Protein-based Degrader Strategies against Oncogenic RAS

by

Keith M. Cheah

M.Eng., Imperial College London (2018)

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Author .....
Keith M. Cheah
Department of Chemical Engineering
April 5, 2023

Certified by .....
K. Dane Wittrup, Ph.D.
C.P. Dubbs Professor of Chemical Engineering & Biological Engineering
Thesis Supervisor

Certified by .....
Ronald T. Raines, Ph.D.
Roger and Georges Firmenich Professor of Natural Products Chemistry
Thesis Supervisor

Accepted by .....
Patrick S. Doyle
Robert T. Haslam Professor of Chemical Engineering
Graduate Officer

Protein-based Degrader Strategies against Oncogenic RAS

by

Keith M. Cheah

Submitted to the Department of Chemical Engineering
on April 5, 2023, in partial fulfillment of the
requirements for the degree of of
Doctor of Philosophy

Abstract

Targeted therapies have emerged as a promising cancer treatment strategy by inhibiting proteins specific to cancer cells while preserving healthy cells. RAS proteins play a crucial role in cancer development and are associated with increased tumor growth and invasion. Mutations in RAS genes, especially KRAS G12D, are present in a large proportion of human cancers. However, these proteins have been deemed "undruggable" due to the absence of binding pockets for small-molecule drugs, and the significant challenge of delivering protein-based inhibitors across the cell membrane to reach intracellular RAS.

This thesis focuses on the development of novel protein-based degrader strategies against KRAS G12D, specifically. We first developed a generalizable solubilization strategy to address the low aqueous solubility of proteins that have undergone bioreversible esterification, a permeation strategy that involves raising the cationicity and hydrophobicity of the protein. We then engineered a cell-permeable KRAS-G12D-targeting degrader that consists of an esterified protein-based KRAS-G12D binder, R11.1.6, conjugated to a small molecule ligand of the VHL E3 ligase, VL1. We confirmed the cytosolic entry of esterified R11.1.6-VL1 and demonstrated efficacy in human cancer cell lines through *in vitro* studies. Although modest efficacy in RAS degradation and growth inhibition was observed, this strategy presents a novel paradigm for targeting previously undruggable proteins.

Finally, we build on previous work on intracellularly expressed KRAS-G12D-targeting biodegraders. Unlike the cell-permeable degrader described above, which consists of a small-molecule component, biodegraders are fully protein-based constructs consisting of R11.1.6 conjugated to an E3 ligase itself. They can thus be genetically expressed in cells, eliminating the need for transmembrane delivery. The development of degraders has largely been limited to a trial-and-error approach, with little understanding of the effects of specific design components like linker length, linker rigidity, and target affinity. We utilized high-throughput fluorescence-based screening and regression modeling to determine the relative importance and effect of

such design components on RAS degradation, offering several rational design principles that will inform the future development of RAS-targeting biodegraders.

Overall, these findings offer a valuable contribution to the ongoing efforts in developing targeted therapies against RAS and potentially enabling RAS to become a more druggable target.

Thesis Supervisor: K. Dane Wittrup, Ph.D.
Title: C.P. Dubbs Professor of Chemical Engineering & Biological Engineering

Thesis Supervisor: Ronald T. Raines, Ph.D.
Title: Roger and Georges Firmenich Professor of Natural Products Chemistry

Acknowledgments

As I reflect on my doctoral journey, I am humbled and deeply grateful for the support, guidance, and encouragement of so many individuals who have helped me reach this important milestone. Each person has played a unique and important role in shaping my research, fostering my intellectual growth, and supporting me through the ups and downs of graduate school. Despite my best efforts, I know that I will not be able to adequately capture the gratitude I feel for each one of you. Your unwavering support have made this achievement possible, and I want to express my sincere appreciation.

I would like to thank my advisor, Dane Wittrup, for taking a chance on me. I came into my PhD with little biology-related research experience, and I was not confident that the skills and experience I had were suited for his lab. Dane, however, encouraged me to focus on my interests, and patiently guided and mentored me as I developed my skills as a protein engineer. Thank you for always supporting and encouraging me, for teaching me the ability to reassess and redirect, and for always looking out for my best interests inside and outside the lab. I am especially grateful for your willingness to explore my interests in data science and machine learning with me, and for affording me the freedom to venture beyond the lab to pursue those interests in various contexts.

Thank you to my co-advisor, Ronald Raines, for your guidance and mentorship over these last five years. While I primarily worked at the Wittrup Lab, I always felt welcomed and at home at the Raines Lab. Thank you for always keeping a close eye on the progress of my research, for the encouraging reminders that failed experiments are as equally valuable as successful ones, and for very patiently and diligently guiding me through the process of writing my very first paper.

Thank you to my thesis committee members, Hadley Sikes, Eric Fischer, and former member Darrell Irvine. Your insightful questions, feedback, and advice throughout the years have been instrumental in my thesis work. I am grateful to each one of you not only for contributing to my scientific progress but also for your consistent check-ins on my personal and professional development.

I am also deeply thankful for many other incredible mentors who have guided me throughout my academic career. Thank you to my MIT Biotech Group industry mentors—Seth Berman, Jon Gilbert, Wendy Lam, and Quinn Johns—for taking the time to help me carefully consider and shape my post-PhD career path. Thank you to several professors, the late Prof. Jim Swan, Prof. Shiv Pillai, Prof. Jónas Jónasson, Anne Quaadgras, Prof. Joseph Doyle, Prof. David Sontag, and Dr. Madhur Nayan, for positively shaping my learning experience at MIT. I am especially grateful to have had the opportunity to collaborate with Madhur on a clinical machine learning project since early 2022, through which I have picked up many skills in data science and machine learning. Thank you for always patiently mentoring and guiding me in an area that was completely new to me just 18 months ago. Thank you to my Practice School directors, Brian Baynes, Brian Stutts, and Doug Harrison, for playing an important role in my professional development. And thank you to my undergraduate mentors, Dr. Andreas Kogelbauer, Prof. Andrew Livingston, Prof. Stephen Buchwald, Prof. Tristan Lambert, Dr. Kurt Armbrust, and Dr. Ludmila Peeva, for your invaluable guidance and mentorship during my early years as an academic researcher.

Thank you to former and current members of both the Raines and Wittrup Labs for shaping my PhD experience in such a positive way. Thank you, Adrienne, Alison, Alli, Anthony, Bill, Bri, Byong, Cora, David, Emi, Henry Z., Jordan, Joseph, Lauren, Liz, Luciano, Megan, Naveen, Noor, Owen, Yash, Sarah, Aniekan, Axel, Bella, Brian Gold, Brian Graham, Brian L., Caglar, Claire, Dilyara, Evans, Forrest, Henry K., Ian, Jessica, Jesús, Jinyi, JoLynn, Kenton, Lucas, Lindsey, Nile, Oscar, Vicky, and Yana – for all the fun, laughter, constructive feedback, and positive affirmations. I am so lucky to have had the chance to work with and become friends with such a brilliant, collaborative, and fun group of scientists and engineers. In particular, I want to thank JoLynn, Henry K., Byong, Emi, Joseph, Noor, and Yash for teaching me the basics of biological scientific research during my first few months in the lab. Thank you, Vicky, for sharing your years of brilliant research expertise with me, for patiently guiding me in writing my very first paper, and for the fun morning coffee chats at Blue Bottle; I look forward to many more. Thank you, Aniekan, for collaborating closely

with me to continually ideate and test out new degrader strategies; it never fails to impress me how quickly you become aware of newly published research. Thank you, Megan, for graciously teaching and sharing with me all your knowledge on biodegrader development, and for your fun company while we grind away in the TC hoods. Thank you, Alli, for the opportunity to collaborate and gain experience in RNA sequencing analysis, for your incredible homemade cakes, and for your encouraging words during the tough periods of my PhD. Thank you, Bri, for the joy and energy you always bring to lab, and for all that you do to improve the graduate experience at MIT. And thank you, Mariann, for always helping me out and keeping our lab running so smoothly. Finally, my research work would not have been possible without the help of several core facilities at the Koch Institute (KI), the Biophysical Instrumentation Facility (BIF), and the Department of Chemistry Instrumentation Facility (DCIF). Thank you, Glenn Paradis, Michele Griffin, Mike Jennings (KI Flow), Brad Turner (BIF), Jeffrey Kuhn, Jeff Wyckoff (KI Microscopy), and Walter Massefski (DCIF), for maintaining the instruments and helping me analyze the data that are critical to my research. And thank you, Melinda, for keeping our facilities in top shape and for all of our fun chats.

As my fellow chemical engineering classmates will know, getting through the first year of the graduate curriculum was leaps and bounds more challenging than we could have imagined. Thank you to each and every one of you for your support and companionship. Thank you, Grace and Leerang, for always organizing a potluck, BBQ, or coffee break to look forward to, for encouraging me to be just a tad more outdoorsy, and for the deep care and love you have for the people around you. Thank you, Billy, for always being so distractingly fun and chatty – it was a real bummer when you moved to the West Coast but I always look forward to seeing you. Thank you, Allen, for being such an easygoing, supportive, and goofy roommate and friend; the way you breezed through the first year of graduate school will never fail to impress me. Thank you, Elvin, Hao-Wei, Marc, Hyunhee, Lauren, Pradeep, and TJ for your wholesome and genuine friendship. I always look forward to our get-togethers and I cannot wait for many more to come. Thank you, Kevin, for introducing me to

the world of points and miles, which has enabled me to affordably fulfill my passion for travel while on a PhD stipend. I am also grateful to have worked with such passionate and inspiring individuals at the MIT Biotech Group: Gurrein, Desmond, Josh, Keegan, Lena, Megan, Meilin, and Morgan. And thank you to my fellow Practice School colleagues, Mo, Lorenzo, and Yuan, for the fun experiences in Houston and Dubai. The friendships I have developed during my time in Boston extend beyond the MIT community. Thank you to my collaborator on the clinical machine learning project, Guangya, for our fun conversations and for patiently teaching me all you know about clinical data science and machine learning. And thank you, Alex K., Alex Z., Brandon, Dan, Jon, Noah, Kevin, Sam, and Tyler, for all of the Friday night games and shows and laughter.

I also want to acknowledge friends and mentors who knew me before graduate school. Thank you to my Orpen House flatmates, Carolyn and Sue Yi, for being the most fun and ridiculous (in the best way) duo I know; I'll always miss your bombardment of jokes and insults. Thank you, Michelle, for always being so genuinely concerned and caring about my well-being, and Ben, for always introducing me to incredible restaurants, bakeries, and cafes, in London and beyond. Thank you, Adam, for being an awesome research partner and friend. Thank you, Sandy, for being your quirky, funny, and artistic self – I am grateful for the many meals we shared at JJ's and Senya, and the cool artsy New York things you introduced me to. Thank you, Marie, for always being so caring and supportive, and for encouraging me to be more comfortable in my own skin. Thank you, Nat, for always reminding me of how much I'm missing out by not being on the West Coast, and for always telling it how it is. Thank you, Li Vern, for being by my side throughout high school, undergrad, and graduate school – your deadpan humor will always crack me up. And thank you, Sam and Angela, for your continued support, candid advice, and constant reminders to enjoy life outside of work.

Importantly, I want to thank my parents for all that they have sacrificed to help me get here today. They have provided far more than I could have ever asked for in pursuit of my passions and dreams. I hope that they can breathe a sigh of relief now

that I'm *finally* done with school. And thank you to my sister Steph and brother-in-law Eric for all the meals, celebrations, Costco/Super 88 grocery runs, homemade focaccia, Tony's, love, advice, and support over these past five years. I cannot wait to over-eat with you both again soon. I am also deeply grateful to my brother, Michael, for always being by my side since we were kids; I always appreciate hearing your honest advice and passionate rants. I would be remiss if I did not thank my extended "Quaran-team" family – Jim, Alfio, and Connie. One day, I hope to host dinner parties as incredible as yours – you've set the bar so high. And I hope to show others the amount of kindness and support you have shown me. And of course, I want to thank Luke, for being with me every step of the way, through all the ups and downs of graduate school and beyond. Thank you for always being so encouraging and supportive when I am faced with frustrations and failure, for celebrating my victories and sharing yours, for exposing me to funky cheeses and the latest judicial opinions, for keeping our apartment so impeccably clean and tidy, and for sharing my interest in food and travel. Thank you as well, Amy, Reed, Arthur, Nidhi, and Saachi, for all the love, support, and encouragement you have shown me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cancer is a complex disease that arises from genetic mutations that result in abnormal cell growth. Traditional treatments for cancer include chemotherapy and radiation therapy, which can cause significant side effects due to their non-specific nature. In recent years, targeted therapies have emerged as a promising approach to cancer treatment. These therapies are designed to specifically target molecules that are involved in the development and progression of cancer while sparing healthy cells.

The advent of targeted therapies in the treatment of cancer is based on a better understanding of the molecular mechanisms that drive cancer development. Cancer cells often have mutations that cause them to produce proteins that are not present in normal cells or that are produced at much higher levels. These proteins, also known as targets, are essential for the survival and growth of cancer cells. Targeted therapies aim to inhibit the function of these proteins or the signaling pathways they activate, thereby preventing the growth and spread of cancer cells. While small-molecule targeted inhibitors have the ability to target both extracellular and intracellular targets, protein-based targeted inhibitors such as monoclonal antibodies are generally limited to extracellular targets.

One example of a target in cancer therapy is RAS, a family of intracellular proteins that play a crucial role in cell signaling pathways. Mutations in RAS genes are present in about 30% of all human cancers [3, 4], and are associated with increased tumor growth, invasion, and drug resistance. KRAS represents the most frequently mutated

isoform in RAS-driven cancers (86%) [5]; among patients with cancers of the lung, colon, and pancreas — three of the deadliest forms of cancer — 31%, 45%, and 95% of them harbor mutations in KRAS, respectively. Given its prevalence, there have been significant efforts to develop RAS inhibitors, albeit with limited success. For decades, RAS proteins have been deemed "undruggable" due to the absence of binding pockets for small-molecule drugs [6], and the significant challenge of delivering protein-based inhibitors across the cell membrane to reach intracellular RAS.

KRAS mutations are dominated by single-base missense mutations, with 89% of mutations occurring at codon 12 (G12), 9% at codon 13 (G13), and 1% at codon 61 (Q61) [7]. Among G12 mutations, the most prevalent is the G12D mutation, accounting for approximately 36% of all KRAS mutations, followed by G12V at 23% and G12C at 14%. This means that roughly 8% [1] of human cancers are driven by a KRAS G12D mutation.

The discovery of an allosteric site in KRAS G12C has led to the development of several small-molecule KRAS G12C inhibitors, two of which were FDA-approved within the last two years. However, other forms of RAS including KRAS G12D remain largely undruggable. Protein-based KRAS G12D inhibitors, such as R11.1.6 developed by the Wittrup Lab in 2017, have shown high affinity and specificity for KRAS G12D. However, the genetically encoded intracellular expression of R11.1.6 in tumor cells failed to disrupt downstream RAS signaling. These results prompted us to develop degrader-based strategies against KRAS G12D, which involves the recruitment of an E3 ubiquitin ligase (E3) to induce degradation of the KRAS G12D target.

We first developed a cell-permeable KRAS-G12D-targeting degrader by (1) conjugating a small-molecule E3 ligand (VL1) to R11.1.6, then (2) chemically modifying the protein via bioreversible esterification, a permeation technique developed by the Raines Lab in 2017 [8]. To esterify a protein, the carboxyl groups of the protein are esterified by reaction with a hydrophobic diazo compound, enabling entry of the protein into the cytosol of a mammalian cell, where endogenous esterases hydrolyze the

[1] $30\% \times 86\% \times 89\% \times 36\% = 8\%$

nascent ester groups. The low aqueous solubility of esterified proteins was, however, a major practical challenge. In chapter 2, we address this challenge by developing a solubilization strategy for esterified proteins. Through a solubility screen, we found that the addition of the macrocycle β-cyclodextrin (β-CD) can increase the recovery of esterified green fluorescent protein (est-GFP) by 10-fold. We confirmed via ^{1}H NMR titration experiments that β-CD encapsulates the hydrophobic tolyl group of ester conjugates with $K_a = 321\ M^{-1}$, and discovered a correlation between K_a and est-GFP recovery among a variety of macrocycles. Combining L-arginine and sucrose with β-CD also enabled the nearly quantitative recovery of est-GFP, demonstrating a potentially generalizable strategy to overcome the insolubility of esterified proteins. Chapter 2 was a collaborative effort between the Wittrup lab (Keith Cheah) and the Raines lab (Vicky Joomyung Jun).

In chapter 3, we discuss the development of esterified R11.1.6-VL1 (est-R11-VL1) and its ability to degrade KRAS G12D and cause growth inhibition of KRAS-G12-expressing human tumor cells. We confirmed the cytosolic entry of esterified R11.1.6 and demonstrated modest efficacy of est-R11-VL1 in degrading RAS (up to 40%) and inhibiting the growth (up to 35%) of certain cell lines. Even though the exogenous delivery of a large-molecule degrader was only modestly effective against KRAS G12D, our results support a novel paradigm for targeting previously undruggable proteins. Chapter 3 was a collaborative effort between the Wittrup lab (Keith Cheah) and the Raines lab (Aniekan Okon).

In chapter 4, we build on previous work on an alternative degrader-based approach against KRAS G12D known as biodegraders [9]. A biodegrader consists of the target binder attached via a peptide linker to the E3 ligase itself (instead of a small molecule E3 ligand), i.e. it is a full protein construct; KRAS degradation is induced via intracellular expression of the biodegrader construct. Currently, the design of degraders (whether small molecule or protein-based) remains a highly empirical process, primarily due to the large number of design features involved. The relative nascency of the degrader field also means that little is known about how each feature affects degradation efficiency. Through high-throughput fluorescent-based screening and re-

gression modeling, we developed several rational design principles by identifying and ranking several design features that are important predictors of KRAS G12D degradation efficiency. We also show that weakened affinity of R11.1.6 to KRAS G12D compromises degradation efficiency. Chapter 4 was a collaborative effort with Megan Hoffman (Wittrup Lab).

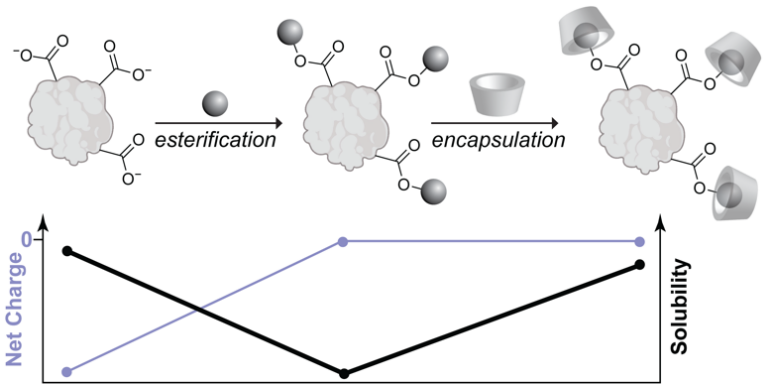
The development of targeted therapies for cancer has revolutionized cancer treatment by allowing for more specific and effective treatment options. However, the challenge of developing inhibitors for "undruggable" targets, such as KRAS G12D, remains a major obstacle. The development of degrader-based strategies, such as the cell-permeable KRAS G12D-targeting degrader described in this thesis, may offer a promising approach to overcoming this challenge. The solubilization strategy for esterified proteins presented in chapter 2 also has the potential for overcoming the practical challenges associated with the use of esterified proteins in various applications. The rational design principles developed in Chapter 4 have the potential to significantly impact the design of future biodegraders against other targets beyond KRAS G12D, enabling a more streamlined and efficient approach to designing degraders for a range of proteins. These developments highlight the importance of continued research in the field of targeted therapies for cancer.

Chapter 2

Host–Guest Complexation by β-Cyclodextrin Enhances the Solubility of an Esterified Protein

This chapter, with some modifications, was published in ACS Molecular Pharmaceutics in 2022 [10].

2.1 Abstract



The carboxyl groups of a protein can be esterified by reaction with a diazo compound, 2-diazo-2-(p-methylphenyl)-N,N-dimethylacetamide. This esterification enables the entry of the protein into the cytosol of a mammalian cell, where the nascent

ester groups are hydrolyzed by endogenous esterases. The low aqueous solubility of the ensuing esterified protein is, however, a major practical challenge. Solubility screening revealed that β-cyclodextrin (β-CD) is an optimal solubilizing agent for esterified green fluorescent protein (est-GFP). Its addition can increase the recovery of est-GFP by 10-fold. α-CD, γ-CD, and cucurbit-7-uril are less effective excipients. ^{1}H NMR titration experiments revealed that β-CD encapsulates the hydrophobic tolyl group of ester conjugates with $K_a = 321\ M^{-1}$. Combining L-arginine and sucrose with β-CD enables the nearly quantitative recovery of est-GFP. Thus, the insolubility of esterified proteins can be overcome with excipients.

2.2 Introduction

α-Aryl-α-diazoacetamide **1** has been shown to esterify the carboxyl groups of proteins in water and enable their vectorial delivery across cellular membranes [11, 8, 12].The key attribute of this reagent is its tuned basicity, which allows the ready abstraction of a proton from a carboxylic acid but not from water. This permeation strategy is analogous to that of ester prodrugs, wherein carboxyl groups are "cloaked" as esters that are hydrolyzed by endogenous esterases upon entry into the cytosol [13, 14, 15, 16]. Likewise, the ability to cloak protein carboxyl groups enables their cytosolic delivery in a traceless manner.

At any given pH, the net charge on a protein is determined by the pK_avalues of its ionizable groups [17]. The net charge on a protein is zero at its isoelectric point (pI), positive at pHs below its pI, and negative at pHs above its pI. The value of log Csat, where Csat is the concentration of a saturated solution, increases with the square of the net charge on the protein [18, 19, 20]. Consequently, proteins tend to be least soluble at pHs near their pI.

The esterification of an anionic protein with an α-aryl-α-diazoacetamide compromises solubility by increasing its pI to near-physiological pH. For example, the green fluorescent protein (GFP) has Z = -9 (where Z refers to the number of Arg+Lys-Asp-Glu). Moreover, esterification with compound 1 replaces a charged functional

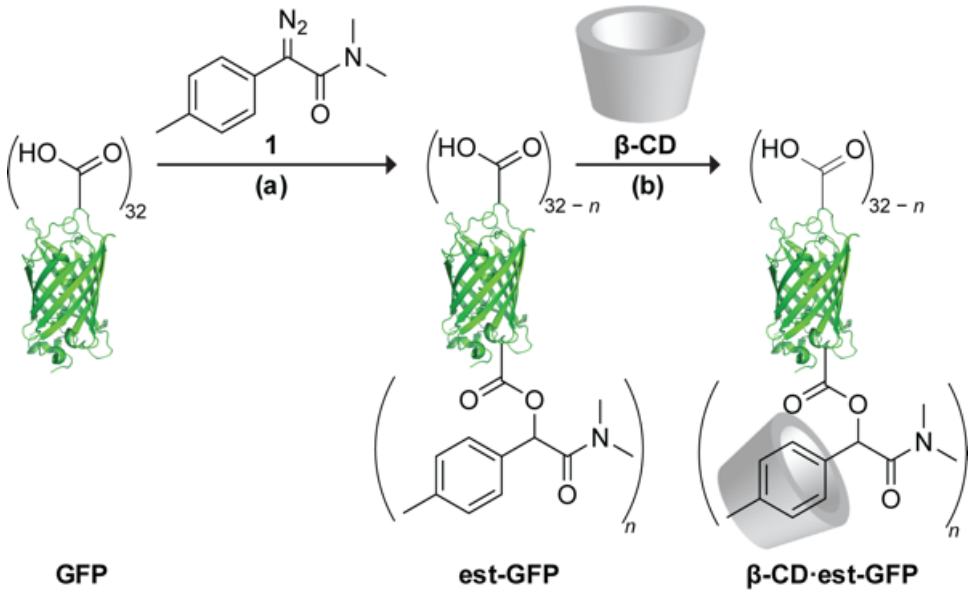


Figure 2-1: Schematic representation of the esterification of GFP with diazo compound **1** to yield esterified GFP (est-GFP) and its encapsulation by β-CD. (a) 10 mM Bis-Tris buffer, pH 6.5, containing CH_3CN (20% v/v) for 4 h at 37 °C; (b) 5 mM β-CD in PBS, pH 7.3 at room temperature.

group with a hydrophobic moiety (Figure 2-1). Accordingly, the protein esterification is confounded by insolubility (vide infra). Herein, we use esterified GFP (est-GFP) as a model protein to generate a solubilization strategy for esterified proteins. We reasoned that cyclodextrins (CDs [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38]) and cucurbiturils (CBs [39, 40, 41, 42, 43]), which are known to solubilize proteins and peptides, could likewise solubilize est-GFP. CDs are a family of cyclic oligosaccharides consisting of glucose subunits linked by α-1,4 glycosidic bonds. The three most common CDs are α-CD, β-CD, and γ-CD, which have six, seven, and eight glucose subunits, respectively. Cucurbit[n]urils (CBn) are composed of glycoluril monomers linked by methylene bridges, where n is the number of glycoluril units. The common cucurbiturils (CBs) are CB5, CB6, CB7, and CB8. Both CBs and CDs have a hollow structure consisting of a hydrophilic exterior and a hydrophobic cavity. We find that the addition of β-CD greatly enhances the aqueous solubility of the conjugated protein, est-GFP, by encapsulating the hydrophobic aryl moiety of the ester conjugates (Figure 2-1).

2.3 Results and Discussion

2.3.1 Solubility Screening

To esterify GFP, the protein was mixed with α-aryl-α-diazoacetamide 1 in Bis-Tris buffer containing 20% v/v acetonitrile. Est-GFP was then exchanged into PBS. The yield of this buffer-exchange step was only (5 ± 1)%, motivating the need for a solubilization strategy.

A variety of additives have been shown to increase the solubility of proteins. [44, 45, 46, 47] We performed an initial solubility screen to determine which of them enhance the aqueous solubility of est-GFP. The experimental design of this solubility screen is shown in Figure 2-2A. Precipitated est-GFP was removed by filtration. The concentration of est-GFP was determined by measuring the absorbance at 488 nm. The number of esters per protein was determined by matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) mass spectroscopy. On average, 11 of the 32 carboxyl groups of GFP were esterified with diazo compound 1, and the yield of this reaction was (84 ± 4)%. est-GFP was then split equally and exchanged into PBS or additive-supplemented PBS, which were adjusted to pH 7.3. The concentrations of additives were chosen based on values reported in the literature. [47] Precipitated est-GFP was removed by filtration, and the concentration and number of esters per protein were determined by mass spectrometry. For each additive, the recovery was calculated as the amount of est-GFP that remained soluble in additive-supplemented aqueous buffer (PBS) as a percentage of the amount of est-GFP pre-exchange (Figure 2-2B).

The macrocycles cyclodextrins and cucurbiturils were chosen based on their potential capability to mask the hydrophobic moieties that decorate the surface of est-GFP. Specifically, we chose to test a family of cyclodextrins, α-CD, β-CD, and γ-CD, as well as cucurbit[7]uril (CB7). Among the family of CBs (CB5, CB6, CB7, and CB8), only CB7 was studied because CB5 has too small of a cavity to form inclusion host-guest complexes, [48] and CB6 and CB8 have poor aqueous solubility. The cavity sizes of α-CD, β-CD, and γ-CD, and CB7 are listed in Table 2.2.

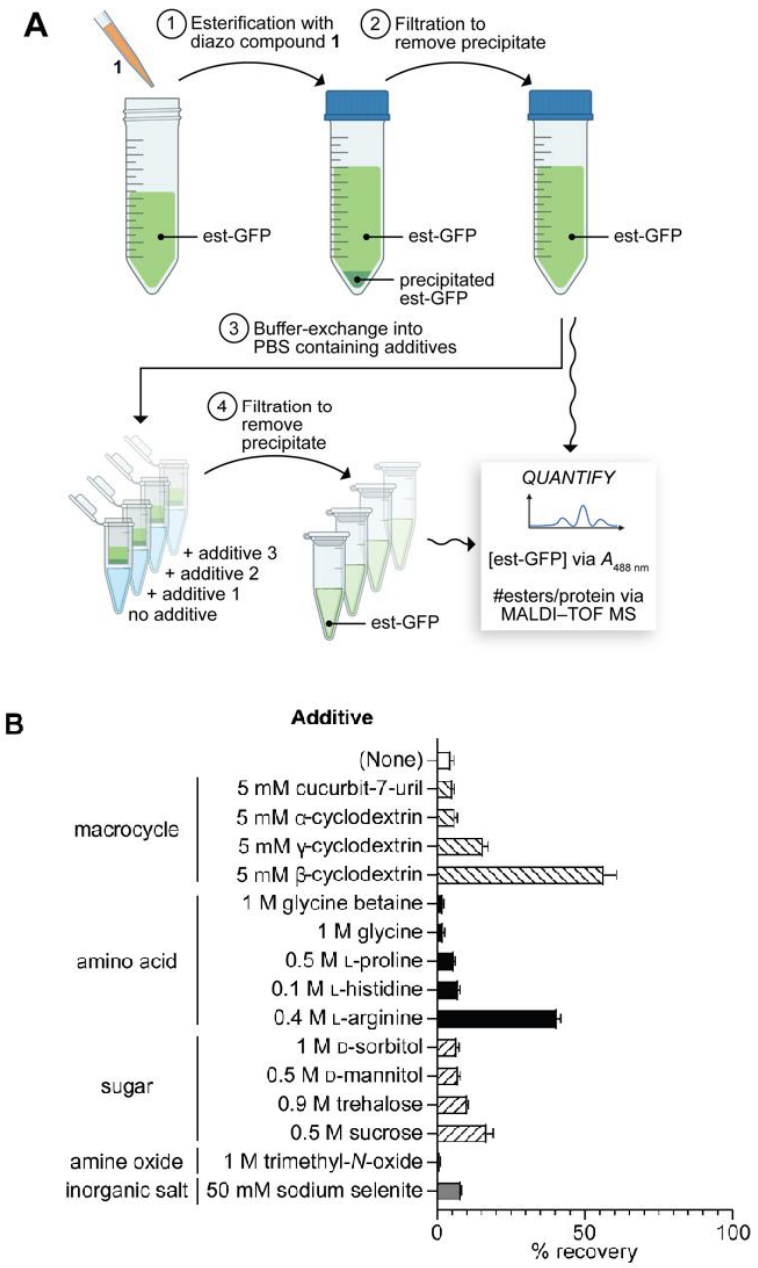


Figure 2-2: Solubility screening of additives. (A) Flowchart of the experimental design. (B) Bar graph of the recovery of est-GFP obtained with different additives. % Recovery is defined as the amount of est-GFP that remains soluble after step 4 in (A) as a percentage of the amount of est-GFP after step 2. Experiments were performed in duplicate.

The osmolytes glycine betaine [49, 50, 51, 52, 53] and trimethyl-N-oxide (TMAO), [51, 54, 55, 56, 57] which have been used previously as stabilizing additives, resulted in lower recovery compared to the no-additive control at $(5 \pm 1)\%$; this reduction in

recovery was also observed with the amino acid glycine. Among other top-performing additives in the literature (e.g., L-proline, [49, 52, 58, 59, 60] L-arginine, [61, 62, 63, 64] D-sorbitol, [50, 65, 66, 67] D-mannitol, [50] trehalose, [57, 65, 50, 68, 69, 70, 71] and sucrose [65, 72, 73, 74]), only L-arginine conferred a greater than 2-fold increase in recovery at $(41 \pm 1)\%$. Its use as an aggregation suppressor has been well documented in the literature. [61, 62, 63, 64] As a neutral crowder, L-arginine has an affinity for isolated protein molecules, thus increasing the free energy of protein-protein assemblies. [61] The macrocycles showed the greatest overall increase in the recovery of est-GFP, with β-CD achieving more than a 10-fold increase in recovery over the no-additive control. Finally, among the sugars and sugar alcohols, only sucrose, a well-known stabilizing agent, [65, 72, 73, 74] resulted in a greater than 2-fold increase in recovery. Sucrose molecules are preferentially excluded from the protein surface and thus raise the free energy of the unfolded state. [72] In the est-GFP that was recovered in the final step, greater recovery was associated with a larger average number of ester labels per protein (as determined by MALDI-TOF mass spectrometry). For example, est-GFP exchanged into PBS had a median of only seven labels (versus 11 labels pre-exchange), whereas est-GFP exchanged into PBS that was supplemented with 5 mM β-CD had a median of nine labels. These results suggest that est-GFP molecules with a larger number of hydrophobic ester labels were precipitating and that the addition of certain additives was keeping highly labeled est-GFP in solution. Overall, the additive screen demonstrated the difficulty in predicting suitable solubility enhancers a priori and motivated the need for further characterization.

2.3.2 Interactions of est-MGA with β-CD

Given the significant enhancement in the aqueous solubility of est-GFP in PBS supplemented with β-CD, we sought to explore the interactions, if any, between ester groups and β-CD. Because β-CD has the ability to encapsulate hydrophobic aromatic groups, we hypothesized that β-CD forms inclusion complexes with the tolyl groups that decorate the surface of est-GFP (Figure 2-1). To test this hypothesis,

we conducted ^{1}H NMR titration experiments. We synthesized a proxy for est-GFP by esterifying a small water-soluble carboxylic acid, O-(2-methoxyethyl)glycolic acid (MGA), with diazo compound 1. The product, est-MGA, is readily soluble in D_2O. The NMR titration experiments were performed in D_2O containing est-MGA (3.33 mM) and increasing concentrations of β-CD (0.00-3.00 equiv). The resulting NMR spectra are shown in Figure 2-3B.

The chemical shift results (Table 2.1) confirmed host-guest interactions between β-CD and the benzylic (H1, H4) and aromatic protons (H2, H3) of est-MGA. The chemical shift for H1 moved upfield as the concentration of β-CD increased, consistent with greater shielding due to the encapsulation of H1 in the β-CD cavity. Likewise, the chemical shifts of aromatic protons H2 and H3 started to diverge, indicative of distinct chemical environments upon complexation with β-CD. This divergence was also apparent with benzylic proton H4, as the gradual addition of β-CD causes that single proton resonance peak to split into two peaks of equal integrated intensities.

Table 2.1: ^{1}H NMR Chemical Shift Values for the Complexation of β-CD with est-MGA in D_2O (400 MHz)

[β-CD]/[est-MGA]	**H1**	$\Delta\delta^a$	**H2**	$\Delta\delta^a$	**H3**	$\Delta\delta^a$	**H4**	$\Delta\delta^a$
0.00	2.284		7.265		7.276		6.377	
0.10	2.282	-0.002	7.258	-0.008	7.277	0.001	6.374	-0.003
0.25	2.279	-0.005	7.245	-0.020	7.278	0.002	6.369	-0.008
0.50	2.275	-0.009	7.228	-0.037	7.282	0.006	6.362	-0.015
1.00	2.269	-0.015	7.202	-0.063	7.287	0.011	6.351	-0.026
1.50	2.264	-0.020	7.181	-0.085	7.291	0.015	6.344	-0.033
2.00	2.261	-0.022	7.168	-0.097	7.293	0.017	6.338	-0.039
3.00	2.257	-0.027	7.150	-0.115	7.296	0.020	6.330	-0.047

[a] $\Delta\delta = \delta(\text{complex}) - \delta(\text{free})$

Because proton H4 is at the stereogenic center of est-MGA, the splitting of this peak suggests increasingly distinct interactions between each stereoisomer and β-CD

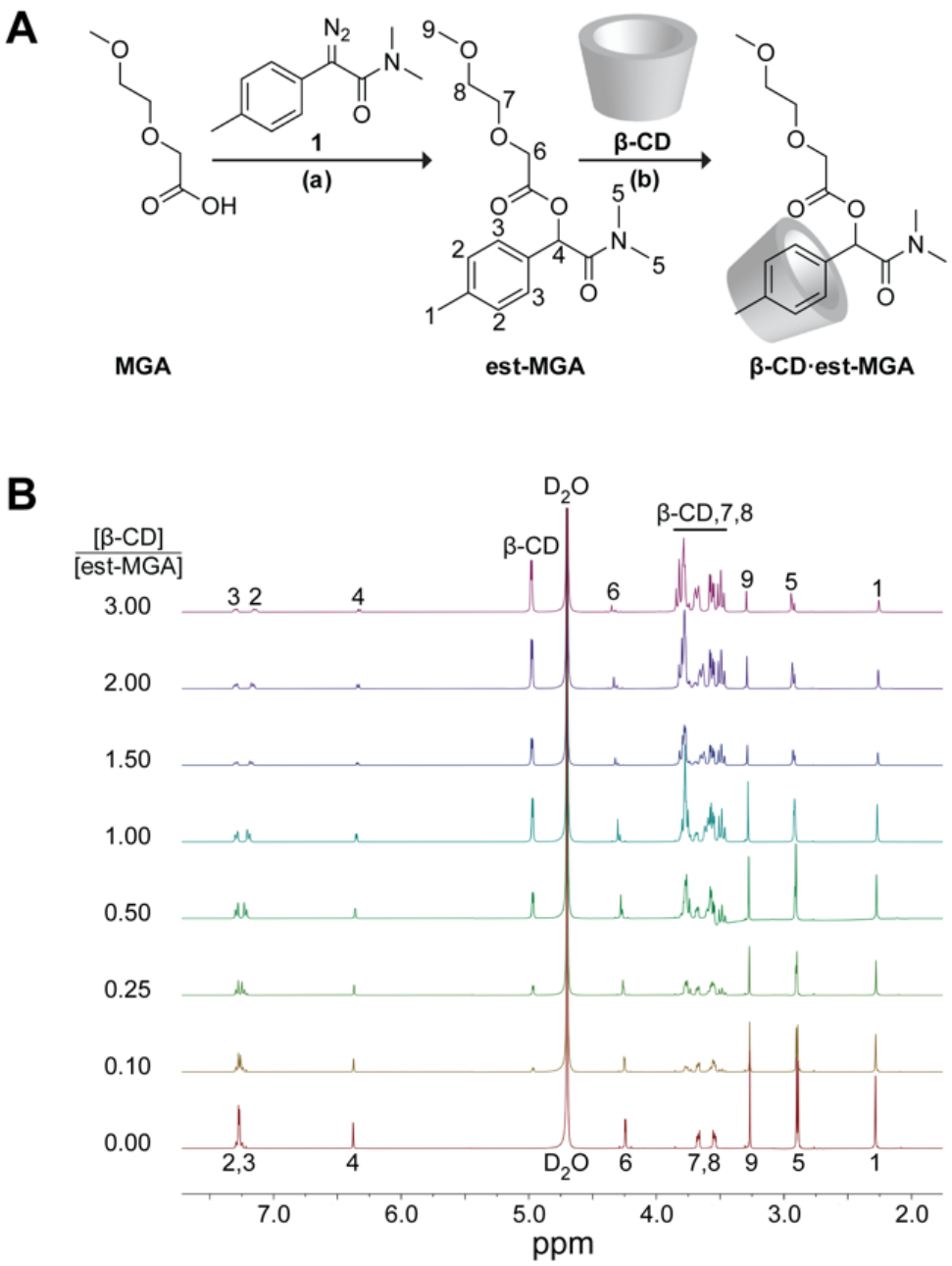


Figure 2-3: (A) Schematic representation of the esterification of MGA with diazo compound 1 to yield esterified MGA (est-MGA), which is a water-soluble small-molecule proxy for est-GFP, and its encapsulation by β-CD. (a) 1:1 CH_3CN/MES-HCl buffer (pH 6.0), at room temperature; (b) 3.33 mM est-MGA with β-CD (0.00-3.00 equiv) in D_2O at room temperature. (B) ^{1}H NMR spectra upon the addition of β-CD (0.00-3.00 equiv) to a 3.33 mM solution of est-MGA in D_2O. The numbering of est-MGA protons is depicted in (A).

(Figure 2-16). The chemical shifts of aliphatic protons (e.g., H5, H6, H7, H8, and H9) remained downfield with increasing equivalents of β-CD, confirming that the

glyme chain does not embed itself within the cavity of β-CD (Figures 2-25 and 2-26). Finally, a binding constant of $K_a = 321 \pm 4\ M^{-1}$ (i.e., $K_d = 3.11 \pm 0.04$ mM) was generated from curves of the chemical shifts of encapsulated protons H1, H2, H3, and H4 of est-MGA versus [β-CD] with Thordarson's fitting program. [75]

The encapsulation of the tolyl group of est-MGA within the β-CD cavity was confirmed with ^{1}H-^{1}H ROESY NMR spectroscopy (Figures 2-23 and 2-24). The 2D spectrum (Figure 2-24) indicates the existence of cross-peaks between the aromatic protons (H2, H3) of est-MGA and the cavity protons (Hx, Hy) of β-CD. On the other hand, no cross-peaks were detected between the aliphatic protons of est-MGA (H5, H6, H7, H8, H9) and the cavity protons (Hx, Hy) of β-CD.

2.3.3 Interactions of est-MGA with Other Macrocycles

Table 2.2: Affinity of Macrocycles for est-MGA and Recovery of est-GFP

Additive	**Cavity Diameter (Å)**[a,b]	**K_a (M^{-1})**[c]	**K_d (mM)**[c]	**[Additive] (mM)**	**est-GFP Recovery (%)**[d]
None	N/A	N/A	N/A	N/A	5 ± 1
CB7	5.4[a]	69 ± 1	14.5 ± 0.2	5	5 ± 0
α-CD	4.7[b]	42.2 ± 0.2	23.7 ± 0.1	5	6 ± 1
β-CD	6.0[b]	321 ± 4	3.11 ± 0.04	5	57 ± 4
γ-CD	7.5[b]	84 ± 3	11.8 ± 0.4	5	16 ± 1
β-CD				5	
+ L-Arginine	N/A	N/A	N/A	250	96 ± 2
+ Sucrose				750	

[a] from ref [76]. [b] from ref [77]. [c] Determined by the nonlinear 1:1 fitting of the NMR peak positions of protons H1, H2, H3, and H4 versus the concentration of macrocycle. [d] Determined by the solubility screen depicted in Figure 2-2.

Upon confirming the encapsulation of the tolyl group of est-MGA within the β-CD cavity, we sought to determine if similar interactions would be observed with other

macrocycles and whether a higher binding affinity between a particular macrocycle and est-MGA corresponds to an enhanced ability of the macrocycle to solubilize est-GFP. Accordingly, we repeated the NMR titration experiments with α-CD, γ-CD, and CB7. The chemical shift results (Tables 2.3, 2.4, and 2.5) confirmed encapsulation of the tolyl group of est-MGA within the hydrophobic cavities of the larger macrocycles, γ-CD, and CB7, but not α-CD. For both γ-CD and CB7, the peaks of the benzylic proton H1 and the aromatic protons (H2 and H3) of est-MGA were all upfield, with the aromatic proton peaks also exhibiting splitting similar to that seen with β-CD. Interestingly, splitting and an upfield shift of the H4 proton peak of est-MGA were observed with γ-CD but not with CB7, suggesting a shallower encapsulation of the tolyl group by CB7; this observation is consistent with the smaller cavity size of CB7 compared to those of β-CD and γ-CD. Finally, for both γ-CD and CB7, the aliphatic protons (H5, H6, H7, H8, H9) of est-MGA exhibited either a downfield shift or none at all, confirming that only the tolyl group is being encapsulated. As with β-CD, the association constants were determined for γ- CD and CB7 using the chemical shifts of protons H1, H2, H3, and H4 (Table 2.2). The values of $K_a$for the association of est- MGA with γ-CD and CB7 were 84 and 69 M^{-1}, respectively.

With α-CD, the smallest of the CDs, all nine proton peaks of est-MGA exhibited a downfield shift with increasing equivalents of the macrocycle, suggesting the complete lack of encapsulation of est-MGA within the α-CD cavity. For consistency, the association constant of α-CD with est-MGA was determined using the downfield chemical shifts of protons H1, H2, H3, and H4, as with the other three macrocycles. Among the four tested macrocycles, α-CD had the lowest affinity with a $K_a$value of 42 M^{-1}.

Having determined the binding affinities of four macrocycles with est-MGA, we sought to determine if the affinity of a macrocycle for est-MGA correlates with its ability to solubilize est-GFP. The ability to solubilize est-GFP increased in the order CB7 $\approx$ α-CD $<$ γ-CD $<$ β-CD (Table 2.2), with CB7 and α-CD demonstrating an insignificant enhancement in solubility. Thus, $K_a$values do correlate with solubilizing ability.

Table 2.3: ^{1}H NMR Chemical Shift ($\Delta\delta$) Values for the α-CD·est-MGA Complex

[α-CD]/[est-MGA]	**H1**	$\Delta\delta^a$	**H2/H3**	$\Delta\delta^a$	**H4**	$\Delta\delta^a$
0.00	2.289		7.276		6.382	
0.10	2.289	0.001	7.278	0.002	6.383	0.001
0.25	2.291	0.002	7.281	0.006	6.385	0.003
0.50	2.293	0.004	7.287	0.011	6.387	0.005
1.00	2.297	0.008	7.297	0.021	6.392	0.010
1.50	2.300	0.012	7.305	0.030	6.396	0.014
2.00	2.304	0.015	7.314	0.038	6.400	0.018
3.00	2.309	0.021	7.328	0.052	6.406	0.024
4.00	2.314	0.025	7.340	0.064	6.412	0.030

[a] $\Delta\delta = \delta(\text{complex}) - \delta(\text{free})$

Table 2.4: ^{1}H NMR Chemical Shift ($\Delta\delta$) Values for the γ-CD·est-MGA Complex

[γ-CD]/[est-MGA]	**H1**	$\Delta\delta^a$	**H2**	$\Delta\delta^a$	**H3**	$\Delta\delta^a$	**H4**	$\Delta\delta^a$
0.00	2.284		7.265		7.277		6.378	
0.10	2.282	-0.002	7.264	-0.002	7.276	-0.001	6.376	-0.002
0.25	2.280	-0.005	7.260	-0.005	7.275	-0.001	6.374	-0.004
0.50	2.275	-0.009	7.257	-0.008	7.275	-0.002	6.370	-0.008
1.00	2.268	-0.016	7.250	-0.015	7.272	-0.004	6.363	-0.015
1.50	2.262	-0.022	7.245	-0.021	7.270	-0.007	6.356	-0.021
2.00	2.258	-0.026	7.241	-0.025	7.269	-0.007	6.351	-0.026
3.00	2.251	-0.033	7.233	-0.032	7.267	-0.010	6.341	-0.037
4.00	2.247	-0.038	7.228	-0.038	7.265	-0.012	6.332	-0.046

[a] $\Delta\delta = \delta(\text{complex}) - \delta(\text{free})$

For completeness, we explored the utility of combinations of additives to enhance the solubilization of est-GFP. Through additional solubility screens, we found that

Table 2.5: ^{1}H NMR Chemical Shift ($\Delta\delta$) Values for the CB7·est-MGA Complex

[CB7]/[est-MGA]	**H1**	$\Delta\delta^a$	**H2**	$\Delta\delta^a$	**H3**	$\Delta\delta^a$	**H4**	$\Delta\delta^a$
0.00	2.286		7.266		7.278		6.379	
0.10	2.282	-0.004	7.262	-0.005	7.276	-0.002	6.379	0.000
0.25	2.276	-0.009	7.254	-0.013	7.273	-0.005	6.379	0.000
0.50	2.266	-0.020	7.240	-0.027	7.267	-0.011	6.378	-0.001
1.00	2.247	-0.039	7.213	-0.053	7.255	-0.023	6.378	-0.001
1.50	2.231	-0.054	7.192	-0.075	7.247	-0.031	6.377	-0.002
2.00	2.219	-0.067	7.175	-0.092	7.240	-0.038	6.377	-0.002
3.00	2.199	-0.087	7.147	-0.119	7.228	-0.050	6.377	-0.002
4.00	2.179	-0.106	7.120	-0.146	7.217	-0.061	6.376	-0.003

[a] $\Delta\delta = \delta(\text{complex}) - \delta(\text{free})$

the addition of L-arginine and sucrose, that is, the most effective additives among the amino acids and sugars, respectively (Figure 2-2), significantly enhanced the recovery of est-GFP (Table 2.2). Given that all three additives — β-CD, L-arginine, and sucrose — have differing solubilizing mechanisms, we were not surprised that they worked together to enhance protein solubility. When exchanged into PBS supplemented with β-CD (5 mM), L-arginine (250 mM), and sucrose (750 mM), the recovery of est-GFP was 96%. That represents a dramatic increase upon the 5% recovery with PBS alone (Table 2.2). On average, 10 of the 32 carboxyl groups of GFP remain esterified (versus 11 pre-exchange), as determined by MALDI-TOF mass spectrometry. Finally, we confirmed that the inclusion of these additives alters neither the extinction coefficient of GFP (Table 2.6) nor the efficiency of cellular uptake (Figure 2-4).

2.4 Conclusions

Our data demonstrate that the esterification of a protein molecule with diazo compounds followed by exchange into an aqueous solution supplemented with β-CD pro-

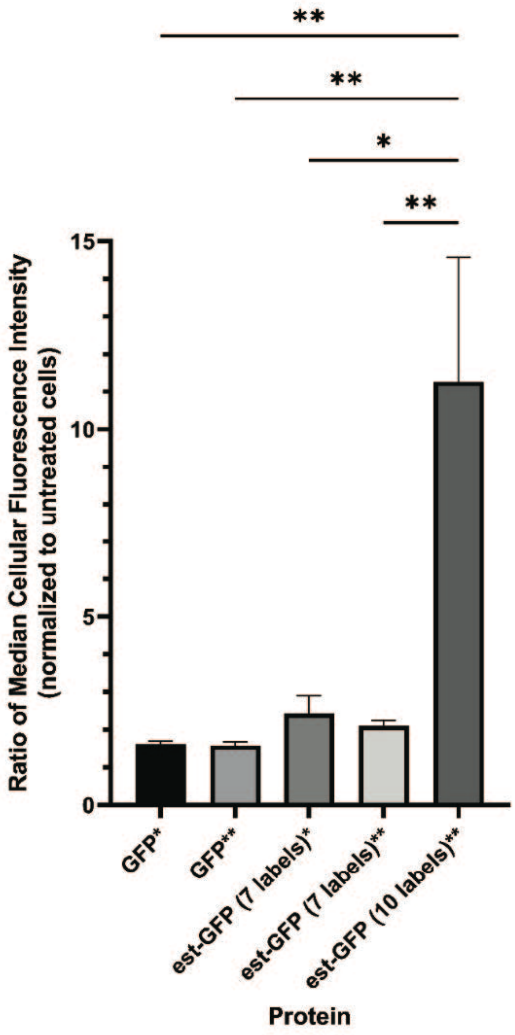


Figure 2-4: Cellular internalization of unmodified GFP and est-GFP. Data were obtained with flow cytometry. CHO-K1 cells were incubated for 4 h at 37 °C with (1) unmodified GFP, (2) esterified GFP with a median of 7 ester labels, or (3) esterified GFP with a median of 10 ester labels in either additive-free (*) or additive-supplemented (**) medium. Median cellular fluorescence intensities are normalized to that of untreated cells. Values are the mean ± SD with two technical replicates per protein. Statistical analysis by ordinary one-way ANOVA. *$p \leq 0.05$, **$p \leq 0.01$.

vides a general means to esterify the carboxyl groups of a protein without compromising its solubility. Furthermore, we show that the ability of a particular macrocycle to solubilize a conjugated protein can be determined by studying interactions between the macrocycle and a small-molecule mimetic of the conjugate (instead of the entire protein molecule). This strategy of using macrocycles as solubilizing agents could

also extend to esterified proteins formed with other diazo compounds. [78]

2.5 Materials and Methods

Materials. Reagent chemicals were obtained from commercial sources and used without further purification. Phosphate-buffered saline (PBS) contained Na_2HPO_4 (5.60 mM), KH_2PO_4 (1.06 mM), and NaCl (154 mM) at pH 7.3. 2-Diazo-2-(p-methylphenyl)-N,N-dimethylacetamide (1) was synthesized as described previously. [78] Esterified O-(2-methoxyethyl)glycolic acid (est-MGA) served as a proxy for est-GFP and was synthesized as described below.

Conditions. All procedures were performed in air at ambient temperature (∼22 °C) and pressure (1.0 atm) unless indicated otherwise.

Mass Spectrometry. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectra for protein characterization were acquired with a microflex LRF instrument (Bruker).

Nuclear Magnetic Resonance (NMR) Spectroscopy. ^{1}H and ^{13}C NMR spectra were acquired with a Bruker Avance Neo 400 MHz or Bruker Avance Neo 500 MHz spectrometer at the MIT Department of Chemistry Instrumentation Facility (Figures S1-S19). Proton chemical shifts are reported in parts per million (ppm, δ scale) and are relative to residual protons in the deuterated solvent ($CDCl_3$: δ 7.26; D_2O-d_2: δ 4.79). Carbon chemical shifts are reported in parts per million (ppm, δ scale) and are relative to the carbon resonance of the solvent ($CDCl_3$: δ 77.16). $CDCl_3$ and D_2O-d_2 were from Sigma-Aldrich. Multiplicities are abbreviated as s (singlet), d (doublet), dd (doublet of doublets), and m (multiplet). We further assigned each peak of est-MGA and β-CD using heteronuclear multiple bond correlation (HMBC) and heteronuclear single quantum correlation (HSQC) experiments. Rotating-frame nuclear Overhauser effect spectroscopy (ROESY) spectra were acquired for the encapsulation study.

Protein Preparation. The "superfolder" variant of GFP was prepared as described previously. [8] The protein was dialyzed into 10 mM Bis-Tris buffer, pH 6.5, prior to esterification. The concentration of the protein was determined by measuring absorbance at 488 nm (A488 nm) with a NanoDrop 2000c spectrophotometer (Thermo Scientific), using the theoretical extinction coefficient of ϵ = 81,760 M^{-1} cm^{-1}.

Protein Esterification. GFP was esterified essentially as described previously. [8] Briefly, one unit volume of α-aryl-α-diazoacetamide 1 in acetonitrile (31.9 mM) was added to four unit volumes of 60 μM GFP in 10 mM Bis-Tris buffer, pH 6.5, and the resulting solution was incubated for 4 h at 37 °C. The ratio of moles of diazo compound to GFP is 133:1 ($\sim$4 equiv per GFP carboxyl group). Precipitated protein was removed by centrifugal filtration with a 0.22 μm cellulose acetate centrifuge tube filter (Corning Costar Spin-X). The concentration of the soluble, esterified protein (est-GFP) was determined by measuring absorbance at 488 nm (A488 nm) with a NanoDrop 2000c spectrophotometer (Thermo Scientific). The yield of GFP esterification was calculated by dividing the number of moles of est-GFP (concentration $\times$ volume) post-esterification by the number of moles of GFP pre-esterification. A MALDI-TOF mass spectrum of est-GFP was acquired, and the number of esters per protein was assigned from the mass of the peak with the highest relative intensity in the MALDI-TOF mass spectrum (Figure 2-5).

Solubility Screening. Additives were screened for their ability to solubilize est-GFP. Est-GFP (in 80% v/v 10 mM Bis-Tris buffer, pH 6.5, and 20% v/v acetonitrile) was exchanged into either PBS, pH 7.3, or additive-supplemented PBS, pH 7.3, with Amicon Ultra centrifugal filter tubes (Millipore) having a molecular weight cutoff of 10 kDa and capacity of 500 μL. A 100 μL aliquot of the est-GFP solution was added to the filter tube and diluted to 500 μL with the desired buffer. The buffer was forced through the membrane by centrifugal filtration at 4,300g until the volume of the retentate protein solution dropped to 100 μL. The protein solution was again

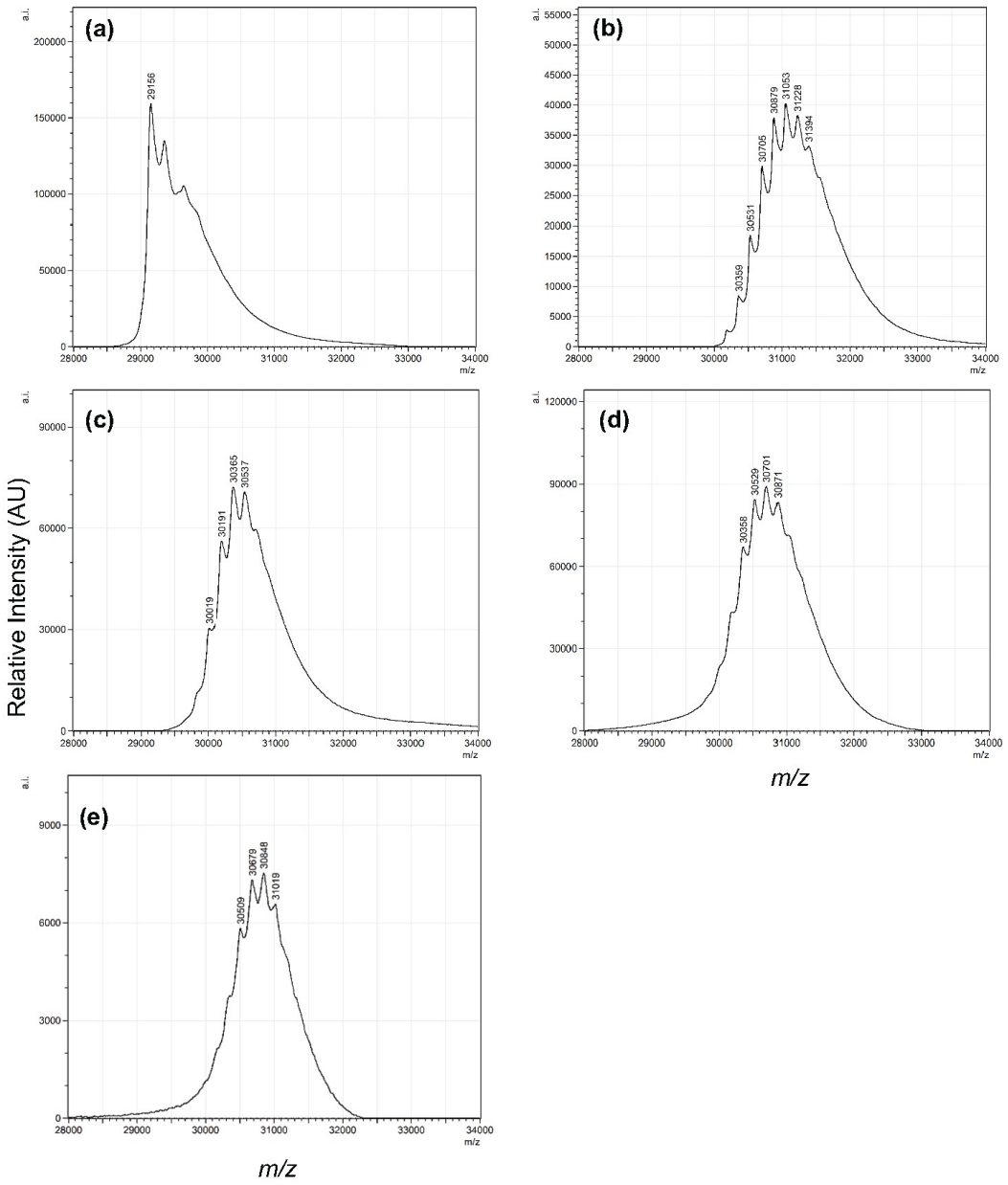


Figure 2-5: Representative MALDI–TOF spectra of unmodified GFP and est-GFP. (a) GFP; (b) GFP esterified with diazo compound 1 (est-GFP) in 10 mM Bis-Tris buffer, pH 6.5, containing CH_3CN (20% v/v); (c) est-GFP exchanged into PBS, pH 7.3; (d) est-GFP exchanged into PBS supplemented with 5 mM β-cyclodextrin; and (e) est-GFP exchanged into PBS supplemented with 5 mM β-cyclodextrin, 250 mM L-arginine, and 750 mM sucrose. Expected m/z: 29,343 + 175 per ester group.

diluted to 500 μL in the desired buffer and concentrated down to 100 μL. The dilution and centrifugation process was performed a total of four times (i.e., $54 = 625\times$

dilution). The retentate was collected by upside-down centrifugation at 1000g for 2 min. Precipitated protein was removed by centrifugal filtration with a 0.22 μm cellulose acetate centrifuge tube filter (Corning Costar Spin-X). The concentration of the soluble, esterified protein ("est-GFP") was determined by measuring the absorbance of the solution at 488 nm (A488 nm) with a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). A MALDI-TOF mass spectrum of est-GFP was acquired, and the number of esters per protein was assigned from the mass of the peak with the highest relative intensity in the MALDI-TOF mass spectrum. "% Recovery" was calculated by dividing the number of moles of est-GFP post-PBS-exchange by the number of moles of GFP pre-esterification.

Synthesis of est-MGA. To access est-MGA, α-aryl-α-diazoacetamide 1 (20 mg, 0.1 mmol, 1 equiv) was mixed with O-(2-methoxyethyl)glycolic acid (67 mg, 0.5 mmol, 5 equiv) in 990 μL of 1:1 acetonitrile/10 mM MES-HCl buffer, pH 6.0. The reaction mixture was stirred at room temperature until the color of the solution turned from red to colorless, indicative of the consumption of the diazo moiety by either esterification or hydrolysis reaction (Figure 2-6). Once the complete consumption of the starting material (1) was confirmed by liquid chromatography-mass spectrometry (LC-MS), the crude product was purified by reversed-phase chromatography on a C-18 column (12 g Sfär C18 D, Duo, 100 Å, 30 μm) (Figure 2-6). Fractions containing est-MGA were collected and lyophilized overnight to afford est-MGA as a clear oil (15.1 mg, 0.05 mmol, 49%).

NMR of est-MGA are shown in Figures 2-7 to 2-11. ^{1}H NMR (400 MHz, D_2O, δ): 7.27 (d, J = 3.3 Hz, 4H), 6.38 (s, ^{1}H), 4.24 (d, J = 2.5 Hz, 2H), 3.76-3.65 (m, 2H), 3.60-3.51 (m, 2H), 3.27 (s, 3H), 2.89 (d, J = 4.8 Hz, 6H), 2.28 (s, 3H). ^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.42-7.32 (m, 2H), 7.22 (d, J = 7.9 Hz, 2H), 6.27 (s, ^{1}H), 4.37-4.22 (m, 2H), 3.79 (dd, J = 5.2, 3.9 Hz, 2H), 3.60 (dd, J = 4.6, 2.0 Hz, 2H), 3.39 (s, 3H), 2.97 (d, J = 12.9 Hz, 6H), 2.38 (s, 3H), 1.78 (s, ^{1}H). ^{13}C NMR (101 MHz, D_2O, δ):

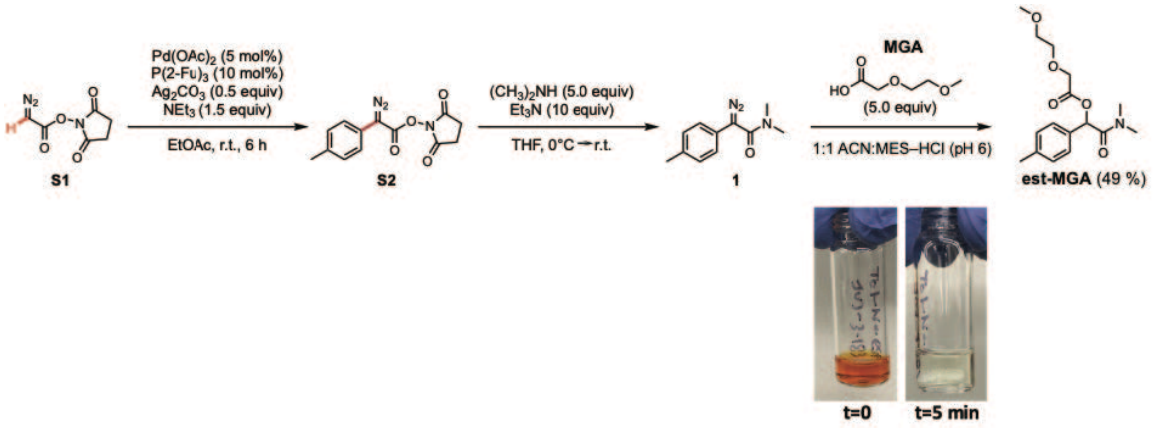


Figure 2-6: Previously reported synthetic route to diazo compound 1 followed by the esterification reaction used herein to yield est-MGA

171.5, 169.6, 140.7, 129.9, 129.2, 128.4, 74.1, 70.9, 70.2, 67.6, 58.0, 36.7, 35.9, 20.3. HRMS-ESI (m/z): [M + H]+ calculated for $C_{16}H_{24}NO_5$, 310.1649; found, 310.1643. **NMR of β-CD** are shown in Figures 2-12 to 2-15.

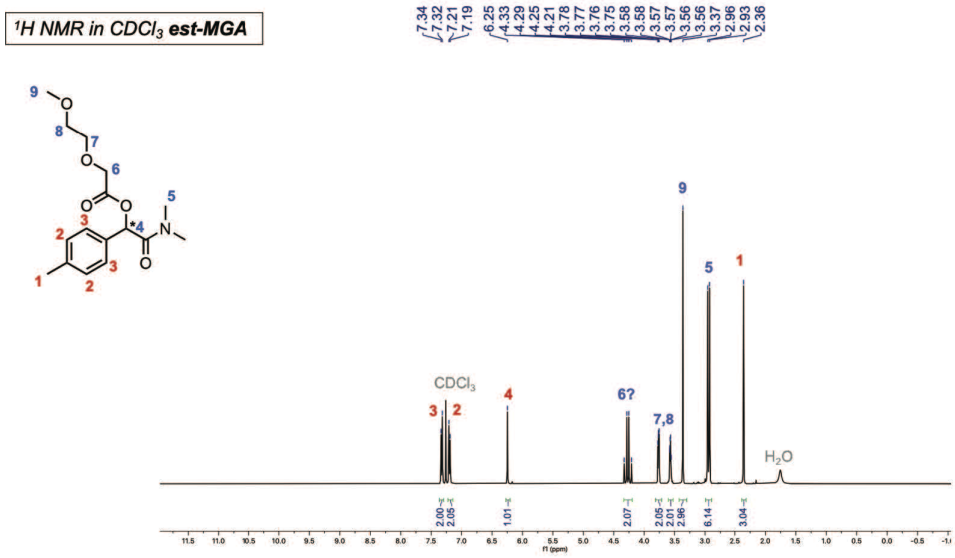


Figure 2-7: ^{1}H NMR spectrum (400 MHz) of est-MGA in $CDCl_3$ at 25 °C.

NMR Encapsulation Studies. The encapsulation of est-MGA was assessed with ^{1}H NMR spectroscopy. The concentration of est-MGA in PBS was kept constant at 3.33 mM, while the concentration of the host was increased gradually (0-10 mM for β-CD; 0-13.33 mM for α-CD, γ-CD, and CB7). The ^{1}H NMR peak positions for protons H1, H2, H3, and H4 of est-MGA were plotted as a function of the

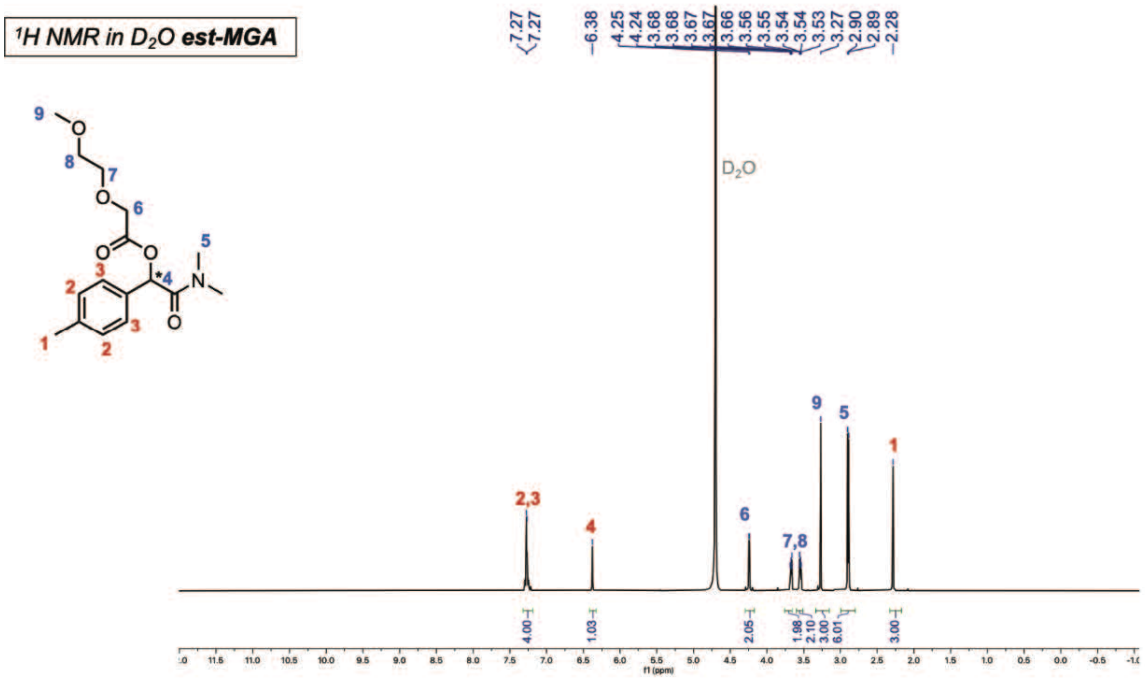


Figure 2-8: ^{1}H NMR spectrum (400 MHz) of est-MGA in D_2O at 25 °C.

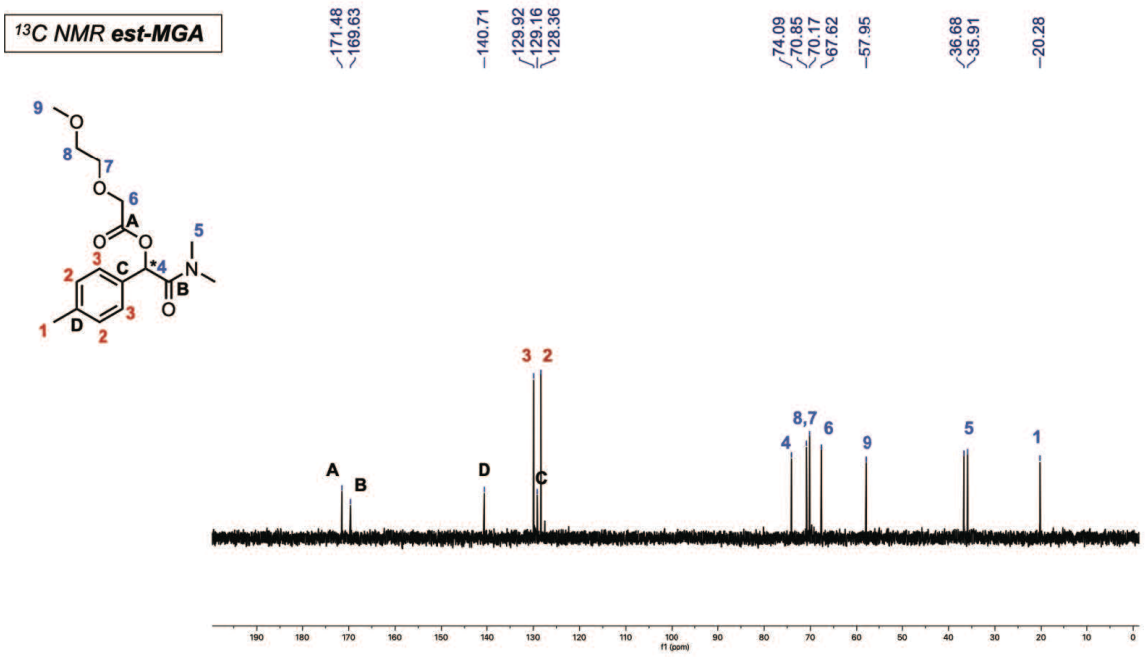


Figure 2-9: ^{13}C NMR spectrum (400 MHz) of est-MGA in D_2O at 25 °C.

total concentration of the host, and values of the equilibrium association constant (K_a) were determined by nonlinear 1:1 fitting curves generated using the method of

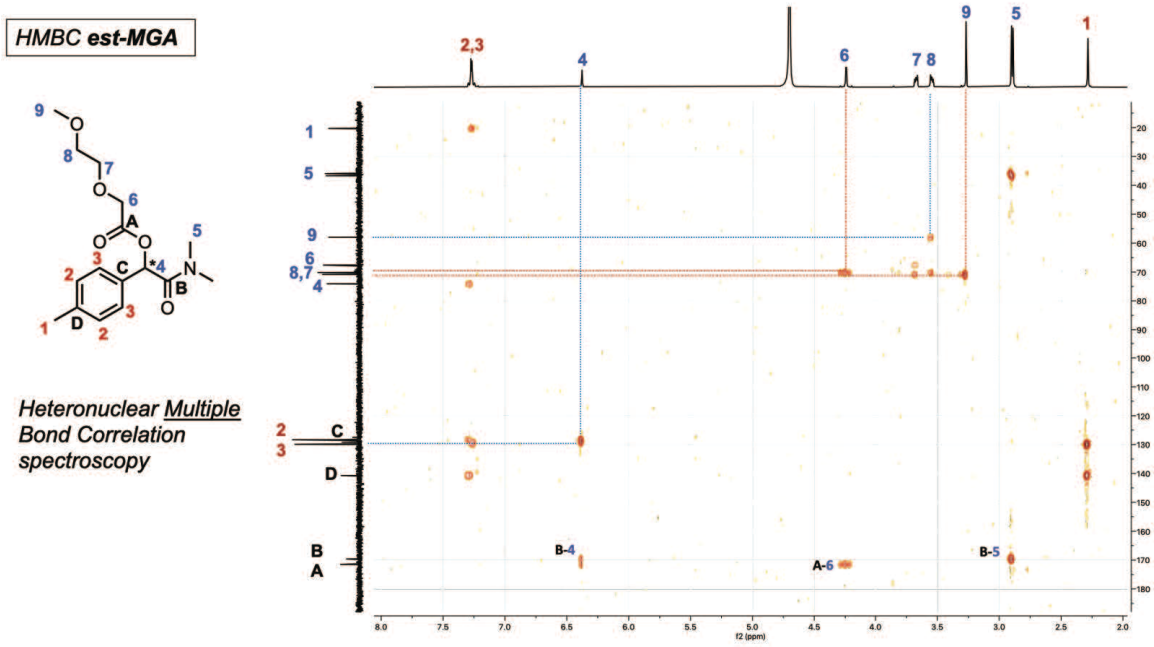


Figure 2-10: HMBC NMR spectrum (500 MHz) of est-MGA in D_2O at 25 °C.

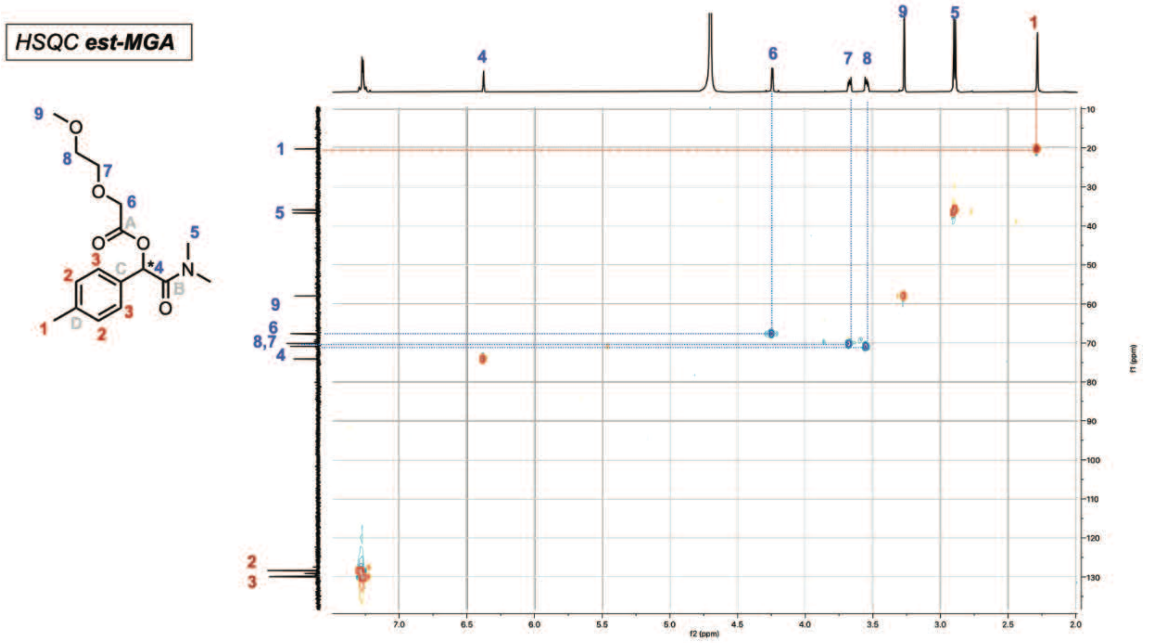


Figure 2-11: HSQC NMR spectrum (500 MHz) of est-MGA in D_2O at 25 °C.

Thordarson.[75]

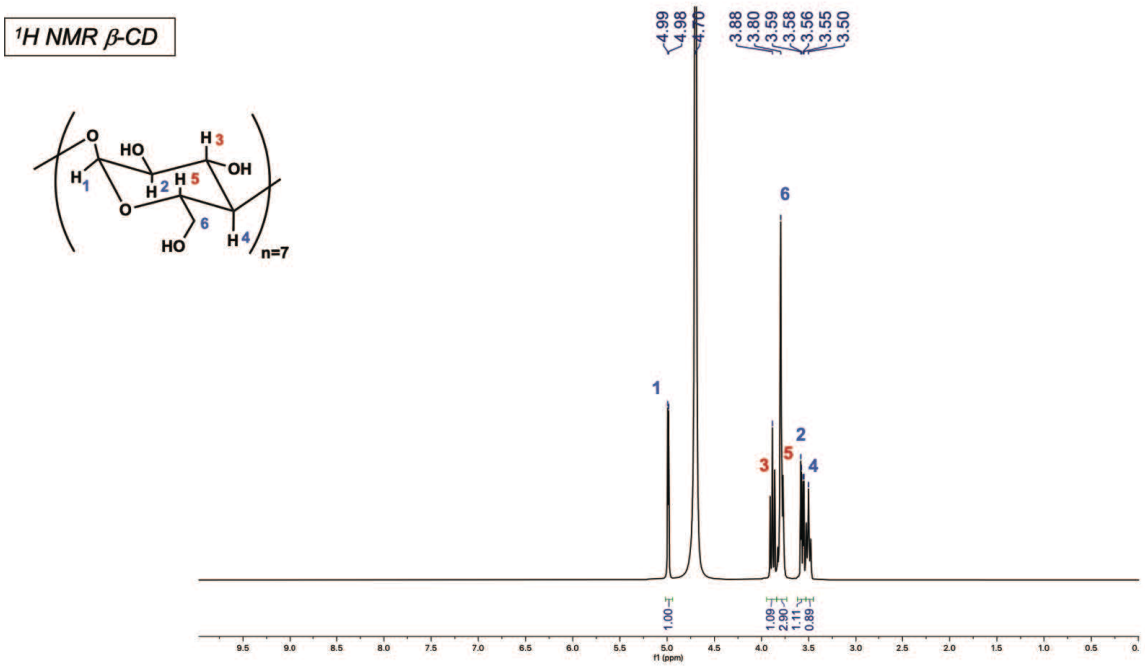


Figure 2-12: ^{1}H NMR spectrum (400 MHz) of β-CD in D_2O at 25 °C.

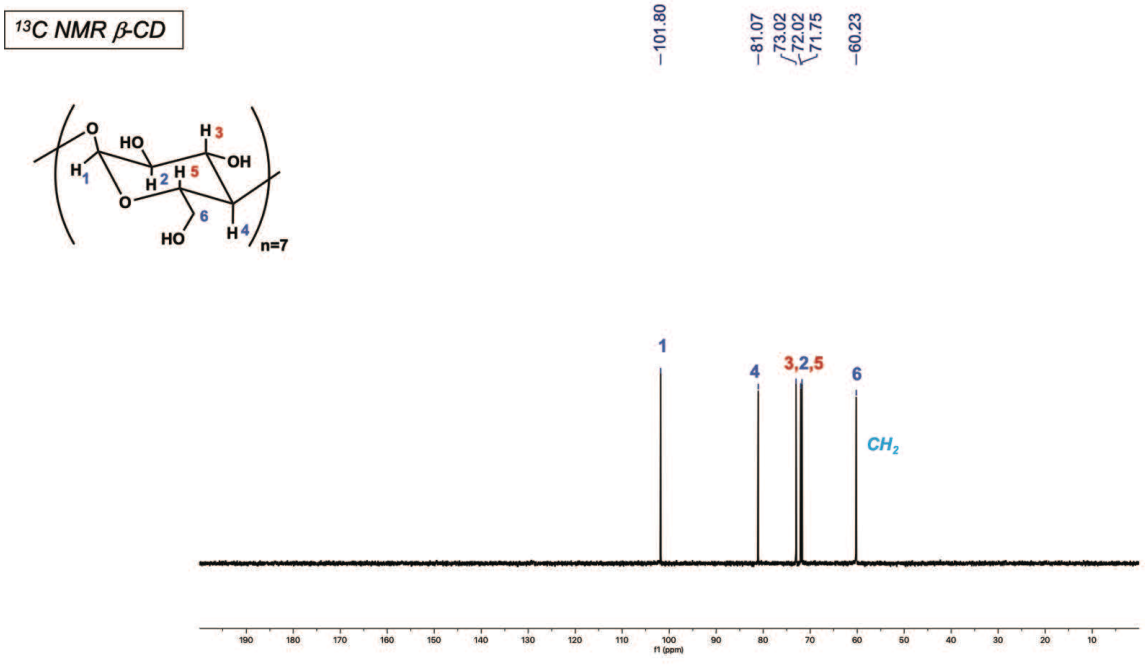


Figure 2-13: ^{13}C NMR spectrum (400 MHz) of β-CD in D_2O at 25 °C.

Effect of Additives on the UV Absorbance of est-GFP. UV absorbance served as a proxy for the structural integrity of GFP in the presence of additives. A

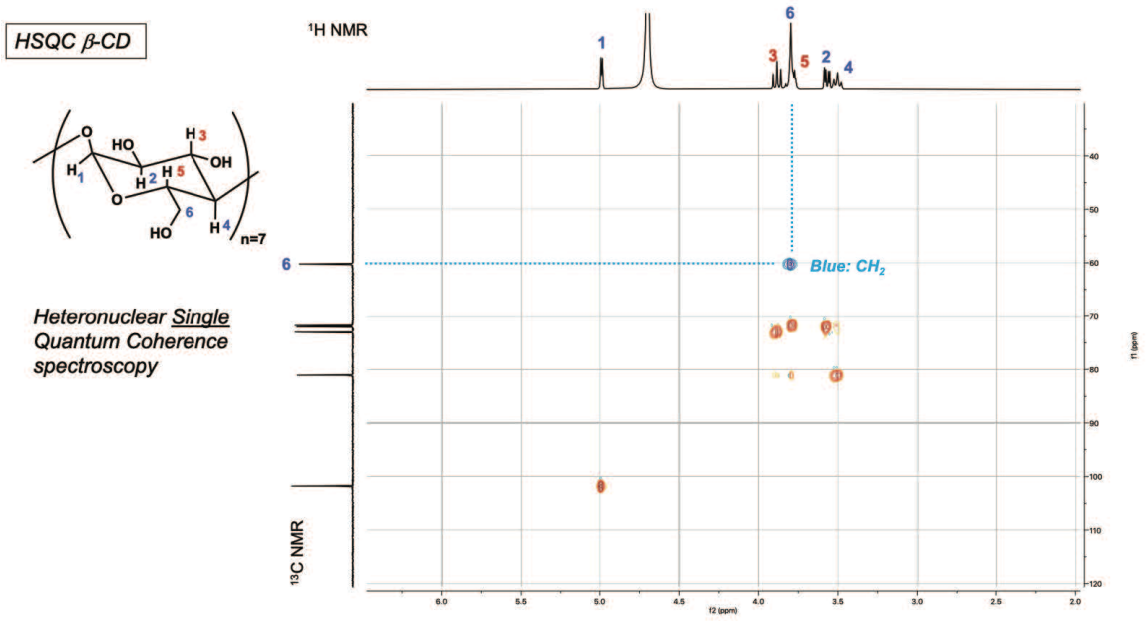


Figure 2-14: HSQC NMR spectrum (500 MHz) of β-CD in D_2O at 25 °C.

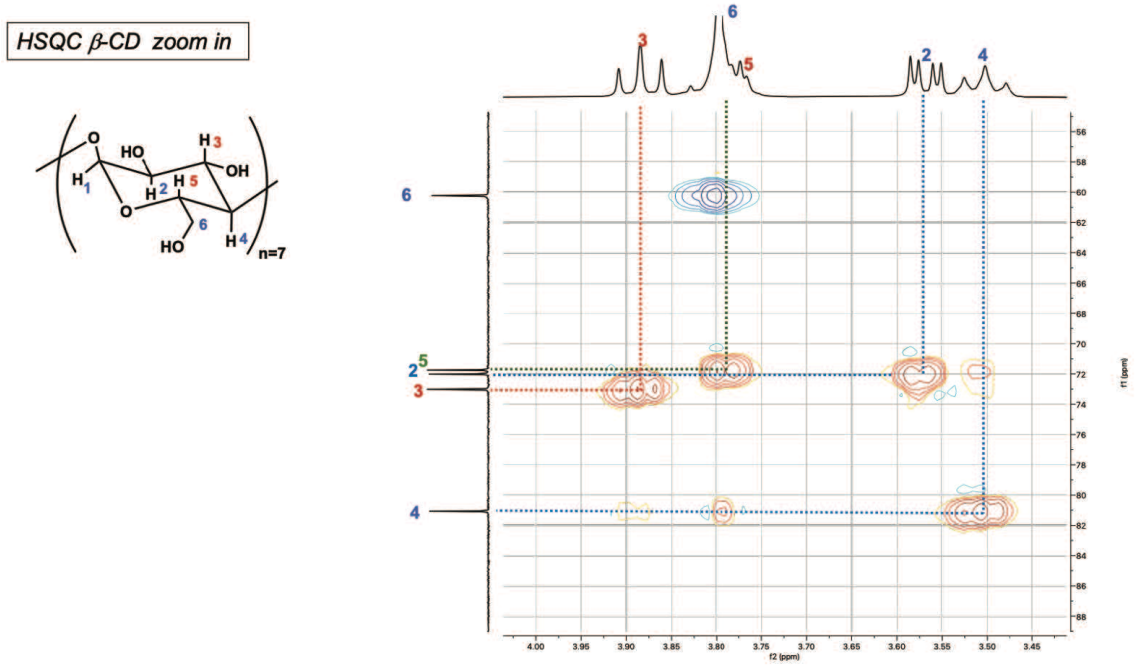


Figure 2-15: HSQC NMR spectrum (500 MHz) of β-CD in D_2O at 25 °C, zoomed in.

54 μM solution of est-GFP in PBS was prepared as described above. The est-GFP had a median of seven esters based on its MALDI-TOF mass spectrum. In separate tubes, 30 μL of the est-GFP solution was diluted with 30 μL of either PBS, pH 7.3, or

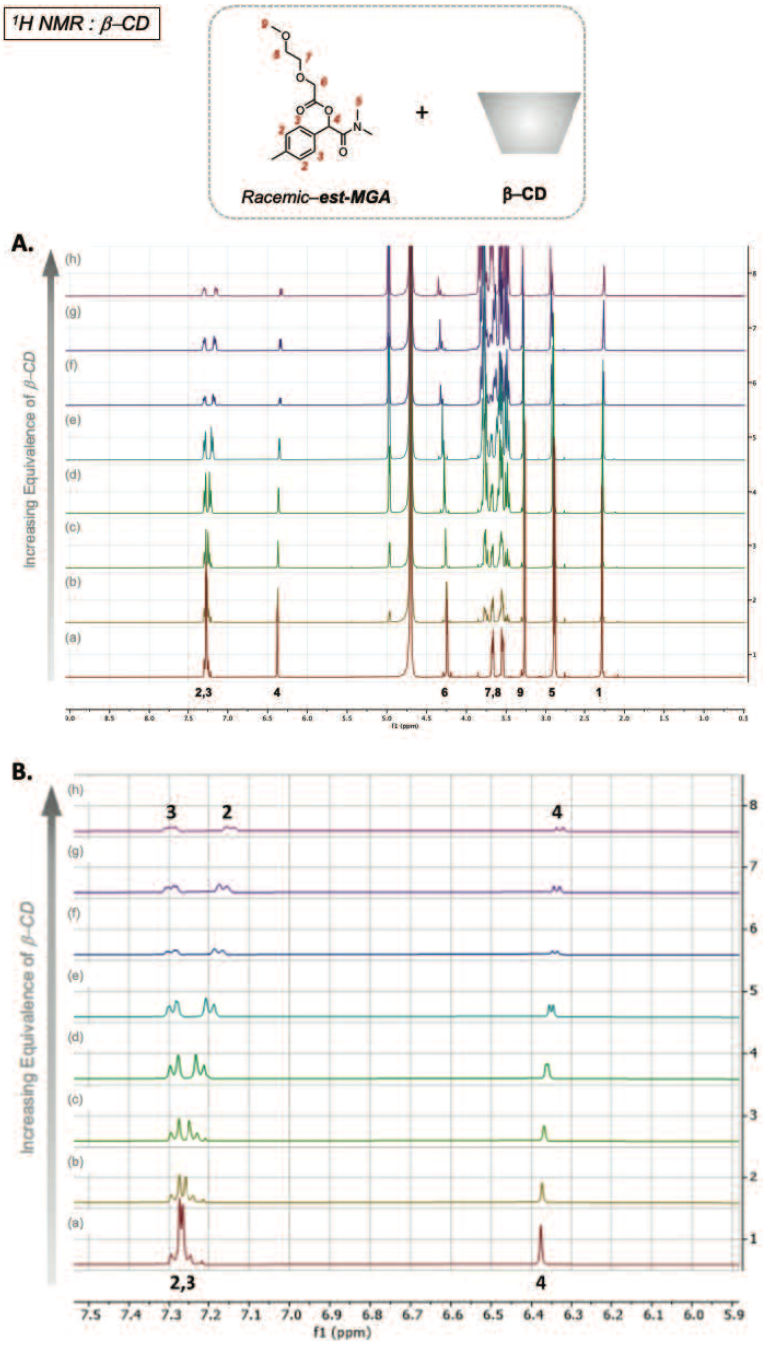


Figure 2-16: ^{1}H NMR spectra (400 MHz) of est-MGA with (a) 0.0 equiv, (b) 0.1 equiv, (c) 0.25 equiv, (d) 0.5 equiv, (e) 1.0 equiv, (f) 1.5 equiv, (g) 2.0 equiv, and (h) 3.0 equiv of β-CD in D_2O at 25 °C. A. Full view; B. 5.9 ppm to 7.5 ppm showing only peaks 2, 3, and 4.

additive-supplemented PBS, pH 7.3 to obtain final concentrations of additives that are the same as those reported in Figure 2-2. The volumes of the resulting solutions were calculated by dividing their weights (determined by taking the difference in weights

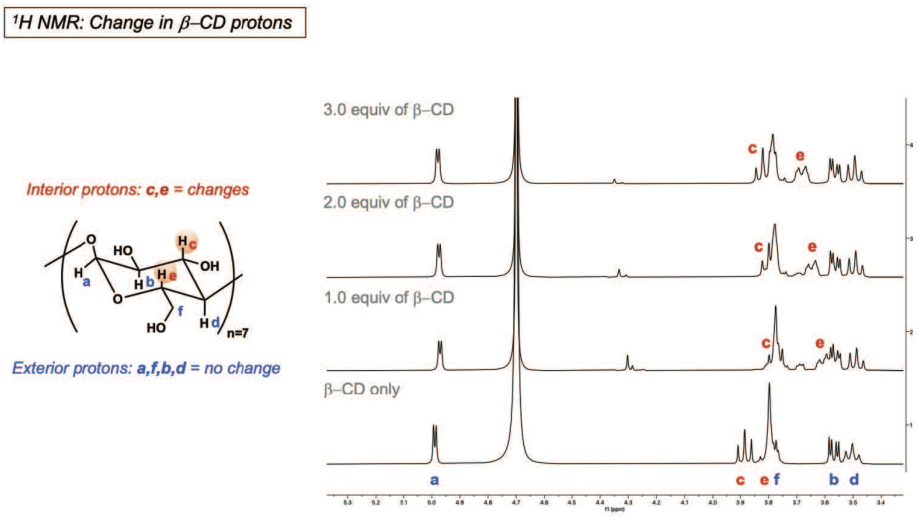


Figure 2-17: ^{1}H NMR spectra (400 MHz) of est-MGA with 0.0 equiv, 1.0 equiv, 2.0 equiv, and 3.0 equiv of β-CD inclusion complex were shown to demonstrate the change in internal protons (Hc and He) of β-CD only in D_2O at 25 °C.

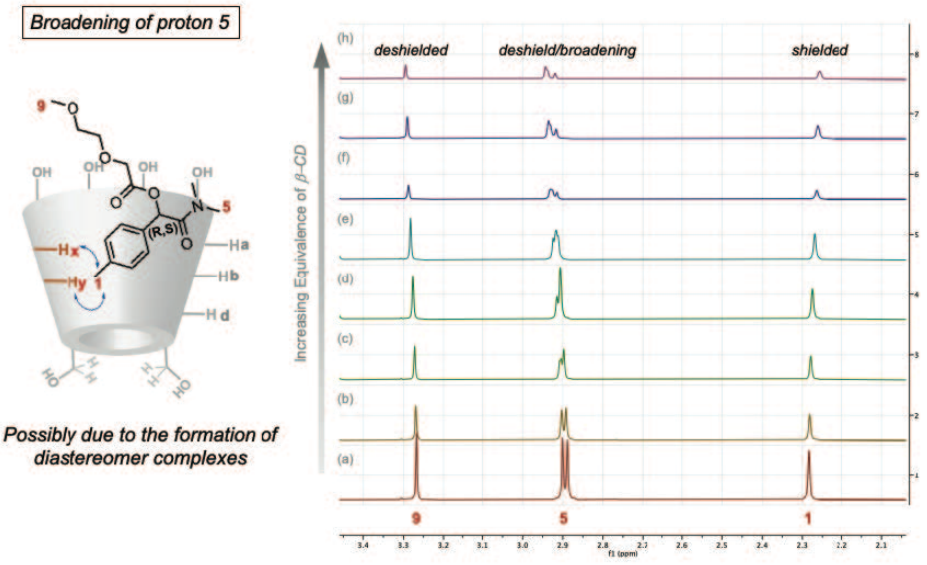
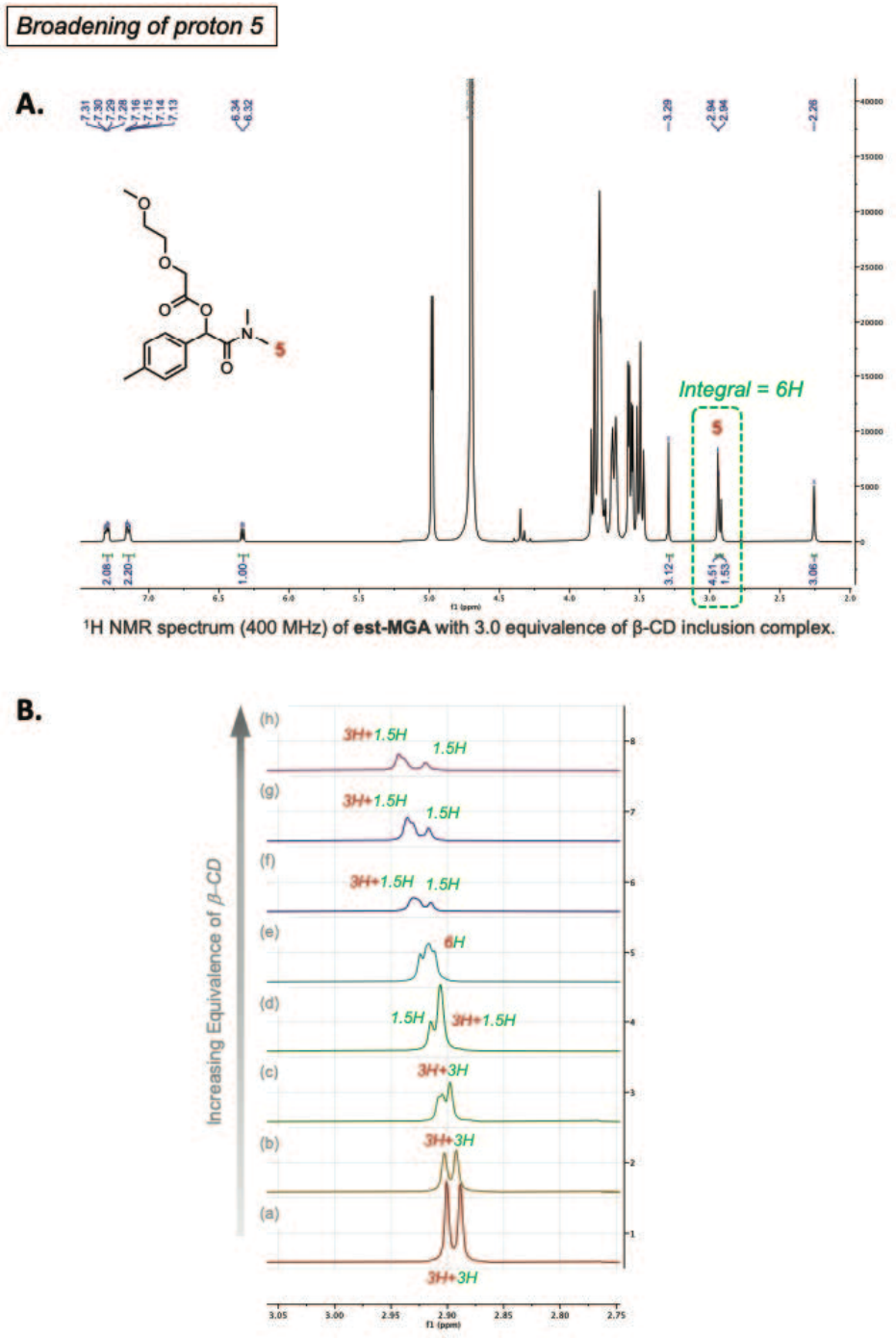


Figure 2-18: ^{1}H NMR spectra (400 MHz) of est-MGA with (a) 0.0 equiv, (b) 0.1 equiv, (c) 0.25 equiv, (d) 0.5 equiv, (e) 1.0 equiv, (f) 1.5 equiv, (g) 2.0 equiv, and (h) 3.0 equiv of β-CD inclusion complex in D_2O at 25 °C.

of the tubes with and without solution) by their densities (determined by measuring the weight of 1.0 mL of protein-free buffer). The concentration of est-GFP in each sample was calculated as 54 μM × 30 μL/calculated volume. The absorbance of each solution at 488 nm (A488 nm) was measured with a NanoDrop 2000c spectropho-



^{1}H NMR spectrum (400 MHz) of **est-MGA** with 3.0 equivalence of β-CD inclusion complex.

Figure 2-19: (A) ^{1}H NMR spectrum of est-MGA with 3.0 equiv of β-CD. (B) 1H NMR of broadening peak of H5 of est-MGA with (a) 0.0 equiv, (b) 0.1 equiv, (c) 0.25 equiv, (d) 0.5 equiv, (e) 1.0 equiv, (f) 1.5 equiv, (g) 2.0 equiv, and (h) 3.0 equiv of β-CD. The broadening of this peak (proton H5) is consistent with increasingly distinct interactions between each stereoisomer of est-MGA and β-CD

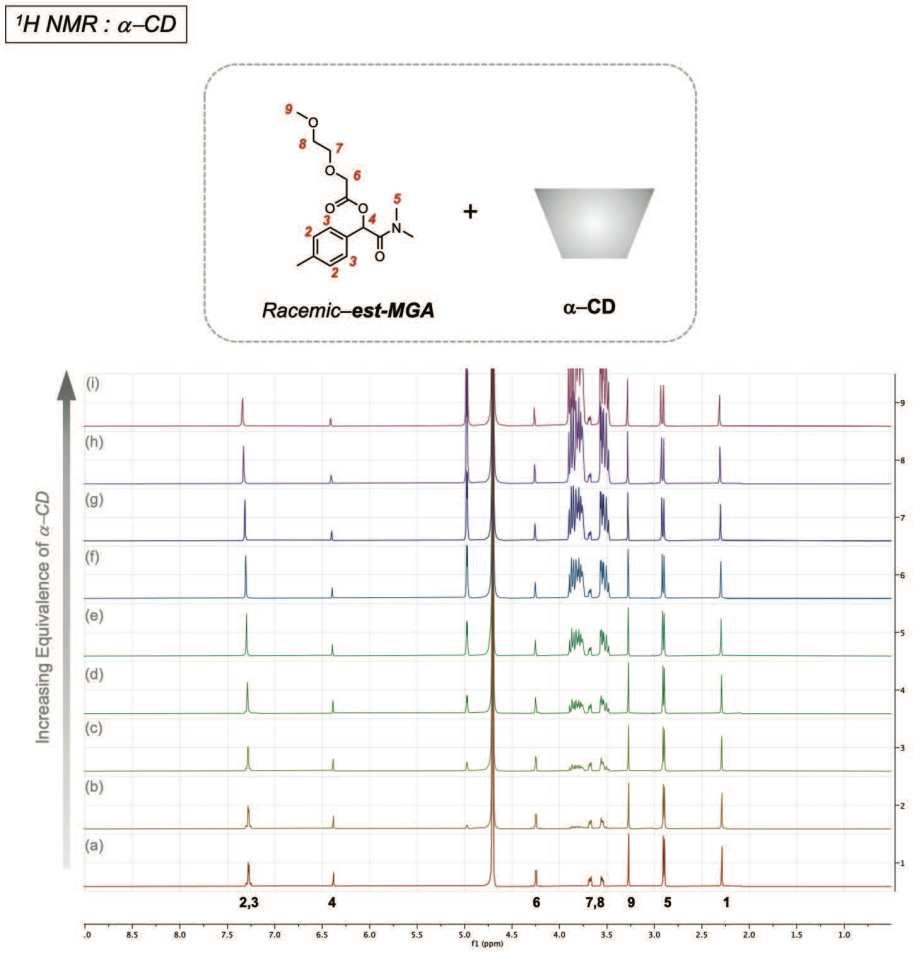


Figure 2-20: ^{1}H NMR spectra (400 MHz) of est-MGA with (a) 0.0 equiv, (b) 0.1 equiv, (c) 0.25 equiv, (d) 0.5 equiv, (e) 1.0 equiv, (f) 1.5 equiv, (g) 2.0 equiv, (h) 3.0 equiv, and (i) 5.0 equiv of α-CD inclusion complex were conducted at in D_2O at 25 °C.

tometer (Thermo Fisher Scientific). An extinction coefficient was calculated for each solution by dividing the A488 nm value by the expected concentration of est-GFP, and the % error of the calculated extinction coefficients (versus the theoretical value of 81760 M^{-1} cm^{-1}). Two replicate experiments were performed for each additive solution. The calculated extinction coefficients of est-GFP in additive-supplemented PBS are shown in Table 2.6.

Cell Culture. Chinese hamster ovary-K1 (CHO-K1) cells (obtained from ATCC) were cultured in F12K nutrient medium supplemented with 10% Fetal Bovine Serum

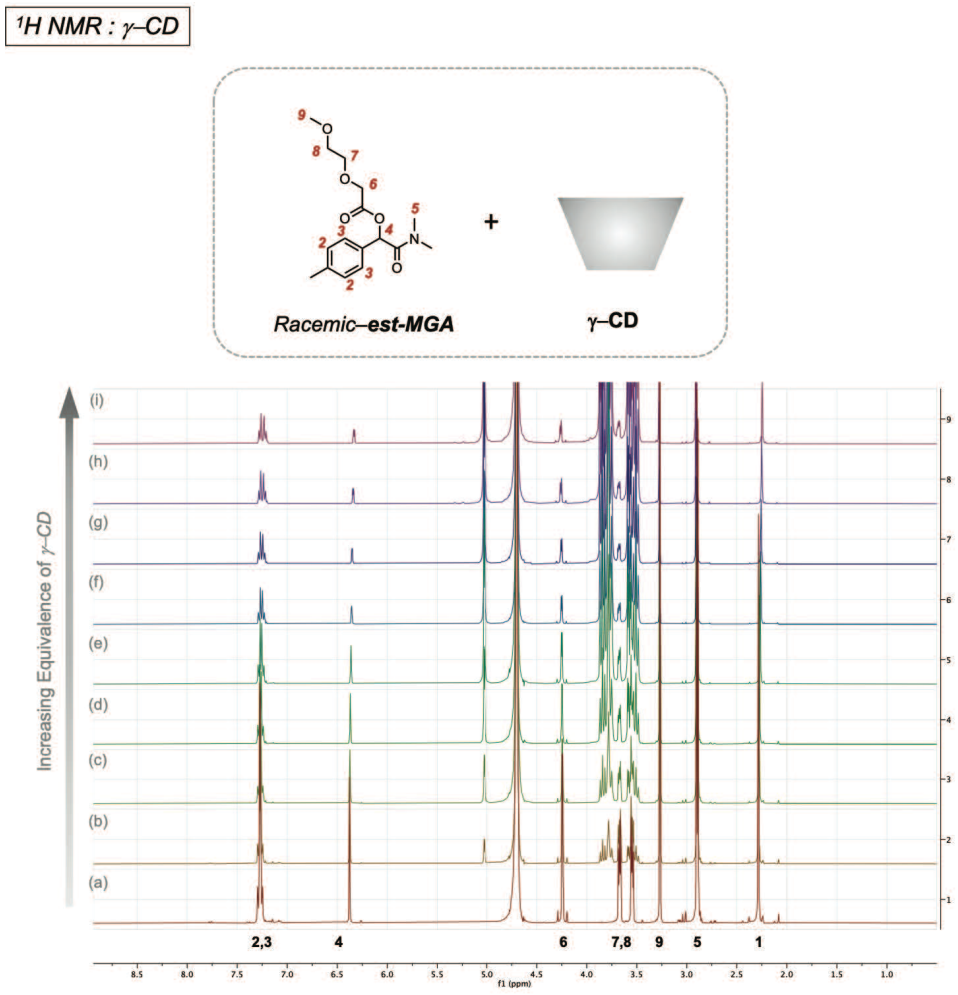


Figure 2-21: ^{1}H NMR spectra (400 MHz) of est-MGA with (a) 0.0 equiv, (b) 0.1 equiv, (c) 0.25 equiv, (d) 0.5 equiv, (e) 1.0 equiv, (f) 1.5 equiv, (g) 2.0 equiv, (h) 3.0 equiv, and (i) 5.0 equiv of γ-CD inclusion complex were conducted at in D_2O at 25 °C.

(Life Technologies), 100 U/mL penicillin, and 100 μg/mL streptomycin (Gibco), unless stated otherwise. The cells were grown in sterile culture flasks in a cell culture incubator at 37 °C under CO_2(g) (5% v/v). Cell lines were passaged a minimum of five times and up to twelve times before use. The cells were counted to determine seeding density using a hemacytometer.

Cellular Internalization of est-GFP. CHO-K1 cells were seeded in culture medium (which was F12K medium supplemented with 10% v/v fetal bovine serum, 100 units

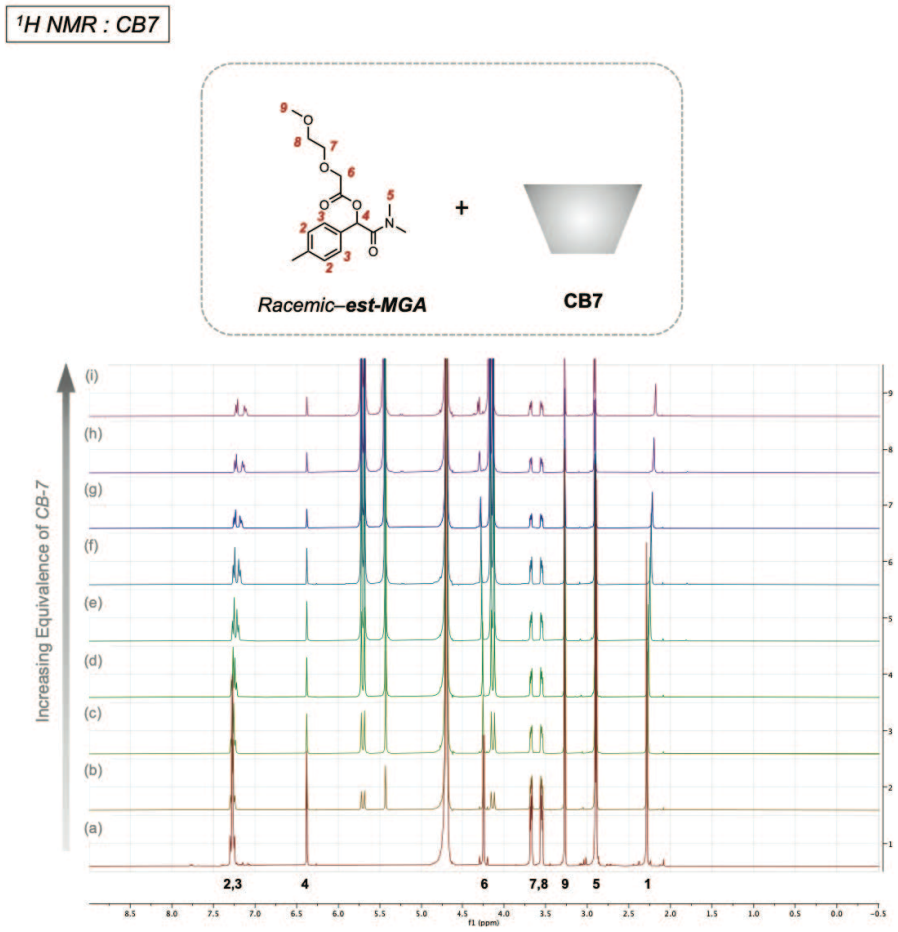


Figure 2-22: ^{1}H NMR spectra (400 MHz) of est-MGA with (a) 0.0 equiv, (b) 0.1 equiv, (c) 0.25 equiv, (d) 0.5 equiv, (e) 1.0 equiv, (f) 1.5 equiv, (g) 2.0 equiv, (h) 3.0 equiv, and (i) 4.0 equiv of CB7 inclusion complex were conducted at in D_2O at 25 °C.

mL^{-1} penicillin, and 100 μg mL^{-1} streptomycin) at a density of 50,000 cells $well^{-1}$ in a sterile 24-well plate at 24 h prior to treatment. The following five solutions of unmodified GFP and est-GFP were prepared to contain 64 μM protein: (a) unmodified GFP in PBS; (b) unmodified GFP in PBS supplemented with β-cyclodextrin (5 mM), L-arginine (250 mM), and sucrose (750 mM); (c) esterified GFP (with a median of 7 ester labels) in PBS (d) esterified GFP (with a median of 7 ester labels) in PBS supplemented with 5 mM β-cyclodextrin, 250 mM L-arginine, and 750 mM sucrose; (e) esterified GFP (with a median of 10 ester labels) in PBS supplemented with β-cyclodextrin (5 mM), L-arginine (250 mM), and sucrose (750 mM). Incuba-

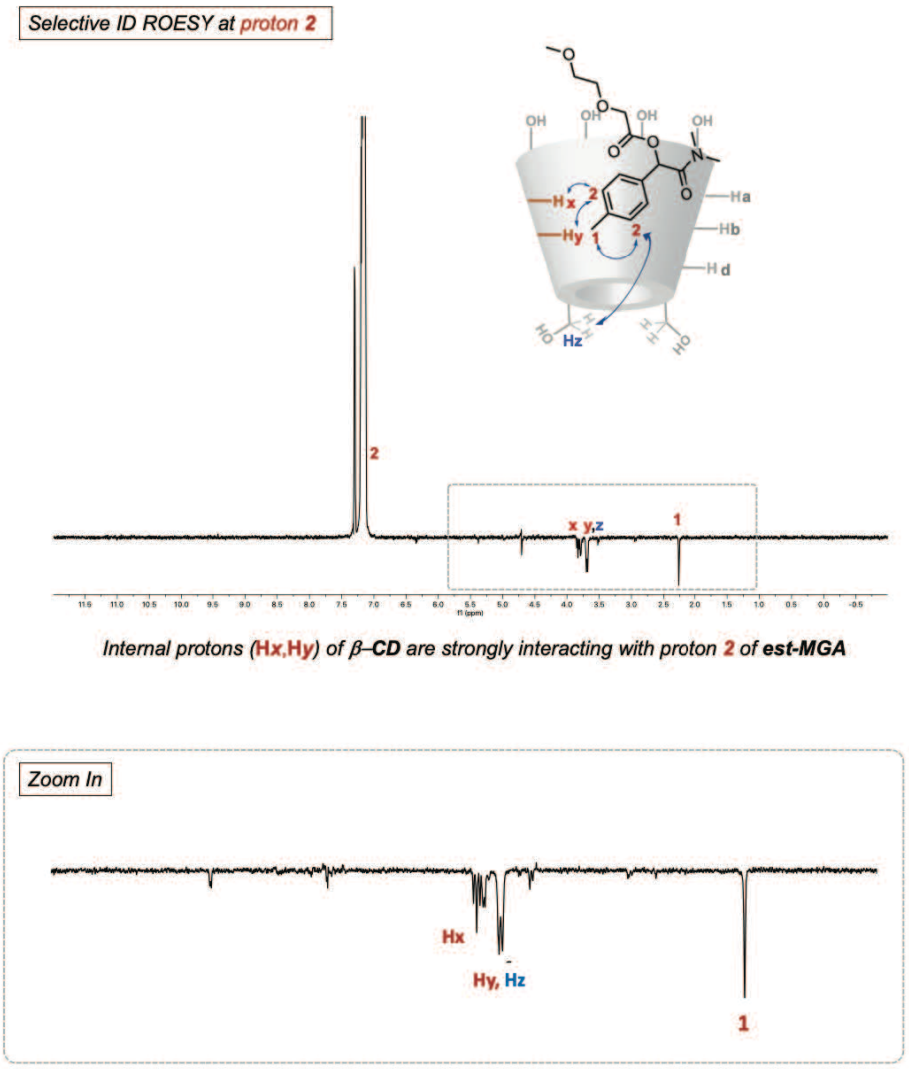


Figure 2-23: Selective 1D ROESY NMR spectrum of the inclusion complex of β-CD and est-MGA in D_2O at 25 °C. The box below has shown the interaction between aromatic ring of est-MGA and Hx and Hy inside the cavity of β-CD. Possible topology of the β-CD·est-MGA inclusion complex in aqueous solution is shown above.

tion solutions were then prepared by diluting all five solutions (a-e) 8-fold into culture medium to obtain a final GFP or est-GFP concentration of 8 μM; the final concentrations of additives in the additive-supplemented solutions (b, d, e) were 0.625 mM β-cyclodextrin, 31.25 mM L-arginine, and 93.75 mM sucrose. These final concentrations were determined to be noncytotoxic to CHO-K1 cells, as described below. The culture medium was removed, and cells were incubated with 300 μL of the incubation solutions for 4 h at 37 °C. After removal of the incubation solutions, the cells were

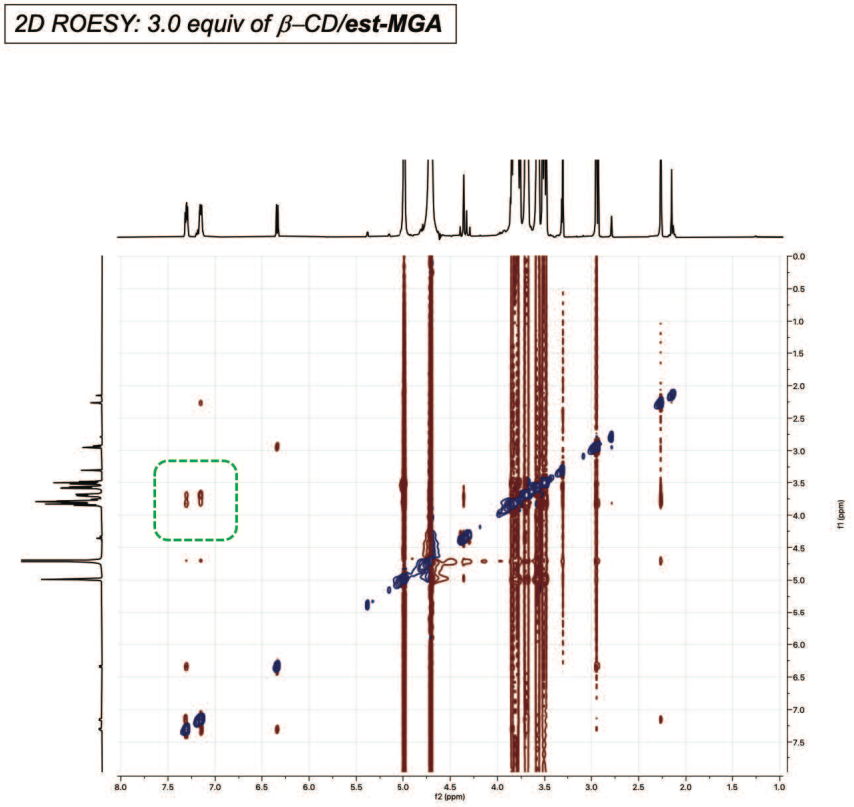


Figure 2-24: 2D ROESY NMR spectrum of the β-CD·est-MGA inclusion complex in D_2O at 25 °C. The green box has shown the interaction between aromatic ring of est-MGA and Hx and Hy inside the cavity of β-CD.

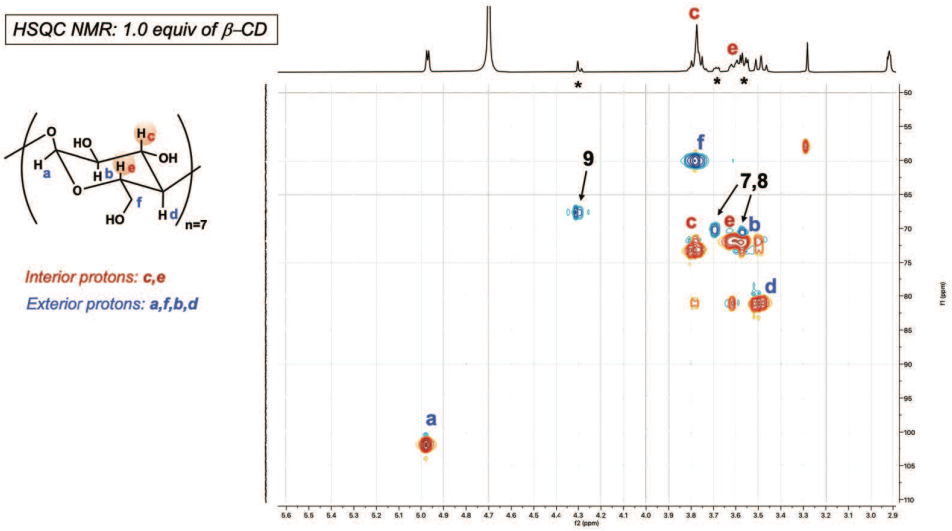


Figure 2-25: HSQC NMR spectra (500 MHz) of 1:1 β-CD/est-MGA in D_2O at 25 °C to assign protons of interest. c is equal to Hx and e is equal to Hy.

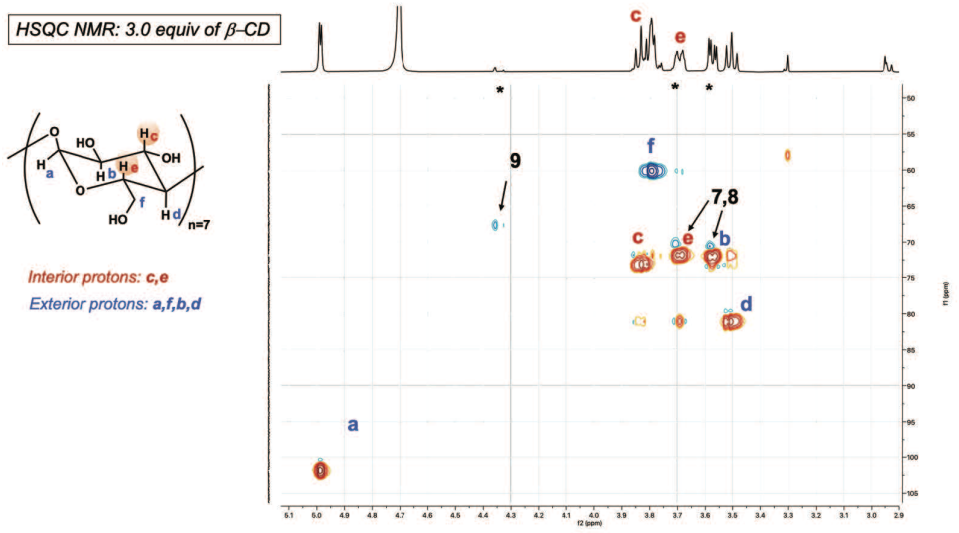


Figure 2-26: HSQC NMR spectra (500 MHz) of 3:1 β-CD/est-MGA in D_2O at 25 °C to assign protons of interest. c is equal to Hx and e is equal to Hy.

rinsed twice with PBS and released from the plate with 200 μL of warmed 0.25% v/v trypsin-ethylenediaminetetraacetic acid (EDTA) mix. Trypsin was quenched by the addition of 400 μL of the medium. The cells were then subjected to centrifugation for 5 min at 200g followed by aspiration of the supernatant. The cells were washed twice by resuspension in 500 μL of ice-cold PBS, centrifugation at 200g for 5 min, and aspiration of supernatant. The cells were resuspended in 500 μL of ice-cold PBS supplemented with bovine serum albumin (2% w/v) and propidium iodide (2 μg mL^{-1}) and kept on ice until the time of analysis. The fluorescence intensity of at least 10,000 events was measured with a BD FACS Celesta flow cytometer. The median fluorescence intensity of live, single cells is shown in Figure 2-4.

Cytotoxicity Assay. The cytotoxicity of the additives to CHO-K1 cells was assessed by quantifying ATP using a CellTiter-Glo 2.0 assay kit from Promega according to the manufacturer's instructions. CHO-K1 cells were plated at a density of 5000 cells well^{-1} in a sterile 96-well plate 24 h prior to treatment. The medium was replaced with either 100 μL well^{-1} of vehicle (PBS) or 100 L well^{-1} of additive-containing PBS for 6 h at 37 °C. The plates were equilibrated at room temperature for 30 min before the addition of CellTiter-Glo 2.0 reagent (100 μL well^{-1}). The contents were mixed

Table 2.6: Calculated Extinction Coefficients of est-GFP in Additive-Supplemented PBS

Additive	**[Additive] (mM)**[a]	**Weight (mg)**[b]	**Density (g/mL)**[c]	**Volume (μL)**[d]	**Expected [est-GFP] (μM)**[e]	ϵ_{calc} **($M^{-1}cm^{-1}$)**[f]	**Error of ϵ_{calc} (%)**[g]
None	N/A	60	1.00	60.0	27.0	82963 ± 524	1.5 ± 0.6
Sucrose	750	65	1.08	60.2	26.9	82104 ± 1576	0.4 ± 1.9
L-Arginine	250	61	1.02	59.8	27.1	83061 ± 522	1.6 ± 0.6
CB7	5	60	1.00	60.0	27.0	81667 ± 786	-0.1 ± 1.0
α-CD	5	60	1.00	60.0	27.0	81296 ± 1833	-0.6 ± 2.2
β-CD	5	60	1.00	60.0	27.0	81111 ± 524	-0.8 ± 0.6
γ-CD	5	60	1.00	60.0	27.0	80370 ± 1048	-1.7 ± 1.3
β-CD	5						
+ L-Arginine	250	67	1.11	60.4	26.8	81412 ± 1317	-0.4 ± 1.6
+ Sucrose	750						

[a] Post-dilution with est-GFP. [b] Determined by taking the difference in the weights of the tubes with and without solution. [c] Determined by measuring the weight of a 1-mL solution of protein-free buffer. [d] Determined by dividing the weight of each solution by their densities. [e] Determined by dividing the amount of est-GFP by the calculated volumes, e.g., 54 μM × 30 μL/60.2 μL = 26.9 μM for sucrose. [f] Determined by dividing the A488 nm value by the expected [est-GFP]. Values are the mean ± SD for two replicates. [g] Determined by calculating the % difference with the theoretical value of 81760 $M^{-1}cm^{-1}$ for GFP. Values are the mean ± SD for two replicates.

for 2 min on an orbital shaker to induce cell lysis. The plate was incubated at room temperature for 10 min before the luminescence was measured with a Tecan Infinite 200 Pro plate reader. Cell viability was expressed as a percentage relative to the vehicle control. In accordance with ISO 10993-5, cell viability above 80% is considered to be noncytotoxic. The viability of CHO-K1 cells in the presence of additives is depicted in Figure 2-27.

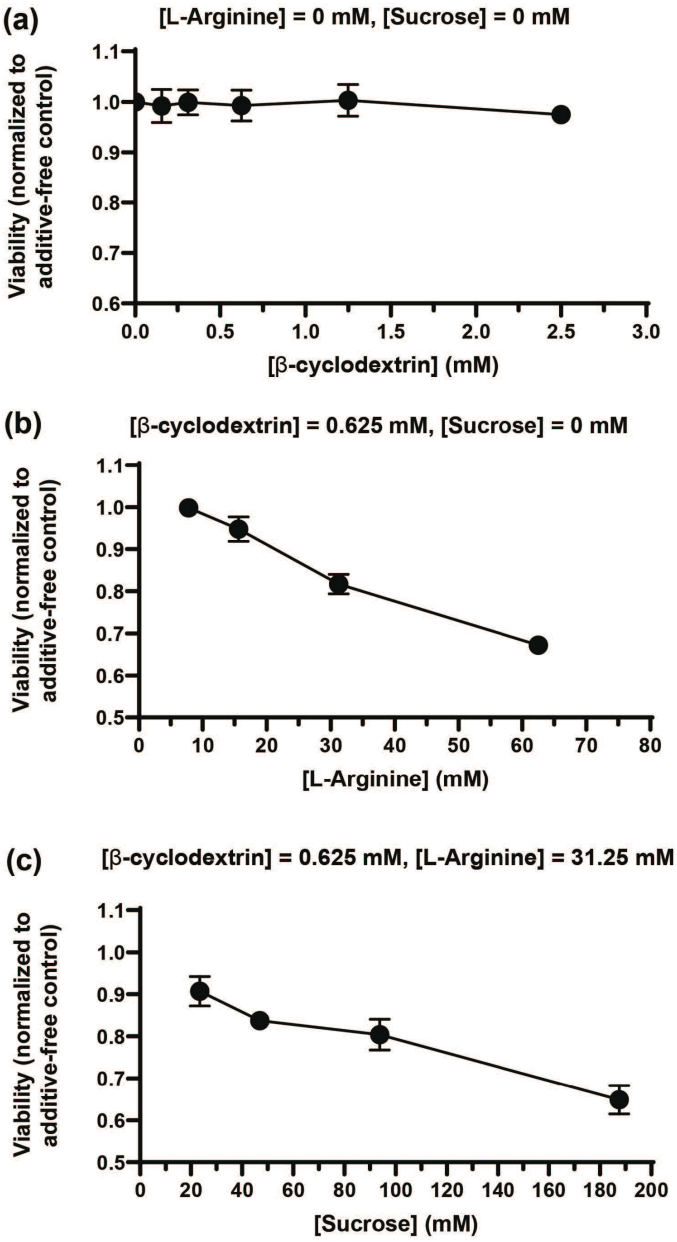


Figure 2-27: Effect of additives on the viability of CHO-K1 cells. CHO-K1 cells were incubated for 6 h at 37 °C with (a) varying concentrations of β-cyclodextrin in PBS, pH 7.3; (b) 0.625 mM β-cyclodextrin and varying concentrations of L-arginine in PBS, pH 7.3; (c) 0.625 mM β-cyclodextrin, 31.25 mM L-arginine, and varying concentrations of sucrose in PBS, pH 7.3. Cell viability was measured with a CellTiter Glo 2.0 Assay from Promega according to the manufacturer's instructions. Viability is normalized to that of the additive-free vehicle control. Values are the mean ± SD with three technical replicates per concentration.

Chapter 3

Development of a Cell-Permeable RAS-targeting Degrader

3.1 Introduction

Given the successful cytosolic delivery of several proteins via bioreversible esterification [11, 8, 12], we sought to extend this strategy to a therapeutically-relevant protein. To this end, we specifically chose to work with R11.1.6, a Raf-competitive inhibitor against KRAS G12D that was derived, via yeast surface display, from a charge-neutralized Sso7d library [1].

3.1.1 Sso7d Protein Library

The wild-type Sso7d protein is derived from the hyperthermophilic archaeon *Sulfolobus solfataricus*. As a DNA-binding protein, it has high cationicity (pI = 9.66). Traxlmayr *et al.* developed yeast-displayed, charge-neutralized libraries of Sso7d (1) to reduce nonspecific binding to anionic mammalian cell surfaces and (2) to circumvent the issue of the ring of positive charges near the binding surface which could inhibit binding with positively-charged antigens or epitopes. Its design as a yeast-display library enables efficient screening via flow cytometric sorting to isolate the yeast-displaying variants with desired properties, e.g. that bind to a desired target.

Kauke *et al.* developed R11.1.6 based on this Sso7d scaffold library [79]. Sso7d has a nine-residue binding surface, highlighted in magenta in Figure 3-1.

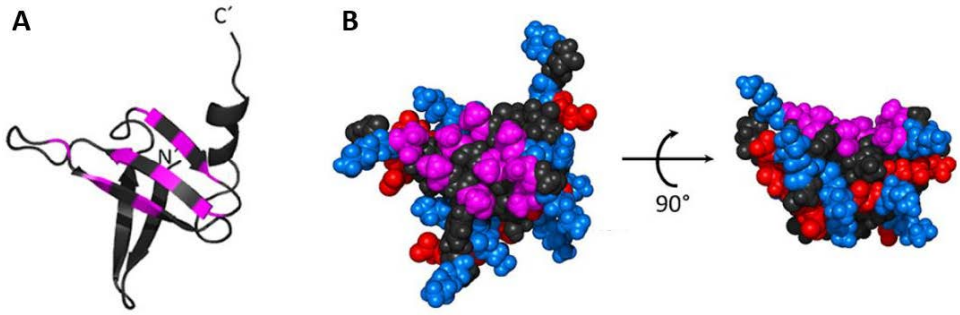


Figure 3-1: Structure of wild-type Sso7d. The nine-residue binding residues are colored in magenta. In the space-filling model (B), the cationic residue — arginine and lysine — are colored in blue, and the negatively charged residues — aspartate and glutamate — are colored in red. Figure adapted from [1]

3.1.2 Why R11.1.6?

R11.1.6 has several characteristics that lend itself favorably to the esterification-based permeation strategy:

1. As an Sso7d protein, R11.1.6 is a relatively small protein (∼7 kDa) and thus has a greater potential to traverse the membrane compared to larger proteins like GFP [80];

2. Since R11.1.6 is a cationic protein (with net charge of +3.2 at physiological pH and a pI of 9.26), its aqueous solubility is unlikely to be compromised upon esterification. As described in section 2.2, the esterification of anionic proteins like GFP compromises solubility by increasing its pI to near-physiological pH; on the other hand, the esterification of cationic proteins like RNase1 [12] do not detrimentally affect its aqueous solubility since its pI would be raised further away from physiological pH;

3. R11.1.6 targets an intracellular target, KRAS G12D. This serves as good motivation for its cytosolic delivery;

4. R11.1.6 does not contain any Asp or Glu residues, i.e. esterification sites, as part of its nine-residue binding surface, preventing the risk that esterification might negatively affect its binding to KRAS G12D.

Additionally:

1. R11.1.6 is well-characterized;

2. As an Sso7d protein, it has high thermal stability, making it more amenable to protein engineering efforts that could potentially introduce destabilizing mutations;

3. R11.1.6 targets KRAS G12D, an important and prevalent target that remains undruggable to this day.

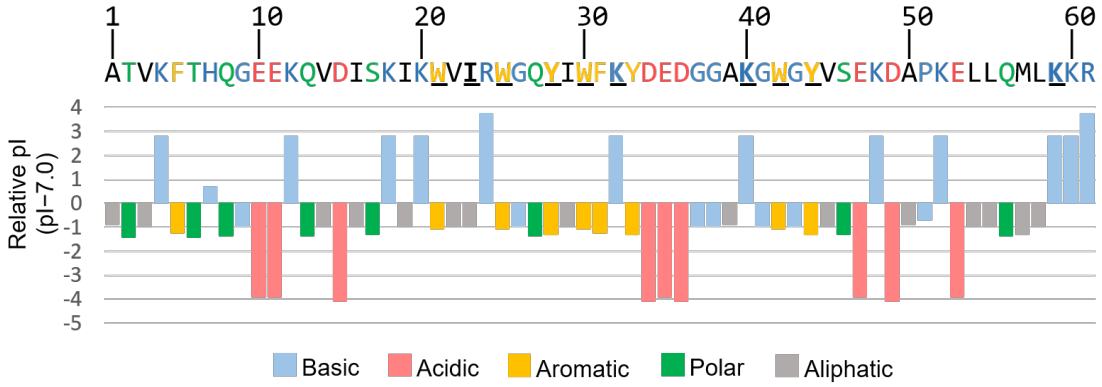


Figure 3-2: Sequence of R11.1.6 and the relative pI (pI - 7.0) of each amino acid. The residues that contribute to the affinity and specificity of R11.1.6 to KRAS G12D — nine Sso7d binding residues (W21, I23, W25, Y28, W30, K32, K40, W42, Y44) and K59 — are bolded and underlined. Each amino acid is also color-coded based on its side-chain properties (basic, acidic, aromatic, polar, aliphatic).

The sequence of R11.1.6, along with plots of the relative pI (normalized to 7.0) of each amino acid, is shown in Figure 3-2. Acidic and basic residues (in red and blue, respectively) have relative pI values that deviate the most from 7.0, consistent with their high hydrophilicity and charged nature at physiological pH. The nine residues that constitute the binding site of R11.1.6, and an additional lysine residue (K59) that was identified to give rise to the specificity of R11.1.6 for the G12D mutant over

WT K-RAS [79], are bolded and underlined (referred to herein as the ten "binder residues"). R11.1.6 also contains nine aspartates (D) and glutamates (E), i.e. nine potential esterification sites.

3.1.3 Why KRAS G12D?

RAS proteins (KRAS, NRAS, and HRAS) are membrane-associated guanosine triphosphatases (GTPases) involved in cell signaling pathways and are the most frequently mutated family of oncoproteins in human cancers [3, 4]. Its mutants are present in ∼30% of human cancers, with KRAS representing the most frequently mutated isoform in RAS-driven cancers (86%) (cf. 11% for NRAS and 3% for HRAS) [5]; KRAS predominates in lung, colorectal, and pancreatic cancers [4]. In normal quiescent cells, RAS proteins are mostly GDP-bound and inactive. However, mutations in RAS genes render RAS proteins persistently bound to GTP with exceptionally high affinity [4]. This mutationally-activated form of RAS binds effector proteins including Raf and PI3K [81, 82] which results in uninhibited proliferation of cancer cells and promotes their migration and invasion [83, 84, 85].

KRAS mutations are dominated by single-base missense mutations, with 89% of mutations occurring at codon 12 (G12), 9% at codon 13 (G13), and 1% at codon 61 (Q61) [7]. Among G12 mutations, the G12D mutation is the most prevalent mutation (36%; cf. 14% for G12C and 23% for G12V), with the G12D mutation predominating in colorectal and pancreatic cancers and the G12C mutation in Non-small cell lung cancer (NSCLC). For decades, KRAS was deemed a challenging therapeutic target ("undruggable") given its considerably smooth and shallow surface [6], but discovery of an allosteric site of KRAS-G12C has led to the development of several small-molecule inhibitors including ARS-1620, MRTX-849 (adagrasib) and AMG-510 (sotorasib), all of which covalently react with cysteine 12 of KRAS-G12C; sotorasib and adagrasib were also FDA-approved in 2021 and 2022, respectively [6]. Several KRAS-G12C-targeting small-molecule PROTACs have also been reported in the literature and will be discussed further in Section 3.2.3.

Despite these major advances, the covalent strategy has been restricted to the

G12C mutation, which represents only a small proportion of mutations in RAS-driven cancers. There is thus a strong motivation to develop therapeutics against other KRAS variants like KRAS G12D. There are still no effective small-molecule or protein-based inhibitors against KRAS G12D [86], but recent works [87, 9] have demonstrated success using degradation-inducing proteolysis technologies, which will be discussed further in Section 3.1.4.

3.1.4 PROTAC-based Design

R11.1.6 binds KRAS G12D with single-digit nanomolar affinity and showed an almost 10-fold greater specificity over KRAS WT. A later study, however, revealed that genetically encoded intracellular expression of R11.1.6 failed to functionally antagonize downstream signaling through either the MAPK or PI3K pathways. [88] This was due to insufficient expression of endogenous R11.1.6 to fully inhibit the K-RAS/B-Raf interaction; mathematical modeling showed that $\sim 10^7$ molecules of R11.1.6 would be required to achieve stoichiometric inhibition of KRAS G12D.

A preliminary study to quantify esterified GFP (est-GFP) internalization in CHO-K1 cells showed entry of 10^5 est-GFP molecules per cell (see section 3.4). So, a cell-permeable variant of R11.1.6 would have to be roughly 100 times more efficient at cytosolic entry vs. est-GFP to achieve the requisite intracellular concentration for disruption of RAS signaling. This hints at a potential roadblock in the pursuit of a cell-permeable stoichiometric inhibitor of KRAS G12D, and prompted us to develop a PROTAC-like molecule that could catalytically degrade KRAS G12D instead.

Proteolysis Targeting Chimeras, or "PROTACs", were first reported in 2001 by the Crews Lab [89]. They are bifunctional molecules that eliminate a protein of interest (POI) inside cells. A PROTAC molecule consists of three components: (1) a ligand that binds to the POI ("POI binder"), (2) a second ligand at the opposite end that binds to an E3 ligase ("E3 ligand"), and (3) a crosslinker that joins the two ligands together ("linker"). The simultaneous binding of the ligands to the POI and E3 ligase brings the two proteins close together, triggering the ubiquitination of POI lysine residues, essentially marking the POI for degradation by the proteasome.

Upon degradation of the POI, the process of POI-ubiquitination and degradation starts again; PROTACs are, in essence, catalytic protein degraders, and can therefore function at sub-stoichiometric concentrations [90]. Importantly, the POI binder does not need to inhibit the function of the protein. A schematic of targeted protein degradation by PROTACs is shown in Figure 3-3.

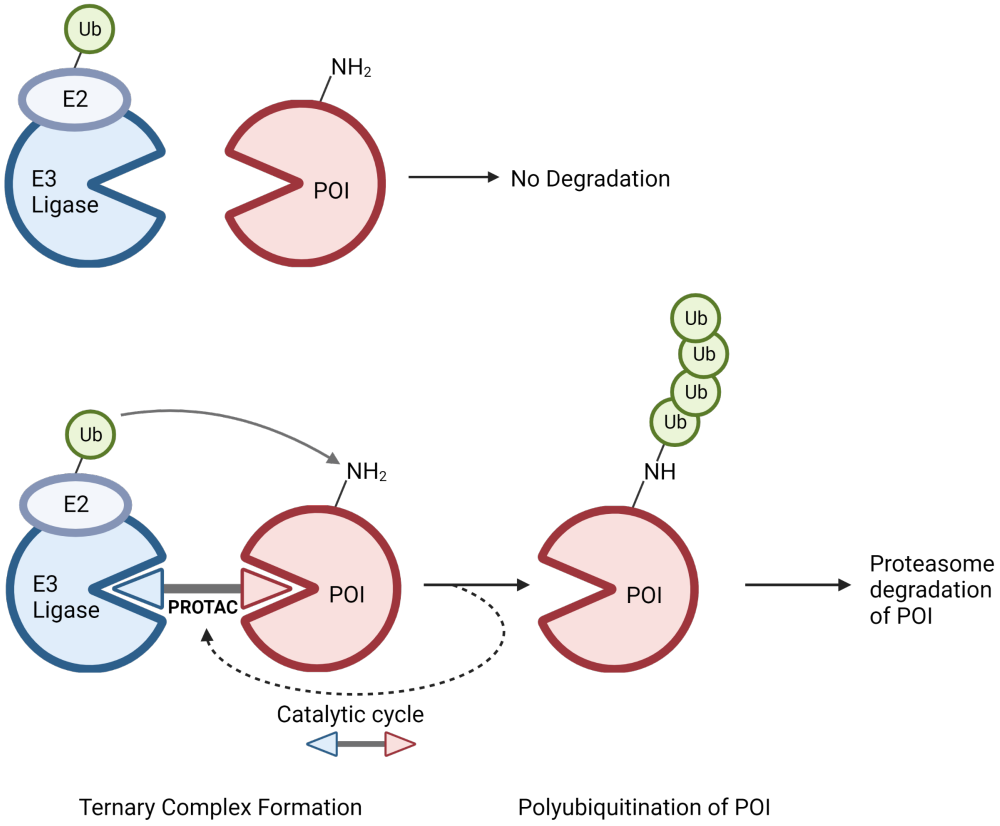


Figure 3-3: Targeted protein degradation by PROTACs.

We reasoned that sub-stoichiometric intracellular levels of a cell-permeable, large-molecule, KRAS G12D-targeting PROTAC, i.e. an esterified R11.1.6 protein molecule conjugated with a small-molecule E3 ligand, might be able to degrade sufficient amounts of KRAS G12D to antagonize downstream signaling.

High-affinity noncovalent ligands against KRAS G12D have proven futile due to the lack of small molecule binding pockets on the protein surface, but recent work from the Partridge Lab has demonstrated the successful degradation of KRAS G12D in cells that express protein-based bioPROTACs, or biodegraders, i.e. proteins that

fuse a high-affinity large-molecule binder (e.g. RBD, K27, R11.1.6, and NS1) to an E3 ligase (e.g. SPOP) [9], which we will discuss further in section 4.

3.2 Results and Discussion

3.2.1 Esterification of R11.1.6

Unlike inherently fluorescent GFP, R11.1.6 had to be fluorescently labeled to enable its detection. A cysteine was introduced at the solvent-exposed C-terminus of R11.1.6 via site-directed mutagenesis (the C-terminal Arg61 was substituted with Cys), followed by maleimide conjugation with BODIPY FL ("BDP"; MW: 414.2 Da; ex. 503 nm, em. 512 nm), a small ($<$ 0.5 kDa) and agnostic (uncharged and hydrophobic) dye. The high specificity of maleimide chemistry [91] ensures that only one dye molecule is conjugated to each R11.1.6 molecule, minimizing any changes that the dye molecule might inflict on the properties and behavior of R11.1.6 ($\sim$7.3 kDa). BDP also lacks a carboxyl group, precluding its conversion into an ester during the esterification of the protein, which could potentially alter the fluorescence of the dye.

R11.1.6-BDP was then esterified following a similar protocol for GFP esterification (in collaboration with Dr. Aniekan Okon from the Raines Lab) to obtain esterified R11.1.6-BDP (est R11.1.6-BDP). On average, 3 to 4 of the 9 available carboxyl groups are esterified, as confirmed by MALDI-TOF spectrometry. est-R11-BDP was then incubated with cells, and cellular internalization was characterized by confocal microscopy (Figures 3-4 and 3-5). With CHO-K1 cells, the est R11.1.6-BDP seemed to get taken up by cells via endocytosis (punctate pattern of fluorescence); unesterified R11.1.6-BDP did not appear to get internalized by CHO- K1 cells. On the other hand, with SW48 G12D cells (a KRAS-G12D-expressing human colon cancer cell line), both unesterified and esterified R11.1.6-BDP seemed to have achieved cytosolic entry to about the same degree (qualitative, as determined via confocal), although flow cytometry analysis (Figure 3-15) revealed a two-fold higher rate of cellular entry for the esterified protein over the unesterified protein. The microscopy

results suggest that the uptake of est R11.1.6-BDP is cell-line-dependent and that esterification of R11.1.6-BDP might not be necessary to engender cytosolic uptake (possibly due to the cationicity and small size of R11.1.6). We also tried to conduct microscopy experiments with HPAF-II and LS180 cells (both of which are adherent KRAS-G12D-expressing human cell lines) but it was difficult to image these cell lines because they grow in clumps (data not shown). Also, given that (1) we are interested in the degradation of KRAS G12D in tumor cell lines and (2) est-R11-BDP is primarily endocytosed in CHO-K1 cells (non-tumor cell line), we only performed confocal microscopy experiments in SW48 G12D cells going forward.

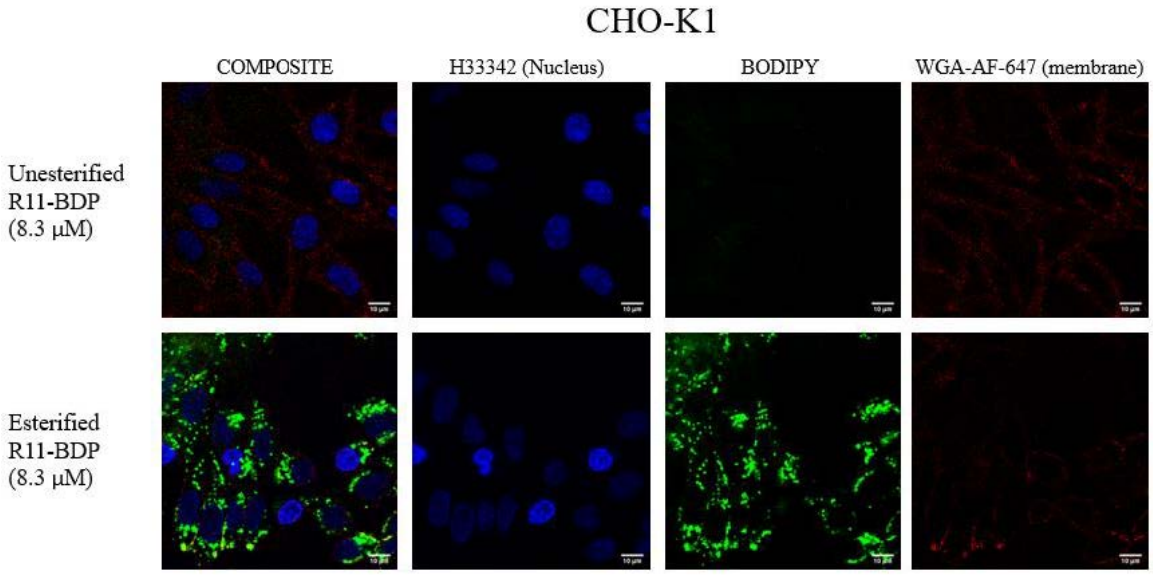


Figure 3-4: Live-cell fluorescence microscopy images of the cellular internalization of R11.1.6-BDP or est R11.1.6-BDP in CHO-K1 cells. Cells were incubated with protein (8.3 μM) in FBS-free F12K medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 4 h at 37 °C. The protein solution was then replaced with F12K medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed, stained with Hoechst 33342 and wheat germ agglutinin (WGA)-Alexa Fluor 647, and imaged by confocal microscopy. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BDP: ex. 503 nm, em. 512 nm. Scale bars: 10 μm.

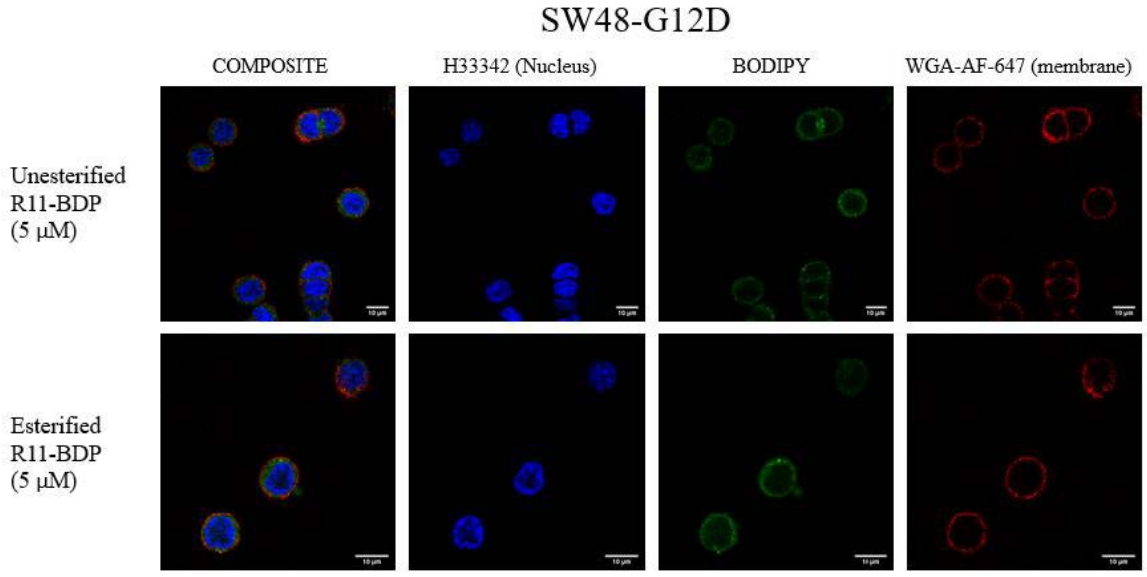


Figure 3-5: Live-cell fluorescence microscopy images of the cellular internalization of R11.1.6-BDP or est R11.1.6-BDP in SW48-G12D cells. Cells were incubated with protein (5.0 μM) in FBS-free RPMI 1640 medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin for 4 h at 37 °C. The protein solution was then replaced with RPMI 1640 medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed, stained with Hoechst 33342 and wheat germ agglutinin (WGA)-Alexa Fluor 647, and imaged by confocal microscopy. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BDP: ex. 503 nm, em. 512 nm. Scale bars: 10 μm.

3.2.2 Complementary permeation strategies

In addition to esterification, we also explored site-directed mutagenesis as a complementary protein engineering strategy to enhance the permeability of R11.1.6. Site-directed mutagenesis to introduce positively charged residues is a strategy that has been employed in many previous works [92, 93, 94]. However, there has been varying success in delivering these mutated proteins into the cytosol of mammalian cells, and it remains uncertain how the surface distribution of cationic residues on a protein affects its cytosolic internalization. The pentamutant GFP designed by Fuchs & Raines [92], which has a cationic patch of arginine and lysine residues and a net charge of +1, was internalized by mammalian cells but was shown to localize in endosomes.

Similarly, supercharged +36 GFP, which has a much higher net charge of +36

and a more diffuse pattern of positive charges across its surface (of mostly lysines), was reported to enter cells via an endocytic pathway; however, its fusion to a variety of proteins was reported to successfully engender cytosolic entry of these proteins, at much higher rates compared to fusions with known protein transduction domains like TAT and penetratin [94]. Quach *et al.* observed no correlation between the clustering of positive charges and increased cytosolic uptake: a fivefold variation in cytosolic uptake was measured among six miniature protein variants with the same charge but different relative placements of the penta-arginine motif (a distinct array of five arginine residues on a folded α-helix), with peptides having a cluster of arginine residues ranking among the best and worst in terms of intracellular concentration. The mixed results of these studies suggest that a straightforward substitution of residues with positive residues will not necessarily promote cellular internalization, let alone true entry into the cytosol. They do, however, demonstrate that this approach holds such potential.

Lysine-free (Arginine-rich) variants

Lys-free variants were of particular interest in the downstream development of R11.1.6-based large-molecule PROTACs. We reasoned that the presence of surface-accessible Lys (i.e. ubiquitin attachment sites) on the POI binder would lead to some level of self-degradation of the PROTAC, and we could improve POI degradation efficiency by removing most or all of the Lys from the POI binder. See section [REF] for further discussion. We also hypothesized that by replacing Lys with Arg (both are basic amino acids with similar chemical structures), we would not only minimize disruption of the biophysical properties of R11.1.6 but also improve the rate of membrane translocation [92, 93, 94].

The first variant we cloned and expressed was the lysine-free variant of R11.1.6 (R11-XK), where all ten lysines (Lys or K) were substituted with arginine (Arg or R). Its sequence is shown in Table 3.1. This raises the number of Arg residues from 2 to 12, i.e. 20% of the full protein sequence. The net charge of the protein (at physiological pH) is maintained at +3.2 but the pI is raised to 10.4 due to the higher

pK_a of Arg vs. Lys. Given that three of the ten binder residues (K32, K40, K59) were mutated from Lys to Arg (see Figure 3-2), it was surprising to observe only a slight loss of binding affinity for KRAS G12D. R11-XK had an affinity of 21 nM for KRAS G12D (c.f. R11.1.6 which has an affinity of 6 nM for KRAS G12D).

The yield of unconjugated R11-XK was comparable to that of R11.1.6 (3.0-4.0 mg/L-culture) when a high concentration of salt (0.5 M NaCl) was maintained in all purification buffers, a common strategy used in purifying highly cationic proteins. However, upon labeling with hydrophobic BDP dye and exchanging into PBS (to ensure compatibility with *in vitro* assays), R11-XK-BDP fully precipitated; the addition of 0.25 M L-Arginine and 0.75 M sucrose into the PBS buffer helped maintain the solubility of R11-XK-BDP. The yield of R11-XK-BDP was ∼0.5 mg/L-culture in additive-supplemented PBS (c.f. 1.5 mg/L-culture for R11.1.6-BDP in regular PBS). Moreover, despite the high specificity of maleimide labeling, the resulting R11-XK-BDP remained poorly labeled (dye:protein molar ratio of 0.08 instead of 1.00). Further optimization of the labeling protocol is still required. Nevertheless, the poorly-labeled R11-XK-BDP (i.e. a combination of 8% of R11-XK-BDP with 92% of unlabeled R11-XK) was incubated with cells, and cellular internalization characterized via confocal microscopy.

Nuclear entry of R11-XK-BDP was observed with both CHO-K1 and SW48-G12D cells (Figure 3-6). The morphology of cells incubated with R11-XK-BDP differs from those in the protein-free control, with the former being flatter ($\sim 8\ \mu$m vs. the typical $\sim 16\mu$m) and more "web-like". There was concern that this change in morphology and protein localization (from cytosolic to nuclear) was driven by the potential cytotoxicity of the excipients (L-Arginine and sucrose) or of R11-XK. However, viability assays proved otherwise; SW48 G12D cells incubated with the excipients, R11.1.6, or R11-XK, all maintained full viability relative to protein-free control (Figures 3-8 and 3-9).

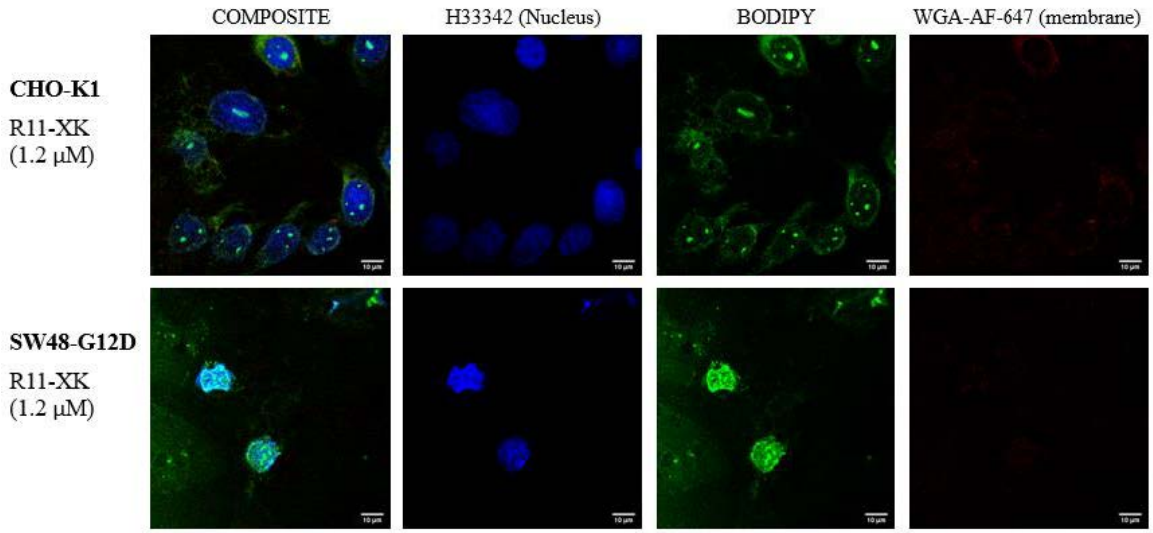


Figure 3-6: Live-cell fluorescence microscopy images of the cellular internalization of R11-XK-BDP in CHO-K1 and SW48 G12D cells. Cells were incubated with protein (1.2 μM) in FBS-free F12K medium (for CHO-K1 cells) or RPMI 1640 medium (for SW48 G12D cells) supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 4 h at 37 °C. The protein solution was then replaced with medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed, stained with Hoechst 33342 and wheat germ agglutinin (WGA)-Alexa Fluor 647, and imaged by confocal microscopy. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BDP: ex. 503 nm, em. 512 nm. Scale bars: 10 μm.

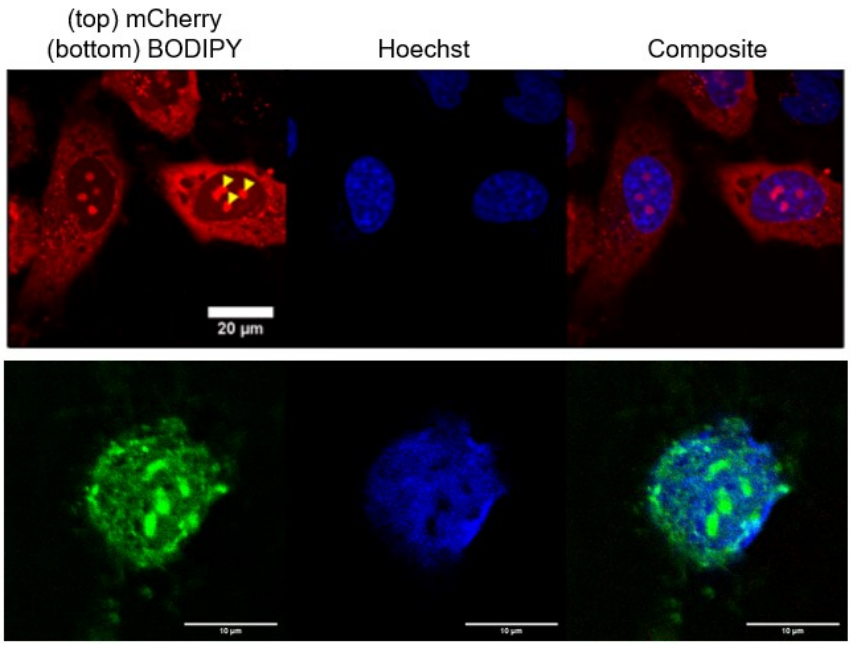


Figure 3-7: (top) Live-cell fluorescence microscopy images of HeLa cells incubated with cR10-conjugated mCherry protein, for 1 h at 37 °C, washed, stained with Hoechst 33342. cR10 is a cyclic cell-penetrating peptide containing 10 Arg residues. Nucleoli of one cell are highlighted with yellow arrowheads. Figure adapted from [2]. (bottom) Live-cell fluorescence microscopy images of SW48-G12D cells incubated with R11-XK-BDP (1.2 μM) in FBS-free RPMI 1640 medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin for 4 h at 37 °C. The protein solution was then replaced with RPMI 1640 medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed, and stained with Hoechst 33342 and wheat germ agglutinin (WGA)-Alexa Fluor 647. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BDP: ex. 503 nm, em. 512 nm. Scale bars: 10 μm.

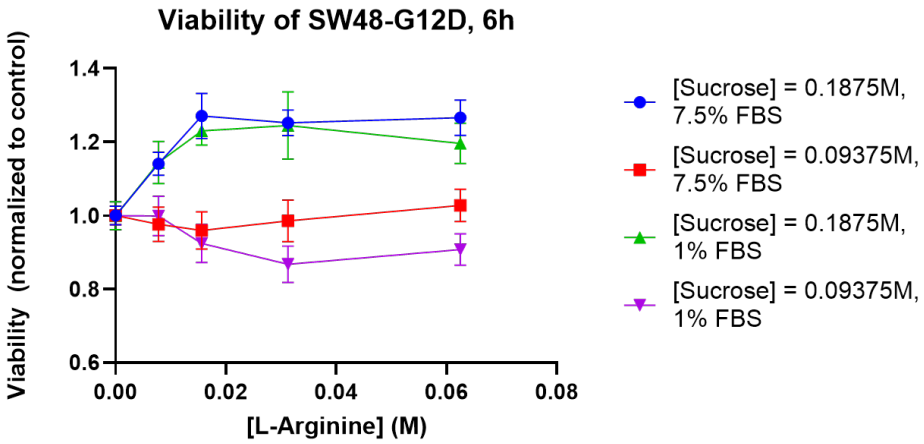


Figure 3-8: Effect of L-Arginine and sucrose on the viability of SW48 G12D cells. SW48 G12D cells were incubated for 6 h at 37 °C with varying concentrations of L-Arginine and sucrose in RPMI 1640 media supplemented with either 1% or 7.5% fetal bovine serum (FBS), pH 7.3. Cell viability was measured with a CellTiterGlo 2.0 Assay from Promega according to the manufacturer's instructions. Viability is normalized to that of the additive-free vehicle control. Values are the mean ± SD with three technical replicates per concentration

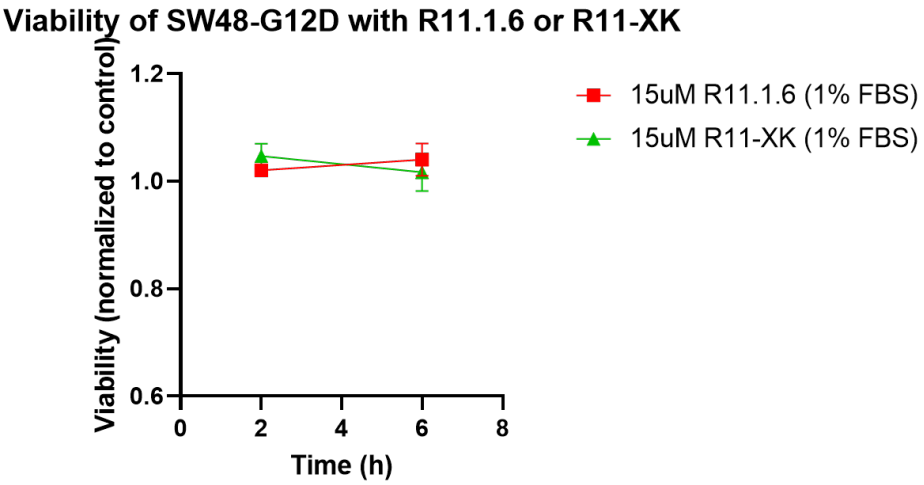


Figure 3-9: Effect of R11.1.6 and R11-XK on the viability of SW48 G12D cells. SW48 G12D cells were incubated for either 2 h or 6 h at 37 °C with either R11.1.6 or R11-XK in RPMI 1640 media supplemented with 1% fetal bovine serum (FBS), pH 7.3. Cell viability was measured with a CellTiterGlo 2.0 Assay from Promega according to the manufacturer's instructions. Viability is normalized to that of the additive-free vehicle control. Values are the mean ± SD with three technical replicates per data point.

Table 3.1: Sequences and K_d values of R11.1.6 and its variants. Mutated residues are in red. The C-terminal residue is originally an Arg (R) but is substituted to a Cys (C) if the protein is conjugated via maleimide chemistry to a small molecule like BODIPY

		K_d (nM)	
Variant	**Sequence**	KRAS G12D	KRAS WT
R11.1.6*	ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKKR	6	17
R11-XK	ATVRFTHQGEERQVDISRIRWVIRWGQYIWFRYDEDGGARGWGYVSERDAPRELLQMLRRR	21	89
R11-XK-R1218K	ATVRFTHQGEEKQVDISKIRWVIRWGQYIWFRYDEDGGARGWGYVSERDAPRELLQMLRRR	n.d.	n.d.
R11-XK-RbK	ATVRFTHQGEERQVDISRIRWVIRWGQYIWFKYDEDGGAKGWGYVSERDAPRELLQMLKRR	3	11
R11-PP4.8	ATVKFTHQGEEKQVDIRRIRWVIRWGQYIWFKYRRDGGAKGWGYVSEKDAPKELLQMLKKR	n.d.	n.d.
R11-PP5.5	ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSERRAPRRLLRMLKKR	n.d.	n.d.
R11-PP6.2	RRVRFTHQGEEKRVRISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKKR	n.d.	n.d.
R11.1.6-BDP	ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKKC-[BDP]	31	155
R11-XK-BDP	ATVRFTHQGEERQVDISRIRWVIRWGQYIWFRYDEDGGARGWGYVSERDAPRELLQMLRRC-[BDP]	n.d.	n.d.
R11-XK-R1218K-BDP	ATVRFTHQGEEKQVDISKIRWVIRWGQYIWFRYDEDGGARGWGYVSERDAPRELLQMLRRC-[BDP]	n.d.	n.d.
R11-XK-RbK-BDP	ATVRFTHQGEERQVDISRIRWVIRWGQYIWFKYDEDGGAKGWGYVSERDAPRELLQMLKRC-[BDP]	95	470
R11-PP4.8-BDP	ATVKFTHQGEEKQVDIRRIRWVIRWGQYIWFKYRRDGGAKGWGYVSEKDAPKELLQMLKKC-[BDP]	n.d.	n.d.
R11-PP5.5-BDP	ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSERRAPRRLLRMLKKC-[BDP]	n.d.	n.d.
R11-PP6.2-BDP	RRVRFTHQGEEKRVRISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKKC-[BDP]	n.d.	n.d.

∗ original characterization by [79]: 4 nM (G12D) and 41 nM (WT).

Interestingly, the punctate, nucleolar staining of R11-XK-BDP in the nuclei of both CHO-K1 (Figure 3-6, top) and SW48 G12D cells (Figure 3-7, bottom) look very similar to that observed with HeLa cells incubated with cR10-conjugated mCherry protein (Figure 3-7, top). cR10 is a cyclic cell-penetrating peptide consisting of 10 Arginines. Overall, the results suggest that an Arginine-rich mutant of R11.1.6 is capable of non-cytotoxically localizing in the nucleus.

Given that the 12-Arg R11-XK variant exhibited (1) strong nuclear entry and (2) low aqueous solubility upon conjugation of BODIPY, we reasoned that reverting some of the arginine mutations to lysines might (1) confer improved cytosolic uptake (vs. R11.1.6) without nuclear localization and (2) improved aqueous solubility. To this end, two variants were expressed, purified, and tested *in vitro*: (1) the 10-Arg R11-XK-R1218K, where R12 and R18 (the two least solvent-accessible residues; see Figure 3-10) are mutated back to Lys, and (2) the 9-Arg R11-RbK, where R32, R40, and R59 (the three Lys residues that bind to KRAS G12D) are mutated back to Lys. Both sequences are shown in Table 3.1. The rationale for designing R11-XK-R1218K is that a solvent-inaccessible residue (i.e. a buried residue) is more likely than a solvent-accessible one to be important in maintaining the structural stability of the protein; by reverting the solvent-inaccessible R12 and R18 back to Lys, we would improve the expression yield and aqueous solubility vs. R11-XK. The rationale for designing R11-XK-RbK is to restore full affinity to KRAS G12D.

A	T	V	K	F	T	H	Q	G	E	E	K	Q	V	D	I
0.5	0.3	0.0	0.7	0.1	0.6	0.9	0.9	0.9	0.8	0.9	0.4	0.6	0.0	0.5	0.1

S	K	I	K	W	V	I	R	W	G	Q	Y	I	W	F	K
0.7	0.3	0.1	0.7	0.5	0.2	0.4	0.5	0.7	0.7	0.8	0.4	0.0	0.3	0.0	0.5

Y	D	E	D	G	G	A	K	G	W	G	Y	V	S	E	K
0.1	0.5	0.6	0.3	1.0	0.7	0.6	0.8	0.3	0.7	0.1	0.4	0.1	0.2	0.3	0.9

D	A	P	K	E	L	L	Q	M	L	K	K	R			
0.7	0.2	0.2	0.7	0.4	0.0	0.2	0.7	0.4	0.1	0.9	1.0	0.9			

Figure 3-10: Relative solvent accessible surface area (SASA) of R11.1.6 residues. SASA values range from 0.0 (least solvent-accessible) to 1.0 (most solvent-accessible). Positive residues (Lys, K, and Arg, R) are highlighted in blue. SASA values were estimated using the get_sasa_relative function of PyMOL

The yields of R11-XK-RbK-BDP and R11-XK-R1218K-BDP were ∼ 1.5 mg/L-culture and ∼ 0.7 mg/L-culture respectively. As with the purification of R11-XK, a high salt concentration (0.5 M NaCl) was maintained in all purification buffers. Surprisingly, the restoration of solvent-inaccessible lysines (i.e. R11-XK-R1218K)

worsened the expression yield (1.5 for R11-XK to 0.7 for R11-XK-R1218K). Unlike R11-XK-BDP which fully precipitates in additive-free PBS, both R11-XK-RbK-BDP and R11-XK-R1218K-BDP are moderately soluble in additive-free PBS.

Both variants were incubated with SW48 G12D cells and cellular internalization was characterized via confocal microscopy. It was observed that both R11-XK-RbK-BDP and R11-XK-R1218K-BDP confer a mixture of nuclear and diffuse cytosolic uptake within the same incubation well (Figures 3-11 and 3-12). In a given well, around 90% of the cells would exhibit nuclear entry of the protein while ∼10% mainly exhibited cytosolic entry of the protein. Cells that exhibit nuclear entry of the protein had the same morphology as those with R11-XK-BDP described earlier (flatter and "web-like"; Figure 3-6); on the other hand, cells that exhibit cytosolic entry had a similar morphology to those incubated with est-R11-BDP (Figure 3-5). The confocal microscopy results of the Arg-rich variants (R11-XK, R11-XK-RbK, R11-XK-R1218K) suggest that the more Arginine-rich a variant, the greater the extent to which the nuclear membrane of a cell is compromised. With R11-XK, all cells observed have nuclear localization of protein; with R11-XK-RbK and R11-XK-R1218K, $\sim$ 90% of cells observed have nuclear localization of protein. The affinities of R11-XK-RbK for KRAS G12D and KRAS WT are 3 nM and 11 nM, respectively. We did not produce enough R11-XK-R1218K for downstream characterization via BLI, so its affinity is not determined.

Positive-patch variants

We also briefly explored if the clustering of Arg residues on the protein surface (as opposed to a more diffuse distribution of Arg, as is the case with the R11-XK, R11-XK-RbK, and R11-XK-R1218K variants) would confer greater cytosolic entry. We called these variants "positive patch" (PP) variants. In designing these PP variants, we adhered to a few design rules: (1) Avoid mutations of solvent-inaccessible residues, i.e. residues with relative SASA < 0.4; (2) Avoid mutations of Gly residues at loops to maintain the flexibility of loop structures [95]; (3) Avoid mutations of binding residues to maintain binding affinty to KRAS G12D.

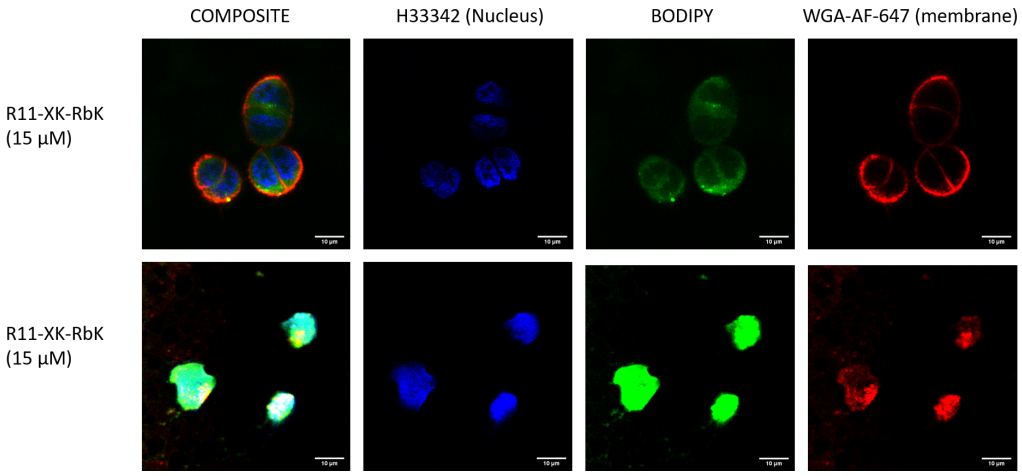


Figure 3-11: Live-cell fluorescence microscopy images of the cellular internalization of R11-XK-RbK-BDP in SW48 G12D cells. Cells were incubated with protein (15.0 μM) in FBS-free RPMI 1640 medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 4 h at 37 °C. The protein solution was then replaced with RPMI 1640 medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed, stained with Hoechst 33342 and wheat germ agglutinin (WGA)-Alexa Fluor 647, and imaged by confocal microscopy. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BDP: ex. 503 nm, em. 512 nm. Scale bars: 10 μm.

Three variants with a 5-Arg patch and net charge of +8.2 — PP4.8, PP5.5, PP6.2 — were designed, expressed, and purified (see Table 3.1 and Figure 3-13). Our initial attempt at purifying these PP variants led to very low yields. This observation is consistent with previous literature showing a correlation between protein insolubility and the presence of Arg-rich surface patches [96]. Since the aggregation of proteins with positive patches is often exacerbated by the presence of nucleic acids [96], we explored the use of DNases and a high-salt concentration (1 M NaCl) in the purification process to remove nucleic acids. These changes resulted in improved yields of $\sim$ 1.0 mg/L-culture for the unconjugated proteins (c.f. 3.0-4.0 mg/L-culture for unconjugated R11.1.6), but downstream conjugation with BODIPY still led to heavy precipitation. The final yield of the BDP-labeled products ranged from 0.12 to 0.40 mg/L-culture. We had enough R11-PP4.8-BDP and R11-PP5.5-BDP for confocal microscopy imaging in SW48 G12D cells. Microscopy imaging revealed a greater level of endosomal entry of these PP variants compared to est-R11.1.6 (Figure 3-5), sug-

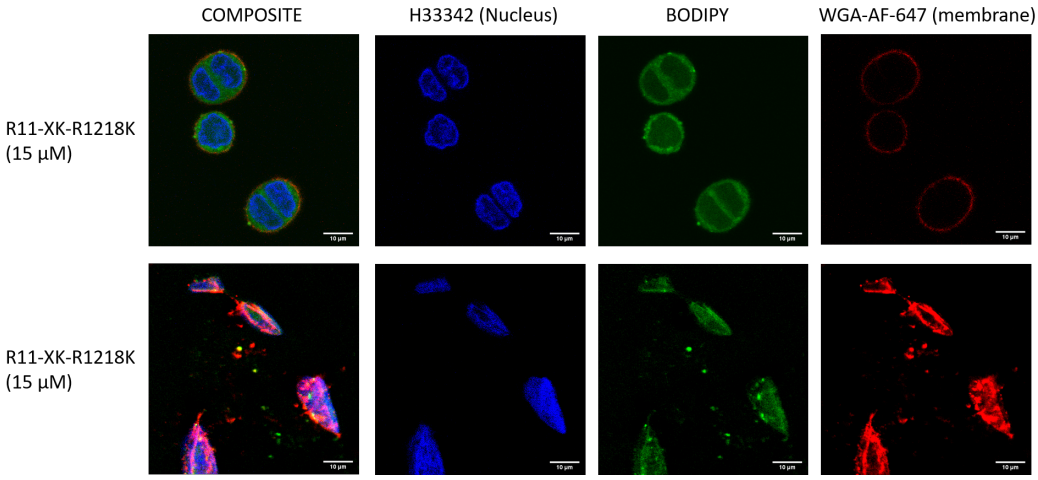


Figure 3-12: Live-cell fluorescence microscopy images of the cellular internalization of R11-XK-R1218K-BDP in SW48 G12D cells. Cells were incubated with protein (15.0 μM) in FBS-free RPMI 1640 medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 4 h at 37 °C. The protein solution was then replaced with RPMI 1640 medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed, stained with Hoechst 33342 and wheat germ agglutinin (WGA)-Alexa Fluor 647, and imaged by confocal microscopy. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BDP: ex. 503 nm, em. 512 nm. Scale bars: 10 μm.

gesting that the clustering of positive charges confer a high level of endosomal entry (vs. cytosolic entry).

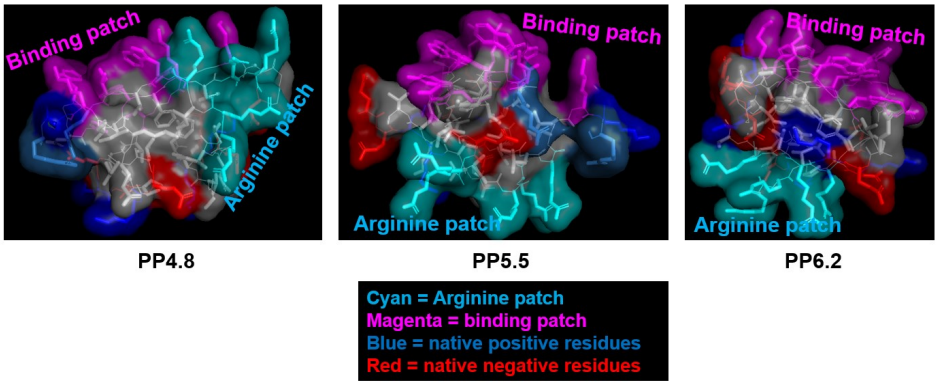


Figure 3-13: Structure of PP variants. Binding residues are colored magenta; Arg patch is colored cyan; basic residues (Arg, Lys) are colored blue; acidic residues (Asp, Glu) are colored red.

The level of cellular entry of each variant in SW48 G12D cells was quantitatively

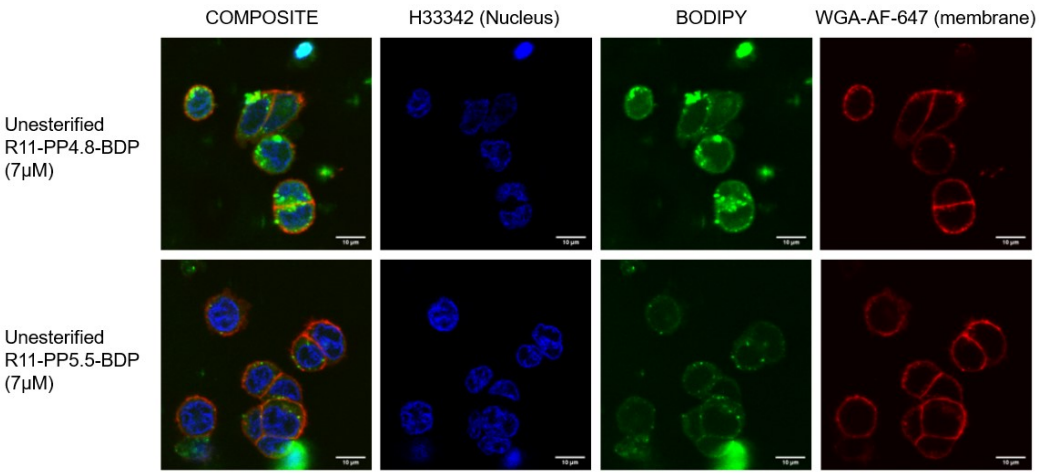


Figure 3-14: Live-cell fluorescence microscopy images of the cellular internalization of R11-PP4.8-BDP and R11-PP5.5-BDP in SW48 G12D cells. Cells were incubated with protein (7.0 μM) in FBS-free RPMI 1640 medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 4 h at 37 °C. The protein solution was then replaced with RPMI 1640 medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed, stained with Hoechst 33342 and wheat germ agglutinin (WGA)-Alexa Fluor 647, and imaged by confocal microscopy. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BDP: ex. 503 nm, em. 512 nm. Scale bars: 10 μm.

determined via flow cytometry (Figure 3-15). Note: YW1 is a non-binding variant of R11.1.6 that serves as a negative control [79, 88]. We see that esterified R11.1.6 and esterified YW1 achieve a statistically significantly higher level of cellular internalization compared to the unesterified control. Additionally, we observe a non-significant increase in cellular entry for the lysine-poor variants (R11-XK-RbK and R11-XK-R1218K) compared to R11.1.6. R11-PP4.8 achieves a far higher degree of cellular entry than R11.1.6, but it is important to realize that a fluorescence readout via flow cytometry does not distinguish between cytosolic and endosomal entry of protein. As seen via confocal microscopy (Figure 3-14), PP variants get taken up by cells via an endosomal pathway, suggesting that the three-fold higher MFI value for R11-PP4.8 is primarily endosomal. Additionally, PP variants suffered from very poor yield owing to their severe insolubility, making it an impractical choice going forward.

Given the difficulty of assessing the extent of cytosolic entry of R11.1.6 variants via confocal microscopy (an inherently non-quantitative method), we decided that the

development of an R11.1.6-based large-molecule PROTAC along with activity-based assays would be a more efficient use of our resources.

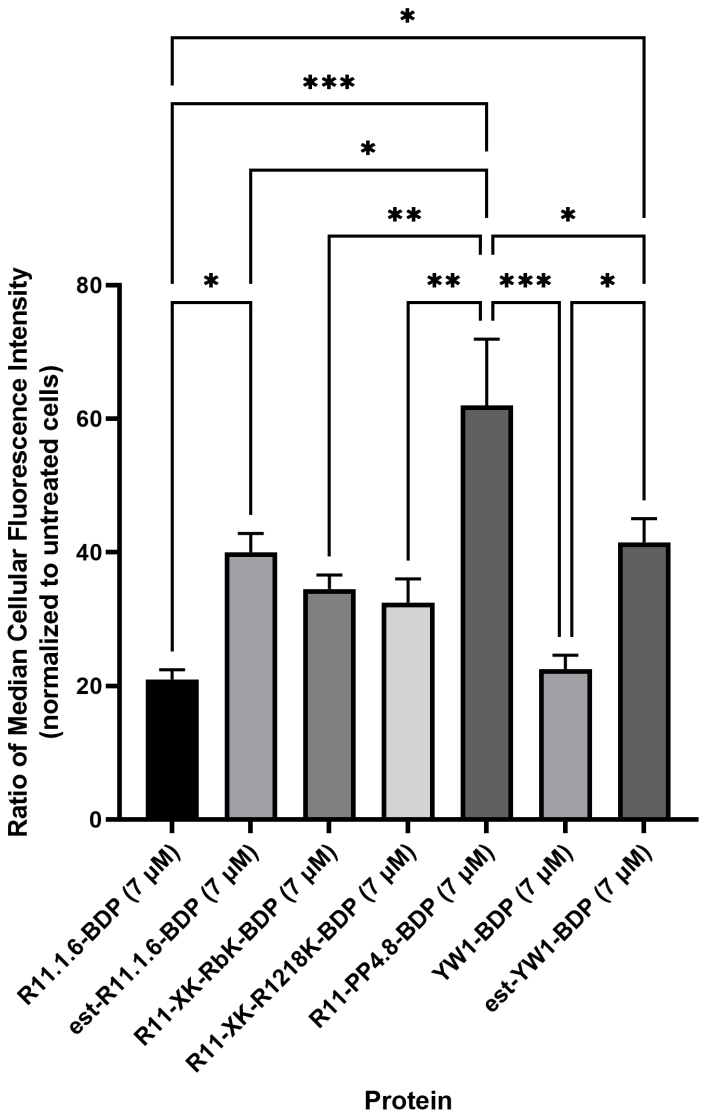


Figure 3-15: Cellular internalization of BDP-labeled R11.1.6, esterified R11.1.6, and R11 variants. Data were obtained with flow cytometry. SW48 G12D cells were incubated with protein in FBS-free RPMI 1640 medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin for 2 h at 37 °C, washed, and extracellular BDP quenched with α-BODIPY antibody for 30 min. Median cellular fluorescence intensities are normalized to that of untreated cells. Values are the mean ± SD with two technical replicates per protein. Statistical analysis by ordinary one-way ANOVA. **$p \leq 0.01$.

3.2.3 Development of an R11.1.6-based PROTAC

Motivation for large-molecule PROTACs

Early work on PROTACs led to the development of several peptide-based PROTACs (i.e. PROTACs with peptide E3-binding- and POI-ligands), although many of these suffered from low potency in the micromolar range, likely due to poor cell penetration ([90, 97]). Given this limitation, efforts have been largely focused on the generation of small-molecule PROTACs instead, because of their increased cell permeability. Specifically, two E3 ligases have been the target of most small-molecule PROTACs designed until now: VHL (Von Hippel-Lindau) and CRBN (cereblon). Thus far, small-molecule PROTACs have shown enhanced target selectivity relative to their small-molecule inhibitors, with many of them showing nanomolar potency and some showing in vivo efficacy as well [98, 99]. However, limitations of small molecules (compared to proteins) persist in the field of PROTACs: off-target effects have been observed in some studies of small-molecule PROTACs [99, 100, 101] and certain targets like KRAS G12D have remained "undruggable" using small-molecule PROTACs.

High-affinity noncovalent ligands against, KRAS G12D have proven futile due to the lack of small molecule binding pockets on the protein surface, but recent work from the Partridge Lab has demonstrated the successful degradation of KRAS G12D in cells that intracellularly express protein-based biodegraders, i.e. proteins that fuse a high-affinity large-molecule binder (e.g. RBD, K27, R11.1.6, and NS1) to an E3 ligase (e.g. SPOP) (Lim et al. 2021). Notably, in their work, R11.1.6 fused to VHL was highly effective at degrading endogenous RAS. The development of an R11.1.6-based large-molecule PROTAC that would be delivered across the cell membrane (i.e. our work) would serve as a complementary strategy to biodegraders and would be, to our knowledge, the first-ever exogenous delivery of a large-molecule PROTAC.

Design of an R11.1.6-based large-molecule PROTAC

Equipped with the cell-internalization strategies described earlier, we sought to design a cell-permeable KRAS-G12D-targeting PROTAC molecule with R11.1.6 or R11-XK-RbK (hereinafter referred to as "RbK") as the POI ligand. We chose the small-molecule VHL Ligand 1 (VL1) as the E3 ligase ligand given (1) its extensive use in small molecule-based targeted degradation and (2) the success of R11.1.6-VHL in the degradation of endogenous RAS [9]. VL1 was conjugated to the POI ligand (R11.1.6) via maleimide chemistry and a PEG-based linker was used (Figure 3-16).

To develop a cell-permeable KRAS-G12D-targeting R11.1.6 PROTAC, we employed the following design cycle:

1. Step 1: Express and purify the POI ligand (R11.1.6 or RbK). Assess its aqueous stability and affinity to KRAS G12D and KRAS WT.

2. Step 2: Produce the degrader construct (R11.1.6-VL1 or RbK-VL1) by conjugating VL1 to the POI ligand. Assess its aqueous stability and affinity to KRAS G12D and KRAS WT. The binding affinities of all constructs produced and tested are shown in Table 3.3

3. (For R11.1.6-VL1 only) Esterify the degrader construct if desired. When esterified, each degrader molecule is labeled with an average of 3 to 5 esters.

4. Test degradation *in vitro* efficiency in KRAS-G12D-expressing cell lines (Table 3.2). Cells are incubated with the degrader construct for 4 hours at 37 °C followed by 20, 44, or 68 hours in degrader-free medium before the relevant metric is read out.

A note on N-terminal vs. C-terminal conjugation. When conjugating (via maleimide chemistry) BDP or VL1 to the Sso7d protein, we had a choice of whether to introduce an N-terminal or C-terminal cysteine to the Sso7d protein. We found the expression yield and purity of the C-terminal variant noticeably higher than that of the N-terminal variant (yield of ∼1.5 mg/L-culture and purity of ∼95% for the C-terminal variant vs. yield of ∼0.8 mg/L-culture and purity of ∼80% for the N-terminal

variant). However, we were concerned that the conjugation of a small molecule at the C terminus of the Sso7d protein would impair binding to KRAS G12D. As we suspected, R11-VL1, which has VL1 conjugated at the C terminus of R11, has a K_d value against KRAS G12D (60 nM; see Table 3.3) that is several-fold higher than that of R11 (6 nM). On the other hand, VL1-R11, which has VL1 conjugated at the N terminus of R11, maintains its affinity for KRAS G12D (3 nM; see Table 3.3). We reasoned that the loss in affinity upon the conjugation of a small molecule at the C terminus is likely due to the proximity of the C terminus to the binding site of R11 (see Figure 3-1); the small molecule might sterically hinder binding to KRAS G12D. The N terminus, on the other hand, is on the opposite face of the binding site, so conjugation of a small molecule at the N terminus does not affect binding to KRAS G12D. There was thus a trade-off between yield and purity vs. affinity. We decided it was more practical to favor yield and purity, so the C-terminal variants were produced and tested for almost all experiments.

After incubation with the construct, two assays were performed to assess performance:

a. Western Blot of KRAS G12D and RAS

b. Viability assay (to assess growth inhibition of cells)

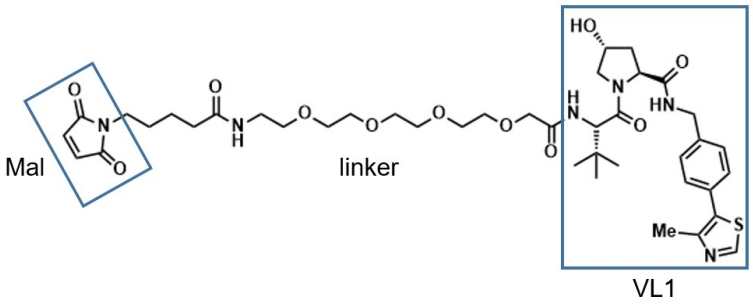


Figure 3-16: Structure of VL1-PEG-Maleimide. VL1 = VHL Ligand 1. MW = 843.01 g/mol.

Western blot analysis

Western blot analysis of HPAF-II, SW48 G12D, and AsPC-1 cells is shown in Figures 3-17 to 3-18. Cells were incubated with degrader (either RbK-VL1 or est-R11-VL1)

Table 3.2: Human cancer cell lines that express KRAS G12D

Cell line	**Tissue**	**Number of mutant alleles**
AsPC-1	Pancreas	2
HPAF-II	Pancreas	1
LS180	Colon	1
SW48 G12D	Colon	1

Table 3.3: Degrader constructs and its K_d values.

		K_d (nM)	
Construct	**Sso7d terminus of conjugation**	**KRAS G12D**	**KRAS WT**
R11*	-	6	17
R11-VL1	C	60	148
VL1-R11	N	6	36
RbK	-	3	11
RbK-VL1	C	22	102

$*$ original characterization by [79]: 4 nM (G12D) and 41 nM (WT).

at various concentrations, along with relevant negative controls. For est-R11-VL1, the negative controls are the no-protein control (NPC) and unesterified R11-VL1; for RbK-VL1, the negative control is the NPC. Cells were lysed and western blots were performed. Degradation of KRAS G12D or pan-RAS was detected using a KRAS G12D and pan-RAS antibody, respectively. Beta-actin (BA) was used as the loading control. Densitometry analysis was performed using ImageJ software. For each lane, the density of the KRAS G12D or RAS band is normalized to the density of the BA band, i.e. one calculated value for each lane; the BA-normalized densities of all lanes are then normalized again to the BA-normalized value of only the NPC lane (which explains the consistent value of 1.0 for all NPC values shown in Figures 3-17 to 3-18).

Disappointingly, of all 12 combinations of cell lines (x3), degraders (x2), and RAS isoforms detected (x2), only HPAF-II cells revealed statistically significant degradation of pan-RAS with the est-R11-VL1 degraded (Figure 3-18). The large variation

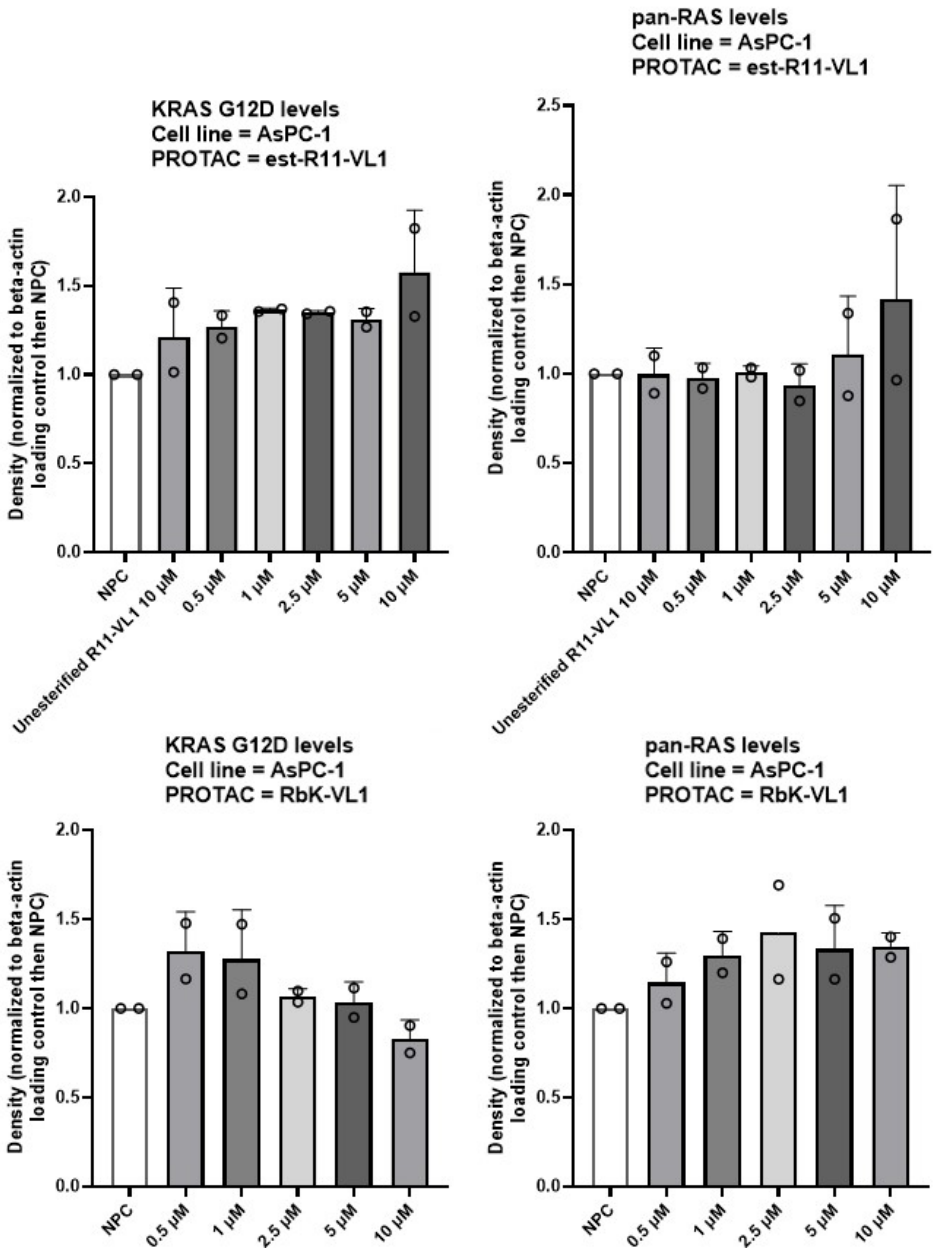


Figure 3-17: Western blot analysis of AsPC-1 wells incubated with est-R11-VL1 or RbK-VL1. Degradation of KRAS G12D or pan-RAS was detected using a KRAS G12D and pan-RAS antibody, respectively. For each sample/lane, densities are first normalized to that of the beta-actin band, then normalized again to the BA-normalized density of the untreated (NPC) cells. Values are the mean ± SD with two technical replicates per sample. Statistical analysis by ordinary one-way ANOVA did not reveal any statistically significantly different populations.

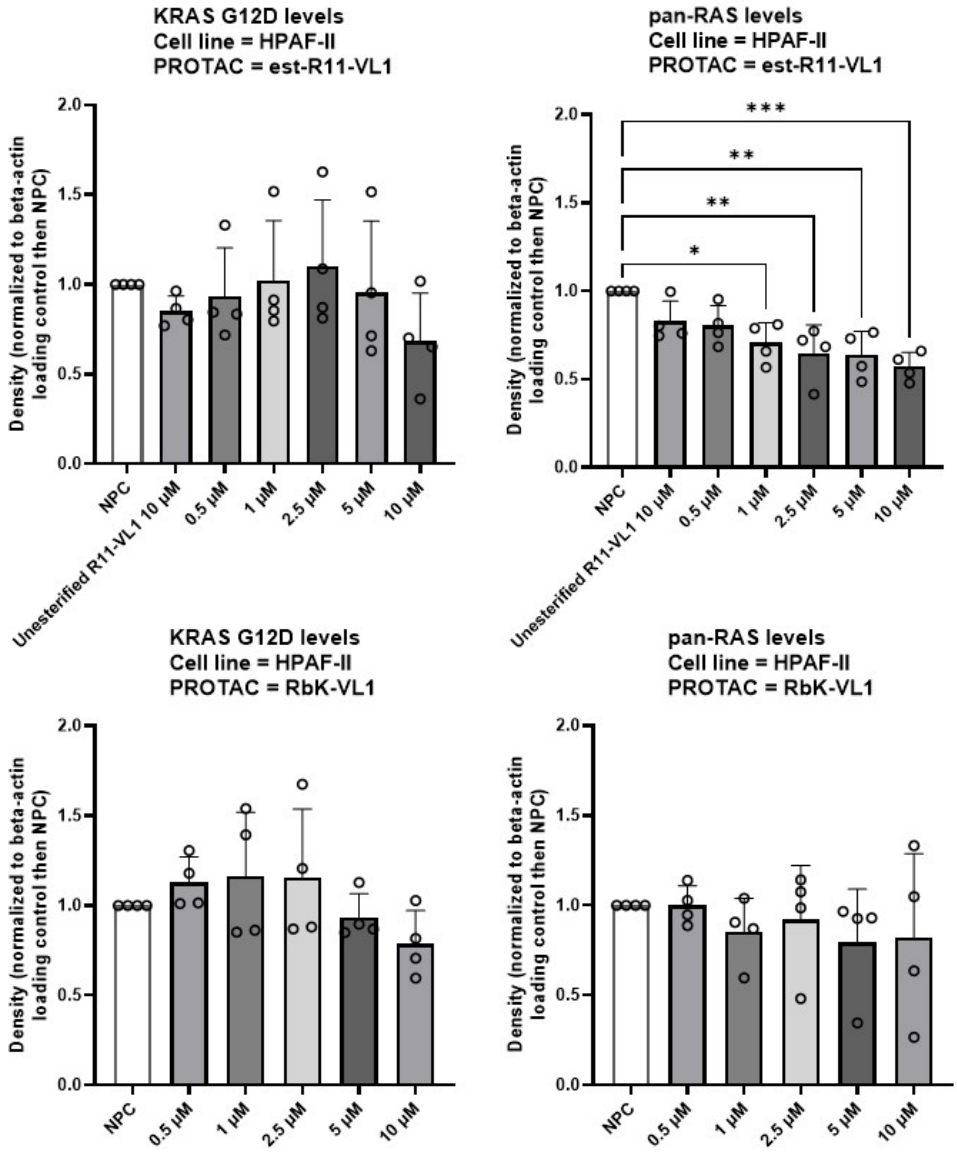


Figure 3-18: Western blot analysis of HPAF-II wells incubated with est-R11-VL1 or RbK-VL1. Degradation of KRAS G12D or pan-RAS was detected using a KRAS G12D and pan-RAS antibody, respectively. For each sample/lane, densities are first normalized to that of the beta-actin band, then normalized again to the BA-normalized density of the untreated (NPC) cells. Values are the mean ± SD with two technical replicates per sample. Statistical analysis by ordinary one-way ANOVA. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.005$.

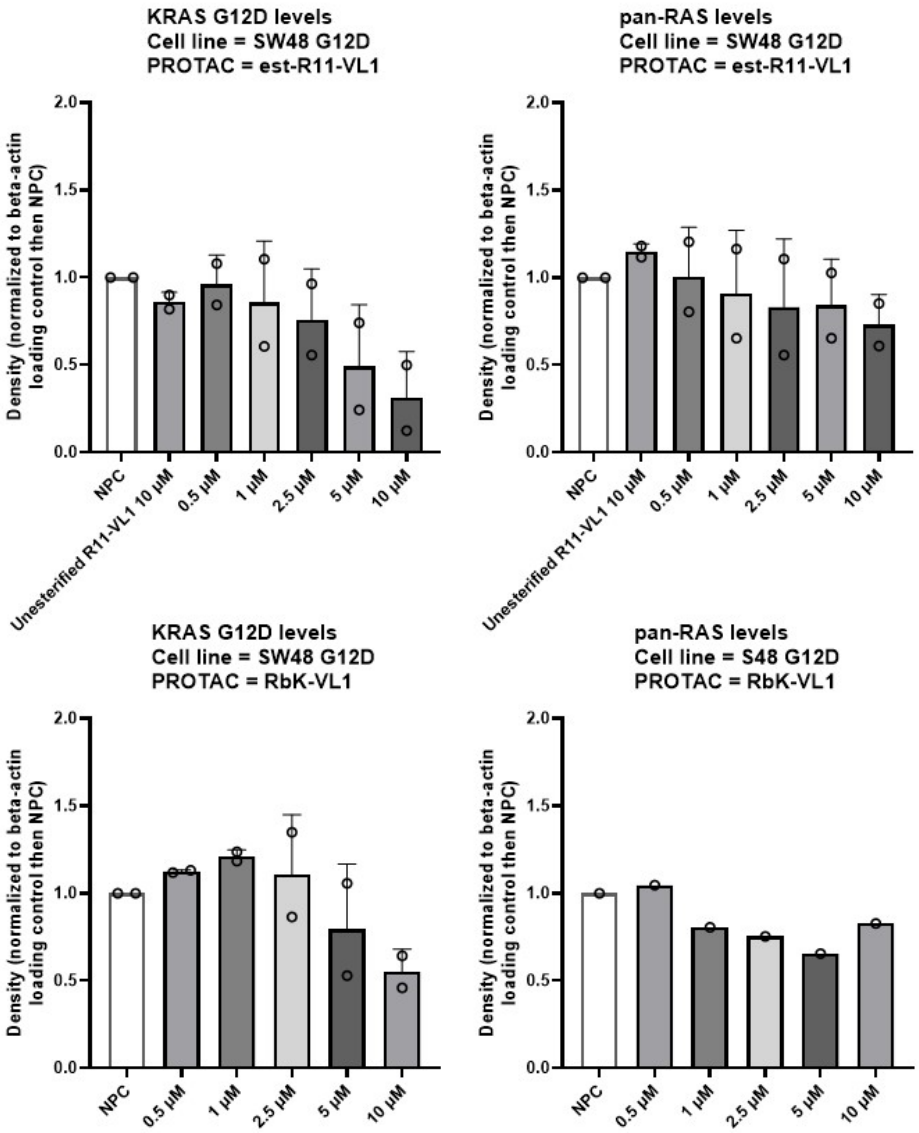


Figure 3-19: Western blot analysis of SW48 G12D wells incubated with est-R11-VL1 or RbK-VL1. Degradation of KRAS G12D or pan-RAS was detected using a KRAS G12D and pan-RAS antibody, respectively. For each sample/lane, densities are first normalized to that of the beta-actin band, then normalized again to the BA-normalized density of the untreated (NPC) cells. Values are the mean ± SD with two technical replicates per sample (except for pan-RAS/RbK-VL1 in the bottom right, which only has one technical replicate). Statistical analysis by ordinary one-way ANOVA did not reveal any statistically significantly different populations.

in normalized densitometry values (notice the large error bars in many of the plots) highlights the inherent "noisiness" of western blots, prompting us to employ a more quantitative method such as growth inhibition assays.

Growth inhibition assays

Cell TiterGlo 2.0 was used for growth inhibition assays, where cells incubated with degraders are lysed and their ATP (a proxy for live cells) measured to determine the number of living cells in each well relative to a degrader-free control well. Growth inhibition assays were performed for all 4 KRAS-G12D-expressing cell lines for various incubation times and a range of concentrations. Cells were treated with degraders (or controls) for 4 hours in FBS-free medium at 37 °C, then 68 hours in FBS-containing medium at 37 °C. Cell viability was then measured with a CellTiter Glo 2.0 Assay from Promega according to the manufacturer's instructions. The results of several key experiments are shown in Figures 3-20 and 3-21.

Figure 3-20 shows the effect of the degrader, est-R11-VL1, on the growth of four different cell lines. The negative controls that were used are described in Table 3.4. Promisingly, we observed statistically significant growth inhibition in LS180 (32% inhibition), HPAF-II (12% inhibition), and AsPC-1 cells (23% inhibition), relative to the no-protein control (NPC). Note that lower concentrations of degrader (0.1 μM and 1.0 μM) did not achieve statistically significant growth inhibition. Growth inhibition was not observed in SW48 G12D cells for any of the constructs, suggesting that insufficient RAS was degraded to suppress its growth. We also hypothesize that SW48 G12D cells possibly do not rely on RAS proteins for survival, since we do observe a clear reduction (albeit not statistically significant) in KRAS G12D levels at 10 μM of est-R11-VL1 (top left quadrant of Figure 3-19). On the other hand, it was surprising to observe a statistically significant inhibition of growth of AsPC-1 cells given the lack of RAS degradation observed via western blot analysis (Figure 3-17).

It was also observed that esterified R11 alone did not cause any growth inhibition in all tested cell lines (est-R11 was not tested with AsPC-1 cells), suggesting that exogenous delivery of a stoichiometric inhibitor (est-R11) is ineffective at suppressing

growth. Moreover, unesterified R11-VL1 did not cause growth inhibition in all four cell lines, suggesting that the improved cytosolic entry conferred by esterification (see Figure 3-15) was responsible for delivering sufficient levels of degrader to suppress growth. Non-RAS-binding constructs (YW1, YW1-VL1, est-YW1-VL1) were also found not to cause any growth inhibition in all tested cell lines (untested with AsPC-1 cells), confirming that growth inhibition is driven specifically by RAS targeting.

The N-terminal variant, VL1-R11, was also expressed, purified, and esterified for testing via growth inhibition assays (Figure 3-21). VL1-R11 has a similar affinity for KRAS G12D as does R11 (6 nM), whereas the C-terminal variant, R11-VL1, has a log-fold weaker affinity of 60 nM (see Table 3.3). A greater level of growth inhibition is observed with est-VL1-R11 in HPAF-II cells (18% inhibition for est-R11-VL1, 36% inhibition with est-VL1-R11), but not AsPC-1 cells (23% inhibition with both est-R11-VL1 and est-VL1-R11) (Figure 3-21). This suggests that increased affinity to KRAS G12D may play a role in the extent of RAS degradation. Additional studies with N-terminal variants and other cell lines are required to conclusively compare the C- vs. N-terminal variants. Finally, we also observed statistically significant growth inhibition of AsPC-1 and HPAF-II cells treated with the lysine-poor degrader variant, RbK-VL1 (Figure 3-21). However, the extent of growth inhibition (13% with HPAF-II cells, 8% with AsPC-1 cells) was less than that achieved by est-R11-VL1 and est-VL1-R11.

Besides growth inhibition assays, we also attempted to quantitatively assess the level of intracellular KRAS G12D via intracellular flow cytometry (data not shown). We performed a few pilot experiments with two degrader constructs (est-R11-VL1, RbK-VL1), four different anti-KRas antibodies (one pan-KRAS, three specific to KRAS-G12D), and three cell lines (AsPC-1, HPAF-II, SW48 G12D). Unfortunately, we did not observe a consistent reduction in fluorescence intensities with the addition of degraders (which might actually be due to the lack of degradation conferred by these constructs; we did not have a positive control, i.e. a functional KRas-G12D degrader, so it is difficult to say). This technique might not be sensitive enough to detect changes in the levels of endogenous KRAS. Further troubleshooting of the

protocol will be necessary.

Table 3.4: Degraders and negative controls

Construct	Binds	Does not bind
R11	KRAS G12D	VHL
R11-VL1	KRAS G12D, VHL	-
est-R11-VL1	KRAS G12D, VHL	-
YW1	-	KRAS G12D, VHL
YW1-VL1	VHL	KRAS G12D
est-YW1-VL1	VHL	KRAS G12D

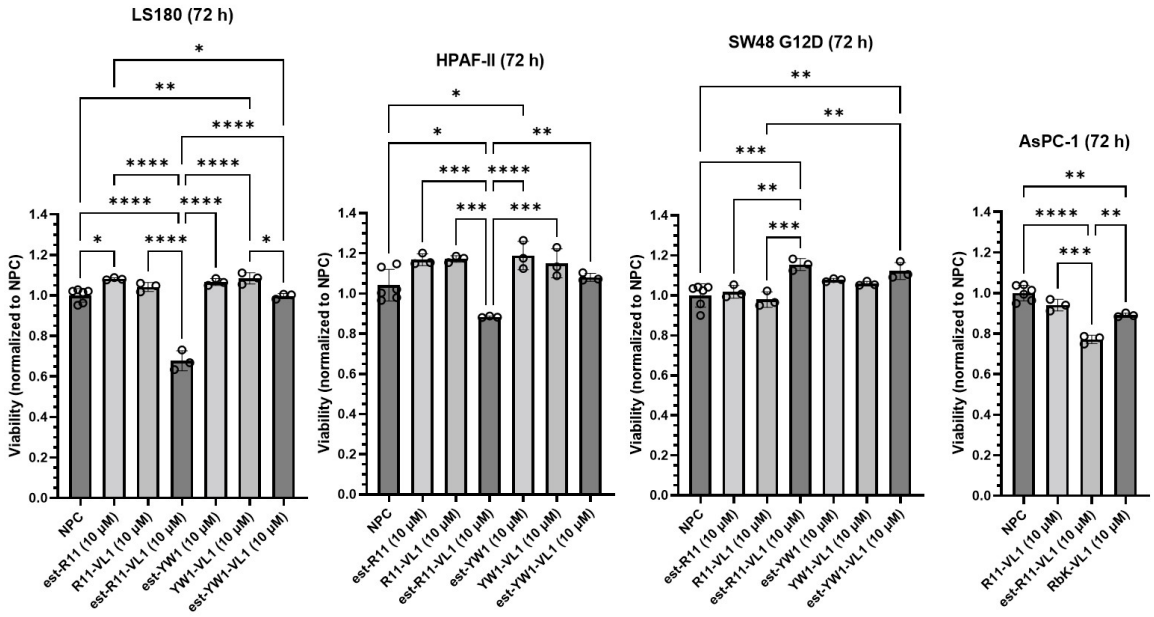


Figure 3-20: Effect of degraders on the viability of cells. Each graph represents a separate experiment. Cells were incubated for 4 h at 37 °C with degraders or controls in FBS-free medium, followed by 68 h at 37 ° with degrader/protein-free FBS-containing medium. Cell viability was measured with a CellTiter Glo 2.0 Assay from Promega according to the manufacturer's instructions. Viability is normalized to that of the no-protein control (NPC). Values are the mean ± SD with three technical replicates per sample (six for NPC). Statistical analysis by ordinary one-way ANOVA. *ple0.05, **p≤0.01, ***p≤0.005, ****p≤0.001.

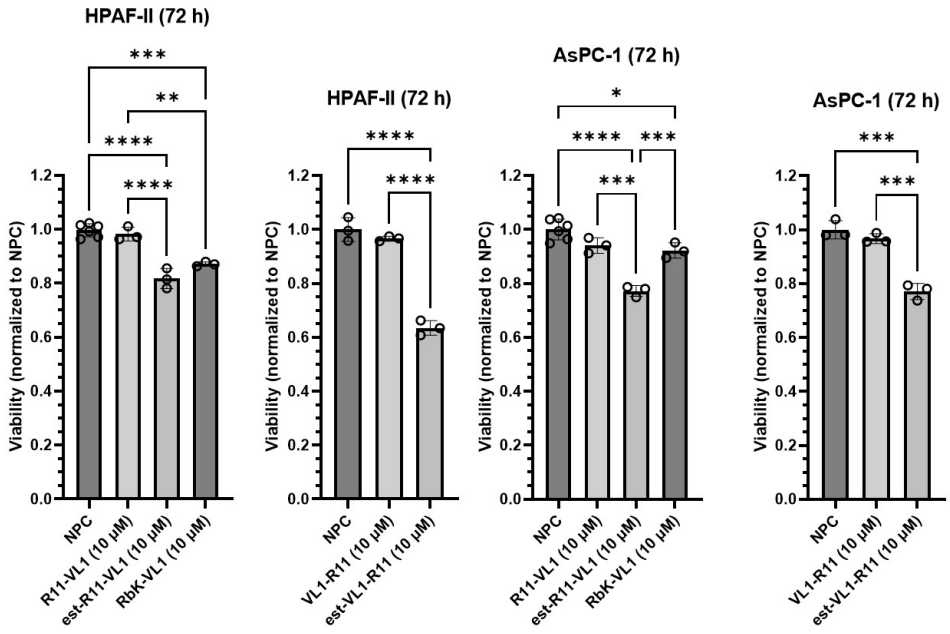


Figure 3-21: Effect of degraders on the viability of cells. Each graph represents a separate experiment. Cells were incubated for 4 h at 37 °C with degraders or controls in FBS-free medium, followed by 68 h at 37 ° with degrader/protein-free FBS-containing medium. Cell viability was measured with a CellTiter Glo 2.0 Assay from Promega according to the manufacturer's instructions. Viability is normalized to that of the no-protein control (NPC). Values are the mean ± SD with three technical replicates per sample (six replicates for some NPC samples). Statistical analysis by ordinary one-way ANOVA. *p≤0.05, **p≤0.01, ***p≤0.005, ****p≤0.001.

3.3 Conclusions

Esterified R11-VL1 was shown to be mildly efficacious at degrading intracellular RAS in tumor cell lines, with statistically significant results achieved in specific cell lines. Pan-RAS degradation was achieved in HPAF-II cells treated with 10 μM est-R11-VL1, and growth inhibition was observed with HPAF-II, AsPC-1, and LS180 cells treated with 10 μM est-R11-VL1.

Additionally, given the challenges of reliably detecting RAS degradation via western blots, we propose the use of more quantitative methods in the future development of large-molecule PROTACs. Specifically, the creation of tumor cell lines that stably express HiBiT-tagged KRAS G12D (via CRISPR knock-in) [102] would allow for more robust and quantitative detection of KRAS G12D levels. Briefly, KRAS-G12D-

HiBiT-expressing tumor cells treated with a degrader of interest would be lysed, chased with LgBiT, and its luminescence read out using a plate reader; luminescence is generated via structural complementation of LgBiT protein to the HiBiT-tag.

Even though the exogenous delivery of a large-molecule PROTAC was only modestly effective against KRAS G12D, our results support a novel paradigm for targeting previously undruggable proteins. This strategy might be effective against a more weakly-expressed protein target; RAS was possibly too difficult a target given its high expression levels.

3.4 Materials and Methods

Materials. Buffers were Phosphate-buffered saline (PBS) (Corning, 21-040-CV), Dulbecco's phosphate-buffered saline (DPBS) with calcium and magnesium (Gibco, 14040117), HEPES (Corning, 25-060-CI). 2-Diazo-2-(p-methylphenyl)-N,N-dimethylacetamide was synthesized as described previously [78].

Conditions. All procedures were performed in air at ambient temperature (∼22 °C) and pressure (1.0 atm) unless indicated otherwise.

Esterification of Sso7d. Sso7d proteins were esterified essentially as described previously [8]. Briefly, proteins were buffer exchanged into 20 mM HEPES buffer (pH 5.5). To the buffer exchanged protein was added α-aryl-α-diazoacetamide in acetonitrile (100 equiv.) and the acetonitrile concentration was adjusted to 20%. The final concentration of protein in all reactions was 110 μM. The resulting solution was incubated overnight at 37 °C. The esterification reaction was quenched by adding twice the volume of PBS (1x, pH 7.4) to the reaction volume, and the solution was filtered through a 0.22 μm cellulose acetate centrifuge tube filter (Corning Costar Spin-X) to remove any precipitated proteins. Esterified proteins were further purified with a PD-10 desalting column to remove unreacted esterification reagent, and proteins were eluted off the column with PBS (1x, pH 7.4). The eluted proteins were further

concentrated with an Amicon Ultra spin concentrator (Millipore). The concentration of the soluble, esterified protein was determined using Pierce BCA Protein Assay. A Q-TOF mass spectrum of esterified proteins was acquired, and the number of esters per protein was assigned from the mass of the peak with the highest relative intensity in the Q-TOF mass spectrum.

Cell culture. HPAF-II (obtained from ATCC [79]) and LS180 (obtained from ATCC [79]) cells were cultured in Eagle Minimum Essential Medium (EMEM). SW48 G12D (kind gift from the White Lab, Massachusetts Institute of Technology [79]) and AsPC-1 (kind gift from the Irvine Lab, Massachusetts Institute of Technology) were cultured in RPMI 1640. CHO-K1 (obtained from ATCC) cells were cultured in F12K medium. All media were supplemented with 10% Fetal Bovine Serum (Life Technologies), 100 U/mL penicillin, and 100 μg/mL streptomycin (Gibco), unless stated otherwise. The cells were grown in sterile culture flasks in a cell culture incubator at 37 °C under CO_2(g) (5% v/v). Cell lines were passaged a minimum of five times and up to twelve times before use. The cells were counted to determine seeding density using a hemacytometer.

Quantification of est-GFP internalization Quantification of est-GFP internalization in CHO-K1 cells was performed via flow cytometry. CHO-K1 cells were seeded in FBS-free culture medium (which was F12K medium supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin) at a density of 50,000 cells $well^{-1}$ in a sterile 48-well dish at 24 h prior to treatment. On the day of treatment, an esterified GFP solution (64 μM; with a median of 10 ester labels) in PBS supplemented with β-cyclodextrin (5 mM), L-arginine (250 mM), and sucrose (750 mM) was diluted 8-fold into FBS-free culture medium to obtain a final GFP or est-GFP concentration of 8 μM; the final concentrations of additives were 0.625 mM β-cyclodextrin, 31.25 mM L-arginine, and 93.75 mM sucrose. These final concentrations were determined to be noncytotoxic to CHO-K1 cells, as described in section 2.5. The culture medium was removed, and cells were incubated with 300 μL of the incubation solutions for

4 h at 37 °C. After removal of the incubation solutions, the cells were rinsed twice with PBS and released from the plate with 200 μL of warmed 0.25% v/v trypsin-ethylenediaminetetraacetic acid (EDTA) mix. Trypsin was quenched by the addition of 400 μL of the medium. The cells were then subjected to centrifugation for 5 min at 200g followed by aspiration of the supernatant. The cells were washed twice by re-suspension in 500 μL of ice-cold PBS, centrifugation at 200g for 5 min, and aspiration of supernatant. The cells were resuspended in 500 μL of ice-cold PBS supplemented with bovine serum albumin (2% w/v) and propidium iodide (2 μg mL^{-1}) and kept on ice until the time of analysis. The GFP fluorescence intensity of at least 10,000 events was measured with a BD FACS Celesta flow cytometer. The median fluorescence intensity (MFI) of live, single cells was recorded. Flow cytometry data were analyzed using FlowJo software.

To determine the number of GFP molecules the MFI corresponds to, we utilized AcGFP flow cytometer calibration beads (TakaraBio 632594) according to the manufacturer's instructions. It was determined that the MFI of the live, single cells was equivalent to 302,701 AcGFP molecules. However, since est-GFP and AcGFP have different fluorescent intensities, we first needed to determine the ratio of their fluorescence intensities so we can adjust the MFI-to-molecules calculation accordingly. The ratio of fluorescence intensities was determined by measuring the fluorescence (with a Tecan Infinite 200 Pro plate reader) of known concentrations of purified AcGFP (obtained commercially from TakaraBio 632502) and superfolder GFP (sfGFP, i.e. the variant of GFP used in all our exogenous delivery experiments, expressed and purified using the protocol described in section 2.5) in PBS, as shown in Table 3.5. The fluorescence intensity (FI) of a protein-free well of only PBS was subtracted from each measurement ("Blank-subtracted fluorescence intensity"). For each protein (AcGFP and sfGFP), a linear regression line (x = protein concentration in μM; y = blank-subtracted FI) was fitted with a pre-set intercept of 0. This generated regression lines of FI_{AcGFP}=3061×[AcGFP, μM] and FI_{sfGFP}=4624×[sfGFP, μM] for AcGFP and sfGFP, respectively. The FI ratio of sfGFP to AcGFP is thus 4624:3061 = 1.51. Therefore, the MFI of the live, single cells incubated with est-GFP is equivalent to

(302,701/1.51=)200,382 sfGFP molecules.

Table 3.5: Fluorescence intensities of AcGFP and sfGFP

AcGFP	**μM**	8.90	4.45	2.23	1.11	0.56
	Blank-subtracted fluorescence intensity	27126	14826	4998	3202	1523
sfGFP	**μM**	7.41	3.71	1.85	0.93	0.46
	Blank-subtracted fluorescence intensity	34035	17733	8347	3757	3179

Calculation of pI and extinction coefficient values. pI values were determined using the Expasy Compute pI/MW tool. Extinction coefficient values were determined using the Expasy ProtParam tool.

Plasmids and cloning. For site-directed mutagenesis within a 30 bp range, primers were designed using NEBaseChanger (https://nebasechanger.neb.com/) and synthesized by Integrated DNA technologies (IDT). PCR amplification, ligation, and transformation were performed in accordance with the New England Biolabs Q5 Site-Directed Mutagenesis Kit (E0554) Quick Protocol. For the lysine-free/-poor plasmids (R11-XK, R11-XK-RbK, R11-XK-R1218K), gBlocks of the Sso7d sequence were synthesized by IDT and cloned into the pE-SUMO backbone using the CloneAmp HiFi PCR Premix and the In-Fusion® HD Cloning Kit as per the manufacturers' protocols.

Protein expression and purification of Sso7d proteins.

R11.1.6 and its variants (R11-XK, R11-XK-RbK, R11-XK-R1218K, R11-PP4.8, R11-PP5.5, R11-PP6.2, YW1) were expressed as fusion proteins consisting of an N-terminal hexahistidine (H6) tag, followed by small ubiquitin-like modifier (SUMO) and Sso7d, using the pE-SUMO-vector previously described by [79]. The proteins were produced in Rosetta 2 (DE3) E. coli cells. Luria Broth (LB) containing kanamycin (50 μg/mL) and carbenicillin (100 μg/mL) was inoculated with transformed BL21 cells and grown overnight at 37 °C. On the following day, LB containing kanamycin (50 μg/mL) and carbenicillin (100 μg/mL) was inoculated (1:100 v/v ratio) with the

overnight culture. Cell growth was monitored until the optical density (O.D.) reached 0.6-0.8. Protein expression was then induced by adding IPTG to a final concentration of 1.0 mM. The temperature was reduced to 20 °C, and protein expression was allowed to continue overnight. Rosetta cells were pelleted by centrifugation at 4,300 g for 15 min at 4 °C and the cell pellet flash frozen in liquid nitrogen and stored at -80 °C.

For protein purification, frozen cell pellets were thawed on ice then resuspended in 10 mL sonication buffer (300mM NaCl, 50 mM sodium phosphate, 10% glycerol, cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche 11873580001; amount according to manufacturer's instructions), pH 7.3) per 500 mL culture, and lysed by sonication. The lysate was then clarified by centrifugation at 10,000 g for 25 minutes at 4 °C to remove unbroken cells and other extraneous cell debris, and filtered (0.45 μm pore membrane vacuum filtration) to yield clear lysate.

A column of TALON Superflow Metal Affinity Resin (Takara 635669) (2 mL of suspended resin per 100 mL of Rosetta culture) was equilibrated with 10 column volumes (CVs) of wash buffer 1 before the supernatant was loaded. H6-SUMO-R11.1.6 bound to the resin. The resin was washed with 15 CVs of wash buffer 1 (PBS (5.60 mM Na_2HPO_4, 1.06 mM KH_2PO_4, 154 mM NaCl), pH 7.3) followed by 15 CVs of wash buffer 2 (PBS (5.60 mM Na_2HPO_4, 1.06 mM KH_2PO_4, 154 mM NaCl), 10 mM imidazole, pH 7.3). Finally, 5 CVs of elution buffer (PBS (5.60 mM Na_2HPO_4, 1.06 mM KH_2PO_4, 154 mM NaCl), 150 mM imidazole, pH 7.3) were used to elute the product, H6-SUMO-Sso7d. The product was buffer-exchanged to PBS (to remove excess imidazole which would affect the activity of the protease Ulp1) and concentrated to 2-5 mg/mL. The N-terminal H6-SUMO tag was cleaved with overnight digestion with protease Ulp1 (SUMO protease 1) at 4 °C. The digestion product was loaded onto a pre-washed (with 10 CVs of wash buffer 1) column of TALON Superflow Metal Affinity Resin (Takara 635669) (2 mL of suspended resin per 100 mL of Rosetta culture). The flow-through was collected and re-loaded onto the column; this was repeated twice more. Digested SUMO, non-digested SUMO-fusion proteins, and SUMO protease 1 (all containing a H6 tag) bound to the resin. The flow-through,

only containing the Sso7d protein, was collected, buffer-exchanged to PBS at pH 7.3, concentrated to 2-5 mg/mL, and flash frozen in liquid nitrogen to be stored at -80 °C.

For Sso7d proteins conjugated to a small molecule, i.e. BODIPY-Maleimide or VL1-PEG-Maleimide (see Figure 3-16), a modified purification strategy was employed. To allow for maleimide conjugation of the small molecule to the Sso7d protein, the Sso7d proteins have a single cysteine introduced at either the N- or C-terminus. The cysteine is "capped" (oxidized) throughout the purification process through the addition of 1 mM 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB, or Ellman's Reagent) in every buffer used in all steps of the purification process (adjusted to pH 7.3). DTNB reacts with the Sso7d protein to form Sso7d-TNB and free TNB^-; the solution turns yellow upon this reaction because TNB^- is dark yellow in color. The purified Sso7d-TNB is buffer exchanged into PBS at pH 6.5 (adjusted with $NaHCO_3$; slightly acidic conditions prevent reactions between maleimide and lysines), its concentration is determined using a BCA Assay, and reacted with BODIPY-Mal or VL1-PEG-Mal (see section "Conjugation of BODIPY and VL1-PEG-Mal".) We have found that Sso7d-TNB stored at 4 ° up to six months remains stable.

Additionally, for Arg-rich proteins (R11-XK, R11-XK-RbK, R11-XK-R1218K, R11-PP4.8, R11-PP5.5, R11-PP6.2), the NaCl concentration of all purification buffers was increased to 0.5 M NaCl. The purified proteins were maintained in PBS adjusted to 0.5 M NaCl for all downstream assays as well.

Protein expression and purification of KRAS proteins.

KRAS WT and KRAS G12D (G domain, residues 1-166) were expressed and purified for biolayer interferometry experiments. Both proteins were expressed in BL21(DE3) E. coli cells as a fusion protein consisting of an N-terminal hexahistidine (H6) tag, followed by the first 166 amino acids of human KRAS isoform 4B using the pE vector. Note: we tested both H6-SUMO-fusions and H6-fusions and found that the yield of SEC-purified, GppNHP-loaded KRAS was lower for the H6-SUMO-fusions ($\sim$ 2 mg/L-culture accounting only for the molecular weight of KRAS) than

the H6-fusions ($\sim$ 6 mg/L-culture accounting only for the molecular weight of KRAS).

The purification protocol was very similar to that previously described [79]. The proteins were produced in Rosetta 2 (DE3) E. coli cells. Luria Broth (LB) containing kanamycin (50 μg/mL) and carbenicillin (100 μg/mL) was inoculated with transformed BL21 cells and grown overnight at 37 °C. On the following day, LB containing kanamycin (50 μg/mL) and carbenicillin (100 μg/mL) was inoculated (1:100 v/v ratio) with the overnight culture. Cell growth was monitored until the optical density (O.D.) reached 0.6-0.8. Protein expression was then induced by adding IPTG to a final concentration of 1.0 mM. The temperature was reduced to 30 °C, and protein expression was allowed to continue overnight. Rosetta cells were pelleted by centrifugation at 4,300 g for 15 min at 4 °C and the cell pellet flash frozen in liquid nitrogen and stored at -80 °C.

For protein purification, frozen cell pellets were thawed on ice then resuspended in 40 mL HisBuffer 1T (50 mM Tris/HCl, 300 mM NaCl, 10 mM imidazole, 5 mM MgCl2, EDTA-free Protease Inhibitor Cocktail (Roche 11873580001; amount according to manufacturer's instructions), pH 7.5) per 2 liters of culture volume, and lysed by sonication. The lysate was then clarified by centrifugation at 10,000 g for 25 minutes at 4 °C to remove unbroken cells and other extraneous cell debris, and filtered (0.45 μm pore membrane vacuum filtration) to yield clear lysate.

A column of TALON Superflow Metal Affinity Resin (Takara 635669) (2 mL of suspended resin per 100 mL of Rosetta culture) was equilibrated with 10 column volumes (CVs) of HisBuffer 1T before the supernatant was loaded. H6-KRAS bound to the resin. The resin was washed with 15 CVs of HisBuffer 1T and eluted with HisBuffer 2T (50 mM Tris/HCl, 300 mM NaCl, 150 mM imidazole, 5 mM $MgCl_2$, pH 7.5). Immediately following elution from the column, DL-Dithiothreitol (DTT; 1 mM final concentration) and cOmplete EDTA-free protease inhibitor cocktail solution (Roche) were added. The protein samples were purified on a Superdex 75 10/300 GL column (GE Healthcare), pre-equilibrated in RAS Phosphatase Buffer (32 mM Tris/HCl, 200 mM $(NH_4)_2SO_4$, 1 mM DTT, 0.5 mM NaN_3, 1 μM $ZnCl_2$, pH 8.0). The purified protein was concentrated to 150-300 μM using Amicon Ultra Centrifugal Fil-

ter Units (EMD Millipore). After adding 0.5 μL calf intestinal alkaline phosphatase (New England Biolabs) per 100 μL of protein solution and GppNHp (Sigma-Aldrich) to three times the protein concentration, the KRAS solution was incubated at 22 °C for 90 minutes. Subsequently, the protein samples were purified on a Superdex 75 10/300 GL column pre-equilibrated in RAS Storage Buffer (20 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, pH 7.5). After SEC purification, the proteins were biotinylated using EZ-Link Sulfo-NHS-LC-Biotin (Life Technologies). Protein samples were supplemented with glycerol (10% final concentration), shock frozen in liquid nitrogen, and stored at -80 °C.

Conjugation of BODIPY and VL1-PEG-Mal. Maleimide chemistry was used to conjugate a small molecule (either BODIPY-Maleimide or VL1-PEG-Maleimide) to the Sso7d protein. The Sso7d-TNB protein (in PBS at pH 6.5) is diluted to a concentration of 500 μM. The disulfide bond of Sso7d-TNB is reduced through the addition of a 10x molar ratio (5 mM) of Tris (2-carboxyethyl) phosphine (TCEP) at room temperature followed immediately by the addition of 5x molar ratio (2.5 mM) of BODIPY-Mal or VL1-PEG-Mal (stocks of both compounds are dissolved in DMSO and stored at -20°C). Additional DMSO was added (as necessary) prior to the addition of BODIPY-Mal or VL1-PEG-Mal such that the final DMSO v/v% is 10%. It was then flushed with nitrogen and allowed to react overnight at room temperature. TCEP, TNB^-, and excess BODIPY-Mal or VL1-PEG-Mal were removed by passing it through a PD-10 Desalting Column (Cytiva 17085101) twice. Sso7d-VL1 and Sso7d-BODIPY were concentrated to 2-5 mg/mL. For Sso7d-BODIPY, the labeling efficiency (dye:protein ratio) was determined by measuring absorbance at 280 nm (to determine Sso7d molar concentration) and 509 nm (to determine BODIPY molar concentration; ϵ = 92000 L·mol^{-1}cm^{-1}) with a NanoDrop 2000c spectrophotometer (Thermo Scientific). For Sso7d-VL1, ∼100% labeling efficiency was confirmed by the presence of only the labeled peak on a Q-TOF mass spectrum.

Bio-layer interferometry (BLI). Samples were analyzed with the ForteBio (Sar-

torius) Octet RED-96 Biolayer Interferometry system using stabilization buffer (same as above) supplemented with 0.1% BSA, 20 μL/L Tween-20, and 10 μM GppNHp. Biotinylated GppNHp-loaded K-Ras WT or G12D was immobilized onto streptavidin-coated BLI-tips (Sartorius 18-5020). Association was analyzed at various concentrations of proteins (1:3 dilutions starting from 200 nM to 0.82 nM), followed by measuring dissociation in buffer. Two reference samples were included for every run: (1) RAS-loaded tips without addition of binder and (2) unloaded tips with addition of 200 nM of binder. The average binding curve from the two reference wells was subtracted from the data. Dissociation constant (K_d) values were obtained from steady-state binding analysis.

Cellular Internalization of R11 variants. SW48 G12D cells were seeded in FBS-free culture medium (which was RPMI 1640 medium supplemented with 100 units mL^{-1} penicillin, and 100 μg mL^{-1} streptomycin) at a density of 50,000 cells $well^{-1}$ in a sterile 48-well dish at 24 h prior to treatment. On the day of treatment, incubation solutions were prepared by diluting proteins to obtain a final concentration of 7 μM. The final concentration of 7 μM is that of the labeled protein; for example, an R11.1.6-BDP solution with a total R11.1.6 (labeled and unlabeled) concentration of 100 μM (as measured by A280) and R11.1.6-BDP (i.e. labeled) concentration of 70 μM (as measured by A503) would have a labeling efficiency of 70%. To dilute it to a final concentration of 7 μM, the solution would be diluted ten-fold (70 μM$\rightarrow$M 7 μM of labeled protein). Right before treatment, the culture medium was removed, and cells were incubated with 200 μL of the incubation solutions for 2 h at 37 °C. After removal of the incubation solutions, the cells were rinsed twice with PBS and released from the plate with 200 μL of warmed 0.25% v/v trypsin-ethylenediaminetetraacetic acid (EDTA) mix. Trypsin was quenched by the addition of 400 μL of the medium. The cells were then subjected to centrifugation for 5 min at 200g followed by aspiration of the supernatant. The cells were washed twice by resuspension in 500 μL of PBS, centrifugation at 200g for 5 min, and aspiration of supernatant. The cells were resuspended in 200 μL of 200 μg/mL α-BDP antibody solution in PBS (BODIPY FL

Polyclonal Antibody, Invitrogen A-5770) at room temperature away from light. The antibody solution was then removed. The cells were washed twice by resuspension in 500 μL of PBS, centrifugation at 200g for 5 min, and aspiration of supernatant. Finally, the cells were resuspended in 500 μL of ice-cold PBS supplemented with bovine serum albumin (2% w/v) and propidium iodide (2 μg mL^{-1}) and kept on ice until the time of analysis. The fluorescence intensity of at least 10,000 events was measured with a BD FACS Celesta flow cytometer. The median fluorescence intensity of live, single cells is shown in Figure 3-15.

Confocal microscopy. Cells were seeded in culture medium at a density of 50,000 cells/well in a sterile eight-well dish (Nunc Lab-Tek II Chamber Slide System) at 24 hours prior to treatment. On the day of treatment, culture medium was aspirated and replaced with BODIPY-labeled protein in FBS-free culture medium (supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin) for 4 h at 37 °C. The protein solution was then replaced with medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 2 h at 37 °C. Cells were then washed twice with DPBS and nuclei stained with Hoechst 33342 (2 μg/mL in DPBS) for 5 min at 37 °C. The cells were kept on ice for 5 minutes, Hoechst 33342 solution removed, and cell membranes stained by incubation with wheat germ agglutinin (WGA)-Alexa Fluor 647 (5 μg/mL in DPBS) for 15 minutes on ice. Cells were then washed twice with ice-cold DPBS and kept in DPBS with HEPES (20 mM) on ice until the time of analysis. Live cells were examined using the Olympus FV1200 Scanning Confocal Microscope. Hoechst 33342: ex. 405 nm, em. 450 nm; WGA-Alexa Fluor 647: ex. 647 nm, em. 700 nm; BODIPY: ex. 503 nm, em. 512 nm. Image acquisition and processing (performed using ImageJ software) settings were maintained between all samples.

Growth Inhibition Assay. The viability of cells was assessed by quantifying ATP using a CellTiter-Glo 2.0 assay kit from Promega according to the manufacturer's instructions. Cells were seeded in culture medium at a density of 5000 cells well^{-1}

in a sterile 96-well plate 24 h prior to treatment. On the day of treatment, culture medium was aspirated and replaced with the PROTAC in FBS-free culture medium (supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin) for 4 h at 37 °C. The protein solution was then replaced with medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 68 h at 37 °C. The plates were equilibrated at room temperature for 30 min before the addition of CellTiter-Glo 2.0 reagent (100 μL $well^{-1}$). The contents were mixed for 2 min on an orbital shaker to induce cell lysis. The plate was incubated at room temperature for 10 min before the luminescence was measured with a Tecan Infinite 200 Pro plate reader. Cell viability was expressed as a percentage relative to the vehicle control.

Western blot analysis. Cells were seeded in culture medium at a density of 150,000 cells $well^{-1}$ in a sterile 12-well plate 24 h prior to treatment. On the day of treatment, culture medium was aspirated and replaced with the PROTAC in FBS-free culture medium (supplemented with 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin) for 4 h at 37 °C. The protein solution was then replaced with medium supplemented with 10%v/v FBS, 100 units mL^{-1} penicillin and 100 μg mL^{-1} streptomycin, for 20 h at 37 °C. Cells were lysed in ice-cold 1x RIPA lysis buffer (supplemented with 1x Protease Inhibitor (Thermo 78425)) and incubated on ice for 30 minutes, with periodical vortexing (10 min apart). Lysates were centrifuged at 20,000 g, 4 °C for 15 min, and supernatants were transferred to a new tube. Protein concentration was determined using the Pierce BCA protein assay kit (Thermo Scientific, 23227). 10-20 μg of protein extract was subjected to SDS-PAGE (Any kD Mini-PROTEAN TGX Gels, Bio-Rad), transferred onto PVDF membranes (iBlot 2 PVDF Mini Stacks IB24002, Invitrogen), using the iBlot gel transfer system (Invitrogen). Membranes were trimmed at ∼35 kDa. The upper half (actin portion) was washed 3 times (5 minutes each) with TBS (1x) and blocked with 5% milk in TBS for one hour at room temperature. The bottom half (RAS protein portion) was washed four times (five minutes each) with TBST (1x) and blocked for 1 hour at room temperature in 5% milk in TBST. All blots were probed with the appropriate primary antibodies overnight

at 4 °C in blocking buffer (5% milk TBST). The next day, blots were washed thrice (for actin portion) and four times (for RAS protein portion) with TBST, followed by probing with the appropriate secondary antibodies for 1 hour at room temperature in blocking buffer (5% TBST). The blots were washed with TBST (three times for actin blot and four times for RAS protein blot) at room temperature. Fluorescence (dried blots in the dark at ambient temperature) and chemiluminescence were imaged using Amersham Imager (General Electric). Western blot densitometry was measured using ImageJ.

Primary antibodies used were:

1. Ras Rabbit mAb (G12D mutant specific, D8H7, Cell signaling, 14429S; 1:5000)
2. pan-RAS (c-4) FITC (Santa Cruz Biotechnology, sc-166691 FITC, H1221; 1:5000)
3. Anti-h/m/r β Actin IgG (clone 937215, R&D Systems, MA B8929; 1:25,000)

Secondary antibodies used were:

1. Anti-Mouse IgG (H+L) DyLight 680 (Cell signaling, 5470S; 1:4000) for actin detection.
2. Anti-Mouse IgG HRP (Promega, W402B; 1:10,000)
3. Anti-Rabbit IgG HRP (Promega, W401B; 1:10,000)

Sequences.

H6-SUMO-R11.1.6

His Tag-SUMO-R11.1.6-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ATVKFTHQGEEKQVDI SKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKKR

agatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaag aaggagatataccatgggt catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagcc agaagtcaagcctgagactcacatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaagg ctgatggaagcgttcgctaaaagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccct gaagatttggacatggaggataacgatattattgaggctcaccgcgaacagattggaggt gcaaccgtgaaattcacacaccaaggcgaag aaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggtttaaatatgatgaagatggtggtgccaaaggtt

ggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga taa tctagaggatccgaattcgagctccgtcg
acaagcttgcggccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgcca
ccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaa
tgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgct
cctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctt
acggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggag
tccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgattt
cggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcgg
ggaaatgtgcgcggaaccccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcat
caaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcag
ttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggtt
atcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagcca
ttacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaag
gacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaata
cctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataa
attccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggct
tcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccattgttggaatttaat
cgcggcctagagcaagacgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatgacca
aaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatct
gctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcag
cagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgc
taatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtc
gggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcg
ccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaa
acgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaa
aacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataacc
gtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctg
atgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagcc
agtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctc
ccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgc
ggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtc
tggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaat
gataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactgg
cggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagc
agcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattc
atgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccc
cgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaa
cgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctcca
gcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcga
cgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagct
aacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggag
aggcggtttgcgtattgggcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagag
ttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggta
tcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaa
ccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatc

ggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatt
tgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagaga
catcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactga
cgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatc
ggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgc
cagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggtt
caccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactc
tcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattag
gaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggcca
cggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatatagg
cgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcg

H6-SUMO-R11-XK

Note: for the DNA sequence, all bases before the HisTag and after the stop codon are the same as those for H6-SUMO-R11.1.6.

HisTag-SUMO-R11-XK-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA
KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ATVRFTHQGEERQVDI
SRIRWVIRWGQYIWFRYDEDGGARGWGYVSERDAPRELLQMLRRR

catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactc
acatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaa
aagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagga
taacgatattattgaggctcaccgcgaacagattggaggt gctacggttcgtttcactcaccaaggcgaagaacgtcaagtagacatttcccg
tatccgttgggtcattcgttggggtcaatacatttggttccgctatgacgaggacggtggtgctcgtgggtggggatacgtctccgagcgtgac
gcgccccgtgagcttttacagatgttgcgtcgtcgt taa

H6-SUMO-R11-RbK

Note: for the DNA sequence, all bases before the HisTag and after the stop codon are the same as those for H6-SUMO-R11.1.6.

HisTag-SUMO-R11-RbK-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA
KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ATVRFTHQGEERQVDI
SRIRWVIRWGQYIWFKYDEDGGAKGWGYVSERDAPRELLQMLKRR

catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactc
acatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaa
aagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagga
taacgatattattgaggctcaccgcgaacagattggaggt gctacggttcgtttcactcaccaaggcgaagaacgtcaagtagacatttcccg
tatccgttgggtcattcgttggggtcaatacatttggttcaaatatgacgaggacggtggtgctaaagggtggggatacgtctccgagcgtgac
gcgccccgtgagcttttacagatgttgaaacgtcgt taa

H6-SUMO-R11-R1218K

Note: for the DNA sequence, all bases before the HisTag and after the stop codon are the same as those for H6-SUMO-R11.1.6.

HisTag-SUMO-R11-R1218K-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ATVRFTHQGEEKQVDI SKIRWVIRWGQYIWFRYDEDGGARGWGYVSERDAPRELLQMLRRR

catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactc acatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaa aagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagga taacgatattattgaggctcaccgcgaacagattggaggt gctacggttcgtttcactcaccaaggcgaagaaaaacaagtagacatttccaa aatccgttgggtcattcgttggggtcaatacatttggttccgctatgacgaggacggtggtgctcgtgggtggggatacgtctccgagcgtgac gcgccccgtgagcttttacagatgttgcgtcgtcgt taa

H6-SUMO-R11-PP4.8

Note: for the DNA sequence, all bases before the HisTag and after the stop codon are the same as those for H6-SUMO-R11.1.6.

HisTag-SUMO-R11-PP4.8-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ATVKFTHQGEEKQVDI RRIRWVIRWGQYIWFKYRRDGGAKGWGYVSEKDAPKELLQMLKKR

catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactc acatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaa aagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagga taacgatattattgaggctcaccgcgaacagattggaggt gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattcgtc gtatccgttgggtaatccgttggggccagtacatttggtttaaatatcgtcgtgatggtggtgccaaaggttggggttatgtgagcgaaaaaga tgcaccgaaagaactgctgcagatgctgaaaaagcga taa

H6-SUMO-R11-PP5.5

Note: for the DNA sequence, all bases before the HisTag and after the stop codon are the same as those for H6-SUMO-R11.1.6.

HisTag-SUMO-R11-PP5.5-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ATVKFTHQGEEKQVDI SKIKWVIRWGQYIWFKYDEDGGAKGWGYVSERRAPRRLLRMLKKR

catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactc acatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaa

aagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagga
taacgatattattgaggctcaccgcgaacagattggaggt gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagc
aaaatcaagtgggtaatccgttggggccagtacatttggtttaaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaacgt
cgtgcaccgcgtcgtctgctgcgtatgctgaaaaagcga taa

H6-SUMO-R11-PP6.2

Note: for the DNA sequence, all bases before the HisTag and after the stop codon are the same as those for H6-SUMO-R11.1.6.

HisTag-SUMO-R11-PP6.2-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA
KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG RRVRFTHQGEEKRVRI
SKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKKR

catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactc
acatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaa
aagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagga
taacgatattattgaggctcaccgcgaacagattggaggt cgtcgtgtgcgtttcacacaccaaggcgaagaaaaacgtgtgcgtattagca
aaatcaagtgggtaatccgttggggccagtacatttggtttaaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaag
atgcaccgaaagaactgctgcagatgctgaaaaagcga taa

H6-SUMO-YW1

Note: for the DNA sequence, all bases before the HisTag and after the stop codon are the same as those for H6-SUMO-R11.1.6.

HisTag-SUMO-YW1-stop

HHHHHH GSLQ DSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA
KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ATVKFTYQGEEKQVDI
SKIKWVIRWGQWIYFKYDEDGGAKGWGYVSEKDAPKELLQMLGKQ

catcaccatcatcatcac gggtccctgcag gactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactc
acatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaa
aagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagga
taacgatattattgaggctcaccgcgaacagattggaggt gcaaccgtgaaattcacataccaaggcgaagaaaaacaggtggatattagca
aaatcaagtgggtaatccgttggggccagtggatttactttaaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaag
atgcaccgaaagaactgctgcagatgctgggaaagcaa taa

Modifications for Cys-Mal conjugations

For all the variants of R11.1.6 described in this chapter, a Cys residue is introduced to the Sso7d sequence whenever BODIPY-Mal or VL1-PEG-Mal is to be conjugated to it. For N-terminal conjugations of BODIPY-Mal or VL1-PEG-Mal to the Sso7d

protein, the three amino acids ATC are added to the N-terminus of the protein. For example, R11.1.6 would be modified as follows:

ATC ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKKR

gcgacctgcgcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggtttaaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

For C-terminal conjugations of BODIPY-Mal or VL1-PEG-Mal to the Sso7d protein, the C-terminal Arg is substituted for Cys. For example, R11.1.6 would be modified as follows:

ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLKK C

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggtttaaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaag tgt

Chapter 4

Development of Rational Design Principles for RAS-targeting Biodegraders

4.1 Introduction

An exogenously-delivered large-molecule PROTAC would have to achieve: (1) efficient delivery across the cell membrane and into the cytosol and (2) efficient degradation of the POI (KRAS G12D) to achieve functional antagonization of RAS signaling. Given the challenges we faced with developing an exogenously-delivered KRAS-G12D-targeting PROTAC, we explored strategies that focus solely on the degradation aspect of this problem, e.g. the biodegrader approach employed by the Partridge Lab [9]. This workflow relies on the intracellular expression of a biodegrader (POI ligand attached, via a peptide linker, to the E3 ligase itself instead of a small-molecule E3-ligase ligand), thus eliminating the delivery challenge.

Briefly, simultaneous transient transfection of plasmids expressing the biodegrader constructs (with mCherry as a transfection marker) and POI-GFP were performed in HEK293T cells (hereinafter referred to as "HEK cells"), followed by flow cytometry analysis to determine the level of POI-GFP. The lower the GFP signal, the

greater the extent of degradation. Upon identification of biodegrader constructs that achieved a high level of POI-GFP degradation, HEK cells were transiently transfected with only the biodegrader plasmid (again with mCherry as the transfection marker), sorted for mCherry signal (mCherry-positive = transfected), and western blot analysis performed to confirm degradation of endogenous RAS. This would confirm that the loss in POI-GFP signal was indeed driven by ubiquitination of the POI and not by off-target ubiquitination of GFP.

By not having to express and produce each degrader construct (as we were doing before), we could more efficiently screen for degrader constructs that degrade KRAS G12D. Currently, the design of degraders remains a highly empirical process, primarily due to the large number of design features involved; the relative nascency of the degrader field also means that little is known about how each feature affects degradation efficiency. Therefore, we sought to identify and rank the most important features for the design of functional RAS-targeting biodegraders to help rationalize the design process. A DNA library of biodegrader constructs with varying (1) linker lengths, (2) linker rigidity, (2) E3 ligases, and (4) orientation (N- or C-terminal conjugation) was screened. Regression analysis was then performed to rank each feature based on its predictive impact. Western blot assays were performed to check against off-target ubiquitination of GFP. Upon identification of the combination of linker length, linker rigidity, E3 ligase, and orientation that achieves the highest level of endogenous RAS degradation, we next explored the effects of POI affinity and lysine richness on degradation efficiency.

4.2 Results and Discussion

4.2.1 Fluorescence-based Assay Development

The key advantage in the iterative development of biodegraders is that the entire construct — the POI ligand, the peptide linker, and the E3 ligase — is composed of amino acids. Unlike the large-molecule PROTACs described in Chapter 3 (which

comprise of a protein-based POI ligand conjugated to a small-molecule E3 ligand), cells can be genetically encoded to intracellularly express the full biodegrader construct. This rapidly accelerated the iterative design cycle since we no longer needed to express and purify the POI ligand.

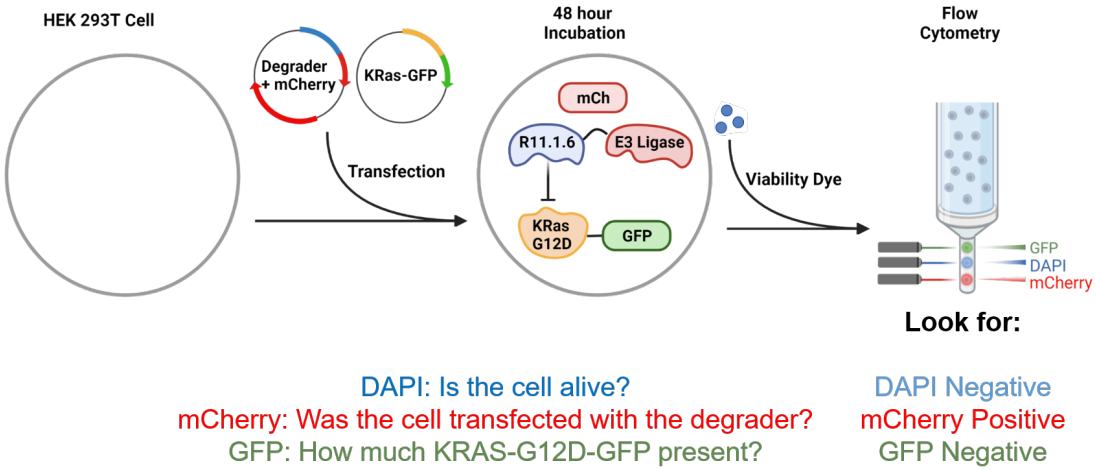


Figure 4-1: Experimental set up to screen biodegrader variants.

To perform high-throughput screening of biodegrader constructs, Lim *et al.* [9] transiently transfected KRAS-G12D-GFP-expressing HEK cells with DNA plasmids that encoded for a biodegrader, and assessed the extent of KRAS-G12D-GFP degradation (i.e. loss in GFP signal) via fluorescence-based flow cytometry. We decided to employ a fluorescence-based approach in screening our biodegrader constructs as well, but in the interest of time, we chose not to generate a KRAS-G12D-GFP-expressing stable HEK cell line. Instead, we co-transfected HEK cells with two plasmids: (1) one encoding for the biodegrader and an mCherry reporter (both proteins are not fused to each other); (2) the other encoding for the KRAS-G12D-GFP fusion protein. The experimental set up is shown in Figure 4-1. Two non-degrader negative controls were developed and used for all experiments (Figure 4-2): (1) R11$_{mut}$-E3, whereby 3 binding residues of R11.1.6 were mutated to Ala (see Table 4.1) and thus does not bind to KRAS G12D; and (2) R11.1.6 (without an E3 ligase component), i.e. it is merely an inhibitor.

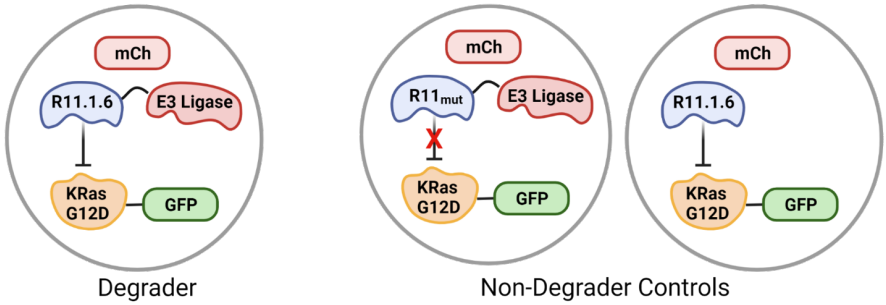


Figure 4-2: Non-degrader negative controls

Flow cytometry is used to determine the extent of degradation of GFP-tagged KRAS G12D. Transfected cells that express both the biodegrader and KRAS-G12D-GFP will be mCherry-positive and therefore reside in quadrants 1 and 2 (Figure 4-3). Cells that were unsuccessfully transfected will remain in Q4. An example of a flow plot for a successful degrader is shown on the left, and that for a non-degrader control is shown on the right. Successful degradation will result in a reduction in GFP signal, as seen in the shift of cells from Q2 to Q1 in the left plot. Cells with no degradation will be retained in Q2. The distribution of GFP fluorescence of mCherry-positive cells treated with the degrader and non-degrader controls is shown in Figure 4-4.

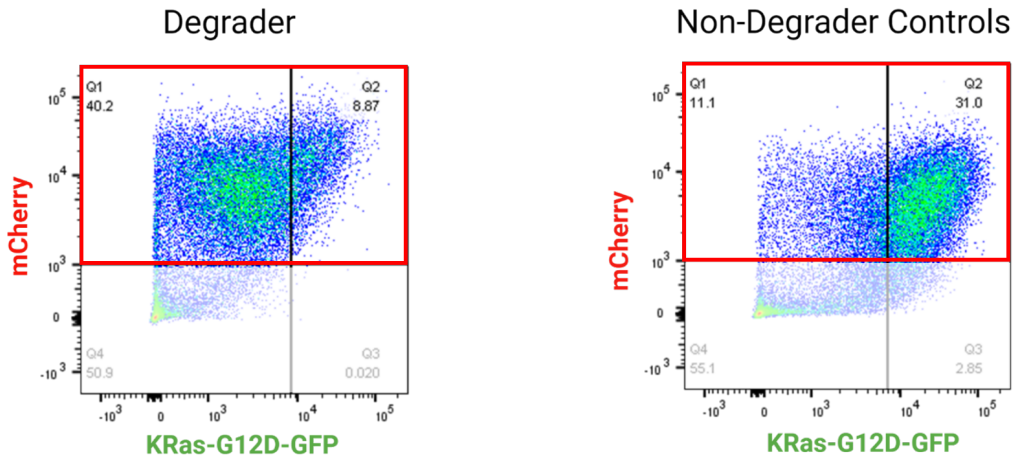


Figure 4-3: Density plots of cells transfected with degrader vs. non-degrader controls and analyzed via flow cytometry. mCherry-positive cells are boxed in red.

Prior to high-throughput screening of the DNA library, we performed several

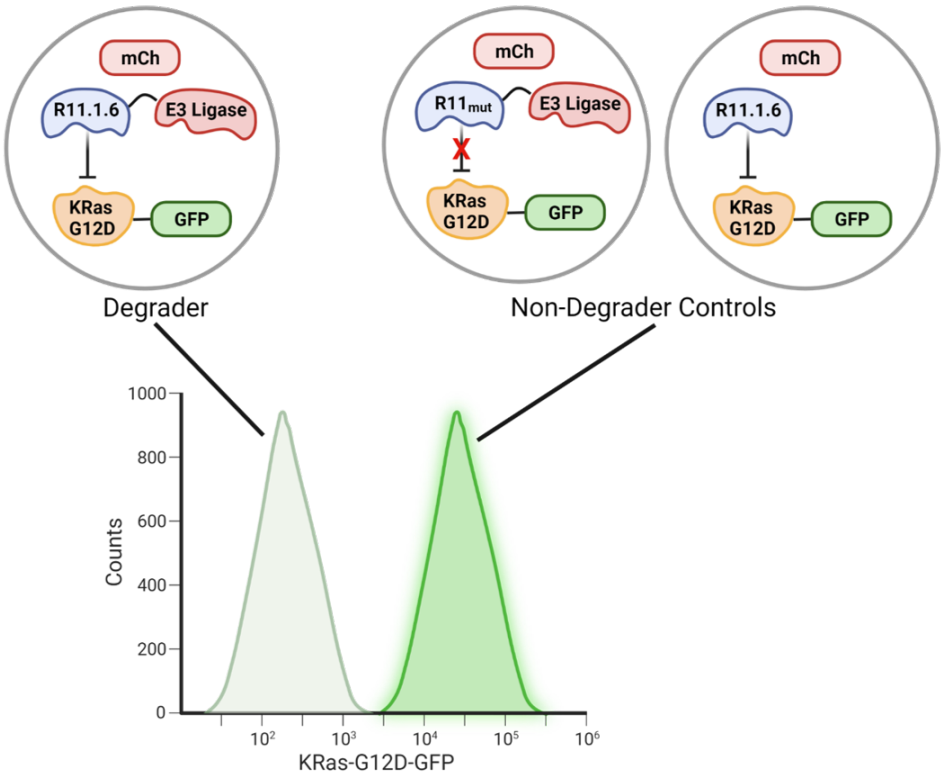


Figure 4-4: Representative GFP fluorescence intensities (as measured via flow cytometry) of mCherry-positive cells transfected with a successful degrader or non-degrader controls.

experiments to optimize transfection efficiencies and to determine an ideal ratio of (ng/μL:ng/μL) biodegrader:KRAS-G12D-GFP plasmids for transfection. The optimized transfection protocol is described in section 4.4. We also performed a titration experiment (Figure 4-5) with seven different biodegrader constructs and found that a 10:1 ratio of biodegrader:KRAS-G12D-GFP plasmids achieves moderate levels of degradation (i.e. drop in median GFP fluorescence intensity) in many of the constructs, allowing us to tease apart differences between the efficacy of biodegrader variants when performing high-throughput screening.

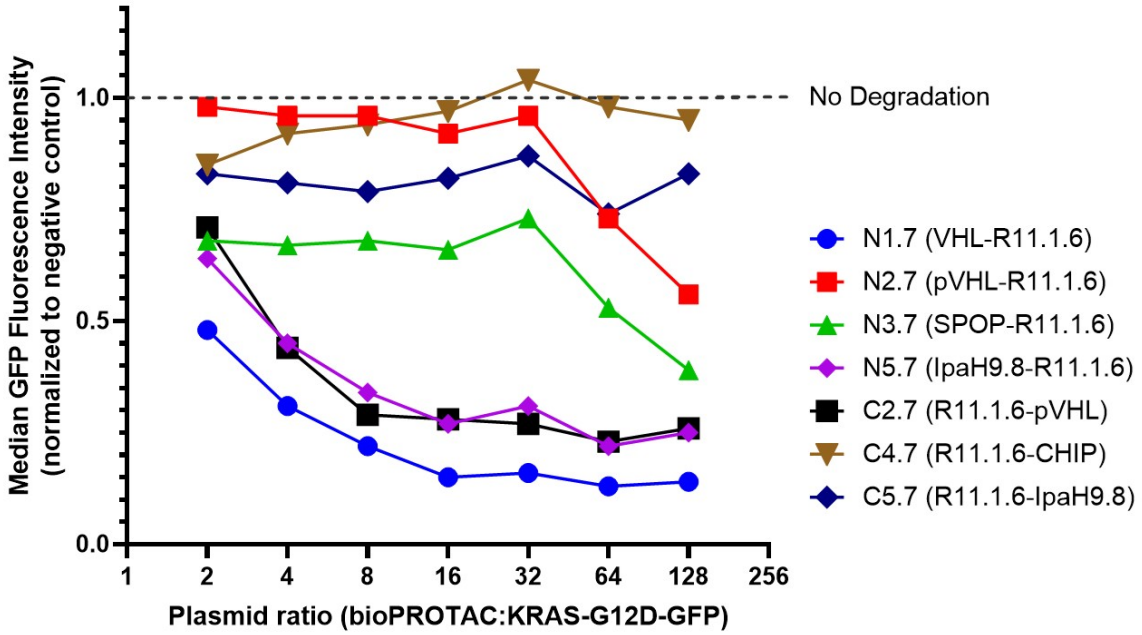


Figure 4-5: Level of KRAS-G12D-GFP degradation in HEK cells transfected with varying ratios of biodegrader to KRAS-G12D-GFP plasmids. For each ratio, the median GFP Fluorescence Intensities of every sample is normalized to that of the non-degrader control, i.e. R11.1.6 only (no E3 ligase component). Values are that of one technical replicate.

4.2.2 Feature Importance: Orientation, E3 Ligases, Linker Length, Linker Rigidity

A DNA library of 90 constructs was generated, with full factorial combinations of the following parameters (Figure 4-6):

1. E3 Ligases:
    - VHL
    - pVHL
    - SPOP
    - CHIP
    - IpaH9.8
2. Peptide linker length (number of amino acids):

- Short (2)
- Medium (7)
- Long (12)

3. Peptide linker rigidity (amino acid sequence for short/medium/long linkers):
   - Flexible (GS / $(GS)_3G$ / $(GS)_6$). Collectively referred to as "GS" linkers.
   - Rigid (AA / A(EAAAK)A / $A(EAAAK)_2A$). Collectively referred to as "AEK" linkers.
   - Rigid (AP / $(AP)_3A$ / $(AP)_6$). Collectively referred to as "AP" linkers.
4. Terminus, i.e. the terminus of R11.1.6 at which the E3 Ligase is conjugated:
   - N
   - C

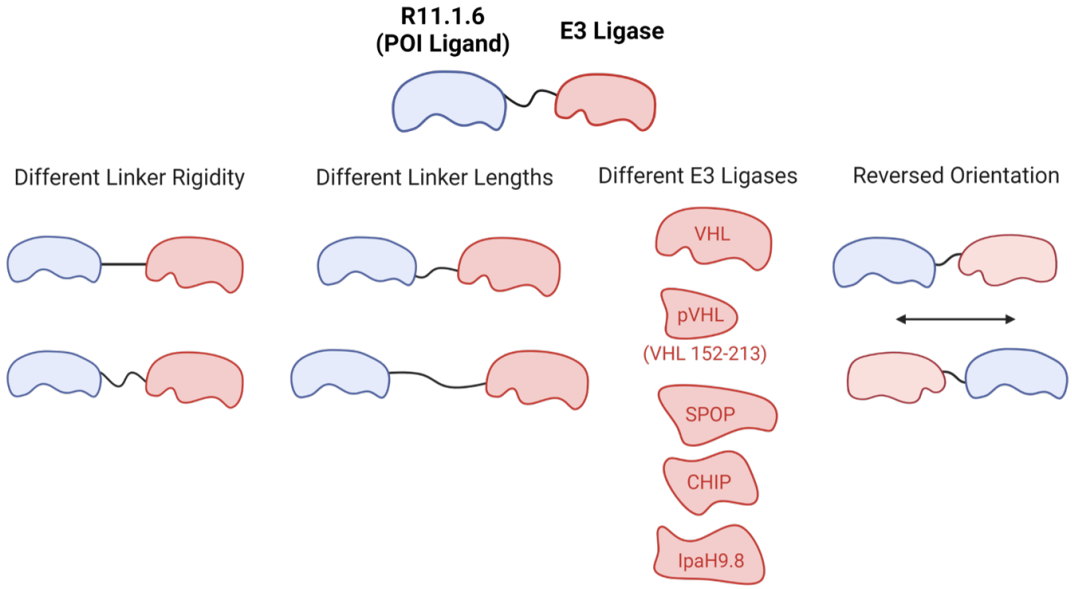


Figure 4-6: Parameters in the design of the 90-construct biodegrader DNA library

The E3 ligases pVHL, SPOP, and CHIP were chosen based on its successful degradation of KRAS-GFP, as reported by Lim *et al.* (see Figure 1c in [9]); note that the

"VHL" E3 ligase reported by Lim *et al.* is equivalent to the "pVHL" we are describing in our work. pVHL, or partial-VHL, is the truncated version of full-length VHL, whereby the natural substrate-binding domain of VHL is removed. In the present work, we decided to test both full-length VHL ("VHL") and truncated VHL ("pVHL"). Finally, the bacterial E3 mimic IpaH9.8 was chosen based on its demonstrated efficacy in a biodegrader context [103].

In addition to the flexible glycine-serine (GS) linker, we chose to explore $(AP)_n$- and $A(EAAAK)_nA$-based linkers because of their increased rigidity [104, 105]. Note: Lim *et al.* had tested 2 of the 90 constructs in our library of R11.1.6-based biodegraders: R11.1.6-$(GS)_3$G-pVHL and R11.1.6-$(GS)_3$G-SPOP. In the present work, we sought to explore the parameter landscape in more detail to identify the relative importance of each parameter.

Figure 4-7 summarizes the results of our flow-cytometry-based screening of the 90-construct DNA library. Four technical replicates were performed for each construct (except for C,VHL/GS-2, which had three technical replicates). All GFP MFI values were normalized to that of the $R11_{mut}$-E3 negative control. Theoretically, the normalized GFP MFI values can range form 0 to 1, implying full degradation or no degradation of KRAS-G12D-GFP, respectively. Highlighted in green are constructs with the highest normalized MFI values, i.e. the worst KRAS-G12D-GFP degraders; in white are constructs with the lowest normalized MFI values, i.e. the best KRAS-G12D-GFP degraders.

With the lowest normalized GFP MFI of 0.13, C, pVHL/GS-2 (R11.1.6-GS-pVHL) is the best KRAS-G12D-GFP degrader (Figure 4-7). Several trends can also be observed from the data. Figure 4-8 groups the data based on features for easier visualization. It seems that the choice of which terminus to place an E3 ligase is as important as the choice of E3 ligase itself (Figure 4-8C). For example, placing VHL or IpaH9.8 at the N terminus of R11.1.6 allows for substantially higher levels of KRAS-G12D-GFP degradation (0.48 & 0.65, respectively) than placing it at the C terminus (0.86 & 0.99, respectively); an N-terminal placement of SPOP and CHIP also confers higher levels of degradation. On the other hand, C-terminal pVHL leads

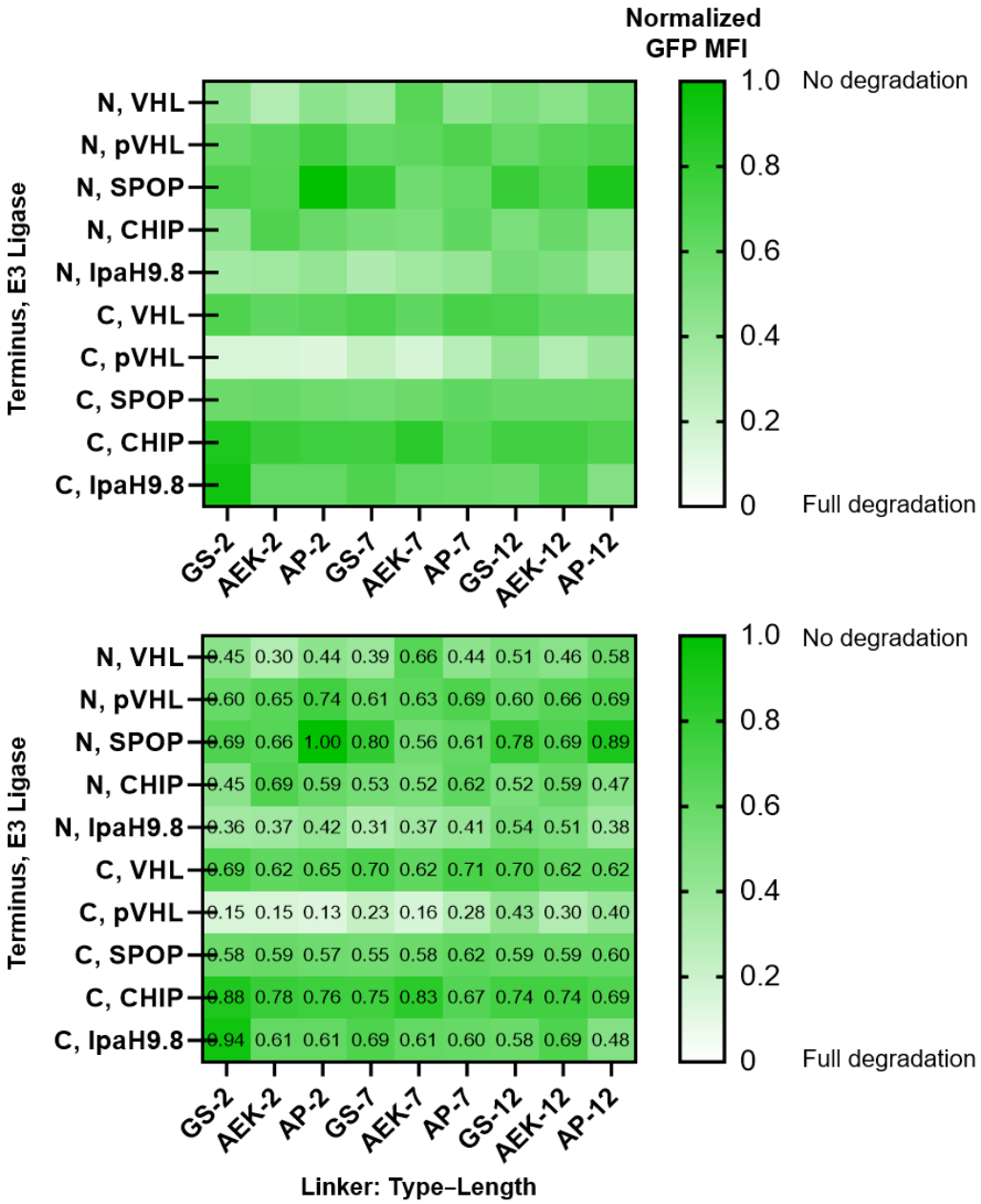


Figure 4-7: Heatmap of the median GFP fluorescence intensity of HEK cells treated with each of the 90 biodegrader constructs. GFP MFI values were normalized to that of the negative control, R11$_{mut}$-E3. A normalized GFP MFI of 1.0 indicates no degradation of KRAS-G12D-GFP; a value of 0.0 indicates full degradation. Both heatmaps are identical; individual values for each construct are shown in the bottom heatmap. The individual values shown are the mean of four technical replicates (except for C,VHL/GS-2, which had three technical replicates); the standard deviation of each value had a mean of ±0.04 (full data shown in section 4.4).

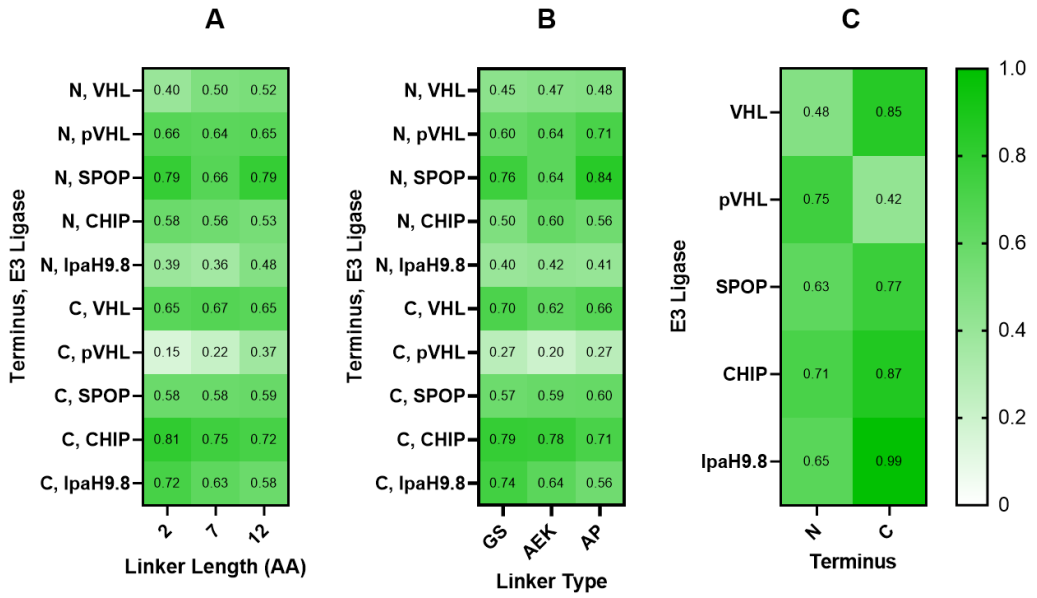


Figure 4-8: Data from Figure 4-7 categorized based on features.

to substantially greater degradation (0.42) than N-terminal pVHL (0.75). These results suggest that the extent of degradation is not driven by the choice of E3 ligase alone; since degradation is driven by proximity-induced ubiquitination of the target, the orientation of the E3 ligase at the N- or C-termini plays a significant role as well.

For each combination of terminus and E3 ligase, we also grouped the data to assess the dependence on linker length (Figure 4-8A) and linker type/rigidity (Figure 4-8B). The effect of linker length is not generalizable across all constructs (4-8A); instead, it depends specifically on the combination of terminus and E3 ligase chosen. N/VHL and C/pVHL constructs perform better with short linkers; N/CHIP, C/CHIP, and C/IpaH9.8 perform better with long linkers; N/SPOP and N/IpaH9.8 perform better with medium-length linkers; linker length does not seem to matter for the rest. Similarly, the effect of linker type is dependent on the specific terminus and E3 ligase chosen; for most of them, linker type does not seem to play much of a role.

Random Forest Regression

To more quantitatively determine which features are most important in predicting KRAS-G12D-GFP degradation, we built a Random Forest (RF) regression model

that predicts GFP MFI from the four categorical features: (1) terminus, (2) E3 ligase, (3) linker length, and (4) linker type. To perform multi-categorical regression, we first converted all features into binary features via dummy encoding. The basic idea is to represent each category as a binary feature that takes a value of 0 or 1, depending on whether the category is present or absent in the data. In our case, the four features were split into 13 features:

- Terminus:
  - `terminal_n`
  - `terminal_c`
- E3 Ligase:
  - `e3_ligase_vhl`
  - `e3_ligase_pvhl`
  - `e3_ligase_spop`
  - `e3_ligase_chip`
  - `e3_ligase_ipa`
- Linker Length:
  - `linker_length_short`
  - `linker_length_medium`
  - `linker_length_long`
- Linker Type:
  - `linker_type_gs`
  - `linker_type_aek`
  - `linker_type_ap`

For example, for the C/pVHL/GS-2 construct, `terminal_c`, `e3_ligase_pvhl`, `linker_length_short`, and `linker_type_gs` would have values of 1; every other feature would have a value of 0.

An RF model is a machine learning algorithm used for classification and regression tasks. It is an ensemble method that combines multiple decision trees and improves their performance by reducing overfitting and increasing accuracy.

Decision trees are a type of supervised learning algorithm that is used to predict the target variable based on a set of input features. It works by recursively splitting the data into smaller subsets based on the features that best separate the target variable. When a decision tree is being built, the algorithm first selects the best feature to split the data based on a metric called "impurity". For a parent node that is being split into child nodes, the goal is to choose the feature that minimizes the impurity (i.e. maximizes the homogeneity) within each of the child nodes. In other words, the "information gain" is maximized at a split. The earlier (or higher up) a feature appears in a decision tree, the more important it is in predicting the target variable.

The RF model works by creating a set of decision trees on randomly sampled subsets of the training data (in our case, the combination of the four features). Each decision tree in the forest independently predicts the target variable (in our case, normalized GFP MFI) from a randomly sampled subset of training data. The final output is the aggregated prediction of all the trees. In the case of regression, the aggregation is performed by taking the average of the predicted values of every tree. The randomness in the RF model comes from two sources:

1. Random sampling of the training data: Each tree is trained on a randomly selected subset of the training data, which ensures that each tree is trained on a different set of samples.

2. Random selection of features: At each node in the decision tree, a random subset of features is considered for splitting, which ensures that each tree is trained on a different set of features.

This randomness helps to reduce the correlation between the trees and improves the performance of the model. In our case, we built and trained an RF regression model to predict normalized GFP MFI. The accuracy of the model was quantified by R^2 and root mean squared error (RMSE). Before training our model, we first split the data (359 data points: 4 technical replicates for each of the 90 constructs except C,VHL/GS-2 which had three technical replicates) into a training set (80% of the data) and a test set (20% of the data). We then trained the model on the training set and evaluated its performance on the test set. We also performed tuning of the following three hyperparameters using grid search, to find the optimal values that maximize the R^2 value of the test set:

1. `n_estimators`: this refers to the number of decision trees in the the RF model. Increasing the number of trees generally improves the accuracy and reduces overfitting, but it also increases the computational cost and memory requirements. A good rule of thumb is to set `n_estimator`s to a value that is large enough to achieve good performance but not too large to cause excessive resource usage or training time.

2. `max_depth`: It refers to the maximum depth of each decision tree in the forest. The depth of a tree is the number of levels from the root node to the terminal nodes. Increasing the depth of the tree can increase the model's complexity and accuracy, but it can also cause overfitting and poor generalization of new data. It is good practice to set `max_depth` to a value that allows the tree to capture the relevant patterns in the data without memorizing the noise or outliers.

3. `min_samples_leaf`: It refers to the minimum number of samples required to form a leaf node in each decision tree. The leaf nodes are the terminal nodes that represent the final prediction of the model. Increasing the minimum number of samples per leaf can improve the stability and robustness of the tree and reduce the variance, but it can also decrease the model's sensitivity and flexibility to the data. A reasonable value for `min_samples_leaf` depends on the size and complexity of the dataset and the desired trade-off between bias and variance.

Ideal values for `n_estimators`, `max_depth`, and `min_samples_leaf` were determined to be 100, 9, and 0.01. Note: a 'max_depth' value greater than 9 is redundant. Remember that there are four feature categories — (1) linker length, (2) linker type, (3) E3 ligase, and (4) orientation — and each unique biodegrader construct must have exactly these four categories specified. This means that each construct consists of exactly 4 of the 13 possible features: 1 of 3 possible linker lengths, 1 of 3 possible linker types, 1 of 5 possible E3 ligases, and 1 of 2 possible orientations. And the mutual exclusivity of features *within* each feature category (e.g. for linker type, it can only be GS or AEK or AP) also means that each unique biodegrader construct can be determined using a maximum of (13 - 4 =) 9 yes/no questions. For example, (1) GS? No → (2) AEK? No → must be **AP**. (3) Length=2? No → (4) Length=7? No → Length must be **12**. (5) VHL? No → (6) pVHL? No → (7) SPOP? No → (8) CHIP? No. → Must be **IpaH9.8**. (9) N-terminal? No → Must be **C-terminal**.

Finally, we evaluated the importance of each of the 13 features in predicting normalized GFP MFI. In an RF model, the feature importance values are generated by analyzing the contribution of each feature in reducing the impurity or increasing the information gain in the decision tree. The higher the contribution of a feature in the decision tree, the higher its importance in the model. From the trained model, we used the `feature_importances_` attribute to obtain the feature importance values (Figure 4-9).

Analyzing the importance of the features allows us to gain invaluable insights into which features are important in the design of a KRAS-G12D-GFP-targeting biodegrader. As seen in Figure 4-9, the choice of which terminus to place the E3 ligase is the most important predictor, followed by the choice of E3 ligase. Linker design appears to be less relevant, with linker type mattering less than linker length.

Endogenous RAS degradation

In screening for biodegraders against KRAS-G12D-GFP, it was important to ensure that the observed degradation of KRAS-G12D-GFP was not driven by off-target ubiquitination of GFP. In a previous study, a KRAS-G12C-targeting PROTAC was suc-

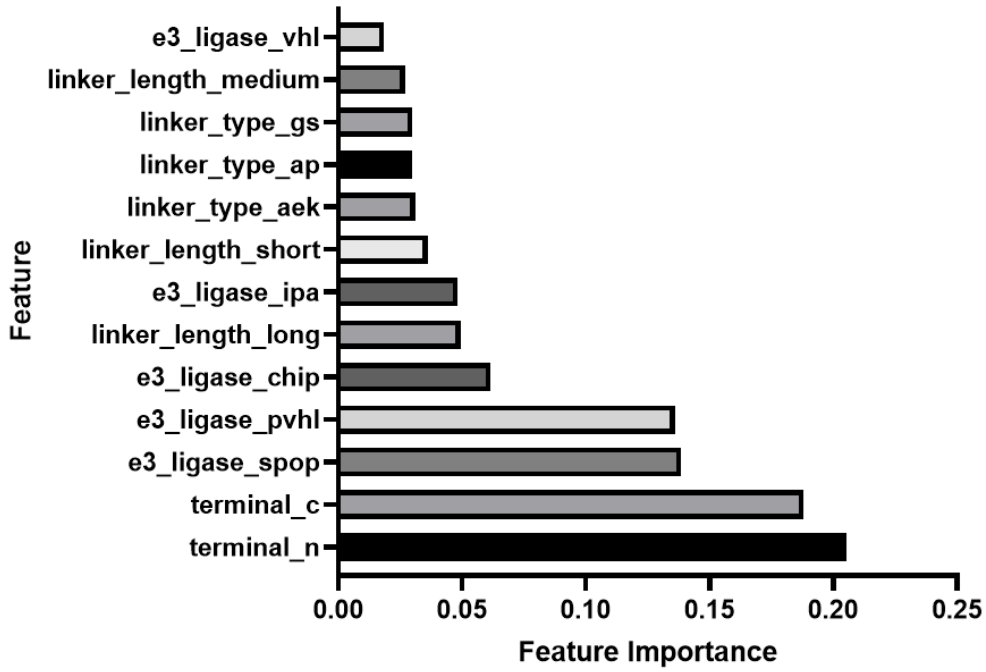


Figure 4-9: Feature importance derived from Random Forest Regression Modeling of the screening data

cessful in degrading KRAS-G12D-GFP but not the endogenous form (KRAS G12C) [106].

To this end, we transfected HEK cells with only the biodegrader-mCherry plasmid, sorted for transfected (mCherry+) cells, and performed western blot analysis to detect endogenous RAS levels (Figure 4-10). For this study, eight out of the 90 biodegrader constructs were chosen based on their enhanced ability to degrade KRAS-G12D-GFP (Figure 4-11). As expected, the endogenous RAS levels of the negative controls (R11, R11$_{mut}$-pVHL, VHL-R11$_{mut}$, IpaH9.8-R11$_{mut}$) were similar. A significant degree of endogenous RAS degradation was achieved with the R11.1.6-pVHL constructs, consistent with their low normalized GFP MFI values of 0.13 to 0.16. On the other hand, despite a substantial level of KRAS-G12D-GFP degradation (as measured by normalized GFP MFI values of 0.31 to 0.44), the VHL-R11.1.6 and IpaH9.8-R11.1.6 constructs did not degrade endogenous RAS. This suggests that these two constructs were likely inducing off-target ubiquitination of GFP, and highlights the importance

of testing for endogenous RAS degradation.

With these results, we identified R11.1.6-AP-pVHL as the best construct to move forward with further characterization and testing. R11.1.6-AP-pVHL achieved the greatest level of KRAS-G12D-GFP degradation (normalized GFP MFI of 0.13; Figure 4-7) and endogenous RAS degradation in HEK cells (Figure 4-11).

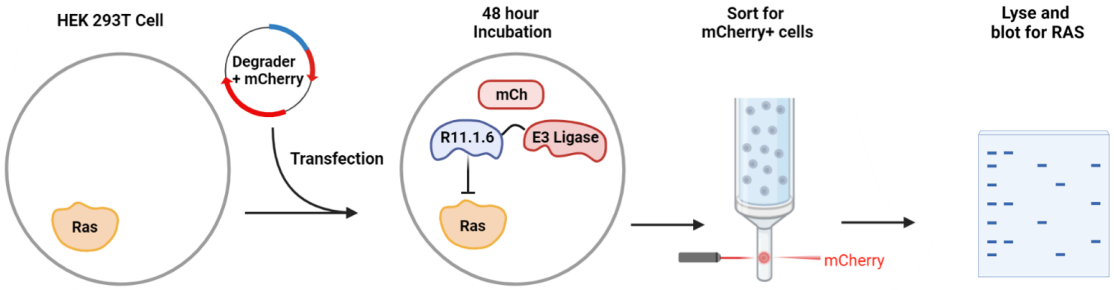


Figure 4-10: Experimental setup to determine the extent of degradation of endogenous RAS in HEK cells transfected with biodegrader.

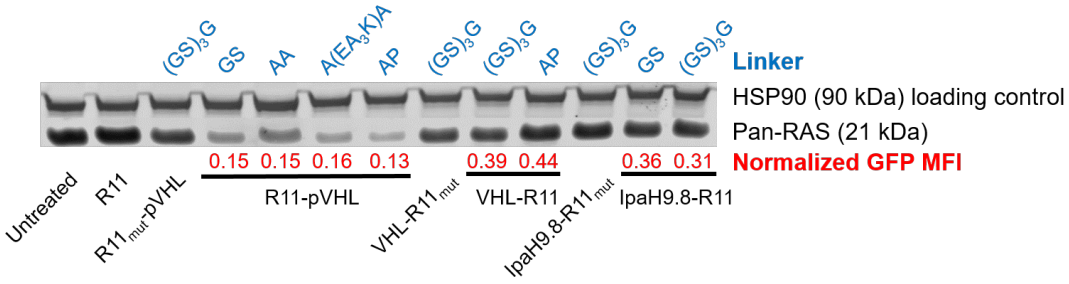


Figure 4-11: Degradation of endogenous RAS in HEK cells, as detected via western blot. Normalized GFP MFI values (in red) are derived from the flow-cytometry-based KRAS-G12D-GFP screening shown in Figure 4-7.

4.2.3 Exploring the Effects of POI Affinity

Previous studies on small-molecule PROTACs have established that a high affinity of the POI ligand to the POI does not necessarily translate to an effective PROTAC [107, 108]. Scheepstra *et al.* suggests that a PROTAC with weak binding will not offer the E3 ligase enough time to ubiquitinate the POI, whereas very tight binding would slow down the traveling of the PROTAC between different POI copies [109]. Bondeson

et al. also showed that some very potent PROTAC kinase binders did not effectively degrade its target POI, whereas other binders with even double-digit micromolar affinity managed to elicit effective degradation of the POI [107]. We hypothesized that the single-digit nanomolar affinity of R11.1.6 for KRAS G12D might not necessarily be ideal in conferring maximal degradation and sought to explore the degradation efficiencies of R11.1.6 variants with weaker affinities.

To create variants of R11.1.6 with weaker affinities, we performed an alanine scan of R11.1.6 by substituting each binding residue with alanine. We expressed, purified, and characterized the KRAS G12D affinities of these variants via biolayer interferometry. We then repeated this process by mutating up to two or three binding residues to alanine until we generated a family of variants with binding residues spanning 2 orders of magnitude (Table 4.1).

These weak-binding variants of R11.1.6-AP-pVHL were screened for their ability to degrade KRAS-G12D-GFP via the flow-based setup described earlier (Figure 4-1). The results are shown in Figure 4-12. The Spearman rank correlation (a measure of monotonicity of the relation between two variables) between POI affinity and normalized GFP MFI is 0.84, suggesting a high positive correlation between both variables. These results suggest that at least for R11.1.6-based biodegraders with affinities ranging from single-digit nanomolar to single-digit micromolar, higher affinity positively correlates with increased degradation efficiency of KRAS G12D. It remains untested if we would see a loss in degradation efficiency for constructs with picomolar affinity for KRAS G12D.

4.2.4 Exploring the Effects of Lysine Richness

Given that ubiquitination of a protein occurs at lysine residues, and that R11.1.6 is rich in lysines (10 of 61 residues are lysines), we reasoned that there might be some level of self-ubiquitination (and therefore self-degradation) of the biodegrader itself. By tuning the number of lysines on R11.1.6, we sought to explore how the lysine richness of the POI ligand affects degradation efficiency.

To this end, we developed a variety of lysine-poor and lysine-free variants of

Table 4.1: Weak-binding variants of R11.1.6

Variant	**Mutation(s)**	Affinity for KRAS G12D (nM)	Number of Lysines in Sso7d sequence
R11.1.6	-	6	10
R11-NB1	W21A	291	10
R11-NB2	I23A	114	10
R11-NB4	Y28A	74	10
R11-NB7	Y44A	81	10
$R11_{mut}$	W25A W30A W42A	>10,000	10
R11-NB11	Y28F	18	10
R11-NB12	Y44F	70	10
R11-NB15	W21A Y28F	209	10
R11-NB16	W21A Y44F	175	10
R11-NB17	I23A Y28F	214	10
R11-NB18	I23A Y44F	472	10
R11-NB19	W42A Y28F	2260	10

R11.1.6-AP-pVHL (Table 4.2) by substituting some or all of the Lys residues of R11.1.6 with either Arg or Ala. R11-XK and R11-RbK are as described in section 3.2.2, with their sequences shown in Table 3.1. The lysine-free R11-XK had all ten Lys residues of R11.1.6 substituted for Arg, whereas R11-RbK had all Lys except K32, K40, and K59 (three of the ten binder residues) substituted for Arg. We tested these constructs in an Sso7d-AP-pVHL format using the fluorescence-based method

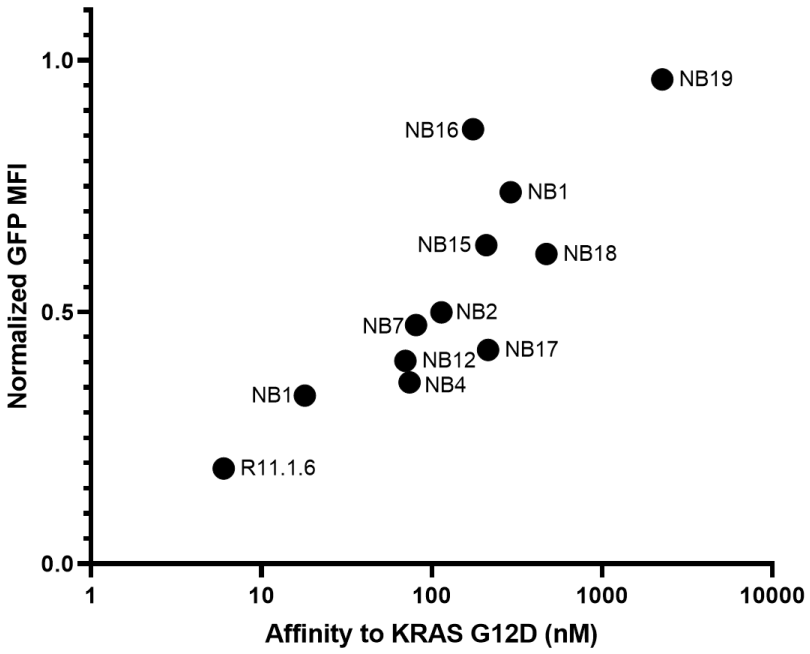


Figure 4-12: Normalized GFP MFI of weak-binding variants of R11.1.6-AP-pVHL. GFP MFI values are normalized to that of $R11_{mut}$-AP-pVHL.

described earlier (Figure 4-1). Surprisingly, we observed that R11-RbK-AP-pVHL and R11-XK-AP-pVHL essentially failed to degrade KRAS-G12D-GFP, with normalized GFP MFI values of 0.88 and 0.98, respectively. R11-RbK-AP-pVHL performed slightly better than R11-XK-AP-pVHL possibly due to its slightly higher affinity for KRAS G12D. We hypothesized that these Arg-rich variants perform poorly because they do not co-localize with cytosolic KRAS; we saw in section 3.2.2 that BODIPY-labeled R11-XK and R11-RbK localize in the nucleus instead.

This prompted us to develop and screen variants that had Lys residues substituted to Ala instead. We reasoned that the increased richness of uncharged Ala residues would not drive nuclear localization. R11-nbKA and R11-XKA are analogous to R11-RbK and R11-XK respectively, with Lys residues being substituted for Ala instead of Arg. Disappointingly, we found that the Ala-rich variants only had slightly improved degradation of KRAS-G12D-GFP (0.73 and 0.83), but still achieved much less degradation of KRAS-G12D-GFP relative to the original R11.1.6-AP-pVHL (0.16). It has yet to be determined whether these Ala-rich variants (1) maintain binding to

KRAS G12D and (2) localize in the cytosol. Given the abundance of Lys residues in R11.1.6, we reason that substitutions of Lys residues significantly affect the localization, binding, and stability of the protein. This makes it difficult to isolate the effect of lysine richness alone on degradation efficiency, warranting further studies to tackle this interesting question.

Table 4.2: Lysine-poor and lysine-free variants of R11.1.6

Sso7d variant	**Mutations**	**Affinity for KRAS G12D (nM)**	**Number of Lysines***	**Normalized GFP MFI****
R11.1.6	N/A	6	10	0.16
R11-RbK	All non-binding K→R	3	3	0.88
R11-XK	All K→R	21	0	0.98
R11-nbKA	All non-binding K→A	n.d.	3	0.73
R11-XKA	All K→A	n.d.	0	0.83

* Number of lysines in the Sso7d sequence only (not pVHL).

** Median GFP fluorescence intensity (GFP MFI) of the Sso7d-AP-pVHL construct tested using the fluorescence-based method (Figure 4-1), normalized to the GFP MFI of $R11_{mut}$-AP-pVHL.

4.3 Conclusions

In this chapter, we expanded upon previous research on KRAS-G12D-targeting biodegraders [9]. We utilized high-throughput fluorescence-based screening and Random Forest regression modeling to develop several rational design principles for the creation of R11.1.6-based KRAS-G12D-targeting biodegraders. Our results indicated that the order of importance for the design features are: orientation of the components, choice of E3 ligase, linker length, and linker type. Moreover, we found that the efficiency of degradation is compromised when the affinity of R11.1.6 for KRAS G12D is weakened. The effect of lysine richness on degradation efficiency was difficult to isolate and remains unclear; further work is required to understand its effect. The findings discussed in this chapter will be valuable in informing the future development of biodegraders against RAS and beyond.

4.4 Materials and Methods

Plasmid preparation. For site-directed mutagenesis within a 30 bp range, primers were designed using NEBaseChanger (https://nebasechanger.neb.com/) and synthesized by Integrated DNA technologies (IDT). PCR amplification, ligation, and transformation were performed in accordance with the New England Biolabs Q5 Site-Directed Mutagenesis Kit (E0554) Quick Protocol. For the lysine-free/-poor plasmids (R11-XK, R11-XK-RbK), gBlocks of the Sso7d sequence were synthesized by IDT and cloned into the backbone using the CloneAmp HiFi PCR Premix and the In-Fusion® HD Cloning Kit as per the manufactures' protocols.

GFP-based flow cytometry screen. HEK293T cells were seeded in antibiotic-free culture medium (supplemented with 10% V/V FBS) at a density of 16,000 cells well^{-1} in a sterile 96-well plate (surface area per well is 0.32 cm^2, so seeding density is 50,000 cells per cm^2) 24 h prior to transfection. On the day of transfection, degrader/mCherry and KRAS-G12D-GFP DNA plasmids were diluted in

OptiMEM to a concentration of 9.75 ng/μL. For experiments described in section 4.2.1, 39.1 μL of degrader/mCherry and 3.9 μL of KRAS-G12D-GFP DNA plasmids were mixed (i.e. a 10:1 v/v ratio of degrader/mCherry:KRAS-G12D-GFP plasmids) before the addition of 1.40 μL of transfection reagent, polyethylenimine (PEI). For experiments described in section 4.2.3 and 4.2.4, 35.8 μL of degrader/mCherry and 7.2 μL of KRAS-G12D-GFP DNA plasmids were mixed (i.e. a 5:1 v/v ratio of degrader/mCherry:KRAS-G12D-GFP plasmids) before the addition of 1.40 μL of transfection reagent, polyethylenimine (PEI). DNA/OptiMEM/PEI mixture was allowed to incubate at room temperature for 20 minutes before 8.4 μL of it is added per well of cells. The final DNA(μg):surface area(cm^2) ratio = 0.250; the final OptiMEM(μL):surface area(cm^2) ratio = 26.3; the final PEI(μL):DNA(μg) ratio = 3.5. 48 hours transfection, culture medium was removed, cells were rinsed twice with PBS and released with warmed 0.25% v/v trypsin-ethylenediaminetetraacetic acid (EDTA) mix. Trypsin was quenched by the addition of culture medium (supplemented with 10% FBS, 100 units mL^{-1} penicillin, and 100 μg mL^{-1} streptomycin). Cells were centrifuged for 5 min at 500g followed by aspiration of the supernatant. Cells were incubated with DAPI (0.1 μg/mL in PBS) for 10 min at 37 °, centrifuged for 5 min at 500g, followed by aspiration of the supernatant. The cells were washed twice by resuspension in ice-cold PBS, centrifugation for 5 min at 200g, and aspiration of supernatant. The cells were resuspended in ice-cold PBS supplemented with bovine serum albumin (2% w/v). The cell suspension was passed through a 35 μm nylon mesh to dissociate aggregates and kept on ice before analysis on a BD LSR II Flow Cytometer. The fluorescence intensities of DAPI, GFP, and mCherry of at least 10,000 events were measured. Data were analyzed in FlowJo; the median GFP fluorescence intensity (GFP MFI) of DAPI negative (live cells) and mCherry positive (transfected cells) cells were recorded. For experiments described in section 4.2.1, GFP MFIs of constructs with N-terminal VHL, pVHL, SPOP, CHIP, and IpaH9.8 were normalized to the GFP MFI of VHL-$(GS)_3$G-R11$_{mut}$, pVHL-$(GS)_3$G-R11$_{mut}$, SPOP-$(GS)_3$G-R11$_{mut}$, CHIP-$(GS)_3$G-R11$_{mut}$, and IpaH9.8-$(GS)_3$G-R11$_{mut}$, respectively; and constructs with C-terminal VHL, pVHL, SPOP, CHIP, and IpaH9.8 were

normalized to the GFP MFI of R11$_{mut}$-(GS)$_3$G-VHL, R11$_{mut}$-(GS)$_3$G-pVHL, R11$_{mut}$-(GS)$_3$G-SPOP, R11$_{mut}$-(GS)$_3$G-CHIP, and R11$_{mut}$-(GS)$_3$G-IpaH9.8, respectively. For experiments described in section 4.2.3 and 4.2.4, GFP MFIs of all constructs were normalized to the GFP MFI of R11$_{mut}$-AP-pVHL.

Random Forest Regression Analysis. Random Forest (RF) Regression Analysis was used to model the relationship between the four biodegrader features (orientation, linker length, linker type, and E3 ligase) and the normalized GFP MFI. The four categorical features were converted to 13 binary-valued dummy variables as described in section 4.2.1; for each biodegrader construct, 4 of the 13 features would have a value of 1 while the rest would have a value of 0. The data was split randomly into training (80%)and test (20%) sets. Hyperparameter tuning of 'n_estimators', 'max_depth', and 'min_samples_leaf' was carried out while training on the test set and evaluating against the R^2 value of the actual vs. predicted target variable of the test set. The optimized hyperparameter values for 'n_estimators', 'max_depth', and 'min_samples_leaf' were determined to be 100, 9, and 0.01.

Cell culture HEK293T cells were cultured in DMEM. SW48 G12D (kind gift from the White Lab, Massachusetts Institute of Technology [79]) and AsPC-1 (kind gift from the Irvine Lab, Massachusetts Institute of Technology) were cultured in RPMI 1640. All media were supplemented with 10% Fetal Bovine Serum (Life Technologies), 100 U/mL penicillin, and 100 μg/mL streptomycin (Gibco), unless stated otherwise. The cells were grown in sterile culture flasks in a cell culture incubator at 37 °C under CO_2(g) (5% v/v). Cell lines were passaged a minimum of five times and up to twelve times before use. The cells were counted to determine seeding density using a hemacytometer.

Endogenous RAS degradation. HEK293T cells were seeded in antibiotic-free culture medium (supplemented with 10% V/V FBS) at a density of 750,000 cells well^{-1} in a sterile 60 mm dish 24 h prior to treatment. On the day of treatment,

DNA plasmids expressing the degrader construct were transfected with PEI according to the manufacturer's protocol. 48 hours after transfection, culture medium was removed, cells were rinsed twice with PBS and released with warmed 0.25% v/v trypsin-ethylenediaminetetraacetic acid (EDTA) mix. Trypsin was quenched by the addition of culture medium (supplemented with 10% FBS, 100 units mL^{-1} penicillin, and 100 μg mL^{-1} streptomycin). Cells were centrifuged for 5 min at 500g followed by aspiration of the supernatant. The cells were washed twice by resuspension in ice-cold PBS, centrifugation for 5 min at 500g, and aspiration of supernatant. The cells were resuspended in ice-cold PBS supplemented with bovine serum albumin (2% w/v). The cell suspension was passed through a 35 μm nylon mesh to dissociate aggregates before analysis on SONY MA900-1 Cell Sorter. Distinct populations of cells with low (non-transfected) or high (transfected) mCherry fluorescence intensities were observed. 1 million cells with high mCherry fluorescence intensities were sorted into PBS supplemented with bovine serum albumin (2% w/v) and immediately kept on ice. Sorted, mCherry+ cells were then centrifuged for 5 min at 200g followed by aspiration of the supernatant. The cells were washed once by resuspension in ice-cold PBS, centrifugation for 5 min at 200g, and aspiration of supernatant. Cell pellets were flash-frozen in liquid nitrogen and stored at -20 ° until the time of western blot analysis.

Western blot analysis. Cells were lysed in ice-cold 1x RIPA lysis buffer (Abcam ab156034) supplemented with 1x Protease Inhibitor (Thermo 78425) and 100 mM dithiothreitol (DTT), and incubated on ice for 30 minutes. Lysates were centrifuged at 18,000 g, 4 °C for 15 min, and supernatants were transferred to a new tube and flash-frozen in liquid nitrogen. Protein concentration was determined using the BCA protein assay kit (Pierce). 15-20 μg of protein extract was separated on 4-12% Bis-Tris plus gels (15-well, 1.0 mm) transferred onto nitrocellulose membranes using the iBlot gel transfer system (Invitrogen), and blocked for 45 minutes at room temperature with blocking buffer (Intercept Blocking buffer in tris-buffered saline (TBS) (Licor 927-60001)). Blots were probed with the appropriate primary antibodies overnight at

4 °C in blocking buffer supplemented with 0.1% v/v Tween-20, followed by secondary antibodies IRDye® 680RD donkey anti-mouse IgG (Li-Cor; 1:20,000) or IRDye® 800CW donkey anti-rabbit IgG (Li-Cor, 1:20,000) for 1 hour at room temperature in blocking buffer supplemented with 0.1% v/v Tween-20 and 0.01% SDS. Fluorescent signals were imaged and quantified using Odyssey Imager (Li-Cor). Primary antibodies used were: pan-RAS (Santa Cruz Biotechnology, sc-166691; 1:1000), HSP90 (Cell Signaling Technology, C45G5; 1:1000), FLAG-tag (Cell Signaling Technology, D6W5B; 1:1000). Results of western blot analysis of endogenous RAS degradation among the best biodegrader constructs are shown in Figure 4-11.

Protein expression and purification of Sso7d proteins The expression and purification protocol of Sso7d proteins is as previously described in section 3.4.

Bio-layer interferometry (BLI). Samples were analyzed with the ForteBio (Sartorius) Octet RED-96 Biolayer Interferometry system using stabilization buffer (same as above) supplemented with 0.1% BSA, 20 μL/L Tween-20, and 10 μM GppNHp. Biotinylated GppNHp-loaded K-Ras WT or G12D was immobilized onto streptavidin-coated BLI-tips (Sartorius 18-5020). Association was analyzed at various concentrations of proteins (1:3 dilutions starting from 200 nM to 0.82 nM), followed by measuring dissociation in buffer. Two reference samples were included for every run: (1) RAS-loaded tips without addition of binder and (2) unloaded tips with addition of 200 nM of binder. The average binding curve from the two reference wells was subtracted from the data. Dissociation constant (K_d) values were obtained from steady-state binding analysis.

Sequences. The biodegrader constructs described in this chapter were designed modularly by combining different sets of KRAS-binding Sso7d (i.e. R11.1.6 or its variants), linkers, and E3 Ligases.

Biodegrader constructs with the E3 Ligase at the C terminus of Sso7d have

the following plasmid sequence:

FLAG-KRAS-binding-Sso7d-linker-E3-Ligase—(IRES2-backbone)—mCherry

M DYKDDDDK GSAGGGGS [KRAS-binding-Sso7d] [linker] [E3-Ligase] EF stop ------ M ATT
MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLP
FAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD
GEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYD
AEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK stop

catgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatgg
ctgactaatttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggaggctttttggaggcctaggct
tttgcaaagatcgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggag
aggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttg
tcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctg
tgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccga
gaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcga
gcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccag
gctcaaggcgagcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttct
ggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaa
tgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcggga
ctctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcgg
aatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccctagggggaggctaactgaaaca
cggaaggagacaataccggaaggaacccgcgctatgacggcaataaaaagacagaataaaacgcacggtgttgggtcgtttgttcataaacg
cggggttcggtcccagggctggcactctgtcgataccccaccgagaccccattggggccaatacgcccgcgtttcttccttttccccaccccacc
ccccaagttcgggtgaaggcccagggctcgcagccaacgtcggggcggcaggccctgccatagcctcaggttactcatatatactttagattga
tttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactg
agcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgcta
ccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttct
agtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtg
gcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagc
ccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggac
aggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggttt
cgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcc
tggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgccatgcattagttattaatagta
atcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgac
ccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaa
ctgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccca
gtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgg
gcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttc
caaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctggtttagtgaacc
gtcagatccgaggtggccgccacc atg gactataaggatgacgacgataaa ggaagcgcgggtggcggcggatcc [Sso7d] [Linker
] [E3Ligase] gaattc tga gcccctctccctcccccccccctaacgttactggccgaagccgcttggaataaggccggtgtgcgtttgtcta
tatgttattttccaccatattgccgtcttttggcaatgtgagggcccggaaacctggccctgtcttcttgacgagcattcctaggggtctttcccct
ctcgccaaaggaatgcaaggtctgttgaatgtcgtgaaggaagcagttcctctggaagcttcttgaagacaaacaacgtctgtagcgaccctttt
gcaggcagcggaacccccccacctggcgacaggtgcctctgcggccaaaagccacgtgtataagatacacctgcaaaggcggcacaaccccag
tgccacgttgtgagttggatagttgtggaaagagtcaaatggctctcctcaagcgtattcaacaaggggctgaaggatgcccagaaggtaccc

cattgtatgggatctgatctggggcctcggtacacatgctttacatgtgtttagtcgaggttaaaaaaacgtctaggccccccgaaccacgggg
acgtggttttccttgaaaaacacgatgataat atg gccacaacc atggtgagcaagggcgaggaggataacatggccatcatcaaggagt
tcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacc
cagaccgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacg
tgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcg
tggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgt
aatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctg
aagctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaa
catcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatgg
acgagctgtacaag tga gcggccgcgactctagatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacac
ctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcac
aaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaaggcgtaaattgtaagcgttaata
ttttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatag
accgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctat
cagggcgatggcccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagc
ccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggca
agtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcaggtggcactttcggggaaatgtg
cgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaa
aaggaagagtcctgaggcggaaagaaccagctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtat
gcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaag

Biodegrader constructs with the E3 Ligase at the N terminus of Sso7d have the following plasmid sequence:

FLAG-E3-Ligase-linker-KRAS-binding-Sso7d—(IRES2-backbone)—mCherry

M DYKDDDDK GSATG [E3-Ligase] [linker] [KRAS-binding-Sso7d] ASGGG stop ------ M ATT
MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLP
FAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD
GEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYD
AEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK stop

catgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatgg
ctgactaatttttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggaggctttttttggaggcctaggct
tttgcaaagatcgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggag
aggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttg
tcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctg
tgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccga
gaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcga
gcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccag
gctcaaggcgagcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttct
ggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaa
tgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcggga
ctctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcgg
aatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccctagggggaggctaactgaaaca
cggaaggagacaataccggaaggaacccgcgctatgacggcaataaaaagacagaataaaacgcacggtgttgggtcgtttgttcataaacg
cggggttcggtcccagggctggcactctgtcgataccccaccgagaccccattggggccaatacgcccgcgtttcttccttttccccacccccacc

ccccaagttcgggtgaaggcccagggctcgcagccaacgtcggggcggcaggccctgccatagcctcaggttactcatatatactttagattga
tttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactg
agcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgcta
ccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttct
agtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtg
gcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagc
ccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggac
aggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggttt
cgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcc
tggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgccatgcattagttattaatagta
atcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgac
ccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaa
ctgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccca
gtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgg
gcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttc
caaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctggtttagtgaacc
gtcagatccgaggtggccgccacc atg gactataaggatgacgacgataaa ggaagcgcgaccggt [E3-Ligase] [Linker] [Sso7
d] gctagcggcggcgga tga gcccctctccctcccccccccctaacgttactggccgaagccgcttggaataaggccggtgtgcgtttgtcta
tatgttattttccaccatattgccgtcttttggcaatgtgagggcccggaaacctggccctgtcttcttgacgagcattcctaggggtctttcccct
ctcgccaaaggaatgcaaggtctgttgaatgtcgtgaaggaagcagttcctctggaagcttcttgaagacaaacaacgtctgtagcgacccttt
gcaggcagcggaacccccccacctggcgacaggtgcctctgcggccaaaagccacgtgtataagatacacctgcaaaggcggcacaaccccag
tgccacgttgtgagttggatagttgtggaaagagtcaaatggctctcctcaagcgtattcaacaaggggctgaaggatgcccagaaggtaccc
cattgtatgggatctgatctggggcctcggtacacatgctttacatgtgtttagtcgaggttaaaaaaacgtctaggccccccgaaccacgggg
acgtggttttcctttgaaaaacacgatgataat atg gccacaacc atggtgagcaagggcgaggaggataacatggccatcatcaaggagt
tcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacc
cagaccgccaagctgaaggtgaccaagggtggcccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacg
tgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcg
tggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgt
aatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctg
aagctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaa
catcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatgg
acgagctgtacaag tga gcggccgcgactctagatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacac
ctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcac
aaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaaggcgtaaattgtaagcgttaata
ttttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatag
accgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctat
cagggcgatggcccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagc
ccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggca
agtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcaggtggcacttttcggggaaatgtg
cgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaa
aaggaagagtcctgaggcggaaagaaccagctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtat
gcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaag

And the protein and DNA sequences of each "module" (KRAS-binding Sso7d, linker,

or E3 ligase) are shown below.

Protein and DNA sequences of KRAS-binding Sso7d variants.

R11.1.6

ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggttt
aaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-XK

ATVRFTHQGEERQVDISRIRWVIRWGQYIWFRYDEDGGARGWGYVSERDAPRELLQMLR
RR

gctacggttcgtttcactcaccaaggcgaagaacgtcaagtagacatttcccgtatccgttgggtcattcgttggggtcaatacatttggttccg
ctatgacgaggacggtggtgctcgtgggtggggatacgtctccgagcgtgacgcgccccgtgagcttttacagatgttgcgtcgtcgt

R11-RbK

ATVRFTHQGEERQVDISRIRWVIRWGQYIWFKYDEDGGAKGWGYVSERDAPRELLQMLK
RR

gctacggttcgtttcactcaccaaggcgaagaacgtcaagtagacatttcccgtatccgttgggtcattcgttggggtcaatacatttggttcaa
atatgacgaggacggtggtgctaaagggtggggatacgtctccgagcgtgacgcgccccgtgagcttttacagatgttgaaacgtcgt

R11-XKA

ATVAFTHQGEEAQVDISAIAWVIRWGQYIWFAYDEDGGAAGWGYVSEADAPAELLQMLA
AR

gcaaccgtggcgttcacacaccaaggcgaagaagcgcaggtggatattagcgcgatcgcgtgggtaatccgttggggccagtacatttggttt
gcgtatgatgaagatggtggtgccgcgggttggggttatgtgagcgaagcggatgcaccggcggaactgctgcagatgctggcggcgcga

R11-nbKA

ATVAFTHQGEEAQVDISAIAWVIRWGQYIWFKYDEDGGAKGWGYVSEADAPAELLQMLK
AR

gcaaccgtggcgttcacacaccaaggcgaagaagcgcaggtggatattagcgcgatcgcgtgggtaatccgttggggccagtacatttggttt
aaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaagcggatgcaccggcggaactgctgcagatgctgaaagcgcga

R11-NB1

ATVKFTHQGEEKQVDISKIKAVIRWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaaggcggtaatccgttggggccagtacatttggtt
taaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB2

ATVKFTHQGEEKQVDISKIKWVARWGQYIWFKYDEDGGAKGWGYVSEKDAPKELLQML
KKR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtagcgcgttggggccagtacatttggtt
taaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB4

ATVKFTHQGEEKQVDISKIKWVIRWGQAIWFKYDEDGGAKGWGYVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccaggcgatttggtt
taaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB6

ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGAGYVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggttt
aaatatgatgaagatggtggtgccaaaggtgcgggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB7

ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGAVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggttt
aaatatgatgaagatggtggtgccaaaggttggggtgcggtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11$_{mut}$

ATVKFTHQGEEKQVDISKIKWVIRAGQYIAFKYDEDGGAKGAGYVSEKDAPKELLQMLKK
R

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgtgcgggccagtacattgcgtt
taaatatgatgaagatggtggtgccaaaggtgcgggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB11

ATVKFTHQGEEKQVDISKIKWVIRWGQAIWFKYDEDGGAKGWGYVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtttatttggttt
aaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB12

ATVKFTHQGEEKQVDISKIKWVIRWGQYIWFKYDEDGGAKGWGAVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggttt
aaatatgatgaagatggtggtgccaaaggttggggttttgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB15

ATVKFTHQGEEKQVDISKIKAVIRWGQFIWFKYDEDGGAKGWGYVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaaggcggtaatccgttggggccagtttatttggttt
aaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB16

ATVKFTHQGEEKQVDISKIKAVIRWGQYIWFKYDEDGGAKGWGFVSEKDAPKELLQMLK

KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggttt
aaatatgatgaagatggtggtgccaaaggttggggttttgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB17

ATVKFTHQGEEKQVDISKIKWVARWGQFIWFKYDEDGGAKGWGYVSEKDAPKELLQML
KKR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtagcgcgttggggccagtttatttggttt
aaatatgatgaagatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB18

ATVKFTHQGEEKQVDISKIKWVARWGQYIWFKYDEDGGAKGWGFVSEKDAPKELLQML
KKR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtagcgcgttggggccagtacatttggtt
taaatatgatgaagatggtggtgccaaaggttggggttttgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

R11-NB19

ATVKFTHQGEEKQVDISKIKWVIRWGQFIWFKYDEDGGAKGAGYVSEKDAPKELLQMLK
KR

gcaaccgtgaaattcacacaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtttatttggttt
aaatatgatgaagatggtggtgccaaaggtgcgggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga

Protein and DNA sequences of Linkers.

GS: GS / ggcagc

(GS)$_3$G: GSGSGSG / ggcagcggcagcggcagcggc

(GS)$_6$: GSGSGSGSGSGSGS / ggcagcggcagcggcagcggcagcggcagcggcagc

AA: AA / gcggcg

A(EAAAK)A: AEAAAKA / gcggaagcggcggcgaaagcg

A(EAAAK)$_2$A: AEAAAKEAAAKA / gcggaagcggcggcgaaagaagcggcggcgaaagcg

AP: AP / gcgccg

(AP)$_3$A: APAPAPA / gcgccggcgccggcgccggcg

(AP)$_6$: APAPAPAPAPAP / gcgccggcgccggcgccggcgccggcgccggcgccg

Protein and DNA sequences of E3 Ligases.

VHL

MPRRAENWDEAEVGAEEAGVEEYGPEEDGGEESGAEESGPEESGPEELGAEEEMEAGRP

RPVLRSVNSREPSQVIFCNRSPRVVLPVWLNFDGEPQPYPTLPPGTGRRIHSYRGHLWLFR
DAGTHDGLLVNQTELFVPSLNVDGQPIFANITLPVYTLKERCLQVVRSLVKPENYRRLDIVR
SLYEDLEDHPNVQKDLERLTQERIAHQRMGD

atgccccgccgcgcagagaattgggatgaggcagaagtaggagccgaggaagcaggcgttgaagaatatgggcctgaagaggacggtggcg
aagaaagtggtgcagaggaatcaggccctgaggagagcggtccagaggaactcggcgcagaggaggagatggaagctgggagaccccgtc
cggtcctcaggtctgtaaatagtcgcgaaccgagtcaagttatttttttgtaatagatccccccgcgtcgtcctgccagtctggcttaactttgatg
gggagcctcaaccgtatcctactctgccaccaggaacgggccggcgcatacactcttatcgaggtcatctctggcttttccgcgatgctggcac
acacgacggactcctcgtaaatcagacagaactgttcgtaccctctcttaacgttgatgggcaaccgatttttgccaacataacgctcccagtat
atactctcaaggaacgctgcctccaagtcgtgaggtctttggttaagcccgaaaactatcgacgcctcgatatcgttcgatctctgtacgaggat
cttgaagaccatccaaatgtacagaaggaccttgagcgcttgacccaagaacgcatagcccatcaaagaatgggcgat

pVHL

TLPVYTLKERCLQVVRSLVKPENYRRLDIVRSLYEDLEDHPNVQKDLERLTQERIAHQRMG
D

acgctcccagtatatactctcaaggaacgctgcctccaagtcgtgaggtctttggttaagcccgaaaactatcgacgcctcgatatcgttcgatc
tctgtacgaggatcttgaagaccatccaaatgtacagaaggaccttgagcgcttgacccaagaacgcatagcccatcaaagaatgggcgat

SPOP

SVNISGQNTMNMVKVPECRLADELGGLWENSRFTDCCLCVAGQEFQAHKAILAARSPVFS
AMFEHEMEESKKNRVEINDVEPEVFKEMMCFIYTGKAPNLDKMADDLLAAADKYALERL
KVMCEDALCSNLSVENAAEILILADLHSADQLKTQAVDFINYHASDVLETSGWKSMVVSHP
HLVAEAYRSLASAQCPFLGPPRKRLKQS

agcgtgaacatctccggccagaacacaatgaacatggtcaaggtgcccgagtgcagactggccgacgagctgggaggactgtgggagaactc
caggtttaccgactgctgcctgtgcgtggccggccaagagttccaagcccacaaagccatcctggccgctaggtcccccgtgttcagcgccatg
ttcgagcacgagatggaggagtccaagaagaacagagtggagattaacgatgtggagcccgaggtgttcaaagaaatgatgtgcttcatcta
caccggcaaggcccccaacctggataaaatggccgatgacctgctggccgccgccgataagtacgccctggagagactgaaggtgatgtgcg
aggacgctctgtgttccaacctgtccgtggaaaatgccgccgagatcctcatcctggccgacctgcatagcgccgaccagctgaaaacccagg
ccgtggacttcatcaactatcacgcttccgacgtgctggagaccagcggatggaagagcatggtggtgagccatccccatctcgtggccgaag
cctacaggagcctggcaagcgcccagtgtccctttctgggccctcccaggaagagactgaaacagagc

CHIP

RLNFGDDIPSALRIAKKKRWNSIEERRIHQESELHSYLSRLIAAERERELEECQRNHEGDEDD
SHVRAQQACIEAKHDKYMADMDELFSQVDEKRKKRDIPDYLCGKISFELMREPCITPSGITY
DRKDIEEHLQRVGHFDPVTRSPLTQEQLIPNLAMKEVIDAFISENGWVEDY

cggctgaacttcggggacgacatccccagcgctcttcgaatcgcgaagaagaagcgctggaacagcattgaggagcggcgcatccaccagga
gagcgagctgcactcctacctctccaggctcattgccgcggagcgtgagagggagctggaagagtgccagcgaaaccacgagggtgatgagg
acgacagccacgtccgggcccagcaggcctgcattgaggccaagcacgacaagtacatggcggacatggacgagctttttttctcaggtggatg
agaagaggaagaagcgagacatccccgactacctgtgtggcaagatcagctttgagctgatgcgggagccgtgcatcacgcccagtggcatc
acctacgaccgcaaggacatcgaggagcacctgcagcgtgtgggtcattttgaccccgtgacccggagccccctgacccaggaacagctcatc
cccaacttggctatgaaggaggttattgacgcattcatctctgagaatggctgggtggaggattac

IpaH9.8

LADAVTAWFPENKQSDVSQIWHAFEHEEHANTFSAFLDRLSDTVSARNTSGFREQVAAWL
EKLSASAELRQQSFAVAADATESCEDRVALTWNNLRKTLLVHQASEGLFDNDTGALLSLGR
EMFRLEILEDIARDKVRTLHFVDEIEVYLAFQTMLAEKLQLSTAVKEMRFYGVSGVTANDL
RTAEAMVRSREENEFTDWFSLWGPWHAVLKRTEADRWAQAEEQKYEMLENEYPQRVAD
RLKASGLSGDADAEREAGAQVMRETEQQIYRQLTDEVLALRLSENGSQLHHS

ctggctgatgccgtgacagcatggttcccggaaaacaaacaatctgatgtatcacagatatggcatgcttttgaacatgaagagcatgccaac
acctttccgcgttccttgaccgcctttccgatacagtctctgcacgcaatacctccggattccgtgaacaggtcgctgcatggctggaaaaact
cagtgcctctgcggagcttcgacagcagtctttcgctgttgctgctgatgccactgagagctgtgaggaccgtgtcgcgctcacatggaacaat
ctccggaaaaccctcctggtccatcaggcatcagaaggccttttcgataatgataccggcgctctgctctccctgggcagggaaatgttccgcct
cgaaattctggaggatattgcccgggataaagtcagaactctccattttgtggatgagatagaagtctacctggccttccagaccatgctcgca
gagaaacttcagctctccactgccgtgaaggaaatgcgtttctatggcgtgtcgggagtgacagcaaatgacctccgcactgccgaagccatg
gtcagaagccgtgaagagaatgaatttacggactggttctccctctggggaccatggcatgctgtactgaagcgtacggaagctgaccgctgg
gcgcaggcagaagagcagaaatatgagatgctggagaatgagtaccctcagagggtggctgaccggctgaaagcatcaggtctgagcggtg
atgcggatgcggagagggaagccggtgcacaggtgatgcgtgagactgaacagcagatttaccgtcagctgactgacgaggtactggccctg
cgattgtctgaaaacggctcacaactgcaccattca

Protein and DNA sequence of the negative control that lacks an E3 Ligase:

FLAG-R11.1.6—(IRES2-backbone)—mCherry

M DYKDDDDK GSAGGGGS [KRAS-binding-Sso7d] ASGGGTGEF stop ------ M ATT MVSKGE
EDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS
PQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKV
KLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTY
KAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK stop

catgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatgg
ctgactaatttttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggaggctttttttggaggcctaggct
tttgcaaagatcgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggag
aggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttg
tcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctg
tgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccga
gaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcga
gcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccag
gctcaaggcgagcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttct
ggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaa
tgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcggga
ctctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcgg
aatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccacccctaggggaggctaactgaaaca
cggaaggagacaataccggaaggaacccgcgctatgacggcaataaaaagacagaataaaacgcacggtgttgggtcgtttgttcataaacg
cggggttcggtcccagggctggcactctgtcgataccccaccgagaccccattggggccaatacgcccgcgtttcttccttttccccacccccacc
ccccaagttcgggtgaaggcccagggctcgcagccaacgtcggggcggcaggccctgccatagcctcaggttactcatatatactttagattga
tttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactg
agcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgcta
ccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttct

agtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtg
gcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagc
ccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggac
aggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtt
tcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttc
ctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgccatgcattagttattaatagt
aatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacga
cccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaa
actgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccc
agtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatg
ggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactt
tccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctggtttagtgaa
ccgtcagatccgaggtggccgccacc atg gactataaggatgacgacgataaa ggaagcgcgggtggcggcggatcc gcaaccgtgaaa
ttcacaccaccaaggcgaagaaaaacaggtggatattagcaaaatcaagtgggtaatccgttggggccagtacatttggtttaaatatgatgaa
gatggtggtgccaaaggttggggttatgtgagcgaaaaagatgcaccgaaagaactgctgcagatgctgaaaaagcga gctagcggcggcg
gaaccggtgaattc tga gcccctctccctcccccccccctaacgttactggccgaagccgcttggaataaggccggtgtgcgtttgtctatatg
ttattttccaccatattgccgtcttttggcaatgtgagggcccggaaacctggccctgtcttcttgacgagcattcctaggggtctttcccctctcg
ccaaaggaatgcaaggtctgttgaatgtcgtgaaggaagcagttcctctggaagcttcttgaagacaaacaacgtctgtagcgacccctttgca
ggcagcggaaccccccacctggcgacaggtgcctctgcggccaaaagccacgtgtataagatacacctgcaaaggcggcacaaccccagtgc
cacgttgtgagttggatagttgtggaaagagtcaaatggctctcctcaagcgtattcaacaaggggctgaaggatgcccagaaggtaccccat
tgtatgggatctgatctggggcctcggtacacatgctttacatgtgtttagtcgaggttaaaaaaacgtctaggccccccgaaccacggggacg
tggttttcctttgaaaaacacgatgataat atg gccacaacc atggtgagcaagggcgaggaggataacatggccatcatcaaggagttca
tgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccag
accgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtga
agcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtgg
tgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaat
gcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaag
ctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgccccggcgcctacaacgtcaacat
caagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacg
agctgtacaag tga gcggccgcgactctagatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctc
cccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaa
atttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaaggcgtaaattgtaagcgttaatattt
tgttaaaattcgcgttaaatttttgttaaatcagctcatttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagacc
gagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcag
ggcgatggcccactacgtgaaccatcaccctaatcaagtttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccc
cgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagt
gtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcaggtggcacttttcggggaaatgtgcgc
ggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaag
gaagagtcctgaggcggaaagaaccagctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgca
aagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaag

Protein and DNA sequence of KRAS-G12D-EGFP:

HA-tag-KRAS-G12D-EGFP

M ASS YPYDVPDYA SLGGPSGST MTEYKLVVVGADGVGKSALTIQLIQNHFVDEYDPTIE
DSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYRE
QIKRVKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIPFIETSAKTRQGVDDAFYTL
VREIRKHKEKMSKDGKKKKKKSKTKCVIM GGGGSGGGGS VSKGEELFTGVVPILVELDG
DVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKL
EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLS
TQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK stop

tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagac
aagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccata
tgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcagattggctattggccattgcatacgttgtatccatatcataatatg
tacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcat
agcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatga
cgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagt
gtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttccta
cttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacgg
ggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgcccc
attgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccgtcagatcgcctggagacgccatcc
acgctgttttgacctccatagaagacaccgggaccgatccagcctccgcggccgggaacggtgcattggaacgcggattccccgtgccaagag
tgacgtaagtaccgcctatagactctataggcacacccctttggctcttatgcatgctatactgtttttggcttggggcctatacacccccgcttc
cttatgctataggtgatggtatagcttagcctataggtgtgggttattgaccattattgaccactcccctattggtgacgatactttccattacta
atccataacatggctctttgccacaactatctctattggctatatgccaatactctgtccttcagagactgacacggactctgtatttttacagga
tggggtcccatttattatttacaaattcacatatacaacaacgccgtcccccgtgcccgcagtttttattaaacatagcgtgggatctccacgcg
aatctcgggtacgtgttccggacatgggctcttctccggtagcggcggagcttccacatccgagccctggtcccatgcctccagcggctcatggt
cgctcggcagctccttgctcctaacagtggaggccagacttaggcacagcacaatgcccaccaccaccagtgtgccgcacaaggccgtggcgg
tagggtatgtgtctgaaaatgagcgtggagattgggctcgcacggctgacgcagatggaagacttaaggcagcggcagaagaagatgcagg
cagctgagttgttgtattctgataagagtcagaggtaactcccgttgcggtgctgttaacggtggagggcagtgtagtctgagcagtactcgtt
gctgccgcgcgcgccaccagacataatagctgacagactaacagactgttcctttccatgggtcttttctgcag atg gcttcaagt tatccat
atgatgtgcccgactatgcc tcacttggaggaccctctggatccacc atgactgaatataaacttgtggtagttggagctgacggcgtaggca
agagtgccttgacgatacagctaattcagaatcattttgtggacgaatatgatccaacaatagaggattcctacaggaagcaagtagtaattg
atggagaaacctgtctcttggatattctcgacacagcaggtcaagaggagtacagtgcaatgagggaccagtacatgaggactggggagggc
tttctttgtgtatttgccataaataatactaaatcatttgaagatattcaccattatagagaacaaattaaaagagttaaggactctgaagatgt
acctatggtcctagtaggaaataaatgtgatttgccttccagaacagtagacacaaaacaggctcaggacttagcaagaagttatggaattcc
ttttattgaaacatcagcaaagacaagacagggtgttgatgatgccttctatacattagttcgagaaattcgaaaacataaagaaaagatgag
caaagatggtaaaaagaagaaaaagaagtcaaagacaaagtgtgtaattatg ggcggaggtggctccggtggcggaggttcc gtgagcaa
gggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcg
agggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgac
ctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcg
caccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagg
gcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcag
aagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccat
cggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatgg
tcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaag tga taggtcgacacgtgtgatcagatatcgcgg
ccgctctagaccaggcgcctggctcgagatcacttctggctaataaaagatcagagctctagagatctgtgtgttggttttttgtggatctgctg
tgccttctagttgccagccatctgttgtttgccccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatga

ggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcagcacagcaaggggaggattgggaagacaata
gcaggcatgctggggatgcggtgggctctatgggtacctctctctctctctctctctctctctctctctctctctcggtacctctctctctctct
ctctctctctctctctctctctctcggtaccaggtgctgaagaattgacccggttcctcctgggccagaaagaagcaggcacatccccttctctgt
gacacaccctgtccacgcccctggttcttagttccagccccactcataggacactcatagctcaggagggctccgccttcaatcccacccgctaa
agtacttggagcggtctctccctccctcatcagcccaccaaaccaaacctagcctccaagagtgggaagaaattaaagcaagataggctatta
agtgcagagggagagaaaatgcctccaacatgtgaggaagtaatgagagaaatcatagaatttcttccgcttcctcgctcactgactcgctgc
gctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaac
atgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcac
aaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcct
gttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcg
gtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaa
cccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaa
gtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctct
tgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcct
ttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagat
ccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcaccta
tctcagcgatctgtctatttcgttcatccatagttgcctgactccgggggggggggcgctgaggtctgcctcgtgaagaaggtgttgctgactc
ataccaggcctgaatcgccccatcatccagccagaaagtgagggagccacggttgatgagagctttgttgtaggtggaccagttggtgattttg
aacttttgctttgccacggaacggtctgcgttgtcgggaagatgcgtgatctgatccttcaactcagcaaaagttcgatttattcaacaaagccg
ccgtcccgtcaagtcagcgtaatgctctgccagtgttacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaactgca
atttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggc
aagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaat
caccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatc
aaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaac
aggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgtt
ttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccag
tttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcga
tagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagca
agacgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgt
gcaatgtaacatcagagattttgagacacaacgtggctttcccccccccccccattattgaagcatttatcagggttattgtctcatgagcggata
catatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatc
atgacattaacctataaaaataggcgtatcacgaggccctttcgtc

Chapter 5

Conclusions and Outlook

This thesis presents the development of novel degrader-based large-molecule targeted therapeutics for the treatment of RAS cancers, with each of the three projects focused on addressing different aspects of this undertaking. Although FDA-approved small-molecule inhibitors have been successfully developed against KRAS G12C, KRAS G12D remains an elusive target. KRAS G12D is one of the most common KRAS mutations, driving $\sim$ 8% of human cancers worldwide, yet there are currently no approved drugs that can effectively target this mutation.

To tackle this challenge, we developed a cell-permeable KRAS-G12D-targeting degrader, esterified R11.1.6-VL1 (est-R11-VL1), in chapter 3. est-R11-VL1 consists of a small-molecule E3 ligand (VL1) conjugated to esterified R11.1.6 (est-R11); R11.1.6 is a large-molecule binder to KRAS G12D and it was esterified to improve its transmembrane permeability. To address the practical challenge of low aqueous solubility of esterified proteins, we developed a solubilization strategy involving common pharmaceutical excipients in chapter 2. Our results demonstrate that the esterification of a protein molecule followed by exchange into an aqueous solution supplemented with β-CD provides a general means to esterify the carboxyl groups of a protein without compromising its solubility. We also show that the ability of a particular macrocycle to solubilize a conjugated protein can be determined by studying interactions between the macrocycle and a small-molecule mimetic of the conjugate (instead of the entire protein molecule). This strategy of using macrocycles as solubilizing agents could

also extend to esterified proteins formed with other diazo compounds.

In chapter 3, we demonstrate that est-R11-VL1 has modest efficacy at degrading intracellular RAS and inhibiting the growth of certain tumor cell lines. Even though the exogenous delivery of a large-molecule degrader was only modestly effective against KRAS G12D, we believe that this strategy presents a novel paradigm for targeting previously undruggable proteins. We also suggest that this approach may be effective against a more weakly-expressed protein target, as RAS might be too highly expressed. Additionally, given the challenges of reliably detecting RAS degradation via western blots, we recommend the use of more quantitative methods, such as the employment of the luminescence-based HiBit/LgBiT system, in the future development of exogenously-delivered large-molecule degraders.

In chapter 4, we expanded upon the previous research on KRAS-G12D-targeting biodegraders. Unlike est-R11-VL1, which is a hybrid molecule, biodegraders are fully protein-based constructs that consist of the KRAS G12D binder and an E3 ligase attached via a peptide linker. We utilized high-throughput fluorescence-based screening and regression modeling to develop several rational design principles for the creation of R11.1.6-based KRAS-G12D-targeting biodegraders. Our results indicated that the order of importance for the design features are: orientation of the components, choice of E3 ligase, linker length, and linker type. Moreover, we found that the efficiency of degradation is compromised when the affinity of R11.1.6 for KRAS G12D is weakened. Interestingly, due to the relatively high lysine richness of R11.1.6 (10 out of 61 residues are lysines), we suspect that there may be some level of self-degradation of the biodegrader molecule, but pilot studies were inconclusive in isolating and characterizing the effects of lysine richness on degradation efficiency; future work may provide invaluable insights for the development of more effective biodegraders.

Our findings offer a valuable contribution to the ongoing efforts in developing targeted therapies against RAS, which is a prevalent oncogene in human cancers. Our results suggest that large-molecule degrader-based strategies have the potential to be a promising avenue for developing therapies against previously considered undruggable targets such as RAS. We hope that these findings will inform further research and

exploration in this direction, and contribute towards potentially enabling RAS to become a more druggable target.

Bibliography

[1] Michael W. Traxlmayr, Jonathan D. Kiefer, Raja R. Srinivas, Elisabeth Lobner, Alison W. Tisdale, Naveen K. Mehta, Nicole J. Yang, Bruce Tidor, and K. Dane Wittrup. Strong enrichment of aromatic residues in binding sites from a charge-neutralized hyperthermostable sso7d scaffold library. *The Journal of biological chemistry*, 291:22496–22508, 10 2016.

[2] Anselm F.L. Schneider, Antoine L.D. Wallabregue, Luise Franz, and Christian P.R. Hackenberger. Targeted subcellular protein delivery using cleavable cyclic cell-penetrating peptides. *Bioconjugate Chemistry*, 30, 2019.

[3] Kevin M. Haigis. Kras alleles: The devil is in the detail. *Trends in Cancer*, 3, 2017.

[4] Andrew M. Waters and Channing J. Der. Kras: The critical driver and therapeutic target for pancreatic cancer. *Cold Spring Harbor Perspectives in Medicine*, 8, 2018.

[5] Ian A. Prior, Paul D. Lewis, and Carla Mattos. A comprehensive survey of ras mutations in cancer. *Cancer Research*, 72, 2012.

[6] Lamei Huang, Zhixing Guo, Fang Wang, and Liwu Fu. Kras mutation: from undruggable to druggable in cancer. *Signal Transduction and Targeted Therapy*, 6, 2021.

[7] Kevin Wood, Thomas Hensing, Raeva Malik, and Ravi Salgia. Prognostic and predictive value in kras in non-small-cell lung cancer. *JAMA Oncology*, 2, 2016.

[8] K A Mix, J E Lomax, and R T Raines. Cytosolic delivery of proteins by bioreversible esterification. *Journal of the American Chemical Society*, 139:14396–14398, 2017.

[9] Shuhui Lim, Regina Khoo, Yu-Chi Juang, Pooja Gopal, Huibin Zhang, Constance Yeo, Khong Ming Peh, Jinkai Teo, Simon Ng, Brian Henry, and Anthony W Partridge. Exquisitely specific anti-kras biodegraders inform on the cellular prevalence of nucleotide-loaded states. *ACS Central Science*, 7:274–291, 2021.

[10] Keith M Cheah, Joomyung V Jun, K Dane Wittrup, and Ronald T Raines. Host–guest complexation by β-cyclodextrin enhances the solubility of an esterified protein. *Molecular Pharmaceutics*, 19:3869–3876, 2022.

[11] K A Mix and R T Raines. Optimized diazo scaffold for protein esterification. *Organic Letters*, 17:2358–2361, 2015.

[12] V T Ressler, K A Mix, and R T Raines. Esterification delivers a functional enzyme into a human cell. *Acs Chemical Biology*, 14:599–602, 2019.

[13] B Testa and J M Mayer. *Hydrolysis in Drug and Prodrug Metabolism: Chemistry, Biochemistry, and Enzymology*. Verlag Helvetica Chimica Acta, 2003.

[14] B M Liederer and R T Borchardt. Enzymes involved in the bioconversion of ester-based prodrugs. *Journal of Pharmaceutical Sciences*, 95:1177–1195, 2006.

[15] L D Lavis. Ester bonds in prodrugs. *Acs Chemical Biology*, 3:203–206, 2008.

[16] J Rautio, J Kärkkäinen, and K B Sloan. Prodrugs—recent approvals and a glimpse of the pipeline. *European Journal of Pharmaceutical Sciences*, 109:146–161, 2017.

[17] C Tanford. The interpretation of hydrogen ion titration curves of proteins. *Adv. Protein Chem.*, 17:69–165, 1962.

[18] C Tanford. *Physical Chemistry of Macromolecules*. John Wiley & Sons, 1961.

[19] E Huckel P Debye. Zur theorie der elektrolyte. i. gefrierpunktserniedrigung und verwandte erscheinungen. *Phys. Z.*, 24:185–206, 1923.

[20] Eugène Fredericq and Hans Neurath. The interaction of insulin with thiocyanate and other anions. the minimum molecular weight of insulin1. *Journal of the American Chemical Society*, 72:2684–2691, 5 2002.

[21] M Hagenlocher and R Pearlman. Use of a substituted cyclodextrin for stabilization of solutions of recombinant human growth hormone. *Pharmaceutical Research*, 6:S30, 1989.

[22] R J Prankerd, H W Stone, K B Sloan, and J H Perrin. Degradation of aspartame in acidic aqueous-media and its stabilization by complexation with cyclodextrins or modified cyclodextrins. *International Journal of Pharmaceutics*, 88:189–199, 1992.

[23] A K Banga and R Mitra. Minimization of shaking-induced formation of insoluble aggregates of insulin by cyclodextrins. *Journal of Drug Targeting*, 1:341–345, 1993.

[24] S A Charman, K L Mason, and W N Charman. Techniques for assessing the effects of pharmaceutical excipients on the aggregation of porcine growth-hormone. *Pharmaceutical Research*, 10:954–962, 1993.

[25] J Horsky and J Pitha. Inclusion complexes of proteins—interaction of cyclodextrins with peptides containing aromatic-amino-acids studied by competitive spectrophotometry. *J. Inclusion Phenom. Mol. Recognit. Chem.*, 18:291–300, 1994.

[26] M Katakam and A K Banga. Aggregation of proteins and its prevention by carbohydrate excipients - albumins and gamma-globulin. *Journal of Pharmacy and Pharmacology*, 47:103–107, 1995.

[27] M Katakam and A K Banga. Aggregation of proteins and its prevention by carbohydrate excipients—albumins and γ-glubulin. *Journal of Pharmacy and Pharmacology*, 47:103–107, 1995.

[28] A Lovatt, A Cooper, and P Camilleri. Energetics of cyclodextrin-induced dissociation of insulin. *European Biophysics Journal with Biophysics Letters*, 24:354–357, 1996.

[29] T Loftsson and M E Brewster. Pharmaceutical applications of cyclodextrins .1. drug solubilization and stabilization. *Journal of Pharmaceutical Sciences*, 85:1017–1025, 1996.

[30] R A Rajewski and V J Stella. Pharmaceutical applications of cyclodextrins. 2. in vivo drug delivery. *Journal of Pharmaceutical Sciences*, 85:1142–1169, 1996.

[31] L Sharma and A Sharma. Influence of cyclodextrin ring substituents on folding-related aggregation of bovine carbonic anhydrase. *European Journal of Biochemistry*, 268:2456–2463, 2001.

[32] Y Dotsikas and Y L Loukas. Kinetic degradation study of insulin complexed with methyl-beta cyclodextrin. confirmation of complexation with electrospray mass spectrometry and 1h nmr. *Journal of Pharmaceutical and Biomedical Analysis*, 29:487–494, 2002.

[33] D E Otzen, B R Knudsen, F Aachmann, K L Larsen, and R Wimmer. Structural basis for cyclodextrins' suppression of human growth hormone aggregation. *Protein Science*, 11:1779–1787, 2002.

[34] M Khajehpour, T Troxler, V Nanda, and J M Vanderkooi. Melittin as model system for probing interactions between proteins and cyclodextrins. *Proteins-Structure Function and Bioinformatics*, 55:275–287, 2004.

[35] S Tavornvipas, S Tajiri, F Hirayama, H Arima, and K Uekama. Effects of hydrophilic cyclodextrins on aggregation of recombinant human growth hormone. *Pharmaceutical Research*, 21:2369–2376, 2004.

[36] S Tavornvipas, F Hirayama, S Takeda, H Arima, and K Uekama. Effects of cyclodextrins on chemically and thermally induced unfolding and aggregation of lysozyme and basic fibroblast growth factor. *Journal of Pharmaceutical Sciences*, 95:2722–2729, 2006.

[37] E Bajorunaite, A Cirkovas, K Radzevicius, K L Larsen, J Sereikaite, and V A Bumelis. Anti-aggregatory effect of cyclodextrins in the refolding process of recombinant growth hormones from escherichia coli inclusion bodies. *International Journal of Biological Macromolecules*, 44:428–434, 2009.

[38] A A Haroun and N R El-Halawany. Encapsulation of bovine serum albumin within β-cyclodextrin/gelatin-based polymeric hydrogel for controlled protein drug release. *Irbm*, 31:234–241, 2010.

[39] Meghan E Bush, Nicole D Bouley, and Adam R Urbach. Charge-mediated recognition of n-terminal tryptophan in aqueous solution by a synthetic host. *Journal of the American Chemical Society*, 127:14511–14517, 2005.

[40] Lisa M Heitmann, Alexander B Taylor, P John Hart, and Adam R Urbach. Sequence-specific recognition and cooperative dimerization of n-terminal aromatic peptides in aqueous solution by a synthetic host. *Journal of the American Chemical Society*, 128:12574–12581, 2006.

[41] Mikhail V Rekharsky, Hatsuo Yamamura, Young Ho Ko, N Selvapalam, Kimoon Kim, and Yoshihisa Inoue. Sequence recognition and self-sorting of a dipeptide by cucurbit[6]uril and cucurbit[7]uril. *Chemical Communications*, page 2236, 2008.

[42] S J Barrow, S Kasera, M J Rowland, J del Barrio, and O A Scherman. Cucurbituril-based molecular recognition (vol 115, pg 12320, 2015). *Chemical Reviews*, 116:12651–12652, 2016.

[43] R K K Sanku, O O Karakus, M Ilies, and M A Ilies. Inclusion complexes in drug delivery and drug targeting: Formation, characterization, and biological applications. *Targeted Nanosystems for Therapeutic Applications: New Concepts, Dynamic Properties, Efficiency, and Toxicity*, 1309:187–221, 2019.

[44] S N Timasheff. Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. *Advances in Protein Chemistry*, 51:355–432, 1998.

[45] H Hamada, T Arakawa, and K Shiraki. Effect of additives on protein aggregation. *Current Pharmaceutical Biotechnology*, 10:400–407, 2009.

[46] Satoshi Ohtake, Yoshiko Kita, and Tsutomu Arakawa. Interactions of formulation excipients with proteins in solution and in the dried state. *Advanced Drug Delivery Reviews*, 63:1053–1073, 2011.

[47] D J Leibly, T N Nguyen, L T Kao, S N Hewitt, L K Barrett, and W C Van Voorhis. Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *Plos One*, 7:13, 2012.

[48] J Mohanty, S Dutta Choudhury, N Barooah, H Pal, and A C Bhasikuttan. Mechanistic aspects of host–guest binding in cucurbiturils: Physicochemical properties. *Comprehensive Supramolecular Chemistry II*, pages 435–457, 2017.

[49] Paleg L G.; Stewart G R.; Bradbeer J W. Proline and glycine betaine influence protein solvation. *Plant Physiol.*, 75:974–978, 1984.

[50] Jai K Kaushik and Rajiv Bhat. Why is trehalose an exceptional protein stabilizer? *Journal of Biological Chemistry*, 278:26458–26465, 2003.

[51] E S Courtenay, M W Capp, C F Anderson, and M T Record. Vapor pressure osmometry studies of osmolyte-protein interactions: Implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro. *Biochemistry*, 39:4455–4471, 2000.

[52] T O Street, D W Bolen, and G D Rose. A molecular mechanism for osmolyte-induced protein stability. *Proceedings of the National Academy of Sciences*, 103:13997–14002, 2006.

[53] L R Singh, T A Dar, S Rahman, S Jamal, and F Ahmad. Glycine betaine may have opposite effects on protein stability at high and low ph values. *Biochim Biophys Acta*, 1794:929–935, 2009.

[54] Qin Zou, Brian J Bennion, Valerie Daggett, and Kenneth P Murphy. The molecular mechanism of stabilization of proteins by tmao and its ability to counteract the effects of urea. *Journal of the American Chemical Society*, 124:1192–1202, 2002.

[55] Cecilia C Mello and Doug Barrick. Measuring the stability of partly folded proteins using tmao. *Protein Science*, 12:1522–1529, 2003.

[56] Char Y Hu, Gillian C Lynch, Hironori Kokubo, and B Montgomery Pettitt. Trimethylaminen-oxide influence on the backbone of proteins: An oligoglycine model. *Proteins: Structure, Function, and Bioinformatics*, pages NA–NA, 2009.

[57] Anannya Bandyopadhyay, Kanika Saxena, Neha Kasturia, Vijit Dalal, Niraj Bhatt, Asher Rajkumar, Shuvadeep Maity, Shantanu Sengupta, and Kausik Chakraborty. Chemical chaperones assist intracellular folding to buffer mutational variations. *Nature Chemical Biology*, 8:238–245, 2012.

[58] Ravindra Singh Prajapati, Mili Das, Sridhar Sreeramulu, Minhajuddin Sirajuddin, Sankaranarayanan Srinivasan, Vaishnavi Krishnamurthy, Ranganathan Ranjani, C Ramakrishnan, and Raghavan Varadarajan. Thermodynamic effects of proline introduction on protein stability. *Proteins: Structure, Function, and Bioinformatics*, 66:480–491, 2006.

[59] M T Fisher. Proline to the rescue. *Proc. Natl. Acad. Sci. USA*, 103:13265–13266, 2006.

[60] Z Ignatova and L M Gierasch. Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proceedings of the National Academy of Sciences*, 103:13357–13361, 2006.

[61] B M Baynes, D I C Wang, and B L Trout. Role of arginine in the stabilization of proteins against aggregation. *Biochemistry*, 44:4919–4925, 2005.

[62] Dong Xin Zhao and Zhong Xian Huang. Effect of arginine on stability of gst–znf191(243–368). *Chinese Chemical Letters*, 18:355–356, 2007.

[63] Ranendu Ghosh, Sunny Sharma, and Krishnananda Chattopadhyay. Effect of arginine on protein aggregation studied by fluorescence correlation spectroscopy and other biophysical methods. *Biochemistry*, 48:1135–1143, 2009.

[64] Curtiss P Schneider, Diwakar Shukla, and Bernhardt L Trout. Arginine and the hofmeister series: The role of ion–ion interactions in protein aggregation suppression. *The Journal of Physical Chemistry B*, 115:7447–7458, 2011.

[65] Joan F Back, David Oakenfull, and Malcolm B Smith. Increased thermal stability of proteins in the presence of sugars and polyols. *Biochemistry*, 18:5191–5196, 1979.

[66] K.; Morikawa Gekko T. Preferential hydration of bovine serum albumin in polyhydric alcohol–water mixtures. *J. Biochem.*, 90:39–50, 1981.

[67] Ashutosh Tiwari and Rajiv Bhat. Stabilization of yeast hexokinase a by polyol osmolytes: Correlation with the physicochemical properties of aqueous solutions. *Biophysical Chemistry*, 124:90–99, 2006.

[68] R Liu, H Barkhordarian, S Emadi, C Park, and M Sierks. Trehalose differentially inhibits aggregation and neurotoxicity of beta-amyloid 40 and 42. *Neurobiology of Disease*, 20:74–81, 2005.

[69] Nishant Kumar Jain and Ipsita Roy. Effect of trehalose on protein structure. *Protein Science*, pages NA–NA, 2008.

[70] Florence Béranger, Carole Crozet, Andrew Goldsborough, and Sylvain Lehmann. Trehalose impairs aggregation of prpsc molecules and protects prion-infected cells against oxidative damage. *Biochemical and Biophysical Research Communications*, 374:44–48, 2008.

[71] Nishant Kumar Jain and Ipsita Roy. Trehalose and protein stability. *Current Protocols in Protein Science*, 59:4.9.1–4.9.12, 2010.

[72] J C Lee and S N Timasheff. The stabilization of proteins by sucrose. *J. Biol. Chem.*, 256:7193–7201, 1981.

[73] M E Ressing, W Jiskoot, H Talsma, C W Vaningen, E C Beuvery, and D J A Crommelin. The influence of sucrose, dextran, and hydroxypropyl-beta-cyclodextrin as lyoprotectants for a freeze-dried mouse igg2a monoclonal-antibody (mn12). *Pharmaceutical Research*, 9:266–270, 1992.

[74] B S. Kendrick; Carpenter J. F.; Cleland J. L.; Randolph T. W. A transient expansion of the native state precedes aggregation of recombinant human interferon-gamma. *Proc. Natl. Acad. Sci. USA*, 95:14142–14146, 1998.

[75] P Thordarson. Determining association constants from titration experiments in supramolecular chemistry. *Chemical Society Reviews*, 40:1305–1323, 2011.

[76] Jae Wook Lee, S Samal, N Selvapalam, Hee-Joon Kim, and Kimoon Kim. Cucurbituril homologues and derivatives: New opportunities in supramolecular chemistry. *Accounts of Chemical Research*, 36:621–630, 2003.

[77] J Szejtli. Introduction and general overview of cyclodextrin chemistry. *Chemical Reviews*, 98:1743–1753, 1998.

[78] J V Jun and R T Raines. Two-step synthesis of α-aryl-α-diazoamides as modular bioreversible labels. *Organic Letters*, 23:3110–3114, 2021.

[79] Monique J. Kauke, Michael W. Traxlmayr, Jillian A. Parker, Jonathan D. Kiefer, Ryan Knihtila, John McGee, Greg Verdine, Carla Mattos, and K. Dane Wittrup. An engineered protein antagonist of k-ras/b-raf interaction. *Scientific Reports*, 7, 2017.

[80] Nicole J. Yang and Marlon J. Hinner. Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods in molecular biology (Clifton, N.J.)*, 1266:29–53, 2015.

[81] Shonna A. Moodie, Berthe M. Willumsen, Michael J. Weber, and Alan Wolfman. Complexes of ras·gtp with raf-1 and mitogen-activated protein kinase kinase. *Science*, 260, 1993.

[82] Meg Trahey and Frank Mccormick. A cytoplasmic protein stimulates normal n-ras p21 gtpase, but does not affect oncogenic mutants. *Science*, 238, 1987.

[83] Adrienne D. Cox and Channing J. Der. Ras history: The saga continues. *Small GTPases*, 1, 2010.

[84] Adrienne D. Cox, Stephen W. Fesik, Alec C. Kimmelman, Ji Luo, and Channing J. Der. Drugging the undruggable ras: Mission possible? *Nature Reviews Drug Discovery*, 13, 2014.

[85] Andrew G. Stephen, Dominic Esposito, Rachelc G. Bagni, and Frank McCormick. Dragging ras back in the ring. *Cancer Cell*, 25, 2014.

[86] Soonsil Hyun and Dongyun Shin. Small-molecule inhibitors and degraders targeting kras-driven cancers. *International Journal of Molecular Sciences*, 22, 2021.

[87] Satoshi Imanishi, Lijuan Huang, Shoko Itakura, Masamichi Ishizaka, Yoichi Iwasaki, Tomohiro Yamaguchi, and Etsuko Miyamoto-Sato. In vivo kras g12d/v degradation mediated by canddy using a modified proteasome inhibitor. *bioRxiv*, 2022.

[88] Monique J. Kauke, Alison W. Tisdale, Ryan L. Kelly, Christian J. Braun, Michael T. Hemann, and K. Dane Wittrup. A raf-competitive k-ras binder can fail to functionally antagonize signaling. *Molecular Cancer Therapeutics*, 17, 2018.

[89] Kathleen M. Sakamoto, Kyung B. Kim, Akiko Kumagai, Frank Mercurio, Craig M. Crews, and Raymond J. Deshaies. Protacs: Chimeric molecules that target proteins to the skp1-cullin-f box complex for ubiquitination and degradation. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 2001.

[90] Ashton C. Lai and Craig M. Crews. Induced protein degradation: An emerging drug discovery paradigm. *Nature Reviews Drug Discovery*, 16, 2017.

[91] João M.J.M. Ravasco, Hélio Faustino, Alexandre Trindade, and Pedro M.P. Gois. Bioconjugation with maleimides: A useful tool for chemical biology. *Chemistry - A European Journal*, 25, 2019.

[92] Stephen M. Fuchs and Ronald T. Raines. Arginine grafting to endow cell permeability. *ACS Chemical Biology*, 2, 2007.

[93] Kim Quach, Jonathan LaRochelle, Xiao Han Li, Elizabeth Rhoades, and Alanna Schepartz. Unique arginine array improves cytosolic localization of hydrocarbon-stapled peptides. *Bioorganic and Medicinal Chemistry*, 26, 2018.

[94] David B. Thompson, James J. Cronican, and David R. Liu. Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. *Methods in Enzymology*, 503, 2012.

[95] Florian Krieger, Andreas Möglich, and Thomas Kiefhaber. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *Journal of the American Chemical Society*, 127, 2005.

[96] Pedro Chan, Robin A. Curtis, and Jim Warwicker. Soluble expression of proteins correlates with a lack of positively-charged surface. *Scientific Reports*, 3, 2013.

[97] Sainan An and Liwu Fu. Small-molecule protacs: An emerging and promising approach for the development of targeted therapy drugs. *EBioMedicine*, 36, 2018.

[98] Chung Hyo Kang, Dong Ho Lee, Chong Ock Lee, Jae Du Ha, Chi Hoon Park, and Jong Yeon Hwang. Induced protein degradation of anaplastic lymphoma kinase (alk) by proteolysis targeting chimera (protac). *Biochemical and Biophysical Research Communications*, 505, 2018.

[99] Daniel P. Bondeson, Alina Mares, Ian E.D. Smith, Eunhwa Ko, Sebastien Campos, Afjal H. Miah, Katie E. Mulholland, Natasha Routly, Dennis L. Buckley, Jeffrey L. Gustafson, Nico Zinn, Paola Grandi, Satoko Shimamura, Giovanna Bergamini, Maria Faelth-Savitski, Marcus Bantscheff, Carly Cox, Deborah A. Gordon, Ryan R. Willard, John J. Flanagan, Linda N. Casillas, Bartholomew J. Votta, Willem Den Besten, Kristoffer Famm, Laurens Kruidenier, Paul S. Carter, John D. Harling, Ian Churcher, and Craig M. Crews. Catalytic in vivo protein knockdown by small-molecule protacs. *Nature Chemical Biology*, 11, 2015.

[100] Mikhail M. Savitski, Nico Zinn, Maria Faelth-Savitski, Daniel Poeckel, Stephan Gade, Isabelle Becher, Marcel Muelbaier, Anne J. Wagner, Katrin Strohmer, Thilo Werner, Stephanie Melchert, Massimo Petretich, Anna Rutkowska, Johanna Vappiani, Holger Franken, Michael Steidel, Gavain M. Sweetman, Omer Gilan, Enid Y.N. Lam, Mark A. Dawson, Rab K. Prinjha, Paola Grandi, Giovanna Bergamini, and Marcus Bantscheff. Multiplexed proteome dynamics profiling reveals mechanisms controlling protein homeostasis. *Cell*, 173, 2018.

[101] Mette Ishoey, Someth Chorn, Natesh Singh, Martin G. Jaeger, Matthias Brand, Joshiawa Paulk, Sophie Bauer, Michael A. Erb, Katja Parapatics, André C. Müller, Keiryn L. Bennett, Gerhard F. Ecker, James E. Bradner, and Georg E. Winter. Translation termination factor gspt1 is a phenotypically relevant off-target of heterobifunctional phthalimide degraders. *ACS Chemical Biology*, 13, 2018.

[102] Kristin M. Riching, Sarah Mahan, Cesear R. Corona, Mark McDougall, James D. Vasta, Matthew B. Robers, Marjeta Urh, and Danette L. Daniels. Quantitative live-cell kinetic degradation and mechanistic profiling of protac mode of action. *ACS Chemical Biology*, 13, 2018.

[103] Morgan B Ludwicki, Jiahe Li, Erin A Stephens, Richard W Roberts, Shohei Koide, Paula T Hammond, and Matthew P DeLisa. Broad-spectrum proteome editing with an engineered bacterial ubiquitin ligase mimic. *ACS central science*, 5(5):852–866, 2019.

[104] R. Arai, H. Ueda, A. Kitayama, N. Kamiya, and T. Nagamune. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering*, 14, 2001.

[105] Dhan G. Bhandari, Barry A. Levine, Ian P. Trayer, and Mary E. Yeadon. 1h-nmr study of mobility and conformational constraints within the proline-rich n-terminal of the lc1 alkali light chain of skeletal myosin: Correlation with

similar segments in other protein systems. *European Journal of Biochemistry*, 160, 1986.

[106] Mei Zeng, Yuan Xiong, Nozhat Safaee, Radosław P. Nowak, Katherine A. Donovan, Christine J. Yuan, Behnam Nabet, Thomas W. Gero, Frederic Feru, Lianbo Li, Sudershan Gondi, Lincoln J. Ombelets, Chunshan Quan, Pasi A. Jänne, Milka Kostic, David A. Scott, Kenneth D. Westover, Eric S. Fischer, and Nathanael S. Gray. Exploring targeted degradation strategy for oncogenic krasg12c. *Cell Chemical Biology*, 27, 2020.

[107] Daniel P. Bondeson, Blake E. Smith, George M. Burslem, Alexandru D. Buhimschi, John Hines, Saul Jaime-Figueroa, Jing Wang, Brian D. Hamman, Alexey Ishchenko, and Craig M. Crews. Lessons in protac design from selective degradation with a promiscuous warhead. *Cell chemical biology*, 25:78, 1 2018.

[108] Haixiang Pei, Yangrui Peng, Qiuhua Zhao, and Yihua Chen. Small molecule protacs: An emerging technology for targeted therapy in drug discovery. *RSC Advances*, 9, 2019.

[109] Marcel Scheepstra, Koen F.W. Hekking, Luc van Hijfte, and Rutger H.A. Folmer. Bivalent ligands for protein degradation in drug discovery. *Computational and structural biotechnology journal*, 17:160–176, 1 2019.