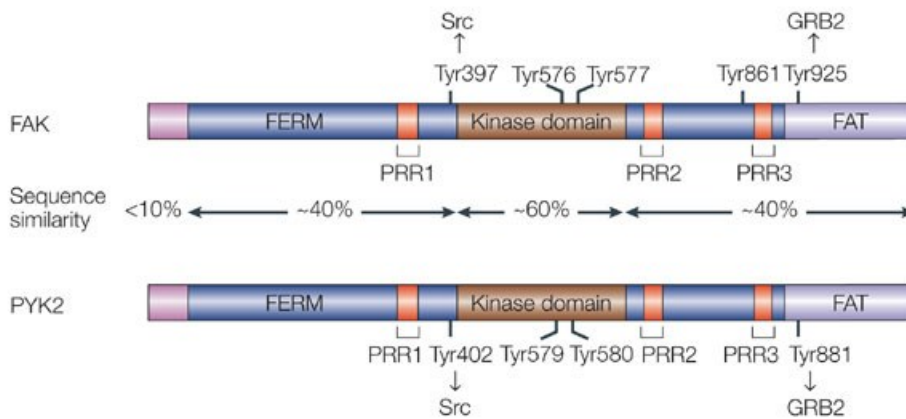


Figure 4-3. Domains, motifs, and phosphorylation sites in FAK family tyrosine kinases FAK and Pyk2. Adapted from Mitra et al 2005, [6].

Pyk2 shares a similar domain arrangement with FAK, with 60% sequence identity in the central kinase domain, conservation of proline-rich regions (PRRs), and identical positions of four tyrosine phosphorylation sites. Pyk2 tyrosines 402, 579, 580 and 881 correspond to FAK tyrosines 397, 576, 577 and 925, respectively. Phosphorylation of Pyk2 Tyr402 and Tyr881 create Src-homology-2 (SH2) binding sites for Src and Grb2, respectively. Pyk2 contains a C-terminal focal adhesion targeting (FAT) domain that binds to paxillin [6].

Analysis of Pyk2 activation events has shown the importance of phosphorylation at four tyrosine residues Tyr402, Tyr 579, Tyr580, and Tyr881 [11]. Autophosphorylation of Pyk2 is a trans event and occurs solely at Tyr402 which is the initiating step for activation of the additional tyrosine phosphorylation sites [12]. The literature consistently states that upon autophosphorylation at Tyr402, Src is recruited through binding to phosphorylated Tyr402 via the Src-SH2 domain, which stimulates the activation of Src [7, 12]. Activated Src then phosphorylates Pyk2 on Tyr579 and Tyr580 which enhances the catalytic activity of Pyk2 [7]. The FAT domain also contains Tyr881 which when phosphorylated by Src serves as a binding site for the adaptor protein Grb2, providing the linkage to activation of the MAP kinase signaling pathway [5, 13] (Figure 4-4). LPA and bradykinin stimulation in PC12 cells triggers Pyk2 autophosphorylation and binding of Src, linking G-protein coupled receptors with Grb2 Sos to activate the MAP kinase pathway [14]. In the bioactive peptide signaling peptide pathway Pyk2 is shown to be calcium sensitive and responsible for activation of P38 and JNK. In this pathway Src, an important downstream kinase, is named as an essential intermediate in the activation Ras and the MAP kinase pathways by signaling through GPCRs [1]. Pyk2 is shown to constitutively interact with known substrates paxillin and p130cas in B cells [15, 16]. These two proteins connect Pyk2 to the cytoskeleton indicating a role in cell adhesion and focal adhesion dynamics [7, 16, 17].

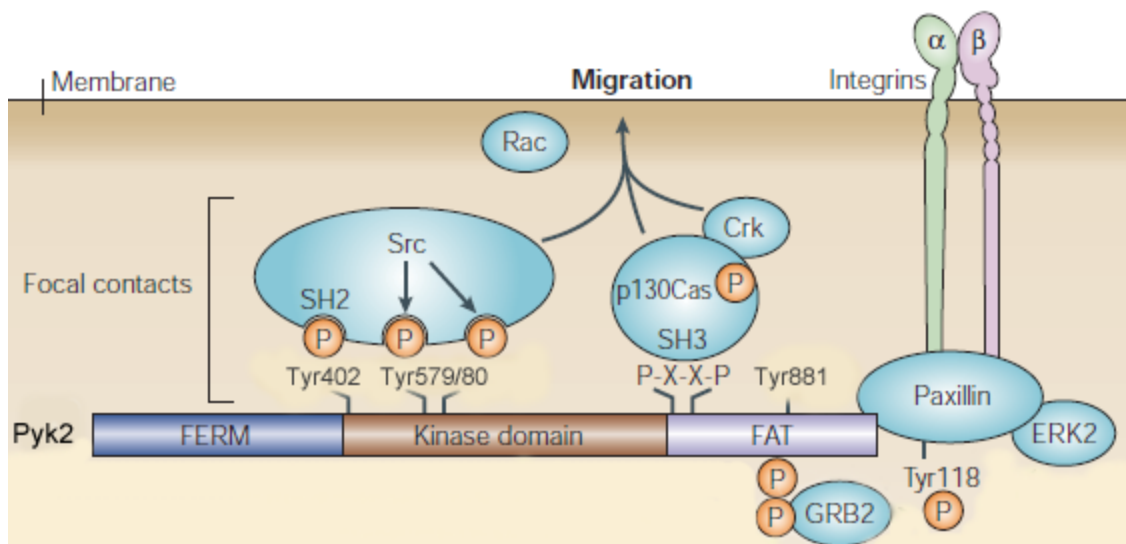


Figure 4-4. Pyk2 signaling within focal contacts. Adapted from Mitra et al 2005, [6].

Integrin activation or growth factor signaling promotes Pyk2 autophosphorylation at Tyr402, which creates a binding site for the SH2 domain of Src. Src-mediated phosphorylation of Pyk2 at Tyr579 and Tyr580 promotes full catalytic activity of Pyk2. Active Pyk2–Src facilitates SH3-mediated binding of p130Cas to Pyk2 and its subsequent phosphorylation. Crk binding to phosphorylated p130Cas facilitates Rac activation, lamellipodia formation and cell migration. Paxillin binding to the Pyk2 focal adhesion targeting (FAT) domain is important for Pyk2 focal contact localization. Src-mediated phosphorylation of Pyk2 at Tyr881 creates an SH2 binding site for the Grb2 adaptor protein, which leads to the activation of Ras and ERK2 cascade. Within focal contacts, Pyk2–Src promotes phosphorylation of paxillin at Tyr118 [6].

The studies in Chapter 4 describe our efforts to characterize the role of SUMOylation in the dynamics of Pyk2 activation and signaling, identify potential crosstalk mechanisms

between phosphorylation and SUMOylation, as well as establish a phenotype associated with Pyk2 SUMOylation. We have determined that Pyk2 is SUMOylated by SUMO1 and SUMO2 by PIAS1 and PIAS4 which confirms the results of our proteome microarray screen. We have shown SUMOylation of Pyk2 enhances autophosphorylation at Tyr402. Importantly, we our results demonstrates that SUMOylation of Pyk2 stimulates the interaction with Src in a phosphorylation independent manner. We have also determined SUMOylation of Pyk2 enhances its ability to phosphorylate paxillin. Finally, we examined the ability for SUMOylation of Pyk2 to enhance cell migration after wounding in metastatic breast cancer cells.

4.1. Results

4.1.1. Validation of Pyk2 SUMOylation in HeLa Cells

The results of the global E3 ligase substrate screen indicated Pyk2 is SUMO modified under numerous conditions: high concentration E1 and E2 with SUMO1, PIAS1 SUMO1, PIAS2 SUMO1, and PIAS4 SUMO2 (Figure 4-5).

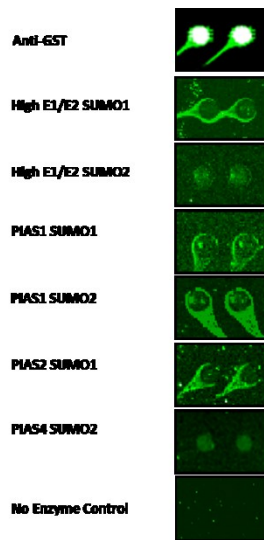


Figure 4-5. Pyk2 SUMOylation on Human Proteome chip

Pyk2 is SUMO modified under several conditions in the Human Proteome chip E3 ligase assays. Raw images from chip-based assays are shown.

To confirm the results from the protein microarray experiment Pyk2 was subcloned into a V5 mammalian expression vector and co-transfected with MYC-tagged SUMO, plus and minus the appropriate E3 ligase. SUMOylation status was assessed by V5 immunoprecipitation followed by western blotting for MYC-tagged SUMO. The cell based assays confirmed E3 ligase specificity, however, there were differences in paralog specificity. We can attribute variation in paralog specificity in the *in vitro* versus *in vivo* assays to potential differences in localization of pools of SUMO in the cell, the involvement of additional proteins either in complex with the substrate or SUMOylation machinery, or the activity of the sentrin proteases. In the cell based assays with expression of SUMO1 and in the absence of E3 overexpression, Pyk2 can be modified

with SUMO1. The addition of PIAS1 showed very strong stimulation with SUMO1 but not SUMO2, PIAS2 was virtually inactive, and PIAS4 caused robust SUMOylation with SUMO1. Overall, Pyk2 seems to be preferentially modified by SUMO1 (Figure 4-6). Ironically, in a paper where FAK was reported to be SUMOylated by PIAS1 with SUMO1 at K152, Pyk2 failed to show SUMOylation under the conditions used in their assays [18].

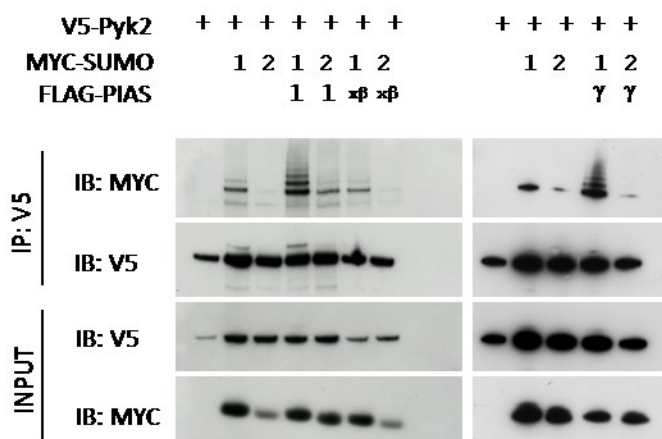


Figure 4-6. Validation of Pyk2 SUMOylation and E3 ligase specificity

HeLa cells were transfected with plasmids encoding for V5-tagged Pyk2, MYC-tagged SUMO1 or SUMO2, along with FLAG-tagged PIAS1, PIAS2, and PIAS4. V5-Pyk2 was immunoprecipitated and analyzed by western blotting with V5 and MYC antibodies as indicated.

4.1.2. Mapping SUMOylation Sites in Pyk2

To generate a tool to assess the function of SUMOylation on Pyk2, we must first identify the sites of SUMO attachment to generate a SUMO deficient mutant for comparative functional studies. SUMO attaches to lysine residues and thus it is possible to predict potential lysine attachment sites, generate arginine mutants, and perform SUMOylation assays. Mass spectrometry can also reveal the sites of attachment to the substrate. There are serious caveats to consider with both approaches. If the substrate protein contains only a few lysines, prediction-directed mutagenesis is a reasonable approach. SUMO site prediction programs can be very effective at narrowing down potential lysines for mutation, however, this approach may be too cumbersome if the substrate is a large protein with dozens of lysine residues. The likelihood of selecting the correct lysine or combination of lysines becomes very low and the manual work required may prove prohibitive. Mass spectrometry for SUMOylation site mapping also has inherent challenges. When a SUMO1 modified protein undergoes tryptic cleavage, the result is the substrate fragment covalently attached to a 19 amino acid SUMO1 fragment. Unlike non-peptide PTMs (e.g. phosphorylation and acetylation), further fragmentation of this product creates multiple overlapping b- and y- fragment ions that are uninterpretable using standard peptide sequencing software [19]. Accurate identification of the site(s) relies on manual calculations which may be burdensome.

Pyk2 contains 68 lysine residues throughout the amino acid sequence. At the outset of the site mapping effort, we used SUMOsp prediction software to identify high confidence lysines for mutation. Lysine 35, 145, and 895 (a consensus site) lysine were mutated to create a lysine (K) to arginine (R) mutant with all three candidate lysines mutated in one

construct. Transfection of HeLa cells with WT and 3KR Pyk2 plasmids showed reduced SUMOylation by SUMO1 on 3KR (Figure 4-7). This indicates that at least one of the lysines mutated in the 3KR mutant is a SUMO attachment site but other lysines in the protein remain SUMO modified.

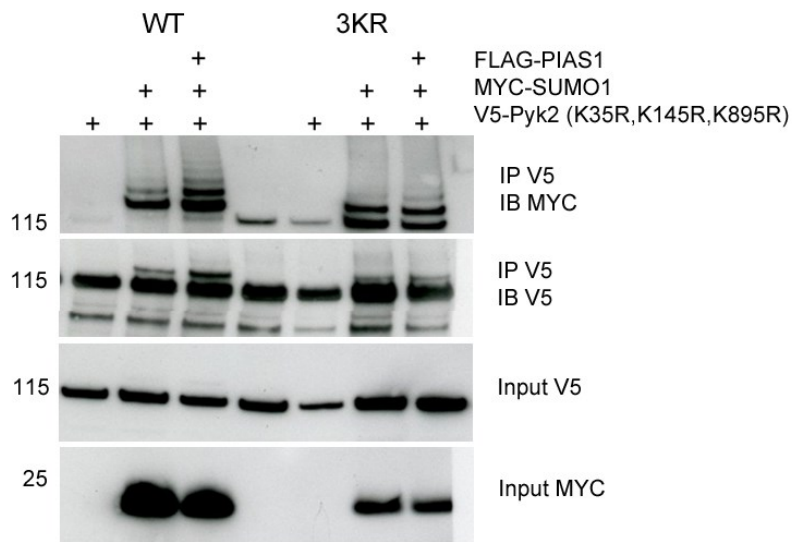


Figure 4-7. PIAS1 directed modified 3KR mutant shows impaired SUMOylation compared to wild type.

HeLa cells were transfected with plasmids encoding for V5-tagged Pyk2 or V5-Pyk2 containing three lysine to arginine mutations (3KR). V5-Pyk2 constructs were co-transfected with MYC-tagged SUMO1 along with FLAG-tagged PIAS1 as indicated. V5-Pyk2 was immunoprecipitated and analyzed by western blotting with V5 and MYC antibodies.

The decision to discontinue hunting for SUMOylated lysines with SUMOsp forced us to pursue mass spectrometry. Because the abundance of the SUMOylated form of the

protein is low, we chose to perform a double immunoprecipitation, to enrich for the SUMOylated species. Approximately 4×10^7 HeLa cells were transfected with V5-Pyk2, MYC-SUMO1, and FLAG-PIAS1. First, anti-V5 agarose was used to immunoprecipitate the total population of Pyk2. This was washed and eluted from the beads with a V5 peptide, MYC-agarose was used to efficiently pull down the SUMOylated fraction. The eluate from this pulldown was precipitated and concentrated. The recovered Pyk2-SUMO1 sample was separated by SDS-PAGE then silver stained. Four prominent bands in the silver stained gel corresponded to the molecular weight of multiply SUMOylated Pyk2 (Figure 4-8). These bands were excised, destained, then subjected to tryptic digest followed by tandem mass spectrometry. The resulting spectra was analyzed by Mascot and SEQUEST, however, both programs failed to identify SUMOylated Pyk2 peptides. Rigorous manual calculations were required create interpretable composite mass predictions for potential SUMOylated peptides. This analysis revealed lysine 581 as the only potentially SUMOylated lysine residue (Figure 4-9). It is interesting to note that the tyrosine residues, Tyr579/80, that Src phosphorylates are directly adjacent to this lysine. We hypothesized that, if indeed a true attachment site, SUMOylation as K581 site may affect the Src interaction.

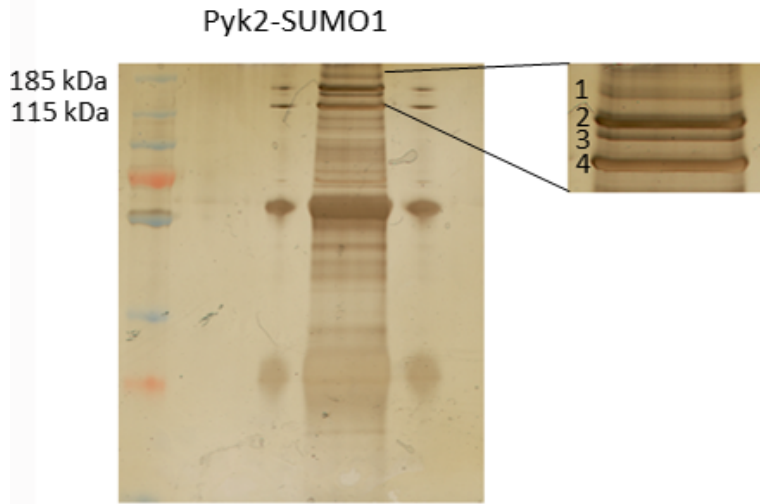


Figure 4-8. Pyk2-SUMO1 immunoprecipitate visualized by silver staining

V5-Pyk2 coexpressed with MYC-SUMO1 in HeLa cells. Pyk2-SUMO1 was sequentially immunoprecipitated first by V5 affinity agarose, eluted, then pulled down with MYC-agarose. The sample was separated by SDS-PAGE and visualized by silver staining. Four high molecular weight bands correspond to SUMO-modified forms of Pyk2.

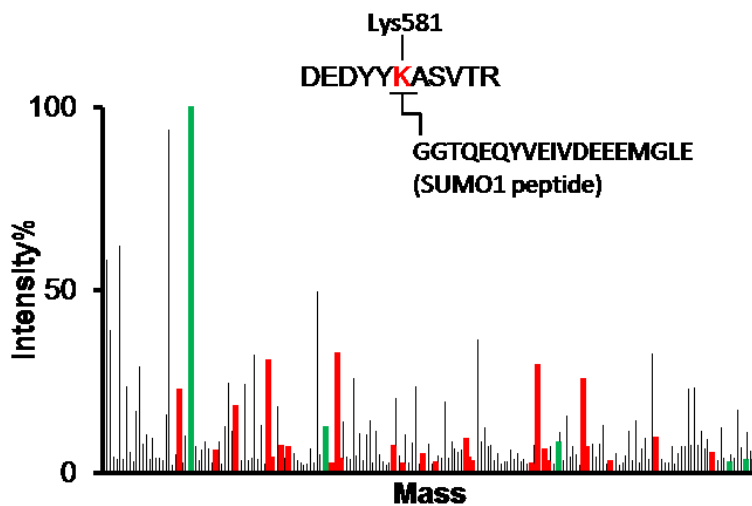


Figure 4-9. K581 SUMO1 modification spectra

Co-expression of a lysine 581 single mutant with SUMO1 resulted in equivalent SUMOylation as WT. The evidence from the 3KR mutant suggested that while, at least one K->R mutation reduced SUMOylation, additional sites are still targeted by SUMOylation. We elected to introduce a fourth lysine mutation at K581 to the 3KR mutant construct. Co-transfection of the 4KR in parallel with 3KR and WT revealed dramatically reduced SUMOylation on 4KR Pyk2 (Figure 4-10). We were able to detect reduced SUMOylation with the anti-MYC antibody as well as anti-SUMO1 (21C7). This result gave us confidence that 4KR Pyk2 would serve as SUMOylation deficient mutant for functional studies

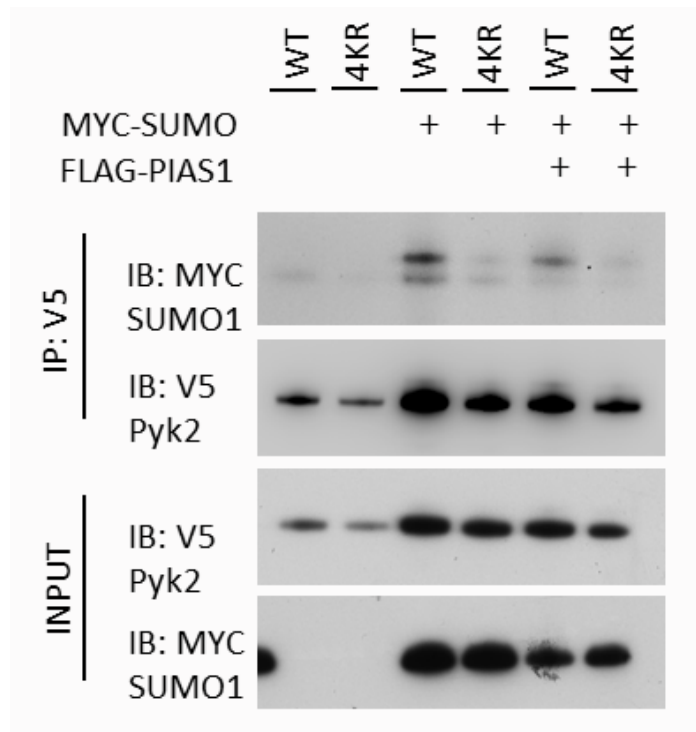


Figure 4-10. 4KR Pyk2 mutant exhibits significantly reduced SUMOylation compared to wild type

HeLa cells were transfected with plasmids encoding for V5-tagged Pyk2 or V5-Pyk2 containing four lysine to arginine mutations (4KR). V5-Pyk2 constructs were co-transfected with MYC-tagged SUMO1 along with FLAG-tagged PIAS1 as indicated. V5-Pyk2 was immunoprecipitated and analyzed by western blotting with V5 and MYC antibodies.

4.1.3. SUMOylation of Pyk2 enhances autophosphorylation at Tyr402

In the study that validated SUMOylation of FAK, FAK-SUMO1 was associated with higher Tyr397 autophosphorylation signal than non-SUMOylated FAK, however the K152R mutant did not show reduced Tyr397 phosphorylation signal [18]. Because autophosphorylation of Pyk2 is an essential step in its activation and function, we also examined whether it was altered by SUMOylation. We compared the levels of autophosphorylation at Tyr402 with a phosphospecific Tyr402 antibody in the following conditions: non-SUMOylated Pyk2, SUMOylated Pyk2, and 4KR Pyk2 co-transfected with MYC-SUMO1, and FLAG-PIAS1 or PIAS4 (Figure 4-11). WT Pyk2-SUMO1 showed markedly higher autophosphorylation than non-SUMOylated WT Pyk2. Dramatically less pTyr402 signal was detected for 4KR Pyk2 co-transfected with SUMO1 than WT Pyk2-SUMO1. Co-expression with PIAS1 and PIAS4 increases the level of SUMOylation of WT Pyk2 but decreases the level of pTyr402 signal compared to co-expression of SUMO1 without E3 ligases. Autophosphorylation signal was

undetectable with co-expression of 4KR Pyk2 and PIAS1 or PIAS4. These results demonstrate SUMOylation of Pyk2 increased the phosphorylation level of WT Pyk2. Because 4KR Pyk2 shows significantly less autophosphorylation activity we consider increased pTyr402 signal a SUMO conjugation-dependent event. It is important to note that the levels of SUMOylation and autophosphorylation do not entirely correlate, as co-expression with E3 ligases reduces the pTyr402 signal.

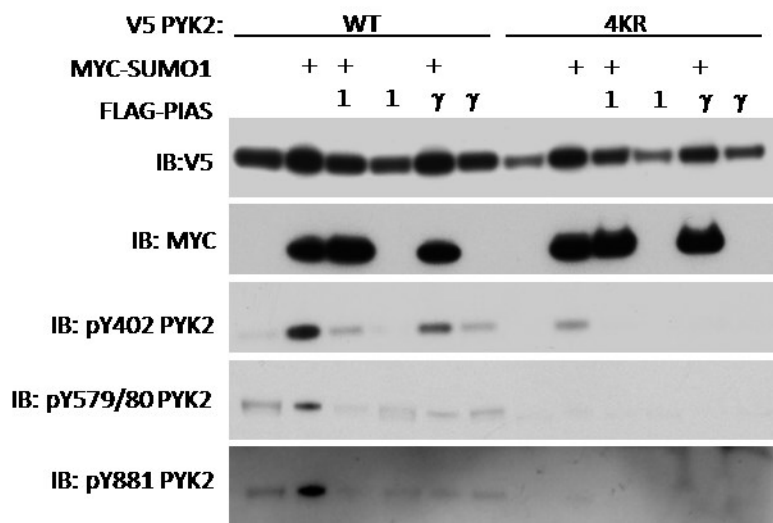


Figure 4-11. Autophosphorylation at Tyr402 is impaired in 4KR mutant

HeLa cells were transfected with plasmids encoding for V5-tagged Pyk2 or V5-Pyk2 containing four lysine to arginine mutations (4KR). V5-Pyk2 constructs were co-transfected with MYC-tagged SUMO1 along with FLAG-tagged PIAS1 or PIAS4 as indicated. Lysates were analyzed for phosphorylation status at Tyr402, Tyr579/80, and Tyr881 using phosphospecific antibodies.

Enhanced autophosphorylation in mammalian cells can result from stimulation of phosphorylation or inhibition of dephosphorylation. To shed light on the mechanism of enhanced autophosphorylation of Pyk2-SUMO1, we conducted an *in vitro* tandem SUMOylation/kinase activity assay. Recombinant purified GST-Pyk2 was SUMOylated by SUMO1 in the absence of an E3 ligase. Autophosphorylation was examined for the following conditions: untreated, autophosphorylated (kinase assay), SUMOylated, SUMOylated then autophosphorylated (tandem assay). The samples were resolved by SDS-PAGE and subjected to western blotting with anti-pTyr402 antibody. The pTyr402 signal confirmed SUMOylation indeed enhances autophosphorylation more than the kinase assay with ATP alone (Figure 4-12). The strongest autophosphorylation signal was observed when Pyk2 was SUMOylated then subject to autophosphorylation reaction. SUMOylation alone produced a stronger signal than the autophosphorylation reaction alone. This is likely because SUMOylation is an ATP dependent mechanism and thus the kinase may autophosphorylate during the SUMOylation step. Naïve untreated protein did not have any basal pTyr402 signal. These results demonstrate that SUMOylation reliably increases the intrinsic autophosphorylation activity of Pyk2.

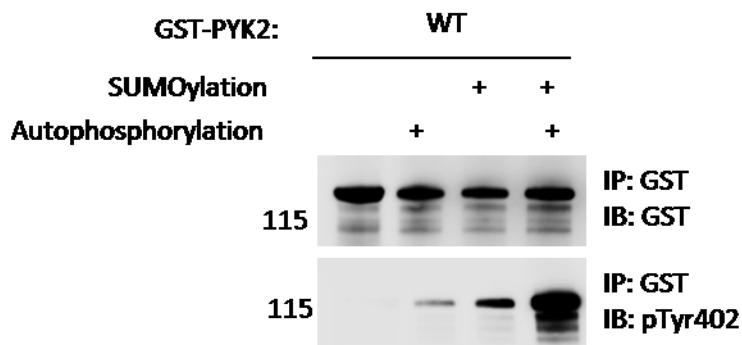


Figure 4-12. SUMOylation stimulates autophosphorylation of Pyk2 *in vitro*

Recombinant GST-Pyk2 was SUMOylated by SUMO1 in the absence of an E3 ligase. Autophosphorylation was examined for the following conditions: untreated, autophosphorylated (kinase assay), SUMOylated, SUMOylated then autophosphorylated (kinase assay). The samples were resolved by SDS-PAGE and subjected to western blotting with anti-pTyr402 antibody.

Pyk2 autophosphorylation on Tyr402 is known to create a binding site for the Src homology 2 domain (SH2) of Src, which in turn phosphorylates other Pyk2 tyrosine residues including Tyr579 and Tyr580 in the kinase domain, and Tyr881 in the FAT domain [20]. We also examined the effect of SUMOylation on phosphorylation of Tyr579, Tyr580, and Tyr881 with phosphosite-specific antibodies. Pyk2-SUMO1 was the only condition that showed strong phosphorylation signal at all four tyrosine sites. No phosphorylation signal was detected for Tyr 579, 580, and 881 for Pyk2 4KR. Based on these results we can conclude SUMOylation of Pyk2 stimulates autophosphorylation of Tyr402 and subsequent phosphorylation at Tyr579, Tyr580, and Tyr881 (Figure 4-11).

4.1.4. SUMOylation stimulates Src association with Pyk2

The established mechanism of interaction between Pyk2 and Src states that autophosphorylation of Pyk2 on Tyr402 leads to binding of the SH2 domain of Src and its subsequent activation [14]. A mutant form of Pyk2 in which Tyr402 is mutated to phenylalanine is autophosphorylation deficient and was unable to bind to Src through its SH2 domain [14]. As SUMOylation of Pyk2 directly affects phosphorylation at Tyr402,

we chose to further examine the role of SUMOylation in the association of Pyk2 and Src. Expression of WT Pyk2, 4KR Pyk2, and a Y402F mutant with and without SUMO1 resulted in stimulation of Src interaction with Pyk2-SUMO1 but non-SUMOylated Pyk2, 4KR, or 4KR co-expressed with SUMO1 (Figure 4-13). SUMO1 modification of Pyk2 by PIAS1 results in higher levels of SUMOylation, less autophosphorylation than Pyk2-SUMO in the absence of PIAS1 (consistent with earlier findings). In continuing with this trend, Src interaction with Pyk2-SUMO1 modified by PIAS1 is less than Pyk2-SUMO1 in the absence of PIAS1. We tested SUMOylation of the Pyk2 Y402F and found that it becomes SUMOylated as efficiently as WT Pyk2. Thus we can conclude that Y402F phosphorylation is not required for SUMOylation of Pyk2. A very surprising result was Pyk2 Y402F-SUMO1 was able to rescue the association with Src, despite the lack of autophosphorylation. This represents the first time Pyk2-Src interaction is not autophosphorylation dependent and instead SUMOylation dependent.

Pyk2-SUMO1-Src association stimulated activation of Src at Tyr416, however, pTyr416 signal was not present in the Pyk2 Y402F-SUMO1-Src interaction. We interpret that while SUMOylation may stimulate the interaction between Pyk2 and Src, it is not sufficient to initiate the activation of Src. To achieve activation of Src and Pyk2, SUMOylation and autophosphorylation are necessary. This represents a novel example of crosstalk between SUMOylation and phosphorylation, where both modifications are required for the full effect.

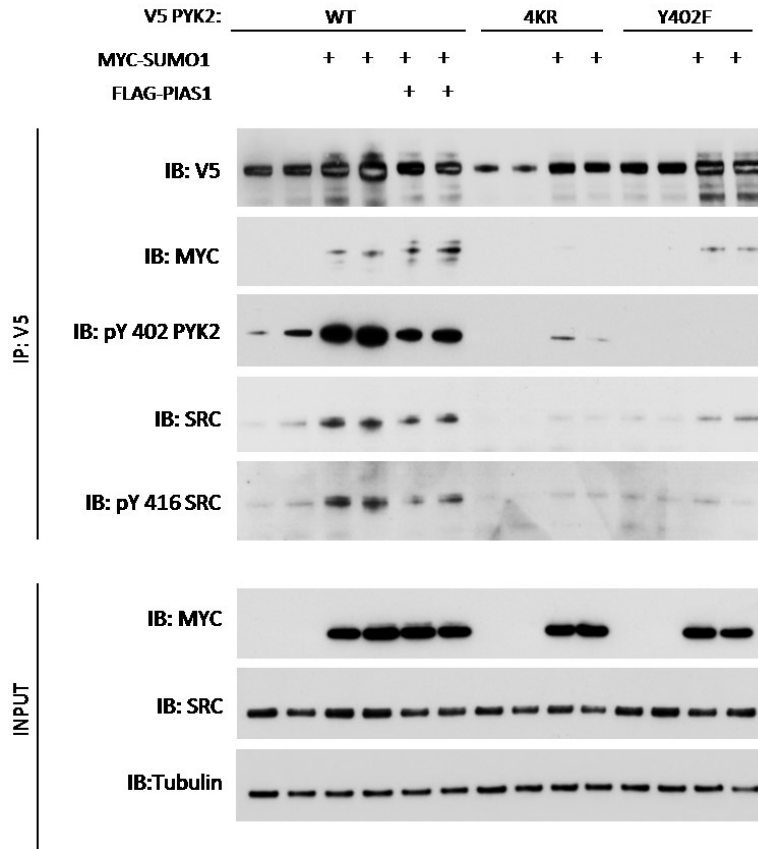


Figure 4-13. SUMOylation of WT Pyk2 but not 4KR promotes association with endogenous Src

HeLa cells were transfected with plasmids encoding for V5-tagged Pyk2, V5-Pyk2 containing four lysine to arginine mutations (4KR), and V5-Pyk2 Y402F. V5-Pyk2 constructs were co-transfected with MYC-tagged SUMO1 along with FLAG-tagged PIAS1. V5-Pyk2 was immunoprecipitated under non-denaturing conditions to assess the interaction with endogenous Src. Phosphorylation status of Pyk2 and Src were also analyzed using phosphospecific antibodies. Inputs controls for MYC-SUMO1, endogenous Src and Tubulin are also shown.

4.1.5. Role of SUMO modification in Pyk2-mediated Paxillin Phosphorylation

Paxillin is a focal adhesion protein, constitutive interactor with Pyk2 and FAK, and a known *in vivo* target of Pyk2 phosphorylation [9, 16, 21]. We wanted to examine whether Pyk2 SUMOylation also has the ability to stimulate tyrosine phosphorylation of downstream targets of Pyk2 that are involved in focal adhesion dynamics. WT Pyk2, 4KR Pyk2, and Pyk2 Y402F were co-expressed in HeLa cells with SUMO1, plus and minus the E3 ligases PIAS1 and PIAS4. Pyk2 and mutant versions were immunoprecipitated and immunoblotted for paxillin and Tyr118 with a phosphospecific antibody. Paxillin constitutively co-immunoprecipitated with Pyk2 and its mutant forms. Immunoblotting for tyrosine phosphorylation of paxillin at Tyr118 revealed that WT Pyk2 SUMOylation enhances downstream paxillin phosphorylation. Paxillin Tyr118 phosphorylation was strongest in the presence of Pyk2-SUMO1 and absent with the 4KR and Y402F mutants. Phosphorylation levels were similar between non-SUMOylated Pyk2 and 4KR co-expressed with SUMO1 (Figure 4-14). We can conclude that Pyk2 SUMOylation not only enhances autophosphorylation and association with Src, but it also stimulates phosphorylation of at least one downstream target, paxillin.

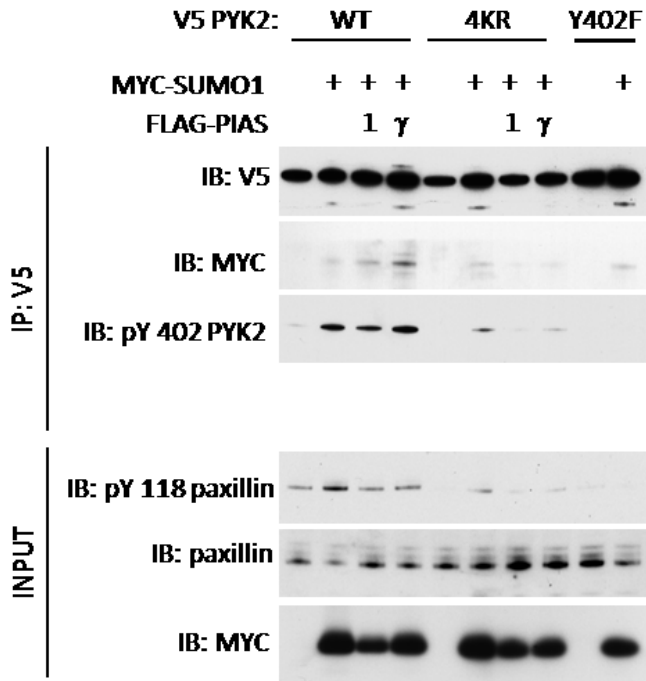


Figure 4-14. SUMOylation of Pyk2 promotes phosphorylation of paxillin

HeLa cells were transfected with plasmids encoding for V5-tagged Pyk2, V5-Pyk2 containing four lysine to arginine mutations (4KR), and V5-Pyk2 Y402F. V5-Pyk2 constructs were co-transfected with MYC-tagged SUMO1 along with FLAG-tagged PIAS1. Phosphorylation status of Pyk2 adaptor protein paxillin was determined by western blotting for pTyr118.

4.1.6. Pyk2 SUMOylation mutant, 4KR, shows reduced signaling through ERK1/2 axis

Paxillin serves an important role as a scaffold protein for MAPK pathway activation, critical for regulating cytoskeletal dynamics and cell migration [22, 23]. Paxillin associates constitutively with either the active or inactive forms of MEK, however, activated Ras and inactive ERK preferentially interact with the scaffold [22]. ERK directly associates with the ¹¹⁸YSFP region of paxillin but displays higher affinity for the

phosphor-tyrosine form (pTyr118). This scaffolding mechanism may serve to recruit inactive ERK to site of new focal contact in order to mediate ERK activation by MEK at these discrete locations. Upon activation ERK in turn phosphorylates paxillin at Serine 83, which promotes its interaction with activated FAK under stimulation with HGF [24]. Paxillin mutants, ASVA and S83A, cannot reproduce increased association with activated FAK, suggesting recruitment of activated FAK is dependent on paxillin phosphorylation [24]. An analogous mechanism for the association of paxillin with Pyk2 has not been described. Evaluation of the phenotype associated with disrupting the paxillin-ERK in mIMDC-3 cells, resulted in failure to engage in cell spreading.

Based on our result that Pyk2-SUMO1 stimulates phosphorylation of paxillin at Tyr118, within the ERK interaction site, we chose to examine the effect of Pyk2-SUMOylation on ERK activation. To better evaluate the role of SUMOylation within this signaling axis, we transfected HEK293 cells with WT Pyk2 and 4KR, with and without SUMO1 co-transfection. These experiments reveal that overexpression of WT Pyk2 stimulates ERK1/2 activation, whereas the 4KR mutant alone does not result in phosphorylation of ERK1/2. With both Pyk2 WT and 4KR, co-expression with SUMO1 caused a modest increase in ERK1/2 activation (Figure 4-15). Pyk2 has also been shown to be necessary for P38 activation in HUVEC endothelial cells [25], thus we were interested to test whether SUMO1 co-expressed with Pyk2 WT or 4KR could enhance P38 phosphorylation. Over expression of Pyk2 WT showed a significant increase in P38 activation, however, we do not observe a SUMO-dependent increase in activation. These experiments reveal that disruption of SUMOylation with the 4KR mutant specifically impairs ERK1/2 activation. Collectively, our results suggest that the SUMOylation-

dependent Pyk2-Src interaction recruits Grb2 Sos, activating Ras and the MAP kinase pathway.

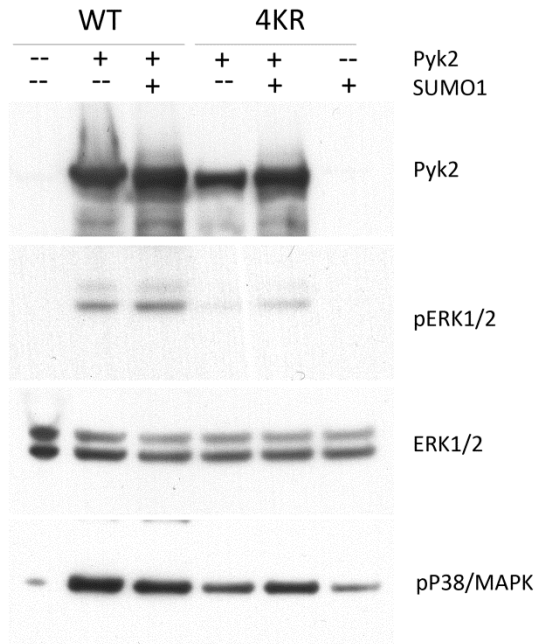


Figure 4-15. Pyk2 4KR mutant inhibits activation of ERK1/2 but not P38/MAPK

HEK293 cells were transfected with either WT or 4KR versions of Pyk2, plus and minus SUMO1 co-expression. Whole cell lysates were subject to western blotting to probe for activation of ERK1/2 (pERK1/2), total ERK1/2, activated P38 MAPK (pP38/MAPK), and Pyk2 expression.

4.1.7. SUMOylation of Pyk2 affects paxillin-mediated cell migration

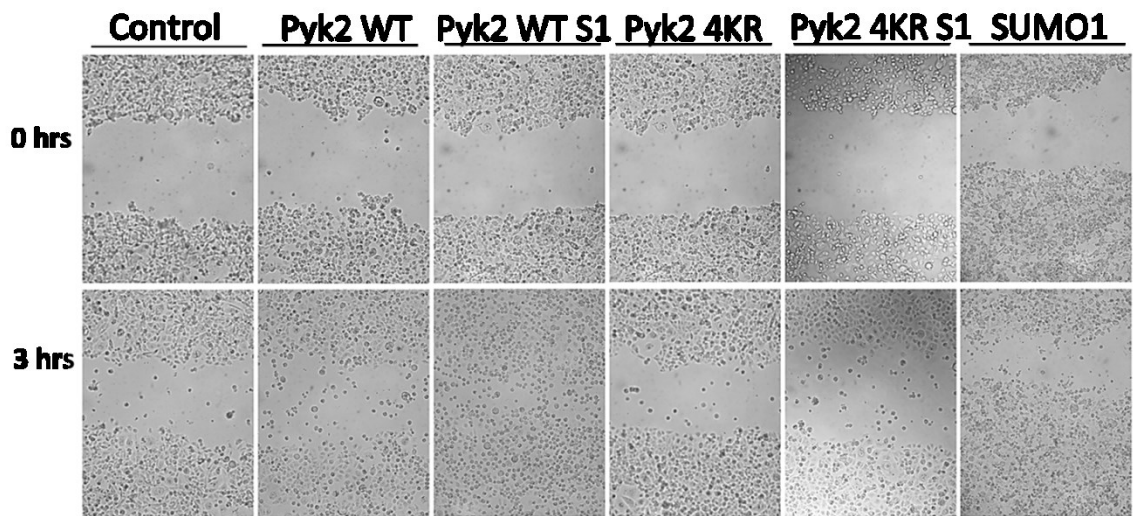
Paxillin is a focal adhesion protein that serves to recruit other cytoskeletal proteins such as vinculin, talin, and Crk [21]. Paxillin is truly a multiadaptor protein, as exemplified by

its scaffold function in the MAPK kinase pathway. Paxillin serves to coordinate activation of ERK at sites of focal adhesion to potentiate cell migration [22-24, 26, 27]. Paxillin undergoes heavy tyrosine phosphorylation during dynamic focal adhesion events like cell adhesion, cytoskeletal reorganization, and growth signaling [21]. Phosphorylation of paxillin at Tyr31 and Tyr118 serve as a docking sites for Crk through its SH2 domain [21, 28]. These sites are highly phosphorylated during cell migration and the phosphorylated species is localized at the leading edges of migrating epithelial cells [29]. The paxillin-Crk complex recruits DOCK180, an SH3 domain containing protein downstream target of integrin signaling [28, 30, 31]. DOCK180 can directly interact with and activate small GTPase Rac1 [32], and thereby promotes the migration and invasion of mammary epithelial cells [33]. In addition, Paxillin Tyr118 phosphorylation correlates with cell migration in rat bladder carcinoma NBT-II cells [28]. Furthermore, Jnk and Brk, two cytoplasmic kinases have been shown to regulate cell migration through paxillin phosphorylation in MDA-MB-231 breast cancer cells [34, 35].

With this evidence, combined with our finding that Pyk2 SUMOylation triggers paxillin phosphorylation, we hypothesized that SUMOylation of Pyk2 is an upstream event that promotes cell migration through phosphorylation of paxillin. To test this hypothesis we performed a cell migration assay using MDA-MD-231 metastatic epithelial breast cancer cells. Adenovirus encoding for WT Pyk2 or Pyk2 4KR with and without co-infection of SUMO1, were used to cause over expression of Pyk2-SUMO1 in breast cancer cells. A confluent monolayer of cells expressing our constructs was wounded and cell migration was monitored over the next 3 hours. Expression of Pyk2 stimulated cell migration into the wound. Co-infection with SUMO1 showed enhanced migration over control, SUMO1

alone or Pyk2 alone. Pyk2 4KR expression showed slightly more migration than the control, however, migration was impaired compared to WT Pyk2 (Figure 4-15). Co-infection of SUMO1 with Pyk2 4KR also showed an increase in migration over 4KR alone. The number of migratory cells in the WT Pyk2 SUMO1 condition is significantly larger than the untreated control cells (p -value 0.009). Also, Pyk2 WT shows significantly more migration than the Pyk2 4KR mutant (p -value 0.015). We suspect that 4KR shows some enhanced migration compared to the control because although its activation is impaired, it is not kinase dead. This provides potential for low levels of kinase activity. This study offers evidence that SUMOylation of Pyk2 plays a positive role in cell migration likely through tyrosine phosphorylation of focal adhesion protein paxillin.

A.



B.

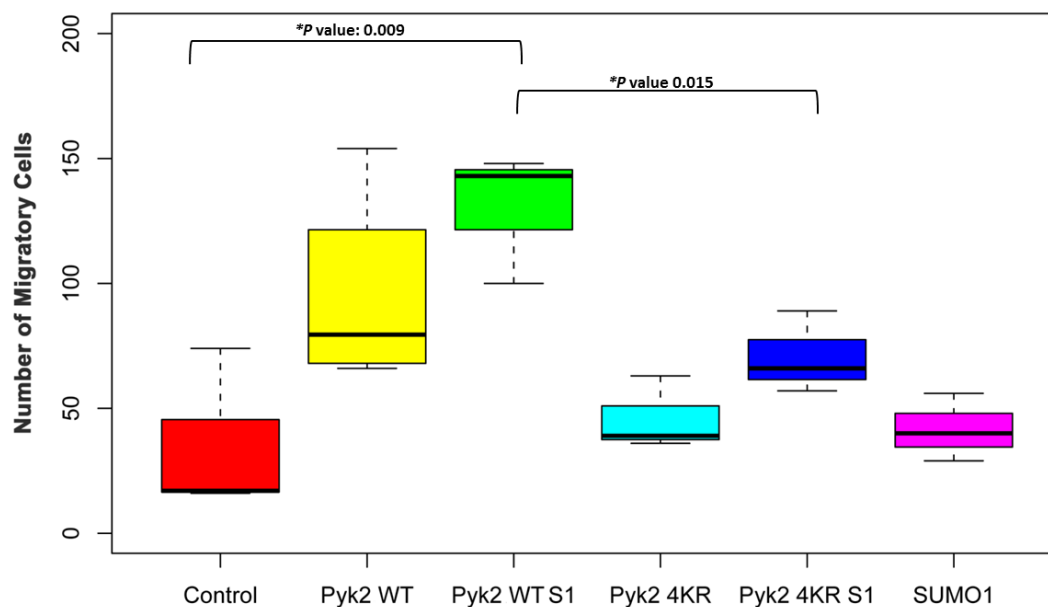


Figure 4-16. SUMOylation of Pyk2 stimulates cell migration in MDA-MB-231 breast cancer cells

A. Adenovirus encoding for Pyk2 WT or Pyk2 4KR with and without co-infection of SUMO1, were used to express Pyk2-SUMO1 in breast cancer cells. A confluent monolayer of cells expressing our constructs was wounded and cell migration was monitored over the next 3 hours. B. Box/Whisker plots were generated using R. The number of migratory cells in the Pyk2 WT SUMO1 condition is significantly larger than the untreated control cells (p -value 0.009). Also, Pyk2 WT shows significantly more migration than the Pyk2 4KR mutant (p -value 0.015).

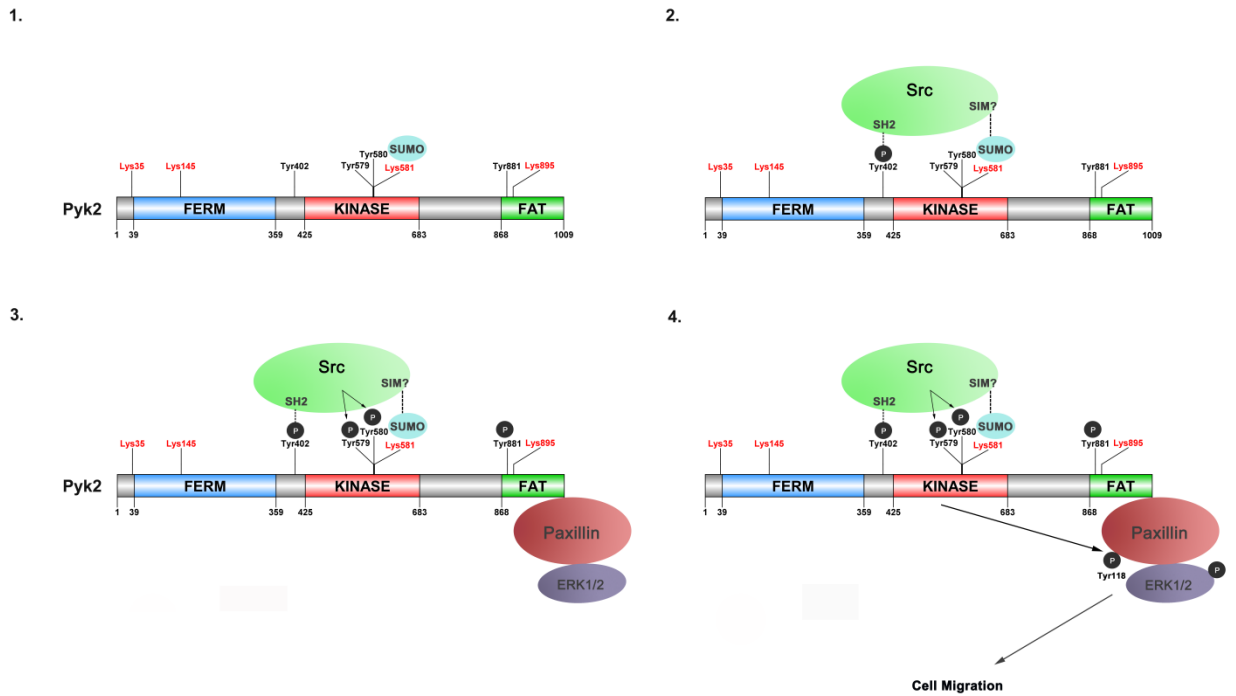


Figure 4-17. Crosstalk-mediated activation of Pyk2 model

In step 1, SUMOylation of Pyk2 occurs at Lys581 as well as other acceptor lysine sites. SUMOylation triggers autophosphorylation at Tyr402, which stimulates interaction with Src through its SH2 domain as well as potential SIM-mediated interactions (step 2). Src phosphorylates Pyk2 at Tyr579, Tyr580, and Tyr881 resulting in full catalytic activity of Pyk2 (step 3). Pyk2-SUMO1 phosphorylates focal adhesion protein paxillin at Tyr118. pTyr118 is linked to cell migration likely through activation of the MAP kinase pathway (step 4). SUMOylation of Pyk2 uncovers a novel crosstalk-mediated mechanism for kinase activation and function.

Discussion

We report for the first time, SUMO1 modification of Pyk2 is stimulated by PIAS1 and PIAS4. This finding is of particular importance because a previous study attempted to examine whether Pyk2 is a SUMO substrate yet failed to show its SUMOylation [18]. We were able to demonstrate SUMOylation by numerous methods including protein microarray assays, *in vitro* solution based assays, and in HeLa cells co-transfected with Pyk2, SUMO1, and E3 ligases. While it is well known that Pyk2 autophosphorylation is involved in various cellular signaling processes only one mechanism of activation is known or accepted. Our study has revealed a novel paradigm for the activation mechanism of Pyk2. SUMOylation is able to trigger autophosphorylation of Pyk2 without an upstream stimulus from integrin signaling, G protein activation, or calcium. SUMOylation and phosphorylation are known to cooperate to activate certain transcription factors as well as Aurora kinase [36-38], however this is the first time SUMOylation has been shown to affect the intrinsic activity of a kinase. We found that the role of SUMOylation extended beyond intramolecular activity as SUMOylation of a kinase dead mutant (Y402F) was able to recruit bona fide interaction partner Src, which was assumed to only interact with autophosphorylated Pyk2 only through its SH2 domain. This interaction does not produce full activation of the enzymes thus these findings illustrate a mechanism where the two PTMs cooperate to generate full activation of the Pyk2 at Tyr402, Tyr579, Tyr580, and Tyr881.

In our assays it is clear that PIAS1 and PIAS4 stimulate higher levels of SUMOylation than expression of SUMO1 alone, either by modifying numerous lysines on the protein or generating SUMO chains, but likely a combination of the two. We consistently observe an inverse relationship between the level of SUMOylation and autophosphorylation signal whereby PIAS1 and PIAS4 SUMOylation decrease pTyr402 signal and reduce the association with Src. It is unclear why higher levels of SUMOylation impair kinase activity. Based on our data we cannot infer if it is a direct effect of SUMOylation or a possible secondary effect. Detailed biochemical studies are required to answer these questions. It is important to note that, while performed in cells, our results are based on overexpression studies and thus it is unlikely that under basal conditions PIAS1 or PIAS4 would achieve comparable levels of SUMOylation to what we have observed. Thus it is not sounds to draw the conclusion that PIAS1 and PIAS4 reduce Pyk2 autophosphorylation and/or Src association *in vivo*. To answer this question, we can conduct *in vitro* SUMOylation of Pyk2 with increasing concentrations of E1 and E2 or in the presence of recombinant E3, then monitor the correlation between levels of SUMOylation and autophosphorylation signal. Binding assays with Src following *in vitro* SUMOylation with increasing levels of E1, E2, and/or E3 may clarify the relationship between SUMOylation levels and Src interaction.

Another result to note is our Pyk2 4KR mutant shows dramatically impaired autophosphorylation and Src interaction though both events are detectable at extremely low levels. The results indicate that Pyk2 4KR is SUMOylation deficient yet not kinase dead; it can autophosphorylate in an inefficient manner. Its trans autophosphorylation

mechanism amplifies autophosphorylation activity and allows the potential for activated kinase [11, 39].

SUMOylation is known to affect protein-protein interactions, as we observed with Src, however association with focal adhesion protein paxillin was not affected by SUMOylation as the interaction was constitutive as previously reported [9, 16]. There may be additional protein-protein interactions that are promoted or diminished due to SUMOylation that we can examine through direct immunoprecipitation and immunoblotting or mass spectrometry in future studies. Although SUMOylation of Pyk2 did not affect the interaction with paxillin, it stimulated phosphorylation of paxillin at pTyr118. Notably, phosphorylation of paxillin at Tyr118 triggers binding of inactive ERK [22]. Paxillin serves as a MAP kinase scaffold whereby it also contains binding sites for RAF and MEK, to potentiate the phosphorylation cascade. Paxillin bound ERK becomes activated by MEK phosphorylation, leading to phosphorylation of dozens of downstream molecules involved in cytoskeletal dynamics [22, 24, 40]. Previous studies have shown that Pyk2 acts as an upstream regulator of ERK activity in the context of specific G-protein ligands such as histamine, bradykinin, and lysophosphatidic acid [14, 41]. In addition, Pyk2 has been shown to regulate P38/MAPK activity in endothelial cells under Vascular Endothelial Growth Factor (VEGF) stimulation [25]. Our assays were able to establish a connection between SUMOylation of Pyk2 and activation of paxillin scaffold kinase ERK1/2, but not P38/MAPK. The SUMO deficient 4KR mutant was unable to stimulate phosphorylation of ERK1/2, indicating Pyk2 SUMOylation is critical for tyrosine phosphorylation of paxillin and activation of ERK1/2. These results also suggest SUMOylation of Pyk2 triggers the signaling events associated with G-protein

activation and not growth factor activation. Pyk2 is known to function as a point of convergence between integrin and G-protein coupled cascades. As such, our results put forward the possibility that SUMO modification serves to regulate downstream signaling specificity and discriminate phosphorylation substrates.

FAK, Src, paxillin, and ERK are all known to play a large role in cell motility and focal adhesion turnover [12]. More recently, Pyk2 was also implicated in promoting cell migration, proliferation, and invasiveness through its association with Src and paxillin [17, 27, 41-43]. Pyk2 is involved in cell migration in various tissue types such as glioma, breast cancer, hepatocellular, and B cells under many stimuli including heregulin, integrin signaling, VEGF [27, 43-45], although in many cases over expression of Pyk2 is sufficient to stimulate cell motility. On the basis of these previous studies, together with our finding that Pyk2 SUMOylation triggers phosphorylation of paxillin at Tyr118 and activation of ERK1/2, we sought to test whether Pyk2 SUMOylation could promote cell migration in the context of wound healing. We have strengthened here not only the role of Pyk2 in cell migration but also identified a novel function of SUMOylation, as Pyk2-SUMO1 correlated with enhanced motility of MDA-MB-231 metastatic breast cancer cells. In the context of our results, over expression of Pyk2 promotes cell migration, and co-expression with SUMO augments the migration phenotype. Identifying the endogenous dynamics of Pyk2 SUMOylation is critical for understanding the importance of this finding. Likely, PIAS1 or PIAS4 mediates SUMO modification of Pyk2 *in vivo*, stimulating its autophosphorylation, association with Src, and phosphorylation of paxillin to elicit cell migration. Based on our collective results, we propose a crosstalk-mediated model whereby SUMOylation of Pyk2 stimulates its autophosphorylation activity,

interaction with Src, paxillin phosphorylation, and ERK activation resulting in initiation of cell migration pathways (Figure 4-17).

It is possible that SUMO plays a broader role in cell migration than simply mediating Pyk2 dynamics. Rac1 is a member of the Rho GTPase family that is known to regulate cell migration, adhesion dynamics, and cytoskeleton remodeling [46]. SUMOylation was demonstrated to function in cell migration by modifying and activating GTPase Rac1 in MEF cells [47]. PIAS3 was identified as an E3 ligase for Rac1 and downregulation of PIAS3 resulted in impaired migration compared to controls [47]. Phosphorylation of P38, which is a known downstream mediator of Rac signaling as well as a MAPK SUMOylation substrate in our assays, was also impaired in PIAS3-downregulated cells [47].

Group Modification of Higher-Order Assemblies

The hypothesis that assemblies of physically and functionally connected proteins are SUMO modified as a group, is a provocative and interesting theory that may apply to focal adhesion signaling complexes [48]. The theory explains the rare occurrence of phenotypes linked to individual SUMO deficient mutants by proposing that many proteins in a functional group are modified, where SUMO serves as a glue to promote the collective function of the complex [48]. For example, proteins involved in nucleotide excision repair, ribosomal activity, and homologous recombination are SUMO modified as a “group” [48-51]. Similarly, our studies have revealed many functional connections between SUMOylated proteins, their interactors, and downstream targets involved in focal adhesions and cell migration. To highlight the players, Pyk2 and FAK are

SUMOylated, and both interact with Src and paxillin. In turn, paxillin serves a scaffold/adaptor protein to coordinate MAP kinase signaling, bridging the connection between the FAK family and the MAP kinases. Moreover, our screen identified more than 13 SUMOylated MAP kinases, including ERK1, ERK2, and P38/MAPK, and A-RAF. A report by Kubota et al demonstrates both MEK1, MEK2, and MEKK1 are SUMO modified as well [52]. The majority of these SUMO modified proteins act together, physically and functionally, in the context of focal adhesions [22, 23, 41], implicating SUMO as the dynamic glue that supports and stabilizes the complex. Pyk2, Src, Paxillin, and many MAPKs contain predicted SIMs which may provide additional affinity within the assembly to efficiently coordinate signal transduction. This example illustrates so-called “higher-order signaling machines” which rely on proximity driven enzyme activation to generate signal amplification and possibly temporal spatial regulation of signal transduction [53]. Characterization of Pyk2 SUMOylation describes a novel mechanism where in SUMO modification drives amplification of autophosphorylation, interaction with Src, phosphorylation of paxillin, activation of ERK1/2, and cell migration.

Material and Methods

Pyk2 SUMOylation Validation in HeLa cells

Pyk2 was subcloned into the PCAGIG-V5 mammalian expression vector, SUMO1 and SUMO2 were subcloned into PCAGIG-MYC mammalian expression vector, and E3 ligases were subcloned into PSG5-FLAG mammalian expression vector. V5-Pyk2 and

Myc-SUMO were transfected together with and without FLAG-E3 ligases constructs with Fugene 6 (Promega) transfection reagent into HeLa cells seeded at 2×10^5 cells per well. After 48 hours, the cells were washed with phosphate buffered saline and lysed in RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% Triton-X-100) containing 0.5 mM PMSF, Roche Protease Inhibitor Cocktail, 20mM *N*-ethylmaleimide and 1% SDS, to inhibit deSUMOylation and dissociate non-covalent protein complexes. After initial lysis in 1% SDS, the lysate was diluted to achieve a 0.2% SDS concentration to accommodate immunoprecipitation. The lysate was incubated with anti-V5- agarose (Sigma-Aldrich) for two hours to immunoprecipitate the substrate. The beads were washed three times with RIPA buffer containing 0.1% SDS then incubated with 2X LDS sample buffer containing β -mercaptoethanol then heated at 100°C for 10 minutes. Samples were resolved in parallel on 4-12% NuPage Bis-Tris gels (Life Technologies) then subject to western blotting using anti-V5-HRP and anti-MYC-HRP antibodies (Life Technologies).

Pyk2-SUMO1 site mapping

4×10^7 HeLa cells were co-transfected with V5-Pyk2, MYC-SUMO1, and FLAG-PIAS1 using Fugene 6 transfection reagent (Promega). After 48 hours, cells were rinsed with warm PBS then lysed with RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% Triton-X-100) containing 0.5 mM PMSF, Roche Protease Inhibitor Cocktail, 20mM *N*-ethylmaleimide and 1% SDS. The concentrated lysate was spun down to clear debris and the supernatant was diluted with RIPA without SDS, to a final concentration of 0.2% SDS. V5-Pyk2 was immunoprecipitated with Anti-V5-agarose (Sigma Aldrich), washed three times with RIPA containing 0.1% SDS. Pyk2 was eluted using V5 peptide

(0.5 mg/ml). Eluate was incubated with anti-MYC agarose to isolate Pyk2-SUMO1. Beads were again washed three times with RIPA containing 0.1% SDS. Pyk2-SUMO1 was eluted from beads in 0.4 M ammonium bicarbonate and 8 M urea. TCA precipitation was used to concentrate the eluate and the pellet was resuspended in 8M urea. The sample was heated to 100°C for 10 minutes in LDS with β -mercaptoethanol, then fresh iodoacetamide was added to 50 mM concentration, the mixture was incubated in the dark at room temperature for 30 minutes. The sample was resolved by SDS -PAGE using a 12% NuPage Bis-Tris gel (Life Technologies) then silver stained with the SilverQuest kit (Life Technologies). Four bands corresponding to Pyk2-SUMO1 were excised from the gel, destained, and subject to tryptic digestion (1:20 trypsin/substrate ratio). The resulting peptides were separated on a Dionex Ultimate 3000 RSLCnano system (Thermo Scientific) with a 75 μ m x 15 cm Acclaim PepMap100 separating column (Thermo Scientific) protected by a 2 cm guarding column (Thermo Scientific). Mobile phase consisted of 0.1% formic acid in water (A) and 0.1% formic acid 95% acetonitrile (B). The gradient profile was set as following: 4-30% B for 40 min, 30-45% B for 10 min, 45-95% B for 10 min. MS analysis was performed using an Orbitrap Velos Pro mass spectrometer (Thermo Scientific). The spray voltage was set at 2.2 kV. Spectra ($AGC \times 10^6$) were collected from 400-1800 m/z at a resolution of 60,000 followed by data-dependent HCD MS/MS (at a resolution of 7500, collision energy 35%, activation time 0.1 ms) of the 10 most abundant ions. MS/MS spectra were searched against a human IPI reference database (V3.87) using the SEQUEST engine in Proteome Discover 1.3. Searching parameters included mass tolerance of precursor ions (± 20 ppm) and product ion (± 0.06 Da), dynamic modification of carboxyamidomethylated Cys (+ 57.0215 Da),

dynamic mass shifts for oxidized Met (+ 15.9949 Da), and dynamic modification of SUMO1 C-terminal peptide (ELGMEEEDVIEVYQEQTGG) or target peptide terminal attached to the modified K. Only b and y ions were considered during the database match.

Pyk2 *in vitro* kinase assay

GST-Pyk2 WT was purified from yeast using glutathione sepharose 4B (GE Healthcare). Following washes Pyk2 was left on beads and separated into two aliquots. One aliquot was SUMOylated under standard conditions with 90 nM E1, 300 nM E2, 5 mM ATP in 50 mM HEPES, pH 7.4, 100 mM NaCl, 10 mM MgCl₂, and 0.1 mM dithiothreitol for 1 hour at 37°C with gentle shaking. The reaction mixture was removed, the beads were washed two times with SUMOylation buffer without enzymes or ATP, a final wash was performed with kinase buffer lacking ATP. The second aliquot served as a control and was incubated with SUMOylation reaction buffer lacking enzymes, SUMO, and ATP. In the second phase of this reaction, the beads were again divided in two, creating four reaction conditions. Kinase buffer (50 mM HEPES, pH 7.4, 10 mM MgCl₂, 10 mM MnCl₂, 300 mM KCl, and 0.5% Nonidet P-40) with 1 mM cold ATP was added to one aliquot that had previously undergone SUMOylation and one aliquot that had only been incubated with buffer. The autophosphorylation reaction was allowed to proceed for 1 hour at 30°C with gentle shaking then all reactions were washed three times with kinase buffer lacking ATP. The beads were heated in 2X LDS sample buffer containing β-mercaptoethanol at 100°C for 10 minutes, then resolved by SDS-PAGE and immunoblotted with anti-GST (Millipore), anti-SUMO1 (21C7) (Matunis laboratory), and anti-Pyk2 pTyr402 (Life Technologies) antibodies.

SUMO dependent Pyk2-Src interaction

V5-tagged WT Pyk2, 4KR Pyk2, and Pyk2 Y402F were all transfected with and without MYC-SUMO1 in HeLa cells. V5-tagged WT Pyk2 and 4KR Pyk2 were also co-transfected with flag-PIAS1 and MYC-SUMO1. 4KR Pyk2 and Y402F were generated using the QuikChange II site directed mutagenesis kit (Agilent Technologies). Cells were rinsed with warm phosphate buffered saline 48 hours after transfection and lysed with Kamiya buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM MgCl₂, 1 mM EDTA, 1% Triton-X-100). The lysate was incubated with anti-V5-agarose for two hours. Beads were washed three times with TBST then boiled with 2X LDS containing β -mercaptoethanol. The immunoprecipitates were resolved by SDS-PAGE followed by immunoblotting with antibodies to Src (Cell Signaling) and paxillin (Santa Cruz). Phosphospecific antibodies for paxillin pTyr118 and Src pTyr416 (Cell Signaling) were used to assess phosphorylation status.

ERK activation Assay

HEK293 cells were transfected with Pyk2 WT and 4KR versions. SUMO was co-transfected, or transfected alone as a control, as indicated. 48 hours after transfected cells were lysed in 2X LDS with β -mercaptoethanol and boiled for 10 minutes. Lysate samples were subject to SDS-PAGE and western blotting with anti-Pyk2 (Cell Signaling), pERK1/2 (Cell Signaling), Total ERK1/2 (Santa Cruz), and pP38 (Cell Signaling) antibodies.

Wound Healing Assay

Adenovirus constructs for WT Pyk2, Pyk2 4KR, and SUMO1 were generated using ViraPower™ Adenoviral Gateway™ Expression Kit (Life Technologies) according to the

manufacturer's instructions. MDA-MB-231 breast cancer cells were either infected with Pyk2 WT or Pyk2 4KR adenovirus alone or co-infected with SUMO1 adenovirus. After 48 hours cells reached confluency and were wounded using a 10 μ l tip then washed with growth factor media and allowed to grow for three hours in a 37°C incubator. Images were acquired at time 0 hours and 3 hours.

References

- [1] Della Rocca GJ. Pleiotropic Coupling of G Protein-coupled Receptors to the Mitogen-activated Protein Kinase Cascade. ROLE OF FOCAL ADHESIONS AND RECEPTOR TYROSINE KINASES. *J Biol Chem.* 1999;274:13978-84.
- [2] London R. Identification and Characterization of a Novel Related Adhesion Focal Tyrosine Kinase (RAFTK) from Megakaryocytes and Brain. *J Biol Chem.* 1995;270:27742-51.
- [3] Li J, Avraham H, Rogers RA, Raja S, Avraham S. Characterization of RAFTK, a novel focal adhesion kinase and its integrin-dependent phosphorylation and activation in megakaryocytes. *Blood.* 1996;88:417-28.
- [4] Geiger B, Spatz JP, Bershadsky AD. Environmental sensing through focal adhesions. *Nat Rev Mol Cell Biol.* 2009;10:21-33.
- [5] Lipinski CA, Loftus JC. Targeting Pyk2 for therapeutic intervention. *Expert Opin Ther Targets.* 2010;14:95-108.
- [6] Mitra SK, Hanson DA, Schlaepfer DD. Focal adhesion kinase: in command and control of cell motility. *Nat Rev Mol Cell Biol.* 2005;6:56-68.
- [7] Avraham H, Park S-Y, Schinkmann K, Avraham S. RAFTK/Pyk2-mediated cellular signalling. *Cell Signal.* 2000;12:123-33.
- [8] Lev S, Moreno H, Martinez R, Canoll P, Peles E, Musacchio JM, et al. Protein tyrosine kinase PYK2 involved in calcium-induced regulation of ion channel and MAP kinase functions. *Nature.* 1995;376:737-45.
- [9] Avraham H. Tyrosine Phosphorylation of the Related Adhesion Focal Tyrosine Kinase in Megakaryocytes upon Stem Cell Factor and Phorbol Myristate Acetate Stimulation and Its Association with Paxillin. *J Biol Chem.* 1997;272:10804-10.
- [10] Zheng C. Differential Regulation of Pyk2 and Focal Adhesion Kinase (FAK). THE C-TERMINAL DOMAIN OF FAK CONFERS RESPONSE TO CELL ADHESION. *J Biol Chem.* 1998;273:2384-9.

- [11] Park SY, Avraham HK, Avraham S. RAFTK/Pyk2 activation is mediated by trans-acting autophosphorylation in a Src-independent manner. *J Biol Chem.* 2004;279:33315-22.
- [12] Schlaepfer DD, Hauck CR, Sieg DJ. Signaling through focal adhesion kinase. *Biophysics and Molecular Biology.* 1999;71:435-78.
- [13] Blaukat A. Adaptor Proteins Grb2 and Crk Couple Pyk2 with Activation of Specific Mitogen-activated Protein Kinase Cascades. *J Biol Chem.* 1999;274:14893-901.
- [14] Dikic I, Tokiwa G, Lev S, Courtneidge SA, Schlessinger J. A role for Pyk2 and Src in linking G-protein-couple receptors with MAP kinase activation. *Nat Aust.* 1996;383:547-50.
- [15] Avraham H. The Related Adhesion Focal Tyrosine Kinase Is Tyrosine-phosphorylated after beta 1-Integrin Stimulation in B Cells and Binds to p130cas. *J Biol Chem.* 1997;272:228-32.
- [16] Avraham S. The Related Adhesion Focal Tyrosine Kinase Forms a Complex with Paxillin in Hematopoietic Cells. *J Biol Chem.* 1996;271:31222-6.
- [17] Taniyama Y, Weber DS, Rocic P, Hilenski L, Akers ML, Park J, et al. Pyk2- and Src-Dependent Tyrosine Phosphorylation of PDK1 Regulates Focal Adhesions. *Mol Cell Biol.* 2003;23:8019-29.
- [18] Kadare G, Toutant M, Formstecher E, Corvol JC, Carnaud M, Bouterin MC, et al. PIAS1-mediated sumoylation of focal adhesion kinase activates its autophosphorylation. *J Biol Chem.* 2003;278:47434-40.
- [19] Pedrioli PG, Raught B, Zhang XD, Rogers R, Aitchison J, Matunis M, et al. Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. *Nat Methods.* 2006;3:533-9.
- [20] Riggs D, Yang Z, Kloss J, Loftus JC. The Pyk2 FERM regulates Pyk2 complex formation and phosphorylation. *Cell Signal.* 2011;23:288-96.
- [21] Nishiya N, Tachibana K, Shibanuma M, Mashimo JI, Nose K. Hic-5-reduced cell spreading on fibronectin: competitive effects between paxillin and Hic-5 through interaction with focal adhesion kinase. *Mol Cell Biol.* 2001;21:5332-45.

- [22] Ishibe S, Joly D, Zhu X, Cantley LG. Phosphorylation-dependent paxillin-ERK association mediates hepatocyte growth factor-stimulated epithelial morphogenesis. *Mol Cell*. 2003;12:1275-85.
- [23] Kolch W. Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nat Rev Mol Cell Biol*. 2005;6:827-37.
- [24] Ishibe S, Joly D, Liu ZX, Cantley LG. Paxillin serves as an ERK-regulated scaffold for coordinating FAK and Rac activation in epithelial morphogenesis. *Mol Cell*. 2004;16:257-67.
- [25] McMullen M, Keller R, Sussman M, Pumiglia K. Vascular endothelial growth factor-mediated activation of p38 is dependent upon Src and RAFTK/Pyk2. *Oncogene*. 2004;23:1275-82.
- [26] Hu CT, Cheng CC, Pan SM, Wu JR, Wu WS. PKC mediates fluctuant ERK-paxillin signaling for hepatocyte growth factor-induced migration of hepatoma cell HepG2. *Cell Signal*. 2013;25:1457-67.
- [27] Sun CK, Man K, Ng KT, Ho JW, Lim ZX, Cheng Q, et al. Proline-rich tyrosine kinase 2 (Pyk2) promotes proliferation and invasiveness of hepatocellular carcinoma cells through c-Src/ERK activation. *Carcinogenesis*. 2008;29:2096-105.
- [28] Petit V, Boyer B, Lentz D, Turner CE, Thiery JP, Valles AM. Phosphorylation of Tyrosine Residues 31 and 118 on Paxillin Regulates Cell Migration through an Association with CRK and NBT-II Cells. *The Journal of Cell Biology*. 2000;148:957-69.
- [29] Nakamura K, Yano H, Uchida H, Hashimoto S, Schaefer E, Sabe H. Tyrosine phosphorylation of paxillin alpha is involved in temporospatial regulation of paxillin-containing focal adhesion formation and F-actin organization in motile cells. *J Biol Chem*. 2000;275:27155-64.
- [30] Hasegawa H, Kiyokawa E, Tanaka S, Nagashima K, Gotoh N, Shibuya M, et al. DOCK180, a major CRK-binding protein, alters cell morphology upon translocation to the cell membrane. *Mol Cell Biol*. 1996;16:1770-6.
- [31] Kiyokawa E. Evidence That DOCK180 Up-regulates Signals from the CrkII-p130Cas Complex. *J Biol Chem*. 1998;273:24479-84.

- [32] Kiyokawa E, Hashimoto Y, Kobayashi S, Sugimura H, Kurata T, Matsuda M. Activation of Rac1 by a Crk SH3-binding protein, DOCK180. *Genes Dev.* 1998;12:3331-6.
- [33] Kelly PJ, Westwick JW, Whitehead IP, Der CJ, Parise LV. Cdc42 and Rac1 induce integrin-mediated cell motility and invasiveness through PI(3)K. *Nature.* 1997;390:632-6.
- [34] Huang C, Rajfur Z, Borchers C, Schaller MD, Jacobson K. JNK phosphorylates paxillin and regulates cell migration. *Nature.* 2003;424:219-23.
- [35] Chen HY, Shen CH, Tsai YT, Lin FC, Huang YP, Chen RH. Brk activates rac1 and promotes cell migration and invasion by phosphorylating paxillin. *Mol Cell Biol.* 2004;24:10558-72.
- [36] Fernandez-Miranda G, Perez de Castro I, Carmena M, Aguirre-Portoles C, Ruchaud S, Fant X, et al. Mitotic Kinase Aurora-B is regulated by SUMO-2/3 conjugation/deconjugation during mitosis. *J Cell Sci.* 2011;123:2823-33.
- [37] Ban R, Nishida T, Urano T. Mitotic kinase Aurora-B is regulated by SUMO-2/3 conjugation/deconjugation during mitosis. *Genes Cells.* 2011;16:652-69.
- [38] Hietakangas V, Anckar J, Blomster HA, Fujimoto M, Palvimo JJ, Nakai A, et al. PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A.* 2006;103:45-50.
- [39] Polley S, Huang DB, Hauenstein AV, Fusco AJ, Zhong X, Vu D, et al. A structural basis for I κ B kinase 2 activation via oligomerization-dependent trans auto-phosphorylation. *PLoS Biol.* 2013;11:e1001581.
- [40] Yoon S, Seger R. The extracellular signal-regulated kinase: multiple substrates regulate diverse cellular functions. *Growth Factors.* 2006;24:21-44.
- [41] Litvak V, Tian D, Shaul YD, Lev S. Targeting of PYK2 to focal adhesions as a cellular mechanism for convergence between integrins and G protein-coupled receptor signaling cascades. *J Biol Chem.* 2000;275:32736-46.

[42] Block ER, Tolino MA, Klarlund JK. Pyk2 activation triggers epidermal growth factor receptor signaling and cell motility after wounding sheets of epithelial cells. *J Biol Chem*. 2010;285:13372-9.

[43] Lipinski CA, Tran NL, Menashi E, Rohl C, Kloss J, Bay RC, et al. The Tyrosine Kinase Pyk2 Promotes Migration and Invasion of Glioma Cells. *Neoplasia*. 2005;7:435-45.

[44] Zrihan-Licht S, Fu Y, Settleman J, Schinkmann K, Shaw L, Keydar I, et al. RAFTK/Pyk2 tyrosine kinase mediates the association of p190 RhoGAP with RasGAP and is involved in breast cancer cell invasion. *Oncogene*. 2000;19:1318-28.

[45] Tse KW, Dang-Lawson M, Lee RL, Vong D, Bulic A, Buckbinder L, et al. B cell receptor-induced phosphorylation of Pyk2 and focal adhesion kinase involves integrins and the Rap GTPases and is required for B cell spreading. *J Biol Chem*. 2009;284:22865-77.

[46] Nobes CD, Alan H. Rho GTPases Control Polarity, Protrusion, and Adhesion during Cell Movement. *The Journal of Cell Biology*. 1999;144:1235-44.

[47] Castillo-Lluva S, Tatham MH, Jones RC, Jaffray EG, Edmondson RD, Hay RT, et al. SUMOylation of the GTPase Rac1 is required for optimal cell migration. *Nat Cell Biol*. 2010;12:1078-85.

[48] Psakhye I, Jentsch S. Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell*. 2012;151:807-20.

[49] Silver HR, Nissley JA, Reed SH, Hou YM, Johnson ES. A role for SUMO in nucleotide excision repair. *DNA Repair (Amst)*. 2011;10:1243-51.

[50] Finkbeiner E, Haindl M, Muller S. The SUMO system controls nucleolar partitioning of a novel mammalian ribosome biogenesis complex. *EMBO J*. 2011;30:1067-78.

[51] Finkbeiner E, Haindl M, Raman N, Muller S. SUMO routes ribosome maturation. *Nucleus*. 2011;2.

[52] Kubota Y, O'Grady P, Saito H, Takekawa M. Oncogenic Ras abrogates MEK SUMOylation that suppresses the ERK pathway and cell transformation. *Nat Cell Biol.* 2011;13:282-91.

[53] Wu H. Higher-order assemblies in a new paradigm of signal transduction. *Cell.* 2013;153:287-9

CURRICULUM VITAE FOR Ph.D. CANDIDATES

The Johns Hopkins University School of Medicine

Ijeoma K. Uzoma

March 24, 2014

Educational History

Ph.D. expected	2014	Program in the Department of Pharmacology and Molecular Sciences Mentor: Dr. Heng Zhu	Johns Hopkins University School of Medicine
B.A.	2006	Biology	Tufts University

Other Professional Experience

Post Baccalaureate Research Fellow	2007	Department of Pharmacology Mentor: Dr. David L. Rimm	Yale University School of Medicine
------------------------------------	------	---	------------------------------------

Scholarships and Fellowships

(2011-2013) NIH Ruth L. Kirschstein National Research Service Awards for Individual Predoctoral Fellows (F31) NIGMS

(2012) Keystone Symposia Scholarship- Proteomics, Interactomes, Stockholm, Sweden

Publications

Uzoma I. Characterizing SUMO E3 Ligase substrate specificity and the significance of SUMO modification on mechanisms of Focal Adhesion Kinase signaling. Manuscript in Preparation. 2014.

Uzoma I, Zhu H. Interactome mapping: using protein microarray technology to reconstruct diverse protein networks. Genomics Proteomics Bioinformatics. 2013 Feb;11(1):18-28.

Horst B, Gruvberger-Saal SK, Hopkins BD, Bordone L, Yang Y, Chernoff KA, **Uzoma I,** et al. Gab2-mediated signaling promotes melanoma metastasis. Am J Pathol. 2009 Apr;174(4):1524-33.

Posters

Uzoma I, Cox E, et al. 2012 Constructing Global SUMOylation Networks. Poster Presentation, Keystone Symposia Proteomics, Interactomes. Stockholm, Sweden. May 2012.

Uzoma I, Cox E, et al. 2011. Globally constructing SUMOylation networks utilizing protein microarrays. Poster Presentation, 10th HUPO World Congress. Geneva, Switzerland. September 2011.

Uzoma I, Cox E, et al. 2011. A Protein Microarray-Based Approach To Globally Construct SUMOylation Networks. Poster Presentation, Cold Spring Harbor Meeting - The Ubiquitin Family. Cold Spring Harbor, NY. May 2011.

Uzoma I, Cox E, et al. 2009. Global Analysis of SUMO E3 Ligase Substrates. Poster Presentation, National Technology Centers for Networks and Pathways — All Hands Meeting. Bethesda, MD. March 2010.

Uzoma I, Cox E, et al. Identification of E3 Ligase Dependent SUMO Substrates. Poster Presentation, International Conference on SUMO, Ubiquitin, UBL Proteins: Implications for Human Diseases. Houston, TX. February 2010.

Service and Leadership

(2012-2013) Henrietta Lacks Dunbar Health Science Scholarship Selection Committee/Mentor- Evaluated applications and interviewed Dunbar High School scholarship nominees. Served as a mentor for first Henrietta Lacks Scholarship recipient, focusing on college applications, SAT preparation, and general preparation for health science careers

(2012-2013) Initiated Danger-Avoidance Training Workshops and coursework for students and post docs at the Johns Hopkins Medical Institute. Secured funding from numerous groups to subsidize cost for student body

(2009-2013) Graduate Student Peer Mentor- Created School of Medicine-wide seminars on “How to choose a lab” and “How to prepare for Orals” for graduate students. Also provided confidential one-on-one mentoring for graduate student peers