Small-Molecule Gated Artificial Regulatory Domains: A Novel Tool For Dissecting Signaling Pathways

Daniel Thomas Cunningham-Bryant

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Dustin J. Maly, Chair

Michael Gelb

Champak Chatterjee

Program Authorized to Offer Degree:

Department of Chemistry

University of Washington

Abstract

Small Molecule Gated Artificial Regulatory Domains: A Novel Tool For Dissecting Signaling Pathways

Daniel Thomas Cunningham-Bryant

Chair of the Supervisory Committee:

Professor Dustin J Maly
Department of Chemistry

The activities of the proteins involved in eukaryotic signaling networks are controlled by a complex system of allosteric regulatory domains, in which auto-inhibitory interacting domains repress protein activity. The field of synthetic biology aims to harness such biological systems to study the roles of individual signaling nodes, and more generally, to obtain spatio-temporal control over protein activity. To this end, we have developed a novel, modular, chemical genetic method for conferring small molecule control over the activity of signaling proteins. This system replaces their auto-inhibitory regulatory domains with an artificial protein-protein interaction engineered from the Hepatitis C Virus protease (NS3-4A). This artificial regulatory mechanism can be disrupted with a bio-orthogonal, cell-permeable, and clinically approved small molecule. This system confers spatio-temporal control over protein activity in a dose-dependent and reversible manner.

Initial engineering efforts were conducted by conferring our switch in two distinct approaches. Initial characterization entailed application of this system, in an inter-molecular fashion, to the regulation of AKT pathway activation in a localization dependent manner. A second mode of application was demonstrated using this switch in an intra-molecular system. The artificial regulator domains were applied to the guanine nucleotide exchange factor Son of Sevenless, an activator of the GTPase Ras. A computational method using Rosetta Remodel directed the replacement of the endogenous regulatory modules with the HCV-Protease based switch, resulting in the successful generation of a bio-orthogonal chemically inducible activator of Ras.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 1

Chapter 2

List of Abbreviations

A3	A-385358
Aa	Amino Acid
ABT	Abbot Laboratories
Akt	Protein Kinase B
Ala	Alanine
Asp	Aspartic Acid
Asu	Asunaprevir
ATP	Adenosine triphosphate
BAD	Bcl-2 Associated Death Promoter
Bcl-2	B-cell Lymphoma 2
Bcl-xL	B-cell Lymphoma-extra large
BFP	Blue Fluorescent Protein
BH3	Bcl-2 Homology 3
BirA	Biotin Ligase
Cas9	CRISPR associated protein 8
CASPase-3	Cysteine-dependent asparte-directed protease 3
Cdc25	Cell-division cycle 25 phosphatase
CDD	Chemical Disruptor of Dimerization
CIAR	Chemically Inducible Activator of Ras
CID	Chemical Inducer of Dimerization
COM	Center of Mass
Cre	Cre-Recombinase
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
Dox	Doxycycline
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
EGFP	Enhanced Green Fluorescent Protein
EGFR	Epidermal Growth Factor Receptor
ER	Endoplasmic Reticulum
ERK	Extracellular Signal Related Kinase
FAM/FITC	Fluoroscein
FBS	Fetal Bovine Serum
FRET	Förster Resonance Energy Transfer
FKBP	FK506 Binding Protein
FRB	FKBP12-Rapamycin Binding Domain
GAP	GTPase Activating Protein
GDP	Guanosine diphosphate
GEF	Guanine Nucleotide Exchange Factor
GFP	Green Fluorescent Protein
Glu	Glutamic Acid
Gly	Glycine
GPCR	G-Protein Coupled Receptor
GST	Glutathione S-Transferase
GTP	Guanosine triphosphate

HCV	Hepatitis C-Virus
HEK293T	Human Embryonic Kidney 293 cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	Hexahistidine
HPLC	High Performance Liquid Crystallography
HSQC	Heteronuclear single quantum coherence spectroscopy
IC_{50}	Half-maximal inhibitory concentration
IPTG	Isopropyl β-D-1-thiogalactopyranoside
iSH2	Inter SH2
JNK2	C-Jun N-Terminal Kinase 2
K_D	Dissociation Constant
LB	Luria Bertani Broth
LOV2	Light-oxygen-voltage-sensing domain
MAPK	Mitogen Activation Protein Kinase
MEK	Dual specificity mitogen-activation protein kinase kinase
Mod. RIPA	Modified Radioimmunoprecipitation buffer
NC	Negative Control
NLS	Nuclear Localization Signal
NMR	Nuclear Magnetic Resonance
NS2	Nonstructural protein 2
NS3	Nonstructural protein 3
NS4A	Nonstructural protein 4A
O-CIAR	Bio-Orthogonal Chemically Inducible Activator of Ras
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PBSA	PBS Supplemented Fetal Bovine Serum
PCR	Polymerase Chain Reaction
PDK1	Pyruvate Dehydrogenase Kinase 1
pERK	Phosphorylated Extracellular Signal Related Kinase
PH	Pleckstrin Homology Domain
PIP2	Phosphatitylinositol 4,5-bisphosphate
PMSF	Phenyylmethylsulfonyl fluoride
Ras	Rat Sarcoma GTPase
REM	Ras Exchanger Motif
RhoA	Ras homolog gene gamily member A
RNAi	RNA Interference
RTK	Receptor Tyrosine Kinase
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfide
Ser	Serine
SH2	Src Homology 2
SH3	Src Homology 3
SOS	Son of Sevenless
SOScat	Catalytic Domain of Son of Sevenless
TBS	Tris Buffered Saline
Thr	Threonine
Zdk	Zdark

Units

°C	Degrees Celsius
g	grams
kDa	KiloDalton
L	Liter
m	milli; meter;
M	Molar
Min	minutes
mol	moles
n	nano
s	seconds
µ	micro

Acknowledgements

The work presented here would not have been possible without the constant support of my family, friends, advisors, and teachers. First, I want to thank my parents for their unwavering support and belief in me. The near limitless faith they've shown me has been the driving force in all my personal and professional accomplishments. Your tireless work ethic has been an inspiration and provided me with the ability to preserve no matter how challenging the obstacle. I would also like to thank my brother and sister for all of their love and laughter.

I would also like to thank all of my co-workers and friends. To the entire Maly Lab, you made working a blast even if the toughest of times. To all of my friends, you provided a constant source of support, camaraderie, and friendship throughout the years.

This work could not have been done without the guidance of my advisor, Dustin Maly. Thank you for all of your hard word and guidance. To the rest of my committee, thank you for all of the input and time.

Hannah, this work simply would not have been possible without you. Your love, compassion, and support has allowed me to achieve my aspirations. Your love of exploration, your work ethic, your positivity, and your creativity have all been an inspiration to me and for that, I will be forever grateful. I truly could not have done this without you.

I would also like to say thank you the people in the past who have supported me. I would like to thank Mark Reddington and Hans Johansson, their mentorship, guidance, and patience. Thank you for you help. I would also like to thank my undergraduate advisor Bakthan Singaram, Jinsoo Kim and the rest of the lab at UCSC, their help throughout my time there was greatly appreciated.

Lastly, I would like to give a special thank you to my high school physics teacher Del France. His words of encouragement in the end of my senior year gave me the confidence necessary to pursue a career in the sciences, for that I am grateful, thank you.

Chapter 1: Design and Development of a Bio-Orthogonal Chemical Disruptor of Dimerization

I. Introduction

The ability of cells to transduce extracellular and intracellular signals into the vast array of phenotypic outputs necessary to sustain themselves is staggering. Fundamentally, changes in a eukaryotic cell's environment trigger a cascade of intracellular signaling networks comprising signaling proteins, enzymes, scaffolds, secondary messengers, and an array of other molecules. Every component has activity that is tightly regulated. The activation and regulation of such networks is responsible for the downstream phenotypic output. The precise balance between each signaling pathway is critical to the maintenance of cellular homeostasis as well as for the diverse array of complex behaviors observed, from cellular differentiation to replication, motility, and apoptosis[1,2]. In particular, the spatial and temporal control over signaling proteins play a critical role in dictating the specific outcome[3,4].

Classically, signaling pathways have been studied in a relatively linear manner: relying on dissecting individual members of signaling networks and deciphering their order and role in a linear progression. These methods have demonstrated that pathways utilize an array of scaffolds, GTPases, kinases, and phosphatases to regulate the degree of flux through the pathway[4]. Such methodologies have yielded tremendous insights into the diversity of signaling proteins and their targets, but so far have proved insufficient in elucidating the

complexity of feedback loops as well as the interplay between individual pathways[3,5].

Methods currently available to probe signaling pathways can be broadly separated into two categories: *pre-translational* methods target the protein encoding gene, and *post-translational* methods that target the protein's activity after it has been expressed. Methods that target the gene directly, such as RNAi knockdown or complete gene knockout, offer exquisite target specificity due to targeting the gene directly. By removing the protein entirely, the resulting loss of function phenotype can be observed. However, due to RNAi taking a matter of days for significant knockdown of the protein of interest, these methods lack any significant degree of temporal resolution.

Targeting a protein directly via small-molecule inhibitors allows for rapid and reversible inhibition as well as dose-dependent control over the protein's activity. While small-molecule inhibitors allow for temporal control, significant drawbacks and shortcomings exist. Specifically, such inhibitors lack any spatial resolution by targeting all copies of the protein within a cell. An additional hurdle for small-molecule inhibitors is that they are difficult to develop and often have a significant number of off-target effects.

The constraints of modern biological techniques have led scientists to develop novel chemical genetic techniques that merge protein engineering with pharmacology to allow for the conditional regulation of protein activity[6,7]. An ideal system would combine the best attributes of both systems: the rapid, reversible, and dose-dependent control from small-molecules while maintaining the target

specificity and spatial resolution of genetically encoded methods. Such systems have been developed that utilize small molecule gated inter-molecular switches. These switches use genetically encoded small-molecule inducible protein dimerization domains to control a protein's activity in a drug-regulated manner (Figure 1).

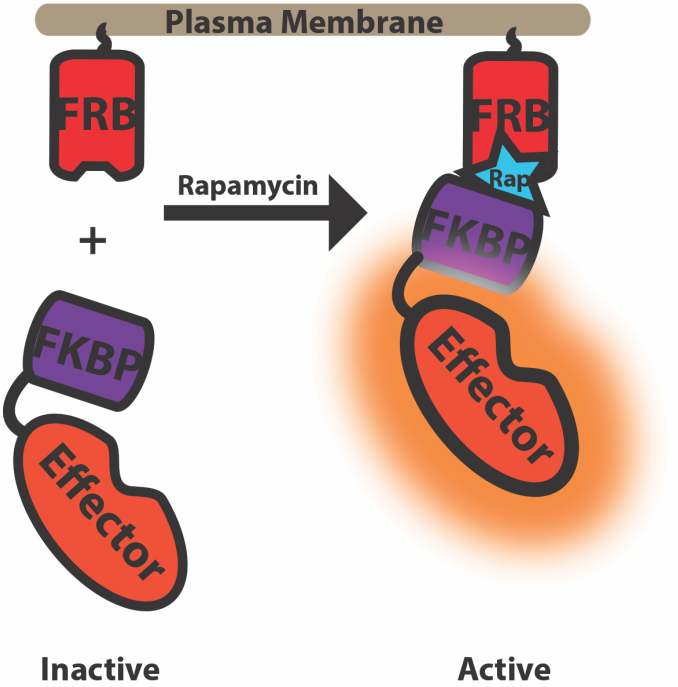


Figure 1: **Schematic representation of a CID system based on Rapamycin**. Chemically inducible dimerizers utilize a set of protein tags. A classic example is the FKBP/FRB Rapamycin system. A protein of interest, whose activity is localization dependent, is tagged with one member and upon addition of Rapamycin, the protein of interest is recruited to the localized binding partner.

Chemical Inducers of Dimerization (CIDs) utilize two genetically encoded protein domains that can be dimerized by the addition of a small molecule. The first such system was developed over 20 years ago[9], taking advantage of a class of natural products termed immunophilins. These natural products, consisting of the small molecules Rapamycin, FK506, and Cyclosporin, bind in a manner that

dramatically increases the drug/protein complex's binding affinity for a second protein. For instance, the mammalian protein FK506 Binding Protein 12 (FKBP12) exhibits no discernable binding to the FKBP12-Rapamycin Binding Domain (FRB)[10]. In the absence of FKBP, FRB has a K_D of 26 μM for Rapamycin. However, FKBP binds Rapamycin with a K_D of 0.2 nM and the FKB:Rapamycin complex has an affinity for FRB of 12 nM, a ~2000-fold increase in affinity[10].

The complete lack of affinity between the two protein domains in the absence of drug coupled with the potent and rapid formation of the FKBP:Rapamycin:FRB complex led to a series of elegant experiments demonstrating the utility of such a switch. Original targets included associating receptor molecules, induction of the death of specific cell types in living animals through activation of the programmed cell death pathway[11], recruitment of proteins to subcellular compartments[12], and activating transcription by inducing a DNA/Binding-Protein domain with activation domain[13], clearly demonstrating the power of the CID system.

Since the initial discovery[9], characterization[10], and application of the FKBP/FRB system, numerous other CID switches have been developed[8] to deal with the issue of bio-orthogonality. Initial efforts focused on chemically modifying Rapamycin and mutating FKBP/FRB accordingly in a "bump/hole" approach[14]. While successful, truly orthogonal switches not only provide for a truly inert CID system but also allow for the use of multiple CID systems in parallel.

II. NS3/4A Protease: A Bio-Orthogonal Chemical Disruptor of Dimerization

The effectiveness and broad utility of the CID system illustrates the potential functionality of small molecule regulated protein tags. The obvious companion to the CID toolkit would be a Chemical *Disruptor* of Dimerization (CDD). A CDD switch would comprise a protein-protein interaction that forms a dimer upon expression, yet can be readily and reversibly disrupted by the addition of a small-molecule While the CID system can translocate an effector of interest to a specified location, a CDD system would pre-localize the desired effector and allow for rapid dissociation away from the site of activity (Figure 2).

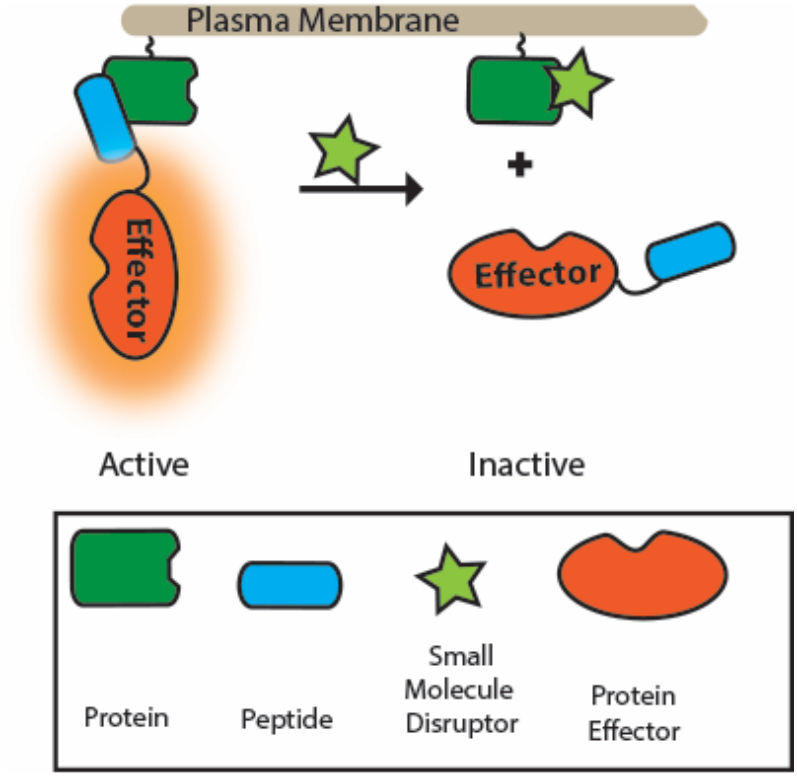


Figure 2: Schematic representation of a CDD switch. A chemical disruptor of dimerization (CDD) would require two protein domains with a high affinity for each other. Formation of this dimer would pre-localize an effector protein tagged with one half of the binding system. Addition of a small molecule that disrupts this protein-protein interaction would result in translocation of the effector protein away from the localization site.

While no CDD systems have yet been developed, recently an opto-genetic disruptor of dimerization[6] was described. Wang et al, engineered a variant of Protein A's ZDK domain that binds to the dark state of LOV2. LOV2 is a photosensing domain isolated from the phototropin 1 protein of *Avena sativa*. Irradiation of cells expressing these two domains by blue light induces a conformation change in LOV2 that induces dissociation from the engineered ZDK domain. ZDK was fused to the GTPases RhoA, Rac1, and the Guanine Nucleotide Exchange Factor (GEF) VAV2 and used to sequester each protein at the mitochondria. In the dark, the sequestration of these protein domains away from the cytosol and plasma membrane masks their activity. Irradiation with blue light induces the release of the ZDK-GTPase or ZDK-GEF fusion protein, and phenotypic changes such as changes in cell edge velocity, ruffling, and protrusion distributions were measured. While this opto-genetic system clearly has multiple uses, they are limited due to the limitations of light irradiation instruments. These instruments are not readily applicable to large screens, are not available to most labs, and are not capable of being used in animal models.

The need for a *chemically* induced disruptor of dimerization led us to consider what elements are necessary for a successfully switch. An ideal CDD would comprise two non-mammalian protein domains with high affinity for each other that exhibit no known activity in mammalian cells. Additionally, this protein-protein interaction must be readily disrupted by a cell-permeable small molecule that has no off-targets in mammalian cells and, ideally, is clinically approved. These criteria necessitated the use of non-mammalian proteins with clinically

approved competitive inhibitors. Such criteria provide a significant hurdle, as small molecules that competitively disrupt a protein-protein interaction are rare.

Peptide based inhibitors are rarely used to target intra-cellular proteins. Poor cell permeability typically limits such inhibitors to targeting protein domains exposed extra-cellularly[15]. However, peptide-based inhibitors offer a unique opportunity for designing a binding partner for a CDD switch. A known, non-mammalian protein that has a clinically approved small-molecule inhibitor presents a target for selection based evolution of a peptide which binds the protein, yet can be displaced by the drug. Fortunately, such a peptide was developed that targets a protease produced by the Hepatitis C Virus[15].

The Hepatitis C Virus (HCV) virion encodes a polyprotein of approximately 3000 amino acids, varying based on the isoform[16]. Upon infection, the polyprotein is both co- and post-translationally processed by viral proteases as well as endogenous proteases. This processing produces the mature viral proteins necessary for completing the infection cycle.

The HCV virion contains a series of structural proteins that are processed by two non-structural proteases, NS2-3 and NS3-4A. One of the roles of the NS2-3 protease is to liberate NS3-4A from NS2-3 and the NS3-4A protease complex cleaves all remaining junctions[16,17]. The NS3 protease domain is a chymotrypsin-like serine protease comprising the first 180 amino acids of the NS3 protein. This protease domain forms a heterodimeric complex with the non-structural protein NS4A, critical for proper folding of the NS3 active site[18]. While

NS4A is a 54 residue long peptide, this can be reduced to a 14 residue segment (residues 21-34) that still forms the NS3-4A complex[19].

While NS3 and NS4A form a heterodimeric complex during viral infection, biochemical characterization and drug development assays necessitated the development of a simplified, single chain protein. This protein, NS3-4A, comprises the minimal NS3 protease domain fused with the minimal NS4A peptide to the N-terminus of NS3[20]. This construct has been used for biochemical and structural studies for drug development as well as a variety of chemical genetic tools such as degrons[21].

Numerous drugs are currently available which target the NS3 protease. All of these drugs bind to the protease active site in peptide-mimetic manner. However, most of these drugs only span the S4-S1' portion of the active site. This relatively narrow window results in a low bar for drug resistance mutations, as evidenced by the numerous relevant mutations seen in the clinic[22]. To expand the target space with which new small-molecule drugs could potentially be developed, Kugler et al developed a novel peptide inhibitor scaffold, called Cp5-46-A (Cp5), that bind the protease in a manner that prevents its cleavage by the protease. Additionally, these peptides utilize a novel binding motif that has yet to be targeted in drug development, a tyrosine finger (Figure 3A-C).

Cp5 variants were developed by phage display, resulting in several potent iterations of varying affinity (Table 1). An ideal CDD switch would allow for tunable affinities. This would allow for an additional degree of control, allowing signal modulation of the desired protein of interest based on the need of the

individual experiment. These Cp5 variants span a range of affinities covering ~100x range, allowing for a wide range of tunability. The NS3-4A protein, Cp5 peptide, and small molecule HCV protease inhibitors provided a reasonable starting point for a bio-orthogonal CDD switch.

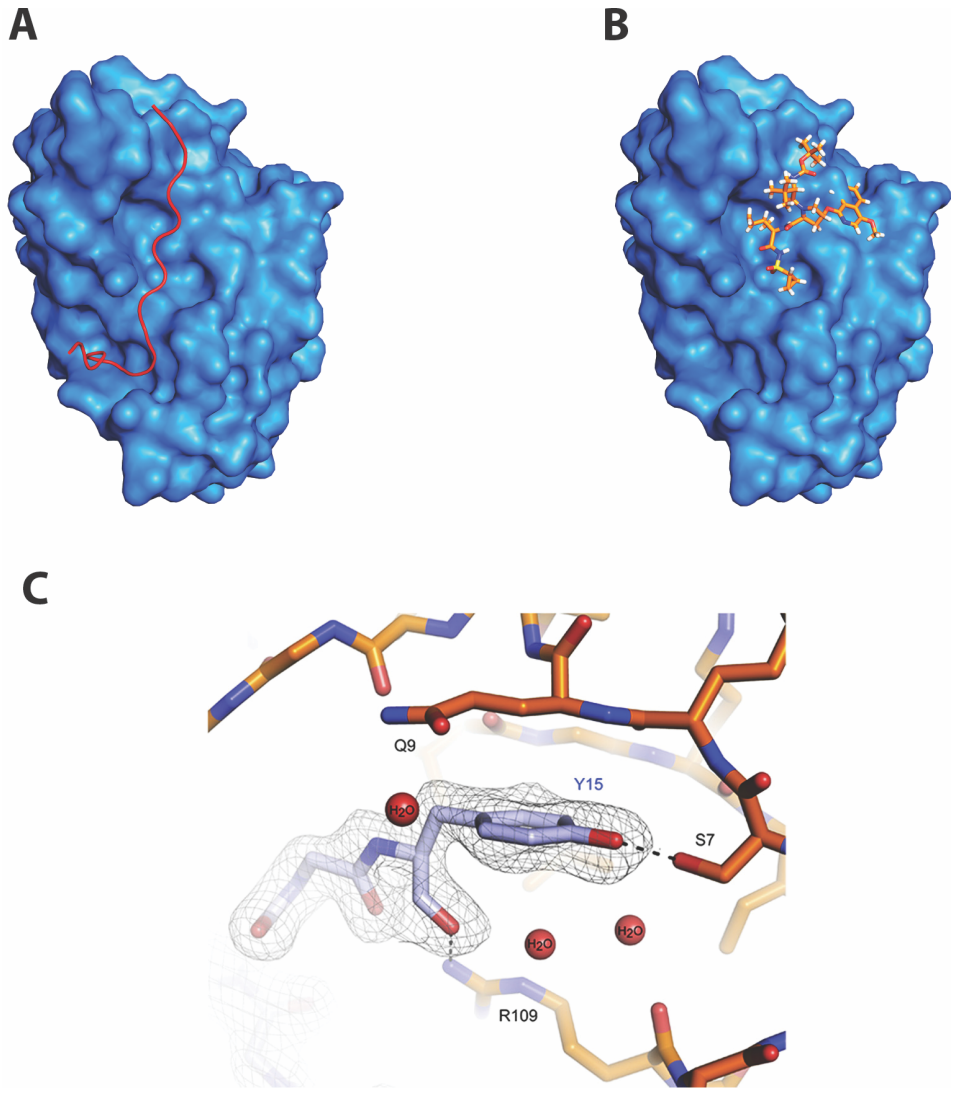


Figure 3: Crystal structures of NS3-4A bound to the Cp5 peptide and Asunaprevir. A and B: Aligned crystal structures of NS3-4A bound to Cp5 (A, PDB 4A1X) and Asunaprevir (B, PDB 4WF8) illustrating that while Cp5 binds the same surface as Asunaprevir, it also occupies additional binding pockets distal to the catalytic core. **C:** Image adapted from Kugler et al showing the tyrosine residue of Cp5 that forms the novel tyrosine-finger binding motif.

III. Results and Discussion

A. Bio-Physical Optimization of NS3-4A

Prior to determining the capability of NS3-4A and Cp5 to function as a CDD, we set out to confirm that Cp5 binds and inhibits NS3-4A protease activity in our hands. NS3-4A was expressed as a fusion to SNAP-Tag to increase its solubility. Cp5-46-A (GELGRLVYLLDGPGYDPIHSD) was expressed as a GST-fusion. Two NS3-4A constructs were generated, the NS3-1b genotype and a variant where helix-α0 was mutated from LLGCII to AAGAAA. Constructs were assayed using a simple proteolysis assay in which a fluorescently labeled peptide contains a fluorescence quencher distal to the fluorophore. In the absence of a protease, low fluorescence is observed. Upon cleavage of the peptide, an increase in fluorescence can be monitored over time (Figure 4A).

As a control, the potent, covalent NS3-4A protease inhibitor Telaprevir was used as a standard for complete protease inhibition (Figure 4B). Each peptide was subsequently titrated against both NS3-4A variants to determine the concentration necessary to inhibit 50% of enzyme activity, (IC_{50}'s, Figure 4C). Telaprevir potently inhibits NS3-4A. Interestingly, Cp5 inhibits NS3-4A-1b potently, yet a significant loss of potency is observed for the alanine mutant. Fundamentally, this result implies that NS3-4A and Cp5 are candidates for a bio-orthogonal CDD switch.

To successfully utilize the NS3-4A protease as a component of the CDD switch, it must behave well bio-physically. Specifically, it must not non-specifically aggregate or undergo significant degradation upon expression in mammalian

cells. The Cp5 peptide was selected against the NS3-1b HCV genotype, a particularly prevalent variety. However, the NS3-1b protease has been shown to be particularly difficult to express using bacterial expression systems[20]. Aggregation of NS3-4A constructs fused to effector proteins in mammalian cells could lead to aberrant phenotypes. To determine whether NS3-1b fused to NS4A would potentially aggregate, NS3-4A-1b was assayed for aggregation using yeast display.

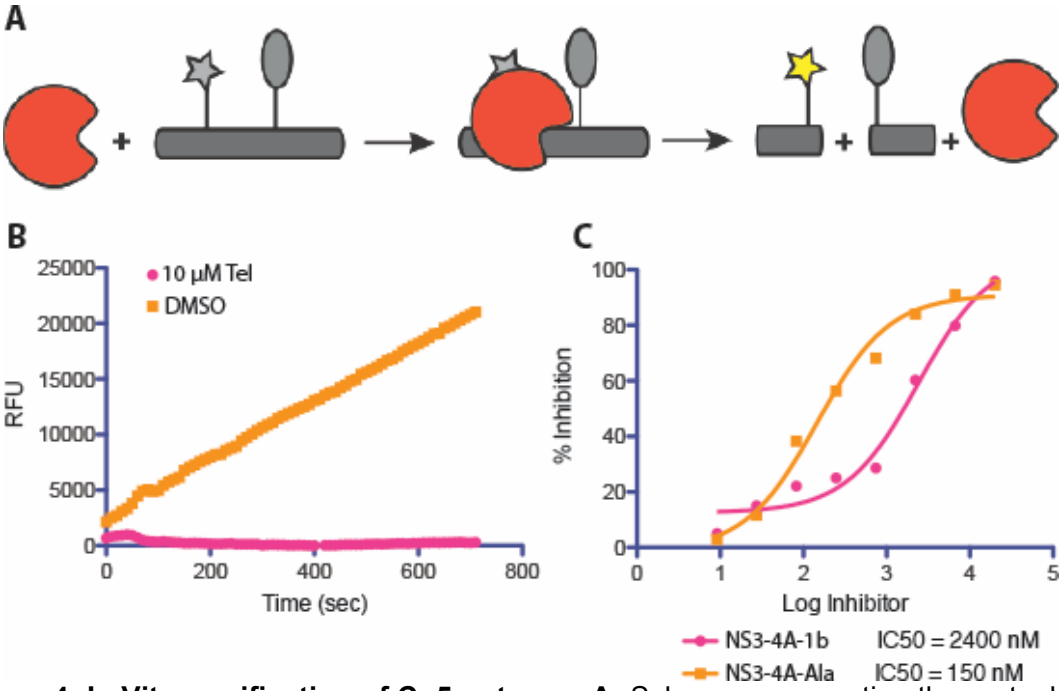


Figure 4: In Vitro verification of Cp5 potency: A: Scheme representing the proteolysis assay used to verify Cp5 inhibition. **B:** Raw plot over time showing that treatment of NS3-4A-1b with 10 µM Telaprevir completely inhibits protease activity. **C:** IC_{50} plots depicting the potency of Cp5 against two NS3-4A constructs.

Yeast expressing BirA and an AviTagged NS3-4A-1b, Cp5, or a non-binding control protein D5, were treated with streptavidin-phycoerythrin. Yeast were then exposed to myc-tagged NS3-4A-1b or Cp5 that were fluorescently

labeled with FITC labeled anti-myc antibodies. Yeast were sorted via FACS (Figure 5).

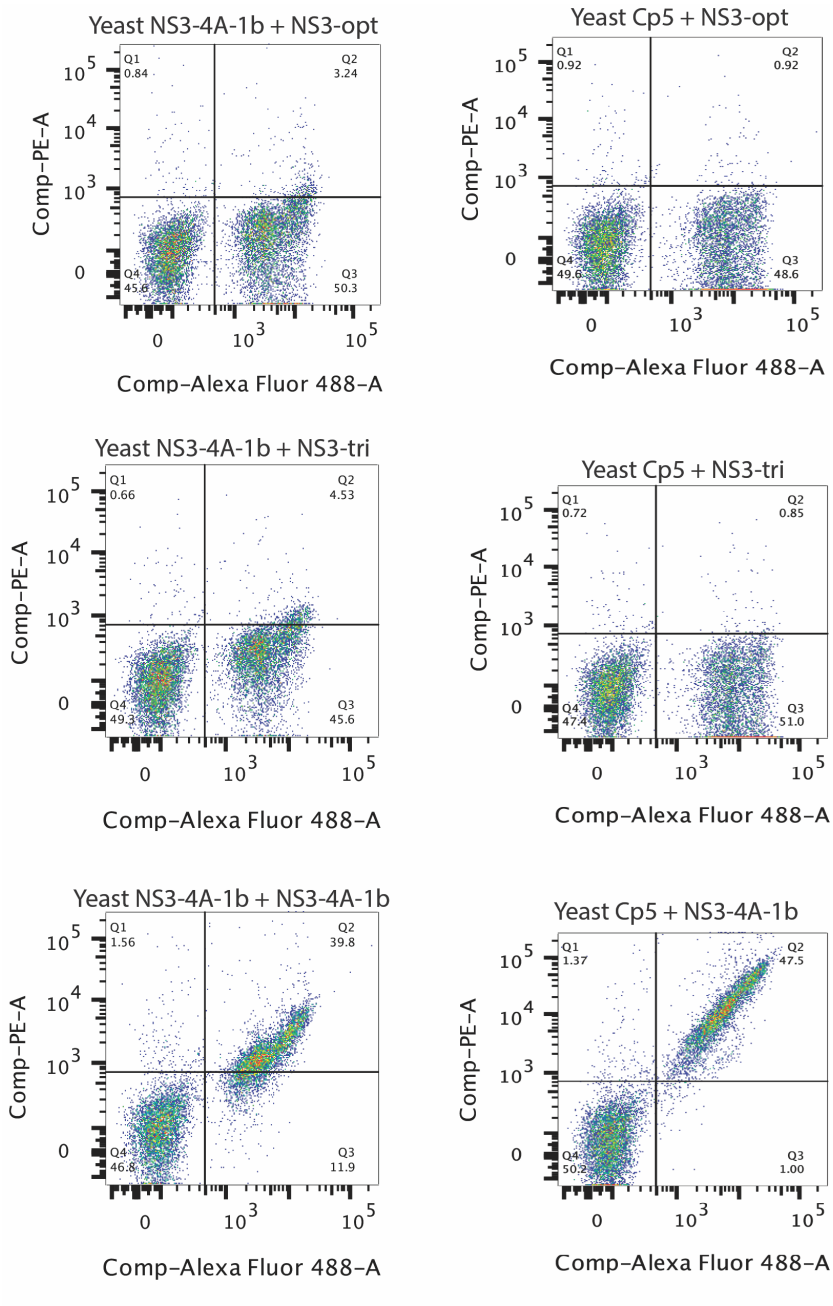


Figure 5: FACS data illustrating NS3-4A-1b's aggregation propensity. Yeast expressing surface exposed NS3-4A-1b shows no aggregation when exposed to NS3-opt or NS3-tri. However, NS3-4A-1b aggregates when exposed to another copy of NS3-4A-1b. Yeast expressing surface exposed Cp5 fails to bind NS3-opt and surprisingly NS3-tri, while clearly binding NS3-4A-1b.

Our yeast display results clearly show that NS3-4A-1b aggregates to such a degree as to cause concern in its use as an intra-cellular component of a CDD. Therefore, we next characterized an aggregation resistant, solubility optimized NS3-4A variant (NS3-opt) that had been[20] developed for biochemical, biophysical, and structural studies. Significant portions of the NS3 protease domain are mutated in NS3-opt, which increase solubility but also may negatively impact the binding of Cp5 (Figure 6). Of particular concern were NS3 residues that appear to make contact with two Cp5 binding motifs. The tyrosine finger binding motif (Figure 3), which was a key component in the design of Cp5 binding to NS3-4A, has lost these three critical contacts in NS3-opt.

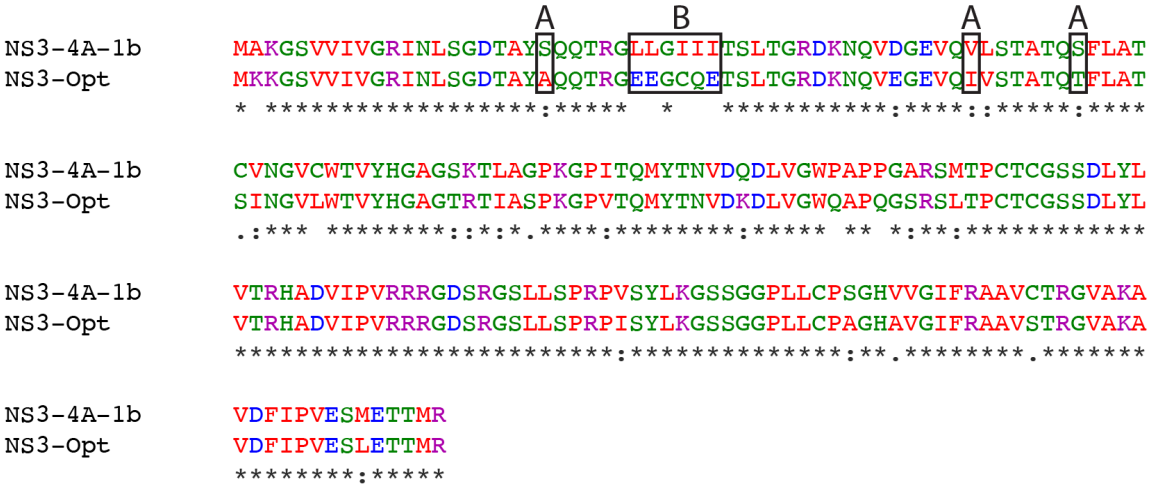


Figure 6: Sequence alignment of NS3-4A-1b and NS3-Opt: The sequence of NS3-4A-1b was aligned to the solubility optimized NS3 variant NS3-opt. Tyrosine finger binding motif interacting residues are shown in boxes labeled 'A' and the hypothetical helix interacting motif is shown in the box labeled "B".

NS3-opt and an NS3-opt variant where the tyrosine finger binding motif was restored (NS3-tri) were assayed with the same yeast display assay (Figure 5). Myc-tagged NS3-opt, NS3-tri, and NS3-4A-1b were exposed at a

concentration of 5 μM to yeast displaying Cp5 (Figure 5). As predicted, NS3-opt and NS3-tri show less non-specific binding than NS3-4A-1b (Figure 5).

An additional potential Cp5 binding interaction was observed in the crystal structure. In NS3-opt, an α-helix (helix-α0 of NS3-4A-1b, PDB: 4A1X) known to interact with the plasma membrane[23] during viral infection has been mutated to enhance solubility. This α-helix in the wild-type NS3-1b sequence consists of

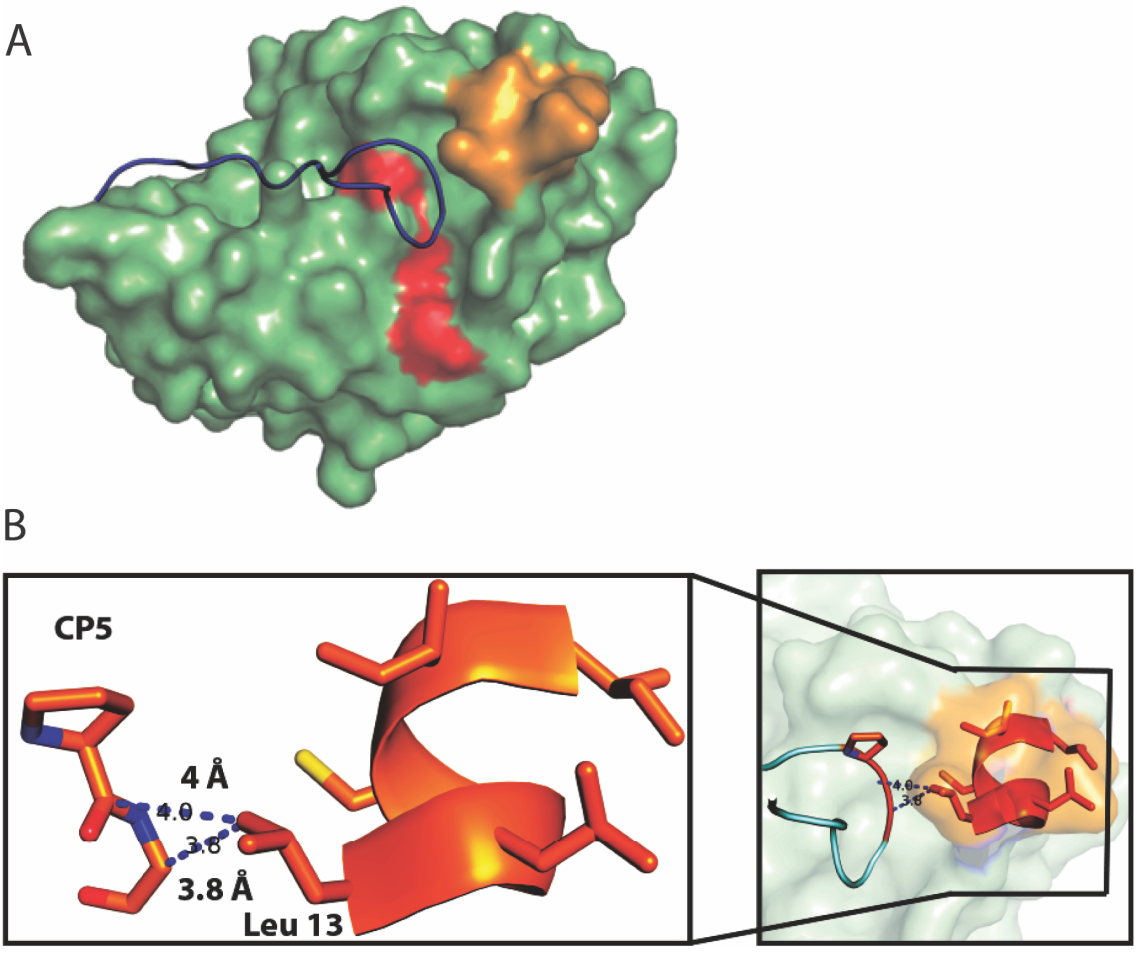


Figure 7: Interactions between Helix-α0 and Cp5: A) Co-crystal structure (PDB: 4A1X) of Cp5 (orange/cyan) and NS3-4A-1b (red/light green). **B)** Close up view of the Helix-α0 interaction with Cp5. Leucine 13 of Helix-α0 (orange) is close proximity to Cp5 (Red/Blue).

hydrophobic residues LLGCII and was mutated to EEGCQE in NS3-opt. The crystal structure of the Cp5/NS3-4A-1b complex indicated that the leucine in the first position of helix-α0 appears to be in close proximity to Cp5 (Figure 7A and B) and its conversion to glutamate in NS3-opt may negatively impact binding.

To identify an NS3-4A construct that shows reduced aggregation but retains Cp5 binding, a panel of NS3-4A variants was generated. All constructs utilized the NS3-opt sequence as the backbone. To maintain the solubility optimized characteristics of NS3-opt but to restore Cp5 binding, the tyrosine finger binding motif was re-introduced and individual residues from wild-type helix-α0 were introduced one at a time (Table 1). Yeast display provided information regarding propensity for aggregation, however it is not ideal for measuring affinity in solution. Furthermore, a functional intra-cellular assay is highly useful for prioritizing construct because a broadly useful CDD system must be functional in mammalian cells. To this end, each of the NS3-4A chimeras were assayed using a fluorescence co-localization assay in mammalian cells. The ability of an EGFP-Cp5 construct to co-localize with each mitochondrially-localized NS3-4A variant was determined. In addition, the ability of a small molecule drug to disrupt co-localization was also determined.

NIH 3T3 cells were co-transfected with each component of the CDD switch. EGFP was tagged with two copies of the Cp5 peptide or a weaker NS3 binding peptide (DECA), separated by flexible 25 amino acid linkers composed of glycine, serine, and threonine. NS3-4A constructs (Table 1) were tagged at their C-terminus with mCherry and at their N-terminus with the mitochondrial

localization domain of the Mitochondrial Import Receptor Subunit Tom20 (Tom20). Cells were treated with DMSO or 10 μM of the clinically approved NS3 protease drug Asunaprevir (which exhibits sub-nanomolar affinity for wild-type NS3-1b) for 5 minutes. Cells were fixed and imaged by confocal fluorescence microscopy (Figure 8A-C) and the degree of co-localization was determined for each cell.

Helix Name	Sequence
NS3-4A-H1	LEGCQE
NS3-4A-H2	ELGCQE
NS3-4A-H3	LLGCQE
NS3-4A-H4	LLGCIE
NS3-4A-H5	LLGCII
NS3-4A-H6	LEGCIE
NS3-4A-H7	EEGCQE

Table 1: Sequence variants tested in the co-localization assay. The crystal structure suggested that the first leucine residue contacts Cp5. Residues from NS3-4A-1b's helix were reintroduced to specific positions to try and recapitulate the binding affinity observed in wild-type NS3-4A-1b.

Chimeras H1-H4 exhibit significant co-localization with EGFP when treated with DMSO (Figure 8). This implies that reverting the first leucine at the N-terminus of helix-α0 is sufficient to restore Cp5 binding in a biologically relevant context. Treatment with 10 μM Asunaprevir induces a loss of fluorescence co-localization that corresponds to the degree to which the first four α-helix residues are restored. Interestingly, constructs in which the majority, or

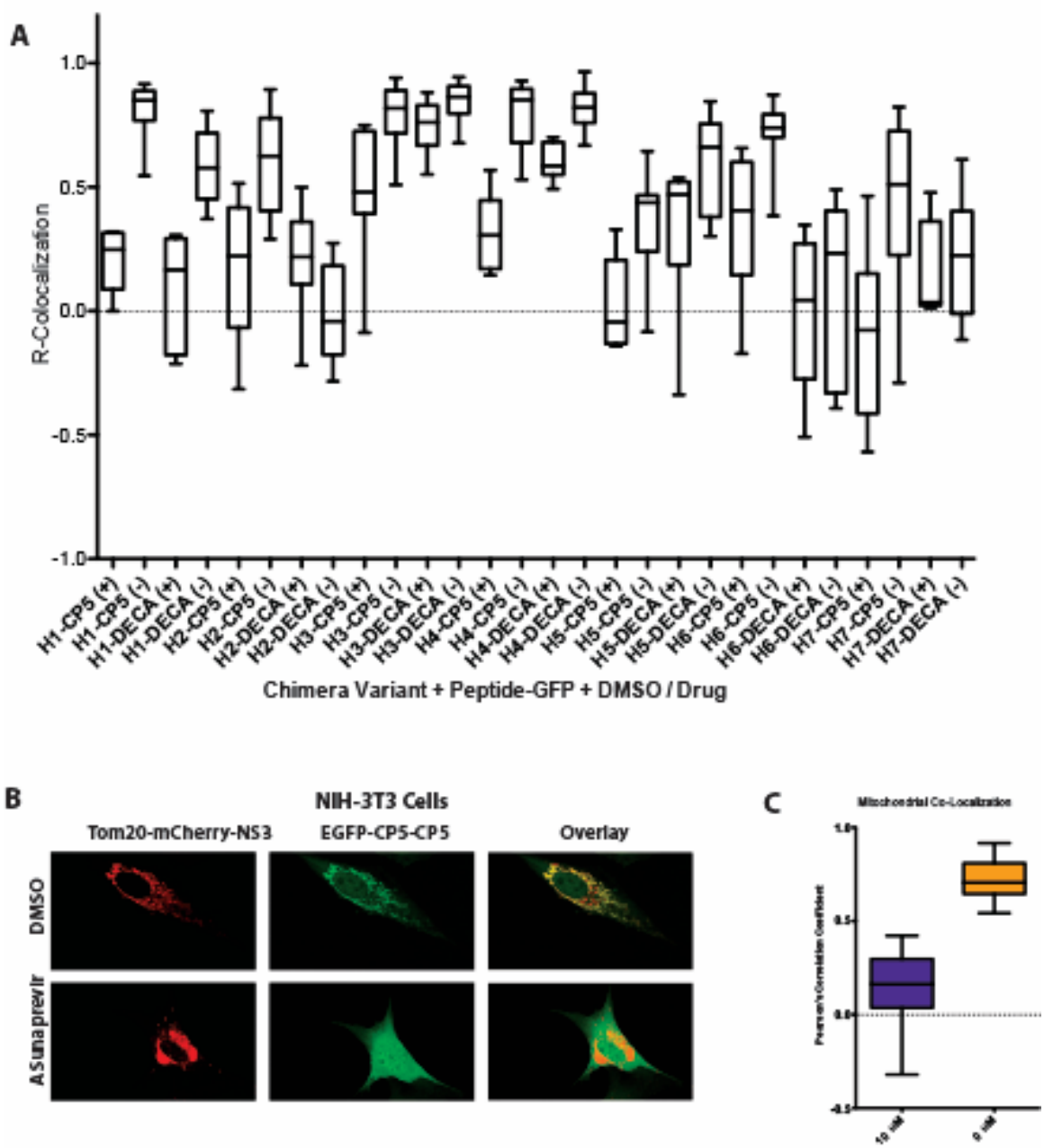
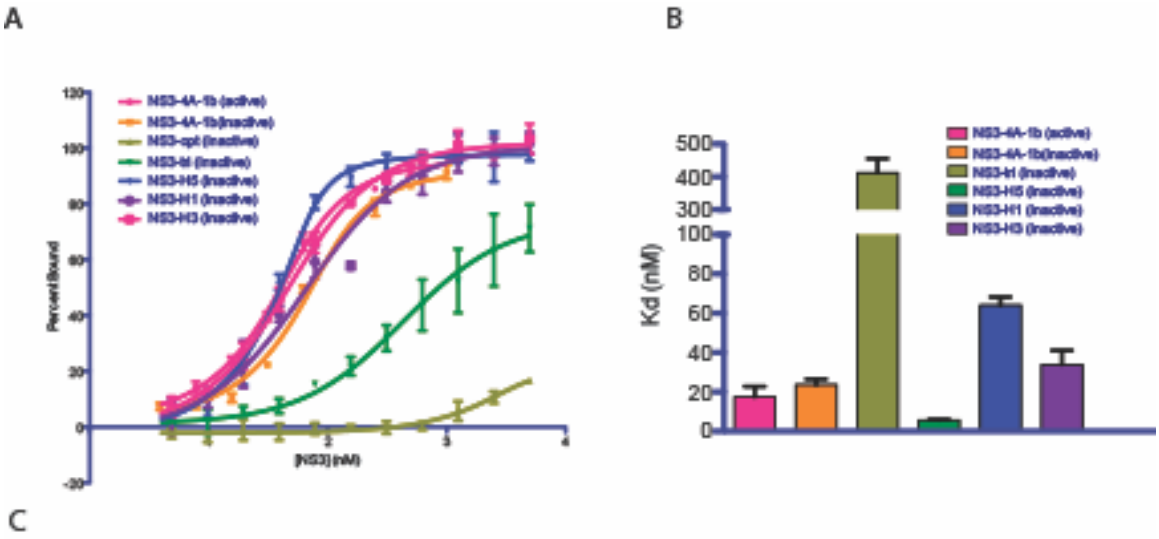


Figure 8: Mitochondrial co-localization of HCV-CDD: A) NS3-chimeras were localized to the mitochondria via a Tom20 tag. EGFP-Cp5-Cp5 was co-transfected, and cells were treated with DMSO or drug for 5 minutes. Several chimeras demonstrate restored Cp5 binding. 15 cells for each variant were imaged and Pearsons R-co-localization correlation coefficients were calculated for each cell using ImageJ. **B)** Representative images of mitochondrial localized tags for cells treated with DMSO or 10 μM Asunaprevir for 5 minutes. **C)** Quantification of NS3-H1/Cp5 co-localization was obtained using 50 cells for each condition at 5 minutes.

all, of the hydrophobic residues are restored display significant decreases in co-localization in the absence of drug as well as a higher degree of variability in co-localization. It is likely that these residues contribute significantly to the aggregation prone behavior displayed in wild-type NS3-1b.

The lower affinity DECA peptide appears to be harder to displace than the Cp5 peptide, despite having a significantly lower affinity for NS3-1b. As DECA makes no contacts to the hydrophobic helix-α0, it is largely uninfluenced by the chimeric mutations. Why this peptide is more difficult to displace than Cp5 when the affinity of the peptide for Ns3-1b is approximately 100x weaker is unknown. However, the high degree of co-localization exhibited by both NS3-H1 and NS3-H3 and the rapidity with which they can be displaced by Asunaprevir clearly demonstrate that these components are capable of functioning as a bio-orthogonal CDD.

To more fully characterize the NS3-chimera variants, the Kd was determined for several of the constructs. Catalytically dead (S139A) NS3-H1, NS3-H3, NS3-4A-1b, NS3-Opt, and NS3-Tri and catalytically active NS3-4A-1b constructs were generated and tested. All of these constructs were expressed and purified without solubility tags. Fluorescently labeled Cp5 peptide was purchased from GenScript. Constructs were assayed using a fluorescence polarization assay by titrating the NS3-4A variant to a constant concentration of Fluorescent Cp5 (Figure 9A-C).



Construct	K_D (nM)
NS3-4A-1b (active)	17 ± 5
NS3-4A-1b (inactive)	24 ± 12
NS3-opt	> 5000
NS3-tri	430 ± 110
NS3-H5	6 ± 3
NS3-H1	64 ± 25
NS3-H3	34 ± 8

Figure 9: Biochemical characterization of the binding affinities of NS3 variants. A) Binding affinities were calculated the anisotropy values and plotted as percent bound. NS3 variants were titrated against FAM-Cp5. **B)** K_D's determined from fluorescence polarization assay. Ns3-opt was not included as it showed no binding to NS3 up to 5000 nM. **C)** K_D's for each construct listed. Note that all constructs are catalytically inactive unless specified in the name. WT construct implies the NS3-opt scaffold with the tyrosine finger and helix-α0 completely restored.

Biochemical characterization of NS3 variants confirms trends observed in fluorescence co-localization assay. Restoration of the tyrosine finger binding motif provided a dramatic increase in affinity, from >5000 nM to 400 nM. Inclusion of the first leucine residue in helix-α0 provides an increase in affinity of another eight-fold, while including the first and second leucine residues almost entirely restores wild-type affinity. Not only do these results verify that we can restore Cp5 binding to a solubility optimized variant, but it also provides an

additional component of this toolkit. As Cp5 has an assortment of affinity variants, having NS3 constructs with tunable affinities increases the utility of this switch.

B. Expanding HCV-CDD to Multiple Sub-Cellular Compartments

Mitochondrial localization of HCV-CDD illustrates the capability of this system to behave as a bio-orthogonal CDD in mammalian cells. To demonstrate the versatility of this system, it must be applicable to other cellular compartments. To this end, we aimed to demonstrate a displaceable co-localization at the plasma membrane and within the nucleus.

An N-terminal myristoylation sequence (**GCGCSSHPEDD**) was used to localize mCherry followed by two repeats of the Cp5 peptide (myr-mCherry-Cp5-Cp5), separated by a 25 amino acid flexible linker, to the plasma membrane. The myr-mCherry-Cp5-Cp5 construct was co-transfected with NS3-H1 tagged EGFP, separated by a flexible 25 amino acid linker. Cells were treated with DMSO or 10 μM Asunaprevir for 5 minutes, fixed, and imaged using confocal fluorescent microscopy (Figure 10A and B). Cp5 was localized to the plasma membrane instead of NS3 for quantification as no co-localization was observed when NS3 was localized via the myristoylation sequence with cytosolic Cp5.

Cells expressing myristoylated Cp5 show a high degree of co-localization with cytosolic NS3-H1-EGFP (Figure 10A and B). When cells were treated with 10 μM Asunaprevir for 5 minutes, significant loss of co-localization was observed. Cells transfected with myr-mCherry-Cp5-Cp5 and EGFP without NS3-4A, no co-localization is observed.

To verify that HCV-CDD can be utilized inside the nucleus, a construct containing a triple nuclear localization sequence (3x-NLS) tagged to Blue Fluorescent Protein and dual repeats of the Cp5 peptide were used. NIH 3T3 cells were co-transfected with the 3x-NLS-BFP-Cp5-Cp5 construct and NS3-H1 tagged with EGFP. Cells were treated with 10 μM Asunaprevir for 5, 15, or 60 minutes or treated with DMSO. Cells were fixed and imaged using confocal fluorescence microscopy (Figure 10C and D). Cells treated with DMSO exhibit significant levels of co-localization. Treatment with 10 μM Asunaprevir results in a time dependent loss of co-localization. Complete loss of co-localization is observed at 60 minutes, with significant loss beginning to occur at 15 minutes.

C. HCV-CDD Regulation of the AKT Pathway

Eukaryotic plasma membranes are rich and diverse environments where proteins interact with lipids, metabolites, scaffolds, enzymes, and a variety of other molecules to regulate signaling networks[24,25]. These membranes act as the barrier at which extra-cellular signals are transduced. Additionally, the lipids themselves play a role in signaling and of particular importance in the phosphoinositide 3-kinase (PI3K) pathway[26]. This pathway has been of particular clinical interest due its promotion of cancer when loss of regulation occurs[27].

The PI3K family of proteins is divided into four classes, Class I, Class II, Class III, and Class IV. Activation of this pathway occurs mainly through upstream activation of G-Protein Coupled Receptors (GPCR's) or Receptor Tyrosine Kinases (RTKs). Activation that results in intra-cellular receptors that are tyrosine phosphorylated leads to recruitment of PI3K via their SH2 domains.

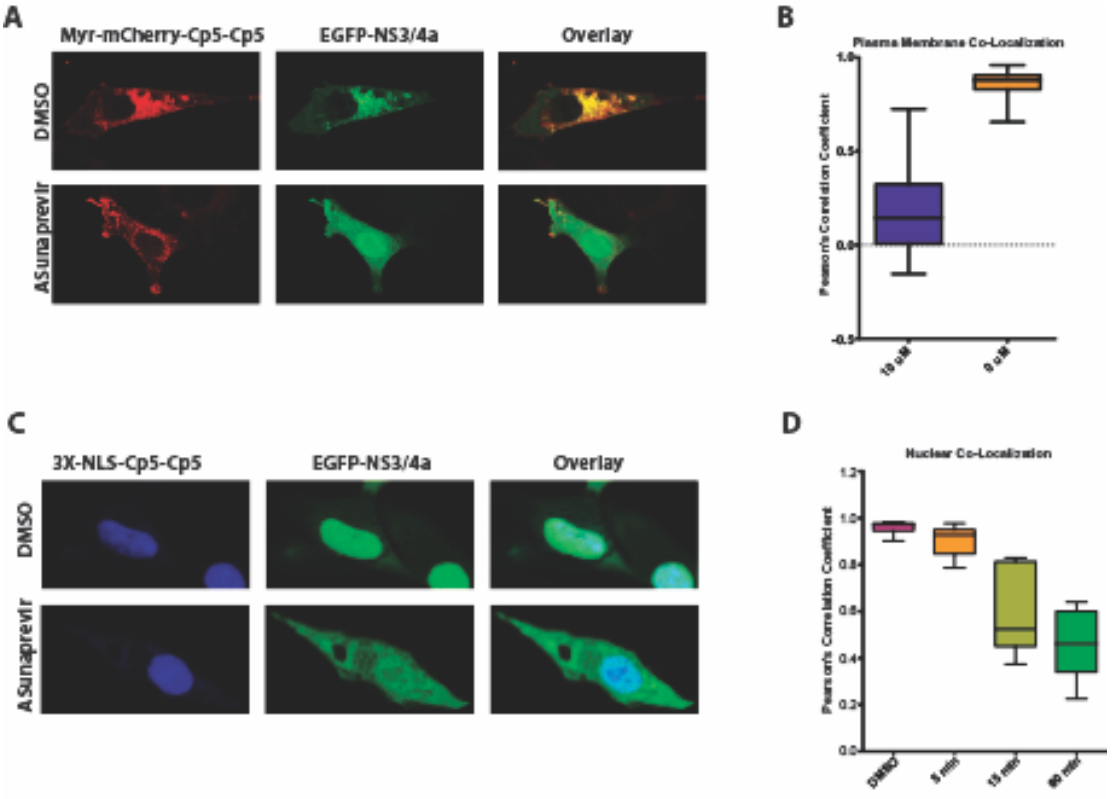


Figure 10: Applying HCV-CDD to other cellular compartments: A) Representative images depicting myristoylated mCherry-Cp5-Cp5 co-transfected with NS3-H1-EGFP after being treated with DMSO or 10 μM Asunaprevir for 5 minutes. **B)** Quantification of plasma membrane co-localization of myr-mCherry-Cp5-Cp5 co-transfected with NS3-H1-EGFP after being treated with DMSO or 10 μM Asunaprevir for 5 minutes. **C)** Representative images depicting nuclear localized 3x-NLS-BFP-Cp5-Cp5 co-transfected with NS3-H1-EGFP after being treated with DMSO or 10 μM Asunaprevir for 60 minutes. **D)** Quantification of nuclear localized HCV-CDD switch at 0, 5, 15, and 60 minutes after treatment with 10 μM Asunaprevir.

Class 1A PI3Ks are composed of a heterodimer between the regulatory p85 subunit and the catalytic p110 subunit. The p85 subunit contains an SH3 domain, a Rho-GAP domain, and two SH2 domains. The two SH2 domains bind to phosphorylated tyrosines on the activated receptor and subsequently recruit p110 to the inter-SH2 (iSH2) domain, turning on the catalytic activity of p110[28]. This recruitment induces p110 to convert PIP2 to phosphatidylinositol(3,4,5)-triphosphate (PtdIns[3,4,5]-P3). This lipid then acts as a platform for recruiting a variety of signaling enzymes that contain Pleckstrin Homology (PH) domains and canonically results in downstream phosphorylation of Protein Kinase B.

Localization of the iSH2 domain alone has been shown to be sufficient for the production of PtdIns[3,4,5]-P3. Protein Kinase B (AKT) binds PtdIns[3,4,5]-P3 and is phosphorylated by mTOR Complex 2 and PDK1, resulting in it's activation. iSH2 has been shown to be capable of activating the AKT pathway by localization via alternative CID systems[29]. Activation through recruitment has been well characterized, yet no means currently exist that can dissociate iSH2. Such a system could be readily used to measure the rate at which this pathway is turned off.

To measure the degree to which pAKT dephosphorylation can be controlled by dissociation of the iSH2 domain, an NS3-4A-H1-iSH2 fusion was generated with a 21 amino acid flexible linker separating the two domains. Cos-7 cells were co-transfected with NS3-4A-H1-iSH2 and myr-mCherry-Cp5-Cp5. Cells were serum starved for 18 hours and treated with 10 μM Asunaprevir for 5, 10, and 15 minutes or treated with DMSO. Cells were harvested and subjected to

a western blot for pSerine473 of AKT, a key marker for AKT activation (Figure 11A and B). Drug untreated cells exhibited a ratio of pAKT/AKT 3-fold greater than background. Addition of 10 μM Asunaprevir led to an immediate decrease in pAKT/AKT. At 5 minutes, complete loss of pAKT signal was observed. Transfection of NS3-H3-iSH2 alone, led to a slight increase in pAKT/AKT ratio that was not affected by Asunaprevir treatment.

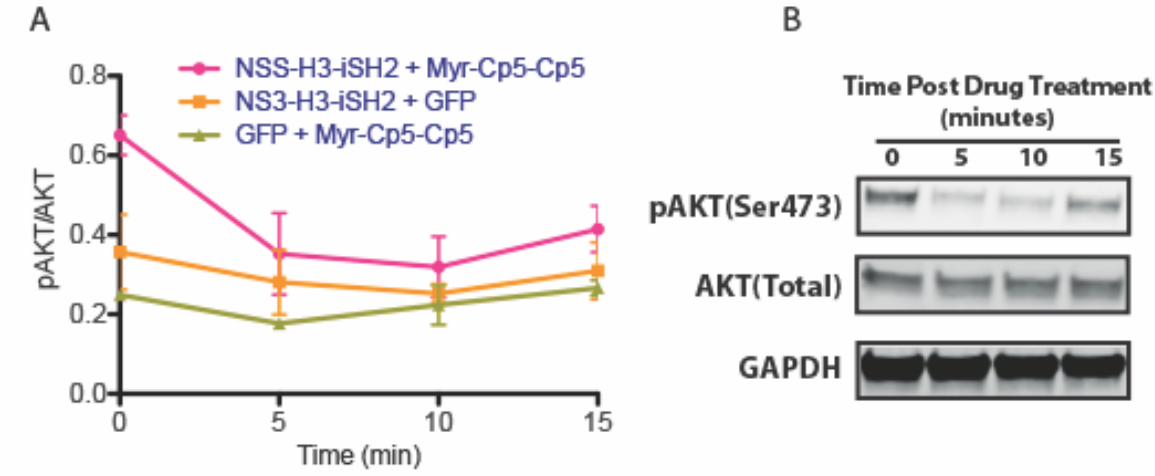


Figure 11: HCV-CDD based regulation of the AKT pathway: A) Time course plotting the ratio of pAKT to total AKT for cells co-transfected with combinations of NS3-H3-iSH2, Myr-Cp5-Cp5, and GFP. **B)** Representative western blots used for quantification of pAKT-AKT ratios.

IV. Conclusion

The need for a bio-orthogonal, chemically inducible disruptor of dimerization led to the development of such a system based on the Hepatitis C Virus Protease NS3. The minimal NS3-4A fusion protein was optimized for bio-physical characteristics while maintaining affinity for the inhibitory peptide Cp5. The protein/peptide dimer was shown to behave well in mammalian cells. This was verified experimentally by localizing NS3-4A to the mitochondria and

recruiting Cp5-tagged EGFP and by localizing Cp5 to the plasma membrane, as well as the nucleus, and recruiting fluorescently tagged NS3-4A to it. To further demonstrate the utility of this switch, the rate of de-phosphorylation of pAKT was measured by recruiting NS3-4A-iSH2 to myristoylated Cp5 and releasing iSH2 by adding the clinically approved drug Asunaprevir. Currently, efforts are underway to further expand this methodology to recruit effector proteins that modulate the amplitude of AKT signaling.

V. Materials and Methods:

1. NS3-4A and Peptide Variant Cloning, Expression, and Purification.

A. Gene Synthesis General Methods

CLONING OF PEPTIDE-GST CONSTRUCTS

The peptide and GST genes were obtained as double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). Peptide and GST genes were subcloned into pMCSG7, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF SNAPTAG-NS3/4A CONSTRUCTS

A mutant form of the human gene O^6-alkylguanine-DNA-alkyltransferase was amplified from the pSS26(b) plasmid (Covalys) using Gibson Assembly (NEB) overhang primers designed in NEBuilder. The NS3/4a-1b gene was obtained as double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). Protease and SNAPtag genes were subcloned into

pMCSG7 a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF NS3/4A VARIANT CONSTRUCTS

The NS3/4A variants were obtained as double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). Protease genes were subcloned into pMCSG7, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

B. Fusion Construct Protein Sequence

Peptide-GST Design:

His_6-$(GS/T)_3$-Cp5-GTGS-GS

His_6-$(GS/T)_3$-DECA-GTGS-GS

Cp5 Sequence:

GELGRLVYLLDGPGYDPIHSD

DECA Sequence:

DELILCPLDL

GST Sequence:

SPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNL

PYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAY

SKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLY

MDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK

SDLVPR

SNAPtag-NS3/41 Design:

SNAPtag-GTGT-NS3/4A-GS-His_6

SNAPtag Sequence:

DKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLG

GPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVVK

FGEVISYQQLAALAGNPAATAAVKTALSGNPVPILIPCHRVVSSSGAVGGYEGG

LAVKEWLLAHEGHRLGKPGLG

NS3-4A-1b Sequence:

AKGSVVIVGRINLSGDTAYSQQTRGAAGTAATSATGRDKNQVDGEVQVLSTAT

QSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGAR

SMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLC

PSGHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR

NS3-Opt Sequence

MKKKGSVVIVGRINLSGDTAYAQQTRGEEGCQETSQTGRDKNQVEGEVQIVST
ATQTFLATSINGVLWTVYHGAGTRTIASPKGPVTQMYTNVDKDLVGWQAPQGS
RSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLC
PAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRSP

NS3-Tri Sequence

MKKKGSVVIVGRINLSGDTAYSQQTRGEEGCQETSQTGRDKNQVEGEVQVVS
TATQSFLATSINGVLWTVYHGAGTRTIASPKGPVTQMYTNVDKDLVGWQAPQG
SRSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGSAGGPLL
CPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETTMRSP

C. Expression and Purification Protocol:

Cp5-GST Expression and Purification

The His_6-Cp5-GST plasmid was transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to inoculate 250 mL of LB both with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 0.5 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For Cp5-GST purification, the pellet was thawed on ice and re-suspended in 10 mL of His_6 Lysis Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20 mM imidazole, 5 mM DTT) supplemented with PMSF (1 mM). The re-

suspended cell pellet was lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of Lysis Buffer and the protein was eluted in 3 mL of Elution Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 200 mM Imidazole, 5 mM DTT). Purified protein was dialyzed twice into 1000 mL Storage Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 5 mM DTT). Protein was stored by snap-freezing aliquots and storing at -80 °C.

DECA-GST Expression and Purification

The His_6-DECA-GST plasmid was transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to inoculate 250 mL of LB both with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 0.5 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. Protein was purified and dialyzed using the same protocol as Cp5-GST.

SNAPtag-NS3/4A-1b Expression and Purification

The SNAPtag-NS3/4A-His_6 plasmid was transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to

inoculate 500 mL of LB both with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 0.25 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For SNAPtag-NS3/4A purification, the pellets were thawed on ice and re-suspended in 10 mL of LS-His_6 Lysis Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20% (w/v) glycerol, 20 mM imidazole, 5 mM DTT). Supplemented with PMSF (1 mM). The re-suspended cell pellet was lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of LS-Lysis Buffer and the protein was eluted in 3 mL of LS-Elution Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20% (w/v) glycerol, 200 mM Imidazole, 5 mM DTT). Purified protein was dialyzed twice into 1000 mL LS-Storage Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20% (w/v) glycerol, 5 mM DTT, 0.6 mM lauryldimethylamine-N-oxide). Protein was stored by snap-freezing aliquots and storing at -80 °C.

NS3-4A Variant (Opt, Tri) Expression and Purification

Proteins were expressed in BL21 (DE3) *E. coli* at 37°C to an O.D.600 of 0.5-1.0, then moved to 18°C and induced to 0.5 mM IPTG overnight. For biotinylated constructs, 12.5 mg D(+)-biotin/L culture was added upon inoculation with overnight culture. After 16-20 hours overnight grown, cultures were harvested, and cell pellets frozen at -80°C. Cell pellets were re-suspended in 20 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole, 1 mM DTT, 0.1% Tween-20. All

buffers for NS3a purifications additionally included 10% v/v glycerol. Cells were lysed by sonication, and supernatant was incubated with Ni-NTA resin (Qiagen) for at least 1 hour at 4°C. Resin was washed with 20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, and proteins were eluted with 20 mM Tris pH 8.0, 500 mM NaCl, 300 mM imidazole. Biotinylated constructs were then further purified by size exclusion chromatography on a Superdex-75 10/300 GL column (GE Healthcare) in a buffer of in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, 10% glycerol.

D. NS3-4A Protease Activity Assay

The potency of the peptide-GST inhibitors against NS3-4A protease was characterized with a FRET assay. Titrations of peptide-GST inhibitors (3-fold serial dilutions starting at 10 μM) were added to a black 96-well plate (Corning, product number 3720) containing 50 nM NS3-4A. Inhibitors were incubated with NS3-4A at room temperature for 1 hour. To each well was simultaneously added substrate M2235 (Bachem) to a final concentration of 5 μM and reactions were monitored by measuring the fluorescence increase every minute for 30 minutes at 22 °C on a Perkin Elmer EnVision fluorimeter (excitation, 360 nm; emission 460 nm). Each measurement was carried out in triplicate. Slopes of the fluorescence increase were compared to a no-protease control. A nonlinear regression model was used to fit curves with GraphPad Prism.

E. NS3-4A Fluorescence Polarization Assay

The affinities of the NS3-4A variants for Cp5 were evaluated using a fluorescence polarization assay. Fluorescently labeled (FAM) Cp5 was obtained as a crude (GenScript) and purified by HPLC. Titration of NS3-4A variants (3-fold serial dilutions starting at 5 μM) were diluted in FP-Buffer (50 mM HEPES, pH 7.8, 100 mM NaCl, 5 mM DTT, 1% Glycerol, 0.01% Tween, 5% v/v DMSO) and were added to a black 96-well plate (Corning, product number 3720) containing 10 nM (final concentration) FAM-Cp5 and incubated at room temperature in the dark for 1 hour. Fluorescence polarization was measured at 22 °C on a Perkin Elmer EnVision fluorimeter (excitation, 495 nm; emission 520 nm). Each measurement was carried out in triplicate. Anisotropy values were obtained and a nonlinear regression model was used to fit curves with GraphPad Prism.

2. Yeast Display of NS3-4A Variants

A. Yeast Display Gene Synthesis General Methods

CLONING OF NS3-4A VARIANTS

The NS3-4A variants were obtained as double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). G-Blocks were subcloned into a pDW363 vector containing an N-terminal AviTag and hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

AviTag-Ns3/4A Variant Design:

AviTag-GGSS-His_6-$(GS)_3$-NS3/4A

AviTag Sequence:

AGGLNDIFEAQKIEWHED

B. Yeast Display and FACS Analysis

Yeast were grown overnight at 30 °C in yeast minimal media (-ura for strain selection, -trp for pETCON selection) supplemented with 2% w/v glucose. Overnights were used to inoculate SGCAA cultures (2% w/v galactose, 0.67% w/v yeast nitrogen base, 0.5% w/v casamino acids, and 0.1 M sodium phosphate, pH 6.6) to an O.D. 600 of 1.0-2.0 and protein expression (Cp5 or NS3-4A-1b) was induced overnight at 30 °C. Before sorting or analysis, cells were pelleted and re-suspended in PBS supplemented with 0.5% w/v bovine serum albumin (PBSA). Protein solutions of biotinylated Myc tagged NS3-4A variants were made in PBSA and incubated with the yeast for 30 min to 1 hour at 22 °C. Cells were washed in cold PBSA and incubated for 15 min on ice with SAPE and fluorescein isothiocyanate-conjugated chicken anti-c-myc (Immunology Consultants Laboratory), both diluted 1:100 in PBSA. After the labeling incubation, cells were washed again in cold PBSA and analyzed on a C6 flow cytometer (Accuri) or a FACS Canto cytometer (BD Biosciences), or sorted on a SH800 (Sony Biotechnology) cell sorter or a FACS Aria III (BD Biosciences) cell sorter. Sorted yeast recovered for 1-2 days at 30 °C in yeast minimal media plus 2% w/v glucose.

3. Fluorescence Co-Localization in Mammalian Cells

A. Eukaryotic Expression Vector Cloning General Methods

CLONING OF MITOCHONDRIAL LOCALIZED NS3-4A VARIANTS

The NS3-4A variants were obtained as mammalian codon optimized double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). G-Blocks were subcloned into a pcDNA5/FRT/TO vector containing an N-terminal Tom20-mCherry tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF PLASMA MEMBRANE LOCALIZED mCHERRY-CP5-CP5

Two repeats of Cp5, separated by a 25 amino acid linker were obtained as mammalian codon optimized double stranded DNA G-Blocks (IDT) containing an Gibson Assembly overhangs designed in NEBuilder (NEB). G-Blocks were subcloned into a pcDNA5/FRT/TO vector containing an N-terminal myristoylation sequence tagged with tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF NUCLEAR LOCALIZED CP5-CP5-BFP

Two repeats of Cp5, separated by a 25 amino acid linker were obtained as mammalian codon optimized double stranded DNA G-Blocks (IDT) containing an Gibson Assembly overhangs designed in NEBuilder (NEB). G-Blocks were subcloned into a pcDNA5/FRT/TO vector containing three NLS repeats tagged with BFP using Gibson Assembly (NEB, product number E2611L).

CLONING OF EGFP-TAGGED NS3-4A

NS3-4A-H1 was obtained as a mammalian codon optimized double stranded DNA G-Block (IDT) containing a set of Gibson Assembly overhangs designed in NEBuilder (NEB). G-Blocks were subcloned into a pcDNA5/FRT/TO vector containing a C-terminal EGFP tag using Gibson Assembly (NEB, product number E2611L).

Tom20-mCherry-Ns3/4A Variant Design:

Tom20-GSGSG-mCherry-$(GSG)_7$-NS3/4A-$(GSG)_6$-FLAG

EGFP-Cp5-Cp5

EGFP-GSGTG-Myc-$(GSGTG)_5$-Cp5-$(GSGTG)_5$-Cp5

Myr-mCherry-Cp5-Cp5 Design:

Myr-$(GS)_4$-mCherry-GSGTG-Myc-$(GSGTG)_5$-Cp5-$(GSGTG)_5$-Cp5

3xNLS-BFP-Cp5-Cp5 Design:

3xNLS-$(GS)_4$-BFP-GSGTG-Myc-$(GSGTG)_5$-Cp5-$(GSGTG)_5$-Cp5

EGFP-NS3-H1

EGFP-GSGTG-Myc-$(GSGTG)_5$-NS3/4A-H1

Tom20 Sequence:

VGRNSAIAAGVCGALFIGYCIYFDRKRRSDPNF

Myristoylation Sequence

GCGCSSHPEDD

3xNLS Sequence

DPKKKRKVDPKKKRKVDPKKKRKV

FLAG-Tag Sequence

DYKDDDDK

mCherry Sequence

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKV

TKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFED

GGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPE

DGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNED

YTIVEQYERAEGRHSTGGMDELYK

EGFP Sequence

VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP

VPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNY

KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNG

IKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEK

RDHMVLLEFVTAAGITLGMDELYK

BFP Sequence

SELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFA
FDILATSFLYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTS
LQDGCLIYNVKIRGVNFTSNGPVMQKKTLGWEAFTETLYPADGGLEGRNDMAL
KLVGGSHLIANIKTTYRSKKPAKNLKMPGVYYVDYRLERIKEANNETYVEQHEV
AVARYCDLPSKLGHKLN

B. NIH-3T3 Cell Culture and Transient Transfection Conditions

NIH-3T3 cells were maintained in DMEM (Gibco, product number 11065092) supplemented with 10% FBS (Gibco, product number A3160602). All transient transfections were done using LipoFectamine3000 (ThermoFisher, product number L3000015) at a ratio of 3:2:1 LipoFectamine3000:p3000Reagent:µgDNA prepared in OptiMem (Gibco, product number 11058021) 16-20 hours after plating of cells. Transfections were allowed to proceed for 24 hours before experiments were performed.

C. Confocal Microscopy to Track Fluorescence Co-Localization

24 hours prior to transfection, $3x10^4$ 3T3 cells were plated onto 18 mm glass cover slips (Fisher, product number 12-546) in a standard 12-well plate. After transfection with the appropriate localized NS3-4A/Cp5 pair, cells were allowed to recover for 24 hours before treatment with 10 µM Asunaprevir (ApexBio, product number A3195) or DMSO (0.5% DMSO final concentration).

Cells were incubated with drug for the stated time points before media was aspirated and cells were washed once with chilled PBS. After washing, cells were fixed in 4% paraformaldehyde (Electron Microscopy Services, product number 15710) prepared in 1x PBS for 15 minutes. Paraformaldehyde was removed and cells were washed twice with chilled PBS. Slides were mounted onto glass cover slips using Fluoromount G (Southern Biotechnology, product number 0100-01) and sealed. Images were generated using a Leica SP8X Confocal Microscope and the appropriate wavelengths to visualize EGFP, mCherry, and BFP. Quantification of co-localization was determined using ImageJ.

4. iSH2 Induced Activation of AKT pathway Assay

A. NS3/4A-iSH2 Expression Vector Cloning General Methods

CLONING OF iSH2-TAGGED NS3-4A

NS3-4A-H3 was obtained as a mammalian codon optimized double stranded DNA G-Block (IDT) containing a set of Gibson Assembly overhangs designed in NEBuilder (NEB). iSH2 was obtained as a mammalian codon optimized double stranded G-Block. G-Blocks were subcloned into a pcDNA5/FRT/TO vector containing a C-terminal FLAG tag using Gibson Assembly.(NEB, product number E2611L).

iSH2-NS3-H1 Layout Scheme

iSH2-$(GSGTG)_5$-NS3/4A(H1)-$(GSG)_6$-FLAG

iSH2 Protein Sequence

RLLYPVSKYQQDQIVKEDSVEAVGAQLKVYHQQYQDKSREYDQLYEEYTRTSQ
ELQMKRTAIEAFNETIKIFEEQGQTQEKCSKEYLERFRREGNEKEMQRILLNSE
RLKSRIAEIHESRTKLEQQLRAQASDNREIDKRMNSLKPDLMQLRKIRDQYLVW
LTQKGARQKKINEWLGIKNETEDQYALMEDEDDLP

B. Cos-7 Cell Culture and Transient Transfection Conditions

Cos-7 cells (ATCC, product number CRL-1651) were maintained in DMEM (Gibco, product number 11065092) supplemented with 10% FBS (Gibco, product number A3160602). All transient transfections were done using TurboFectin8.0 (Origene) at a ratio of 3:1 TurboFectin:µgDNA prepared in OptiMem (Gibco, product number 11058021) 16-20 hours after plating of cells. Transfections were allowed to proceed for 24 hours before experiments were performed.

C. HCV-CID Induced Activation of AKT Pathway Conditions

24 hours prior to transfection, $3x10^5$ Cos-7 cells were plated onto poly-D-lysine 24 well plates. After transfection with the appropriate DNA components, cells were allowed to recover for 24 hours. After 24 hours, media was aspirated and the cells were washed with 1 mL of 37 °C PBS then serum starved with FBS-free DMEM for 18 hours prior to drug treatment. Media were subsequently aspirated and the cells were washed twice with 1 mL chilled PBS, then lysed with 75 µL Mod. RIPA buffer (50 mM Tris, pH 7.8, 1% IGEPAL CA-630, 150 mM

NaCl, 1 mM EDTA, 2 mM Na_3VO_4, 30 mM NaF, Pierce Protease Inhibitor Tablet). Cleared lysates were subjected to SDS-PAGE and transferred to nitrocellulose. Blocking and antibody incubations were done in TBS with 0.1% Tween-20 (v/v) and blocking buffer (Odyssey). Primary antibodies were diluted as follows: Total AKT (1:2000, #2920), pAKT-Ser473 (1:2000, #4060), GAPDH (1:2000, #2118) all from Cell Signaling Technology. Blots were washed in TBS with 0.1% Tween-20. Antibody binding was detected by using near-infrared-dye-conjugated secondary antibodies and visualized on the LI-COR Odyssey scanner. Blots were quantified via densitometry with Image Studio (LI-COR).

VI. References

1. Banaszynski, L.A.; Wandless, T.J. Conditional control of protein function. *Chem. Biol.* **2006**, *13*, 11-21.

2. Dueber, J.E.; Yeh, B.J.; Chak, K.; Lim, W.A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **2003**, *301*, 1904-1908.

3. Buskirk, Allen R.; Liu, David R. Creating small-molecule-dependent switches to modulate biological functions. *Chem. Biol.* **2005**, *12*, 151-161.

4. Kuriyan, J.; Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **2007**, *450*, 983-990.

5. Inoue, T., Heo, W.D., Grimley, J.S., Wandless, T.J. & Meyer, T. An inducible translocation strategy to rapidly activate and inhibit small GTPase signaling pathways. *Nat. Methods*, **2015,** 2, 415–418.

6. Wang, H. et. al. LOVTRAP: an optogenetic system for photoinduced protein dissociation. *Nat. Methods*. **2016**, 13, 755-758.

7. Liang, F.; Ho, Q. W.; Crabtree, G. R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci. Signaling.* **2011**, 4, rs2.

8. DeRose, R., Miyamoto, T., & Inoue, T. Manipulating signaling at will: chemically-inducible dimerization (CID) techniques resolve problems in cell biology. *Pflugers Archiv : European Journal of Physiology*, **2013** *465*(3), 409–417.

9. Crabtree, G. R.; Schreiber, S. L. Three-part inventions: intracellular signaling and induced proximity. *Trends Biochem Sci.* **1996** *21*, 418-422

10. Banaszynski, L. A.; Liu, C. W.; Wandless, T. J. Characterization of the FKBP:Rapamycin:FRB ternary complex. **2005** *127*, 4715-4721

11. Spencer, D. M.; Belshaw, P. J.; Crabtree, G. R.; Schreiber, S. L. Functional analysis of Fas signaling *in vivo* using synthetic inducers of dimerization. *Curr. Biol.* **1996** *6(7)*: 839-848.

12. Belshaw, P. J.; Ho, S. N.; Crabtree, G. R.; Schreiber, S.L. Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins. *PNAS* **1996** *10*: 4604-4607.

13. Ho, S. N.; Biggar, S. R.; Spencer, D. M.; Schreiber, S. L.; Crabtree, G. R. Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature* **1996** *382(6594)*: 822-826.

14. Bayle, J. H. et al. Rapamycin analogs with differential binding specificity permit orthogonal control of protein activity. *Chemistry and Biology*. **2006** *13*, 99-107.

15. Kugler, J. et al. High affinity peptide inhibitors of the Hepatitis C Virus NS3-4A protease refractory to common resistant mutants. *J. Biol. Chem.* **2012**. *287*: 39224-39232.

16. Bartenschlager, R.; Frese, M.; Pietschmann, T. Novel insights into Hepatitis C Virus replication and persistence. *Adv. Virus. Res.* **2004** *63*: 71-180.

17. Bartenschlager, R.; Lohmann, V. Replication of the Hepatitis C Virus. *Baillieres Best Pract. Res. Clin. Gastroenerol.* **2000** *14*: 241-254.

18. Kim, J. L.; Morgenstern, K. A. et al. Crystal structure of the Hepatitis C Virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* **1996** *87*: 343-355

19. Tomei, L.; Failla, C. et al. A central hydrophobic domain of the Hepatitis C Virus NS4A protein is necessary and sufficient for the activation of the NS3 protease. *J. Gen. Virol* **1996** *77*: 1065-1070.

20. Wittekind, M.; Weinheirner, S.; Zhang, Y.; Goldfarb, V. *US Patent 6333186.* **2002**

21. Chung, H. K. et al. Tunable and reversible drug control of protein production via a self-excising degron. *Nat. Chem. Biol.* **2015** *11*: 713-720.

22. Romano, K. P.; Ali, A. et al. The molecular basis of drug resistance against Hepatitis C Virus NS3/4A protease inhibitors. *PLoS Pathog* **2012** *8*:e1002832.

23. Brass, V.; Berke, J. M. et al. Structural determinants for membrane association and dynamic organization of the hepatitis c virus NS3-4a complex. *PNAS* **2008** *105(38)*: 14545-14550

24. Wymann, M. P.; Schneiter, R. Lipid signaling in disease. *Nat. Rev. Mol. Cell Bio.* **2008** *9*: 162-176.

25. Wymann, M. P. Inflammation and Cancer. *Subcell. Biochem.* **2012** *58*: 111-181.

26. Varnai, P.; Thyagarajan, B.; Rohacs, T.; Balla, T. Rapidly inducible changes in phosphatidylinositol 4,5-bisphosphate levels influence multiple

regulatory functions of the lipid in intact living cells. *J. Cell Biol* **2006** *175*: 377-382.

27. Wymann, M. P.; Marone, R. Phosphoinositide 3-kinase in disease: timing, location, and scaffolding. *Curr. Opin. Cell Biol.***2005** *17*: 141-149

28. Wymann, M. P.; Zvelebil, M.; Laffargue, M. Phosphoinositide 3-kinase signaling – which way to target? *Trends Pharmacol. Sci.* **2003** *24*: 366-376.

29. Erhart, D.; Zimmermann, M.; et al. Chemical Development of Intracellular Protein Heterodimerizers. *Chem. Biol.* **2013** *20*: 549-557.

30. Robinson, M. S.; Sahlender, D. A.; Foster, S. D. Rapid inactivation of proteins by Rapamycin-induced rerouting to mitochondria. *Dev. Cell* **2010** *18(2)*: 324-331.

Chapter 2: Design and Development of a Bio-Orthogonal Chemically Inducible Activator of Ras

I. Introduction

A. Synthetic Biology

Modern molecular biology has developed a wide array of tools with which biological pathways can be dissected. The field of synthetic biology aims to expand this toolkit by engineering novel means to regulate these biological systems[1]. These toolkits are designed with the goal of more precisely determining an individual protein's role in biologically relevant contexts. Such techniques have led to the ability to rewire signaling behavior by manipulating interacting proteins as well as individual protein domains[2,3].

Several approaches to specifically resolve the phenotypic outputs of specific signaling network perturbations currently exist. One such approach was demonstrated in the previous chapter. Briefly, effector proteins can be precisely localized to individual subcellular compartments to control the activity of a specific signaling pathway. However an alternative method, one that directly controls the accessibility of a proteins catalytic core rather than just its spatial localization, would allow for more precise temporal control over these pathways. Such a system has recently been developed and was inspired by the modularity of endogenous regulatory domains.

The activity of endogenous effector proteins is tightly regulated. A common theme by which this regulation occurs is through the use of modular allosteric domains. These domains repress the catalytic activity of these effectors

by a combination of *intra*-molecular and *inter*-molecular auto-inhibitory interactions. Individual regulatory domains often appear in a diverse array of effectors, thus acting as modular repressor domains. An interesting approach to engineering signaling pathways is to develop new modular repressor domains that could be applied to effectors of interest. By replacing endogenous regulatory domains with non-native domains, the specific roles of individual protein domains can be elucidated and new cellular behaviors can be developed. Such a system would allow for the catalytic activity of a protein and the subsequent phenotype resulting from this activity to be decoupled from the protein's regulatory, scaffolding, and localization domains[3].

While this approach has clear advantages for a number of applications, the generation of such modular systems is difficult. A major goal of recent years has been to develop a set of artificial regulatory domains to auto-inhibit a protein of interest that can then be subsequently disrupted by the addition of a cell permeable small molecule. Numerous small-molecule gated switches have been developed that focus on co-localization of domains of interest, yet the development of modular regulatory domains has been lacking. Such a system would maintain the control over spatial localization of proteins, with the temporal resolution and dose-dependency encoded by small molecules.

B. GTPases – Hubs of Cellular Regulation

The Ras superfamily of small GTPases play central roles in cellular signaling. These proteins play roles in cell proliferation, cell morphology, cell motility, nuclear and vesicle transport, and a variety of other critical cellular processes. These enzymes paradoxically act as central signaling nodes, governing a diverse array of signaling pathways, while functioning as a simple binary switch. Ras GTPases function as molecular switches by cycling between an inactive GDP-bound and an active GTP-bound state[4-6] (Figure 1). Cycling between these two states is regulated by a two opposing enzymatic activities: Guanine Nucleotide Exchange Factors (GEFs) and GTPase Activation Proteins (GAPs)[5]. GEFs are a diverse class of proteins, with more than 80 being well characterized[6]. Upon recruitment and activation, GEFs bind to a GTPase and

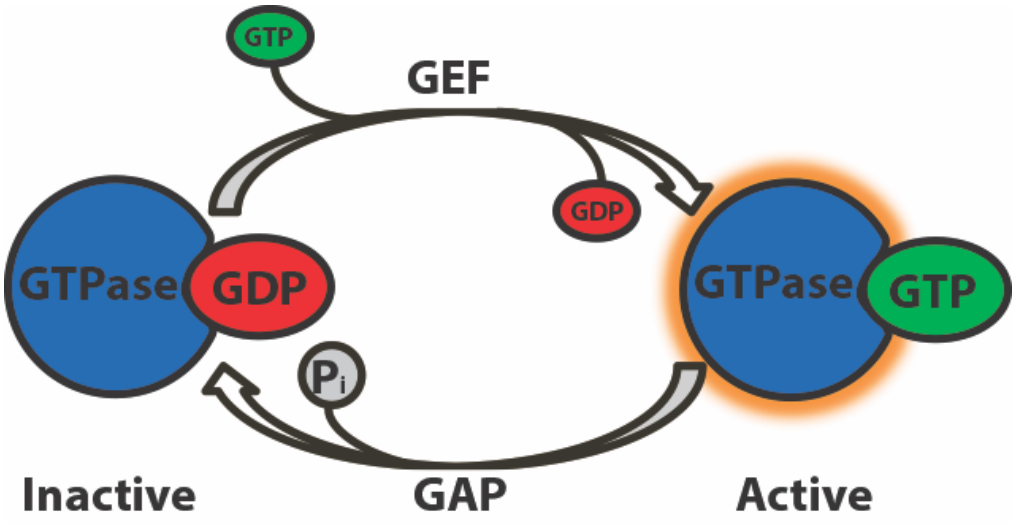


Figure 1: Regulation of GTPase activation: GTPase activity is dictated by whether GTP or GDP is bound. GTPase activation is a tightly regulated process. Activation is initiated by a GEF binding to the GTPase, inducing a conformational change that releases GDP, allowing the more abundant GTP to bind. Deactivation of the GTPase is catalyzed by GAP interactions, accelerating the hydrolysis of GTP to GDP.

catalyze the exchange of GDP for the more abundant GTP, resulting in GTPase activation.

For decades, a major focus in cancer research has been the GTPase Ras. Ras is a central signaling node, specifically involved in numerous cellular pathways. A canonical example of Ras activity is its role in the Epidermal Growth Factor Receptor (EGFR) pathway, in which binding of an extra-cellular growth factor (e.g. EGF), results in the intra-cellular phosphorylation receptor tyrosine kinases. These phosphorylation events recruit a series of signaling proteins that lead to the activation of Ras. The formation of Ras•GTP in turn activates the Raf-Mek-Erk MAPK cascade[7]. Ras has been shown to be the most frequently mutated oncogene and has been shown to govern often-conflicting cellular processes and phenotypes[9]. To this end, the model of Ras acting as a simple on/off switch seems insufficient to explain the complexity of behaviors its activation results in. One explanation for how Ras is capable of achieving such diverse outputs is the specific sub-cellular environment it is activated in. In addition, the amplitude and duration of Ras activation may also play a role in downstream signaling and phenotypic effects.

Classic molecular biology approaches are incapable providing temporal, spatial and rheostatic control over Ras activation, which limits their ability to determine the individual contributions of these factors. Dissection of Ras-mediated pathways would allow for better targeting in drug design as well as more personalized targeting with current therapeutics.

C. Bcl-xL and BH3 –A Novel Artificial Regulatory Domain System

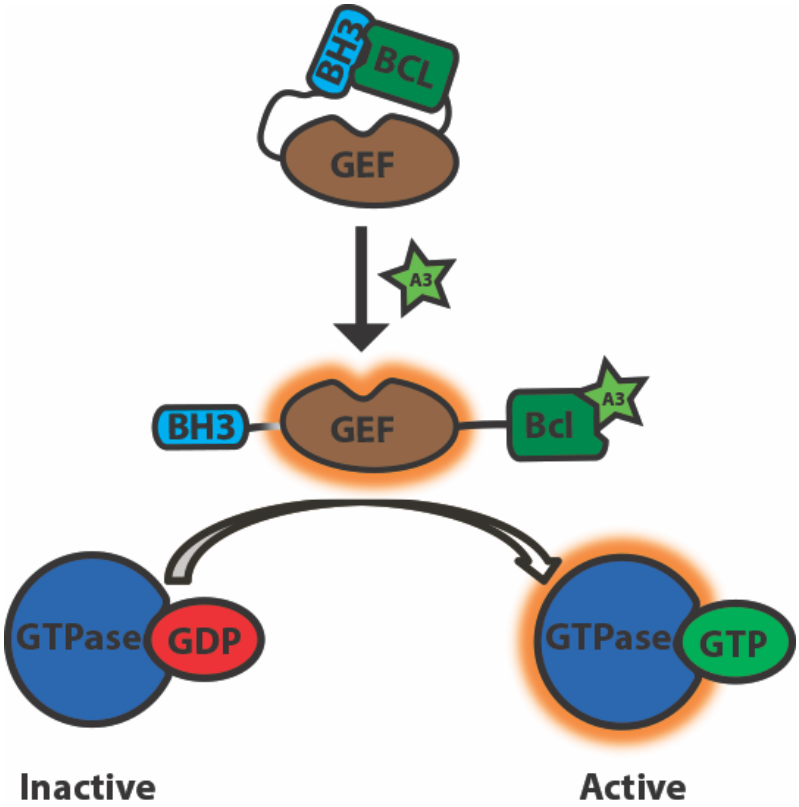


Figure 2: Bcl-xL based auto-inhibition of GEFs: Schematic representation of the strategy used to intra-molecularly gate GEFs. Genetic fusion of Bcl-xL and BH3 to opposite termini of the GEF along with appropriate linker lengths allows for an intra-molecular complex to form that sterically occludes the active site of the GEF.

Our lab has previously generated a general system for controlling the activity of intra-cellular protein activities with cell permeable small molecules[10,11]. The small-molecule regulated artificial regulatory domain system we developed was designed by flanking a core catalytic domain with two domains that form an intra-molecular interaction. The two flanking domains form a complex that sterically occludes the binding site of the core catalytic module. This intra-molecular complex can then be disrupted by the addition of a small molecule that displaces one binding partner, thus exposing the active site of the core catalytic domain (Figure 2).

The system developed in our lab utilizes the apoptotic regulatory proteins Bcl-xL and BH3. Bcl-xL is a member of the Bcl-2 family of anti-apoptotic proteins that can be reduced to a single, 25 kDa binding domain. This protein is largely α-helical with a globular fold that forms a hydrophobic binding groove capable of binding the α-helix of the pro-apoptotic BH3 family of proteins. Previous work has demonstrated that these BH3 family

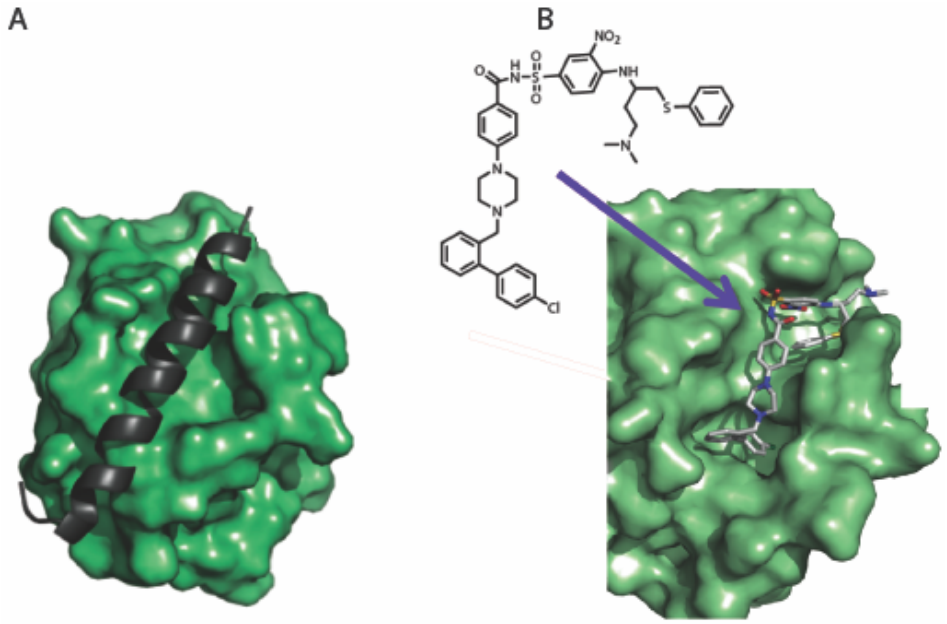


Figure 3: Crystal structures of Bcl-xL bound to a BH3 peptide and ABT-737. A) Crystal structure of Bcl-xL bound to the BH3 peptide 'Bad'. **B)** Crystal structure of Bcl-xL bound to the drug ABT-737, which binds in a similar fashion to A3. The binding groove of the BH3 peptide is the same as the binding groove of ABT-737.

proteins can be reduced to a single α-helical peptide. These peptides retain low nano-molar affinity for Bcl-2 family members. Due to the role that Bcl-2 proteins play in regulating apoptosis, they have been targets of drug development for some time. A drug has been developed and clinically approved that is capable of binding Bcl-2 and out competing BH3 family proteins (Figure 3). Our lab initially

applied this system to gating the activity of a series of guanine nucleotide exchange factors (GEFs), and subsequently to CRE recombinase and Cas9.

D. Chemically Inducible Activator of Ras.

Our lab has recently applied the Bcl-xL/BH3 intra-molecular switch to control the activity of a GEF that activates the GTPase Ras. Evidence suggests that the specific subcellular location of Ras plays a critical role in dictating its downstream activity[8]. Ras has been shown to be present in almost every cellular compartment, including the Golgi, ER, and even the nucleus, and specific Ras isoforms show differential subcellular distributions[7]. Due to the small size of Ras and the fact that a large percentage of its surface area interacts with effector proteins, it is unsuitable for engineering efforts. In addition, over-expression of engineer Ras constructs can perturb their subcellular distribution. To this end, our lab applied the Bcl-xL/BH3 system instead to a direct activator of Ras, the immediately upstream GEF, Son of Sevenless (SOS).

Endogenous SOS is an approximately 152 kDa protein composed of several individual domains. Efforts to directly activate Ras in a SOS induced manner have shown that SOS can be truncated to a minimal, constitutively active catalytic domain (SOScat) of approximately 55 kDa. Computationally guided generation of a fusion that contains the Bcl-xL and the BH3 peptide artificial regulatory domains led to an auto-inhibited SOScat construct. This construct was localized to the plasma membrane via an H-Ras CaaX-motif. Numerous localization tags are available, yet the CaaX of H-Ras was chosen due its well characterized use in localizing SOScat to the plasma membrane. Addition of the

Bcl-xL binding drug A-385358 (A3) to cells expressing this auto-inhibited SOScat construct led to disruption of the Bcl-xL/BH3 peptide complex and the rapid formation of Ras•GTP.

This Chemically Inducible Activator of Ras (CIAR) allows for rapid, dose-dependent activation of Ras signaling. CIAR functions as a Ras pathway rheostat, allowing for interrogation of Ras signaling at a variety of subcellular compartments. CIAR was localized to the plasma membrane via an H-Ras CaaX motif, but the CaaX motif can be varied and the roles of Ras in different sub-cellular environments can be probed. While CIAR is a powerful new tool for probing a specific signaling pathway, it does utilize endogenous mammalian proteins. A3 targets endogenous Bcl-xL, which can lead to confounding effects in some cellular environments. While controls have shown that A3 alone is not responsible for the phenotypes observed by activation CIAR[11], an ideal system would not target mammalian proteins. Such a bio-orthogonal system could not only be used to replace CIAR, but it could also be applied to other GEFs (or other proteins of interest) and used in combination. Such a tool would allow for unprecedented spatio-temporal control over signaling networks. Therefore, we explored the use of the HCV-CDD described in Chapter 1 for the generation of a Bio-Orthogonal Chemically Inducible Activator of Ras (O-CIAR)

2. Results and Discussion

A. Designing an HCV-Based Chemically Inducible Activator of Ras (O-CIAR)

The guanine nucleotide exchange factor (GEF) Son Of Sevenless (SOS) was selected as the first target for the application of HCV-CDD as an intra-molecular regulatory switch. The success of the Bcl-xL/BH3 based CIAR switch[11] provided a clear set of guidelines and benchmarks for the development of an orthogonal activator. The GEF SOS can be reduced to a core, constitutively active catalytic subunit comprising the cdc25 and REM domain (SOScat, res: 574-1020)[12]. Initial demonstrations of CIAR showed that the SOScat domain could be auto-inhibited through steric occlusion of the active site by application of an intra-molecular protein-protein interaction complex. The success of an artificial regulatory switch, composed of Bcl-xL and BH3, implies that a protein-peptide interaction of this size is capable of auto-inhibiting SOScat. The NS3-4A protease and the Cp5 peptide are of comparable size and topology to Bcl-xL and BH3 respectively.

Artificial regulatory switches not only must be of high enough affinity that they form a tight intra-molecular complex, but they must be capable of occluding their target's active. In the context of CIAR, Bcl-xL and BH3 flank the SOScat domain and are joined by a set of flexible linkers composed of glycines, serines, and threonines. The orientation of artificial regulatory domains and the length of these linkers dictate the degree to which the protein-peptide interaction is able to occlude the active site. The sheer number of permutations in N- and C-terminal

linker lengths possible is enormous and as such is prohibitive to experimentally screen.

To simplify the problem, we used RosettaRemodel[13] as a computational framework for determining linker variants that are likely to satisfy three important criteria. First, that the linkers are of sufficient length to allow the formation of an intra-molecular NS3-4A and Cp5 complex. Second, NS3-4A/Cp5 complex formation has an energetic local minima that occludes the active site of SOScat. Third, that the energetic landscape for complex formation is sufficiently 'steep' so as to provide a large enough energetic penalty for the artificial regulatory domain moving away the SOScat active site. Within this design framework, we set the NS3-4A/Cp5 complex as a rigid-body inside a loop that bridges the termini of SOScat. This allows for the simplification of the computational problem to that only concerned with loop closure. This takes into account only the orientation of NS3-4A/Cp5 and the linker lengths separating the regulatory domains from SOScat. The low-energy ensemble resulting from each attempt at chain closure was demonstrated previously to correlate with the spatial localization of the proteins involved.

Initial characterization of the Bcl-xL/BH3 CIAR system demonstrated that RosettaRemodel localization experiments could be quantified by the localization parameter μ-mean[11]. This parameter corresponds to the difference between the average center-of-mass (CoM) of Ras bound to the active site of SOScat and the CoM of the NS3-4A/Cp5 complex. This parameter, coupled with its standard deviation, allows for computational screening of linker variants. The similarity in

size and topology between the NS3-4A/Cp5 and Bcl-xL/BH3 complexes led us to focus on a narrow set of linker variants to decrease the computational simulation time necessary. Constructs were screened with N- and C-terminal NS3-4A with Cp5 at the opposing terminus (Figure 4A-D, Table 1).

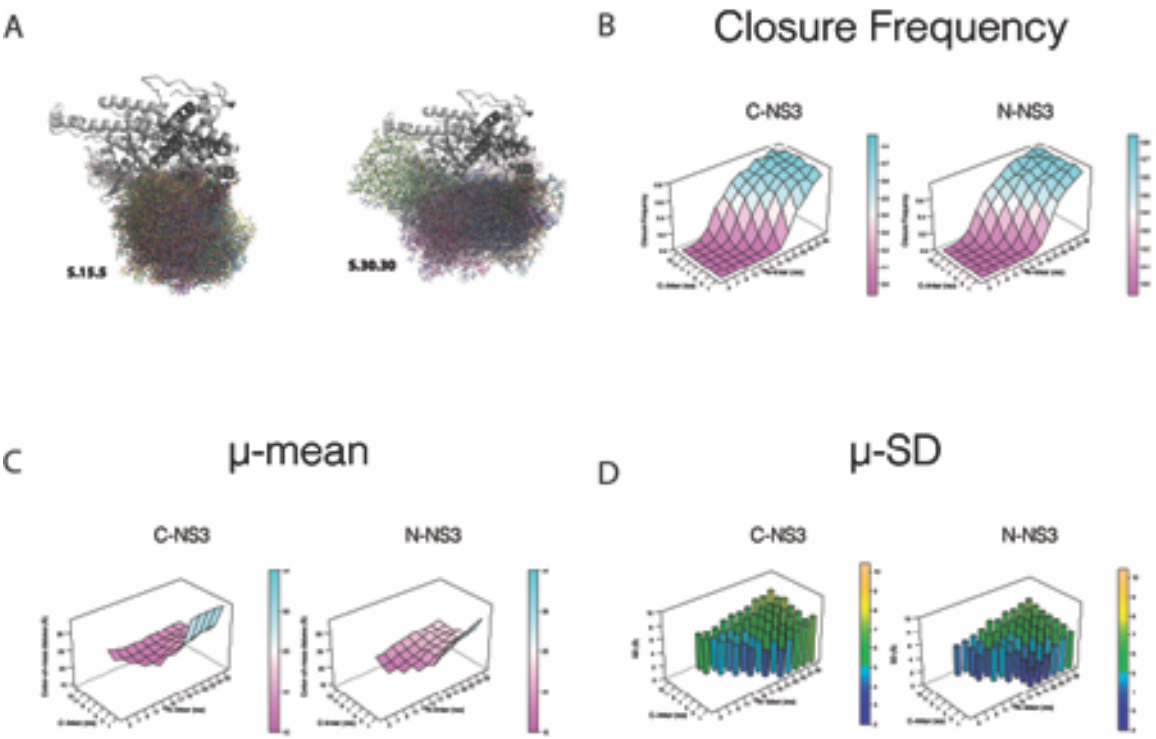


Figure 4: Computational modeling of bio-orthogonal CIAR (O-CIAR). RosettaRemodel was used to predict the likely location an intra-molecular NS3-4A/Cp5 complex. Each modeling trajectory tries to result in a chain closure. One thousand trajectories per linker variant were run, and successful closures were overlaid and analyzed in PyMol. **A)** The density of overlaid trajectories over the Ras binding site are indicative of the likelihood of successful SOScat auto-inhibition. **B)** The frequencies of successful closures were plotted and trajectories that did not result in successful chain closures were discarded. **C)** The average COM distance was plotted as *μ*-mean. **D)** The standard deviation of the calculate *μ*-mean were plotted.

C-NS3				N-NS3			
Linkers	μ-mean	SD	Clos.Frq	Linkers	μ-mean	SD	Clos.Frq
15-5	*17.7502*	*3.9056*	*0.02*	**15-5**	*14.2719*	*2.6690*	*0.018*
15-9	*18.5625*	*4.3824*	*0.265*	**13-7**	*15.0124*	*3.3320*	*0.01*
17-7	*18.6433*	*4.0772*	*0.309*	**11-9**	*15.4144*	*6.5933*	*0.002*
15-7	*18.9188*	*4.4453*	*0.106*	**13-9**	*16.5638*	*3.5832*	*0.076*
17-5	*19.1694*	*3.9838*	*0.14*	**15-7**	*16.8362*	*4.6966*	*0.095*
17-9	*19.2955*	*4.9010*	*0.49*	**15-9**	*16.9553*	*4.0129*	*0.249*
19-9	*19.3501*	*4.9038*	*0.576*	**13-11**	*17.3030*	*4.0944*	*0.221*
19-7	*19.4044*	*4.6649*	*0.5*	**11-11**	*17.3080*	*3.1591*	*0.039*
13-11	*19.4370*	*4.7682*	*0.201*	**17-5**	*17.5187*	*4.010*	*0.137*

Table 1: Summary of values obtained from RosettaRemodel modeling of O-CIAR. Linker variants are shown as numbers (#-#) pertaining to the number of amino acids (G, S, or T) in the N-terminal linker followed by the C-terminal linker.

The *μ*-mean, standard deviation, and closure frequencies show an energy landscape that is less conclusive than that for Bcl-xL/BH3. Based on the similarity between NS3-4A and Bcl-xL in size and topology, the C-terminal NS3-4A constructs were chosen for initial experimental characterization. Additionally, several Cp5 variants exist covering a 100x range in affinities. A critical component of these synthetic regulatory switches is that the drug must be able to disrupt the protein-peptide interaction. To this end, we chose to initially start using the Cp5 variant of lowest affinity.

B. Experimental Verification of HCV-CIAR Construct Functionality

Computational modeling predicted that for constructs with C-Terminal NS3-4A, an ideal linker set would have 17 amino acids between Cp5 and

SOScat and 7 amino acids between SOScat and NS3-4A. A small panel of HCV-CIAR constructs was generated in a mammalian expression vector containing linker set of 17:3, 17:5, 17:7, and 17:9 (Figure 5). Modeling suggested that C-terminal linker variations would be better tolerated (Figure 4C). An additional construct was generated replacing Cp5 with BH3. By replacing Cp5 with a non-binding peptide, the active site of SOScat is exposed and thus constitutively active as a positive control. Linkers were designed to minimize secondary structure by alternating glycine residues between serines and threonines.

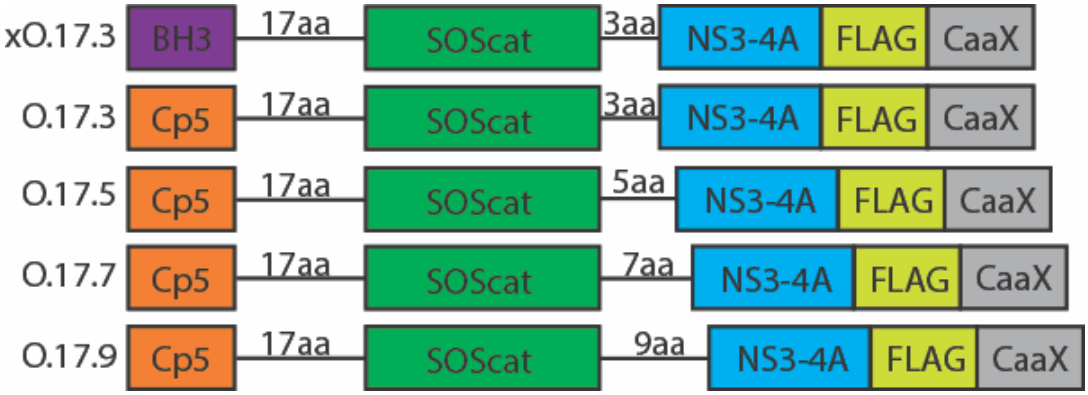


Figure 5: Bio-orthogonal CIAR constructs tested: Modeling suggested an optimal linker set of 17- and 7- amino acids at the N- and C-termini respectively. Bio-Orthogonal CIAR (O-CIAR) constructs were generated that contain subtle variants in their C-terminal linker lengths.

HCV-CIAR constructs characterization must be verified in a cellular context. The CIAR switch has previously been demonstrated to functionally gate SOScat activity in a biochemical context, thus a cellular assay would allow for rapid determination of an optimal linker variant while characterizing HCV-CIAR's exact cellular behavior. Constructs were evaluated using a cellular reporter assay based on measuring the production of phosphorylated ERK, a downstream

product of plasma membrane induced Ras activation. Variants were compared by determining the degree of phosphorylated ERK (pERK) generated upon addition of drug and comparing that to a DMSO treated control. In HEK293T cells, constructs were transiently transfected and cells were serum starved for 18 hours. Cells were subsequently treated with 10 μM Asunaprevir or DMSO for 60 minutes. Cells were harvested and the levels of ERK and pERK were determined by western blot (Figure 6A).

HCV-CIAR constructs exhibit an interesting linker-length dependent pERK activation relationship. A constitutively active control construct that contains BH3 instead of Cp5 (xO.17.3), results in a strong pERK signal in a drug independent manner. Surprisingly, an increase in pERK signal is observed in the drug treated samples as the C-terminal linker length increases. Initial modeling suggested that, if the linkers were too short to allow for intra-molecular closure, constructs would dimerize or oligomerize in an *inter*-molecular fashion. We hypothesized that this would result in a signal similar to that of xO.17.3. The data suggest, however, that such constructs fail to initiate activation of Ras regardless of drug addition. Potential reasons for this observed behavior could include formation of aggregates that, while soluble, sequester SOScat in such a manner that its active site is inaccessible to Ras.

As the linker length is increased to O.17.5 and O.17.7, a drug dependent increase in pERK signal is readily observed, with O.17.7 exhibiting the largest increase in total pERK signal. Interestingly, the fold-increase between DMSO

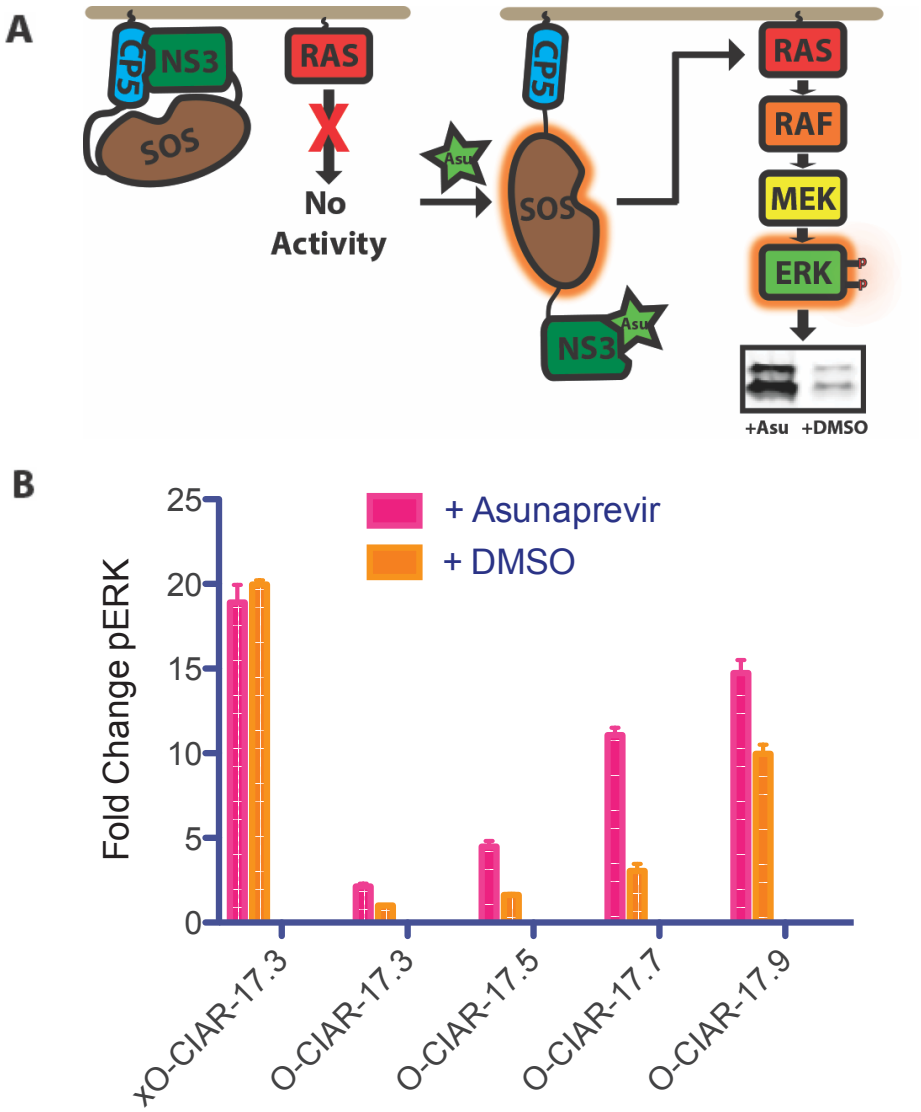


Figure 6: Comparison of O-CIAR linker variants: A) O-CIAR constructs were tested using a simple cellular pERK reporter assay. Addition of drug disrupts the NS3-4A/Cp5 interaction, exposing SOScat's active site. This activates the Ras cascade, resulting in the phosphorylation of ERK in proportion to the level of activated Ras. **B)** The fold change for the increase in pERK signal for both DMSO and drug treated cells was measured.

treated and drug treated pERK signal for O.17.5 and O.17.7 are similar, but the total pERK signal for DMSO or drug treatment varies. The background pERK signal in DMSO treated cells implies that the closed NS3-4A/Cp5 complex incompletely occludes SOScat's active site. Coinciding closely with the

computational model predictions, O.17.9 exhibits strong pERK signal but an almost equal signal in the absence of drug. This implies that the linkers are long enough to form the NS3-4A/Cp5 complex, but that the complex is able to vacate the SOScat active site. This exposes the active site of SOScat and allows for the binding of Ras.

C. O-CIAR Functions as a Ras Pathway Rheostat

Functionally demonstrating that O-CIAR is capable of replacing CIAR as a Ras pathway rheostat necessitates that O-CIAR is capable regulating the amplitude of flux through the pathway of interest. What this requires is that O-CIAR has a strong dose-dependent relationship between drug concentration and the resulting pERK signal. By titrating the amount of drug, the amount of active SOScat and, thus the intra-cellular concentration of active Ras, varies. To this end, this rheostat-like behavior was verified by assaying Flp-In T-REx HEK293T cells stably expressing O.17.7, serum starving for 18 hours, and treating with various drug concentrations. The O-CIAR expressing cells were treated with 20 μM, 2.2 μM, 250 nM, 25 nM, or DMSO for 60 minutes (Figure 7). Predictably, O-CIAR exhibits a clear dose-dependent relationship with Asunaprevir, demonstrating a functional rheostat-like behavior.

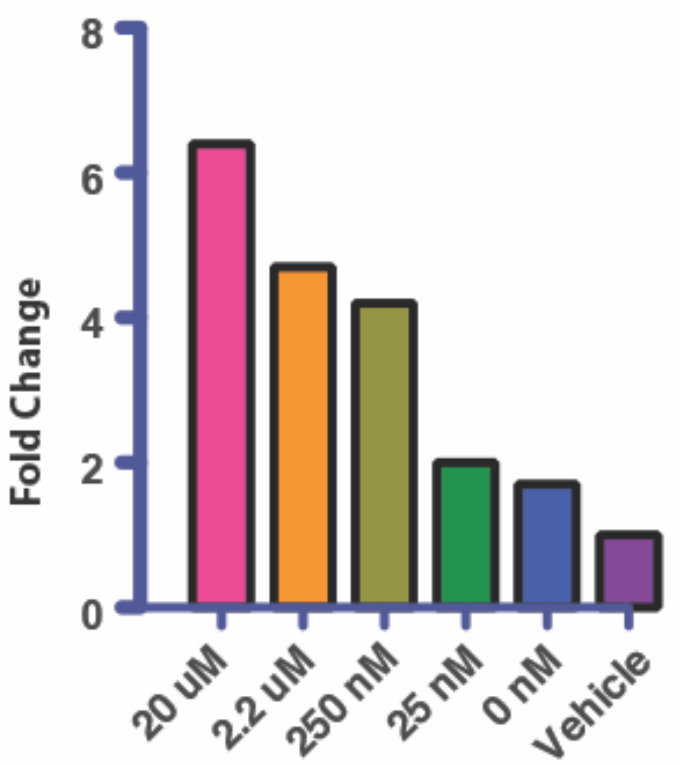


Figure 7: Dose-dependent activation of O-CIAR. T-REx 293T cells stably expressing O-CIAR-17.7 were treated with varying concentrations of Asunaprevir. A strong dose-dependent pERK activation profile is observed.

D: Utilizing the HCV-CDD Tool Kit: Optimization of O-CIAR

Initial characterization demonstrated that an N-terminal linker of 17 amino acids and a C-terminal linker of 7 amino acids gives the largest dose-dependent increase in pERK signal. However, the DMSO treated sample gave a pERK signal 3-fold greater than untransfected cells. To lower this background activation, we systematically increased the affinity of the Cp5 peptide by introducing individual mutations[14]. Converting residue 4 from Gly to Asp and residue 5 from Arg to Glu in the Cp5 peptide results in a 6-fold increase in NS3-4A affinity. O.17.7 constructs containing GΔD, RΔE, or both GRΔDE mutations were generated. Constructs were assayed using the same transient transfection assays in HEK293T cells (Figure 8).

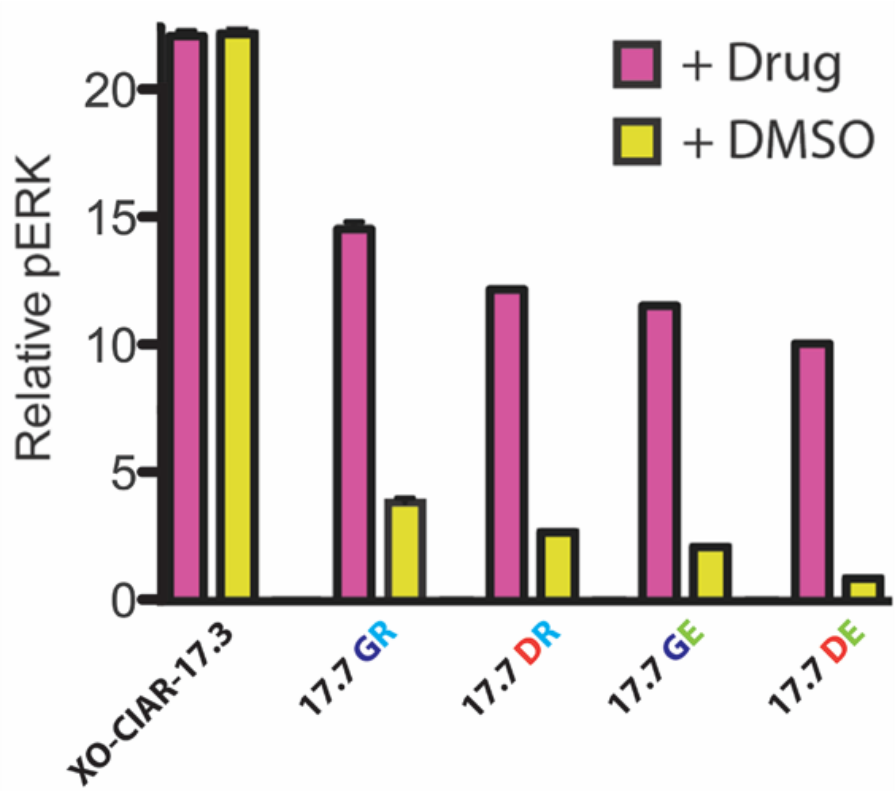


Figure 8: Optimization of O-CIAR with higher affinity Cp5 variants. Increasing the affinity of the Cp5 peptide decreases the maximum pERK signal observed while decreasing background pERK.

The Cp5 variants assayed show a clear relationship between the affinity of the peptide for NS3-4A, and the background pERK signal observed. Interestingly, while higher affinity variants decrease pERK signal in DMSO treated cells, they also decrease the maximum pERK signal in drug treated cells, but to a lesser extent. The highest affinity peptide variant assayed (O.17.7-DE) showed the highest pERK signal with the lowest background signal. The variants with moderate affinity increases show, a moderate decrease in background.

E. O-CIAR Temporally Regulates Ras Pathway Signaling

The O.17.7-DE construct has now been shown to function as a spatially constrained Ras pathway rheostat. However, a rheostat controls not only the amplitude of the flux being directed through the pathway, but also the duration of the signal. To characterize the temporal nature of Ras at the plasma membrane as dictated by SOScat-directed activation, a simple time-course experiment using O.17.7-DE was performed. Flp-In T-REx 293 cells stably expressing O.17.7-DE were generated to precisely assay for time-dependent pERK signal. These cells were serum starved for 18 hours and subsequently treated with 10 μM Asunaprevir or DMSO for 10, 20, or 60 minutes and compared to cells not expressing a construct (Figure 9). The time-dependent increase and plateau in

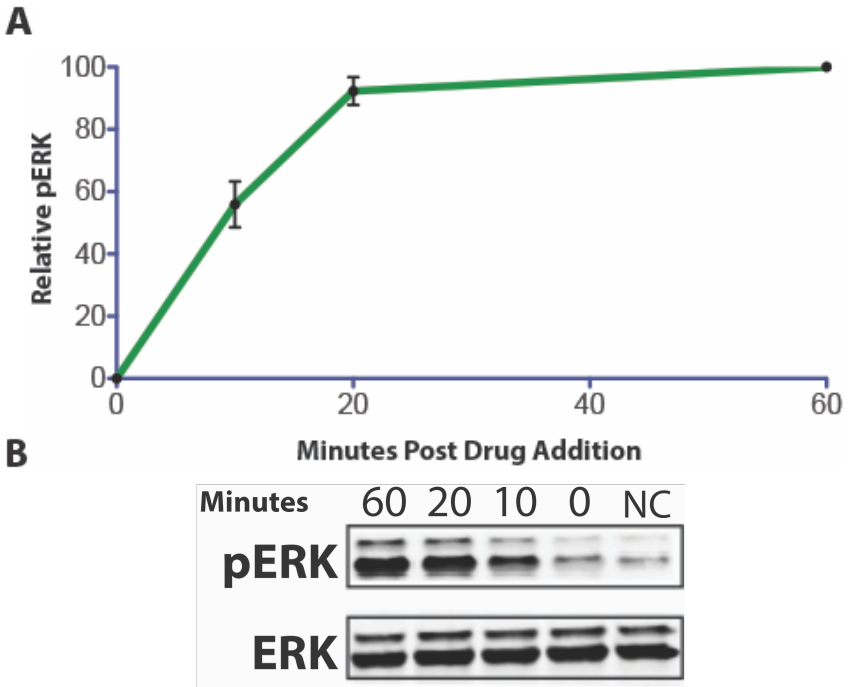


Figure 9: Temporal regulation of Ras activity by O-CIAR: Activation of O-CIAR by Asunaprevir leads to Ras induced ERK phosphorylation in a time dependent manner. Maximum activation of pERK signaling closely matches Bcl-xL regulated CIAR.

pERK signal in these cells closely resembles that observed in the Bcl-xL/BH3 CIAR system. This activation is in opposition to that observed in EGF treated cells. The sustained activation of the Ras pathway is indicative of a lack of a negative feedback loop that is capable of shutting down this pathway, one that would be induced by the recruitment of other factors associated with trans-membrane receptor activation.

III. Conclusion

The HCV-CDD system comprising NS3-4A and the Cp5 peptide has now been shown to behave as an artificial regulatory domain. We have accomplished this be demonstrating the capability of this system to function in an inter-molecular fashion (Chapter 1) and now as an intra-molecular switch. This was carried out using the previously developed Chemically Inducible Activator of Ras system. Our O-CIAR construct is capable of acting as a bio-orthogonal Ras pathway rheostat. This construct can control Ras pathway activation in a spatio-temporal fashion. O-CIAR was localized to the plasma membrane to illustrate spatial control, and the ability to control flux through the Ras pathway was shown via dose-dependent control of the signal amplitude and temporal regulation of ERK phosphorylation.

Importantly, the components in the HCV-CDD switch cover a wide range of affinities. This range of affinities provides a sort of toolkit with which future researchers can tune the binding affinities of the protein/peptide complex components. The utility of this toolkit idea was demonstrated by assaying Cp5

affinity variants to find the optimal balance between maximum signal and background activation. This balance will be different depending on the protein that is being gated. This versatility is a powerful tool for enabling future applications.

IV. Materials and Methods:

1. *Designing an HCV-Based Chemically Inducible Activator of Ras*

A. Conformational Sampling by RosettaRemodel

The NS3-4A-1b/Cp5 complex (PDB 4A1X) was modeled using the same protocol as Bcl-xL/BH3 in the development of CIAR (Rose, J. R, 2017) . Briefly, the auto-inhibitory complex was treated as a single, rigid-body between the N- and C- termini of the SOScat model (residues 574-1,020, PDB 1XD2). This allows for the treatment of the linker lengths as a circular permutation design thus allowing for insertion of the NS3-4A-1b/Cp5 complex across the termini and solving for loop closure. Linker lengths were chosen based on the previously determined optimal Bcl-xL/BH3 CIAR design with variations made in both sets of linkers to expand the analyzed space. 1,000 independent trajectories were sampled in 50 parallel runs. Specific computational methods and protocols were defined previously[11].

2. *Characterization and Optimization of O-CIAR Designs*

A. Eukaryotic Expression Vector Cloning General Methods

CLONING OF O-CIAR-17.3

The NS3-4A-1b construct was obtained as a human codon-optimized double stranded DNA G-Block (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). The SOScat gene containing the linker set 17/3 was cloned out from human codon-optimized Bcl-xL/BH3 CIAR. Cp5 was obtained as a human codon-optimized double stranded DNA G-Block containing Gibson Assembly overhangs designed in NEBuilder. Genes were subcloned into a pcDNA5/FRT/TO vector containing a C-terminal FLAG tag and H-Ras CaaX motif.

CLONING OF O-CIAR LINKER VARIANTS

Linker variants were generated using QuikChange Mutagenesis of the O-CIAR.17.3 plasmid.

CLONING OF xO-CIAR.17.3

xO-CIAR.17.3 was generated by Gibson Assembly. The DNA containing the N-terminal 17 amino acid linker, SOScat, C-terminal 3 amino acid linker, and NS3-4A-1b, was obtained by PCR. A pCDNA5/FRT/TO vector containing N-terminal BH3 and C-terminal FLAG-CaaX(H-Ras) was linearized by PCR. The SOScat-NS3 gene was inserted into the linearized vector by Gibson Assembly.

xO-CIAR.17.3

Bad-$(G/S)_{17}$-SOScat-(GSG)-NS3/4A-$(GS)_6$-FLAG-CaaX

O-CIAR.17.3

Cp5-$(G/S)_{17}$-SOScat-(GSG)-NS3/4A-$(GS)_6$-FLAG-CaaX

O-CIAR.17.5

Cp5-$(G/S)_{17}$-SOScat-(GSGTG)-NS3/4A-$(GS)_6$-FLAG-CaaX

O-CIAR.17.7

Cp5-$(G/S)_{17}$-SOScat-(GSGTGSG)-NS3/4A-$(GS)_6$-FLAG-CaaX

O-CIAR.17.9

Cp5-$(G/S)_{17}$-SOScat-(GSGTGSGTG)-NS3/4A-$(GS)_6$-FLAG-CaaX

BH3 Peptide Sequence

APPNLWAAQRYGRELRRMADEGEGSFK

Cp5 Peptide Sequence

GELGRLVYLLDGPGYDPIHSD

FLAG-Tag Sequence

DYKDDDDK

H-Ras CaaX Sequence

QHKLRKLNPPDESGPGCMSCKCVLS

NS3-4A-1b Sequence:

AKGSVVIVGRINLSGDTAYSQQTRGAAGTAATSATGRDKNQVDGEVQVLSTAT
QSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGAR
SMTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLC
PSGHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR

SOScat (574-1020) Sequence

DVYRFAEPDSEENIIFEENMQPKAGIPIIKAGTVIKLIERLTYHMYADPNFVRTFLT
TYRSFCKPQELLSLIIERFEIPEPEPTEADRIAIENGDQPLSAELKRFRKEYIQPVQ
LRVLNVCRHWVEHHFYDFERDAYLLQRMEEFIGTVRGKAMKKWVESITKIIQRK
KIARDNGPGHNITFQSSPPTVEWHISRPGHIETFDLLTLHPIEIARQLTLLESDLY
RAVQPSELVGSVWTKEDKEINSPNLLKMIRHTTNLTLWFEKCIVETENLEERVA
VVSRIIEILQVFQELNNFNGVLEVVSAMNSSPVYRLDHTFEQIPSRQKKILEEAHE
LSEDHYKKYLAKLRSINPPCVPFFGIYLTNILKTEEGNPEVLKRHGKELINFSKRR
KVAEILGEIQQYQNQPYCLRVESDIKRFFENLNPMGNSMEKEFTDYLFNKSLEIE
P

B. HEK-293T Cell Culture and Transient Transfection Conditions

HEK-293T were maintained in DMEM (Gibco, product number 11065092) supplemented with 10% FBS (Gibco, product number A3160602). All transient transfections were done using TurboFectin8.0 (Origene) at a ratio of 3:1 TurboFectin:µg DNA prepared in OptiMem (Gibco, product number 11058021)

16-20 hours after plating of cells. Transfections were allowed to proceed for 24 hours before experiments were performed.

C. Flp-In T-REx HEK-293T Cell Culture and Stable Cell Line Generation

Flp-In T-REx 293-T (Invitrogen, product number R78007) cells were maintained in DMEM (Gibco, product number 11065092) supplemented with 10% FBS (Gibco, product numberA3160602). Dox-inducible stable 293Ts were generated using the Flp-In T-REx system as described in product documentation, and maintained with 50 μg/mL hygromycin and 15 μg/mL blasticidin after selection.

D. O-CIAR Linker Length and Peptide Variant Assays

24 hours prior to transfection, $3.5x10^5$ HEK 293-T cells were plated onto poly-D-lysine 24 well plates. After transfection with the O-CIAR linker variant, cells were allowed to recover for 24 hours. After 24 hours, media was aspirated and the cells were washed with 1 mL of pre-warmed (37 °C) PBS, then serum starved with FBS-free DMEM for 18 hours prior to drug treatment. Media was subsequently aspirated and the cells were washed twice with 1 mL chilled PBS, then lysed with 60 µL Mod. RIPA buffer (50 mM Tris, pH 7.8, 1% IGEPAL CA-630, 150 mM NaCl, 1 mM EDTA, 2 mM Na_3VO_4, 30 mM NaF, Pierce Protease Inhibitor Tablet). Cleared lysates were subjected to SDS-PAGE and transferred to nitrocellulose. Blocking and antibody incubations were done in TBS with 0.1% Tween-20 (v/v) and blocking buffer (Odyssey). Primary antibodies were diluted as follows: Total ERK (1:4000, #9107), phosphorylated ERK (1:2000, #4370), α-

Tubulin (1:10,000, #3873) all from Cell Signaling Technology. Blots were washed in TBS with 0.1% Tween-20. Antibody binding was detected by using near-infrared-dye-conjugated secondary antibodies and visualized on the LI-COR Odyssey scanner. Blots were quantified via densitometry with Image Studio (LI-COR).

E. O-CIAR Stable Cell Dose Dependency and Time Course Screen

$4.0x10^5$ Flp-In T-REx HEK 293-T cells stably expressing O-CIAR.17.7 were plated in complete DMEM media supplemented with 1 µg/mL doxycycline onto poly-D-lysine 24 well plates. After 24 hours, media was aspirated and the cells were washed with 1 mL of pre-warmed (37 °C) PBS, then serum starved with FBS-free DMEM for 18 hours prior to drug treatment. Media was subsequently aspirated and the cells were washed twice with 1 mL chilled PBS, then lysed with 60 µL Mod. RIPA buffer (50 mM Tris, pH 7.8, 1% IGEPAL CA-630, 150 mM NaCl, 1 mM EDTA, 2 mM Na_3VO_4, 30 mM NaF, Pierce Protease Inhibitor Tablet). Cleared lysates were subjected to SDS-PAGE and transferred to nitrocellulose. Blocking and antibody incubations were done in TBS with 0.1% Tween-20 (v/v) and blocking buffer (Odyssey). Primary antibodies were diluted as follows: Total ERK (1:4000, #9107), phosphorylated ERK (1:2000, #4370), α-Tubulin (1:10,000, #3873) all from Cell Signaling Technology. Blots were washed in TBS with 0.1% Tween-20. Antibody binding was detected by using near-infrared-dye-conjugated secondary antibodies and visualized on the LI-COR Odyssey scanner. Blots were quantified via densitometry with Image Studio (LI-COR).

V. References

1. Drubin, D. A.; Way, J. C.; Silver, P. A. Designing biological systems. *Gene Dev.* **2007** *21*: 242-254.

2. Dueber, J. E.; Yeh, B. J.; Bhattacharyya, R. P.; Lim, W. A. Rewiring cell signaling: the logic and plasticitiy of eukaryotic protein ciruitry. *Curr. Opin. Struc. Biol.* **2004** *14*: 690-699.

3. Yeh, B. J.; Rutligliano, R. J.; Deb, A.; Bar-Sagi, D.; Lim, W. A. Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* **2007** *447*: 596-600

4. Rossman, K. L.; Der, C. J.; Sondek, J. GEF means go: turning on Rho GTPases with guanine nucleotide exchange factors. *Nat. Rev. Mol. Cell Bio*. **2005** *6*: 167-180.

5. Fritz, G.; Kaina, B. Rho GTPases: promising cellular targets for novel anticancer drugs. *Curr. Cancer Drug Targets* **2006** *6*: 1-14

6. Bos, J. L.; Rehmann, H.; Wittinghofer, A. GEFs and GAPs: critical elements in the control of small G proteins. *Cell* **2007** *129*: 865-877.

7. Iversen, L.; Ty, H. L.; Lin, W. C. et al. Ras activation by SOS: Allosteric regulation by altered fluctuation dynamics. *Science* **2014** *345*: 50-54.

8. Omerovic, J.; Hammond, D. E.; Clague, M. J.; Prior, I. A. Ras isoform abundance and signaling in human cancer cell lines. *Oncogene* **2008** *27*: 2754-2762.

9. Karnoub, A. E.; Weinberg, R. A. Ras oncogenes: split personalities. *Nat. Rev. Mol. Cell Biol.* **2008** *9*: 517-531.

10. Goreshnik, I.; Maly, D. J. A small molecule-regulated guanine nucleotide exchange factor. *J. Am. Chem. Soc.* **2009** *132*: 938-940.

11. Rose, J. C.; Huang, P. S.; et al. A computationally engineered RAS rheostat reveals RAS-ERK signaling dynamics. *Nat. Chem. Biol.* **2017** *13*: 119-126.

12. Boriack-Sjodin, P. A.; Margarit, S. M.; Bar-Sagi, D.; Kuriyan, J. The structural basis of the activation of Ras by SOS. *Nature* **1998** *394*: 337-343.

13. Huang, P. S.; et al. RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **2011** *6*: e24109.

14. Kugler, J. et al. High affinity peptide inhibitors of the Hepatitis C Virus NS3-4A protease refractory to common resistant mutants. *J. Biol. Chem.* **2012**. *287*: 39224-39232.

Chapter 3: Small-Molecule Gated Protein Inhibitor Scaffolds

I. Introduction

A. Expanding the Toolkit for Synthetic Biology

The current toolkit for probing and engineering cellular signaling pathways has now been expanded to include small-molecule gated activators of specific pathways. Application of this system to intra-molecularly gate proteins of interest (POI) provides a powerful tool for probing that pathway. However, this approach requires that design and optimization be carried out for each POI. Topological constraints may prevent many proteins from being amenable to an intra-molecular switch thus limiting the ability of this strategy to be generally applied throughout the proteome. Additionally, utilizing inter-molecular switches to activate or repress protein function in a localization dependent manner relies on using POI's whose activity is regulated solely by localization. Generalizable approaches for targeting proteins of interest would assist in overcoming these hurdles.

In a truly generalizable approach for regulating endogenous proteins of interest it would be highly desirable to apply small-molecule gated regulatory domains to a single protein scaffold. If the scaffold backbone and topology could be kept constant while engineering the binding site for selectivity, the design component of applying artificial regulatory domains would only need to be done once. Such protein inhibitor scaffolds now exist as new classes of non-immunoglobulin scaffolds have been identified with *in vitro* display technologies. These scaffolds coupled with the *in vitro* selection methodologies allow for the

rapid identification of potent and selective protein-based inhibitors suitable for intracellular studies[1-3]

Development of these scaffolds have come from two distinct approaches. One started by using an inert scaffold based on a known protein binding motif. By engineering these motifs into a single repeat protein, the binding surface can be varied for select for proteins of interest. Proteins in this category include single chain antibody-like proteins called monobodies[1], as well as a new class of proteins called Designed Ankyrin Repeat Proteins (DARPin)[3-6]. The second approach started with a protein substrate of a class of proteins of interest, and improving the protein substrate's affinity for individual members of that class. The protein scaffold they used was the protein ubiquitin, and the inhibitors generated from this scaffold termed Ubiquitin Variants (Ubv)[2,7.]

B. DARPins: A Unique Class Of Antibody Mimetic Proteins.

DARPins are based off the highly conserved and widely used Ankyrin repeat domain. Ankyrin repeat proteins are present in almost all forms of life and serve as protein binding domains both intra- and extra-cellularly[8]. They contain highly conserved N- and C-terminal domains with varying numbers of Ankyrin domain repeats between these two terminal 'capping' domains. The majority of DARPins currently in use contain between two and four repeats between the capping domains and are labeled according to this number (N2C, N3C, N4C).

Each Ankyrin Repeat has a binding surface composed of a highly variable loop and the solvent exposed portion on an α-helix[9]. DARPins were adapted from the naturally occurring Ankyrin repeat domains to contain a specific

set of seven residues per repeat that are varied during the in-vitro selection process. The remaining portion of the DARPin scaffold was optimized for biophysical behavior, specifically for stability and solubility as well as eliminating all native cysteines to ensure that they are not reduced in the mammalian cytosol. DARPins have been generated as inhibitors of several distinct enzyme families and have shown incredible selectivity and potency with K_D's in the high picomolar to low nanomolar range[10].

C. Ubv's: A Novel Approach For Selectively Targeting De-Ubiquitinases.

Ubiquitin is a 76 amino acid protein that is present in all eukaryotes. The role of Ubiquitin in the cell is that of a post-translational modification (PTM) and is thus used in a diverse set of signaling pathways. Like protein phosphorylation, the location and degree to which a protein is ubiquitinated can dictate how that PTM is recognized[11]. However, unlike protein phosphorylation, ubiquitin is capable of being further modified by additional ubiquitin units. Ubiquitin modifications can thus be further diversified through building both linear and branched chains. The extent to which a protein is ubiquitinylated dictates the specific signaling output. Possible outcomes of the specific ubiquitin modification on a protein can target it for degradation via the proteasome, localize it to individual cellular compartments[12], allosterically modulate enzyme activity[13], or affect specific protein-protein interactions[14].

The process of ubiquitylating a target protein is carried out by a three-step mechanism involving activation, conjugation and ligation by E1-activators, E2-conjugators, and E3-ligators respectively[15]. The diversity in targets and types of

ubiquitin modifications requires a high degree of specificity in target recognition. This specificity is obtained through the increase in the diversity of proteins in each subsequent step of the ubiquitin conjugation process. The human genome encodes two E1 activators, 30 E2 conjugators, and more than 600 E3 ligators[16].

Just as regulation of protein phosphorylation is reversed by phosphatases, ubiquitin modifications are tightly controlled and are removed by their own class of proteins. The family of proteases responsible for the removal of ubiquitin moieties are termed De-Ubiquitinases (DUBs). There are approximately 100 known human DUBs, however their individual roles are quite poorly understood[17]. DUBs have proven difficult to study in the context of cellular pathways as there only exist a few DUB inhibitors. Additionally, they show poor potency and very little selectivity. Knockout studies have yielded some useful information but many DUB knockouts exhibit embryonic lethality in mice[18].

Recently, Ernst and coworkers developed a methodology to use Ubiquitin as a scaffold for developing potent and selective inhibitors of ubiquitin regulatory pathways[2]. The authors utilized the crystal structures of ubiquitin bound to individual DUBs to map the important contact residues. From this, they identified 33 residues that consistently make contacts and appear to dictate specificity. These were then varied using an in-vitro selection method, specifically phage display, and numerous highly selective and potent inhibitors were discovered. By increasing the affinity for a specific DUB, overexpression of a Ubv will bind and outcompete endogenous ubiquitin, thereby function as an inhibitor. These Ubv's target a diverse range of DUBs as well as other members of the Ubiquitin

regulatory network, demonstrating that ubiquitin-based-modulators may potentially be developed against any member of the network.

Application of the intra-molecular systems discussed in previous chapters to these protein inhibitor scaffolds is unlikely to yield a functional switch. Topological constraints either prevent the regulatory domains from sterically occluding the active site or interfere with the binding of targets regardless of drug induced dimer disruption. Generating a small-molecule gated switch that utilizes these protein inhibitor scaffolds requires a different approach.

II. Results and Discussion

A. Split Ubv's: Rapamycin Gated DUB Inhibition

Ubiquitin has been shown to be capable of being expressed as two separate protein halves that will spontaneously re-fold if brought into proximity with the other half[19]. Recently, Pratt et al. demonstrated a system by which the re-folding of these ubiquitin halves could be induced by the small-molecule Rapamycin[20].This was demonstrated by two ubiquitin halves, with the first half containing residues 1-37, and the other half using residues 35-76. The C-terminal di-glycine motif of Ubiquitin is the site of cleavage by DUBs, thus to avoid cleavage of fusion domains, these were mutated to non-cleavable alanines. These ubiquitin halves were shown to be stable and well tolerated in the cell when overexpressed. We envisioned the Ubv inhibitors to be similarly capable of being split into two separate halves. The fusion of each half to FKBP or FRB would enable Rapamycin induced re-folding of the Ubv domain, and thus Rapamycin induced DUB inhibition. The generation of this Chemically Induced

Dimerizering Inhibitor (CIDI) would provide a unique tool for probing the ubiquitin system.

For our proof of principle, we chose Ubv21. Ubv21 is the inhibitor of USP21. Ubv21 differs from wild-type ubiquitin by only three resides, thus we hypothesized that it might be the most readily amenable to the CIDI system. We generated a split Ubv21 using the same orientation and linkers developed and optimized by Pratt et al. This entailed ten amino acid linkers separating the C-terminus of FKBP from the C-terminal Ubv21 fragment, and an eight amino acid linker between the N-terminus of FRB and the N-terminal Ubv21 fragment. This split inhibitor was assayed using a simple protease assay. In this experiment, wild-type ubiquitin is covalently linked to the fluorophore Rhodamine-110. While covalently bound, Rho-110 is non-fluorescent. Upon addition of the DUB, ubiquitin will be cleaved from Rho-110 and the fluorescence signal will increase over time. In the presence of an inhibitor, the DUB's activity will be reduced and we can compare the slope of the inhibited DUB to that of the non-inhibited DUB[21].

Initial biochemical characterization of split Ubv21 showed potent inhibition of the Ubiquitin Specific Protease 21 (USP21) in with Rapamycin (K_D = 14 nM), but still significant inhibition with addition of DMSO (K_D = 550 nM, Figure 1A). This suggests that the Ubv21 fragments retain some residual affinity for each other and will spontaneously re-fold in the absence of Rapamycin. This behavior has been previously characterized in split wild-type ubiquitin and it's been shown that mutation of an isoleucine at position 13 to an alanine or glycine completely

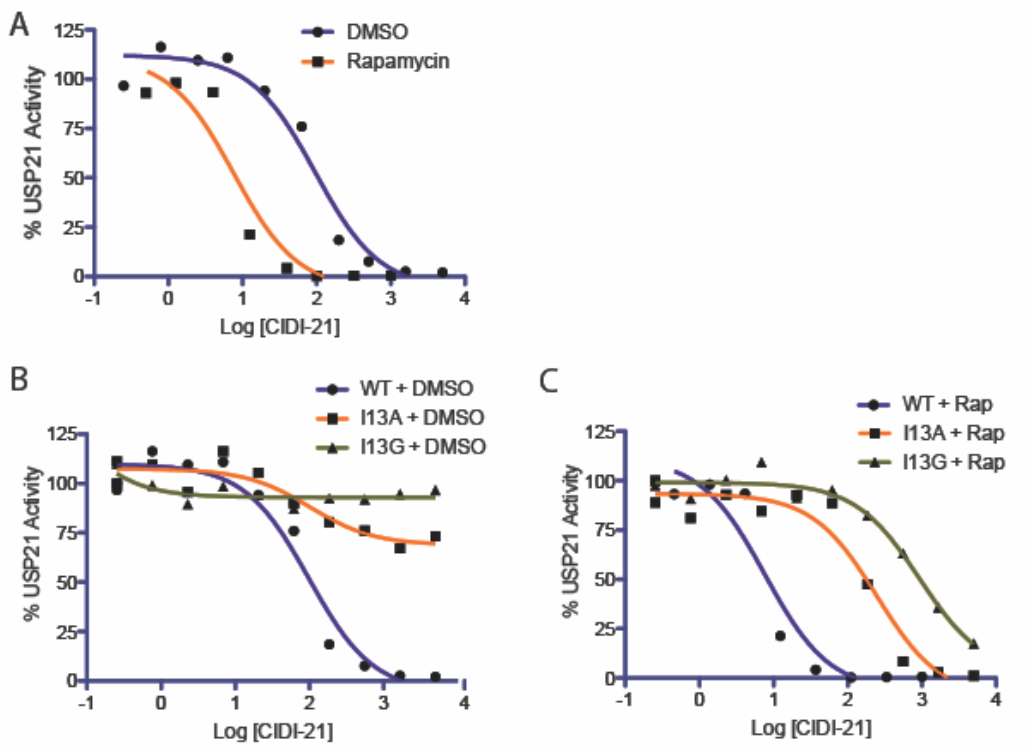


Figure 1: Biochemical characterization of CIDI-21: A) Ubv21 split into two fragments function as a Rapamycin inducible inhibitor or USP21. However, significant inhibition is observed in the absence of Rapamycin. **B and C)** Mutating a key isoleucine residue results in significant loss in potency for both DMSO and Rapamycin treated constructs.

eliminated spontaneous refolding of the two halves at physiologically relevant concentrations[19]. This isoleucine is conserved in all of the ubiquitin variant inhibitors as it does not contact any binding partner. We introduced both I13A and I13G mutations to the split Ubv21 system. The knockback in potency for DMSO treated samples is dramatic, with both constructs exhibiting IC_{50} values greater than 5000 nM. However, the resulting loss of potency for Rapamycin treated samples is significant enough as to render the inhibitors of questionable utility (Figure 1B and C).

The crystal structure of ubiquitin reveals that I13 occupies a deep and large hydrophobic pocket (Figure 2A). The I13G and I13A variants resulted in a

67- and 18-fold loss of potency in when Rapamycin was present respectively. The ability of Ubiquitin and Ubv21 to accommodate such a dramatic residue change in the interior of the protein suggested that a more conservative mutation might be able to restore potency upon drug addition while maintaining minimal spontaneous re-folding. Several mutations covering a range of size and degrees of hydrophobicity were introduced and assayed (Figure 2B**)**.

All Ubv21 variants assayed demonstrate a loss in affinity for Rapamycin treated samples as compared to wild-type Ubv21, yet they show an increase in potency when compared to I13A and I13G variants. The relative size, and hydrophobicity of the residue used appears to correlate strongly with the resulting affinity for USP21. To be relevant in a cellular context, a small-molecule regulated protein inhibitor must be inactive upon expression and potent upon induced formation of the protein-inhibitor complex. To this end, Ubv21 I13F, I13T and I13M appear to provide the low nanomolar potency necessary while essentially eliminating spontaneous complex formation.

The successful generation of a small-molecule induced protein inhibitor of USP21 led us to apply this methodology to several other Ubv's developed by Ernst et al. We attempted to apply the Rapamycin system to Ubv7, Ubv8, UbvB1 and USP2. However, generating split Ubv variants that were soluble enough for *in vitro* characterization proved to be intractable. Future work might necessitate characterization directly in a cellular context, or optimization through application of various solubility-enhancing tags may be necessary.

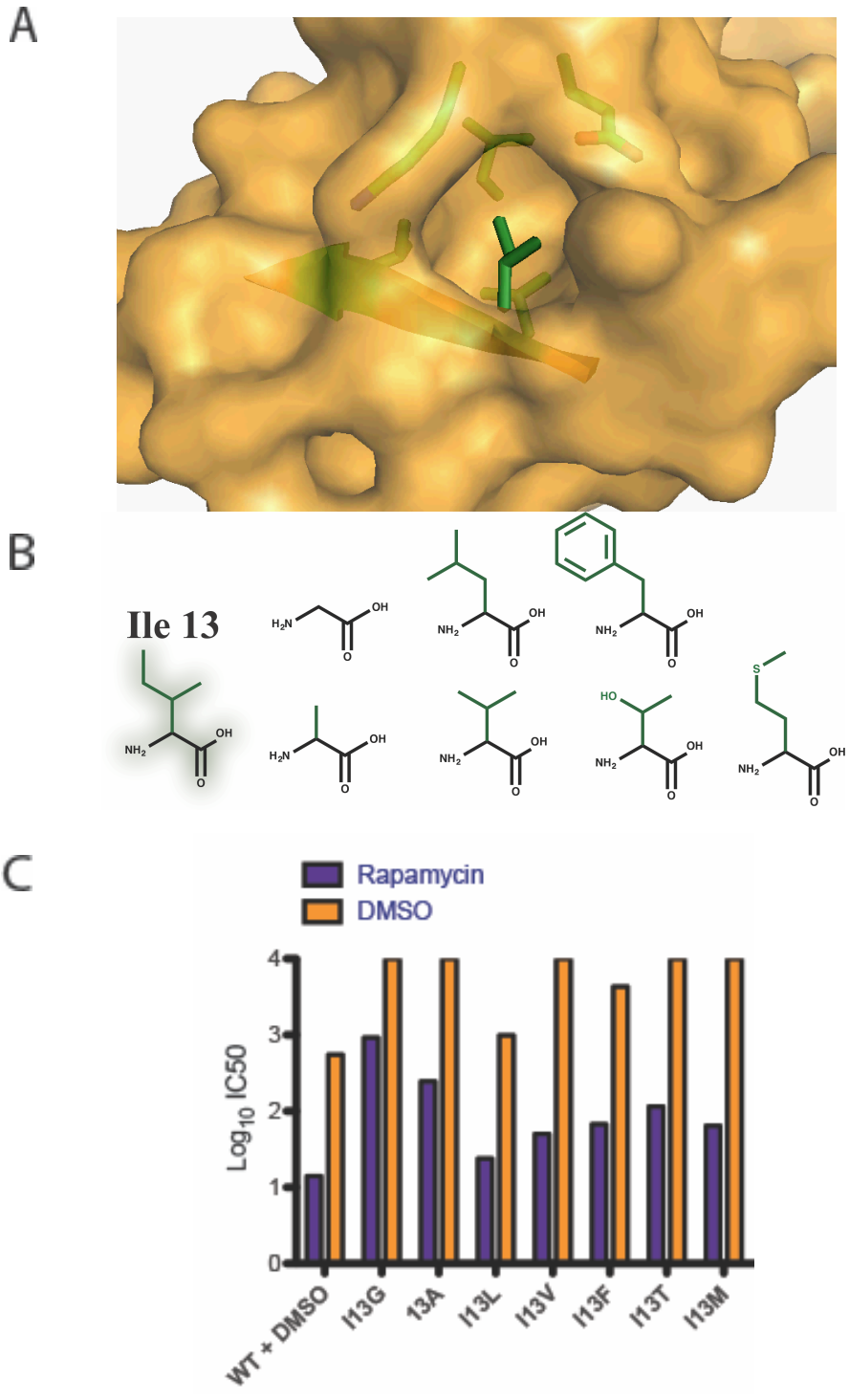


Figure 2: CIDI-21 mutations tune affinity. A) Isoleucine-13 occupies a deep hydrophobic pocket, stabilizing the ubiquitin fold. **B)** Residues used for mutation of I13. **C)** IC_{50}'s for each I13 variant are plotted as Log_{10} values.

B. Development of a Small Molecule Controlled DARPin Scaffold

The ubiquitin scaffold provided an interesting opportunity to probe a specific class of proteins, but a truly generalizable tool would comprise a small-molecule controlled protein inhibitor scaffold that could be used to probe any desired protein in the cell. As such, we envision conferring small molecule control to an antibody-mimetic protein, one that would allow for rapid in-vitro selection against a diverse array of targets and be stable in the mammalian cytosol.

The first-generation of switchable DARPins were designed to use the Bcl-xL/BH3 system described previously. Based on some initial molecular modeling, we designed a series of switchable DARPins with various linker lengths and orientations. Our model system utilized an N2C DARPin developed by Plukthun et al that selectively inhibited the c-Jun N-terminal Kinase 2 (JNK2). JNK2 is a member of the Mitogen Activated Protein Kinase (MAPK) family and DARPins were made that could differentiate between the highly similar isoforms JNK1/2/3. JNK2 is activated via phosphorylation from two upstream kinases, MKK4 and MKK7[22]. These DARPins are believed to inhibit JNK activity by occupying the MKK4/7 binding site, thus preventing phosphorylation and the subsequent activation.

The switchable DARPin designs were assayed using an ELISA screen that probed for phosphorylated JNK2. Addition of the DARPin constructs should not result in inhibition of JNK2 phosphorylation unless A3 is present. Disappointingly, all designs assayed potently inhibited JNK2 activation, regardless of the presence of A3. Using more refined modeling techniques, we

mapped the likely trajectories of the Bcl-xL and BH3 domains on the DARPin and it became clear that these domains, no matter how short the linkers were, would never block the binding surface (Figure 3A).

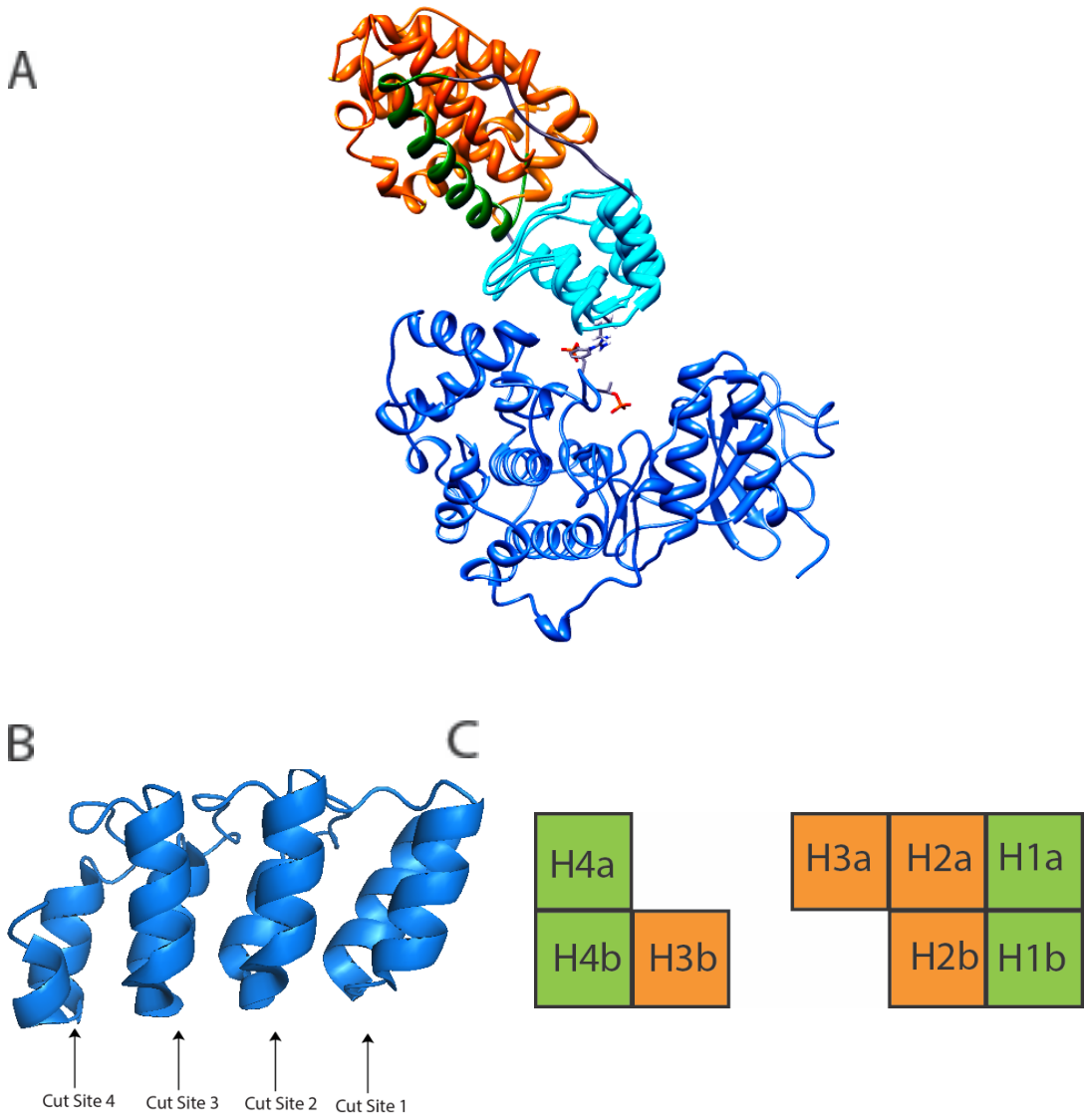


Figure 3: Generation of small-molecule gated DARPins: A) RosettaRemodel suggested that Bcl-xL (Orange) would bind to BH3 (Green), yet not across the DARPin's (cyan) binding surface for target proteins (Blue). **B+C)** The DARPin structure suggested that they may be amenable to a split-inhibitor system. The turn motif between the helical repeats were the initial candidates.

Based on the partial success of the split Ubv system, we chose to split a DARPin using the FRB/FKBP Rapamycin system. Unlike the Ubiquitin variants, no prior literature was found that discussed splitting DARPins. Due to the lack of precedent, we had to choose potential cut sites based on the available information. Based on the crystal structures of the DARPins, we chose to cut at the 'turn' positions between the α-helical repeats in each subunit (Figure 3B) These turns are not involved in the binding interface and are composed of unstructured glycine repeats.

As our model system, we chose a recently developed DARPin that inhibits the Cysteine-dependent Aspartate-directed Protease 3 (CASPase-3)[23]. CASPase-3 is an executioner CASPase that exists as a pro-enzyme. Upon initiation of the apoptotic cascade, CASPase-3 is cleaved into two subunits by the upstream CASPases-8 and -9. These two subunits form a heterodimer that subsequently interacts with another CASPase-3 heterodimer to form the fully active heterotetrameric complex. Upon formation of active CASPase3, the cell is irrevocably committed to apoptosis[24].

The DARPin fragments were assayed using a simple CASPase assay. In this assay a known CASPase substrate, the peptide DEVD, is labeled with the fluorophore 7-amino-4-methylcoumarin (AMC). When covalently bound to the peptide, AMC is non-fluorescent. Upon addition of active CASPase-3, the fluorophore is cleaved from the peptide and becomes fluorescent. By monitoring the rate of fluorescence increase over time, we can determine the IC_{50} of the DARPin fragments in the presence and absence of Rapamycin. Assays were

carried out with a starting concentration of 10 μM of each DARPin fragment and 30 μM Rapamycin in DMSO or the equivalent volume of DMSO. The mixture of Rapamycin, DARPin fragments and CASPase-3 was allowed to incubate for 1 hour before substrate addition (Figure 4A and B).

Splitting the DARPin at cut site 2 and 3 resulted in Rapamycin induced IC_{50}'s of 900 nM and 600 nM respectively. For a useful cellular inhibitor, an IC_{50} in the low nano-molar range is ideal as it allows for a wide dynamic range of control over the protein of interest. Demonstration of the capability of DARPins to be split suggests that increasing their ability to re-form in a proximity induced manner is potentially feasible. Future work toward this optimization includes swapping the orientation of the DARPin fragments (swapping which fragment is attached to either FRB or FKBP), varying the linker lengths, and modifying residues at the binding interface. Specifically, the crystal structures of all of the DARPins show a key histidine contact immediately adjacent to our designed cut site. This histidine interacts with the bottom of the α-helix on the other half. In the native, uncut state, the turn motif positions the histidine in a manner that might not be favored in the cut form due to the lack of the turn (Figure 4C).

Due to the potential for significant optimization of the inter-molecular DARPin switch, we focused effort on generating a DARPin scaffold amenable to the intra-molecular Bcl-xL/BH3 system. The modeling that suggested Bcl-xL/BH3 peptide are incapable of sterically occluding the DARPin binding site also implied that this was due to the orientation of the DARPin termini. This suggested that if

one of the termini were re-orientated, the Bcl-xL/BH3 peptide complex could be re-positioned to the binding interface.

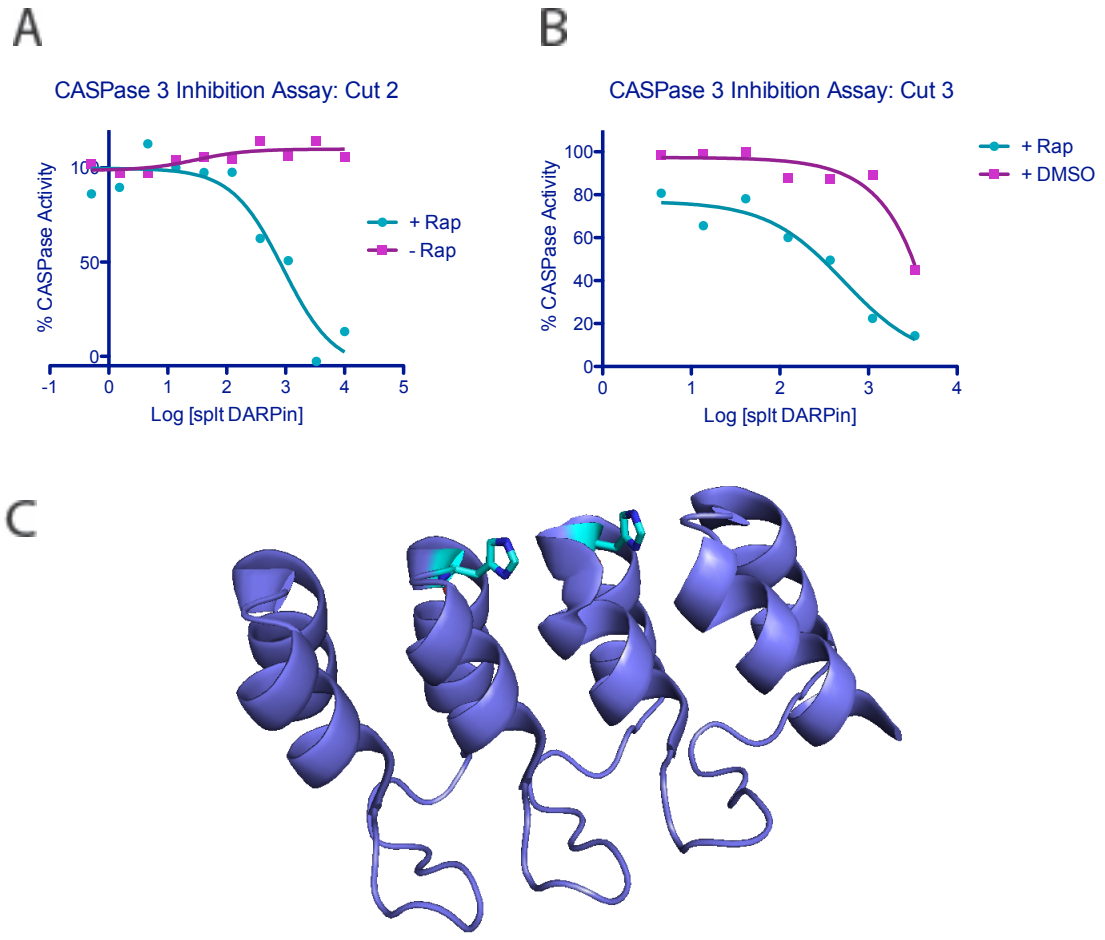


Figure 4: Inhibition of CASPase 3 by the split DARPin system: A and B) IC_{50} curves from the split-inhibitor assay. Inhibition is observed for both cut sites in the presence of Rapamycin, but not DMSO. **C)** The relatively weak inhibition of CASPase 3 is potentially due to a loss of necessary structural contacts in the DARPin. Two specific histidine residues are absent in the split design and may be contributing to this loss of potency

A set of novel single or triple helical bundles,based on the DARPin helices were designed using Rosetta. The addition of one or three helices would re-orient the DARPin termini (Figure 5). Four constructs (Hd7, Hd49, Hd57, Hd58) were selected for characterization. Each was expressed and assayed to ensure

JNK2 phosphorylation remains inhibited. All four retained potency and were thus characterized by HSQC NMR. By structurally probing the DARPin construct and comparing it to the original DARPin scaffold, the new helical domains can be assessed for folding (Figure 6).

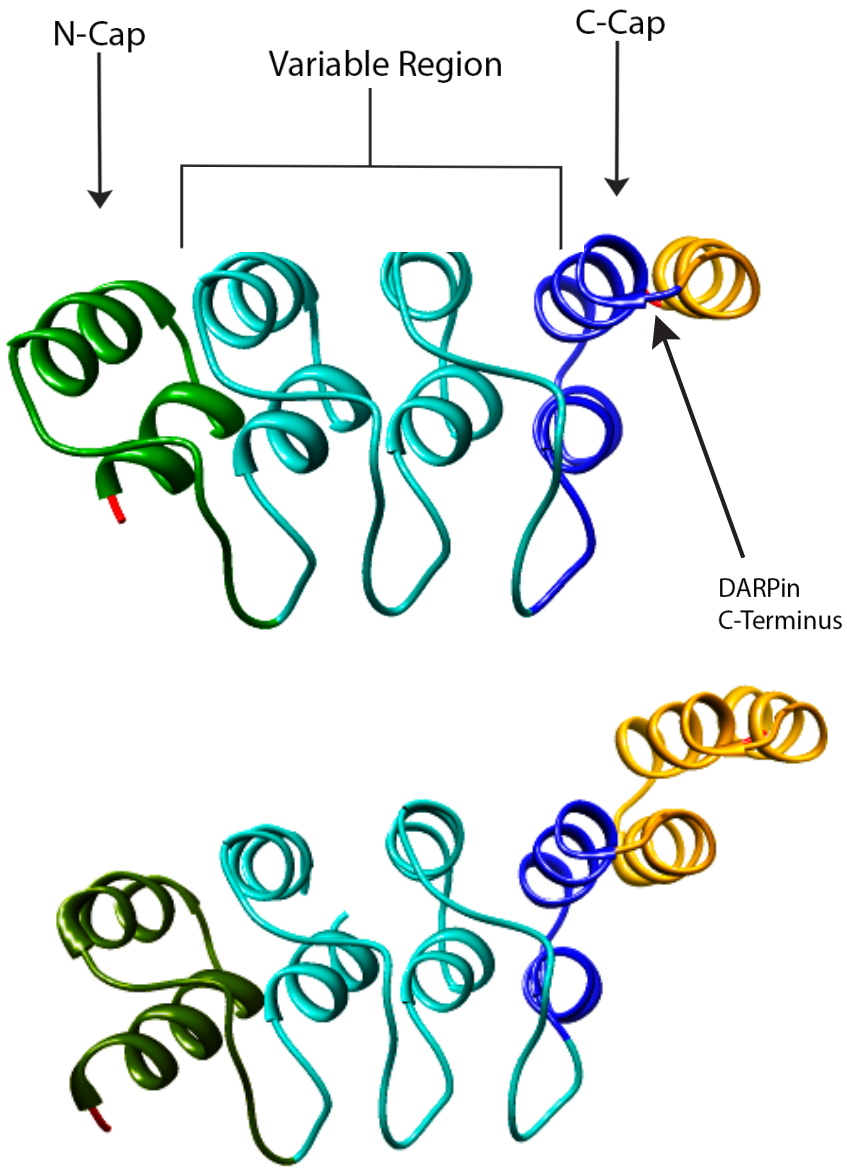


Figure 5: Rosetta designed DARPin termini re-orientation. A) Rosetta modeling predicted the addition of helices (orange) to the C-terminal capping domain (blue) would re-orient the termini and allow for Bcl-xL/BH3 occlusion of the DARPin binding surface. **B)** An additional 3 helical repeat was also suggested by modeling.

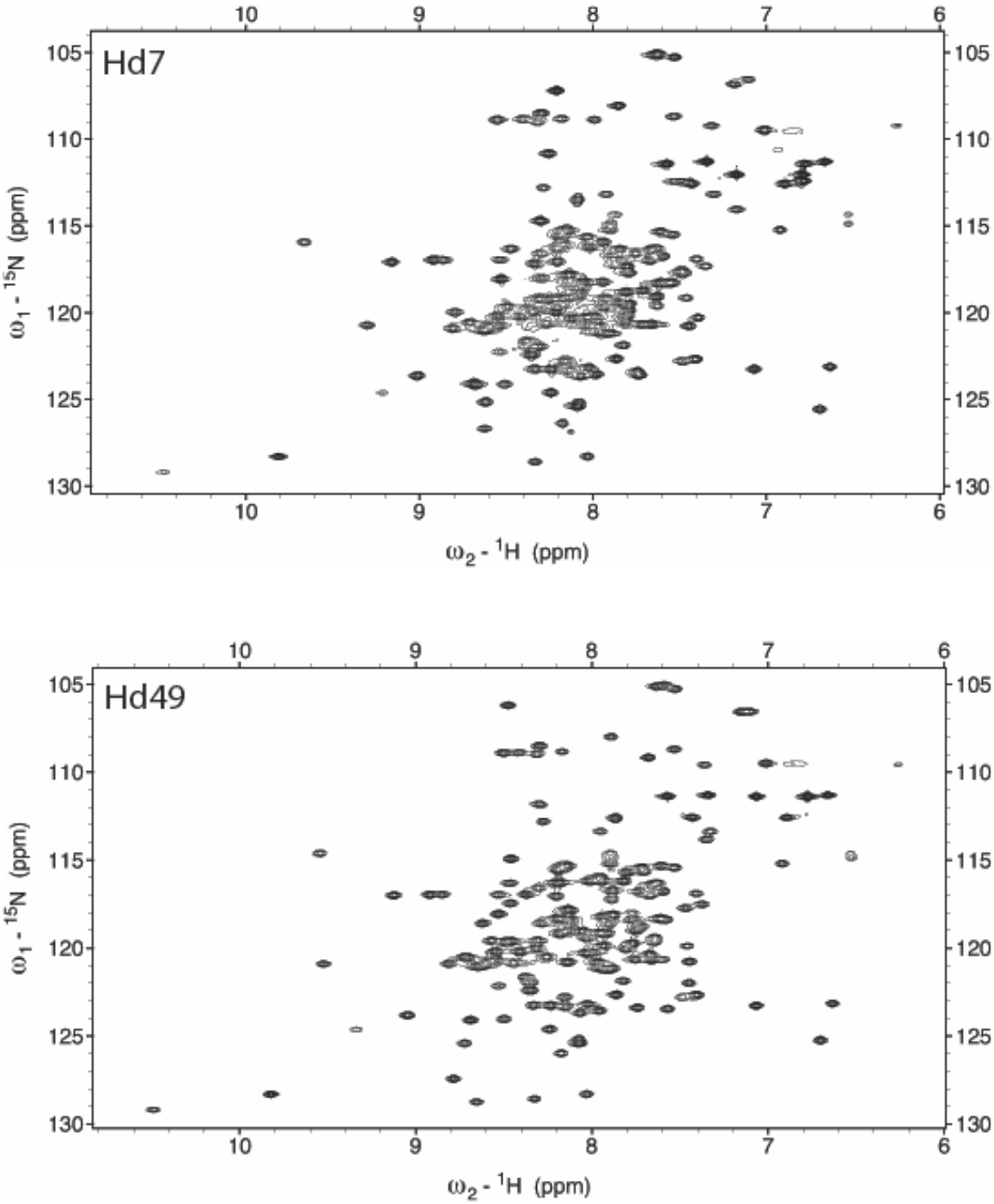


Figure 6A and B: HSQC spectra for designed DARPin helices. All four sets of designed helices produce highly ordered, clearly defined HSQC spectra. All residues are visible, implying all residues are highly ordered. This suggests that the helical designs are well folded as intended.

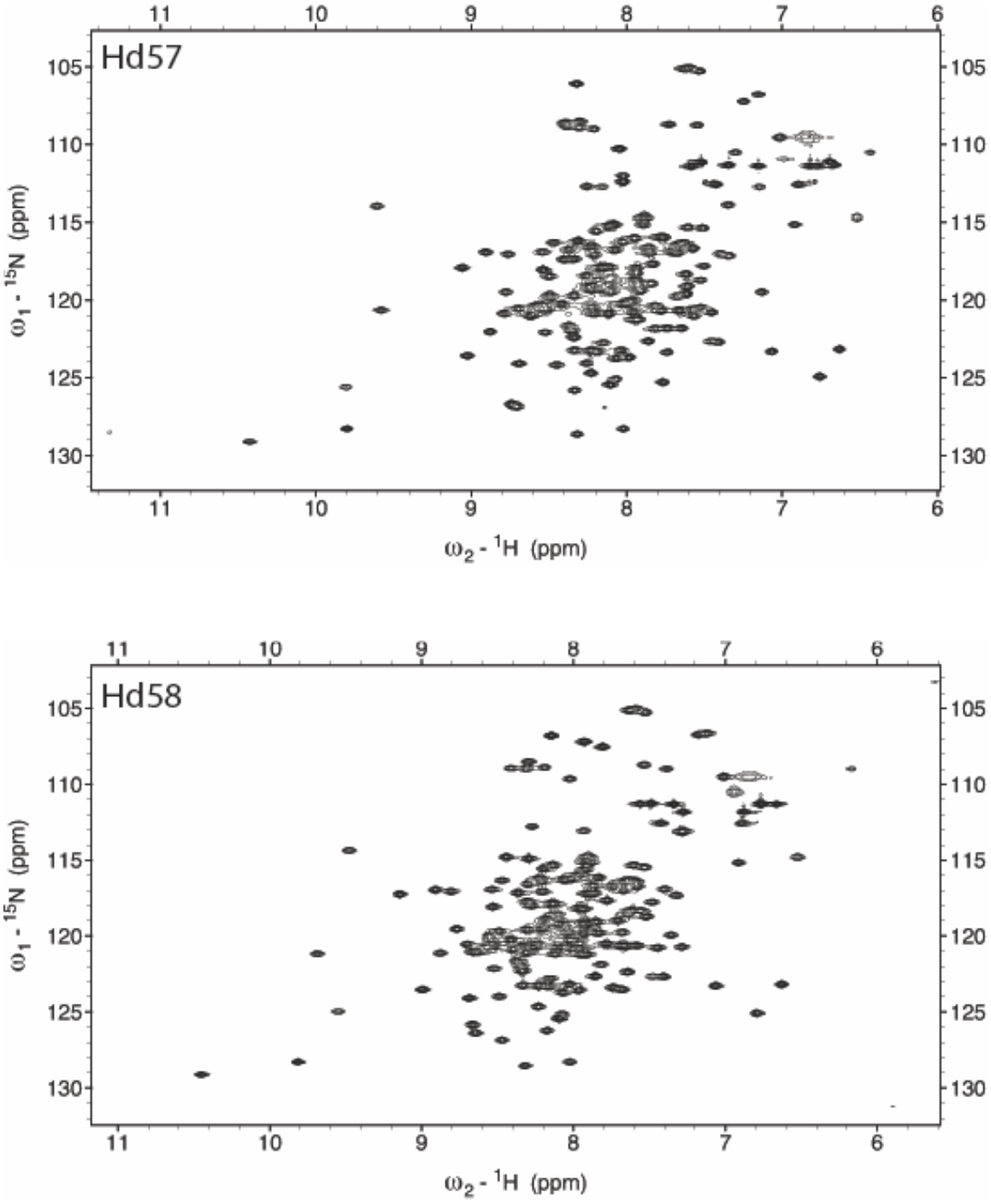


Figure 6C and D: HSQC spectra for designed DARPin helices. All four sets of designed helices produce highly ordered, clearly defined HSQC spectra. All residues are visible, implying all residues are highly ordered. This suggests that the helical designs are well folded as intended.

All DARPins retain potent inhibition for JNK2. Additionally, all four helical constructs demonstrate that they are well folded. This technique does not allow for precise structural determination, simply whether the residues are structurally rigid and their position is well defined. These constructs are ideal candidates for future use in application of the Bcl-xL/BH3 peptide switch.

III. Conclusion.

The toolkit available to chemical biologists is a constantly growing. Chemical genetic techniques are continually being expanded upon and refined, with focus being placed on generalizable tools that can be quickly and simply applied to a variety of unique problems. However, current tools are ill equipped to probe the spatio-temporal dynamics of signaling pathways in the cell. To this end, we attempted to confer small-molecule control to a set of protein inhibitor scaffolds. These scaffolds, once optimized, would allow for the rapid generation of a variety of potent and selective inhibitors using *in vitro* selection methods. Methods such as phage-display and ribosome display have been used to rapidly generate a diverse array of binding and inhibitor proteins off of a single protein scaffold.

The feasibility of this system was demonstrated using two unique scaffolds. The first took advantage of a class of inhibitors of De-Ubiquitinases based on their substrate ubiquitin. By separating Ubv21 into two separate protein tags, we generated an inhibitor capable of targeting USP21 in a Rapamycin

induced manner. However, this switch failed to be generally applicable as other Ubv's were not amenable to *in vitro* characterization.

The second scaffold we targeted was a set of antibody mimetic proteins called DARPins. Initial attempts focused on generating an intra-molecularly gated DARPin, yet this failed. By applying the split-protein methodology we demonstrated for Ubv21, we generated a split DARPin construct capable of inhibiting CASPase3 in a Rapamycin dependent manner. While functional, significant optimization is necessary to restore potency to a level that is of use in a cellular context. Finally, we began work on designing a new DARPin scaffold in which we add structural elements to allow for application of an intra-molecular switch. Initial characterization revealed four constructs that demonstrate folding of the new helical bundles and retain potency for prevention of JNK2 phosphorylation.

IV. Materials and Methods:

1. Ubv21 and CIDI Cloning

A. Gene Synthesis General Methods

CLONING OF UBV'S

The Ubv21, Ubv8, Ubv7, UbvB1, Ubv2 genes were obtained as double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). The genes were subcloned into pMCSG7, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF FRB CONSTRUCTS

The FRB gene was obtained as a double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). The N-terminal half (residues 1-37) of each Ubv were isolated by PCR and contained Gibson overhangs designed in NEBuilder (NEB). The genes were subcloned into pMCSG7, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF FKBP CONSTRUCTS

The FKBP gene was obtained as a double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). The C-terminal half (residues 35-76) of each Ubv were isolated by PCR and contained Gibson overhangs designed in NEBuilder (NEB). The genes were subcloned into pMCSG7, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF USP21

USP21 was purchase from Addgene (plasmid #22574). The USP21 gene was isolated by PCR and designed to contain Gibson overhangs by NEBuilder (NEB). The gene was subcloned into pET28a, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF CASPase DARPin D3.4

CASPase DARPin D3.4 genes were obtained as a double stranded DNA G-Block (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). The

genes were subcloned into pMCSG7, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF FRB CONSTRUCTS

The FRB gene was obtained as a double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). The N-terminal halves (residues 1-46 Cut-N2, residues 1-79 cut-N3) of D3.4 were isolated by PCR and contained Gibson overhangs designed in NEBuilder (NEB). The genes were subcloned into pMCSG7, a bacterial expression vector containing an N-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

CLONING OF FKBP CONSTRUCTS

The FKBP gene was obtained as a double stranded DNA G-Blocks (IDT) containing Gibson Assembly overhangs designed in NEBuilder (NEB). The C-terminal half (residues 46-124 Cut C2, residues 79-124 Cut C3) of D3.4 were isolated by PCR and contained Gibson overhangs designed in NEBuilder (NEB). The genes were subcloned into pMCSG7, a bacterial expression vector containing a C-terminal hexahistidine tag using Gibson Assembly (NEB, product number E2611L).

B. Fusion Construct Protein Sequence

Ubv21 Design

His_6-(GSGT)-Ubv21

Ubv21 Sequence

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTL

SDYNIQKWSTLFLLLRLRAA

Ubv8 Design

His_6-(GSGT)-Ubv8

Ubv8 Sequence

MRIVVKTLMGRTIILEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTL

SDYNIHNHSALYLLLKLRAA

Ubv7 Design

His_6-(GSGT)-Ubv7

Ubv7 Protein Sequence

MQIFVKFRTGKTYTLEVEPSDTIENVKAKIQDKLGIPPDQQWLIFAGKRLEDGRT

LSDYNIQKESTLRGVRRLRAA

Ubv2 Design

His_6-(GSGT)-Ubv2

Ubv2 Protein Sequence

MQIFVNTLSGKHITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTL

SDYNIQKESTLHLVLRLRAA

UbvB1 Design

His_6-(GSGT)-UbvB1

UbvB1 Protein Sequence

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQKLLFARKQLEDGRTL

SDYNIHKESFLYLVLRLRAA

N-Terminal Ubv Fragment FRB Fusion Design

His_6-(GSGT)-Ubv#$_{(1-37)}$-GGSTMAA-FRB

FRB Protein Sequence

MASRILWHEMWHEGLEEASRLYFGERNVKGMFEVLEPLHAMMERGPQTLKET

SFNQAYGRDLMEAQEWCRKYMKSGNVKDLLQAWDLYYHVFRRISKTS

C-Terminal Ubv Fragment FKBP Fusion Design

His_6-(GSGT)-FKBP-GSLEGSTMSG-Ubv#$_{(35-76)}$

FKBP Protein Sequence

ASRGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFM

LGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVEL

LKLETS

USP21 Design

His_6-$(GSG)_3$-USP21

USP21 Protein Sequence

SGHVGLRNLGNTCFLNAVLQCLSSTRPLRDFCLRRDFRQEVPGGGRAQELTE

AFADVIGALWHPDSCEAVNPTRFRAVFQKYVPSFSGYSQQDAQEFLKLLMERL

HLEINRRGRRAPPILANGPVPSPPRRGGALLEEPELSDDDRANLMWKRYLERE

DSKIVDLFVGQLKSCLKCQACGYRSTTFEVFCDLSLPIPKKGFAGGKVSLRDCF

NLFTKEEELESENAPVCDRCRQKTRSTKKLTVQRFPRILVLHLNRFSASRGSIKK

SSVGVDFPLQRLSLGDFASDKAGSPVYQLYALCNHSGSVHYGHYTALCRCQT

GWHVYNDSRVSPVSENQVASSEGYVLFYQLMQEPPR

CASPase DARPin D3.4 Design

His_6-$(GSG)_3$-D3.4

DARPin D3.4 Protein Sequence

DLGKKLLEATRAGQDDEVRILMANGADVNAMDDAGVTPLHLAAKRGHLEIVEVL

LKHGADVNASDIWGRTPLHLAATVGHLEIVEVLLEYGADVNAQDKFGKTAFDISI

DNGNEDLAEILQKLN

N-Terminal D3.4 Fragment FRB Fusion Design

His_6-$(GSG)_3$-FRB-GSLEGSTMSG-$D3.4_{(res1-46,79)}$

C-Terminal D3.4 Fragment FKBP Fusion Design

His_6-(GSGT)-$D3.4_{(46,79-124)}$-GGSTMAAA-FKBP

Designed DARPin Helical Variant Design

D3.4-GS-His_6

Hd7 Protein Sequence

DLGKKLLEAARAGQDDEVRILMANGADVNAYDDNGVTPLHLAAFLGHLEIVEVL

LKYGADVNAADSWGTTPLHLAATWGHLEIVEVLLKHGADVTAQDKFGKTAIDIV

EDNGNQDLLEIMRRFLEKQGGDKTLIEKMEKLVRT

Hd49 Protein Sequence

DLGKKLLEAARAGQDDEVRILMANGADVNAYDDNGVTPLHLAAFLGHLEIVEVL

LKYGADVNAADSWGTTPLHLAATWGHLEIVEVLLKHGADVTAQDKFGKTAVDIV

SDNGNKDLMEILRRFLEKQGGDKTLIEKLEKLIRT

Hd57 Protein Sequence

DLGKKLLEAARAGQDDEVRILMANGADVNAYDDNGVTPLHLAAFLGHLEIVEVL

LKYGADVNAADSWGTTPLHLAATWGHLEIVEVLLKHGADVTAQDKFGKTAVDIA

EDNGNQDLLEILRRFLEKQGGDKTLIEKIRKLVET

Hd58 Protein Sequence

DLGKKLLEAARAGQDDEVRILMANGADVNAYDDNGVTPLHLAAFLGHLEIVEVL

LKYGADVNAADSWGTTPLHLAATWGHLEIVEVLLKHGADVSAQDKFGKTAIDIV

SDNGNQDLLEILRRFLEKQGGDKTLIEKVRELVEK

C. Expression and Purification Protocols:

Ubv Variant Expression and Purification

The His_6-Ubv plasmids were transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to inoculate 250 mL of LB both with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 1.0 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For His6-Ubv purification, the pellets were thawed on ice and re-suspended in 10 mL of His_6 Lysis Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20 mM imidazole). The re-suspended cell pellets were lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of Lysis Buffer and the protein was eluted in 3 mL of Elution Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 200 mM Imidazole). Purified protein was dialyzed twice into 1000 mL Storage Buffer (50 mM HEPES pH 7.8, 100 mM NaCl). Protein was stored by snap-freezing aliquots and storing at -80 °C.

FRB/FKBP Ubv21 Fragment Expression and Purification

The His_6-$Ubv21_{1-37}$-FRB and His_6-FKBP-$Ubv21_{35-76}$ plasmids were transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of

the 5 mL culture was used to inoculate 250 mL of LB both with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 0.25 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For purification, the pellets were thawed on ice and re-suspended in 10 mL of LS-His_6-Lysis Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20 mM imidazole, 10% glycerol). The re-suspended cell pellets were lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of chilled LS-Lysis Buffer and the protein was eluted in 3 mL of Elution Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 200 mM Imidazole, 10% glycerol). Purified protein was dialyzed twice into 1000 mL chilled LS-Storage Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 10% Glycerol). Protein was stored by snap-freezing aliquots and storing at -80 °C.

USP21 Expression and Purification

The USP21 plasmid was transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with Kanamycin (50 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to inoculate 500 mL of LB both with Kanamycin (50 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 0.25 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For purification, the pellets were thawed on ice and re-

suspended in 10 mL of USP-His_6-Lysis Buffer (10 mM Tris-HCl pH 7.0, 500 mM NaCl, 10 mM imidazole, 1 mM beta-mercaptoethanol, 0.1 µM PMSF, 5% glycerol). The re-suspended cell pellets were lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of chilled USP-Lysis Buffer and the protein was eluted in 3 mL of USP-Elution Buffer (10 mM Tris-HCl pH 7.0, 500 mM NaCl, 200 mM imidazole, 1 mM beta-mercaptoethanol, 0.1 µM PMSF, 5% glycerol). Purified protein was dialyzed twice into 1000 mL chilled USP-Storage Buffer (20 mM Tris-HCl pH 7.0, 500 mM NaCl, 2 mM DTT, 5% Glycerol). Protein was stored by snap-freezing aliquots and storing at -80 °C.

DARPin Variant Expression and Purification

The His_6-D3.6 plasmid was transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to inoculate 250 mL of LB both with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 1.0 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For His6-D3.4 purification, the pellets were thawed on ice and re-suspended in 10 mL of His_6 Lysis Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20 mM imidazole). The re-suspended cell pellets were lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified

using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of Lysis Buffer and the protein was eluted in 3 mL of Elution Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 200 mM Imidazole). Purified protein was dialyzed twice into 1000 mL Storage Buffer (50 mM HEPES pH 7.8, 100 mM NaCl). Protein was stored by snap-freezing aliquots and storing at -80 °C.

FRB/FKBP Ubv21 Fragment Expression and Purification

The His_6-FRB-D3.4 and His_6-D3.4-FKBP plasmids were transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB broth with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to inoculate 250 mL of LB both with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 0.25 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For purification, the pellets were thawed on ice and re-suspended in 10 mL of LS-His_6-Lysis Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 20 mM imidazole, 10% glycerol). The re-suspended cell pellets were lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of chilled LS-Lysis Buffer and the protein was eluted in 3 mL of Elution Buffer (50 mM HEPES pH 7.8, 100 mM NaCl, 200 mM Imidazole, 10% glycerol). Purified protein was dialyzed twice into 1000 mL chilled LS-Storage Buffer (50 mM

HEPES pH 7.8, 100 mM NaCl, 10% Glycerol). Protein was stored by snap-freezing aliquots and storing at -80 °C.

DARPin Variant Expression and Purification for NMR Analysis

The His_6-D3.6 plasmid was transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of ^{15}N M9 minimal media (1x M9 Salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% ^{15}N- labeled NH_4Cl, 10 mM glucose) with ampicillin (100 µg/mL). 18 hours post inoculation, the entirety of the 5 mL culture was used to inoculate 250 mL of ^{15}N M9 minimal media with ampicillin (100 µg/mL). Cultures were grown at 37 °C to on OD_{600} of 0.8, cooled to 18 °C and induced with 1.0 mM IPTG. Protein was expressed at 18 °C overnight. Cells were harvested by centrifugation and pellets stored at -80 °C. For His6-D3.4 purification, the pellets were thawed on ice and re-suspended in 10 mL of His_6 NMR-Lysis Buffer (50 mM HEPES pH 7.8, 150 mM NaCl, 20 mM imidazole). The re-suspended cell pellets were lysed via sonication and the lysate was cleared by centrifugation. The cleared lysate was purified using Ni-NTA agarose (Qiagen) by rotating at 4 °C for 1 hour. The resin was subsequently washed with 10 mL of Lysis Buffer and the protein was eluted in 3 mL of Elution Buffer (50 mM HEPES pH 7.8, 150 mM NaCl, 200 mM Imidazole). Purified protein was dialyzed twice into 1000 mL NMR Buffer (50 mM HEPES pH 7.8, 150 mM NaCl, 5% glycerol, 5 mM DTT, in 95% H_2O/ 5% D_2O). Protein was stored by snap-freezing aliquots and storing at -80 °C.

D. USP21 Protease Inhibition Assay

The potency of the split Ubv21 inhibitor against USP21 protease was measured using the cleavage-sensitive fluorogenic substrate Ub-7-amido-4-methylcoumarin (Ub-AMC, Boston Biochem). Titrations of inhibitors (3-fold serial dilutions starting at 10 μM) in the presence of 80 μM Rapamycin or DMSO were added to a black 96-well plate (Corning, product number 3720) containing 20 nM USP21. Inhibitors were incubated with USP21 at room temperature for 1 hour. To each well was simultaneously added substrate Ub-AMC to a final concentration of 250 nM and reactions were monitored by measuring the fluorescence increase every minute for 30 minutes at 22 °C on a Perkin Elmer EnVision fluorimeter (excitation, 340 nm; emission 465 nm). Each measurement was carried out in triplicate. Slopes of the fluorescence increase were compared to a no-protease control. A nonlinear regression model was used to fit curves with GraphPad Prism.

E. CASPase Inhibition Assay

The potency of the split D3.4 DARPin inhibitor against CASPase-3 was measured using the cleavage-sensitive fluorogenic substrate DEVD-amido-4-methylcoumarin (DEVD-AMC, ENZO Diagnostics). Titrations of inhibitors (3-fold serial dilutions starting at 10 μM) in the presence of 80 μM Rapamycin or DMSO were added to a black 96-well plate (Corning, product number 3720) containing 10 nM CASPase-3 (Abchem). Inhibitors were incubated with CASPas-3 at room temperature for 1 hour. To each well was simultaneously added substrate DEVD-

AMC to a final concentration of 50 μM and reactions were monitored by measuring the fluorescence increase every 30 seconds for 30 minutes at 22 °C on a Perkin Elmer EnVision fluorimeter (excitation, 340 nm; emission 465 nm). Each measurement was carried out in triplicate. Slopes of the fluorescence increase were compared to a no-protease control. A nonlinear regression model was used to fit curves with GraphPad Prism.

F. Designed DARPin Helical NMR

A naturally abundant ^{15}N-1H HSQC experiment was collected at 25 °C on a Bruker DRX 500 MHz spectrometer with a cryogenically cooled TCI probe. Data was collected with 256 scans to increase S/N with 2048x256 data points in the direct (^{1}H) and indirect (^{15}N) dimensions respectively. For ^{15}N labeled samples, the same sample concentrations, buffer conditions, NMR experiment, and magnet were used. Data was collected at 25 °C and 37 °C and the total number of scans was decreased to 8 to account for uniformly enriched ^{15}N signal. Data was processed with NMRPIPE and analyzed by SPARKY.

V. References

1. Karatan, E. et al. **(2004)** Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. *Chem. Biol.* 11: 835-844.

2. Ernst, A. et al. **(2013)** A strategy for modulation of enzymes in the ubiquitin system. *Science* 339: 590-595.

3. Binz, H. K. et al. **(2004)** High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotech.* 22: 575-582.

4. Stumpp, M. T.; Binz, K.; Amstutz, P. **(2008)** DARPins: a new generation of protein therapeutics. *Drug Discov. Today* 13: 695-701.

5. Boersma, Y. L. Chao, G.; Steiner, D.; Wittrup, K. D.; Plukthun, A. **(2011)** Bispecific designed ankyrin repeat proteins (DARPins) targeting epidermal growth factor receptor inhibit A431 cell proliferation and receptor recycling. *J. Biol. Chem.* 48: 41273-41285.

6. Parizek, P. et al. **(2012)** Designed ankyrin repeat proteins (DARPins) as novel isoform-specific intra-cellular inhibitors of c-jun n-terminal kinases. *ACS Chem. Biol.* 7: 1356-1366.

7. Zhang, Y. et al. **(2012)** Conformational stabilization of ubiquitin yields potent and selective inhibitors of USP7. *Nat. Chem. Biol.* 9: 51-61.

8. Binz, H. K. et al. **(2003)** Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* 332: 489-503.

9. Forrer, P.; Stumpp, M. T.; Binz, H. K.; Pluckthun, A. **(2003)** A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* 539: 2-6.

10. Plukthun, A. **(2015)** Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol*. 55: 489-511.

11. Welchman, R. L.; Gordon, C.; Mayer, R.J. **(2005)** Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell Bio.* 6: 599-609.

12. Kerscher, O.; Felderbaum, R.; Hochstrasser, M. **(2006)** Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* 22: 159-180.

13. Reyes-Turcu, F. E. et al. **(2006)** The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif in unachored ubiquitin. *Cell* 124: 1197-1208.

14. Hurley, J. H.; Lee, S.; Prag, G. **(2006)** Ubiquitin-binding domains. *Biochem. J.* 399: 361-372.

15. Ciechanover, A. **(2015)** The unravelling of the ubiquitin system. *Nat. Rev. Mol. Cell Biol.* 16: 322-324.

16. Kawabe, H.; Brose, N. **(2011)** The role of ubiquitinylation in nerve cell development. *Nat. Rev. Neurosci.* 12: 251-268.

17. Komander, D.; Clague, M. J.; Urbe, S. **(2009)** Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* 10: 551-563.

18. Wilson, S. M. et al. **(2002)** Synaptic defects in ataxia mice result from a mutation in Usp14, encoding a ubiquitin-specific protease. *Nat. Genet.* 32: 420-425.

19. Johnsson, N.; Varshavsky, A. **(1994)** Split ubiquitin as a sensor of protein interactions in vivo. *PNAS* 91: 10340-10344.

20. Pratt, M. R.; Schwartz, E. C.; Muir, T. W. **(2007)** Small-molecule-mediated rescue of protein function by an inducible proteolytic shunt. *PNAS* 104: 11209-11214.

21. Hassiepen, U. et al. **(2007)** A sensitive fluorescence intensity assay for deubiquitinating proteases using ubiquitin-rhodamine110-glycine as substrate. *Anal. Biochem.* 371: 201-207.

22. Davies, R. J. **(2000)** Signal transduction by the JNK group of MAP kinases. *Cell* 103: 239-252.

23. Schroeder, T. et al. **(2013)** Specific inhibition of caspase-3 by a competitive DARPin: molecular mimicry between native and designed inhibitors. *Structure* 21: 277-289.

24. Porter, A. G.; Janicke, R. U. **(1999)** Emerging roles of caspase-3 in apoptosis. *Cell Death Differ.* 6: 99-104.