Dissection of spatiotemporal intracellular signaling using engineered chemical and genetic tools

Emily M Dieter

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Dustin J. Maly, Chair

Champak Chatterjee

Jesse Zalatan

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Dissection of spatiotemporal intracellular signaling using engineered chemical and genetic tools

Emily M Dieter

Chair of the Supervisory Committee:
Professor Dustin J. Maly
Chemistry

Cellular signaling proteins are responsible for transmitting intracellular signals and enacting the appropriate response to received stimuli. These messages are propagated through networks that rely on allosteric regulation, post-translational modifications, or other mechanisms for modulating signaling protein activity. Scientists seeking to engineer systems that mimic endogenous pathways must be able to control and precisely tune the spatiotemporal activity of the proteins involved in these complex networks. This thesis describes the development of chemical genetic tools that specifically control individual signaling nodes in a cell using three distinct approaches. The first approach utilizes conformation-selective inhibitors that modulate the allosteric regulation of Src Family Kinases (SFKs). SFKs are multi-domain kinases that transmit

cellular signals through their catalytic and non-catalytic activities. Although dysregulation of kinases is often implicated in human diseases, studying individual kinases is difficult due to their conserved catalytic domain. By combining genetics with small molecule probe development, we created potent and specific kinase inhibitors that are capable of allosterically controlling the intramolecular regulation of an individual kinase. We applied this strategy to the SFKs and studied the phenotypic effects that resulted from conformation-selective inhibition. The second approach employed a small molecule-controlled, genetically-encoded rheostatic switch for studying signaling mediated by localized pools of RAS. In the final approach, a second-generation chemically-disrupted proximity (CDP) system was developed. This optimized CDP system can be incorporated into numerous chemical genetic systems that rely on intramolecular and intermolecular regulation. Together, these three approaches represent a broad platform for interrogating and engineering diverse cellular processes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Å	**ANGSTROMS**
ABL	**ABELSON MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1**
ATP	**ADENOSINE TRI-PHOSPHATE**
CD	**CATALYTIC DOMAIN**
CYS	**CYSTEINE**
DAPI	**4',6-DIAMIDINO-2-PHENYLINDOLE**
DFG	**ASPARTATE-PHENYLALANINE-GLYCINE**
DMS	**DEEP MUTATIONAL SCANNING**
DMSO	**DIMETHYLSULFOXIDE**
FYN	**PROTO-ONCOGENE TYROSINE PROTEIN KINASE FYN**
GFP	**GREEN FLUORESCENT PROTEIN**
HCK	**TYROSINE-PROTEIN KINASE HCK**
PDB	**PROTEIN DATA BANK**
RHO	**RAS HOMOLOG FAMILY MEMBER**
ROCK	**RHO-ASSOCIATED COILED-COILED CONTAINING PROTEIN KINASE 1**
RSK2-CTD	**P90 RIBOSOMAL PROTEIN S6 KINASE**
SFK	**SRC FAMILY KINASE**
SRC	**PROTO-ONCOGENE TYROSINE PROTEIN KINASE SRC**
SRC-WT	**SRC-WILD TYPE**
SRC-CYS	**SRC-SENSITIZED CYSTEINE MUTANT**
SYF	**SRC/YES/FYN KNOCKOUT CELLS**
UAS	**UPSTREAM ACTIVATING SEQUENCE**
YES	**PROTO-ONCOGENE TYROSINE PROTEIN KINASE YES**

ACKNOWLEDGEMENTS

If you are reading this, it is probably because you were instrumental to this work being accomplished. I am going to do my best to thank you, but please know that I fully expect my words to fall short. I hope you realize that without you, none of this would have been possible. Thank you for being a part of my life.

First, I would like to thank Dusty for his guidance, patience, knowledge, and support. Thank you for helping me to become the scientist I am today. The members of the Maly Lab, past and present, have creating a fun, supportive work environment which has allowed me to flourish. Dr. Ames Register, Dr. Carrie Gower, Dr. Hannah Feldman, and Dr. Dan Cunningham-Bryant accepted me as a trivia team member when I invited myself along so many years ago, and have continued to hang out with me, even though they have all moved on to bigger and better things. Don't worry, you are never going to get rid of me now. Dr. Chloe Lombard, Almost-Dr. Sujata Chakraborty, and Almost-Dr. Linglan Fang were my Maly Lab cohorts, and I am so glad I got to go through this entire process with you all. I would also like to thank Cindy Wei and Zack Potter for the many many many office chats, and finally thank you to Kyler Radmall for being my desk-mate and continuing work on the ANR project.

Thank you to the wonderful faculty and staff at the University of Washington. My committee especially, comprised of Dr. Jesse Zalatan, Dr. Champak Chatterjee, Dr. Ning Zheng, and Dr. Daniel Chiu, has always been there to provide feedback and moral support, even if it is just a friendly conversation as we pass in the hallway. Dr. Ethan Merritt single-handedly taught me X-Ray crystallography. It was a pleasure to learn from you and thank you for always having my back. Dr. Doug Fowler and Dr. Ethan Ahler collaborated with us during the SFK project, and their expertise was invaluable for crafting our final story.

I would also like to thank the Chemistry, Biology, and Physics departments at Saint Michael's College, specifically Dr. Shane Lamos, Jenn Paone-Vogt, and Dr. Alain Brizard. I would not have pursued grad school without your support and guidance.

Thank you to my family and friends outside of grad school who have helped to keep me sane throughout this entire process. Thank you to Maike and Walter for always having an open door, a fridge fully stocked with beer, and many episodes of Jeopardy to watch. Heidi and Chad Chapman have been especially helpful these past few months, whether it is an extra hand to help move, or a couch to crash on to watch a Lifetime movie together. Dr. Glenna, Almost-Dr. Dillon, and Dingo Foight, literally opened their door and allowed me to live in their attic for the past few months. Thank you for cooking me dinner every night, for listening to my complaints, for teaching me how to complete the Times crossword, and letting me cuddle with your dog. Thank you for letting me be a part of your family, I am going to miss you all so much.

To my parents and my brothers, thank you for your love and support. Even though we are spread across the country, I know that no matter when I call, someone will pick up. From you, I have learned that family always comes first, and that has been a constant source of reassurance in my life. Thank you for your encouragement and for being my biggest cheerleaders. I hope you know that no matter how far away I am, I am always wishing we were closer together.

Lastly, thank you to Tyler Davis. Thank you for being my partner, my roommate, my travel buddy, my baking companion, and most importantly my best friend. Being apart from you has been so rough, and I cannot wait to join you down under. I love you.

And now reader, you are most likely wondering if I cried the while writing this.

Yes, yes I did.

DEDICATION

To Mother, Father, Christopher, and Gregory. I could not have done this without you. I love you.

INTRODUCTION

Cells are constantly responding to received external stimuli, whether it be environmental or autologous. In order to effectively respond to these signals, the cell must be able to react in real-time and provide an appropriately proportional and temporal response. Many times, the responses stem from the complex signaling networks that have evolved in a cell, which propagate these signals and ensure the correct response. Due to the complexities of these networks, researchers have struggled both to understand the role of individual proteins, as well as reproduce these natural signaling pathways in a controlled manner.

My thesis work in Dr. Dustin Maly's lab has focused on the development and utilization of tools to specifically activate individual signaling nodes in a cell. Using these tools, developed by myself and others in the Maly group, I have investigated signal transduction networks in biologically relevant environments, gaining insight into the spatio-temporal roles of individual proteins in disease-relevant signaling networks.

In ***Chapter 1: A Chemical Genetic Method for Kinase Inhibition***, I undertake an investigation of intramolecular regulatory mechanisms in Src family kinases (SFKs) using conformation-selective inhibitors. Kinases are an enzyme family that are responsible for regulating and propagating cellular signals through their catalytic and non-catalytic activities. Thus, kinases are some of the most commonly dysregulated proteins associated with human disease. Due to the structural similarity between human kinases, targeting individual kinases with small molecule inhibitors has proven to be challenging. In this chapter, I describe a method that utilizes selective small-molecule inhibitors with a genetic sensitization strategy to specifically target SFKs and study the cellular consequences of their conformation-selective inhibition.

In ***Chapter 2: Interrogation of localized RAS-mediated signaling using a chemically induced activator of RAS***, I describe a chemically-inducible activator of RAS, and use this technology to interrogate the signaling networks mediated by different pools of localized RAS. RAS GTPases are one of the most frequently mutated genes in human cancers, and although they have been a major focus of research for over forty years, basic questions about RAS biology still remain due to the lack of tools that allow spatiotemporal control over RAS activation. To remedy this problem, we created a rheostatic activator of RAS that can be specifically localized to different sub-populations of RAS. This localized RAS activator allowed us to investigate the dynamics of localized RAS activation. Using global and targeted phospho-proteomics, I was able to dissect the signaling networks that are activated upon activation of different localized pools of RAS.

In ***Chapter 3: Development of a Mammalian Reporter System for the Optimization of a Bio-Orthogonal Chemical Disruptor of Proximity***, I introduce our second-generation chemically-disrupted proximity (CDP) system based on the interaction between the viral protease NS3a and a peptide inhibitor. Although our recently-reported, first-generation CDP system was capable of conferring small molecule control over a number of intermolecularly and intramolecularly-regulated systems, we found that it failed in certain applications. To combat this problem, I set out to further optimize the protein/peptide interaction that is central to this system in mammalian cells. I engineered a transcriptional reporter cell line and tested a single-variant peptide library in it. Hits from the library were collected, and next-generation sequencing was used to obtain peptides that more efficiently activated transcriptional outputs. Individual peptides from the library were generated and tested, revealing an optimized NS3a/peptide pair. These efforts resulted in an optimized peptide that can be used in wide-ranging applications with NS3a.

Chapter 1. A CHEMICAL GENETIC METHOD FOR KINASE INHIBITION

1.1 INTRODUCTION

Cells are constantly receiving environmental signals that can result in diverse cellular processes such as growth, proliferation, differentiation, or apoptosis. These signals are generally received as external stimuli at the plasma membrane and must be transduced to assorted subcellular locations. Cells have developed signaling networks to propagate these messages from the plasma membrane to their final destination of action. Although many different types of proteins are involved in these signaling cascades, one of the major players is a family of enzymes known as protein kinases.

There are over 500 human kinases, which are responsible for regulating intracellular signaling. On a structural level, many kinases contain non-catalytic regulatory domains that can be allosterically linked to the conserved catalytic domain (CD) shared by all kinases. The CD is composed of an N-terminal lobe that contains the αC-helix, a C-terminal lobe made up of β-strands, and an ATP-binding site between the two lobes (**Figure 1.1a**)[1,2]. Kinases are phosphotransferases, catalyzing the transfer of ATP's gamma phosphate to a substrate, which occurs in the ATP-binding site. The phosphorylation of substrates is a post-translational modification that serves as a regulatory mechanism for many signaling pathways.

Crystal structures of kinases have shown the ATP binding site can exist in at least three distinct conformations: active, αC-helix out, or Asp-Phe-Gly (DFG)-out. In the active conformation, which is necessary for catalysis to occur, key catalytic residues, are aligned (**Figure 1.1b**)[2]. The αC-helix out and DFG-out are inactive conformations, where either the αC-helix or

DFG motif are displaced from their active conformations, preventing catalysis from occurring (**Figure 1.1c, 1.1d**)[3]. Although this displacement prevents ATP catalysis, the kinase can still partake in non-catalytic functions, such as scaffolding, substrate recruitment/recognition, and DNA-binding.

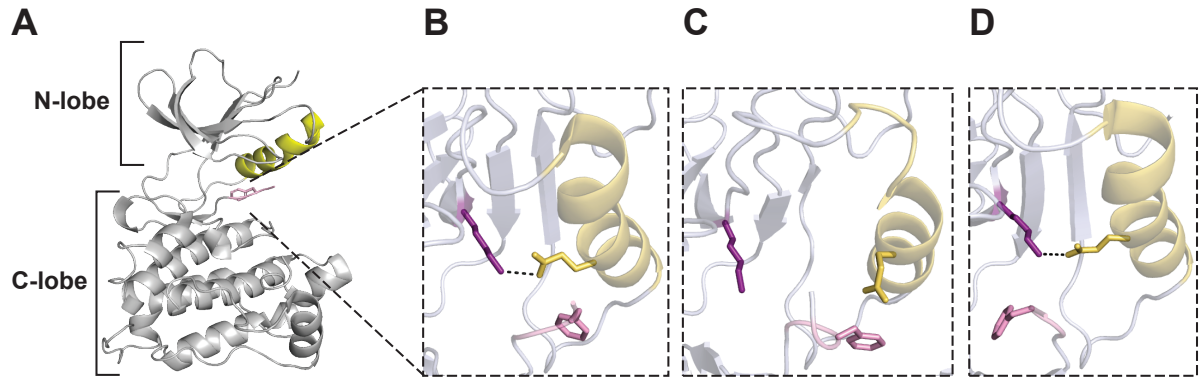

Figure 1.1. ***Catalytic domain structure and ATP-binding site conformations of protein kinases***. (A) The catalytic domain of Src, a canonical protein kinase (PDB ID: 5SYS). The αC-helix is shown in yellow while the DFG motif is shown in pink. (B) Active ATP-binding site conformation, where all key catalytic residues are aligned. (C) αC-helix out ATP-binding site conformation, where the αC-helix is rotated outward, disrupting the salt bridge between the catalytic lysine in the N-lobe and a glutamic acid in the αC-helix. (D) DFG-out ATP-binding site conformation, where the DFG motif is flipped 180º, but the salt bridge between the CD and the αC-helix is maintained.

One example of kinases with known catalytic and non-catalytic functions are the Src Family Kinases (SFKs), a group of non-receptor tyrosine kinases[4]. SFKs regulate cellular processes such as proliferation, growth, survival, motility, and adhesion[5]. As a result, mis-regulation of SFKs has been linked to various illnesses and diseases, making them an enticing drug target. Yet, even though SFKs are some of the most well-studied kinases, questions still remain about their phosphotransferase-independent function, such as what are the roles of some of the less understood regulatory domains? How do the regulatory domains coordinate control of the catalytic domain? To what degree does each domain contribute to intramolecular regulation?

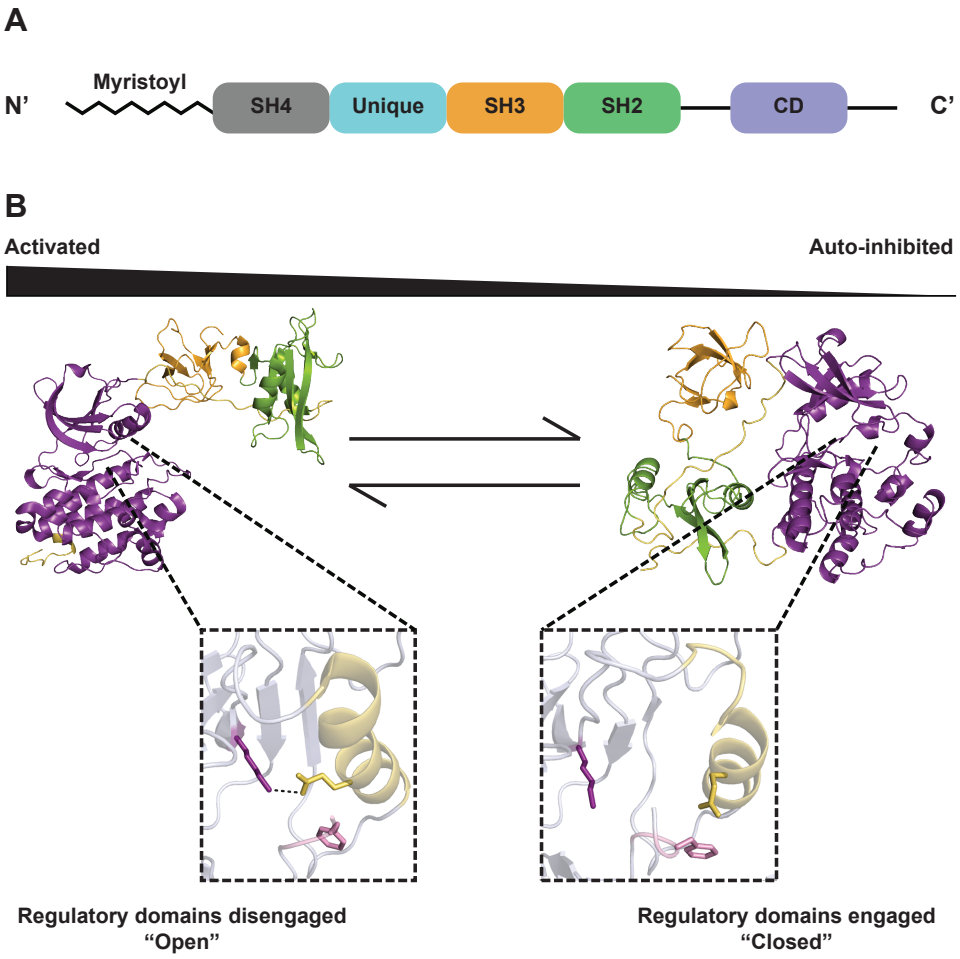

Figure 1.2. ***Structure and conformations of Src family kinases.*** (A) Linear schematic of Src kinase. (B) Crystal structures of the global conformation of Src. On the left, Src is shown in an open global conformation (PDB ID: 1Y57). The regulatory domains are disengaged, and the ATP-binding site is in the active conformation, resulting in a catalytically competent kinase. On the right, the regulatory domains are engaged with the catalytic domain, and the ATP-binding site is in the αC-helix out conformation, resulting in an auto-inhibited kinase.

SFKs are composed of an N-terminal myristoylation site and regulatory domains, followed by a C-terminal catalytic domain (**Figure 1.2a**). These domains are coupled through a bidirectional, allosteric relationship (**Figure 1.2b**)[4]. This allosteric relationship has been exploited in order to better understand intramolecular regulation of SFKs and other kinases through the use of ATP-competitive inhibitors. By stabilizing the ATP-binding site in a specific conformation, the

regulatory domains can also be biased to be more engaged or disengaged from the CD. Our group, and others, have previously used conformation selective ATP-competitive probes *in vitro* to investigate the allosteric relationship of SFKs[6-9]. Stabilization of the ATP-binding site in the inactive αC-helix out conformation strengthens intramolecular engagements, resulting in a globally closed kinase conformation. Stabilization of the active ATP-binding site conformation weakens intramolecular engagements, resulting in a globally open kinase conformation[7].

The obtainment of divergent phenotypes due to differences in kinase conformation has been speculated for a number of kinases but isolating the actions of a single kinase has proven challenging due to the conserved nature of the kinase catalytic domain[3,10,11]. Conventional methods, such as knocking out the kinase, completely removes the kinase from the system, making it impossible to separate catalytic and non-catalytic functions. Another commonly used method is to mutate the kinase. In this method, the protein is still present in the system, but mutations can have unknown effects, such as disruption of protein folding and intramolecular interactions, which interrupt the native functions. An alternative method is to use ATP-competitive inhibitors that stabilize specific active site conformations. By using probes that stabilize certain ATP-binding site conformations, allosteric coupling between the ATP-binding site and distal domains can be manipulated. Yet, because of the conserved overall structure of CD throughout the kinome, this method has been of limited use in cellular studies because it is difficult to identify small molecules that inhibit a single kinase of interest.

To get around the problem of catalytic domain homology, we have created a chemical genetic method to exclusively target an individual kinase. We characterized this method using SFKs as a model kinase, but this is a generalizable strategy that can be used to investigate non-

catalytic functions of kinases across the kinome. Utilization of this method ultimately resulted in discovering a novel intramolecular regulatory interaction in SFKs.

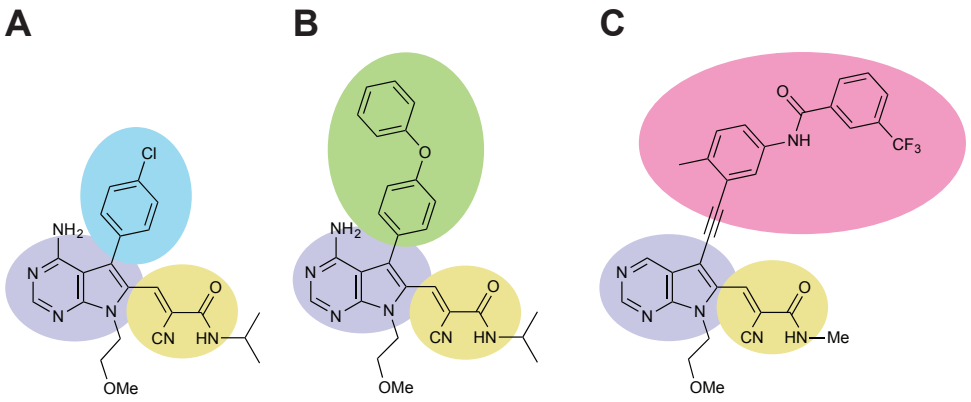

Figure 1.3. ***Conformation-selective probes.*** (A-C) Chemical structure of **1** (A), **2** (B), or **3** (C). The purple circle indicates the pyrrolopyrimidine scaffold, while the yellow region is the Michael acceptor that will covalently bind to the kinase. The variable colored regions indicate a substituent that will stabilize a specific ATP-binding site conformation.

1.2 RESULTS

1.2.1 *Probe Design*

To study kinase non-catalytic functions, ATP-competitive conformation selective probes were used in conjunction with a functionally neutral mutation introduced into a kinase of interest. The neutral mutation is introduced to the active site of a target kinase, sensitizing the kinase to electrophile-containing probes that stabilize one of three ATP-binding site conformations. Probes **1**, **2**, or **3** were developed and predicted to stabilize the active, αC-helix out, or DFG-out ATP-binding site conformation, respectively, based on previous literature (**Figure 1.3a, 1.3b, 1.3c**). Each probe scaffold consists of a pyrrolopyrimidine, which is an adenine mimetic, that participates in hydrogen-bonding between the kinase hinge region and the small molecule in the ATP-binding site. Different substituents were installed at the C-5 position of the pyrrolopyrimidine scaffold to stabilize each of the unique ATP-binding site conformations shown in **Figure 1.1**. **1** is an analog

of PP2, a well-known kinase inhibitor that has a small aryl group that stabilizes the active ATP-binding site conformation (**Figure 1.3a**)[12,13]. **2** has a larger hydrophobic group that had previously been shown to force the αC-helix to rotate, disrupting the salt bridge between the lysine and the glutamic acid, and ultimately stabilizing the αC-helix out conformation (**Figure 1.3b**)[6]. We

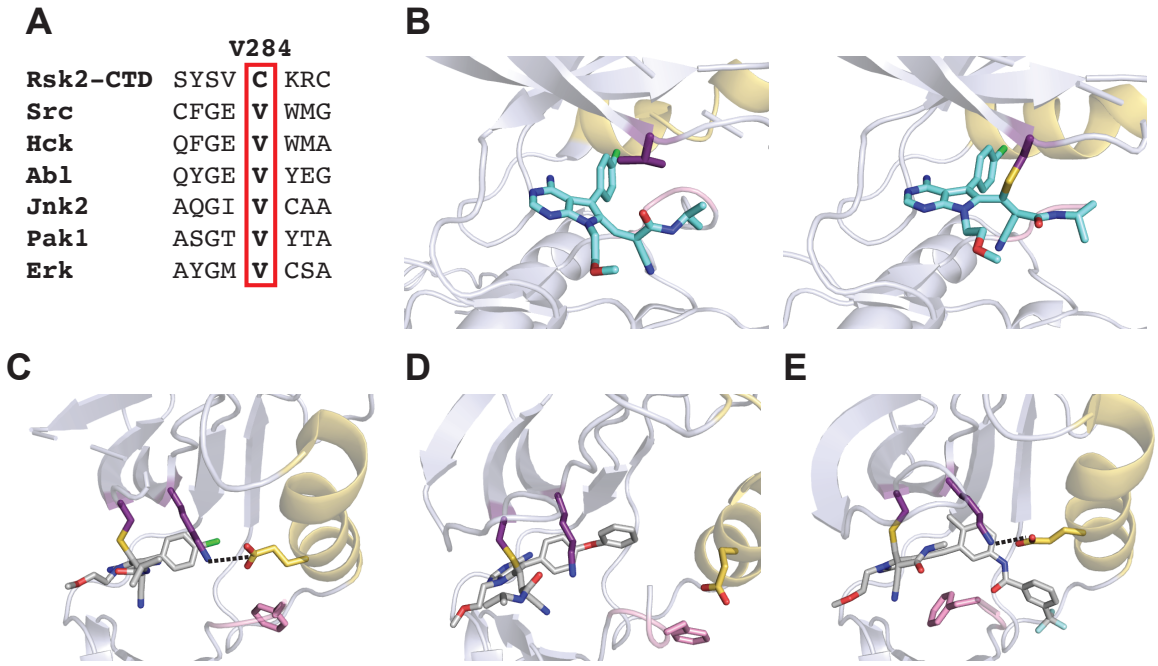

Figure 1.4 ***Conformation-selective probes bound to sensitized Src.*** (A) Sequence alignment of the naturally occurring cysteine in the active site of Rsk2-CTD compared to various other kinases. (B) Src-WT^{CD} (left) and Src-Cys^{CD} (right) bound to **1**. Only in the sensitized kinase is there a covalent bond between the kinase and the probe. (C-D) Crystal structures of Src-Cys^{CD} bound to **1** (C, PDB: 5SWH), **2** (D, PDB: 5TEH), or **3** (E, PDB: 5SYS). The αC-helix is shown in yellow, the DFG motif is shown in pink, and the catalytic lysine and mutated cysteine are shown in purple

predicted that **3** will extend into the DFG pocket while making a hydrogen-bonding interaction with a glutamic acid in the αC helix, stabilizing the DFG-out conformation (**Figure 1.3c**).

In addition to these C-5 substituents, we coupled the conformation-selective strategy to a chemical genetic strategy, which was inspired by work done by Dr. Jack Taunton's group[14]. The Taunton group studied RSK2, a kinase that has a naturally occurring cysteine in its ATP-binding pocket. Electrophile-containing probes were designed that covalently and reversibly bind to the

active site cysteine in RSK2. Sequence alignments showed that most kinases had a valine at this position that was not necessary for catalytic activity (**Figure 1.4a**). We mutated this valine to a cysteine, creating probe-sensitized kinases, and installed a cyanoacrylamide electrophile in our conformation-selective probes. These two components allow for a reversible covalent bond to form between the kinase and our conformation-selective probes (**Figure 1.4b**).

To confirm the covalent nature of the kinase-probe bond, and that the probes stabilized the predicted conformations, crystal structures of the wild type (Src-WT) c-Src catalytic domain (Src-WT^{CD}) with **1** or the sensitized c-Src catalytic domain (Src-Cys^{CD}) with **1**, **2**, or **3** were obtained. In each sensitized Src structure (PDB code: 5SWH(**1**), 5TEH(**2**), and 5SYS(**3**)), there is continuous electron density linking the installed cysteine residue to the reactive carbon in each probe (**Figure 1.4b**). In contrast, the Src-WT structure (PDB code: 5T0P) contains no electron density between probe **1** and the kinase, indicating the lack of a covalent bond (**Figure 1.4b**). In each co-crystal, the pyrrolopyrimidine scaffold is stabilized by at least one hydrogen bond in the hinge region that links the two kinase lobes, and each probe stabilizes the predicted conformation.

In structure 5SWH, (Src-Cys^{CD} bound to **1**), the salt bridge between the catalytic lysine and the glutamic acid is maintained, thus the kinase is in the active conformation (**Figure 1.4c**). In structure 5TEH (Src-Cys^{CD} bound to **2**), the αC-helix rotates 8 Å to accommodate the probe. This rotation disrupts the salt bridge between the catalytic lysine and the glutamic acid in the αC-helix, resulting in the inactive αC-helix out conformation (**Figure 1.4d**). In structure 5SYS (Src-Cys^{CD} bound to **3)**, there is only one hydrogen bond between the pyrrolopyrimidine ring of the probe and the peptide backbone of the kinase, but there are additional hydrogen bonds between the probe and the αC-helix and the DFG-motif (**Figure 1.4e**). The hydrogen bond between the amide linker in the probe and a glutamic acid in the αC-helix stabilizes the helix in the "in" position. Yet due to

the extension of the probe into the DFG pocket, stabilized by a hydrogen bond between the carbonyl of the probe and the peptide backbone of the aspartic acid, the DFG-motif is forced to undergo a 180º rotation, ultimately stabilizing the inactive DFG-out conformation.

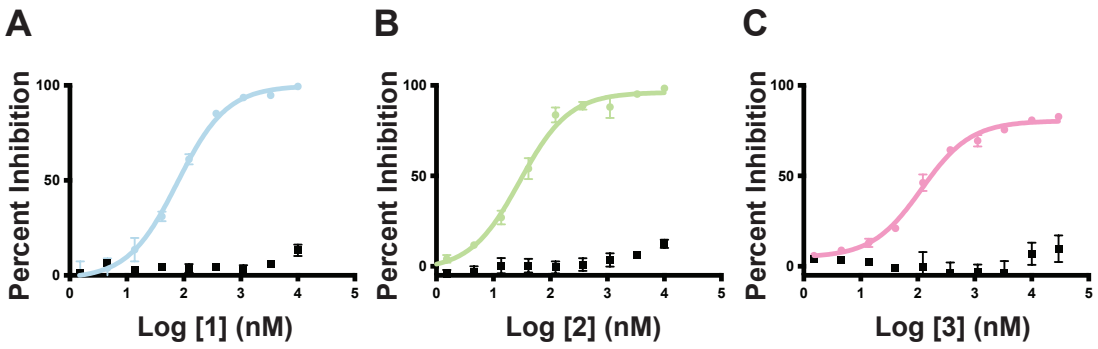

Figure 1.5 ***Ability of conformation-selective probes to target sensitized kinases.*** (A) Titration of **1** against the catalytic domain of Hck-WT (black line) or Hck-Cys (blue line). (B) Titration of **2** against the catalytic domain of Hck-WT (black line) or Hck-Cys (green line). (C) Titration of **3** against the catalytic domain of Hck-WT (black line) or Hck-Cys (pink line).

1.2.2 *Probe Validation*

The wild type or sensitized catalytic domains of kinases known to have non-catalytic functions were recombinantly expressed and purified. The activity of these kinases was assayed using radioactive ATP in the presence of either **1**, **2**, or **3** to determine the selectivity and potency of the probes. For all tested kinases, the sensitized mutants were potently inhibited by the probes, while no significant binding to the wild type kinases were observed, indicating selectivity for the sensitized kinase over its wild type counterpart (**Figure 1.5a, 1.5b, 1.5c Table 1.1**).

Next, selectivity for our sensitized kinases from a pool of endogenous kinases was tested in mammalian cells[15]. HEK293T cells were generated that stably and inducibly express either wild type Hck or sensitized Hck, a member of the SFK family, then grown in either heavy or light media. Following lysis cells grown in light media were incubated with DMSO, while cells grown in heavy media were incubated with **1**, **2**, or **3**. After drug incubation, the cells were incubated with

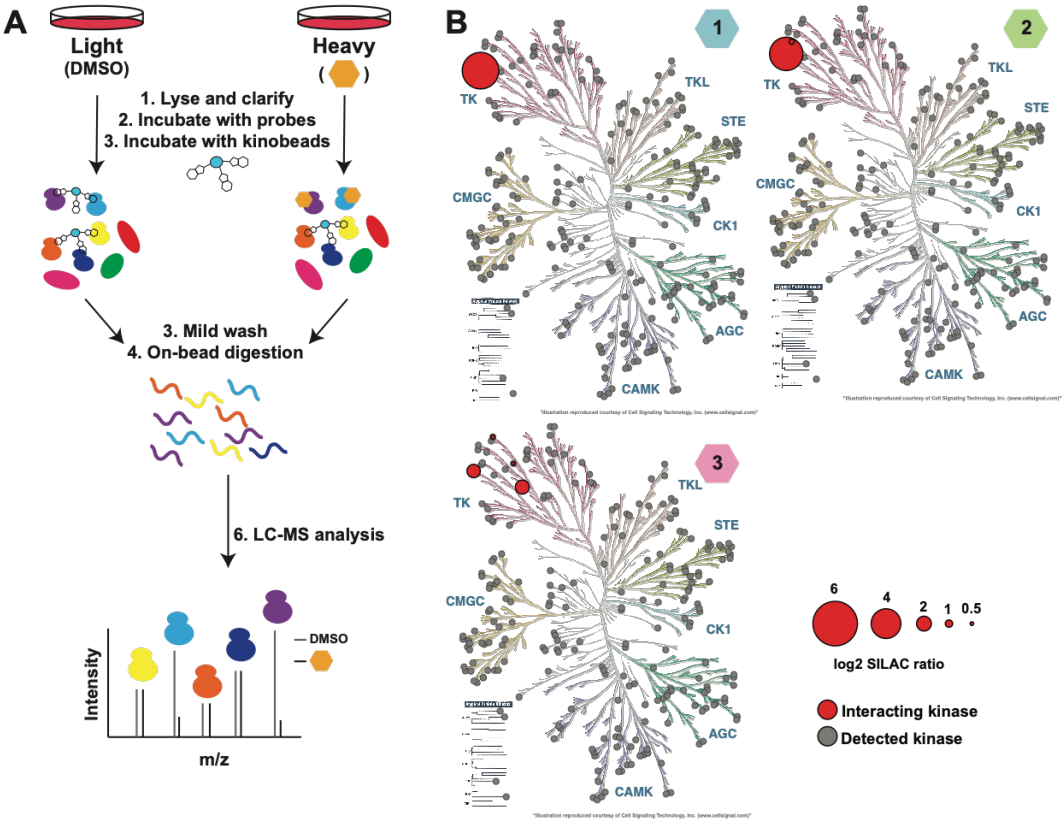

Figure 1.6 ***Kinome profiling of conformation-selective probes to target sensitized kinases.***(A) Schematic of kinobead pulldown. Cells were grown in either heavy or light labeled SILAC media, lysed, then treated with DMSO or a conformation-selective probe (colored hexagon). Following drug incubation, heavy/light labels were combined, and the lysates were incubated with kinobeads, digested, then subjected to LC-MS analysis. If a probe was bound to a kinase, it would prevent the kinase from being enriched by the kinobead matrix, resulting in less quantification of the kinase in probe-treated samples. (B) Results from kinobead pulldown experiments.

kinobeads, which are sepharose beads conjugated to a mixture of seven non-specific ATP-competitive kinase inhibitors. These beads have been shown to pull down over 300 kinases (**Figure 1.6a**)[15]. After incubation unbound lysate is washed away, while anything bound to the kinobeads was trypsinized and subjected to MS/MS analysis. For all inhibitors, ~95% depletion of Hck-Cys was observed in the heavy channel compared to the light channel (Figure 1.6b). This indicates that when the electrophilic probes are present, they engage Hck-Cys, preventing the kinase from being

enriched by non-specific inhibitors in the kinobeads. Although EPHA2 was observed as a significant off-target from the ~220 kinases profiled for probe **3**, Hck-Cys was the most depleted in all experiments. These experiments demonstrate that in a pool of endogenous kinases, high selectivity for a sensitized kinase can be obtained.

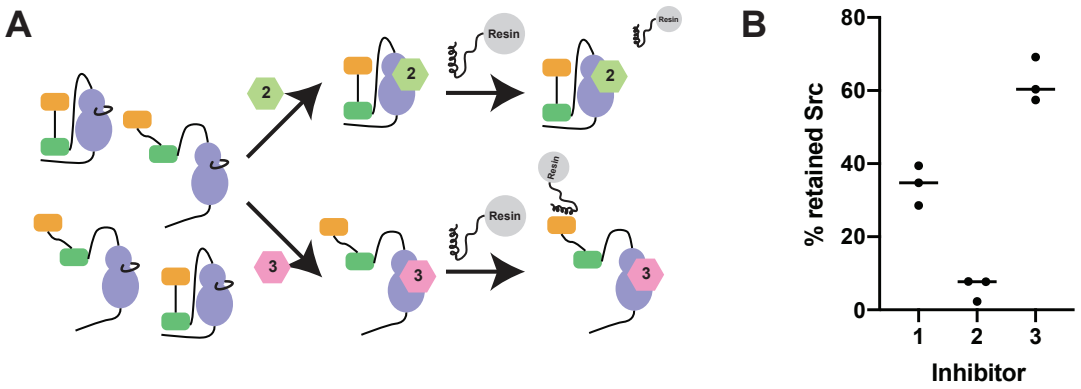

Figure 1.7 ***SH3 pulldown reveals conformation preferences of Src treated with 1, 2, or 3.***(A) ematic of SH3 pulldown. Cells stably expressing flag-tagged Src-Cys were treated with **1**, **2**, or **3**, then incubated with a matrix derivatized with a peptide that binds the SH3 domain of Src. Matrix-bound Src was eluted and analyzed. If Src is in a globally open conformation, it is expected to enriched by the matrix, where when Src is in a globally-closed conformation it will not be retained by the matrix. (B) Results of the SH3 pulldown for cells treated with probes **1**, **2**, or **3**.

Additionally, biochemical assays were performed to confirm that the expected global kinase conformations were stabilized in the presence of the probes. Flag-tagged Src-Cys was stably overexpressed in HEK293T cells then treated with **1**, **2** or **3**. Following treatment, cells were lysed and then incubated with a resin functionalized with a peptide that binds to the SH3 domain of Src (**Figure 1.7a**). If Src is stabilized in a globally open conformation, the SH3-binding peptide would have greater accessibility to the SH3 domain, resulting in greater pulldown by the peptide. Intermediate amounts of Src were pulled down following the incubation of cells with probe **1**, and virtually no Src was pulled down upon incubation of cells with **2**. Incubation of cells with probe **3** resulted in a 5-fold increase in the amount of Src pulled down compared to cells treated with **2** and a 2-fold increase compared to cells treated with **1** (**Figure 1.7b**). These results establish that the

regulatory domains are disengaged from the catalytic domain when the ATP-binding site is stabilized in a DFG-out conformation by probe **3**, indicating a globally open conformation. This is thought to stem from a destabilization of the SH3 and SH2-catalytic domain linker interaction, originating from the stabilized DFG-out conformation[7,8].

1.2.3 *Investigation of the SFK N-terminal tail accessibility*

We next sought to characterize the effect of conformation selective probes on the SFK SH4 domain. SFKs are co-translationally myristoylated, which, along with either a poly-basic cluster of amino acids or palmitoylation in the SH4 domain, allow membrane binding[4]. We hypothesized that if the SH4 domain was allosterically linked to the catalytic domain, our conformation-selective probes would modulate the accessibility of the SH4 domain, which could be visualized through changes in localization.

Src/Yes/Fyn knockout cells (SYFs) were transiently transfected with either Src-WT or Src-Cys that were c-terminally GFP tagged[16,17]. 24 hours post-transfection, the cells were treated with probe or DMSO for 15 minutes, then fixed and stained with DAPI (a nucleus stain) and wheat germ agglutinin conjugated to Alexafluor 647 (WGA, a plasma membrane stain) (**Figure 1.8a**). Incubation with DMSO or **2** lead to diffuse localization of Src-Cys throughout the cell (**Figure 1.8b**). Interestingly, incubation with **3** resulted in a re-localization of Src-Cys primarily to the plasma membrane or the perinuclear region and initiated a blebbing phenotype in the cells, indicating that the SH4 domain acts as a regulatory element (**Figure 1.8b, 1.8c**). When the kinase is treated with **2**, the ATP-binding site is stabilized in the αC-helix out conformation and overall closed conformation, which extends to the SH4 domain. In the closed conformation, the SH4 domain is also intramolecularly engaged, preventing the kinase from interacting with the membrane. When the kinase is stabilized in the DFG-out ATP-binding site conformation, resulting

in an overall open conformation, the SH4 domain is disengaged and free to interact with the membrane, resulting in the change of localization.

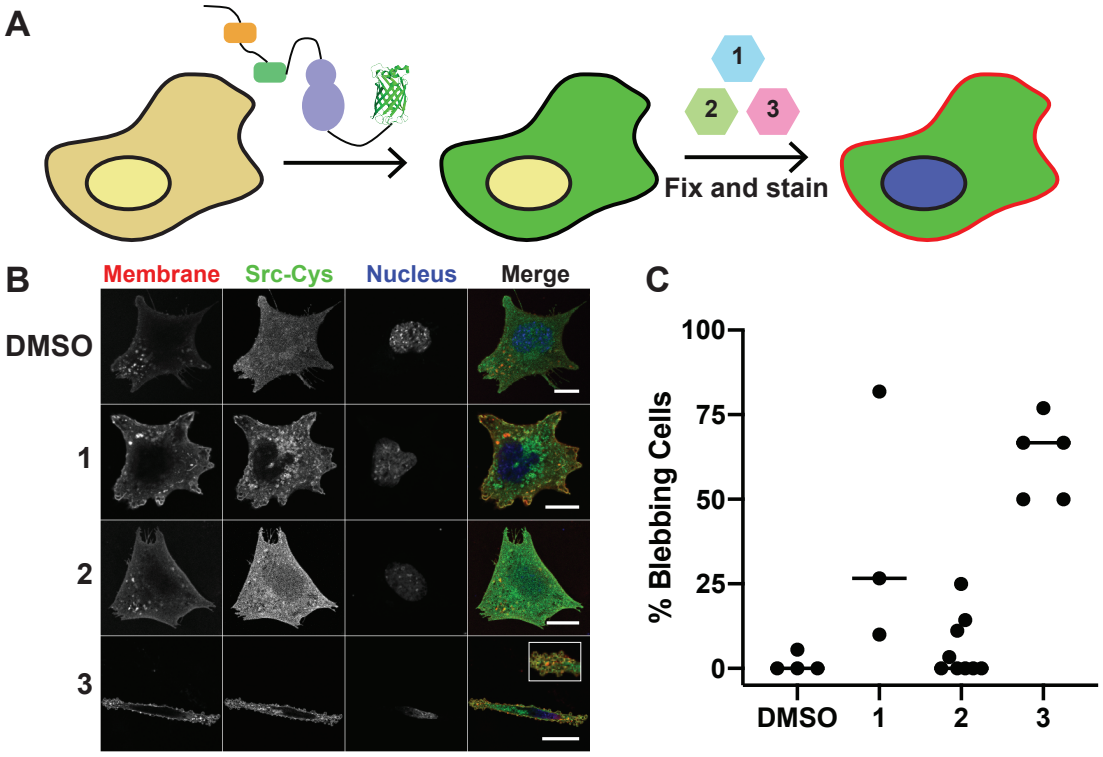

Figure 1.8 ***Confocal microscopy reveals phosphotransferase-independent cellular blebbing driven by conformational changes at the ATP-binding site.*** (A) Experimental workflow for visualization of cells by confocal microscopy. SYF cells were transfected with GFP-tagged Src-Cys, treated with a probe for 15 minutes, fixed and stained, then imaged. (B) Representative images of cells treated with DMSO, **1**, **2**, or **3**. Inset shows membrane blebs. Scale bar represents 10 μm. (C) Percentage of GFP-tagged Src-Cys cells showing blebs. Each point represents a replicate treatment with multiple cells imaged and scored in a double-blind fashion.

1.2.4 *Investigation of the SH4 domain as an auto-inhibitory element*

To investigate the intramolecular interactions which are responsible for the auto-inhibitory element of the SH4 domain, we partnered with Dr. Doug Fowler's group at the University of Washington. We hypothesized that this interaction occurs in a manner similar to that of the kinase Abl, which is also N-terminally myristoylated. Crystal structures of Abl show that there is intermolecular binding between the myristate and a in a pocket in Abl's CD[18]. Previous literature

has examined a homologous pocket in SFKs, but mutational experiments provide evidence that the SFK myristate does not bind in this pocket[17]. To understand the SFK intermolecular interaction, our collaborators used Deep Mutational Scanning (DMS). Every amino acid in the catalytic domain of Src was mutated to every other amino acid, then the effects of these mutations were measured through a yeast growth assay. In particular, we were interested in residues that resulted in a lack of yeast growth because active Src is toxic to yeast[19,20]. If Src is active, these results indicate that the kinase is in an active, globally open conformation, and that the mutant disrupts intramolecular engagements. Hyperactivating residues were grouped into seven clusters based on spatial proximity. Two of these clusters had known autoinhibitory functions: residues in these clusters had previously been shown to interact with the SH2 and SH3 domains[21,22]. Interestingly, one of the largest clusters of hyperactivating mutants had no known intramolecular function. This cluster was located in the αF-pocket of the c-terminal lobe of Src, a well-conserved binding pocket in SFKs. We hypothesized that these residues interact with the SH4 domain and theorized that the hyperactivating mutants found in our DMS data prevented the N-terminal tail from interacting with the residues in the pocket, locking the kinase in the activated, open conformation.

To test this hypothesis, a GFP-tagged Src mutant containing one of the hyperactivating mutants found in the αF pocket was generated, then transiently transfected into SYF cells (**Figure 1.9a, 1.9b**). Similar to Src treated with **3**, we observed that this mutant resulted in a relocalization of Src to cellular membranes and also induced a blebbing phenotype. Follow-up biochemical validation was also performed, and the results of these experiments strongly suggest that the SH4 domain intramolecularly interacts with residues in the αF pocket[20].

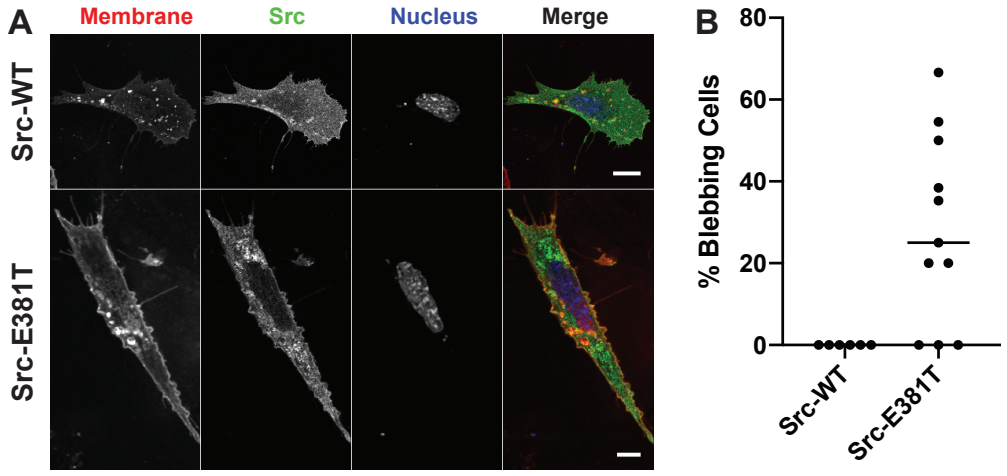

Figure 1.9 ***Src-E381T mutant reveals regulatory interface.*** (A) Representative images of GFP-tagged Src-WT or Src-E381T imaged 24 hours post-transfection. Scale bar represents 10 μm. (B) Percentage of GFP-tagged Src-WT or Src-E381T cells showing blebs. Each point represents a replicate treatment with multiple cells imaged and scored in a double-blind fashion

1.2.5 *Investigation of phenotypic effects of conformation selective Src inhibition*

While observing Src αF pocket mutants and Src-Cys treated with high concentrations of **3** (10 μM), we noticed that the cells exhibited a blebbing phenotype (**Figure 1.8c, 1.9c**). Blebs are rounded protrusions of the cell membrane caused by a destabilization of the actin cytoskeleton[23]. Live-cell videos for over an hour demonstrated that the observed blebs were dynamic and non-apoptotic, which are commonly associated with cell motility and division. Previous literature reported that cells transfected solely with the SH4 domain of different SFKs exhibit non-apoptotic blebbing and demonstrated that the blebs could be halted with a ROCK kinase inhibitor[23]. We hypothesized that our blebs followed a similar pathway. Relocalization of the SFKs to the plasma membrane caused the destabilization of the actin cytoskeleton, resulting in bleb formation which was then contracted through the activity of RHO, ROCK, and myosin. RHO is a GTPase that activates ROCK, a kinase that phosphorylates myosin, which in turn disassembles actin[25]. SYF cells transfected with Src-Cys and treated with a ROCK inhibitor followed by **3**, exhibited

relocalization to membranes, but no blebbing occurred (**Figure 1.10a, 1.10b**). These results lead us to believe that the blebbing phenotype was in fact dependent on RHO, ROCK, and myosin activity.

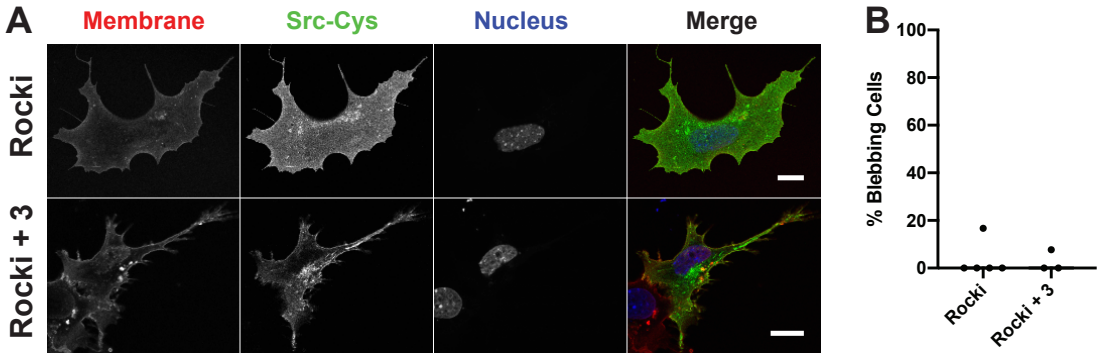

Figure 1.10 ***Rock inhibitor treatment demonstrates RHO GTPase dependence.*** (A) Representative images of GFP-tagged Src-Cys treated with Rock inhibitor GSK429286A for two hours, followed by DMSO or **3** for 15 minutes. Scale bar represents 10 µm. (B) Percentage of GFP-tagged Src-Cys cells showing blebs. Each point represents a replicate treatment with multiple cells imaged and scored in a double-blind fashion.

Next, we questioned if blebbing was dependent on the method of membrane association. All SFKs are N-terminally myristoylated, but an additional element is needed to bind to the plasma membrane. Src has a cluster of poly-basic lysines, while Fyn is palmitoylated in its SH4 domain[26]. A construct that bi-cistronically expressed Fyn-Cys and zsGreen was transfected into SYF cells and treated with DMSO, or **3** for 15 minutes (**Figure 1.11a. 1.11b**). Similar to Src, treatment of Fyn with **3** induced blebbing, demonstrating that different SH4 domains are capable of inducing membrane blebbing in fibroblasts.

1.3 Discussion

In this chapter, I present a generalizable method to selectively target an individual kinase, which can be used to study kinase non-catalytic functions. A panel of covalent probes were generated, which were shown to stabilize different ATP-binding site conformations through X-

ray crystallography. We applied our sensitization strategy to kinases across the kinome and showed that the covalent inhibitors could covalently inhibit the sensitized kinases over their wild type counterparts, demonstrating the generalizability of our method.

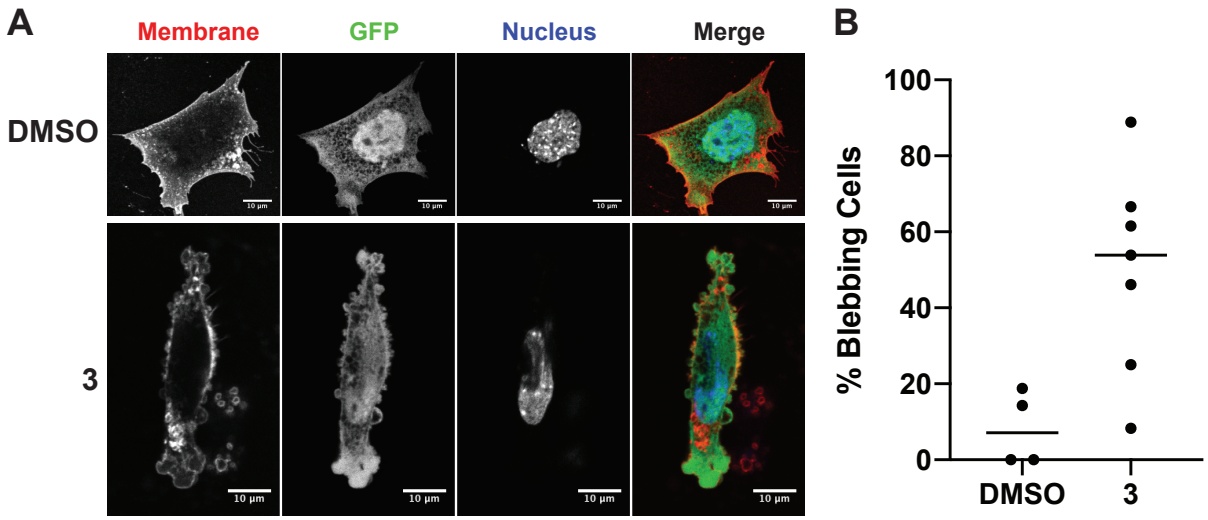

Figure 1.11 ***Fyn-Cys treated with 3 also blebs.*** (A) Representative images of SYF cells expressing Fyn-Cys (untagged) and GFP treated with DMSO or **3**. (B) Percentage of SYF cells exhibiting blebs. Each point represents a replicate treatment with multiple cells imaged and scored in a double-blind fashion.

This method was then specifically applied to Src, one of the SFKs. Biochemical validation demonstrated that by locking the kinase in a certain ATP-binding site conformation, the global kinase conformation could be allosterically controlled. Our chemical genetic strategy, combined with DMS data, identified a novel regulatory patch in the catalytic domain of the kinase. By creating a sensitized Fyn, we also demonstrated that this regulatory interaction is conserved across multiple members of the SFK family.

Future applications of this method can be used to elucidate the non-catalytic functions of diverse kinases, whether it be through inter- or intra-molecular associations. Additionally, by applying the knowledge gained through these experiments, drug developers can determine how stabilization of a specific ATP-binding site conformation can affect downstream signaling. For

example, we demonstrated that by stabilizing the Src ATP-binding site in a DFG-out conformation, cells exhibited increased blebbing, a phenotype typically associated with increased cell motility. Kinases are commonly a drug target in human cancers and creating more mobile cancer cells is something to be avoided. Application of these methods to kinases across the kinome can reveal important elements of intramolecular regulation, which can have serious implications for future drug discovery.

1.4 Materials and Methods

1.4.1 *Synthesis of Inhibitors*

Inhibitors were synthesized and purified by Dr. Rao Vidadala (**1** and **2**) and Linglan Fang (**3**) following the procedure described in Ahler et al., 2019.

1.4.2 *Cloning*

All mutants were made through QuikChange mutagenesis (Agilent) following standard protocols and verified through Sanger sequencing. All subcloning was achieved using Gibson assembly and validated through Sanger sequencing. Please see Table 1.2 for construct amino acid sequences.

Table 1.1. Amino acid sequences of constructs

Construct	Amino acid sequence
pET28a-Src-WT^{CD}	GHMQTQGLAKDAWEIPRESLRLEVKLGQGCFGEVWMGTWNGTTR VAIKTLKPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEPIYIV TEYMSKGSLLDFLKGEMGKYLRLPQLVDMAAQIASGMAYVERMN YVHRDLRAANILVGENLVCKVADFGLARLIEDNEYTARQGAKFPIK WTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVLD QVERGYRMPCPPECPESLHDLMCQCWRKDPEERPTFEYLQAFLEDY FTSTEPQYQPGENL*

pET28a-Src-Cys^{CD}	GHMQTQGLAKDAWEIPRESLRLEVKLGQGCFGE**C**WMGTWNGTTR VAIKTLKPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEPIYIV TEYMSKGSLLDFLKGEMGKYLRLPQLVDMAAQIASGMAYVERMN YVHRDLRAANILVGENLVCKVADFGLARLIEDNEYTARQGAKFPIK WTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVLD QVERGYRMPCPPECPESLHDLMCQCWRKDPEERPTFEYLQAFLEDY FTSTEPQYQPGENL*
pcDNA5-Hck-Cys	MGCMKSKFLQVGGNTFSKTETSASPHCPVYVPDPTSTIKPGPNSHNS NTPGIREAGSEDIIVVALYDYEAIHHEDLSFQKGDQMVVLEESGEWW KARSLATRKEGYIPSNYVARVDSLETEEWFFKGISRKDAERQLLAPG NMLGSFMIRDSETTKGSYSLSVRDYDPRQGDTVKHYKIRTLDNGGF YISPRSTFSTLQELVDHYKKGNDGLCQKLSVPCMSSKPQKPWEKDA WEIPRESLKLEKKLGAGQFGE**C**WMATYNKHTKVAVKTMKPGSMSV EAFLAEANVMKTLQHDKLVKLHAVVTKEPIYIITEFMAKGSLLDFLK SDEGSKQPLPKLIDFSAQIAEGMAFIEQRNYIHRDLRAANILVSASLV CKIADFGLARVIEDNEYTAREGAKFPIKWTAPEAINFGSFTIKSDVWS FGILLMEIVTYGRIPYPGMSNPEVIRALERGYRMPRPENCPEELYNIM MRCWKNRPEERPTFEYIQSVLDDFYTATESQYQQQP*
pcDNA5-Src-Cys-Flag	MGSNKSKPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADG HRGPSAAFAPAAAEPKLFGGFNSSDTVTSPQRAGPLAGGVTTFVALY DYESRTETDLSFKKGERLQIVNNTEGDWWLAHSLSTGQTGYIPSNYV APSDSIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESETTKGAYC LSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSLQQLVAYYSK HADGLCHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQGCFG E**C**WMGTWNGTTRVAIKTLKPGTMSPEAFLQEAQVMKKLRHEKLVQ LYAVVSEEPIYIVTEYMSKGSLLDFLKGETGKYLRLPQLVDMAAQIA SGMAYVERMNYVHRDLRAANILVGENLVCKVADFGLARLIEDNEY TARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPY PGMVNREVLDQVERGYRMPCPPECPESLHDLMCQCWRKEPEERPT FEYLQAFLEDYFTSTEPQYQPGENLGSGTDYKDDDDK*
pef1a-Src-Cys-GFP	MGSNKSKPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADG HRGPSAAFAPAAAEPKLFGGFNSSDTVTSPQRAGPLAGGVTTFVALY DYESRTETDLSFKKGERLQIVNNTEGDWWLAHSLSTGQTGYIPSNYV APSDSIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESETTKGAYC LSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSLQQLVAYYSK HADGLCHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQGCFG E**C**WMGTWNGTTRVAIKTLKPGTMSPEAFLQEAQVMKKLRHEKLVQ LYAVVSEEPIYIVTEYMSKGSLLDFLKGETGKYLRLPQLVDMAAQIA SGMAYVERMNYVHRDLRAANILVGENLVCKVADFGLARLIEDNEY TARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYP GMVNREVLDQVERGYRMPCPPECPESLHDLMCQCWRKEPEERPTFE YLQAFLEDYFTSTEPQYQPGENLGSGTGSGTGATMVSKGEELFTGVV PILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYK TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMAD KQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLST

	QSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK*
pef1a-Src- WT-GFP	MGSNKSKPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADG HRGPSAAFAPAAAEPKLFGGFNSSDTVTSPQRAGPLAGGVTTFVALY DYESRTETDLSFKKGERLQIVNNTEGDWWLAHSLSTGQTGYIPSNYV APSDSIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESETTKGAYC LSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSLQQLVAYYSK HADGLCHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQGCFG EVWMGTWNGTTRVAIKTLKPGTMSPEAFLQEAQVMKKLRHEKLVQ LYAVVSEEPIYIVTEYMSKGSLLDFLKGETGKYLRLPQLVDMAAQIAS GMAYVERMNYVHRDLRAANILVGENLVCKVADFGLARLIEDNEYTA RQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGM VNREVLDQVERGYRMPCPPECPESLHDLMCQCWRKEPEERPTFEYLQ AFLEDYFTSTEPQYQPGENLGSGTGSGTGATMVSKGEELFTGVVPILV ELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT LTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTR AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK QKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLST QSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK*
pef1a-Src- E381T-GFP	MGSNKSKPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASAD GHRGPSAAFAPAAAEPKLFGGFNSSDTVTSPQRAGPLAGGVTTFV ALYDYESRTETDLSFKKGERLQIVNNTEGDWWLAHSLSTGQTGYI PSNYVAPSDSIQAEEWYFGKITRRESERLLLNAENPRGTFLVRESET TKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSLQQ LVAYYSKHADGLCHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEV KLGQGCFGEVWMGTWNGTTRVAIKTLKPGTMSPEAFLQEAQVMK KLRHEKLVQLYAVVSEEPIYIVTEYMSKGSLLDFLKGETGKYLRLPQ LVDMAAQIASGMAYV**T**RMNYVHRDLRAANILVGENLVCKVADFG LARLIEDNEYTARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLT ELTTKGRVPYPGMVNREVLDQVERGYRMPCPPECPESLHDLMCQC WRKEPEERPTFEYLQAFLEDYFTSTEPQYQPGENLGSGTGSGTGAT MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLK FICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEG YVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG HKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQ QNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI TLGMDELYK*
pef1a-Fyn- Cys-ires- zsGreen	MGCVQCKDKEATKLTEERDGSLNQSSGYRYGTDPTPQHYPSFGV TSIPNYNNFHAAGGQGLTVFGGVNSSSHTGTLRTRGGTGVTLFVA LYDYEARTEDDLSFHKGEKFQILNSSEGDWWEARSLTTGETGYIPS NYVAPVDSIQAEEWYFGKLGRKDAERQLLSFGNPRGTFLIRESETT KGAYSLSIRDWDDMKGDHVKHYKIRKLDNGGYYITTRAQFETLQ QLVQHYSEKADGLCFNLTVIASSCTPQTSGLAKDAWEVARRSLCL EKKLGQGCFAE**C**WLGTWNGNTKVAIKTLKPGTMSPESFLEEAQIM KKLKHDKLVQLYAVVSEEPIYIVTEYMNKGSLLDFLKDGEGRALK LPNLVDMAAQVAAGMAYIERMNYIHRDLRSANILVGNGLICKIAD FGLARLIEDNEYTARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGI

	LLTELVTKGRVPYPGMNNREVLEQVERGYRMPCPQDCPISLHELM IHCWKKDPEERPTFEYLQSFLEDYFTATEPQYQPGENL*

1.4.3 *Protein purification*

For X-ray crystallography, the catalytic domain of Src-WT or Src-Cys (residues 261-536) were cloned into the bacterial expression plasmid pET-28a as a N-terminal His_6-Tev tagged construct. This construct was co-transformed in *E. coli* with *Yersinia pestis* tyrosine phosphatase YopH (pCDF-DUET-1 vector) and chaperonin GroEL (pACYDuet-1 vector). Transformed *E. coli* were plated on triple selective plates (Kanamycin/Chloramphenicol/Streptomycin), and a single colony was chosen. The colony was grown in an overnight culture of 50 mL Terrific broth (ThermoFisher) containing all three antibiotics. The next morning, three 2 L cultures were inoculated with the starter culture, grown to an OD_{600} of 1.2 at 37 ºC. Upon reaching optical density, the temperature was dropped to 18 ºC and protein expression was induced with 0.2 M IPTG. The next morning, cells were harvested through centrifugation, lysed by sonification in lysis buffer (50 mM HEPES, pH 8.0, 300 mM NaCl, 1 mM PMSF, 0.1% Triton-X, 20 mM imidazole) and centrifuged for 45 min at 10,000 rpm. Cleared lysates were incubated with 1.5 mL Ni-NTA resin (ThermoFisher) for 1 hour. Lysate was then discarded, and the beads were washed with 20 mL lysis buffer. Bound protein was eluted with lysis buffer containing 300 mM imidazole, the dialyzed overnight at 4 ºC in dialysis buffer (50 mM Tris, pH 8.0, 5% glycerol, 1 mM DTT) with 1:25 ratio of TEV protease:eluted protein (wt:wt). Following dialysis, the solution was filtered through a 0.22-μm filter and loaded onto a Ni-NTA affinity column (GE Lifescience). The column was equilibrated with buffer A (50 mM Tris pH 8.0, 5% glycerol, 1 mM DTT), then washed with five column volumes of buffer A. The protein was eluted with a linear gradient of 0%-35% of

buffer B (50 mM Tris pH 8.0, 1 M NaCl, 5% glycerol, 1 mM DTT), and peak fractions were analyzed by SDS-PAGE. Fractions containing pure Src were pooled and used for crystallization.

Proteins used in the inhibitor titrations were expressed and purified by Dr. Ames Register (Abl, Hck, EphA2) Dr. Chloe Lombard (JNK2), Sujata Chakraborty (Src), and myself (PAK1). For methods to express and purify Hck and Src please see Ahler et al., 2019, while methods to express JNK2 and EphA2 can be found in Fang et al., 2019 (Ahler et al., 2019; Fang et al., 2019). Abl was purified following the same protocol as Hck and Src.

PAK1 V286C/M346T (residues 250-547) was cloned into the bacterial expression construct plasmid pMCSG7 with an N-terminal His-SUMO tag, then transformed in *E. coli* BL21 cells and plated on ampicillin selective plates. From the plate, a single colony was chosen and grown in an overnight culture of 50 mL LB broth with ampicillin. The next day, a 1 L culture was inoculated with the starter culture, grown to an OD_{600} of 0.8, then expression of PAK1 was induced with 0.2 M IPTG at 18 ºC. After allowing the protein to be expressed overnight, the cells were harvested by centrifugation, lysed by sonication in lysis buffer (50 mM HEPES pH 8.0, 300 mM NaCl, 1 mM PMSF, 0.1% Triton-X, 20 mM imidazole), then centrifuged for 45 minutes at 10,000 rm. Cleared lysates were incubated with 0.5 mL N-NTA resin (Thermo) for 1 hr. The supernatant was then discarded, and beads were washed with 20 mL lysis buffer, then protein was eluted with lysis buffer containing 300 mM imidazole. Eluted PAK1 was dialyzed overnight at 4 ºC in dialysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM DTT). Wild type PAK1 was purchased from SignalChem.

1.4.4 *Inhibitor IC_{50} and K_i determination*

Prior to IC_{50} determination, kinase titrations were performed to ensure that the kinase concentration assayed was in the linear range. For inhibitor titration protocols for Src, Hck, and

Abl, please see Ahler et al., 2019. For inhibitor titration protocols for JNK2, EphA2, please see Fang et al., 2019.

PAK1 WT (18 nM) and V286C/M246T (35 nM) catalytic domains were shown to be catalytically active after expression and purification. Proteins were incubated with 1, 2, or 3 at an initial concentration of 10 μM with 3-fold serial dilutions over 10 data points, 4 μM cold ATP, and 0.007 μCi/μL γ32P-ATP for 30 minutes in assay buffer (76 mM HEPES pH 7.5, 5 mM $MgCl_2$, 150 mM NaCl, 3.8 mM EGTA, 0.2 mg/mL BSA, 150 μM sodium orthovanadate, 1 mM betamercaptoethanol). Following the initial incubation, 0.2 mg/mL PAKtide (sequence: RRRLSFAEPG) was added and the entire mixture was allowed to incubate for 2-4 hours. 4.6 μL of the reaction mixture was spotted onto phosphocellulose, then membranes were washed with 0.05% phosphoric acid (3x10 minute washes), air dried, then radioactivity was determined by phosphor-imaging using a GE Typhoon FLA 900 phosphor-scanner. GraphPad Prism software was used to analyze data, and IC_{50} values were determined using the "One-site fit logIC_{50}" option. All assays were done in triplicate.

1.4.5 *Cell culture*

Src/Yes/Fyn (SYF) murine fibroblasts were purchased from ATCC and maintained in DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco). For microscopy experiments, transient transfections were performed 24 hours after plating $2x10^4$ cells/mL on 18 mm glass slides (ThermoFisher) using a 3:1 ratio of μL turbofectin (Origene): μg DNA. 24 hours after transfection, cells were treated with the indicated inhibitors for the specified length of time. For SYF cells treated with the ROCK inhibitor GSK429286A (SelleckChem), 10 μM ROCK inhibitor was added 2 hours prior to conformation-selective inhibitor treatment.

Flp-In™ 293 T-REx™ (ThermoFisher) that had stably incorporated c-terminal flag tagged Hck WT or Cys following manufacturer's instructions were used for proteomic experiments. These cells were maintained in custom made mixtures of SILAC media (Ahler et al., 2019).

1.4.6 *Kinobead proteomic profiling of sensitized kinases*

Cell lysates for these experiments were prepared by Dr. Ames Register, and proteomic profiling, along with the data analysis, was carried out by Dr. Martin Golkowski. For extended methods, please see Ahler et al., 2019.

1.4.7 *SH3 pulldown*

SH3 pulldown assays were performed by Sujata Chakraborty. For extended methods, please see Ahler et al., 2019.

1.4.8 *Microscopy*

Following inhibitor treatment, cells were washed once with PBS then fixed with 4% paraformaldehyde solution (Electron Microscopy Sciences) in PBS for 15 minutes. After fixing, cells were washed twice with PBS, then stained with wheat-germ agluttinin conjugated to Alexafluor647 (ThermoFisher) and NucBlue (Thermofisher), which are membrane and nucleus stains, respectively. Fluoromount G (Southern Biotechnology) was then applied to the slides, which were then mounted and sealed with nail polish, then imaged on a Leica SP8X scope.

Obtained images were brightened for clarity with ImageJ. Images were randomized and blinded, then scored by an unbiased scorer as either blebbing or not blebbing after viewing a training set of images for each condition. Cells were only scored if the majority of the cell was in the image and if they contained a single nucleus and had an intact, continuous membrane. Cells

were considered blebbing if there were 3 or more rounded plasma membrane protrusions. In total, over 2000 cells were scored, and the raw images were uploaded to Mendeley Data.

1.4.9 *Deep Mutational Scanning*

Assembly of the Src catalytic domain library, integration of the library into the yeast growth assay, deep sequencing of the library variants collected from the yeast growth assay, as well as analysis of the deep sequencing data was performed by Dr. Ethan Ahler. For extended methods, please see Ahler et al., 2019.

1.5 REFERENCES

1. Taylor, S. S., Ilouz, R., Zhang, P., Kornev, A. P. (2012). Assembly of allosteric macromolecular switches: lessons from PKA. *Nature Reviews Molecular Cell Biology, 13(10)*, 646–658.

2. Johnson, L. N., Noble, M. E. M., Own, D. J. (1996). Active and Inactive Protein Kinases: Structural Basis for Regulation. *Cell, 85*, 149–158.

3. Kung, J. E., Jura, N. (2016). Structural Basis for the Non-catalytic Functions of Protein Kinases. *Structure, 24(1*), 7–24.

4. Thomas, S. M., Brugge, J. S. (1997). Cellular Functions Regulated by Src Family Kinases. *Annual Review of Cell and Developmental Biology*, 513–609.

5. Sen, B., Johnson, F. M. (2011). Regulation of Src Family Kinases in Human Cancers. *Journal of Signal Transduction, 2*, 1–14.

6. Krishnamurty, R., Brigham, J. L., Leonard, S. E., Ranjitkar, P., Larson, E. T., Dale, E. J., Merritt, E. A., Maly, D. J. (2012). Active site profiling reveals coupling between domains in SRC-family kinases. *Nature Chemical Biology, 9(1*), 43–50.

7. Leonard, S. E., Register, A. C., Krishnamurty, R., Brighty, G. J., Maly, D. J. (2014). Divergent Modulation of Src-Family Kinase Regulatory Interactions with ATP-Competitive Inhibitors. *ACS Chemical Biology, 9(8)*, 1894–1905.

8. Register AC, Leonard SE, Maly DJ. (2014). SH2-catalytic domain linker heterogeneity influences allosteric coupling across the SFK family. *Biochemistry, 53*, 6910-6923.

9. Fang, L., Chakraborty, S., Dieter, E. M., Potter, Z. E., Lombard, C. K., Maly, D. J. (2019). Chemoproteomic Method for Profiling Inhibitor-Bound Kinase Complexes. *Journal of the American Chemical Society*, *141*(30), 11912–11922.

10. Hari, S. B., Merritt, E. A., Maly, D. J. (2014). Conformation-Selective ATP-Competitive Inhibitors Control Regulatory Interactions and Noncatalytic Functions of Mitogen-Activated Protein Kinases. *Chemistry & Biology, 21(5)*, 628–635.

11. Feldman, H. C., Tong, M., Wang, L., Meza-Acevedo, R., Gobillot, T. A., Lebedev, I., Gliedt, M. J., Hari, S. B., Mitra, A. K., Backes, B. J., Papa, F. R., Seeliger, M. A., Maly, D. J. (2016). Structural and Functional Analysis of the Allosteric Inhibition of IRE1α with ATP-Competitive Ligands. *ACS Chemical Biology, 11(8),* 2195–2205.

12. Hanke, J. H., Gardner, J., Dow, R. L., Changelian, P. S., Brisseette, W. H., Weringer, E. J., Pollok, B. A., Connelly, P. A. (1996). Discovery of a Novel, Potent, and Src Family-selective Tyrosine Kinase Inhibitor. *The Journal of Biological Chemistry, 271*, 695–701.

13. Muratore, K. E., Seeliger, M. A., Wang, Z., Fomina, D., Neiswinger, J., Havranek, J. J., Baker, D., Kuriyan, J., Cole, P. A. (2009). Comparative Analysis of Mutant Tyrosine Kinase Chemical Rescue. *Biochemistry, 48(15),* 3378–3386.

14. Serafimova, I. M., Pufall, M. A., Krishnan, S., Duda, K., Cohen, M. S., Maglathlin, R. L., McFarland, J. M., Miller, R. M., Frodin, M., Taunton, J. (2012). Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nature Chemical Biology, 8(5),* 471–476.

15. Golkowski, M., Brigham, J. L., Perera, B. G. K., Romano, G. S., Maly, D. J., Ong, S.-E. (2014). Rapid profiling of protein kinase inhibitors by quantitative proteomics. *Med. Chem. Commun., 5(3*), 363–369.

16. Klinghoffer, R. A., Sachsenmaier, C., Cooper, J. A., Soriano, P. (1999). Src family kinases are required for integrin but not PDGFR signal transduction. *The EMBO Journal*, *18*(9), 2459–2471.

17. Patwardhan, P., Resh, M. D. (2010). Myristoylation and Membrane Binding Regulate c-Src Stability and Kinase Activity. *Molecular and Cellular Biology, 30(17)*, 4094–4107.

18. Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J. (2003). Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell, 112*, 859–871.

19. Fowler, D. M., Fields, S. (2014). Deep mutational scanning: a new type of protein science. *Nature Methods, 11*, 801-807.

20. Ahler, E., Register, A. C., Chakraborty, S., Fang, L., Dieter, E. M., Sitko, K. A., Vidadala, R. S. R., Trevillian, B. M., Golkowski, M., Gelman, H., Stephany, J. S., Merritt, E. A.,

Fowler, D. M., Maly, D. J. (2019). A Combined Approach Reveals a Regulatory Mechanism Coupling Src's Kinase Activity, Localization, and Phosphotransferase-Independent Functions. *Molecular Cell*, 393-408.

21. Xu, W., Harrison, S. C., Eck, M. J. (2017). Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, *385*, 595–602.

22. Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck un- derlies their inactivation by C-terminal tyrosine phosphorylation. Cell *105*, 115–126.

23. Fackler, O. T., Grosse, R. (2008). Cell motility through plasma membrane blebbing. *The Journal of Cell Biology, 181(6)*, 879–884.

24. Tournaviti, S., Hannemann, S., Terjung, S., Kitzing, T. M., Stegmayer, C., Ritzerfeld, J., et al. (2007). SH4-domain-induced plasma membrane dynamization promotes bleb-associated cell motility. Journal of Cell Science, 120(Pt 21), 3820–3829.

25. Charras, G. T. (2008) A short history of blebbing. *J Microsc. 231*, 466-478.

26. Resh, M. D. (1998). Fyn, a Src family tyrosine kinase. *BCB, 11*, 1159-1162.

Chapter 2. INTERROGATION OF LOCALIZED RAS-MEDIATED SIGNALING USING A CHEMICALLY INDUCED ACTIVATOR OF RAS

This chapter is adapted from a manuscript currently in preparation.

2.1 INTRODUCTION

RAS GTPases are involved in the regulation of a multitude of cellular processes, ranging from proliferation and differentiation to apoptosis, through their function as binary molecular switches[1,2]. In addition to their roles in healthy cells, RAS proteins were the first oncogene to be discovered in the 1980's and have since been shown to be the most frequently mutated gene family in human cancers, with ~25% of human tumors containing a RAS mutation [3,4]. Over the last forty years, there have been extensive efforts to understand RAS signaling. While there has been success in obtaining a greater mechanistic understanding of RAS-driven signaling, a number of basic questions about RAS biology still remain, due to the dynamic and complex nature of RAS biology[5]. In particular, the dynamics of RAS-driven processes remain poorly understood.

In healthy cells RAS proteins serve as molecular switches, cycling between an "on" and "off" state. In the "off" state, RAS is bound to guanosine diphosphate (GDP) and does not recruit downstream effectors. GDP-bound RAS is converted to the active, guanosine triphosphate (GTP) bound form by guanine nucleotide exchange-factors (GEFs), which catalyze the exchange of GDP for GTP[6,7,8]. Active RAS is returned to its inactive form through the hydrolysis of GTP to GDP, which is catalyzed by GTPase-activating proteins, or GAPs[9]. GTP-bound RAS mediates downstream signaling through recruitment of cellular effectors. One of the most well-characterized RAS-mediated signaling pathways is the RAS/MEK/ERK kinase phosphorylation cascade, which is initiated by the recruitment of RAF kinases to GTP-bound RAS. In addition to the

RAF/MEK/ERK cascade, there is evidence that GTP-bound RAS is responsible for initiating additional signaling pathways, depending on the RAS isoform and its subcellular localization[10,11,12].

In humans, there are three RAS genes that lead to four RAS protein isoforms: HRAS, NRAS, KRAS4A, and KRAS4B. Although HRAS, NRAS, KRAS4A, and KRAS4B share an effector domain that contains 95% sequence similarity, multiple studies have shown that they are not functionally redundant in cells[11-17]. Additionally, the occurrences of each RAS isoform in different cancers varies. KRAS4B mutations are frequently found in pancreatic (91%), colon (42%), and lung (33%) cancers, while NRAS is predominantly found in melanomas (27%). Oncogenic HRAS and KRAS4A mutations are found much less frequently in cancers[2,11, 18]. Differences in HRAS, NRAS, KRAS4A, and KRAS4B signaling and function is thought to stem from their less conserved hypervariable regions (HVR). RAS HVRs consist of the 19 or 20 amino acids at their C-termini, which all contain C-terminal CAAX motifs. The CAAX motif is post-translationally farnesylated on the cysteine residue at the endoplasmic reticulum (ER), followed by cleavage of the -AAX motif and methylation on the now terminal cysteine[19,20]. After these modifications RAS isoforms are either shuttled directly to the plasma membrane (KRAS4B) or undergo additional palmitoylation in the golgi before ultimately being trafficked to various endomembranes[21].

Until recently, methods for studying RAS-mediated signaling have relied primarily on expression of oncogenic RAS mutants, which are constitutively active, or activation of receptor tyrosine kinases with pleiotropic growth factors. While these methods have provided significant insights into RAS signaling, they are not without shortcomings. Activation of RAS-mediated signaling networks with growth factors allows for temporal control, but multiple signaling

pathways are also activated. This makes it difficult to determine whether the effects observed are due to RAS activation or a parallel signaling pathway. Studies relying on the expression of oncogenic RAS mutants can provide information on isoform-driven cellular processes, but it is often difficult to obtain expression levels that match the endogenous concentrations of these GTPases. In addition, genetic methods lack temporal precision, which does not allow the dynamics of RAS-mediated signaling to be studied. In order to circumvent these issues, we and others have developed methods for activation of endogenous RAS with high temporal precision[22-25].

In this study, we describe the use of a small-molecule controlled protein switch that allows for the rapid activation of endogenous RAS. We refer to this small molecule-controlled protein switch as chemically inducible activator of RAS, or CIAR. CIAR relies on the use of a protein/peptide interaction to autoinhibit the active site of the RAS GEF Son of Sevenless (SOS), preventing it from activating RAS in the basal state. Treatment of cells with a small molecule that disrupts the autoinhibitory protein/peptide interaction activates CIAR's GEF activity, which subsequently promotes the generation of endogenous RAS-GTP. By appending the HVRs from different RAS isoforms to the C-termini of CIAR constructs, we are able to activate different subcellular populations of RAS with a small molecule. In total, we generated four CIAR variants, and the downstream effects of activating each CIAR variant was characterized with quantitative mass spectrometry. This chapter describes the generation and characterization of cell lines expressing our differentially localized CIAR variants.

2.2 RESULTS

2.2.1 *Engineering spatiotemporal control of RAS isoforms*

In order to probe the direct effects of RAS-mediated signaling, a method to rapidly and selectively activate RAS is needed. To that end, we computationally designed and validated an

intramolecular switch called chemically-inducible activator of RAS (CIAR) that specifically activates RAS[23,24]. Briefly, our switch is composed of a RAS GEF that is autoinhibited by a protein/peptide interaction that can be disrupted with a cell permeable small molecule (**Figure 2.1a**). The catalytic domain of the GEF Son of Sevenless (SOScat) protein was chosen as a RAS activator because it has been previously reported that SOS can be reduced to just its Cdc25 and Rem domains, resulting in a constitutively catalytically active protein[6]. To inhibit SOScat's activity, we used the well-characterized interaction between the anti-apoptotic Bcl-xL protein and a BH3 peptide[26]. Linkers connecting Bcl-xL and BH3 were optimized such that in the absence of drug, the Bcl-xL/BH3 interaction occludes the RAS-binding site of SOScat. Numerous small molecule inhibitors are available that selectively target Bcl-xL and disrupt the Bcl-xL/BH3 interaction with minimal effects on most mammalian cell lines that have not been primed for death[27,28]. In this study, we used the Bcl-xL inhibitor A-1155463 (A1), due to its potency and selectivity for Bcl-xL over other Bcl-2 family members[29]. Upon addition of A1 to cells, the autoinhibitory interaction between Bcl-xL and BH3 peptide in CIAR is disrupted, resulting in rapid activation of SOScat. SOScat is then able to promote the formation of RAS-GTP from RAS-GDP.

Although the four RAS isoforms share a highly conserved GTPase domain, their C-terminal ~20 amino acids, called the hypervariable region (HVR), possess less than 15% sequence similarity[21] (**Figure 2.1b**). Because RAS HVRs contain the major determinants for interacting with membranes, this diversity leads to differences in the subcellular distribution of the RAS isoforms[2]. In our initial efforts, we localized our CIAR construct to endomembranes by appending the HVR from HRAS to its C-terminus[24]. In addition to our original HRAS HVR-containing CIAR construct, we generated three CIAR variants: (1) a CIAR construct containing the HVR from KRAS4A (CIAR KRAS4A HVR), (2) a CIAR construct containing the HVR from KRAS4B

(CIAR KRAS4B HVR), and (3) a CIAR construct containing the HVR from NRAS (CIAR NRAS HVR). We hypothesized that each CIAR variant will undergo the same processing and localization as the RAS isoform that its HVR derived from. This differential localization should allow activation of different subcellular pools of RAS.

All four RAS isoforms are expressed in the cytosol, where the C-terminal cysteine on the shared CAAX motif is farnesylated[30]. After farnesylation, RAS proteins are trafficked to the endoplasmic reticulum (ER) where the -AAX motif is cleaved, followed by the methylation of the farnesylated cysteine[31-33]. At this stage, the processing and trafficking of the four RAS isoforms diverge. KRAS4B requires no further post-translational modifications. It contains a polybasic cluster of lysines that promote localization to the plasma membrane[34,35]. NRAS and HRAS have one or two additional cysteines, respectively, in their HVRs that are reversibly palmitoylated in the golgi[35,36]. Upon palmitoylation, NRAS and HRAS translocate to the plasma membrane, where they are depalmitoylated, resulting in shuttling back and forth between the plasma membrane and the golgi[74]. KRAS4A contains both a small polybasic cluster of lysines and a cysteine that is palmitoylated. It is believed that KRAS4A is trafficked through the golgi but has a lesser amount of shuttling between intracellular and plasma membranes compared to HRAS and NRAS[37]. To determine the localization of our CIAR constructs, we created CIAR-mCherry fusions. Flp-In T-REx 293 cell lines that stably and inducibly express each of the CIAR-mCherry fusions were generated. 24 hours after inducing expression of CIAR constructs, cells were then were fixed and imaged (**Figure 2.1c**). We found that CIAR KRAS4A HVR and CIAR KRAS4B HVR demonstrated strong co-localization with AlexaFluor647-labeled Wheat Germ Agglutinin (WGA), suggesting both constructs are predominantly plasma membrane localized. In contrast, we observed very little overlap between CIAR NRAS HVR and CIAR HRAS HVR with

AlexaFluor647-labeled (WGA), demonstrating that both constructs are predominately localized to intracellular endomembranes. These observed patterns of CIAR localization are similar to those that have previously been reported for different RAS isoforms[21,36,38]. These results suggest that each CIAR construct demonstrates similar processing and localization as their RAS isoform counterparts, which we predicted should result in activation of different pools of localized RAS.

To measure the ability of our CIAR variants to directly activate RAS, we used a genetically encoded biosensor called RasRaichu. RasRaichu is a previously reported Förster resonance energy transfer (FRET) sensor that is composed of the RAS GTPase domain, the RAS binding domain of RAF, and two fluorescent proteins that undergo FRET (**Figure 2.1d**)[39]. CIAR variants were co-transfected with RasRaichu constructs that were targeted to either the plasma membrane or the golgi. We observed low FRET in the absence of A1 but higher FRET upon the activation of CIAR (**Figure 2.1e**). We found that both CIAR KRAS4B HVR and CIAR HRAS HVR let to activation of plasma membrane-localized RasRaichu.

Interestingly, CIAR HRAS HVR more efficiently activates golgi-localized RasRaichu than CIAR KRAS4B HVR (*data not shown*). These data demonstrate that our CIAR variants are all able to directly activate RAS. Once our CIAR variants had been validated, we generated four Flp-In T-REx 293 cell lines that each stably express one of the CIAR variants (**Figure 2.1f, 2.1g**). Expression of the CIAR variants can be induced by the addition of doxycycline (dox), and comparison of the different cell lines show similar amounts of each variant being expressed, allowing for comparison between cell lines.

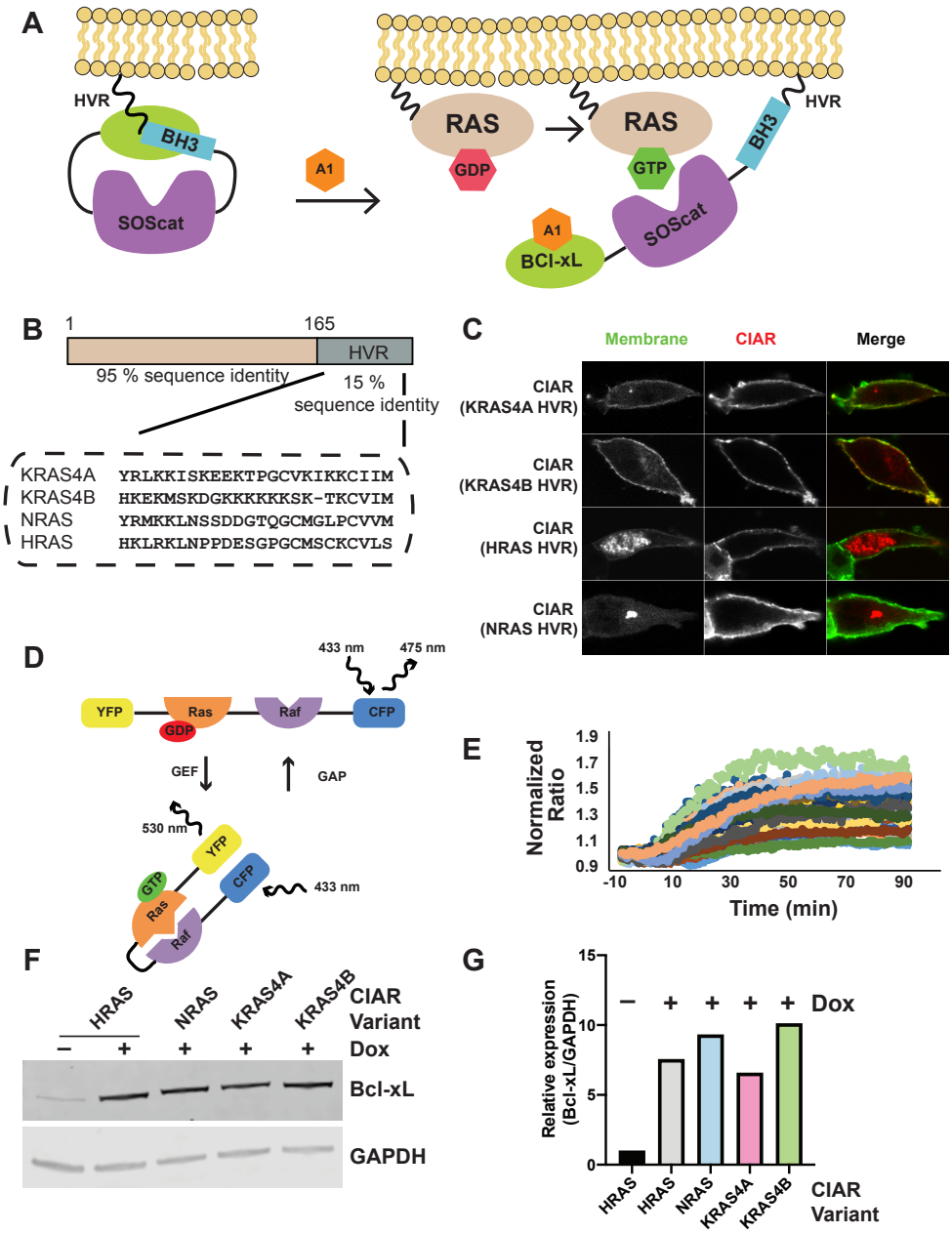

Figure 2.1. ***Characterization of CIAR variants.*** (A) Components of the CIAR system and schematic of CIAR-mediated RAS activation. (B) Comparison of RAS isoform HVR sequences used in the CIAR variants (C) Representative images of mCherry-tagged CIAR variants in cells. Briefly, expression of CIAR constructs were induced with doxycycline, then allowed to express for 24 hours. Cells were then fixed and stained with WGA conjugated to AlexaFluor647 and imaged using a confocal microscope. (D) Schematic of the RasRaichu FRET sensor. (E) Direct activation of plasma membrane-localized RasRaichu by CIAR HRAS HVR. (F) Anti-Bcl-xL blots of HEK293T cells lines stably expressing CIAR variants. (G) Quantitation of CIAR expression following induction of expression. Four Flp-In T-REx 293 cells were generated that each stably and inducibly expressed one of the CIAR variants. 24 hours after plating, cells were induced with dox or left uninduced. The next day cells were lysed and probed for Bcl-xL and GAPDH. Relative expression was calculated by dividing the Bcl-xL intensity values by the GAPDH intensity values.

2.2.2 *Temporal control of CIAR variants results in flux through RAS signaling pathways*

Next, we determined whether each of our CIAR variants are able to activate the canonical RAF/MEK/ERK pathway. To do this, we chose to monitor levels of phosphorylated ERK (pERK), which undergoes dual activation loop phosphorylation upon RAS activation [40]. As an initial test, cells expressing one of our four RAS variants were tested for their ability to activate downstream ERK 30 minutes after addition of 200 nM A1 (Figure 2.2a, 2.2b). Although our CIAR variants are localized to different subcellular locations, each CIAR variant was able to promote similar levels of ERK activation. CIAR-expressing cells were then treated with DMSO or different concentrations of A1. In all cell lines, a dose-dependent increase of phosphorylated ERK (pERK) was observed, demonstrating that CIAR can be used as a rheostat for isoform-specific RAS activation (**Figure 2.2c, 2.2d**). Additionally, we found that each CIAR variant responded similarly to A1, with a half maximal effective concentration (EC_{50}) of ~ 30 nM for CIAR KRAS4A HVR and CIAR NRAS HVR. This demonstrates that similar amounts of these activated CIAR variants lead to similar levels of activated RAS and downstream signaling through the RAF/MEK/ERK pathway.

We then characterized the temporal features of our CIAR variants, again using the canonical RAF/MEK/ERK pathway as a readout. Twenty-four hours after dox induction, CIAR variant-expressing cells were serum starved overnight, followed by treatment with A1 for various amounts of time. We found that ERK was rapidly activated in cells expressing CIAR KRAS4A HVR, CIAR KRAS4B HVR, CIAR HRAS HVR, or CIAR NRAS HVR. (**Figure 2.2e, 2.2f, 2.2g, 2.2h**). Interestingly, cells expressing CIAR KRAS4A HVR, CIAR KRAS4B HVR, or CIAR

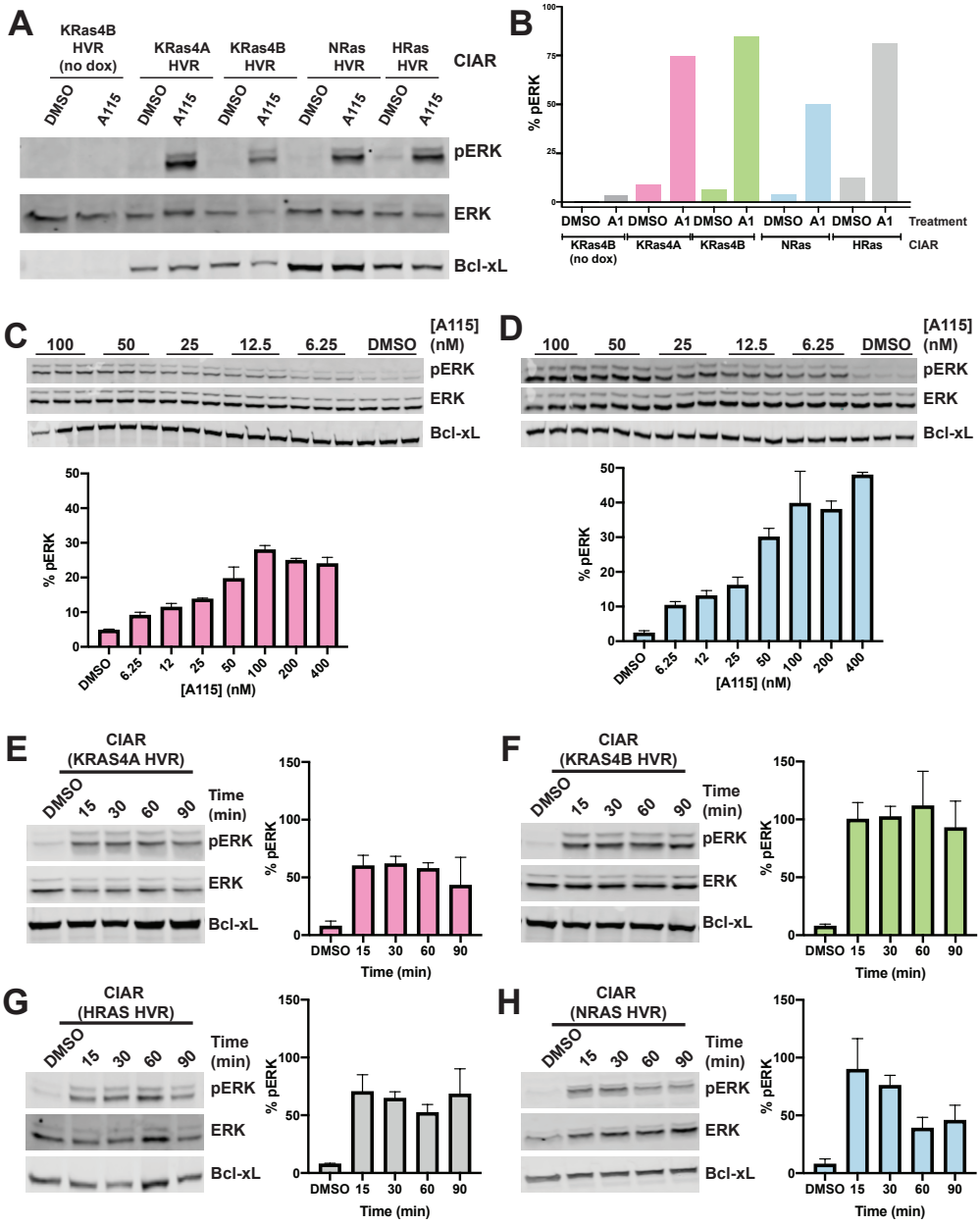

Figure 2.2. ***CIAR variants serve as a RAS rheostat.*** (A) Representative blot of CIAR variants treated with 200 nM A1 for 30 minutes following an overnight serum starvation. (B) Quantitation of % pERK from the blot shown in A. Intensity values of pERK and ERK was measured, with % pERK being calculated by (pERK intensity/ERK intensity) x 100. (C-D) Immunoblot analysis of dose-dependent activation of CIAR KRAS4A HVR (C) or CIAR NRAS HVR (D) in HEK293 cells. Cells were treated with varying concentrations of A1 for 30 minutes following an overnight serum starvation. (E-H) Immunoblot analysis of CIAR KRAS4A HVR (E), CIAR KRAS4B HVR (F), CIAR HRAS HVR (G), or CIAR NRAS HVR (H) treated with 200 nM A1 for the times indicated. Values shown are mean ± SEM (n=1 for (B), and n=3 for (C-H) except for the 200 and 400 nM treatment in (C) and (D) where n=2)

HRAS HVR demonstrated sustained pERK during the timecourse of the experiment but pERK levels diminished after 30 minutes of A1 treatment in CIAR NRAS HVR-expressing cells. Thus, pERK activation kinetics in cells expressing the CIAR NRAS HVR variant were most similar to cells treated with epidermal growth factor (EGF), which exhibit a rapid and transient activation of ERK[41,42]. Activation of CIAR NRAS HVR led to peak pERK levels at 30 minutes, which then decreased over the next hour. In summary, our experiments demonstrate that all four CIAR variants are equally able to activate RAS/RAF/MEK/ERK signaling but that there are some differences in the overall kinetics of activation depending on the localization of CIAR.

2.2.3 *Interrogation of downstream RAS signaling using phosphoproteomics*

After using our four CIAR variants to look at the effects of direct RAS activation on a single signaling node (the RAF/MEK/ERK pathway), we wanted to probe downstream signaling dynamics on a global scale. Specifically, we wanted to examine changes in cell-wide phosphorylation state using phosphoproteomics. To do this, we used label free quantification (LFQ) coupled with an IMAC phosphoenrichment/LC-MS/MS to analyze HEK293 cells stably expressing CIAR KRAS4A HVR, CIAR KRAS4B HVR, or CIAR NRAS HVR. Cells were treated with DMSO, or A1 for 30 or 90 minutes in triplicate following an overnight serum starvation, lysed, digested, and subsequently subjected to IMAC phosphopeptide enrichment (**Figure 2.3a**)[43]. In total, ~6000 unique phosphosites per treatment were identified. Phosphosites that had a -log(p-value) > 1.3 (p-value<0.05) and an enrichment score of >1 or <-1 (>2-fold difference between DMSO and A1-treated conditions) were considered significant. In cells stably expressing CIAR KRAS4A HVR, there were 502 upregulated and 673 downregulated significant phosphosites at 30 minutes and 1095 significantly upregulated sites and 98 downregulated sites at 90 minutes (**Figure 2.3b**). Cells stably expressing CIAR KRAS4B HVR had 286 upregulated and 186 downregulated

significant phosphosites at 30 minutes and 343 upregulated and 338 downregulated significant phosphosites at 90 minutes (**Figure 2.3c**). Cells expressing CIAR NRAS HVR had 999 upregulated and 134 downregulated significant phosphosites at 30 minutes and 862 upregulated and 161 downregulated significant

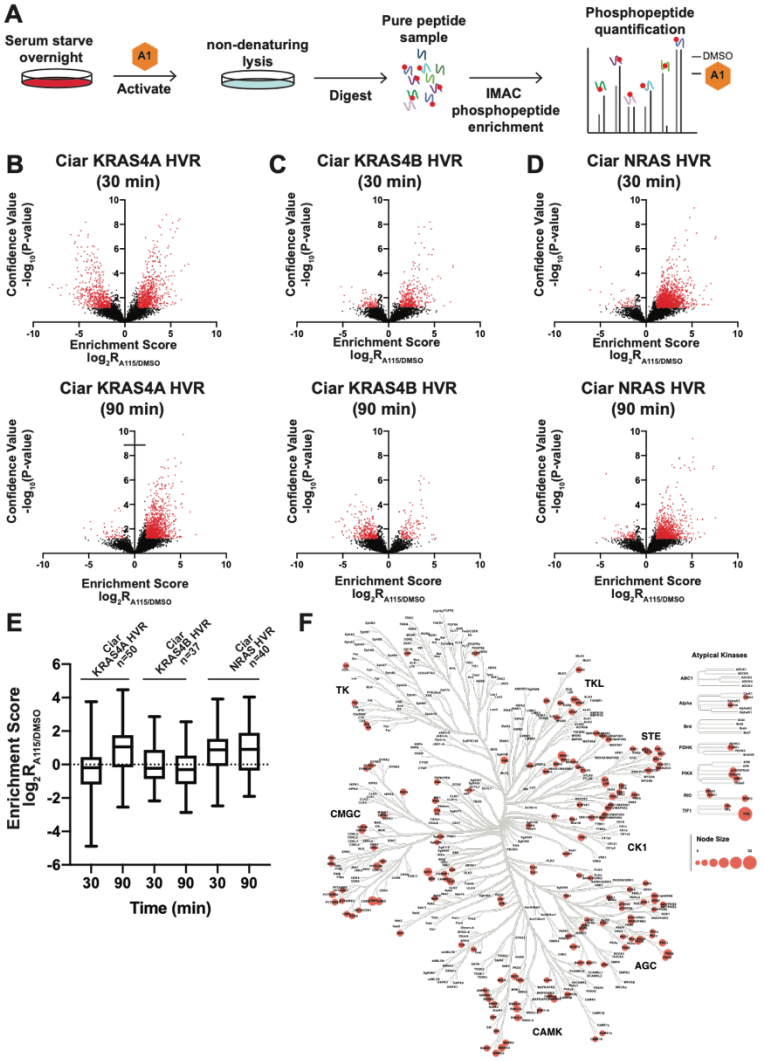

Figure 2.3. ***Total phosphoenrichment of activated CIAR variants.*** (A) Schematic of total phosphoenrichment. (B-C) Volcano plots of phosphosite results following treatment with A1 for 30 (top) or 90 (bottom) minutes of CIAR KRAS4A (B), CIAR KRAS4B (C), and CIAR NRAS (D). Red points are considered significant hits. (E) Log2 LFQ-MS ratios of known ERK1/2 substrate sites (PhosphositePlus database) in response to the indicated treatment. (F) Identified kinases from the total phosphoenrichment for each of the CIAR variants at 30 minutes with identified phosphosites.

phosphosites at 90 minutes (**Figure 2.3d**). Although a substantial number of phosphosites were identified and considered significant, very few were shared between all three CIAR variants. Unsurprisingly, the sites that were shared were on proteins that are abundant in cells, such as cytoskeletal proteins, but also included phosphosites on the activation loop of ERK1/2. These results demonstrate, similar to the immunoblot characterization, that each of the CIAR variants is able to activate the canonical RAF/MEK/ERK signaling pathway, but also illustrate the limitations of this approach. Without any form of enrichment, we mainly observe the phosphorylation state of the abundant proteins in cells, rather than downstream proteins that drive RAS-mediated signaling pathways.

Because ERK activation was observed for all three CIAR variants, we wondered if there was variability in the ability of the variants to activate flux through this pathway. We looked at quantified ERK substrate sites in our samples (50 in CIAR KRAS4A HVR, 37 in CIAR KRAS4B HVR, and 40 in CIAR NRAS HVR) and compared enrichment scores. We found that trends in the substrates were similar to pERK trends from the immunoblot analysis (**Figure 2.3e**). For CIAR KRAS4A HVR, the enrichment score of ERK substrates increased between 30 and 90 minutes, while the enrichment score of ERK substrates identified in CIAR KRAS4B HVR samples remained unchanged between the two time points. For CIAR NRAS HVR, there was a slight decrease in the ERK substrate score between 30 and 90 minutes, like in pERK. A positive shift in enrichment score indicates activation of ERK kinase, whereas a negative shift indicates deactivation through these pathways.

Although the global phosphoproteomics approach was useful in further defining the downstream effects of activating different CIAR variants, the overlap of significant phosphosites between CIAR variants was insufficient to make meaningful comparisons. While a global

phosphoenrichment approach allows for an unbiased measurement of protein phosphosites, we were particularly interested in the phosphorylation state of protein kinases, proteins that are known to be involved in RAS-mediated signaling pathways. Although we identified over 170 kinases, 85% of these had less than 5 identified phosphosites, and only ~5% of our significant hits for each variant and timepoint were kinase phosphosites (**Figure 2.3f**). Kinases are phosphotransferases, and commonly regulate many eukaryotic signaling pathways[44]. Yet, unlike many of their substrates, kinases are not abundant in cells, making them underrepresented in global phosphoproteomic data sets[45]. To circumvent this issue, we chose to pursue an approach that would first allow us to enrich for kinases prior to phosphoenrichment.

2.2.4 *Probing RAS-mediated signaling using kinobeads combined with phosphoproteomics*

Following our efforts to investigate RAS-mediated signaling pathways using total phosphoproteomics, we decided to take a more focused approach, specifically looking at protein kinases. A previously reported protocol demonstrated that the utilization of kinobeads coupled with phosphoproteomics allowed for quantification of system-wide changes in the phosphorylation state of kinases, which many times is directly coupled to their activation state[46]. Kinobeads are multiplexed inhibitor beads, composed of seven nonselective ATP-competitive kinase inhibitors that are able to interact with the majority of the human kinome[47,48]. In addition to enriching kinases, this method has also been shown to pull down hundreds of non-kinase proteins, proteins that do not bind ATP. In order for these proteins to be enriched, we hypothesized that they must be interacting with the kinobead-bound kinases, allowing for identification of multi-protein signaling complexes. These complexes will differ depending on the activation state of the kinase, making it an ideal global readout for signaling pathways activated by our CIAR variants.

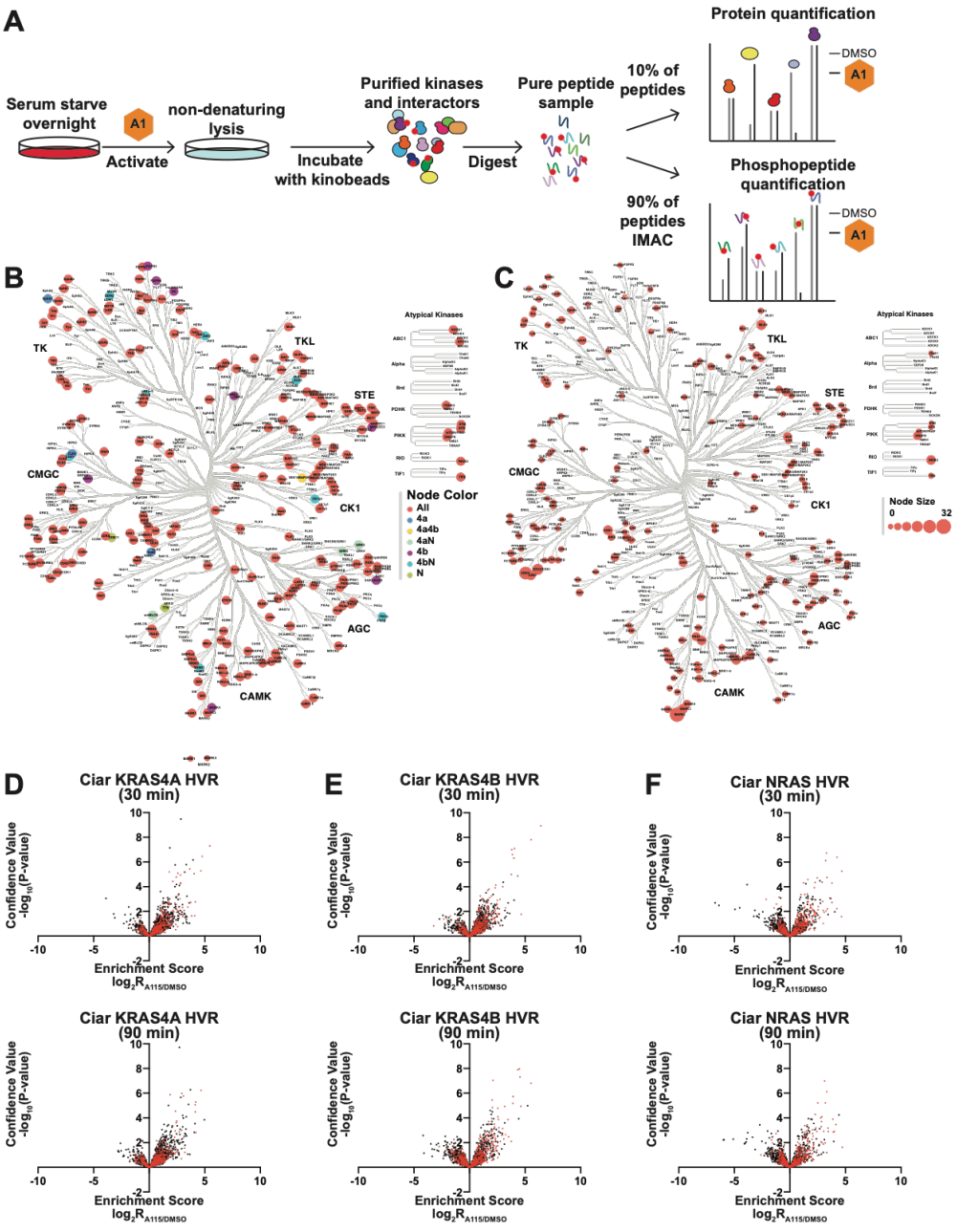

Figure 2.4. ***Kinobead enrichment combined with phosphoproteomic analysis.*** (A) Schematic of kinobead and phosphoenrichment workflow. (B) Kinome tree showing identified kinases from the kinobead pulldown from experiments performed with CIAR variant-expressing cells that have been treated with A1 for 30 minutes following an overnight serum starvation. (C) Kinome tree showing kinases that contain a quantified phosphosite in experiments performed with CIAR variant-expressing cells that have been treated with A1 for 30 minutes. Node size corresponds to number of phosphosites per kinase. (D-F) Volcano plots of phosphosite results for 30 (top) or 90 (bottom) minute A1 treatments of CIAR KRAS4A HVR (B), CIAR KRAS4B HVR (C), and CIAR NRAS HVR (D). Red points are kinase phosphosites.

Similar to the total phosphoenriched samples, we used LFQ and IMAC phosphoenrichment, with an added upfront kinobead enrichment step (**Figure 2.4a)**. Six biological replicates of cells expressing CIAR KRAS4A HVR, CIAR KRAS4B HVR, or CIAR NRAS HVR were treated with A1 for 30 or 90 minutes. After A1 treatment, cells were lysed, incubated with the kinobeads, and enriched proteins were digested on-bead. Following digestion, 10% of the sample was retained for total protein analysis, while the remaining 90% was subjected to the same IMAC phosphoenrichment procedure used in the global phosphoenrichment. To check sample integrity, prior to the kinobead enrichment, lysates were checked for activation of downstream ERK, and found to follow similar trends to those originally observed by initial western blot validation (*data not shown*). These results indicate that the proteomics approach should reflect the previously observed signaling dynamics on a cellular level.

As an initial validation, we focused on the results of the total protein analysis, which quantifies proteins identified in the samples following kinobead pulldown. We theorized that similar amounts and identities of kinases should be pulled down since each CIAR variant was made in the background of HEK293, and the same kinobead mixture was used for each sample. Gratifyingly, we found that between each variant and treatment, ~200 kinases were identified, with the majority of those kinases being identified in all variants (**Figure 2.4b**). Of these 200 kinases, 72% had 5 or less identified phosphosites, and 28% had greater than 5 phosphosites, allowing us to confidently examine activated signaling pathways (**Figure 2.4c**)

Next, we examined the results from the phosphoenrichment following the kinobead pulldown following treatment with A1 for 30 minutes. In the CIAR KRAS4A HVR samples, 1638 phosphosites were identified, with 582 being kinase phosphosites (**Figure 2.4d**). 1992 total phosphosites were identified in CIAR KRAS4B HVR samples, of which 728 were on kinases

(**Figure 2.4e**). CIAR NRAS HVR samples contained 1869 identified phosphosites, with 677 belonging to kinases (**Figure 2.4f**). Significant hits were determined in the same manner as in the total phosphoenrichment, and we found that ~40% of significant hits across all variants and timepoints were kinase phosphosites. Of the 317 total unique significant kinase phosphosites identified, 72 had known functional roles in kinase signaling pathways according to the PhosphositePlus database, which can be used to directly indicate the activation state of the kinase. Additionally, numerous phosphosites were identified for which the kinase-substrate relations were known, which can be used as indirect reporters of kinase activity. This combination of direct and indirect phosphosite evidence demonstrates the utility of a kinobead enrichment approach in identifying flux through signaling pathways. Using these results, with the increased emphasis on signaling proteins, we set out to reconstitute RAS-mediated signaling pathways following activation of the different CIAR variants.

2.2.5 *Phosphoproteomics can be used to examine localized RAS signaling*

One of the utilities of CIAR is that it can be used to specifically activate RAS on biologically relevant time scales. As an initial proof of concept, we chose to look at the canonical RAF/MEK/ERK pathway, a pathway that is conserved between the different RAS isoforms, even though flux through this pathway may differ depending on which pool of RAS is activated by CIAR. Following activation of CIAR KRAS4A HVR, CIAR KRAS4B HVR, and CIAR NRAS HVR for 30 or 90 minutes with A1, 14 phosphosites were identified on RAF isoforms, 7 were identified on MEK isoforms, and 2 were identified on ERK isoforms (**Figure 2.5a**). Although most of these phosphosites had similar intensities between DMSO-treated samples and activated CIAR samples, substantial differences were observed for BRAF pS446/pS447, CRAS pS259, and all

ERK phosphosites (**Figure 2.5b**). Previous literature has reported that BRAF requires phosphorylation on three separate areas on the

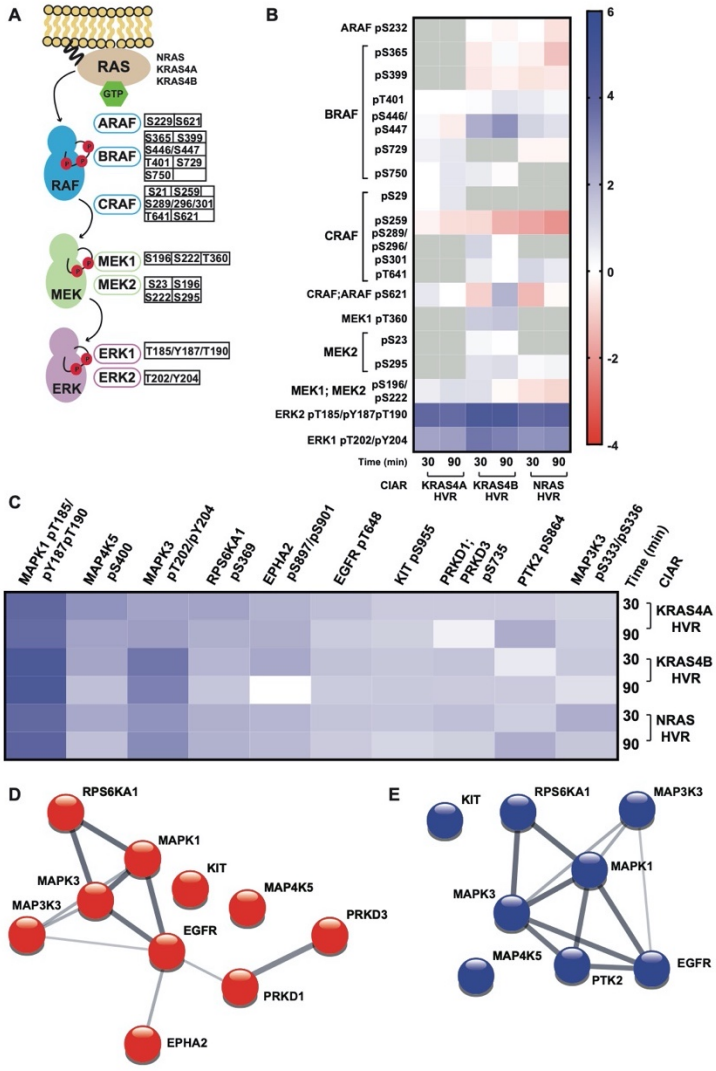

Figure 2.5. ***Signaling pathways activated by all CIAR variants.*** (A) Canonical RAS/RAF/MEK/ERK signaling pathway. Identified phosphosites are shown next to their corresponding protein. (B) Heat map showing the enrichment score of RAF/MEK/ERK phosphosites. Grey indicates the phosphosite was not identified for that variant or time point. (C) Heat map of kinase phosphosites shared between all CIAR variants. (D) String network of the shared kinase phosphosites from CIAR KRAS4A HVR, CIAR KRAS4B HVR, and CIAR NRAS HVR treated with A1 for 30 minutes following an overnight serum starvation (E) String network of the shared kinase phosphosites from CIAR KRAS4A HVR, CIAR KRAS4B HVR, and CIAR NRAS HVR treated with A1 for 90-minute following an overnight serum starvation.

kinase for full activation[85]. Residues Ser446 and Ser447, which exhibited a 4-fold increase in CIAR KRAS4B and a 2-fold increase in CIAR NRAS samples, are part of a SSDD motif, a region immediately N-terminal to the kinase domain, and one of the three subregion that requires phosphorylation for BRAF to be fully active[49]. This phosphorylation is reported to be performed by casein kinase 2 (CSNK2), which was also identified in all kinobead pulldown samples, and had significantly upregulated phosphosites in CIAR KRAS4B HVR samples at both 30 and 90 minute timepoints. On CRAF, phospho-Ser259 was identified in all samples, but interestingly found to have 2-fold higher intensity in DMSO treated samples than in CIAR KRAS4B HVR samples at 90 minutes, indicating that this site was being downregulated upon activation of the CIAR variants. Interestingly, the equivalent site on BRAF (Ser365) was also identified in CIAR KRAS4B HVR and CIAR NRAS HVR samples and was also found to be downregulated in the A1-treated samples. This site, along with Ser621 on CRAF, has been reported to be involved in phosphorylation-dependent binding of RAF to 14-3-3 proteins, which results in suppression of RAF catalytic activity, meaning that phosphorylation is inversely correlated to activation state. This decrease of phosphorylation, combined with the increased phosphorylation of ERK1, ERK2, MEK1, and MEK2 active site residues, which is directly correlated to activation state, is another indication that downstream pathway members are subsequently activated upon activation of CIAR.

Next, we wanted to examine other shared downstream signaling pathways following activation of the CIAR variants. There were 27 phosphosites across both 30- and 90-minute A1-treated timepoints that were considered significant and found in all variants, with 14 belonging to kinases (**Figure 2.5c**). To better visualize how these shared kinase phosphosites interacted, we generated a STRING interaction network, and found that unsurprisingly, the core of these networks

were kinase members of the RAF/MEK/ERK pathway, demonstrating the overlap in the biological function of RAS isoforms in this canonical pathway (**Figure 2.5d, 2.5e**).

2.2.6 *Expansion of signaling networks using non-kinase phosphosites*

Although kinobeads are used as an initial enrichment tool, we see that additional non-kinase proteins and phosphosites are identified in our results. These proteins do not bind ATP, and therefore should not interact with the kinobeads. These proteins must interact with the kinases themselves, suggesting that this method, in addition to looking at kinase signaling pathways, can be used to analyze the composition of assembled kinase signaling complexes. Similar to how phosphorylation state can be indicative of activation state, composition of signaling complexes can also be used as an indicator of activation[50,51].

To examine this, shared non-kinase phosphosites from the kinobead/phosphoenrichment data and non-kinase proteins from the kinobead pulldown data that were significantly changed upon activation of the CIAR variants were added into the shared changed kinase phosphosites (**Figure 2.6a**) Combining these results, the string networks were expanded to show interactions with non-kinases, allowing us to speculate on RAS involvement in the recruitment and regulation of these proteins. For example, at 30 minutes, we see that in addition to the core kinase signaling network, CAD is also added into the connected network (**Figure 2.6b**). CAD is a multifunctional protein that is responsible for initiating and regulating pyrimidine biosynthesis[52]. Although there are residues on CAD that are known to be phosphorylated by ERK kinases, Ser1859 has recently been reported to be phosphorylated by RPS6KA1, which is also included in the core signaling network following activation of the different CIAR variants, in an mTORC1 regulated fashion[53]. mTORC1 has also been reported to be activated by constitutively active RAS mutants, but this

data indicates that this activation is a component of wild-type RAS-mediated signaling that occurs in the first 30 minutes following RAS activation[54].

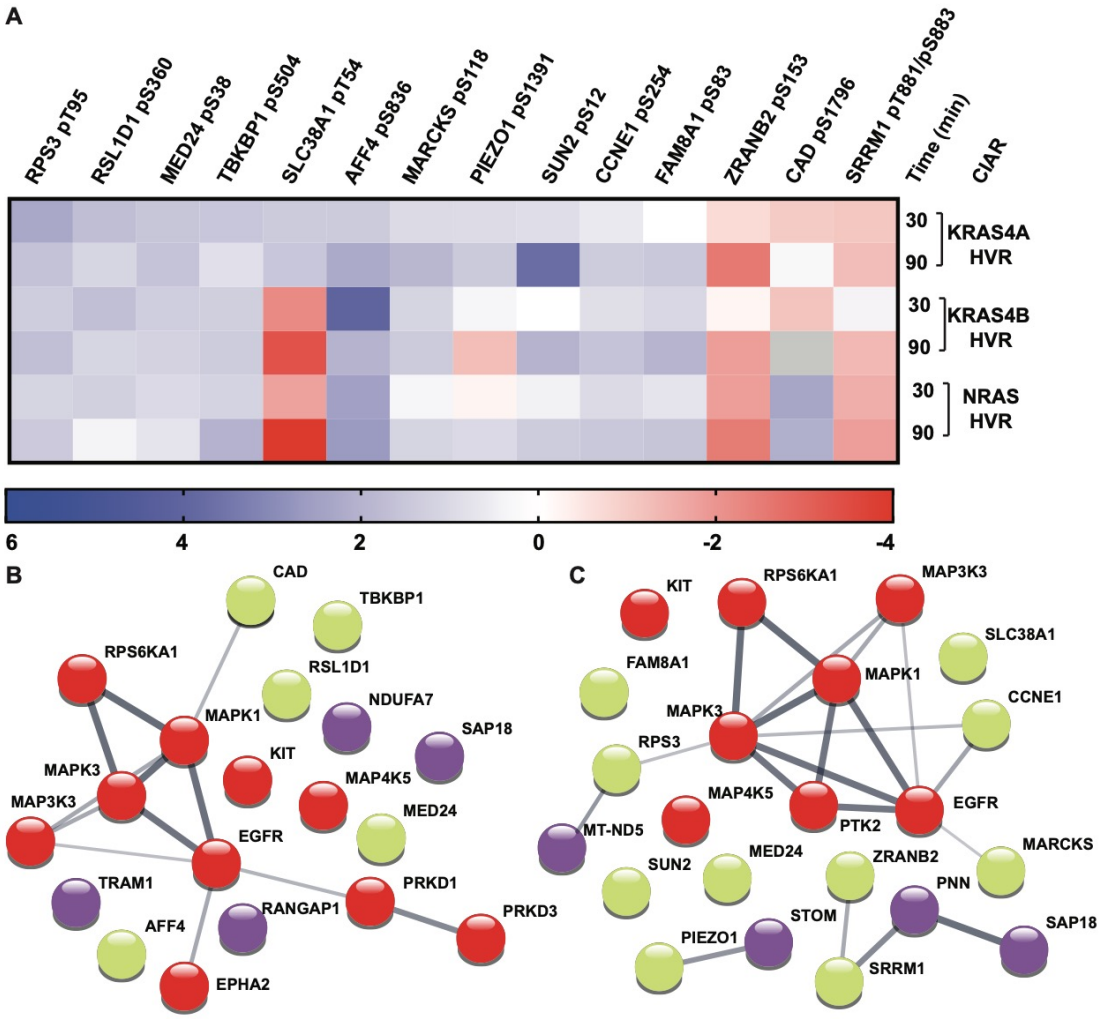

Figure 2.6. ***Non-kinase proteins and phosphosites can be used to expand signaling networks.*** (A) Heat map of shared non-kinase phosphosite enrichment scores for CIAR KRAS4A HVR, CIAR KRAS4B HVR, and CIAR NRAS HVR treated with A1 for 30 or 90 minutes. (B) String network of all shared significantly changed phosphosites from CIAR variants treated with A1 for 30 minutes. Kinase with shared phosphosites are shown in red, proteins that have shared significantly changed phosphosites are shown in yellow, and proteins that are significantly changed between DMSO- and A1- treated samples are shown in purple. (C) String network of all shared hits from the 90-minute time point. Kinase with shared phosphosites are shown in red, proteins that have shared significantly changed phosphosites are shown in yellow, and proteins that are significantly changed between DMSO- and A1- treated samples are shown in purple.

In addition to the kinases in the 90-minute connected network, CCNE1, MARCKS, RPS3, MT-ND5 are also incorporated into the core signaling network (**Figure 2.6c**). While Ser118 on myristoylated alanine-rich C-kinase substrate (MARCKS) has only previously been identified in phosphoproteomic studies, it is known to be involved in regulation of cytoskeletal dynamics. Additionally, previous literature has reported that MARCKS can be myristoylated to engage in plasma membrane binding and contains a calmodulin-binding domain to interact with calmodulin, which has been reported to interact with KRAS[55,56]. A single phosphorylation event on CCNE1, an oncogene, on a peptide containing Ser381, Ser 387, and Thr395 is also upregulated in all CIAR variants at 90 minutes. Ser381 has only been identified as a phosphosite in high-throughput mass spectrometry studies, but Ser387 and Thr395 have been extensively studied and are known to be phosphorylated by CDK2 and GSK3A/GSK3B[57]. Phosphorylation of these phosphosites on CCNE1 are essential for the progression of cells through the G1/S cell cycle transition, a transition which also requires active ERK[58]. Another shared phosphosite, Thr221 on RPS3, is known to be phosphorylated by PKC which causes RPS3 to switch between regulating transcription to DNA repair[59]. Both CCNE1 and RPS3 are localized to the nucleus, demonstrating the far-reaching effects of prolonged RAS activation. Although these phosphosites have not been explicitly linked to RAS-mediated signaling previously, further study of these proteins and their associated phosphosites could elucidate their roles in wild-type RAS-mediated signaling pathways.

2.2.7 *Ras-mediated signaling differences can be identified using phosphoproteomics coupled with kinobead pulldowns*

By examining the proteins and phosphosites that are significantly changed following CIAR activation, we can define the networks that are shared upon activation of all CIAR variants. One of the utilities of localizing CIAR to different subcellular locations is that CIAR can be used to

activate different localized pools of RAS. To take advantage of this feature, we next chose to look at kinase phosphosites that were significantly and uniquely changed in A1-treated samples from a single CIAR variant. As a first step, String was again used to visualize the networks of kinases with unique phosphosites (**Figure 2.7, 2.8**) Unsurprisingly, many of these kinases are involved in a single highly connected network, although smaller networks were identified in CIAR KRAS4A HVR and KRAS4B HVR treated with A1 for 30 minutes, as well as CIAR NRAS HVR treated with A1 for 30 and 90 minutes (**Figure 2.7a, 2.7c, 2.7e, 2.7f**).

Since we have established that the CIAR variants follow similar localization patterns as the RAS isoforms from which their HVR sequences are derived, we looked at the GO: Cellular Component (GOCC) terms associated with each of the kinases to see if there were detectable localization patterns (**Figure 2.8a**). With CIAR KRAS4A HVR and CIAR KRAS4B HVR, we observed that there was a change in the subcellular localization of kinases with unique phosphosites between 30 and 90 minutes, specifically with kinases located at the nucleus and plasma membrane. At 30 minutes, 35% of CIAR KRAS4B HVR and 40% of CIAR KRAS4A HVR kinases identified had associated nuclear terms. At 90 minutes, this percentage increased to 50% and 70% respectively. KRAS4A and KRAS4B are known to be predominately localized to the plasma membrane, so we interpret this change of associated GOCC terms in the propagation of signals throughout the cell. After 30 minutes of A1 treatment, we are observing the primary effects of RAS activation. By 90 minutes, these RAS-mediated signals have propagated throughout the cell, resulting in secondary or tertiary effects of direct RAS activation at subcellular locations further from the initial point of activation. Interestingly, there is a decrease of kinases that have associated plasma membrane GOCC terms between 30 and 90 minutes, for CIAR KRAS4A HVR,

while this number stays relatively the same in CIAR KRAS4B HVR, perhaps indicating increased shuttling of KRAS4A between the plasma membrane and other

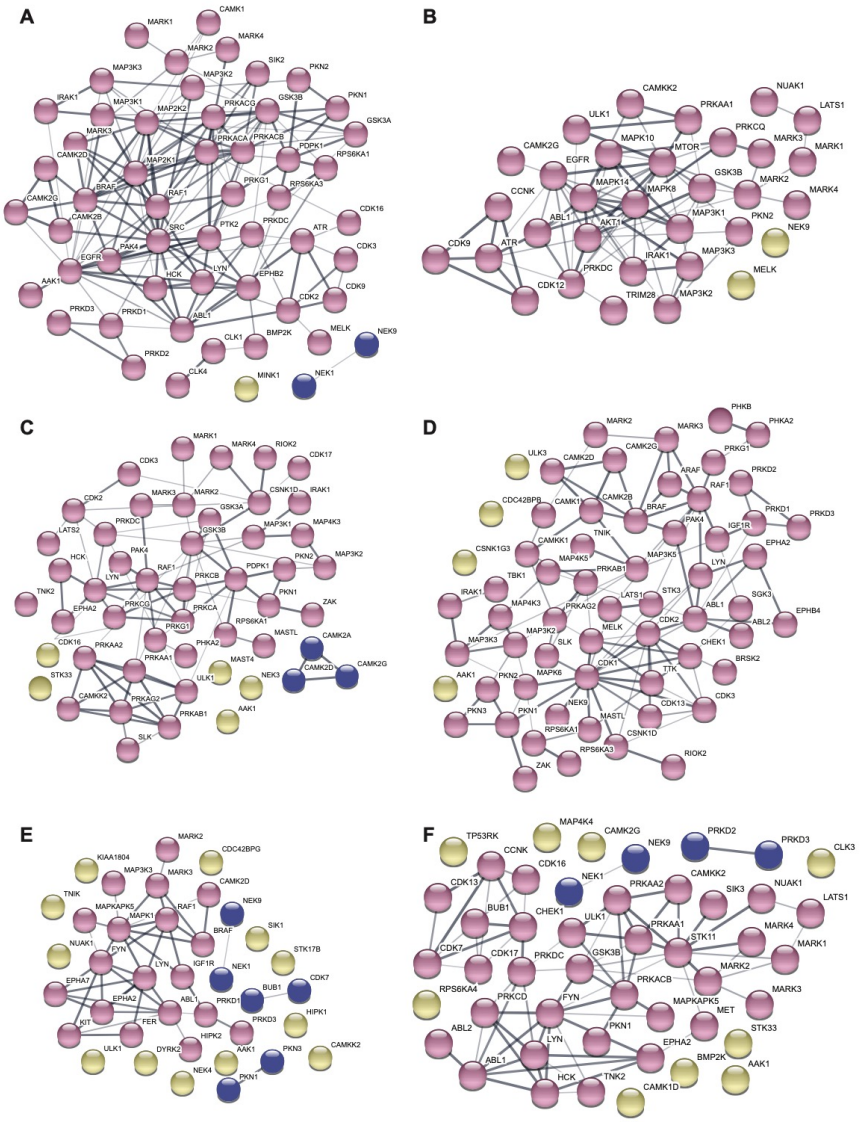

Figure 2.7. ***String networks of kinases with unique phosphosites.*** (A-B) String network of kinases with unique phosphosites identified in CIAR KRAS4A HVR cells treated with A1 for 30 minutes (A) or 90 minutes (B) following an overnight serum starvation. (C-D) String network of kinases with unique phosphosites identified in CIAR KRAS4B HVR cells treated with A1 for 30 (C) or 90 (D) minutes following an overnight serum starvation. (E-F) String network of kinases with unique phosphosites identified in CIAR NRAS HVR cells treated with A1 for 30 (E) or 90 (F) minutes following an overnight serum starvation. Kinases in the core network are colored in pink, kinases that make up mini networks are in blue, while unconnected kinases are in yellow.

intracellular endomembranes, while KRAS4B remains anchored to the plasma membrane, resulting in constant activation of these membrane associated proteins. For kinases with unique phosphosites identified in CIAR NRAS HVR lysates, there is little change in localization between 30 and 90 minutes, perhaps due to the intracellular localization of NRAS, resulting in less necessity for propagation of signals and faster execution of downstream effects.

Next, the associated GO: Biological Processes (GOBP) of the kinases with unique phosphosites were compared. GOBP terms were considered if they included >5 associated proteins from the data set and >5-fold enrichment compared to the expected enrichment for a comparable sized data set (**Figure 2.8b**). Unsurprisingly, all CIAR variants at both 30 and 90 minute timepoints were highly enriched in protein autophosphorylation, as well as other phosphorylation related terms, such as peptidyl tyrosine/serine/threonine phosphorylation/autophosphorylation/modification. Biological processes from both CIAR KRAS4A HVR and CIAR KRAS4B HVR were found to be enriched in terms pertaining to intracellular transport, specifically between the nucleus and the cytoplasm. CIAR KRAS4A HVR and CIAR NRAS HVR also each had unique kinases associated with regulation of cellular senescence, while, interestingly for CIAR KRAS4B HVR unique kinases were found to be enriched in the regulation of the mitotic cell cycle phase. This increased involvement in the cell cycle could explain the relative abundance of KRAS4B mutations over other RAS isoforms in human cancers and could provide important targets for future treatment of these diseases.

Lastly, we wanted to incorporate the non-kinase phosphosites from the kinobead/phosphoenrichment data and the non-kinase proteins from the kinobead pulldown data to expand our signaling networks. Unique hits that showed significant changes between DMSO

and A1 treatment were identified, and again visualized using String (**Figure 2.8c**). Excitingly, with the addition of the unique non-kinase interactors, kinases that were previously exempt from

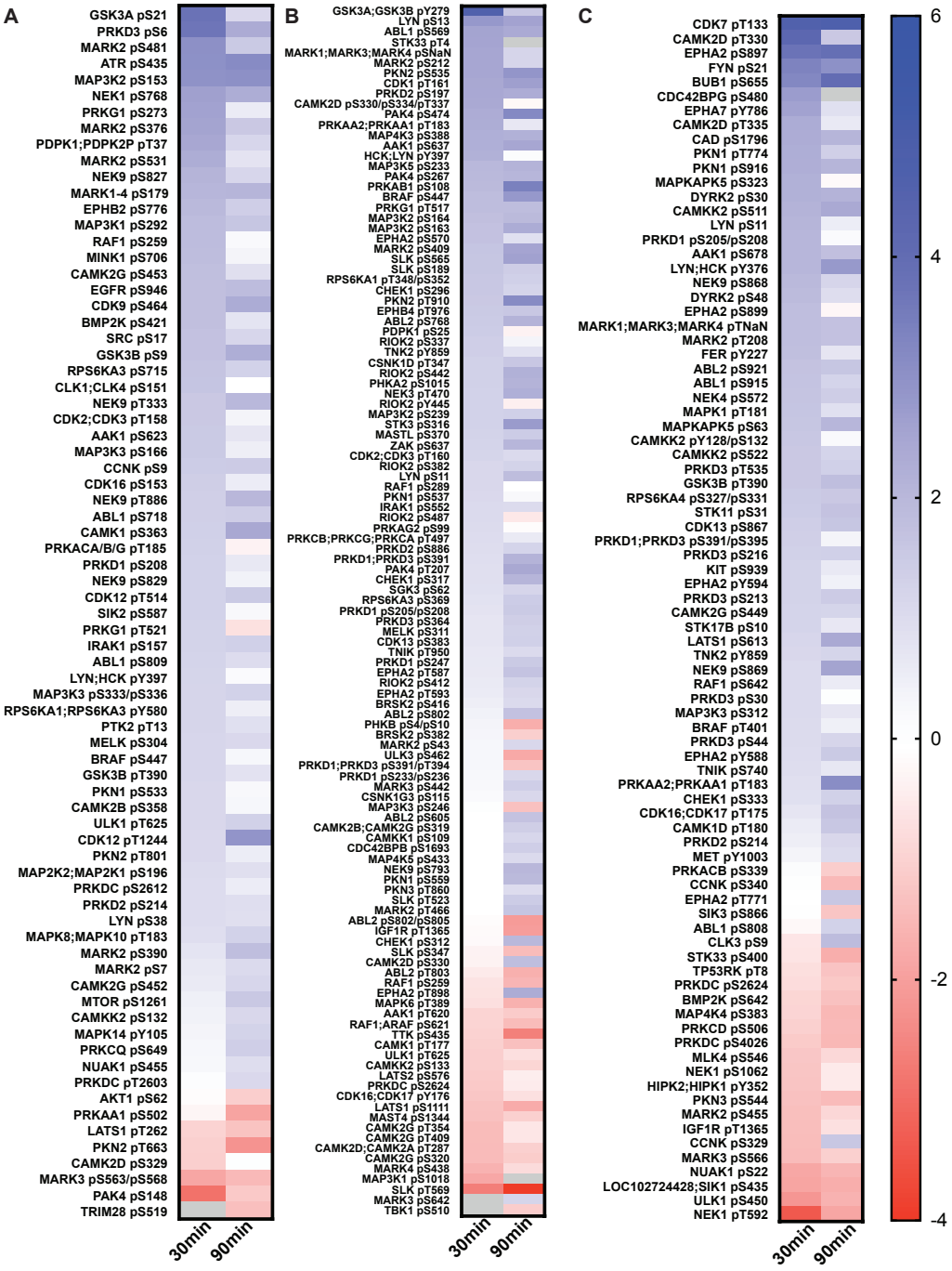

Figure 2.8. ***Heatmaps of unique kinase phosphosites.*** (A) Heatmap of unique kinase phosphosites identified in CIAR KRAS4A HVR samples following treatment with A1 for either 30 or 90 minutes. (B) Heatmap of unique kinase phosphosites identified in CIAR KRAS4B HVR samples following treatment with A1 for either 30 or 90 minutes. (C) Heatmap of unique kinase phosphosites identified in CIAR NRAS HVR samples following treatment with A1 for either 30 or 90 minutes.

the network were now included. For example, in the CIAR KRAS4A HVR 30 minutes data shown, NEK1 and NEK9 previously composed their own mini network, while MELK was only connected to the network through CDK2. With the addition of the non-kinase interactors, NEK1 and NEK9 were incorporated into the network through interactions with a number of nuclear porin proteins (NUPs), while MELK was now at the core of this expanded network. When comparing GOBP terms that were enriched in these expanded networks, the addition of the non-kinase interactors solidified the signaling pathways that were initially identified when only using the unique kinases. For example, many of the CIAR KRAS4B HVR terms were still associated with the regulation of early mitosis, CIAR KRAS4A HVR terms were involved in membrane signaling and intracellular trafficking, and CIAR NRAS HVR terms were involved in Golgi-ER trafficking as well as regulation of the respiratory electron transport chain. The addition of the non-kinase interactors serves as a confirmation of the kinases-based signaling pathways, and these expanded networks can be used to better understand RAS-mediated signaling complexes induced upon direct RAS activation.

2.3 DISCUSSION

In this chapter, I have discussed the utilization of a chemically inducible, intramolecularly regulated RAS activator. RAS mutations are at the root of many human cancers, but without a full understanding of normal RAS signaling, it has been difficult to effectively treat these cancers. Additionally, we know that there are functional differences and divergent subcellular localization patterns between the four different human RAS isoforms, but on a molecular level little is known about how these isoform-specific signaling processes and regulation differ.

Using four CIAR variants, each with an HVR sequence taken from one of the human RAS isoforms, I demonstrate that these intramolecular protein switches are localized in a manner similar

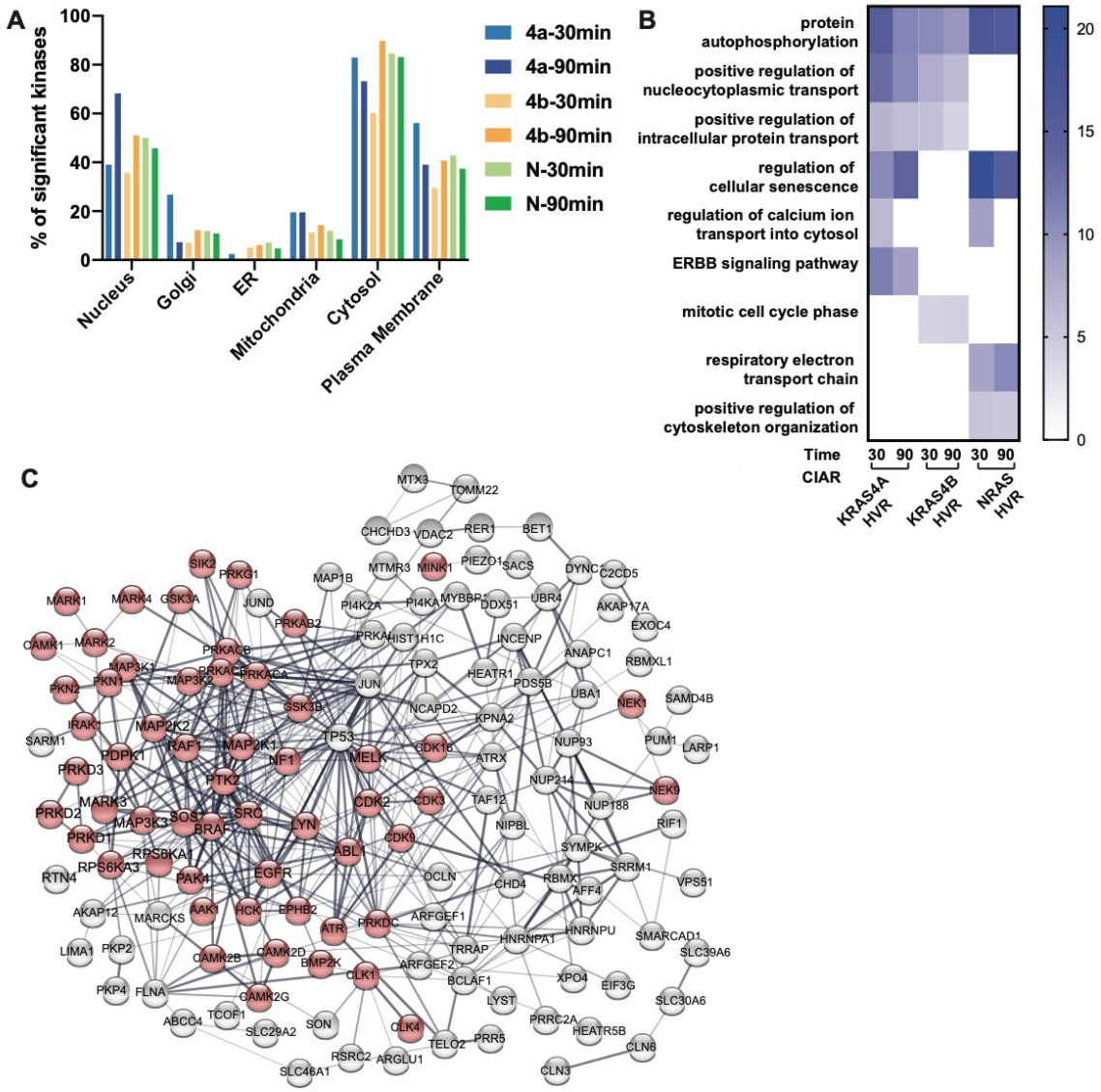

Figure 2.9. ***Analysis and expansion of unique signaling networks.*** (A) GOCC terms associated with kinases that contained a unique phosphosite. GOCC percentage terms were calculated using kinases containing unique phosphosites as a percentage of the total unique kinases in a variant at the specified timepoint. (B) Heatmap of GOBP terms using the fold enrichment calculated by the GOCC database. (C) Expansion of the CIAR KRAS4A 30-minute network. Proteins containing unique phosphosites that were significantly changed were added into the unique kinase network. Kinases are shown in red, while non-kinase interactors are shown in grey.

to that of analogous endogenous RAS proteins. I then looked at the ability of these switches to activate ERK. While ERK is not a direct RAS effector, RAS proteins are known to recruit and activate RAF, which in turn phosphorylates MEK, which finally phosphorylate ERK. We observe

that each of the four CIAR variants is able to induce flux through this pathway in a rheostatic manner, resulting in an increase of phosphorylated ERK that can be controlled rheostatically through the amount of small molecule added.

With the help of proteomics, we expanded our examination of RAS-mediated signaling to the cellular level. Using total phosphoproteomics, we were able to look at the activation state of known ERK effectors, indicating that information on RAS signaling on a global scale was possible. Ultimately, many of the proteins that were identified in our results were proteins that were abundant in the cell, such as those involved in cytoskeletal regulation. To combat this problem, we switched to a phosphoproteomic approach that uses an upstream enrichment of kinases, signaling proteins whose activity is directly related to their phosphorylation state.

The kinobead enrichment method combined with phosphoproteomics allowed us to examine signaling pathways shared upon activation of RAS that had different subcellular localizations. Unsurprisingly, many of these pathways centered around the mitogen-activated protein kinases (MAPKs), but also included receptor tyrosine kinases such as FAK (PTK2) and EPHA2, a logical result since the RAS proteins are all known to be localized at the plasma membrane. Combining the phosphosite information of the non-kinase interactors, as well as the non-kinases identified in our kinobead pulldown, we can expand these networks to speculate on other proteins that might be in complexes with these RAS isoforms.

Finally, we can compare data sets to see which signaling pathways are unique to each of the RAS isoforms. By solely looking at phosphosites and proteins that are uniquely identified following activation of a single CIAR variant, we can identify pathways and subcellular locations that are potentially isoform specific. This information can be used to categorize the effects of each RAS isoform, and can be used for more focused study of RAS in the future. This data has the

potential to expand the amount of known RAS effectors, allowing for a thorough understanding of these molecular mechanisms and ultimately more target options when treating diseases caused by aberrant RAS activity.

2.4 Materials and Methods

2.4.1 *Cell culture, treatment, and harvest*

Flp-In™ T-REx™ 293 cells (ThermoFisher) and HEK293T cells were maintained in DMEM supplemented with 10% FBS (MilliporeSigma) at 37°C with 5% CO_2. For mammalian expression, all constructs were made in pcDNA5/FRT/TO (ThermoFisher). pcDNA5/FRT/TO was cut by HpaI and KpnI and inserts and vector were assembled using Gibson assembly. Constructs were stably integrated into Flp-In™ T-REx™ 293 cells following the manufacturer's protocol using turbofectin (Origene) as the transfection reagent and hygromycin as the selection reagent. Expression of the CIAR variants was induced with 1 µg/mL doxycycline (ThermoFisher) 48 hours before treatment with A-1155463 (A1, ChemieTek), then 18 hours before treatment, cells were serum-starved in FBS-free media. Control treatment was vehicle (DMSO). For harvest of proteomics samples, cells were grown on 15 cm plates, and were washed twice with ice cold PBS before harvesting. Cells used in the global phosphoproteomics analysis were lysed in 350 µL 6 M aq. guanidinium HCl with 100 mM Tris, 5 mM tris(2-carboxyethyl) phosphine hydrochloride (TCEP*HCl) and 10 mM chloroacetamide (CAM), while cells used in kinase-enriched phosphoproteomics analysis were lysed in proteomics modified RIPA buffer (50 mM Tris, 150 mM NaCl, 0.25% Na-deoxycholate, 1% IGEPAL CA630, 1 mM EDTA and 10 mM NaF, pH 7.8) supplemented with HALT protease inhibitor cocktail (100x, Thermo Fisher Scientific) and phosphatase inhibitor cocktail II and III (100x, MilliporeSigma). Western blot samples were lysed

in modified RIPA (50 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, 1% IGEPAL CA630, Pierce protease inhibitor tablet (MilliporeSigma)) supplemented with 2 mM sodium orthovanadate and 30 mM sodium fluoride (100 µL for 12-well plates) after washing once with ice cold PBS. Cells were collected with a cell scraper, lysates were vortexed intermittently at max speed 5x for 3s each, then clarified by centrifugation (20 min at 21,0000 xg and 4 °C). Protein content of lysates was determined using the Pierce 660 nm assay reagent (Thermo Fisher Scientific).

2.4.2 *Extended methods for RAS activation*

1. Flp-In™ 293 T-REx™ cells with stably integrated CIAR are maintained in high glucose DMEM supplemented with 10% FBS at 37 °C with 5% CO_2 until cells are ~80% confluent in a T75 flask.
2. Once cells have reached confluence, remove the DMEM, and wash with 3 mL of DPBS. Add 2 mL of 0.25% Trypsin-EDTA at room temperature until cells detach, then add complete DMEM to quench the trypsin.
3. Move cells to a conical tube, then centrifuge for 3 minutes at 800 rpm to pellet cells. Remove media and resuspend in complete DMEM, making sure to disrupt all cell clumps.
4. Count cells using a hemocytometer or automatic cell counter.
5. Dilute cells to $2.5x10^5$ cells/mL. Plate 1 mL of cells on a poly-d lysine coated plate. Add 1 µg/mL doxycycline in complete DMEM to initiate expression of CIAR.
6. Gently rock the plate back and forth to distribute cells and doxycycline evenly throughout the well.
7. Return the cells to the incubator. Incubate cells overnight at 37 °C with 5% CO_2.
8. Remove cells from step 7 from the incubator 24 hours after plating.

9. Remove the media from the adherent cells. Wash cells twice with 0.5 mL DPBS by gently adding down the side of the well, then gently rock the plate back and forth to cover the entire surface. Remove DPBS from well.
10. Add 1 μg/mL doxycycline in 1 mL high glucose DMEM (no FBS).
11. Return cells to incubator and incubate cells overnight at 37 °C with 5% CO_2.
12. Weigh out 25 mg of A1.
13. Add 3.540 mL DMSO to A1 to make 10 mM stocks.
14. Pipet solution into 50 μL aliquots. Store aliquots at -20 °C.
15. Dilute A1 aliquot to desired concentration in DMSO. For time course studies, we typically dilute the aliquots to 200 μM, resulting in a final concentration of 200 nM A1 when 1 μL is added to the wells.
16. 16-18 hours after switching media, remove the cells from step 11 from the incubator. Add DMSO or 1 μL of the desired concentration of A1 directly to the media.
17. Rock the plate gently back and forth to distribute the A1, then return the cells to the incubator for the desired length of time.
18. After desired length of A1 treatment time, remove plate from step 17 from the incubator.
19. Remove the media and place cells on ice. Wash cells once with 0.5 mL ice cold DPBS, then add 100 μL modified RIPA lysis buffer.
20. Detach cells from well by using a cell scraper or by gently pipetting up and down to release cells from plate. Put lysate in a labeled sterile 1.5 mL Eppendorf tube on ice.
21. Vortex cells on medium speed three times for 3 seconds, then pellet debris by centrifuging in a tabletop centrifuge for 10 min, max speed, 4 °C.

22. Transfer 60 µL supernatant to a clean Eppendorf. Add 30 µL 3x SDS, then boil the samples for 7 minutes at 98 °C.
23. Remove gel from packaging. Set up SDS-PAGE system and add 1x MOPS running buffer. Load 2 µL of the protein ladder and 10 µL of the boiled samples from step 22 on the precast gel. Run the gel at 180 V until the dye front reaches the bottom of the gel, usually about 30 minutes.
24. Assemble the transfer sandwich. Remove the transfer pack from packaging and place the top stack on the bottom tub of the Trans-Blot® Turbo™ cassette.
25. Once the dye front has reached the bottom of the gel, remove the gel from the gel rig. Separate the plastic casing from the gel and place the gel on the nitrocellulose membrane.
26. Place the top stack from the Transblot® Turbo™ Midi Transfer Packs on top of the gel, then close the cassette cover. Place the cassette in the Trans-Blot® Turbo™ and run the MIDI 7-minute program. This will transfer the samples from the gel to the nitrocellulose.
27. Once the program has finished, remove the transfer sandwich from the cassette. Cut the nitrocellulose down to the size of the gel, then remove the gel from the nitrocellulose. Put the nitrocellulose into the western blot incubation box.
28. Add Odyssey® Blocking Buffer to the incubation block and incubate for 1 hour at room temperature with gentle agitation on the orbital shaker.
29. Remove blocking buffer, then add primary antibodies (anti-Bcl-xL antibody (diluted 1:2500), anti-ERK (diluted 1:2500) and anti-phospho-ERK (diluted 1:2500)) in Odyssey® Blocking Buffer and incubate overnight with gentle agitation at 4 °C on an orbital shaker.
30. The next day, remove primary antibody solution, and wash nitrocellulose membrane three times with TBST for five minutes.

31. Add both secondary antibodies (each diluted 1:10,000) in Odyssey® Blocking Buffer. Incubate at room temperature with gentle agitation for 1 hour.
32. Wash membrane three times with TBST for five minutes each, then image the membrane on a Li-Cor Odyssey® infrared imager.
33. Load imaged blot into Image Studio™ Lite software.
34. To quantify bands, select a single channel (700 or 800). Under the "Analysis" tab choose "Draw rectangle." Draw a single rectangle over each band of interest.
35. Under the "Shapes" tab, find the "Signal" value. This corresponds to the background-subtracted band of interest that was in the drawn rectangle.
36. Use signal values to compare samples. When using antibodies for ERK and phospho-ERK we normalize using the following formula:

$$\frac{phospho - ERK\ signal}{ERK\ signal} * 100 = \%\ phospho - ERK$$

2.4.3 *Microscopy*

Flp-In™ T-REx™ 293 cells with fluorescently tagged CIAR variants stably incorporated were plated on 18 mm glass slides (ThermoFisher) in 12-well tissue culture plates at a density of $2x10^4$ cells/mL in DMEM supplemented with FBS and doxycycline. 24 hours after plating, media was removed, cells were washed twice with PBS, then fixed with a 4% paraformaldehyde mixture in PBS (Electron Microscopy Sciences) for 15 minutes. After fixing, cells were washed with PBS, then stained with wheat germ agglutinin conjugated to Alexafluor647 (ThermoFisher) for an additional 15 minutes. Slides were mounted with Fluoromount G (Southern Biotechnology), and sealed with nail polish, then imaged with a Leica SP8X confocal scope.

2.4.4 *Western Blotting*

Clarified cell lysates in modified RIPA buffer were mixed with 3x SDS (240 mM Tris pH 6.8, 30% glycerol, 16% beta-mercaptoethanol, 6% w/v SDS, 0.03% bromophenol blue in water). Samples were heated to 95°C for 7 min, then proteins were separated on Any kD™ Mini-PROTEAN® TGX™ precast polyacrylamide gel (BioRad). Proteins were then electro-transferred onto nitrocellulose membranes using Transblot® Turbo™ Midi Transfer Packs (BioRad). Membranes were blocked with Odyssey Blocking Buffer (LiCor) for 1 hour, then incubated with primary antibodies overnight (1:2500 dilution in Odyssey Blocking buffer). The next day, membranes were washed with TBST, incubated with IRDye® 800CW Goat anti-Rabbit IgG Secondary Antibody and IRDye® 680LT Goat anti-Mouse IgG Secondary Antibody (Li-Cor) for 1 hour, then visualized on a Li-Cor Odyssey® infrared imager. Primary antibodies (all from Cell Signaling Technologies) used were anti-phospho-p44/42 MAPK (Erk1/2)(Thr202/Tyr204) (D13.14.4E) XP Rabbit mAb (Cat #4370), p44/42 MAPK (Erk1/2) (3A7) Mouse mAb (Cat # 9107), and Anti-Bcl-xL Rabbit mAb(Cat #2762).

2.4.5 *Preparation of optimized kinobead mixture*

The seven kinobead affinity reagents used were synthesized in-house as described previously[39.40]. For optimal coverage of the human kinome, an adjusted mixture of the seven kinobead reagents was prepared to make 15 mL of the complete kinobead mixture (Table 1). NHS Act Sepharose® 4 Fast Flow (MilliporeSigma) bead amounts were measured and washed twice with DMSO to remove storage buffer. Beads were resuspended in DMSO to make a 50% slurry. Compounds were weighed out and dissolved in their corresponding Bead/DMSO slurry with triethylamine (MilliporeSigma), then let to react at room temperature in the dark in an end-over-

end rotator. After 20 hours, 50 µL/mL bed volume aminoethanol was added and left to react for another 20 hours. Solutions were then washed with DMSO (10 mL/mL bed volume) followed by 20% ethanol (3x10 mL/bed volume) then combined. Combined beads were washed 3x with 15 mL 1:1 dimethylformamide (DMF): ethanol, then incubated overnight with a 33% DMF/33% ethanol/34% 1 M aminoethanol (in 20 mM acetic acid in 1:1 DMF:ethanol) with 150 µL 1 M EDCI per 0.5 mL bed volume in the dark at room temperature. The next day, beads were washed 3x with 15 mL 1:1 DMF:ethanol, 2x with 15 mL 0.5 M NaCl, then once with 150 mL 20% ethanol. Beads were stored as a 50% slurry in 20% ethanol.

2.4.6 *Preparation of peptides for global phosphoproteomic analysis*

Cell lysates were vortexed briefly, then heated to 95°C for 5 minutes, then sonicated at 100 W for 10 min (30 seconds on, 30 seconds off) on ice. Protein content was measured using the Pierce 660 nM assay reagent (ThermoFisher). Aliquots of 300 µg of protein were diluted 2-fold with 100 mM TEAB pH 8.5, then adjusted to pH 8-9 with 1 N aqueous NaOH. 3 µg MS-grade lysyl endopeptidase Lys-C (Wako) were added and agitated on a thermomixer at 1400 rpm at 37°C. After two hours, the mixture was again diluted 2-fold with 100 mM TEAB, and 3 µg MS-grade trypsin (ThermoFisher) were added. The mixture was agitated on a thermomixer at 800 rpm overnight at 37°C, then acidified with formic acid (1% final). Samples were cleared by centrifugation for 10 min at room temperature and 14,000 xg. Peptides were extracted from the supernatant using Oasis HLB 1cc (10 mg) extraction cartridges (Waters). Cartridges were activated by passing through 200 µL methanol, followed by 200 µL 80% aqueous ACN with 0.1% TFA, then equilibrated with 400 µL 1% formic acid. Peptides were loaded, washed with 400 µL 1% formic acid, then eluted with 300 µL 80% aq. ACN with 0.1% TFA. Following elution, peptides were subjected to IMAC phosphopeptide enrichment (See below).

2.4.7 *Kinase affinity enrichment and on-bead digestion*

Three micro tubes containing 50 µL of a 50% slurry of the kinobead mixture in 20% aqueous ethanol were prepared for each pulldown experiment. The beads were washed twice with 300 µL proteomics modified RIPA buffer. 1 mg of protein extract in proteomics modified RIPA buffer were added to the first tube, then incubated on an end-over-end rotator for 1 hour at 4°C. The beads were spun down rapidly at 2000 rpm on a benchtop centrifuge for 5 seconds, then the supernatant was pipetted into the second tube containing kinobeads. This procedure was repeated for a total of three rounds of affinity enrichment. After removal of the supernatant, the beads were washed twice with 300 µL ice cold proteomics modified RIPA buffer and three times with 300 µL ice cold tris buffered saline (TBS, 50 mM Tris pH 7.8, 150 mM NaCl) to remove detergents. 100 µL of the denaturing buffer (20% trifluoroethanol (TFE), 25 mM Tris containing 5 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP*HCl), and 10 mM chloroacetamide (CAM) pH 7.8) were added to the beads, and the slurries were vortexed briefly at low speeds, then beads from the same experiments were combined. The combined slurries were heated at 95°C for 5 minutes, then diluted 2-fold with 25 mM triethylamine bicarbonate (TEAB). The pH was adjusted to 8-9 by addition of 1N aq. NaOH, then 5 µg MS-grade lysyl endopeptidase Lys-C (Wako) were added and agitated on a thermomixer at 1400 rpm at 37°C. After two hours, 5 µg MS-grade trypsin (ThermoFisher) were added, and the mixture was agitated on a thermomixer at 800 rpm overnight at 37°C. 600 µL of 1% formic acid were added, then the mixture was acidified by adding an additional 6 µL formic acid for a final volume of 1.2 mL peptide solution. 10% of the total volume (120 µL) was aliquoted and desalted using StageTips and analyzed by LC-MS/MS for background protein quanitification[41]. The remaining 90% of the peptide solution was dried under vacuum at room temperature on a SpeedVac. 300 µL of 70% aq. acetonitrile (ACN) with 0.1% TFA was

added to each tube, then the mixture was vortexed and sonicated until dried peptide residue was fully dissolved. If necessary, an additional 0.1% aq. TFA was added in 10 µL increments until all residue was fully dissolved. These samples were then subjected to IMAC phosphopeptide enrichment (See below).

2.4.8 *IMAC phosphopeptide enrichment*

IMAC phosphopeptide enrichment was performed following the published protocol with the following modifications[42]. A 20 µL mixture of a 50% IMAC bead slurry composed of 1/3 commercial PHOS-select iron affinity gel (MilliporeSigma), 1/3 in-house made Fe^{3+}-NTA Superflow agarose and 1/3 in-house made Ga^{3+}-NTA Superflow agarose was washed three times with 10 bed volumes of 80% aq. ACN containing 0.1% TFA then used for phosphopeptide enrichment[43]. Following the enrichment, phosphopeptides were desalted using StageTips with a modified phosphopeptide protocol[41]. StageTips were activated with 50 µL methoanol and 50 µL 80% aq. ACN with 0.1% TFA, then equilibrated with 50 µL 1% aq. Formic acid. Peptides were resuspended in 50 µL 1% aq. formic acid, loaded onto the StageTip, washed with 50 µL 1% aq. formic acid, then eluted with 50 µL 80% aq. ACN containing 0.1% TFA.

2.4.9 *LC-MS/MS analysis*

Following elution from the StageTips, samples were dried on a SpeedVac then resuspended in aq. 0.1% TFA. Peptide samples were separated on a nanoAcquity UPLC instrument with 12 cm long fused silica capillary column (Polymicro Technologies Flexible Fused Silica Capillary Tubing, Inner Diameter 75 µm, Outer Diameter 385 µm) made in house with a laser puller and packed with 5 µm 120 Å reversed phase C18 beads (ReproSil-Pur 120 C18-AQ, Dr. Maisch GmbH HPLC). The LC gradient was 90 minutes long with 10-35% B at 300 nL/min. LC solvent A was

0.1% acetic acid and LC solvent B was 0.1% acetic acid and 99.9% ACN. MS data were collected with a Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer.

2.4.10 *Proteomic data processing*

Raw data files were analyzed by MaxQuant/Andromeda (www.maxquant.org) using the default settings with the addition of variable modifications set to include Phospho (STY) for phosphoenriched samples. MS/MS spectra were searched against the UniProt human database (updated July 22nd, 2015). The data from the Protein Groups and Phospho (STY) results table was processed with Perseus software (http://www.coxdocs.org/doku.php?id=perseus:start#download). Proteins were only considered identified if MaxQuant was able to compute corresponding protein intensity values, and proteins classified as "only identified by site", "reverse", or "contaminant" were filtered out, and proteins with 1 or fewer peptide counts were filtered and removed. Expression columns for both protein and phosphopeptide intensities were log2 transformed and normalized by subtracting the median log2 expression value from each expression value of the corresponding data column. Data imputation was performed using a modeled distribution of MS intensity values downshifted by 1.8 and having a width of 0.2. P-values were calculated using a two-tailed, two-sample T-test. Enrichment scores (log_2[intensity (A1-treated)/intensity(DMSO-treated)]) were calculated using the relative intensity values from each CIAR variant treated with A1 or DMSO with confidence values (-log(P-value)) for each identified protein). Significant hits were defined as protein or phophosites with an enrichment score >1.0 and a confidence value >1.3. Bioinformatics analysis was using the String and PhosphositePlus database.

2.5 REFERENCES

1. Karnoub, A. E., Weinberg, R. A. (2008). Ras oncogenes: split personalities. *Nature Reviews Molecular Cell Biology*, *9*(7), 517–531.

2. Simanshu, D. K., Nissley, D. V., McCormick, F. (2017). RAS Proteins and Their Regulators in Human Disease. *Cell, 170,* 17-33.

3. Papke, B., Der, C. J. (2017). Drugging RAS: Know the enemy. *Science, 355,* 1158-1163.

4. Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J., Der, C. J. (2014). Drugging the undruggable RAS: Mission Possible? *Nature Reviews Drug Discovery, 14,* 828-851.

5. Thompson, H. (2013). US National Cancer Institute's new Ras project targets an old foe. *Nature Publishing Group*, 1–2.

6. Boriack-Sjodin, P.A., Margarit, S.M., Bar-Sagi, D., Kuriyan, J. (1998). The structural basis of the activation of Ras by Sos. *Nature, 394*, 337–343.

7. Cherfils, J., Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiological Reviews, 93,* 269-309.

8. Schlichting, I., Almo, S. C., Rapp, G., Wilson, K., Petratos, K., Lentfer, A., Wittinghofer, A., Kabsch, W., Pai, E. F., Petsko, G. A., Goody, R. S. (1990). Time-resolved X-ray crystallographic study of the conformational change. *Nature*, *345*, 309–315.

9. Hunter, J. C., Manandhar, A., Carrasco, M. A., Gurbani, D., Gondi, S., Westover, K. D. (2015). Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. *Molecular Cancer Research*, *13*(9), 1325–1335.

10. Prior, I. A., Hancock, J. F. (2012). Ras trafficking, localization and compartmentalized signalling. *Seminars in Cell and Developmental Biology*, *23*(2), 145–153.

11. Omerovic, J., Laude, A. J., Prior, I. A. (2007). Ras proteins: paradigms for compartmentalised and isoform-specific signalling. *Cellular and Molecular Life Sciences*, *64*(19-20), 2575–2589.

12. Fehrenbacher, N., Bar-Sagi, D., Philips, M. (2009). Ras/MAPK signaling from endomembranes. *Molecular Oncology*, *3*(4), 297–307.

13. Koera, K., Nakamura, K., Nakao, K., Miyoshi, J., Toyoshima, K., Hatta, T., Otani, H., Aiba, A., Katsuki, M. (1997). K-Ras is essential for the development of the mouse embryo. *Oncogene*, *15*, 1151–1159.

14. Umanoff, H., Edelmann, W., Pellicer, A., Kucherlapati, R. (1995). The murine N-ras gene is not essential for growth and development. *Proceedings of the National Academy of Sciences*, *92*, 1709–1713.

15. Esteban, L. M., Vicario-Abejon, C., Fernandez-Salguero, P., Fernandez-Medarde, A., Swaminathan, N., Yienger, K., Lopez, E., Malumbres, M., McKay, R., Ward, J. M.,

Pellicer, A., Santos, E. (2001). Targeted genomic disruption of H-ras and N-ras, individually or in combination, reveals the dispensability of both loci for mouse growth and development. *Molecular and Cellular Biology*, *21*(5), 1444–1452.

16. Plowman, S. J., Williamson, D. J., O'Sullivan, M. J., Doig, J., Ritchie, A. M., Harrison, D. J., Melton, D. W., Arends, M. J., Hooper, M. L., Patek, C. E. (2003). While K-ras Is Essential for Mouse Development, Expression of the K-ras 4A Splice Variant Is Dispensable. *Molecular and Cellular Biology*, *23*(24), 9245–9250.

17. Johnson, L., Greenbaum, D., Cichowski, K., Mercer, K., Murphy, E., Schmitt, E., Bronson, R. T., Umanoff, H., Edelmann, W., Kucherlapati, R., Jacks, T. (1997). K-ras is an essential gene in the mouse with partial functional overlap with N-ras. *Genes and Development*, *11*, 2468–2481.

18. Hobbs, G. A., Der, C. J., Rossman, K. L. (2016). RAS isoforms and mutations in cancer at a glance. *Journal of Cell Science*, *129*(7), 1287–1292.

19. Ahearn, I. M., Haigis, K., Bar-Sagi, D., Philips, M. R. (2011). Regulating the regulator: post-translational modification of RAS. *Nature Reviews Molecular Cell Biology*, *13*(1), 39–51.

20. Herrero, A., Matallanas, D., Kolch, W. (2016). The spatiotemporal regulation of RAS signalling. *Biochemical Society Transactions*, *44*(5), 1517–1522.

21. Hancock, J. F. (2003). Ras proteins: different signals from different locations. *Nature Reviews Molecular Cell Biology*, *4*(5), 373–385.

22. Cunningham-Bryant, D., Dieter, E. M., Foight, G. W., Rose, J. C., Loutey, D. E., Maly, D. J. (2019). A Chemically Disrupted Proximity System for Controlling Dynamic Cellular Processes. *Journal of the American Chemical Society*, *141*(8), 3352–3355.

23. Rose, J. C., Dieter, E. M., Cunningham-Bryant, D., Maly, D. J. (2018). Examining RAS pathway rewiring with a chemically inducible activator of RAS. *Small GTPases*, *0*, 1–8.

24. Rose, J. C., Huang, P. S., Camp, N. D., Ye, J., Leidal, A. M., Goreshnik, I., Trevillian, B. M., Dickinson, M. S., Cunningham-Bryant, D., Debnath, J., Baker, D., Wolf-Yadlin, A., Maly, D. J. (2017). A computationally engineered RAS rheostat reveals RAS–ERK signaling dynamics. *Nature Chemical Biology*, *13*(1), 119–126.

25. Toettcher, J. E., Weiner, O. D., Lim, W. A. (2013). Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. *Cell*, *155*(6), 1422–1434.

26. Goreshnik, I., Maly, D. J. (2010). A Small Molecule-Regulated Guanine Nucleotide Exchange Factor. *Journal of the American Chemical Society, 132,* 938-940.

27. Konopleva, M., Contractor, R., Tsao, T., Samudio, I., Ruvolo, P. P., Kitada, S., et al. (2006). Mechanisms of apoptosis sensitivity and resistance to the BH3 mimetic ABT-737 in acute myeloid leukemia. *Cancer Cell*, *10*(5), 375–388.

28. Wei, G., Margolin, A. A., Haery, L., Brown, E., Cucolo, L., Julian, B., et al. (2012). Chemical Genomics Identifies Small-Molecule MCL1 Repressors and BCL-xL as a Predictor of MCL1 Dependency. *Cancer Cell*, *21*(4), 547–562.

29. Tao Z. F., Hasvold L., Wang L., Wang, X., Petros, A. M., Park, C. H., et al. (2014). Discovery of a Potent and Selective BCL-XL Inhibitor with in Vivo activity. *ACS Medicinal Chemistry Letters, 5*, 1088-1093.

30. Reiss, Y., Goldstein, J. L., Seabra, M. C., Casey, P. J., Brown, M. S. (1990) Inhibition of purified p21ras farnesyl:protein transferase by Cys-AAX tetrapeptides. *Cell* 62, 81–88.

31. Boyartchuk, V. L., Ashby, M. N., Rine, J. (1997) Modulation of Ras and a-factor function by carboxyl-terminal proteolysis. *Science* 275, 1796–1800.

32. Hrycyna, C. A., Sapperstein, S. K., Clarke, S., Michaelis, S. (1991) The *Saccharomyces cerevisiae STE14* gene encodes a methyltransferase that mediates C-terminal methylation of a-factor and RAS proteins. *EMBO J*. 10, 1699–1709.

33. Dai, Q., Choy, E., Chiu, V., Romano, J., Slivka, S. R., Steitz, S. A., Michaelis, S., Philips, M. R. (1998) Mammalian prenylcysteine carboxyl methyltransferase is in the endoplasmic reticulum. *J. Biol. Chem*. 273, 15030–15034.

34. Hancock, J. F., Paterson, H., Marshall, C. J. (1990) A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell* 63, 133–139.

35. Apolloni, A., Prior, I. A., Lindsay, M., Parton, R. G., Hancock, J. F. (2000) H-ras but not K-ras traffics to the plasma membrane through the exocytic pathway. *Mol. Cell. Biol*. 20, 2475–2487.

36. Choy, E., Chiu, V. K., Silletti, J., Feoktistov, M., Morimoto, T., Michaelson, D., Ivanov, I. E., Philips, M. R. (1999) Endomembrane trafficking of Ras: the CAAX motif targets proteins to the ER and Golgi. *Cell* 98, 69–80.

37. Goodwin, J.S., Drake, K.R., Rogers, C., Wright, L., Lippincott-Schwartz, J., Philips, M.R., Kenworthy, A. K. (2005) Depalmitoylated Ras traffics to and from the Golgi complex via a non- vesicular pathway. J Cell Biol 170, 261–72.

38. Tsai, F. D., Lopes, M. S., Zhou, M., Court, H., Ponce, O., Fiordalisi, J. J., Gierut, J. J., Cox, A. D., Haigis, K. M., Philips, M. R. (2015). K-Ras4A splice variant is widely expressed in cancer and uses a hybrid membrane-targeting motif. *Proceedings of the National Academy of Sciences*, *112*(3), 779–784.

39. Mochizuki, N., Yamashita, S., Kurokawa, K., Ohba, Y., Nagai, T., Miyawaki, A., Matsuda, M. (2001). Spatio-temporal images of growth-factor-induced activation of RAS and Rap1. *Nature*, *411*, 1065–1068.

40. Roberts, P. J., Der, C. J. (2007). Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, *26*(22), 3291–3310.

41. Olsen, J.V. Blagoev, B., Gnad, F., Macek, B., Kuman, C., Mortensen, P., Mann, M. (2006) Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635–648.

42. Tarcic, G., Avraham, R., Pines, G., Amit, I., Shay, T., Lu, Y., Zwang, Y., Katz, M., Ben-Chetrit, N., Jacob-Hirsch, J., Virgilio, L., Rechavi, F., Mavrothalassitis, G., Mills, G. B., Eytan, D., Yarden, Y. (2012) EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. *FASEB J.* 26, 1582–1592.

43. Villen, J.; Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature protocols,* 2008**,** *3* (10), 1630-1638.

44. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science,* 2002, *298* (5600), 1912-34.

45. Daub, H.; Olsen, J. V.; Bairlein, M.; Gnad, F.; Oppermann, F. S.; Korner, R.; Greff, Z.; Keri, G.; Stemmann, O.; Mann, M., Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Molecular cell,* 2008**,** *31* (3), 438-48.

46. Golkowski, Vidadala, V. N., Lau, H.-T., Shoemaker, A., Shimizu-Albergine, M., Beavo, J., Maly, D. J., Ong, S-E. (2020). Kinobead/LC-MS Phosphokinome Profiling Enables Rapid Analyses of Kinase-Dependent Cell Signaling Networks. *J. Proteome Res.*, *2020*, 19, 1235-1247.

47. Golkowski, M., Brigham, J. L., Perera, B. G. K., Romano, G. S., Maly, D. J., Ong, S.-E. (2014). Rapid profiling of protein kinase inhibitors by quantitative proteomics. *Med. Chem. Commun., 5(3*), 363–369.

48. Golkowski, M., Vidadala, R. S. R., Lombard, C. K., Suh, H. W., Maly, D. J., Ong, S.-E. (2017). Kinobead and Single-Shot LC-MS Profiling Identifies Selective PKD Inhibitors. *Journal of Proteome Research*, *16*(3), 1216–1227.

49. Lavoie, H., & Therrien, M. (2015). Regulation of RAF protein kinases in ERK signalling. *Nature Publishing Group*, *16*(5), 281–298.

50. Daub, H., Olsen, J. V., Bairlein, M., Gnad, F., Oppermann, F. S., Korner, R., Greff, Z., Keri, G., Stemmann, O., Mann, M. (2008) Kinase-selective enrichment enables

quantitative phosphoproteomics of the kinome across the cell cycle. *Molecular cell* **31**, 438-448.

51. Fleuren, E. D., Zhang, L., Wu, J., Daly, R. J. (2016) The kinome 'at large' in cancer. *Nat Rev Cancer* **16**, 83-98

52. Sigoillot, F. D., Kotsis, D. H., Serre, V., Sigoillot, S. M., Evans, D. R., Guy, H. I. (2005). Nuclear localization and mitogen-activated protein kinase phosphorylation of the multifunctional protein CAD. *Journal of Biological Chemistry*, *280*(27), 25611–25620.

53. Ben-Sahra, I., Howell, J. J., Asara, J. M., & Manning, B. D. (2020). Stimulation of de Novo Pyrimidine Synthesis by Growth Signaling Through mTOR and S6K1. *Science*, *339*, 1323–1328.

54. Menon, S., Manning, B. D. (2019). Common corruption of the mTOR signaling network in human tumors. *Oncogene*, 1–9.

55. Matsubara, M., Titani, K., Taniguchi, H., Hayashi, N. (2003). Direct involvement of protein myristoylation in myristoylated alanine-rich C kinase substrate (MARCKS)-calmodulin interaction. *Journal of Biological Chemistry*, *278*(49), 48898–48902.

56. Lopez-Alcala, C., Alvarez-Moya, B., Villalonga, P., Calvo, M., Bachs, O., & Agell, N. (2008). Identification of Essential Interacting Elements in K-Ras/Calmodulin Binding and Its Role in K-Ras Localization. *Journal of Biological Chemistry*, *283*, 10621–10631.

57. Welcker, M., Singer, J., Loeb, K. R., Grim, J., Bloecher, A., Gurien-West, M., Clurman, B. E., Roberts, J. M. (2003). Multisite Phosphorylation by Cdk2 and GSK3 Controls Cyclin E Degradation. *Molecular Cell*, *12*(2), 381–392.

58. Chambard, J.-C., Lefloch, R., Pouyssegur, J., Lenormand, P. (2007). ERK implication in cell cycle regulation. *Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research*, *1773*(8), 1299–1310.

59. Kim, T.-S., Kim, H. D., Kim, J. (2009). PKCδ-dependent functional switch of rpS3 between translation and DNA repair. *BBA - Molecular Cell Research*, *1793*(2), 395–405.

Chapter 3. DEVELOPMENT OF A MAMMALIAN REPORTER SYSTEM FOR THE OPTIMIZATION OF A BIO-ORTHOGONAL CHEMICAL DISRUPTOR OF PROXIMITY

3.1 INTRODUCTION

Rational manipulation of cell signaling networks has emerged as a powerful tool to study and control the regulatory systems employed by cells. By using a user-defined input, such as light or a small molecule, the researcher can precisely and temporally control system engagement, mimicking the natural processes found in a cell and providing insight into how a cell controls these dynamic processes. Historically, these engineered control systems can broadly be separated into two categories; pre-translational and post-translational. Pre-translational systems, such as RNAi or gene knockouts, lead to a reduction in transcriptional outputs, allowing for precise removal of a protein of interest from a cell but with the caveat that there is minimal temporal control. In contrast, post-translational control systems target a protein of interest after it has been expressed. One example of post-translational control is the inhibition of a protein of interest with a small molecule. Small molecule inhibition allows for high temporal modulation by controlling when a drug is added. In addition, dose-dependent modulation can also be achieved by tuning the amount of small molecule that is added.

While post-translational control with small molecules is highly desirable, it is often very challenging or impossible to identify small molecules that that have sufficient potency and selectivity to be useful reagents in cells. To overcome this challenge, our group and others have combined chemical and genetic approaches that utilize tools in protein engineering and pharmacology[1-3]. These chemical genetic approaches allow for temporal and, in some cases, spatial

control over intracellular protein activity. Recently, our lab has reported the development of chemical genetic systems, where the activity of a protein interest is regulated with an engineered autoinhibitory intramolecular interaction that can be disrupted with a small molecule[4-6]. These chemical genetic systems that are controlled by chemically-disrupted proximity (CDP) are genetically encoded and can be localized in an autoinhibited state to specific locations within the cell. The activity of the engineered protein is rapidly restored upon small molecule addition, allowing for the study of rapid and graded cellular responses.

Our first-generation CDP system (see Chapter 2) has proven to be a powerful tool for probing temporal aspects of RAS-mediated signaling. However, this CDP system has a significant limitation. Specifically, it relies on the interaction of two endogenous mammalian proteins, making the chemical disrupters used in this system not completely orthogonal to mammalian biology. We developed a second-generation CDP system that uses an engineered Hepatitis C virus protease NS3-4A (NS3a) variant combined with a previously reported peptide inhibitor, which we call *apo* NS3a reader (ANR)[7]. The ANR peptide was initially developed by Kugler et al. using phage display and has been shown to bind to the NS3a active site. Numerous clinically-approved NS3a inhibitors that disrupt the NS3a/peptide interaction are available, which can be used as inputs that are orthogonal to mammalian cellular systems[8].

While our second-generation CDP system based on NS3a could be successfully used in numerous applications, such as creation of an autoinhibited activator of RAS (CIAR) and chemically-disruptable co-localization of fusion proteins, we found that it did not behave as ideally as expected based on computational modeling. In our NS3a-based CIAR system, we found that the level of autoinhibition observed in the absence of a chemical disrupter was lower than in the Bcl-xL-based system. In addition, we found that the NS3a/ANR interaction was inefficient in

intermolecularly recruiting transcriptional activators to the promoters of endogenous genes. To broaden the utility of our second-generation CDP system based on NS3a, I set out to further optimize the ANR peptide directly in mammalian cells. In this section, I will describe how I engineered a fluorescent reporter cell line for screening ANR variants. I also generated a library of ANR single variants that could be screened for their increased efficiency in interacting with NS3a. These ANR variants were screened in the reporter cell line I developed and favorable mutations were identified using next-generation sequencing technology. Finally, ANR single mutants and ANR variants containing multiple favorable mutations were validated in the fluorescent reporter cell line. These efforts resulted in an optimized ANR peptide that can be used in wide-ranging applications with NS3a.

3.2 RESULTS

3.2.1 *Generation of a reporter cell line*

For our second-generation CDP system, we desired a protein-protein interaction that is high affinity, stable in the intracellular environment, and that can be disrupted with small molecules that are orthogonal to mammalian cell biology. We felt that the interaction between the hepatitis C viral protease NS3a and a previously reported peptide inhibitor of NS3a, which we call *apo NS3a reader* (ANR), fulfilled these criteria[4,7]. ANR is a 21-amino acid peptide that was initially optimized for its affinity for catalytically active NS3a through phage display. ANR was found to have low double-digit nanomolar binding affinity for purified NS3a in an *in vitro* assay, which could be disrupted with drugs that bind to the NS3a active site. We found that the NS3a/ANR interaction could be integrated into our autoinhibited CIAR system to allow temporal control over RAS activation and as a chemically-disrupted intermolecular off-switch for controlling transcription[4]. However, we found that NS3a/ANR interaction was not ideal for some

applications. (**Figure 3.1a, 3.1b**). For example, our NS3a-based CIAR switch demonstrated higher basal activity than our Bcl-xL-based CIAR switch, suggesting that the NS3a/ANR interaction is not sufficiently tight to provide full autoinhibition (**Figure 3.1a**). Furthermore, while the NS3a/ANR interaction is capable of providing efficient transcriptional activation through the intermolecular recruitment of a transcriptional activator to an exogenous promoter, this system was found to be less robust at promoting transcriptional activation of endogenous genes (**Figure 3.1b**). These results are somewhat puzzling because the NS3a/ANR interaction should be of sufficiently high affinity for the aforementioned applications.

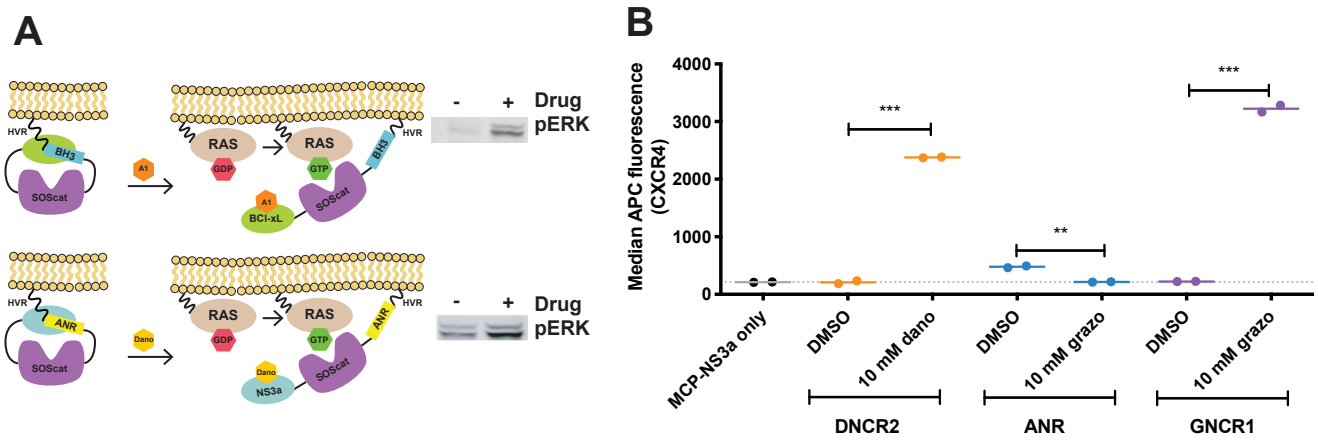

Figure 3.1. ***Characterization of the second-generation CDP system based on NS3a.*** (A) Comparison of the Bcl-xL-based and NS3a-based CIAR systems. Basal pERK levels, demonstrated by the representative pERK blot shown, are higher in cells expressing NS3a-based CIAR compared to Bcl-xL-based CIAR. This suggests that the interaction between NS3a and ANR is insufficiently strong to provide complete autoinhibition in the absence of drug. (B) Comparison of endogenous gene expression with different NS3a-based dimerization systems. Gene expression is robustly induced when DNCR2-VPR or GNCR1-VPR are recruited to promoter-localized NS3a with danoprevir or grazoprevir, respectively. ANR-VPR provides comparatively weak activation in cells expressing promoter-localized NS3a, suggesting that the NS3a/ANR interaction is not capable of robustly recruiting ANR-VPR to promoter-localized NS3a.

Therefore, I set out to further optimize the NS3a/ANR interaction. Because we felt that there was a disconnect between how the NS3a/ANR interaction behaved *in vitro* compared to in the intracellular environment, I developed a mammalian cellular reporter assay for screening ANR variants. I generated a reporter cell line to read out the NS3a/ANR interaction using a Gal4/UAS system[9]. Gal4 is a yeast transcriptional activator that contains a DNA-binding domain (DBD) that recognizes and binds the upstream activating sequence (UAS). In yeast, when full-length Gal4 is bound to the UAS, transcriptional activators are recruited, and transcription of downstream genes is initiated[10]. For use in mammalian systems, Gal4 is truncated to only the DNA binding domain, and recruitment of exogenous transcriptional activators can be used to initiate transcription of genes. By separating these three components, which has been shown to be functional in systems ranging from silkworms to zebrafish, synthetic regulation of endogenous and exogenous genes can be achieved[11,12].

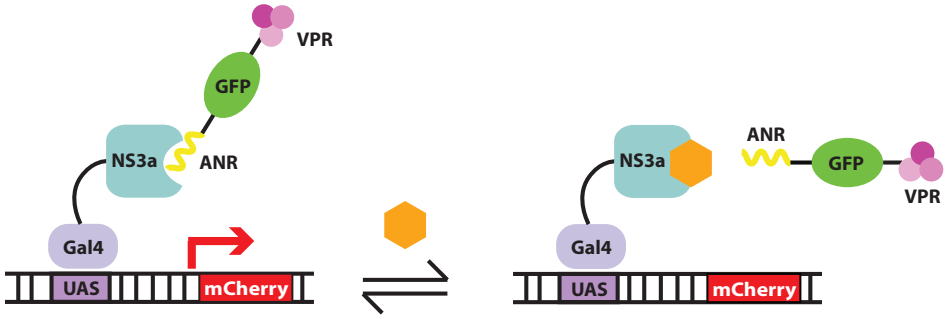

Figure 3.2 ***The fluorescent mammalian reporter system for screening ANR variants.*** Schematic of the fluorescent mammalian reporter system for screening ANR variants. In the absence of drug, ANR-VPR binds to NS3a that is localized to the UAS through Gal4-DBD. The localization of VPR promotes expression of mCherry. Addition of an NS3a inhibitor disrupts the NS3a/ANR interaction and terminates the expression of mCherry.

Our reporter system is composed of three components that we customized to fit our needs (**Figure 3.2**). UAS-minimal CMV (mCMV) was placed upstream of the mCherry gene, allowing for expression of the fluorescent mCherry protein when transcriptional activators are recruited to

this site. The second component of our reporter system is the Gal4-DBD fused to NS3a (Gal4-NS3a). The final component is ANR-VPR, which is ANR linked to VPR, a fusion of three transcriptional activators that have been shown to be functional in mammalian cells. In our reporter system, mCherry expression is low when a transcriptional activator has not been recruited. When ANR-VPR is recruited to GAL4-NS3a, which is bound to the UAS upstream of the mCherry gene, mCherry protein expression is promoted.

As an initial proof of concept, a plasmid containing UAS-mCMV-mCherry, Gal4-NS3a, and ANR-VPR was created and transiently transfected into HEK293 cells (**Figure 3.3**) We observed mCherry expression in the absence of drug, suggesting that NS3a is able to recruit ANR-VPR. The addition of drugs that bind to NS3a and disrupt the ANR/NS3a interaction resulted in a significant decrease in mCherry expression, albeit at higher expression levels than background– cells expressing a VPR fusion (DNCR2-VPR) that is not colocalized with NS3a. These results demonstrate that our three-component system is functional in mammalian cells and that the NS3a/ANR interaction can be measured with the fluorescent mCherry reporter.

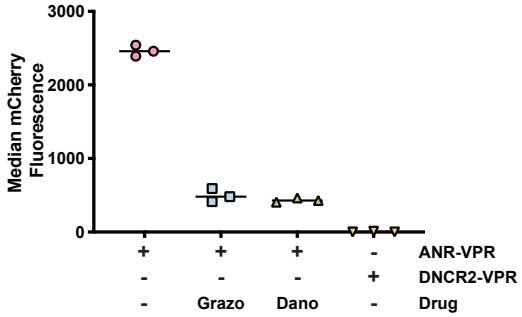

Figure 3.3. ***Characterization of the three-component transcription reporter.*** HEK293 cells were transfected with a construct containing UAS-mCMV-mCherry and Gal4-NS3a, as well as ANR-VPR or DNCR2-VPR. DNCR2 is an engineered protein that only interacts with NS3a that is bound to danoprevir. mCherry expression is elevated in cells expressing ANR-VPR and Gal4-NS3a. Transcription is disrupted in the presence of grazoprevir or danoprevir, which are NS3a inhibitors that disrupt the interaction between ANR and NS3a. In cells expressing DNCR2-VPR and that have not been treated with danoprevir, mCherry expression is low because DNCR2 is not recruited to NS3a.

To obtain an optimized ANR peptide, we decided to design an unbiased library of single mutant ANR variants, but first needed to adapt our reporter system for facile integration of our ANR library. We required that each cell only integrate a single variant, and that expression of the variants should be equal across the population. To achieve this, we used a Bxb1 landing pad cell line developed by the Fowler lab at UW[13]. By co-transfecting a recombinase protein and a mammalian expression vector containing a recombinase binding site and the gene of interest into cells containing a recombination site, a single copy of a gene of interest is incorporated into a defined site into the genome. The Bxb1 landing pad cell lines also contain expression regulators near the recombination site, allowing for inducible expression of the gene that has been integrated into the genome. Although similar systems, like FlpIn cells, are commercially available, the location of the recombination site in the genome is proprietary information, making downstream sequencing of library variants difficult. By using the landing pad cells developed by the Fowler lab, a single copy of each library variant is stably incorporated at the same site in the genome. This should result in similar expression levels of each ANR variant. Because the Bxb1 landing pad cells were engineered by our collaborator, the position of the recombination site is known, resulting in the ability to create sequencing primers over the recombination junction. This provides higher confidence in our sequencing data.

It has been observed that smaller genes are more efficiently incorporated into Bxb1 landing pad cells, so we chose to stably integrate UAS-mCherry and Gal4-DBD-NS3a into cells using lentiviral infection, then use recombination to incorporate ANR-VPR variants. Lentivirus expression constructs were made that contain the UAS sequence upstream of an mCherry gene under the control of a minimal CMV promoter. This lentivirus construct also contains elements for bi-cistronically expressing Gal4-DBD-NS3a fusion protein and iRFP670 gene, both under the

control of a full CMV promoter. Following lentivirus production and subsequent infection of the Bxb1 landing pad cell line, cells that had similar levels of iRFP670 fluorescence, which is under the control of a constitutive CMV promoter, were collected and propagated (**Figure 3.4a, 3.4b, 3.4c**). These cells were then used for the recombination and stable integration of the ANR library.

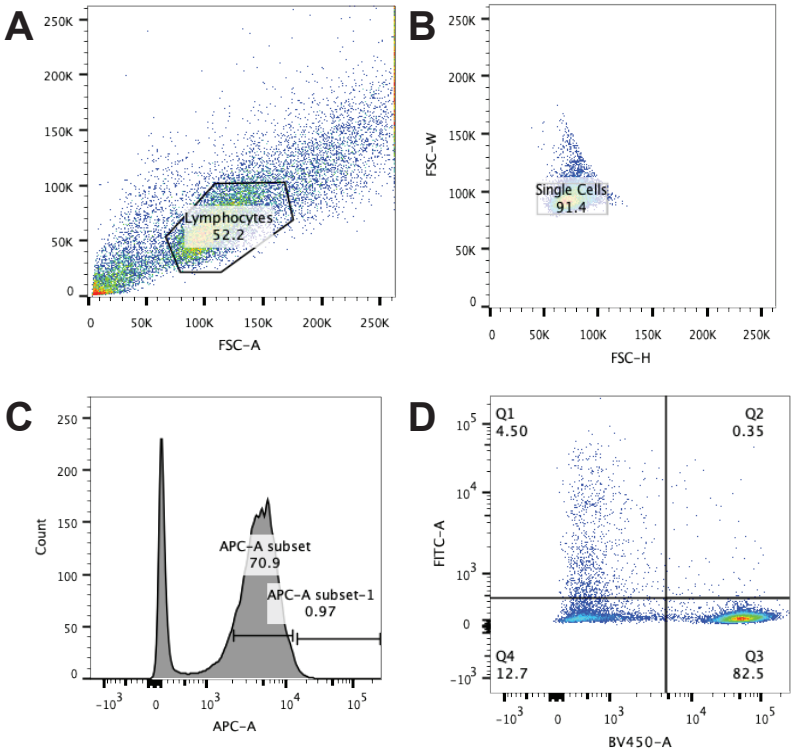

Figure 3.4. ***FACS gating strategy for selecting cells that contain stably integrated reporter components.*** (A) HEK293 landing pad cells were first gated on living cells. (B) Events from single cell populations where then collected. (C) Cells exhibiting similar levels of iRFP670, identified using the APC laser and gated on the peak of the histogram (70.9% of cells shown here), were collected and propagated. (D) Following integration of the ANR library, cells that were BFP-negative and GFP-positive (quadrant 1) were collected and used for future experiments.

3.2.2 *Design, generation and testing of ANR library*

In order to identify a superior binding partner for NS3a, we set out to create an unbiased library of ANR variants. We mutagenized the ANR peptide with a doped oligo strategy[14]. Synthetic oligonucleotides were ordered that contained a custom mix of degenerate nucleotides at every nucleotide position. This mixture contained 3% doping, meaning that at every nucleotide position, 97% of the mixture was composed of the wild type nucleotide, while the remaining 3% consisted of an equal mixture of the other three remaining nucleotides. For the 63-nucleotide sequence encoding ANR, this doping frequency was predicted to give one nucleotide mutation per ANR variant, resulting in mainly single amino acid mutants.

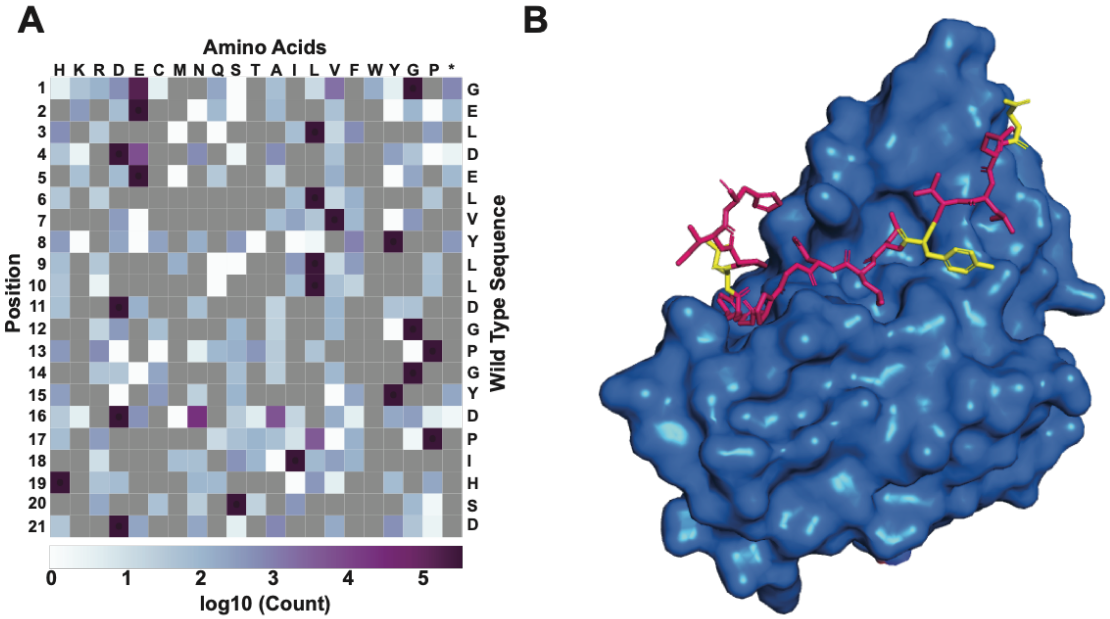

Figure 3.5. ***Deep sequencing results from sorted ANR mutants.*** (A) Heat map showing results from the sequenced ANR variants following the third sort. (B) Structure of wild type ANR (pink) bound to NS3a (blue). Positions of ANR mutations identified from deep sequencing for follow up on are shown in yellow.

Our ANR library was cloned into a plasmid construct containing GFP-tagged VPR, then stably incorporated into the landing pad cell line (**Figure 3.4d**). Upon incorporation of the library, cells that expressed GFP were collected and propagated. Following propagation, these cells were

analyzed, and those that had the highest level of mCherry fluorescence, indicating enhanced ability of the ANR mutant to bind to NS3a, were collected. Five successive sorts of the library were completed, and genomic DNA from the cells was extracted and deep sequenced (**Figure 3.5a**).

Using a previously published structure of WT-ANR bound to NS3a (PDB: 4A1V) we were able to identify the location of mutations in the ANR peptide that were enriched after selection (**Figure 3.5b**). Interestingly, we found that most of selected mutants were in the disordered N- and C-terminal regions of the peptide. We hypothesized that mutations in these amino acids potentially stabilize these regions and create more structured contacts between the ANR peptide and NS3a, which results in the observed higher affinity.

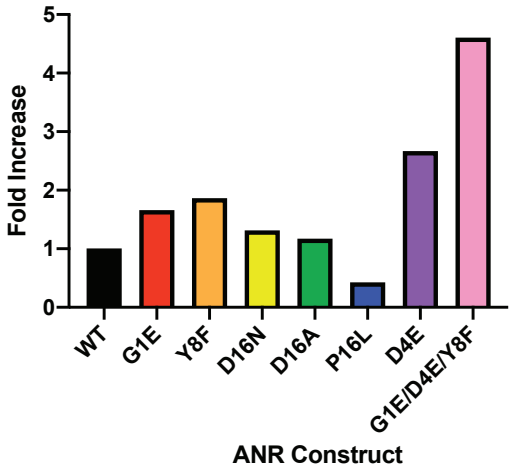

Figure 3.6. ***Functional testing of ANR mutants.*** ANR variants containing single or a combination of enriched mutations. Each individual ANR variant was tested in the fluorescent cell reporter assay. The average fold increase of the mutant over wild-type ANR is shown from duplicate replicates.

ANR variants identified from the library screen were individually tested in our fluorescent cell reporter assay and compared to wild-type ANR (WT-ANR) (**Figure 3.6**). Interestingly, one ANR variant containing a P16L mutation showed decreased fluorescence compared to WT-ANR, even though it had been enriched during the screen. ANR variants containing mutations to the

aspartate at position 16, either to an alanine or an arginine, showed similar binding affinity as that of the WT-ANR to NS3a. We chose to investigate the arginine 16 mutant further in the hopes that it would create increased intermolecular interactions with NS3a when combined with other mutations. From the initial screen, the most promising mutations were a glutamic acid at positions 1 and 4, and a phenylalanine at position 8. We created ANR variants combinations of these mutations, either as double or triple mutants and tested them in our fluorescent cell reporter assay (**Figure 3.6**). We found that an ANR variant containing the G1E, D4E, and Y8F led to the largest increase in mCherry expression, providing a ~5-fold increase over wild-type ANR. This optimized ANR variant will be used as the basis for future libraries of ANR variants.

We optimized our second-generation CDP system by selecting for superior ANR variants. As an alternative approach, we also attempted to optimize NS3a, particularly at positions that have been shown to make close contacts with ANR. NS3a is a viral protease that contains a catalytic serine residue in its active site. Although wild type NS3a has been shown to not cleave the ANR peptide, we have chosen to use a serine to alanine NS3a mutant to absolutely ensure that there is no residual proteolytic activity. In order to potentially increase the interaction surface area between NS3a and ANR, which would be expected to increase the binding affinity of ANR to NS3a, we mutated NS3a's catalytic serine residue. A panel of NS3a mutants were generated and tested against WT-ANR to check if these NS3a variants possessed increased affinity for ANR. We also confirmed that the ANR/NS3a interaction can still be disrupted by NS3a inhibitors, like danoprevir. Unfortunately, from the panel of NS3a mutants that were tested, many seemed to prevent danoprevir from binding, which makes them useless for our system. One mutation that did look promising was replacing NS3a's catalytic serine residue with a threonine. Both serine and threonine contain an alcohol side chain but threonine contains a methyl group that potentially can

make additional hydrophobic interactions with the ANR peptide. Although danoprevir can disrupt the NS3a/ANR interaction, NS3a containing the serine to threonine substitution (NS3a-Thr) demonstrated reduced efficiency in driving mCherry expression, indicating that there is ample room for optimizating its interaction with the ANR peptide. Moving forward, future libraries of ANR variants will be tested against both the original NS3a as well as the NS3a-Thr mutant.

3.3 DISCUSSION

Interestingly, favorable mutations from the first round of our library were only observed in the disordered N- and C-terminal ends of the ANR peptide, rather than the internal portions. Moving forward, we are interested in finding mutations in the internal ANR amino acids (LVYLL) that are able to increase affinity. We wondered if the lack of mutations in the internal region stemmed from that region being "locked down" by contacts made in the terminal regions of the peptide to NS3a. If these regions are stuck in a particular position, that would prevent plasticity of the middle region, resulting in the lack of mutated variants in this region. To get around this, more libraries will be generated using a saturation mutagenesis strategy rather than the previously used doped oligo strategy, allowing for variants with multiple mutations. Each ANR peptide will be encoded by two three-nucleotide stretches that will be composed of NNK degenerate codons, which allow for encoding of all amino acids at that position, while minimizing the number of stop codons, resulting in a library that is expected to have ~400 variants containing double mutations[15]. By creating multiple versions of these double amino acid mutant libraries, we will be able to examine how each mutant works in combination with another in a high-throughput manner, rather than going through and individually testing all the single variant combinations.

Future utilization of this system will be achieved by integrating it with a chemically inducible dimerization system developed by our lab. This system also uses NS3a and clinically

approved NS3a-inhibitors, with the added addition of two computationally designed proteins that specifically recognize NS3a bound to the drugs danoprevir or grazoprevir[43]. By integrating ANR with the two computationally designed proteins and an unbound NS3a, combinatorial control of diverse systems can be attained using the same "receiver" protein, NS3a. This completely bio-orthogonal system has the potential for applications for studying cell signaling networks, as well as integration with other engineered synthetic systems, or even utilization in safer and more effective cell therapies.

3.4 Materials and Methods

3.4.1 *Mammalian Expression constructs*

Three main constructs were used for these experiments (**Table 3.3**). The first contained all components for the transcription-based fluorescence assay and was constructed in a lentiviral vector. The plenti-UAS-miniCMV-mCherry/CMV-Gal4DBD-NS3aH1-P2A-ANR-myc-BFP-VPR (EMD153) was based on a pLenti-UAS-miniCMV-mCherry/CMV-Gal4DBD-ERT2VP16 vector, a gift from Dr. Kenneth Matreyek (derived from addgene plasmid 79130, a gift from Wendell Lim). The vector was digested with SalI and MluI to insert the NS3aH1 from a Maly lab source as well as the P2A-ANR-BFP, which was obtained as a double stranded DNA G-Block (IDT), then assembled using Gibson Assembly. The second construct was also a lentiviral vector but does not contain the ANR-VPR machinery. The pLenti-UAS-miniCMV-mCherry/CMV-Gal4DBD-NS3aH-P2A-iRFP670-T2A-PuromycinResistance construct (EMD157) was derived from plasmid G143-pLenti from Dr. Matreyek and digested with SalI and Tth111i. Inserts consisting of NS3aH1-P2A and iRFP670, both derived from Maly Lab sources, were inserted into the vector and assembled using Gibson Assembly. The plasmid containing the library consisted of 3xNLS-GFP-ANR-VPR-P2A-BlasticidinResistance in the attb vector (EMD175), used for stable

recombination into the TetBxb1 BFP cells (see below). The attb vector, G275B a gift from Dr. Matreyek, was digested with XbaI and SgrAI, then the 3xNLS, GFP, and ANR-VPR were inserted using Gibson Assembly. Plasmids used for lentiviral infection, pCMVdR8.91 and pMD2.G, were a gift from Dr. Jesse Zalatan.

The ANR library was made in the attb-ANR plasmid (EMD175). The WT-ANR plasmid was linearized by pcr on either side of WT-ANR, then the doped oligo library (ordered from IDT) was assembled using Gibson assembly, using a 5:1 ratio of µg insert:vector. Following the Gibson reaction, 1 µL of the assembled plasmids were transformed into 50 µL *E. coli* on ice for 30 minutes, then the cells were heat shocked for 30 seconds. The cells were then added to a 50 mL culture of LB broth with ampicillin and allowed to incubate at 37 ºC with shaking overnight. Prior to overnight incubation, 50 µL of cells were plated on LB agar plates with ampicillin to ensure proper transformation of the library. The next morning, colonies were counted and found to be sufficient for robust library coverage (aiming for 10x the number of expected variants), then plasmid DNA from the liquid culture was obtained using a Qiagen Midiprep kit. This midiprep was used for downstream transfection of the ANR library.

Table 3.2. Amino acid sequences of constructs and primers used for gDNA amplification and sequencing

Construct/Primer	Sequence (Amino acid for constructs, nucleotide for primers)
plenti-UAS-miniCMV-mCherry/CMV-Gal4DBD-NS3aH1-P2A-ANR-myc-BFP-VPR (EMD153)	MKLLSSIEQACDICRLKKLKCSKEKPKCAKCLKNNW ECRYSPKTKRSPLTRAHLTEVESRLERLEQLFLLIFPR EDLDMILKMDSLQDIKALLGTPAAASTAGSGGMAK GSVVIVGRINLSGDTAYSQQTRGLEGCQETSQTGRD KNQVEGEVQVVSTATQSFLATSINGVLWTVYHGAG TRTIASPKGPVTQMYTNVDKDLVGWQAPQGSRSLT PCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPI SYLKGSAGGPLLCPAGHAVGIFRAAVSTRGVAKAV DFIPVESLETTMRSPGSGATNFSLLKQAGDVEENPG PGALSGMGELDELVYLLDGPGYDPIHSDGVLSGSG TGSGTGSGTGTTSGTGTGGSTGEQKLISEEDLGSGS SELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPY EGTQTMRIKVVEGGPLPFAFDILATSFLYGSKTFINH

	TQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDT SLQDGCLIYNVKIRGVNFTSNGPVMQKKTLGWEAF TETLYPADGGLEGRNDMALKLVGGSHLIANIKTTY RSKKPAKNLKMPGVYYVDYRLERIKEANNETYVEQ HEVAVARYCDLPSKLGHKLNGSGSDALDDFDLDML GSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFD LDMLGSPKKKRKVGSQYLPDTDDRHRIEEKRKRTYE TFKSIMKKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAP QPYPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQ VLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVA PPAPKPTQAGEGTLSEALLQLQFDDEDLGALLGNSTDP AVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPE AITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSI ADMDFSALLSQISSGSGSGSRDSREGMFLPKPEAGSAIS DVFEGREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTG PVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDPDEET SQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHL DELTTTLESMTEDLNLDSPLTPELNEILDTFLNDECLLH AMHISTGLSIFDTSLF*
pLenti-UAS-miniCMV-mCherry/CMV-Gal4DBD-NS3aH-P2A-iRFP670-T2A-PuromycinResistance (EMD157)	MILKMDSLQDIKALLGTPAAASTAGSGGMAKGSVVIV GRINLSGDTAYSQQTRGLEGCQETSQTGRDKNQVEGE VQVVSTATQSFLATSINGVLWTVYHGAGTRTIASPKGP VTQMYTNVDKDLVGWQAPQGSRSLTPCTCGSSDLYLV TRHADVIPVRRRGDSRGSLLSPRPISYLKGSAGGPLLCP AGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRSPGS GATNFSLLKQAGDVEENPGPMARKVDLTSCDREPIHIP GSIQPCGCLLACDAQAVRITRITENAGAFFGRETPRVGE LLADYFGETEAHALRNALAQSSDPKRPALIFGWRDGLT GRTFDISLHRHDGTSIIEFEPAAAEQADNPLRLTRQIIART KELKSLEEMAARVPRYLQAMLGYHRVMLYRFADDGSG MVIGEAKRSDLESFLGQHFPASLVPQQARLLYLKNAIRV VSDSRGISSRIVPEHDASGAALDLSFAHLRSISPCHLEFLR NMGVSASMSLSIIIDGTLWGLIICHHYEPRAVPMAQRVA AEMFADFLSLHFTAAHHQRGSGEGRGSLLTCGDVEENP GPTEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHTV DPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAV WTTPESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLL APHRPKEPAWFLATVGVSPDHQGKGLGSAVVLPGVEAA ERAGVPAFLETSAPRNLPFYERLGFTVTADVEVPEGPRTW CMTRK*
attb-3xNLS-GFP-ANR-VPR-P2A-BlasticidinResistance (EMD175)	MDPKKKRKVDPKKKRKVDPKKKRKVGSGVSKGEELFTG VVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT GKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMP EGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID FKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSAL

	SKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSGTGSG TGSGTGTTSGTGTGGSTGGELDELVYLLDGPGYDPIHSDG TTSGTGTGGSTGEFSSAAGTSDALDDFDLDMLGSDALDDF DLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSPKKKR KVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPTD PRPPPRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTM VFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAP APVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDD EDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPH TTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLL SGDEDFSSIADMDFSALLSQISSGSGSGSRDSREGMFLPKPE AGSAISDVFEGREVCQPKRIRPFHPPGSPWANRPLPASLAPT PTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDPDEET SQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDEL TTTLESMTEDLNLDSPLTPELNEILDTFLNDECLLHAMHISTG LSIFDTSLFCSGATNFSLLKQAGDVEENPGPAKPLSQEESTLIE RATATINSIPISEDYSVASAALSSDGRIFTGVNVYHFTGGPCAE LVVLGTAAAAAAGNLTCIVAIGNENRGILSPCGRCRQVLLDLHPGI KAIVKDSDGQPTAVGIRELLPSGYVWEG*
Kam_499	gagaacgtatgtcgaggtaggc
ANRlib_gDNA_revamp1_1	gggttagcaagtggcagccttccaatgcatctgagcctaa
ANRlib_gDNA_fwdamp2_1	aatgatacggcgaccaccgagatctacacgtcatgcttacaagggaagcggtactgg
PTEN_seq_R_generic_a	caagcagaagacggcatacgagatnnnnnnnnggttagcaagtggcagcct
ED_175_read1	cacgacgagcggtacaggaacaggtggtagtacaggt
ED_175_read2	gagcctcccgtcccagtaccactcgttgttcc
EMD_ANRlib_index1	agatgcattggaaggctgccacttgctaaccc
ED_175_index2	cctgtgccgctaccagtaccgcttcccttgta

3.4.2 *Cell culture*

HEK293T cells (ATCC) were used for viral production, and HEK293T TetBxb1 BFP, a gift from Dr. Doug Fowler, were used for library integration and propagation[103]. Both cell lines were maintained in DMEM supplemented with 10% FBS, with the addition of 2 µg/mL doxycycline in the HEK293T TetBxb1 cells. After integration of the EMD157 and EMD175, cells were also maintained in 3 mg/mL puromycin and 5 mg/mL blasticidin.

For lentiviral production, HEK293T cells were plated at a density of $1.5x10^6$ cells/mL in 3 mL in a 6-well plate. The following day, each well was transfected with 3 μg DNA composed of 1.5 μg transfer vector, 1.33 μg pCMVdR8.91, and 0.17 μg pMD2.G using turbofectin (Origene) following manufacturer's instructions. 24 hours post transfection, 2 mL complete media was added to each well, and the target cells (HEK293T TetBxb1 BFP cells) were seeded at a density of $5x10^4$ cells/mL in 0.5 mL media in a 24 well plate. The next day, 4.5 mL of the supernatant from the virally infected cells was collected and centrifuged for 5 min and 400g at room temperature. Without disturbing the viral debris, 1 mL of the supernatant was slowly added to each well of the target cells in the 24-well plate, then the plates were returned to the incubator. The following day, media was removed and replaced with fresh complete media, supplemented with doxycycline and puromycin. After allowing the cells to undergo drug selection, surviving cells were sorted for iRFP670-positive events (constitutively expressed in the EMD157 construct).

HEK293T TetBxb1 BFP cells that had stably incorporated the EMD157 construct were then transfected with the ANR library. Following splitting, $5x10^5$ cells were plated into each well of a six well plate in a total of 2 mL of DMEM supplemented with 10% FBS (no dox). A transfection mixture containing 1.5 μg pCAG-NLS-Bxb1 recombinase plasmid and 1.5 μg ANR library plasmid with 9 μL turbofectin was made following manufacturer's instructions, then added dropwise to the cells, which were then returned to the incubator. The next day, media was changed, cells were induced with doxycycline, and 3 mg/mL puromycin and 10 mg/mL blasticidin was added to the mixture. Following the drug selection (~1 week), cells were sorted, and GFP-positive/BFP-negative events were collected and propagated.

3.4.3 *Library analysis and sorting*

HEK293T TetBxb1 cells that had successfully incorporated EMD157 and the ANR library were expanded into T75 flasks, then treated with DMSO or 2.5 nM danoprevir, a concentration previously shown to decrease the fluorescence of the wild type system by 50%. Cells were gated on successful incorporation of the library (GFP-positive events), and mCherry fluorescence was analyzed on a FACSAria III (BD Biosciences). The top 3% of mCherry-positive cells were collected, and over 50,000 single-cell events were collected per condition. These cells were then propagated, and the process was repeated for five successive sorts. Following each sort, a portion of the population was collected for genomic DNA extraction and next-generation sequencing.

3.4.4 *Library genomic DNA isolation and sequencing*

Genomic DNA was extracted from collected cells using a DNeasy Kit (Qiagen). 50 µL first round PCR reactions were each prepared with a final concentration of 1 µg input genomic DNA, 1x Q5 Master Mix (New England Biolabs), and 0.3 µM each of the KAM499 and ANRlib_gDNA_revamp1_1 forward and reverse primers. The reactions were run on a thermocycler for 30 seconds at 95 ºC, then five cycles of 98 ºC for 10 seconds, 63 ºC for 20 seconds, and 72 ºC for 45 seconds. Following the cycles, the reaction was held at 72 ºC for 3 minutes, then held at 4 ºC. The resulting product was bound to AMPure XP (Beckman Coulter), cleaned, and eluted with 21µL water. The cleaned DNA product was then used in the second-round indexing PCR, which used Q5 Master Mix with Sybr Green (ThermoFisher) and 0.25 µM ANRlib_gDNA_fwdamp2_1 forward primer, and PTEN_seq_R_generic reverse primers to index the reactions. The reaction conditions were 95 ºC for 3 minutes, then 20 cycles of 95 ºC for 15 seconds, 63 ºC for 15 seconds, and 72 ºC for 30 seconds, followed by 72 ºC for 1 minute and a 4 ºC hold. Reactions were cleaned with a Zymo Clean and Concentrator Kit (Zymo Research) and

amplicons were extracted after separation on a 1% TAE/agarose gel using a Zymo Gel DNA Extraction Kit (Zymo Research). Extracted amplicons were quantified using a Qubit dsDNA HS Assay Kit (Life Technologies) and sequenced on a NextSeq using a NExtSeq 500/550 High Output v2 75 cycle kit (Illumina) using primers ED_175_read1, ED_175_read2, EMD_ANRlib_index1, ED_175_index2.

Paired sequencing reads were de-multiplexed with bcl2fastq and joined using the fastq-join tool within the eautils package. Enrich2 was used to identify sequences of collected library members with the help of Jason Stephany. For sequences of primers used in this section, please see Table 4.3

3.5 REFERENCES

1. Barkovich, K. J., Moore, M. K., Hu, Q., Shokat, K. M. (2018). Chemical genetic inhibition of DEAD-box proteins using covalent complementarity. *Nucleic Acids Research*, *46*(17), 8689–8699.

2. Ocasio, C. A., Warkentin, A. A., McIntyre, P. J., Barkovich, K. J., Vesely, C., Spencer, J., Shokat, K. M., Bayliss, R. (2018). Type II Kinase Inhibitors Targeting Cys-Gatekeeper Kinases Display Orthogonality with Wild Type and Ala/Gly-Gatekeeper Kinases. *ACS Chemical Biology*, *13*(10), 2956–2965.

3. Wong, A. W., Urisman, A., Burlingame, A. L., Shokat, K. M. (2019). Chemically reprogramming the phospho-transfer reaction to crosslink protein kinases to their substrates. *The Protein Society*, *28*(3), 654–662.

4. Cunningham-Bryant, D., Dieter, E. M., Foight, G. W., Rose, J. C., Loutey, D. E., Maly, D. J. (2019). A Chemically Disrupted Proximity System for Controlling Dynamic Cellular Processes. *Journal of the American Chemical Society*, *141*(8), 3352–3355.

5. Foight, G. W., Wang, Z., Wei, C. T., Greisen, P. J., Warner, K. M., Cunningham-Bryant, D., et al. (2019). Multi-input chemical control of protein dimerization for programming graded cellular responses. *Nature Biotechnology*.

6. Dieter, E. M., Maly, D. J., (2020). A Chemically-controlled system for activating RAS GTPases, *Methods in Enzymology, 633,* 103-117. Copyright Elsevier (2020).

7. Kügler, J., Schmelz, S., Gentzsch, J., Haid, S., Pollmann, E., van den Heuvel, J., Franke, R., Pietschmann, T., Heinz, D. W., Collins, J. (2012). High affinity peptide inhibitors of the hepatitis C virus NS3-4A protease refractory to common resistant mutants. *The Journal of Biological Chemistry*, *287*(46), 39224–39232.

8. McCauley, J. A.; Rudd, M. T. Hepatitis C virus NS3/4a protease inhibitors. Curr. Opin. Pharmacol. 2016, 30, 84−92.

9. Traven, A., Jelicic, B., Sopta, M. (2006). Yeast Gal4: a transcriptional paradigm revisited. *EMBO Reports*, *7*(5), 496–499.

10. Brand, A. H., Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, *118*, 401–415.

11. Imamura, M., Nakai, J., Inoue, S., Quan, G. X., Kanda, T., Tamura, T. (2003). Targeted Gene Expression Using the GAL4/UAS System in the Silkworm Bombyx mori. *Genetics Society of America*, 1329–1340.

12. Asakawa, K., & Kawakami, K. (2008). Targeted gene expression by the Gal4-UAS system in zebrafish. *Development, Growth & Differentiation*, *50*(6), 391–399.

13. Matreyek, K. A., Stephany, J. J., & Fowler, D. M. (2017). A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Research*, *45*(11), e102–e102.

14. Neylon, C. (2004). Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Research*, *32*(4), 1448–1459.

15. Nov, Y. (2012). When second best is good enough: another probabilistic look at saturation mutagenesis. *Applied and Environmental Microbiology*, *78*(1), 258–262.

VITA

Emily Dieter obtained her Bachelor of Science degree in Biochemistry with minors in Mathematics and Physics at Saint Michael's College in the spring of 2014. As an undergraduate, Emily participated in two summer research programs. During the summer of 2012, Emily synthesized an isotopically labeled biological tagging reagent to be used in LC-MS studies under the direction of Dr. Shane Lamos at Saint Michael's. The following summer, Emily worked under Dr. David Braun at the University of Missouri to fine-map carbohydrate partitioning mutants in maize using molecular genetics.

In the fall of 2014, Emily began her graduate research at the University of Washington with Professor Dustin Maly. While at the University of Washington, Emily was supported by the National Science Foundation Graduate Research Fellowship Program. Following the completion of her doctorate, Emily will continue research as a Postdoctoral Researcher under the guidance of Dr. Karen Mackenzie at the Children's Medical Research Institute in Sydney, Australia.