2022

Construction of molecular tools through protein excision and splicing

BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

CONSTRUCTION OF MOLECULAR TOOLS THROUGH PROTEIN EXCISION AND SPLICING

by

SHANNON NICOLE TUNNEY

B.S., New Mexico Institute of Mining and Technology, 2015
M.S., Boston University, 2019

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2022

Approved by

First Reader

John T. Ngo, Ph.D.
Assistant Professor of Biomedical Engineering

Second Reader

Ahmad S. Khalil, Ph.D.
Associate Professor of Biomedical Engineering

Third Reader

Mary J. Dunlop, Ph.D.
Associate Professor of Biomedical Engineering

Fourth Reader

Allyson E. Sgro, Ph.D.
Assistant Professor of Biomedical Engineering

Fifth Reader

Trevor W. Siggers, Ph.D.
Associate Professor of Biology

DEDICATION

To my partner, who has spent this chapter of our lives far away, yet always there for me.

Thank you for your unwavering support. This wouldn't exist without you.

ACKNOWLEDGMENTS

I owe this to all the people who have been at my side, helping me through this and some pretty hard years. I'm not going to be able to mention them all, but know that all of you are appreciated. My advisor and the group he has put together have been amazing to work with and I can't imagine better colleagues for working through scientific problems or making the best acronyms. I'd like to thank my family whose unwavering support has been so instrumental to my success. I love you, and I'm sorry I didn't call more often. Especially thanks to my grandmother who always believed in me more than I did and loves to be right. I also would like to thank my partner who has been with me through all of this. I can't imagine this without your support and caring ear whenever I needed it and I love you so much. My roommates have supported me and put up with dropped chores when time got tight - I appreciate you all. Grace, thank you especially for giving me a place to stay for the last stage of my time in Boston. It really took a weight off my shoulders. Chengyi, I can't say how much I appreciate your company and friendship. Thank you to you and Samrath for taking me in when the pandemic extended my project. Additionally, my old roommates have stayed close despite me moving across the US and I have always appreciated your wit and suggestions for my next projects; you have given me the motivation to keep moving forward. Thank you also to all the wonderful people I have met in Boston and New England. I hope going forward, we can make some better memories together.

CONSTRUCTION OF MOLECULAR TOOLS THROUGH

PROTEIN EXCISION AND SPLICING

SHANNON NICOLE TUNNEY

Boston University, College of Engineering, 2022

Major Professor: John T. Ngo, Ph.D., Assistant Professor of Biomedical Engineering

ABSTRACT

With the explosion of protein tools as popular platforms for discovery and therapeutics, we see greater need for regulator systems that work congruently within these frameworks, especially safe and effective tools that can be implemented in humans. To this end, we endeavor to create orthogonal, precise and flexible protein modulators that can be easily employed to control protein tools with little need to iterate design for novel contexts. Hepatitis C NS3 protease is employed as a stabilizable linker between protein domains, enabling control over protein localization with FDA approved anti-viral drugs. The power of this tool is demonstrated by controlling gene expression through the controlled tether and release of a transcription factor. Inteins have already been employed to modulate proteins in synthetic contexts, however we observe that natural systems lack the avenues of control necessary to make them indispensable. We employ existing protein tools to construct a system of modular protein association, as well as drug and light inducible schema that reveal gaps in our knowledge of how to repurpose inteins in vivo. Despite this, we use inteins in the construction of a novel cargo delivery platform based on the fusogenic properties of the viral envelope glycoprotein from Vesicular stomatitis virus (VSV-G). We confirm reduced tropism of cargo delivery based on an

intein lock-and-key mechanism that has implications for both biosafety as well as targeted delivery in vivo of natively folded proteins to target cells.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFU	Arbitrary fluorescence units
BFP	Blue fluorescent protein
BU	Boston University
DB	DNA binding domain
DNA	Deoxyribonucleic acid
Dox	Doxycycline
EV	Extracellular vesicles
FDA	Food and Drug Administration
FKBP	FK506 binding protein
FRB	FKBP-rapamycin binding domain
GFP	Green fluorescent protein
Graz	Grazoprevir
h	Hour
H2B	Histone H2B
HCV	Hepatitis C virus
LDL-R	Low density lipid receptor
LInC	Ligand-inducible connection
LOV	Light, oxygen, or voltage domain
MP	Myristoylation-palmitoylation
NS3	NS3/4A Hepatitis C virus protease

NT	Nontransfected
PEST	Peptide rich in proline, glutamic acid, serine, and threonine
PhoCl	Photocleavable protein with sensitivity to 400nm light
RNA	Ribonucleic acid
RT	Room temperature
rTetR	Reverse tetracycline repressor
s.d.	Standard deviation
T7	Escherichia virus T7
TA	Transcriptional activator
TEV	Tobacco etch virus
TMD	Transmembrane domain
VEV	VSV-G mediated extracellular vesicle
VLPs	Virus-like particles
VSV	Vesicular stomatitis virus
VSV-G	Vesicular stomatitis virus glycoprotein
WT	Wild type

Chapter 1: The need for accessible protein systems

1.1 Protein tools for discovery and therapeutics

Genetically incorporated protein tools are powerful systems for both studying the fundamental principles of how life works and for engineering novel activity in synthetic structures. Protein-based systems have recently emerged as a new and exciting field of tools for treating [1,2] and even potentially curing diseases [3-5], dramatically improving quality of life for many patients. These include monoclonal antibodies, which have been used extensively during the Covid-19 pandemic [6-8], and chimeric antigen receptors (CARs) used to treat cancer [9,10]. These protein-based therapies are the product of increasing knowledge of protein structure [11-13], dynamics [14,15], and activity [16-18], and showcase the potential of protein manipulation for therapeutics.

The tools required for synthetic protein manipulation took off in the late 1970s with the elucidation of the lac operon, a system of protein regulation which had been identified the decade earlier [19], and its implementation synthetically [20-22]. The lac operon allows for control of gene expression through mechanisms accessible by researchers; the input or depletion of lactose in the system. Controlled gene regulation, along with the plasmids to carry genes [23] and successful recombinant gene expression [24], demonstrated the flexibility of living systems and their ability to express and utilize recombinant proteins. Protein expression systems and control over expression were used in both discovery platforms, to elucidate properties of unknown proteins [25,26], as well as for therapeutics, i.e., with the expression of recombinant insulin to treat diabetes [27].

In the present day, there are plentiful recombinant protein platforms, as regulating

bodies do not require approval for privately funded recombinant [28]. This makes recombinant protein research a popular platform for discovery and a powerful language for constructing new tools. This has led to more than 390 FDA approved recombinant proteins for diagnostics and therapies [1].

Proteins enjoy particular success as therapeutic molecules due to their biocompatibility with native systems. The most basic pharmacological purpose is replacing deficient systems with a more active analog, as with insulin [27] as well as pancreatic enzyme for pancreatitis [29], growth hormone for deficient children [30], and even lactase for the common lactose intolerance found in adults [31]. While these proteins may be active in their native conformation, and even may be recoverable from natives sources, such as pig pancreatic enzyme replacement therapy [29,32], recombinant protein production allows for added changes to the structure that may improve the therapeutic efficacy. Proteins generally have a short half-life in the human body due to natural proteases and processes, and recombinant changes allow for a longer active time and decrease the dosage rate necessary [33,34]. In addition to replacing deficient proteins, protein therapeutics can augment existing pathways, provide novel activity, interfere with destructive molecules or organisms, or deliver other compounds [35]. Our growing knowledge of protein structure and function allows us to better tailor these proteins to the specific function of interest, such as with tenecteplase, an engineered form of tissue plasminogen activator that has greater specificity than native for plasminogen and therefore more efficiently targets potentially life-threatening blood clots [36,37]. Despite our growing understanding of protein structure and function, there are still

existing challenges to the implementation of proteins as therapies, which can be addressed early in protein tool development.

1.2: The need for precise and orthogonal systems

Proteins have significantly different properties from established therapeutics. Traditional therapies have focused on small molecules, from penicillin and other antibiotics [38] to modern day newly approved cancer therapeutics [39,40]. Even with all the advances in therapeutic platforms during the Covid-19 pandemic, Pfizer's small molecule drug cocktail of nirmatrelvir and ritonovir was heralded as a significant advance in fighting Sars-Cov-2 [41,42]. The "small" aspect of small molecule drugs is so important that it was even published in Lipinski's Rule of Five, stating that therapeutic drugs should be no more than 500 daltons (Da) in size for best pharmacokinetics in the human body, based on a study of the most effective therapeutics [43]. This has influenced the therapeutics that have made it to market [44]. Proteins are very large when compared to small molecule drugs, and so have different adsorption, distribution, metabolism, and excretion (ADME). A single amino acid is, on average, 110 Da, so to be within the rule of five, a protein could only be four amino acids long. Insulin is a small protein at only 51 amino acids long, and yet is still 10 times the Rule of Five's ideal molecular weight at 5.8 kDa [45]. This complicates therapeutic delivery in patients and even now there are efforts to make insulin products with important differences in adsorption and bioavailability to better treat the needs of individuals [46].

Additionally, beyond the importance of size in bioavailability and diffusion,

proteins have more moving pieces than small molecule drugs and this leads to additional complications when administered as a therapeutic. Proteins are defined by their structure but are synthesized as an amino acid strand that can take on many conformations, of which only a small subset have desirable activity [47,12]. Small molecule drugs are chosen for their stability, and in the case of instability there are established strategies to improve upon it [44,48]. Protein therapeutics cannot take advantage of these established guidelines, so new strategies must be devised to get bioactive proteins from production, through transport within the medical system and within the body to the target [34,49,50].

Even when proteins are successfully delivered to the target tissue, there can still be complications due to their extensive biocompatibility. A platform that has garnered particular interest is the CRISPR/Cas9 system [51]. The protein Cas9 can be used in a therapeutic context as a gene editor for genetically inherited diseases [52,53]. Despite the theoretical specificity of Cas9 and other base editors to target the exact sequence of interest, they have become known for off-target DNA cleavage [54-57]. When CRISPR/Cas9 is implemented in a laboratory context, this can be problematic, however in a therapeutic context it is potentially deadly. This has held back advances in medicine using these tools, as scientists have first searched for methods for reducing these off-target effects [57]. In the meantime, 10 million Americans suffer from rare diseases which might be treatable via in vivo gene editing [52]. In order for this protein tool to become a lifesaving technology, the precision of implementation must be improved.

Cas9 is not the only potential protein therapeutic with this complication. Off-target activity is a problem for many recombinant protein systems, including

recombinases [58], protein binding elements [59], and receptors [60]. Regulators of protein activity have been employed to increase the specificity of protein therapeutic elements, for example the Cas9 protein has been split such that it can only be brought together by the small molecule activator rapamycin [61,62]. This temporal control, where the Cas9 is only active at the point in time when the target has been treated with the activator, restricts the off-target activity as determined by deep sequencing [61]. However, this system had to employ multiple inhibition strategies as the simply split protein continued to spontaneously assemble, and more recent work has made improvements to the system [62].

In addition to the limitations in regulatory capability of current strategies for controlling protein activity, many systems have off-target effects as well. This includes the popular regulator rapamycin, as it was first used as a therapeutic itself. Rapamycin is a small molecule drug that binds protein domains FRB and FKBP in the mTor pathway, which is an important regulator of metabolism and cell growth in humans as well as other organisms [63]. The protein binding domains FRB and FKBP are often recombinantly added to proteins that would not otherwise spontaneously associate in order to bring them within interacting distance [64,65]. This system allows temporal control over the association between two proteins by addition or depletion of the small molecule, which makes it a powerful regulator. However, despite its utility and popularity in the laboratory, the use of rapamycin as the activator is a problem in therapeutic contexts, since the small molecule exerts control over metabolism and cell growth, affecting the condition of the patient [63]. Multiple other popular regulators come from therapeutics,

like doxycycline for tet-On and tet-off systems [66], and 4-hydroxytamoxifen for altering estrogen receptor domain binding [67]. All of these tools were employed in mammalian systems due to their known safety and efficacy from therapeutic uses, however, their affinity for off-target elements within a mammalian cell means that they are inappropriate for sensitive applications [67]. This problem has been addressed throughout the years by the development of rapamycin analogs and mutational screening for FKBP and FRB domains that preferentially bind to them, but there is considerable development time required [59,68,69].

In this work, we propose strategies to improve the precision of protein tools using protein regulators that are orthogonal to mammalian biology. We endeavor to employ systems that can be encoded along with the protein tools that only interact with mammalian systems with the single, target activity. We specifically look for protein components that have few-to-no natural interactions with mammalian systems, especially looking to other protein sources such as other species and synthetic sources. While protein therapies show a lot of promise, they also require more precise regulation than traditional medicines and we focus on regulatory tools that can be implemented in a therapeutic context without extraneous interactions.

1.3: The need for plug-and-play design elements

Development in the field of biotechnology has taken off in recent years, especially considering the influx of funding due to the Sars-Cov-2 pandemic. The NIH, for example, was awarded 3 billion more dollars in 2020 for use on Covid-19 related

research than normal budget increases, a 7% increase over the expected budget [70]. The private sector funds a lot of final stage development in biotechnology, and venture capital funding saw an increase from 60 billion dollars in funding in 2019 to 120 billion dollars in 2021 for biotechnology development [71,72]. Not only has the increase in funding increased the resources available, but the expanding knowledge base has shortened the time necessary to take a new tool from technology from therapeutic; for example the Moderna Sars-Cov-2 vaccine was developed January 2020 and given an emergency release for use in humans December 2020 [73]. It is now fully FDA approved in January 2022, a two-year development time from initial work to FDA approval, an unprecedented speed when compared to other modern vaccines which have a 4-to-20-year development time [74]. As the speed of biotechnological research has increased and many development processes are available to us, we come upon one of the main limiting factors: the need to iterate development and testing for implementation of techniques in novel systems [75].

In many cases, domain analysis and recombination of a given protein therapeutic can be conducted to enable it to be used in different contexts. As an example, antibodies are most easily produced by human or mouse hybridomas, an intensive process in need of a highly skilled researcher [76]. In order to meet the increased production demand upon using antibodies as a therapeutic, production can be shifted to bacteria, virus, or yeast, and each of these systems require extensive changes to the antibody structure for successful production [77-79]. In other cases, directed evolution can be harnessed to produce a needed protein with specific activity [80-82]. Even the best of these systems

takes time and expertise to implement, and many directed evolution schemes are complex or lead to cheaters rather than useful outputs [83,84]. It is especially difficult to use known systems and tools in unknown systems, such as when functionalizing a new organism [85].

The ideal protein tool would be consistent in a wide variety of contexts and need very little alteration for employment in the regulation of a novel protein. We venture to create systems that are applicable to as wide a variety of protein contexts as possible, with little need to test elements for each application. This decreases development time and increases ease of use, making this an important design criterion for broad applicability within the scientific community.

1.4: Orthogonal control of protein regulation

In Chapter 2, we look at strategies for precise and orthogonal control of protein regulation using the Hepatitis C virus NS3/4A protease (NS3). We seek orthogonality via the use of a protease and inhibitor pair that act independently of mammalian systems; the protease is an internal viral activator and the small molecule drug inhibitor can be one of many FDA approved treatments for Hepatitis C [86,87]. While the drug is already a therapeutic and therefore has bioactivity, it retains orthogonality because its target is the Hepatitis C protein rather than a mammalian target like rapamycin. Glecaprevir, one of the more recent FDA approved inhibitors, is considered safe in children as young as 12 years old and is only contraindicated in individuals with existing liver disease. Most people who take it have none to mild side effects of headache and fatigue [88]. We

employ this orthogonal pair as a stabilized protein connector.

Hepatitis C NS3 is a cis-acting protease, which allows it to cleave itself from a contiguous strand of amino acids [86]. When this domain is placed between two protein domains, it becomes a drug dependent linker, and the connection can be destroyed by removal of the protease inhibitor and subsequent activation of the protease. We place this between an orthogonal transcription factor with control over a gene of interest and a membrane-bound protein domain to create a "drug-off" system. When the drug is added, the gene turns off. These components are modular, and each domain is self-sufficient, creating a plug-and-play system where any transcription factor could be employed, as well as multiple membrane binding domains; we test a variety. Through this design scheme, we endeavor to produce a gene regulation system that can be used safely and effectively in humans, such as in regulation of a gene in engineered CAR-T cells, and does not require iterative testing to be employed in novel contexts.

1.5: Autonomous protein elements in control schemes

In Chapter 3, we describe our work of employing small active peptides called inteins as autonomous protein regulators. Inteins are single-use protein elements that catalyze the excision of a domain out of a strand of amino acids and splice the adjacent domains together [89]. This conjugation can either be within a single protein or spliced between two separate proteins via split inteins [90]. In the native context, inteins often have intervening sequences that act as gene regulators or cytokines, however these can easily be removed, leaving only splicing activity [91]. Additionally, while inteins have

been found in all domains of life, the inteins found in unrelated species have undetectable cross-reactivity, allowing inteins from yeast and bacteria to be used orthogonally in mammalian contexts [92].

We employ these splicing elements as control systems for protein activity. First, we use the Gp-41-1 split intein to circularize a protein as a regulator of protein stability and therefore activity. We additionally endeavor to use this in conjunction with a degradation tag and a cleavable element, thereby mediating stability via an active protease. Secondly, we endeavor to create inteins with inducible activity. We approach this problem using both the Hepatitis C protease control scheme and a light-sensitive protein cleaving domain. With the protease control scheme, we placed the cleaving domain between the active splicing machinery and associating domains that allow the inteins to splice such that when the protease has cleaved, the inteins are unable to associate in the proper configuration and are inactive. This is a small-molecule drug sensitive splicer. For the light sensitive control scheme, we place the photocleavable protein domain between an active intein and a caging domain. Without light, the intein is caged and inactivated but with the light cleaving the intein free from the cage, the intein is able to activate and splice. A key design element in all of these systems is modularity. We constructed systems where the protein domains external to the intein can be swapped freely to allow for plug-and-play control over protein stability and activity.

1.6: A small peptide lock system for protein delivery

In Chapter 4, we describe a system of protein delivery using inteins as a biological lock. Extracellular vesicles (EVs) have emerged as a promising delivery platform for therapeutic proteins due to the ability of mammalian cells to construct and secrete them, enabling packaging of proteins in their native context [93]. Particularly popular are EVs created by the expression of viral capsid proteins due to their production efficiency and uniformity [94]. Viral EVs have the powerful capability to mediate fusion with the membrane of their target cell, avoiding many complicating cellular processes which impede the delivery of the therapeutic. We propose the addition of a genetically integrated biological lock that can be expressed in the target cells to allow for delivery of a protein specifically to the tissue of interest.

We use one half of a split intein as an intervening sequence between a protein cargo and a membrane binder. Upon entry of the viral EV to the target cell, the other half of the intein expressed by that cell will initiate splicing, leading to targeted release of the protein cargo from the membrane. This biological lock is orthogonal to normal mammalian biology given the use of non-mammalian inteins [92]. Additionally, it is modular, allowing the use of any free-floating protein for delivery to a target tissue, and we propose any protein associated cargo such as Cas9, which contains a protein component and a nucleic acid component.

This system is flexible and allows for the targeted delivery of cargo to any tissue that can be induced to produce the intein release protein. This plug-and-play system takes advantage of the autonomous splicing activity of inteins for targeted release of a cargo.

We endeavor to display the power of protein engineering and synthetic biology to construct powerful and flexible tools that can be employed in many contexts including therapeutics.

Chapter 2: Chemogenetic control of protein localization and gene expression with antiviral drugs

Chapter 2 is adapted from the following manuscript:
EP Tague, HL Dotson, SN Tunney, DC Sloas, JT Ngo. Chemogenetic control of gene expression and cell signaling with antiviral drugs. Nature Methods (2018).

2.1: Background and Motivation

A goal of synthetic biology is to program new functions into cells in ways that can be precisely manipulated for applications in medicine and basic research. Toward this end, researchers have recombined modular domains from natural signaling proteins and genetic control elements to produce new biological 'parts' for cellular engineering [95]. Among the parts that have been especially valuable in synthetic biology are a set of ligand-binding proteins that can be used to engineer drug sensitivity into protein systems, such as the rapamycin system taken from the mammalian mTor pathway [63]. These tools have been used to render numerous cellular processes subject to drug control, including transcription [96], cell signaling [97,98], and genome-editing [99,100]. However, as the goals and capabilities in synthetic biology continue to expand, there remains a need for new inducible platforms. Particularly, there is demand for new inducers few existing interactions with the biological context of the technology.

We set out to devise a new, modular platform that could be not only combined with the extant methods, but also generalized to many different systems. Given recent interest in the development of drug-controllable cell-based therapies, we were especially interested in designing a strategy that allowed precise gating of synthetic gene circuits and signaling pathways by safe and well-characterized small molecules [101]. Seeking to

create an orthogonal system, we targeted inducer compounds that are typically absent from the environment, particularly those with binding preference for nonhuman proteins. To identify such inducers, we turned our attention to antiviral drugs, especially those that have been tested (or approved) for clinical use in humans.

It is a particular focus in the production of antiviral drugs to target the viral proteins rather than the human system surrounding it, which is in contrast to, for example, anti-cancer drugs, which focus on activating and directing the immune system [102, 103]. This approach is successful due to the significantly different biology between virus and host. Viral genomes are, by necessity, very small in order to be easily constructed by the host machinery and packaged safely within the capsid [104]. A sequence shortening measure employed by many viruses is their own protease, which acts to cleave apart separate proteins within the viral polyprotein [105]. This is in contrast to mammalian genetics systems, where each separate protein has their own regulating promoter region and terminator [106]. Inhibitors for these viral proteases therefore target a system completely orthogonal to mammalian biology and therefore attractive for repurposing as a protein tool.

Here we describe our use of the hepatitis C virus (HCV) NS3 protease domain and its various inhibitors as a protein–ligand pair that can be applied to install drug sensitivity into both intracellular and cell-surface proteins. In its natural context, NS3 is a serine cis-protease that excises itself from the HCV polyprotein by cleaving recognition sites that flank it at either end [107]. Because it is essential for HCV replication, numerous inhibitors targeting the viral protease have been developed. In previous work,

cis-cleavage by NS3 was used to conditionally link proteins to imaging tags and degradation sequences, and NS3 trans-proteolysis has been applied to control the activity of T7 RNA polymerase [108-111]. Given its successful application in these non-natural contexts, we asked whether the viral protease could be used to design drug-sensitive systems and gain control over complex cellular processes such as protein localization and transcription.

2.2: A "Turn-on" system as the basis of design

NS3 was used to create drug-sensitive transcription factors (TFs), using the protease as a ligand-inducible connection (LInC) to control the association between modular DNA-binding (DB) and transcriptional-activation (TA) domains. The protease was inserted between minimal DB and TA sequences sourced from the yeast TF Gal4, thus generating $Gal4_{DB}$-NS3-$Gal4_{TA}$ (**Figure 1a**). In this configuration, the viral protease formed a self-immolating connection, excising itself from the fusion construct and, in doing so, separating the DB and TA elements. In the presence of an NS3 inhibitor, self-excision of the protease is blocked, resulting in the preservation of full-length TF capable of activating the expression of targeted genes.

To determine whether Gal4DB-NS3-Gal4TA the protein was expressed in a mammalian cell line containing a stably integrated Gal4-dependent reporter construct (UAS H2B-Citrine). Immunoblotting showed that, although full-length Gal4DB-NS3-GalTA copies were not detected in drug-untreated cells, formation of the intact TF was rapidly induced after exposure to NS3 inhibitor (**Figure 1b**). In addition to preserving

newly made Gal4DB-NS3-Gal4TA copies, drug treatment also induced expression of the H2B-Citrine reporter protein in a dose-dependent manner (**Figure 1c**). This was analyzed by flow cytometry with gating for live HEK293FT cells and a transient transfection marker (**Figure 1d**). This system was tested with various inhibitors and multiple compounds capable of activating robust transcriptional responses were identified, including BILN-2061, asunaprevir, danoprevir, and grazoprevir, some of which are FDA approved (**Figure 2a**) [112]. Together, these results indicate that the LInC strategy can be used to precisely regulate the association of TF elements to achieve inducible control over gene expression.

In addition, TFs with activity against alternative promoters were tested, including one in which the reverse tetracycline repressor (rTetR) was used as the DB element (rTetR-NS3-VP64-p65). In transfected cells, rTetR-NS3-VP64-p65 exhibited 'AND' gate activity, requiring both doxycycline and an NS3 inhibitor to activate transcription from the tetO-containing TRE promoter (**Figure 2b,c**). To complement the 'turn-on' systems described above, we designed a strategy in which NS3 inhibitors could be applied to 'turn-off' gene expression.

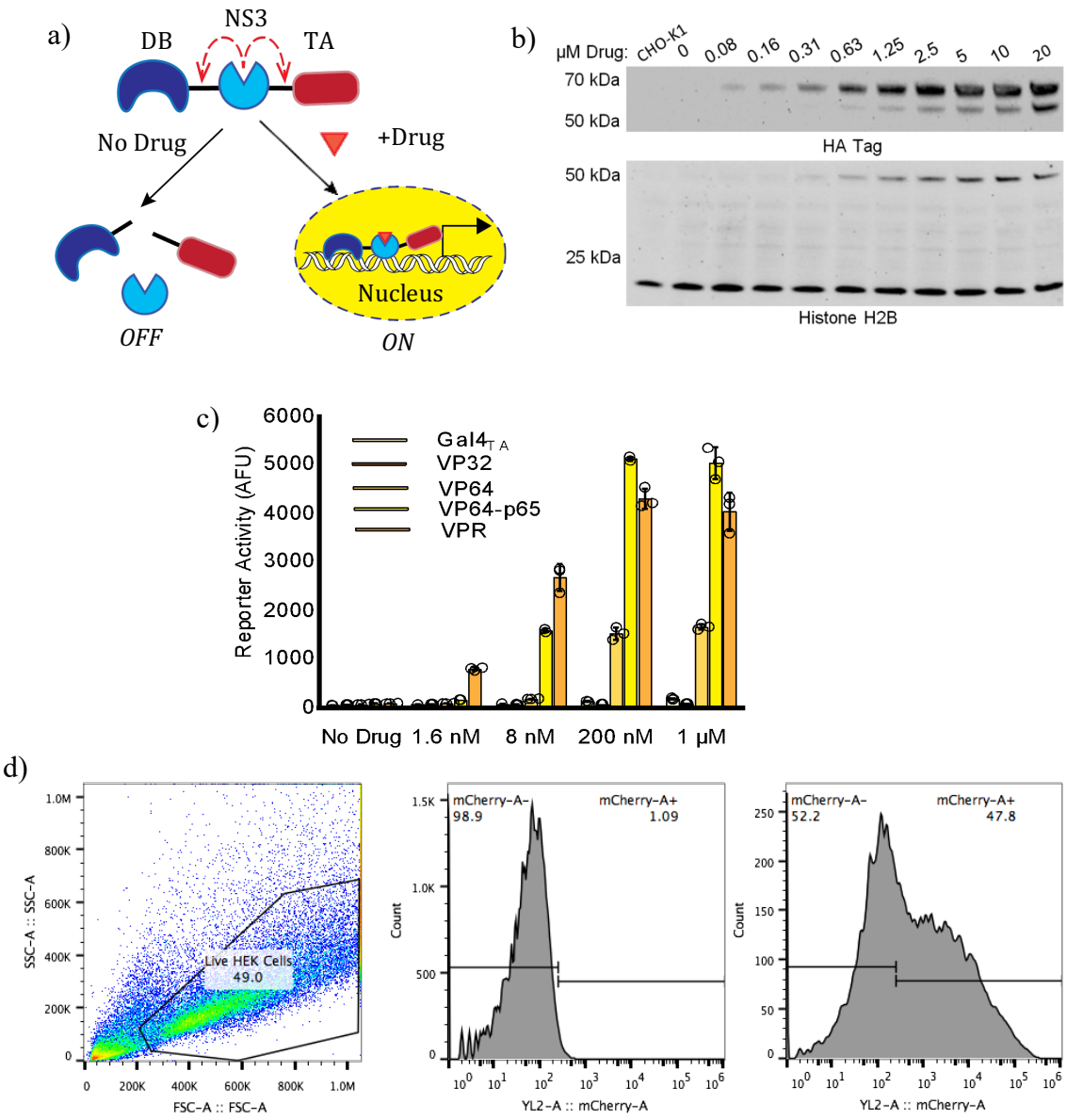


Figure 1: A "Turn-on" drug inducible TF system

a) Design of the drug-inducible turn-on TFs. **b)** Western blot showing accumulation of full length $Gal4_{DB}$-NS3-$Gal4_{TA}$ (anti-HA; Santa Cruz Biotechnology, sc-7392; 60.6 kDa) in response to BILN-2061, and corresponding H2B-Citrine expression (anti-H2B, Cell Signalling 12364; 46.5 kDa). Endogenous histone H2B (13.8 kDa) served as a loading control. **c)** H2B-Citrine fluorescence as determined by flow cytometry in HEK293FT cells carrying reporter DNA (UAS H2B-Citrine) and transiently expressing TFs with the indicated TA domains. Cells were treated with denoted concentrations of BILN-2061. **d)** In this chapter, all live cells were gated using FSC and SSC as represented by the area enclosed within the black line. Non-transfected cells were used to determine the top 1% of fluorescence, and this positive transfection gate was then applied to all transfected cell populations. Geometric mean of reporter fluorescence was then measured from positively transfected cells.

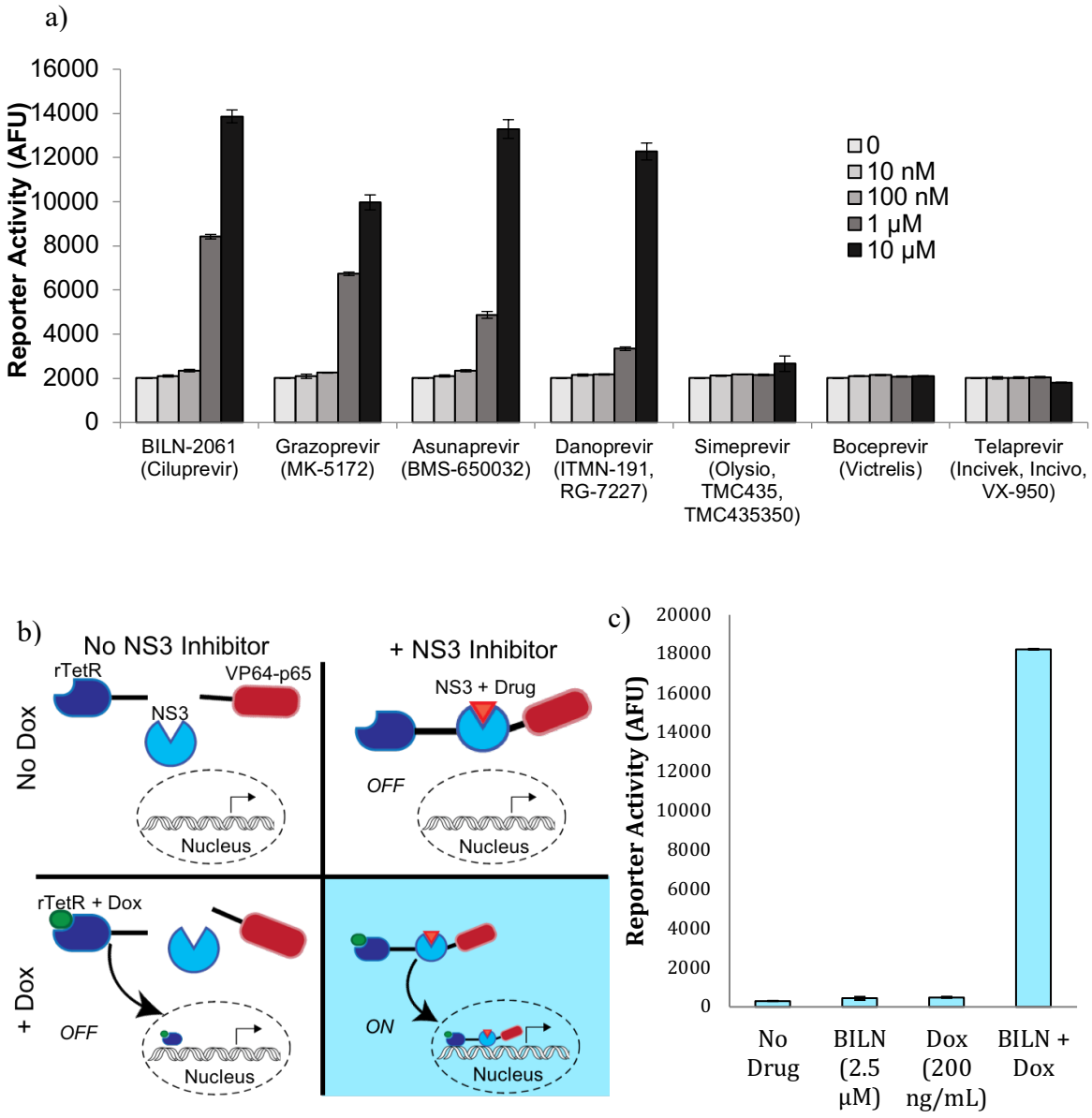


Figure 2: Drug sensitivity and AND-gating of the "Turn-on" drug inducible TF
a) Activation of a UAS H2B-Citrine reporter gene in a Cho-K1 reporter line constitutively expressing $Gal4_{DB}$-NS3-$Gal4_{TA}$ in response to treatment with various NS3 inhibitors. Drug induced expression of H2B-Citrine reporter protein was quantified by flow cytometry. We note that the NS3 sequence used in the LInC module contains the T54A mutation, which confers resistance against telaprevire and boceprevir. Geometric means are displayed as mean ± s.d., as determined by three biologically independent samples. **b)** Schematic of AND gate behavior of rTetR-NS3-VP64-p65; the presence of both doxycycline (to induce rTetR binding to tetO sequences) and an NS3 inhibitor (BILN-2061) are required to activate transcription from the TRE promoter. **c)** BFP expression levels in Cho-K1 cells co-transfected with construct and BFP reporter. Cells were treated with drug for 24 hours prior to flow cytometry. Background activation was assessed through comparison with cells transfected with reporter DNA only.

2.3: Construction of a "Turn-off" System

2.3.1 Structure and Activity

In this approach, we used NS3 to conditionally tether an intact Gal4 (Gal4min) unit to a membrane-targeting domain, with the expectation that protease inhibition could be used to precisely control the amount of soluble versus membrane-bound TF.

Using a type I transmembrane protein as a targeting element, we generated a fusion construct containing NS3 and minimal Gal4 containing just the DNA binding domain and the transcriptional activator elements (Gal4min) as a C-terminal cytosolic domain (TMD-NS3-Gal4min) (**Figure 3a**). Fluorescence imaging of cells expressing a dual-tagged version of the protein (BFP-TMD-NS3-Gal4min-mCherry) showed that Gal4min was released from its BFP-fused transmembrane domain in drug-untreated cells, resulting in a liberated TF unit that localized predominantly to the nucleus (**Figure 3b**). However, in cells in which NS3 activity had been inhibited, the TF remained attached to its targeting element and thus was trafficked to the endoplasmic reticulum and plasma membrane. A version in which an N-terminal myristoylation and palmitoylation substrate was used as the targeting sequence (myr-palm-NS3-Gal4min) exhibited similar behavior, becoming occluded from the nucleus in drug-treated cells (**Figure 3c**). This was also true when the targeting domain of mitochondrial import receptor Tom20 was used to direct the construct to be displayed on the mitochondria (**Figure 3d**).

Given that TFs localize to the nucleus to bind their DNA targets, we anticipated that the LiNC bound TF would facilitate target gene expression in a manner that could be downregulated through NS3 inhibition. We confirmed inhibitable genetic control of these

constructs using flow cytometry for the activation of a UAS driven citrine reporter (**Figure 3e**). This dynamic control of protein activity through control of domain localization was inspired by the activity of the cell signaling protein Notch.

The protein Notch is an important cell-cell communication receptor that is implicated in a number of developmental and essential processes in mammalian biology [113]. When it is activated, it unfolds to reveal a protease cut site, allowing a cytosolic TF to float free and make its way into the nucleus of the cell to regulate gene expression [114]. We anticipated that we could create a facsimile of this activity in a drug dependent way by using a drug stabilized linker. This creates a “turn-off” system whereby the addition of drug turns off the gene regulation via the localization of the TF, which is bound to the outer membrane and therefore cannot move into the nucleus to interact with genetic promoter regions.

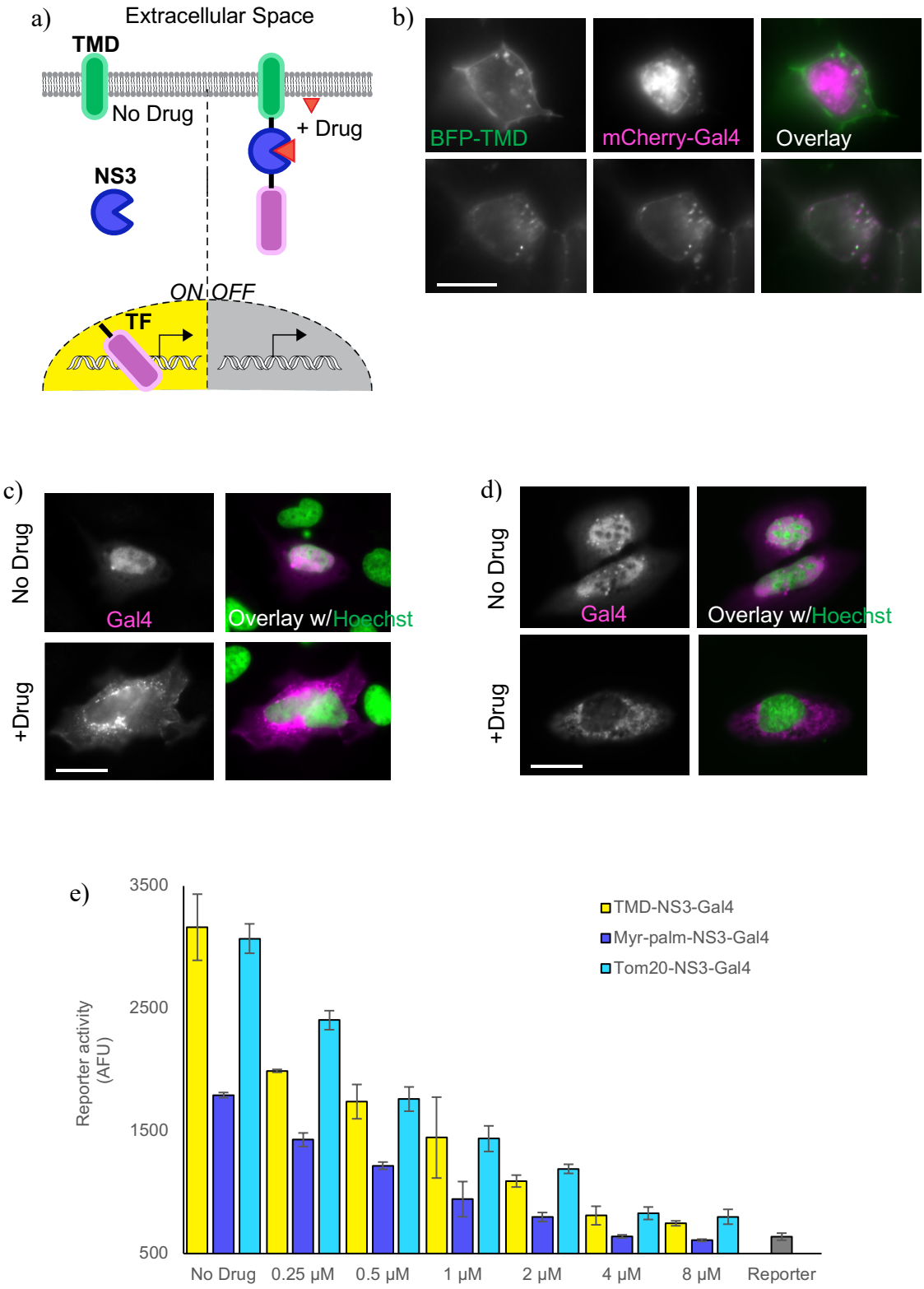

a)
Extracellular Space
TMD
No Drug
+ Drug
NS3
ON/OFF
TF
b)
BFP-TMD
mCherry-Gal4
Overlay
c)
No Drug
+Drug
Gal4
Overlay w/Hoechst
d)
No Drug
+Drug
Gal4
Overlay w/Hoechst
e)
Reporter activity (AFU)
3500
2500
1500
500
TMD-NS3-Gal4
Myr-palm-NS3-Gal4
Tom20-NS3-Gal4
No Drug
0.25 μM
0.5 μM
1 μM
2 μM
4 μM
8 μM
Reporter

Figure 3: Construction of a "Turn-off" TF system
a) General design of the "Turn-off" TFs containing a generic membrane-localizing element (green). **b)** Fluorescence images of HEK293A cells transiently expressing BFP-TMD-NS3-Gal4-mCherry. BFP is N-terminal to the TMD; mCherry is between the DB and TA of Gal4. Scale bar: 10 uM. **c)** Immunostained HeLa cells showing drug-induced membrane targeting of $Gal4_{min}$ via myrpalm-NS3-$Gal4_{min}$. 3 uM BILN-2061 was used. Gal4 was stained with Santa Cruz sc-577 (polyclonal rabbit anti-Gal4) and goat anti-rabbit CF647 (Sigma Aldrich; SAB4600184). Scale bar: 25 uM. **d)** Immunostained HeLa cells showing drug-induced mitochondrial targeting of $Gal4_{min}$ via Tom20-NS3-$Gal4_{min}$. Reagents used are same as previous. **e)** The drug-induced downregulation of the UAS H2B-Citrine construct via TMD-NS3-$Gal4_{min}$, myr-palm-NS3-$Gal4_{min}$, and Tom20-NS3-$Gal4_{min}$ was monitored in transfected HEK 293FT reporter cells treated with grazoprevir at the indicated concentrations. Geometric means are displayed as mean ± s.d., as determined by three biologically independent samples.

2.3.2: Methods

Standard Gibson and restriction cloning procedures were used in the generation of all DNA constructs. The citrine reporter plasmid was a gift from Michael Elowitz (Caltech) and the mTagBFP sequence was a gift from Wilson Wong (Boston University). The MP sequence and Tom20 sequences were taken from native sequences and humanized using IDT's expression optimization tools, then synthesized as a gene fragment (Integrated DNA Technology). MP sequence: ATGGGCTGCATCAAGAG CAAGCGCAAGGACAACCTGAACGACGACGGCGTGGACATGAAG. Tom20 sequence: ATGGGCTGCATCAAGAGCAAGCGCAAGGACAACCTGAACGACG ACGGCGTGGACATGAAG. Plasmid cloning accuracy was confirmed using sanger sequencing (Quitara Biosciences). Asunaprevir and MK-5172 (grazoprevir) were from MedChemExpress. BILN-2061 was a gift from Roger Tsien and Stephen Adams (University of California, San Diego). Concentrated NS3 inhibitor stocks were dissolved in DMSO at concentrations between 3 and 10 mM and diluted into cell culture media at the indicated working concentrations (and no more than 1%). We compare selectivity and cytotoxicity of existing NS3 inhibitors in **Table 1**.

All mammalian cell lines were cultured using standard techniques at 37 ºC with 5% CO_2. HEK293FT cells (ThermoFisher) were cultured in DMEM with 10% FBS and supplemented with nonessential amino acids (Life Technologies), Glutamax (Life Technologies), and G418 (500 μg/ml; Invivogen). HEK293A and HeLa cells were obtained from ATCC and were cultured in DMEM containing 10% FBS and supplemented with Glutamax and penicillin–streptomycin. DNA transfections were

carried out with Lipofectamine 3000 reagent (ThermoFisher) according to the manufacturer's instructions.

Stable cell lines were generated by transfecting with linearized DNA encoding the engineered protein of interest and an antibiotic resistance gene for mammalian selection. Transfections were performed on plates grown to nearly confluent. 48 h post transfection, cells were split and selected via antibiotics. After the elimination of nontransfected control cells (typically 1–2 weeks post addition of antibiotic), surviving cells were transferred to a 96 well plate using a limited dilution procedure to isolate single clones [115].

Cell lysates used in immunoblotting analyses were prepared by lysis of cells in 1 X LDS-PAGE loading buffer (ThermoFisher) after the removal of cell culture media. For drug removal studies, this was performed directly after removal of media in order to denature proteins efficiently, preventing undesired NS3 cis-cleavage in cell lysates. The cell lysates were clarified through sonication followed by centrifugation. The lysates were subsequently analyzed by standard immunoblotting procedures and probed using the antibodies listed following at the indicated dilutions. Detection of labeled antigens was via chemiluminescence via the SuperSignal West Pico PLUS Chemiluminescent Substrate (Pierce).

Primary antibodies used: rabbit anti-histone H2B (Cell Signalling; 12364; 1:1000 dilution for western blotting) mouse anti-HA-HRP (Santa Cruz Biotechnology; sc-7392; 1:1000 dilution for blotting), rabbit anti-GAPDH (Sigma-Aldrich; G9545; 1:3000 dilution for western blotting) and polyclonal rabbit anti-Gal4 (Santa Cruz Biotechnology;

sc-577; diluted 1:500 for western blotting, 1:200 for immunostaining). The following secondary antibodies were used: goat anti-rabbit CF647 conjugate (Sigma-Aldrich; SAB4600184; 1:300 dilution), anti-mouse HRP conjugate (Cell Signalling; 7076; 1:3000 dilution), and anti-rabbit HRP conjugate (BioRad; 170-6515; 1:3000 dilution).

Immunofluorescence staining was performed with cells rinsed with PBS before fixation with formaldehyde (4% v/v diluted into PBS from fresh vials containing 16% solutions (ThermoFisher)). Cells were fixed for 10 minutes at room temperature and then rinsed with PBS (3X) to remove residual fixative. When necessary, cells were permeabilized with Triton-X 100 (0.2% v/v in PBS) for 10 minutes and subsequently rinsed with PBS. Cells were blocked with BSA solution (5% v/v in PBS) for approximately 30 minutes at room temperature (RT) before being stained with primary antibody solution (diluted in PBS) at the dilutions indicated above for 1 h at RT. Cells were stained with secondary antibody solution (in PBS at the dilutions indicated above) for 1 hour at RT before imaging.

Images were taken by epifluorescence microscopy in imaging compatible vessels containing glass coverslip bottoms (MaTek) or optically clear plastic bottoms (ibidi). These vessels were coated with bovine plasma fibronectin (Product #F1141, Sigma-Aldrich) and/or treated for cell adherence by the manufacturer (poly-D-Lysine by MatTek or ibiTreat by ibidi). During imaging. Cells were maintained in PBS, standard culture media, or FluoroBrite DMEM (ThermoFisher). For time-lapse images, cells were imaged in culture media supplemented with 30 mM HEPES diluted from 1 M stock (pH 7.2-7.5; ThermoFisher) and maintained at 37 ºC in a heated imaging chamber throughout

the duration of the time-lapse. Images were acquired with ZEN imaging software (Zeiss). Image files were processed with the ImageJ-based image analysis package Fiji. The images were contrasted uniformly across experiments.

Cells analyzed by flow cytometry were gated for living cells by scatter detection. The geometric mean measured reporter fluorescence levels were reported in arbitrary fluorescence units (AFU). Reporter activation analyses were carried out with stable single clones or cells transiently transfected with DNA encoding the analyzed TF (as indicated in the figure captions). For analyses carried out with transiently expressing cells, plasmid DNA encoding a constitutively expressed fluorescent protein marker was co-delivered at the time of transfection and used to identify positively transfected cell populations. Transfected cells were gated to the top 1% of marker fluorescence of nontransfected control cells under the same condition. Transient expression experiments carried out with the turn-on TFs were gated via detection of an mCherry marker that was expressed via an IRES sequence on the TF-encoding plasmid. Transfected cells were incubated for 24–48 h after transfection in either the presence or the absence of the indicated NS3 inhibitors before being analyzed with an Attune NxT flow cytometer (ThermoFisher).

All flow cytometry data was collected using three biologically independent samples. For fluorescence imaging analyses and immunoblotting, 3 or more images per condition were recorded and representative images or blots were displayed. These methods and reagents are applicable to all following experimental data unless stated otherwise.

Table 1: Table of NS3 Inhibitors. CC_{50} is the reported drug concentrations at which 50% cytotoxicity was observed.

Inhibitor	CC_{50} (μM)	Cells Tested	CC_{50} Method	Selectivity
Asunaprevir (BMS-650032)	11-38 [116]	MT-2, HEK293, HuH-7, HepG2, and MRC5 cell lines	MTS assay for MT-2 cell line; mitochondrial activity assay (resazurin) for all other lines	"IC50 > 50 mM for PPE, human chymotrypsin, human cathepsin B; IC50 > 5 mM human cathepsin A; IC50 > 30 mM trypsin, thrombin, factor VIIa, factor Xa, factor Xia, kallikrein." [116]
BILN-2061 (Ciluprevir)	33 [117]	HuH-7 cell lines	MTT assay	"The inhibition of BILN 2061 was highly specific to the NS3 protease as demonstrated by the lack of significant activity (half-maximal inhibitory concentration (IC50) > 30 μM) against human leukocyte elastase and human liver cathepsin B… A mean 50% cytotoxic concentration (CC50) of 33 μM was observed for BILN 2061, resulting in an apparent selectivity index of 10,000 in Huh-7 cells when compared with the EC50 value obtained for inhibition of subgenomic HCV RNA replication." [117]
Danoprevir (ITMN-191)	75-340 [118]	HuH-7 cell line and primary cultures of human hepatocytes, microvascular endothelial cells, human skeletal muscle myoblasts, human cardiac fibroblasts, human articular chondrocytes, human lung fibroblasts, and renal proximal tubule epithelial cells	ATP detection	"In contrast to the highly potent inhibition of NS3/4A by ITMN-191, none of a panel of 53 proteases was inhibited more than 50% by a 10 μM screening concentration, indicating an IC50 higher than 10 μM against every protease in the panel (Table 2). Ciluprevir inhibited eight proteases and telaprevir inhibited nine proteases in the same panel at levels between 50% and 100%, which indicated that their IC50s against these proteases were 10 μM or less ... Neither ITMN-191 nor telaprevir showed appreciable activity against a broad panel of ion channels, receptors, and transporters, while ciluprevir inhibited human [potassium channel] ERG ... From these data, the specificity index of ITMN-191, defined as the ratio of IC50s against nontarget enzymes (>10 μM) to its IC50 against full-length genotype 1b NS3/4A reference protein (K2040) (0.29 nM) ... was more than 35,000-fold." [118]
Grazoprevir (MK-5172)	>50 [119]	HB1 and JFH replicon lines derived from the HUH-7 cell line	MTS assay	"Grazoprevir (GZR, MK-5172) exhibited >25,000-fold selectivity with respect to HCV protease (GT1, 4 and 6) versus human serine protease inhibition, as

				evaluated with elastase and trypsin (IC50>100 mM) and chymotrypsin (IC50~1.5 μM). In addition, GZR exhibited similarly high selectivity with respect to enzyme and receptor binding of unintended targets with binding activity (IC50 <10 μM) only observed to matrix metalloproteinase-1 & 12 (IC50 ~1.5 & 6.9 μM), lipoxygenase 5-LO (IC50 ~2.8 μM), prostanoid FP (IC50 ~6.5 μM) and hERG (IC50 ~3.3 μM). Clinical significance of this unintended binding/activity of EBR and/or GZR appears unlikely." [120]
Glecaprevir (ABT-493)	>59 [121]	HUH-7, HepG2, and MT-4 human lymphoid T cell line	MTT assay	"When glecaprevir was tested against six human serine proteases (chymase, chymotrypsin type II, chymotrypsin type VII, elastase, kallikrein, and urokinase) and one human cysteine protease (cathepsin B), no inhibition was observed at concentrations up to 200,000 nM. These results indicate that glecaprevir demonstrates a high level of selectivity for the HCV NS3/4A protease over the human proteases tested." [121]

2.3.3: Reversibility

We recognize that the reversibility of this system is complex. We cannot reverse cleavage of a protein tether or draw the TF back to the membrane with this system. Given a system at equilibrium without drug, we found that exposure to NS3 inhibitors led to the accumulation of membrane-tethered Gal4min (**Figure 4a, b**), as well as the gradual depletion of previously cleaved TF copies (**Figure 4c**), suggesting that, while we cannot reverse cleavage of the protein, we can reverse the activity of the targeted gene. We found that the depletion rate of the freed domain could be accelerated through the addition of a small peptide degron on the TF domain called a PEST (rich in proline, glutamic acid, serine and threonine), which we analyzed via western blot (**Figure 4d**). The PEST domain used was derived from the C-terminal region of mouse ornithine decarboxylase, which has previously been used to generate a "destabilized" version of GFP with a reduced half-life of 2 hours [122,123]. We see that the intensity of the full construct bands grew over time, indicating accumulation of the intact proteins following NS3 inhibition. Bands corresponding to cleaved Gal4min and Gal4min-PEST were also observed, the intensities of which diminished over time. The half-life of the Gal4min-PEST was attenuated relative to that of Gal4min from a half-life of between 4 and 8 hours to a half-life of less than 2 hours.

Live-cell imaging showed that the effect of downregulation and depletion of free TF could be reversed after inhibitor withdrawal (**Figure 5a**). We anticipated that this would be a rapid reversal as the depletion of inhibitor was expected to lead to the inhibitor unbinding from the protease, which would activate to cleave and release the TF.

We observed different speeds of reversal with removal of different inhibitors corresponding to the affinity of interaction with the protease (**Figure 5b**). We see faster production of reporter from the removal of danoprevir than grazoprevir. This is somewhat unexpected considering danoprevir's known slow dissociation rate of 2.1×10^{-5} per second, in contrast to the hundred-fold faster dissociation rates of other NS3 inhibitors [124]. However, there is a seven-fold difference in affinity between the two drugs; danoprevir has a K_i of 1.0 nM, while grazoprevir has a K_i of 0.14 nM [125]. From this information, we can hypothesize that the greater contribution to this increase in signal is production of new non-inhibited NS3 proteins that are less likely to become inhibited when in low concentrations of danoprevir as the drug gets depleted from the system. Together, these results indicate that TFs can be conditionally linked to localization signals to permit precise control over their subcellular distributions, leading to control over their activities, and this control scheme is reversible.

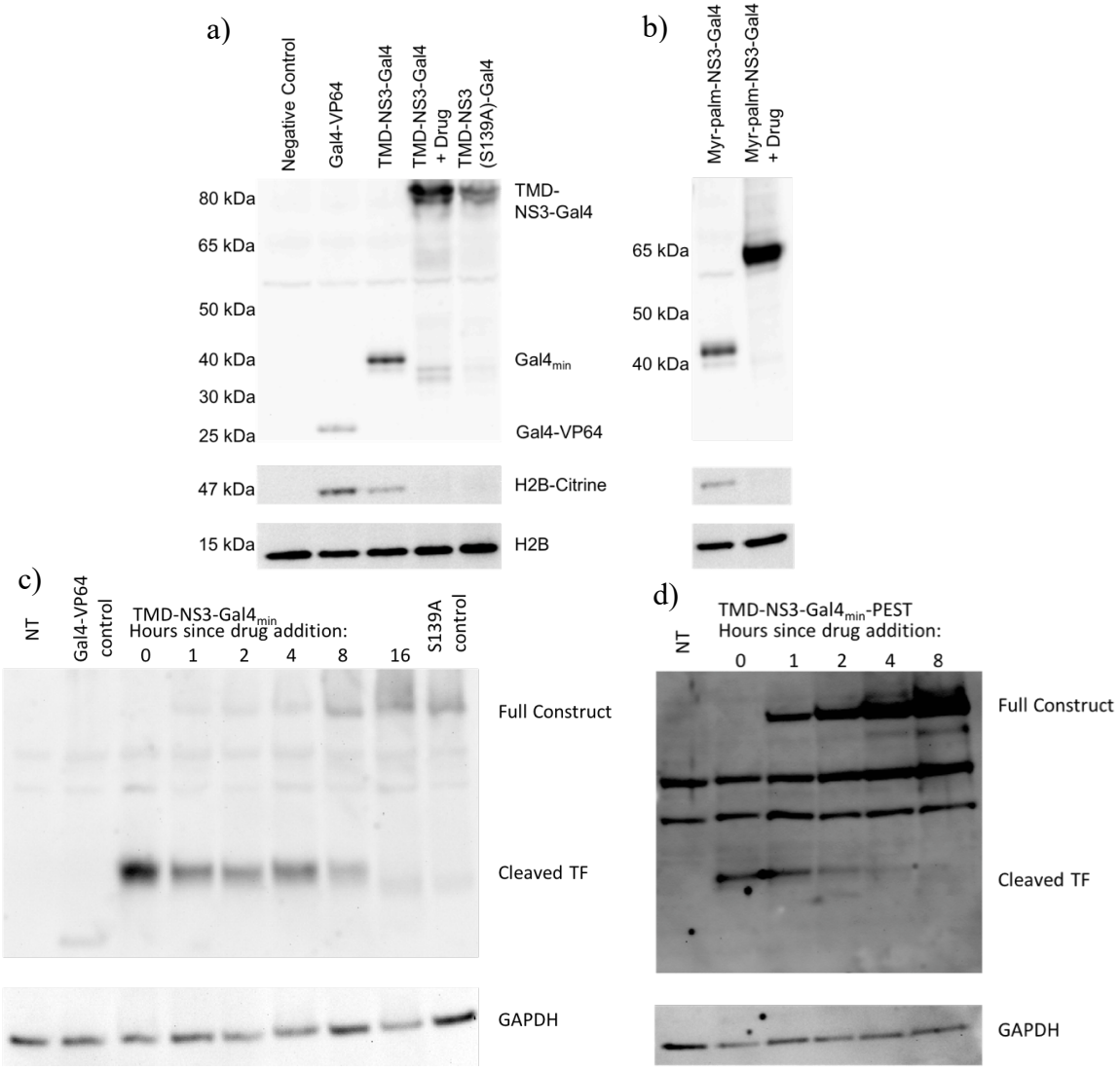


Figure 4: Reversibility of the "Turn-off" TF

Western blotting using an antibody against the Gal4 DB domain (Santa Cruz sc-577, polyclonal rabbit) was used to confirm the drug-induced preservation of intact version of **(a)** TMD-NS3-$Gal4_{min}$ (96.3 kDa) and **(b)** myr-palm-NS3-$Gal4_{min}$ (62.9 kDa) in cells treated with BILN-2061 (3 uM). In the absence of drug, bands corresponding to cleaved $Gal4_{min}$ (29.2 kDa) domains were observed. DNA encoding the TF construct were transfected into HEK 293FT cells containing a stably integrated Gal4-dependent reporter gene (UAS H2B-Citrine). H2B-Citrine (45.5 kDa) was detected via an anti-H2B antibody (anti-H2B, Cell Signalling 12364) and expression of the reporter was observed only in the lysates of drug-untreated cells. Lysates from cells expressing either Gal4-VP64 TF (23.3 kDa) or a proteolytically inactivated mutant of TMD-NS3-$Gal4_{min}$ via a S139A were used as controls. HEK 293FT cells transfected with DNA encoding **(c)** TMD-NS3-$Gal4_{min}$ or **(d)** TMD-NS3-$Gal4_{min}$-PEST were grown without inhibitor until treatment with 3 uM grazoprevir at the indicated times prior to lysis in SDS PAGE loading buffer and western blot. Gapdh was used as a loading control (rabbit anti-Gapdh; Sigma Aldrich G9545).

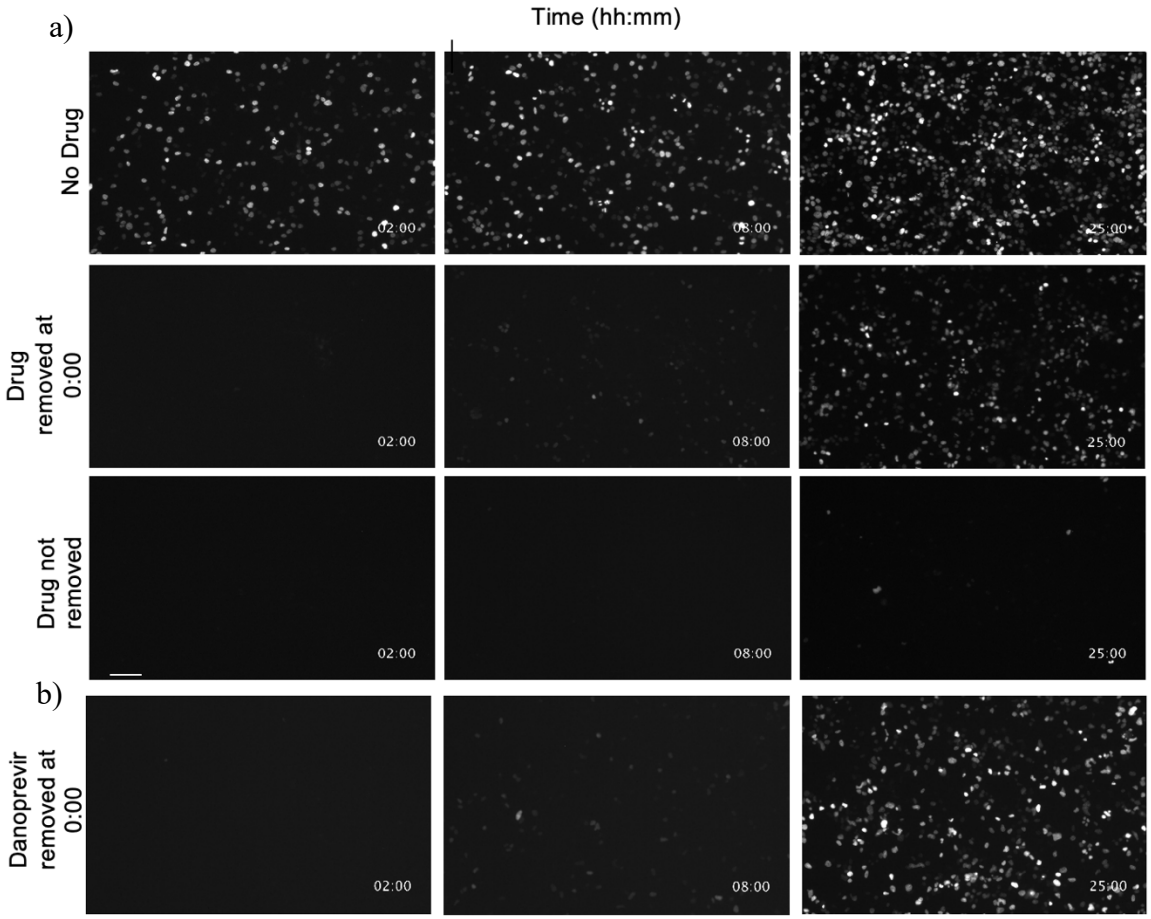


Figure 5: Live-cell imaging of "Turn-off" system reversal
Time lapse fluorescent imaging of HEK 293FT reporter cells (UAS H2B-mCherry) transiently transfected with DNA encoding TMD-NS3-Gal4$_{min}$. Cells were treated with either **(a)** grazoprevir or **(b)** danoprevir at 3 uM immediately at the time of transfection. Live cell imaging was initiated at 24 hours post transfection (t=0), at which point cells were either maintained in drug, or released from NS3 inhibition by exchange into drug-free media. Transfected cells that were not exposed to NS3 inhibitors were used as a control. It can be seen that danoprevir removal triggers a faster release of TF by measure of fluorescence. Scale bar is 100 uM.

2.4 Combined "Turn-off" and "Turn-on" system in a single cell

Recognizing that natural gene expression often involves the synchronized control of multiple target genes, we next combined the 'turn-on' and 'turn-off' systems to create a platform for simultaneous regulation of distinct promoters by a single drug (**Figure 6a**). In cells co-expressing both rTetR-NS3-VP64-p65 and TMD-NS3-Gal4min, we observed coinciding and inverse regulation of TRE- and UAS-controlled reporter constructs (**Figure 6b**). We see that not only can we regulate one gene individually, but the system is robust to regulating multiple genes in tandem. Notably, 'leaky' activation by each TF appeared to be low, comparable to that of control cells containing mismatched reporter–TF pairs as examined by flow cytometry (**Figure 6c**). Taken together, these results indicate that multiple TFs can be controlled via LInC to achieve concurrent and opposing transcriptional changes in response to NS3 inhibition. For these analyses which combine rTetR-NS3-VP64-p65 and TMD-NS3-Gal4min, 125,000 HEK293FT cells were transfected with 25 ng of DNA with DNA mixtures containing a 3:3:2:2 molar ratio of rTetR-NS3-VP64-p65 to TMD-NS3-Gal4 to TRE mTagBFP to UAS H2B-Citrine (as approximated by DNA size).

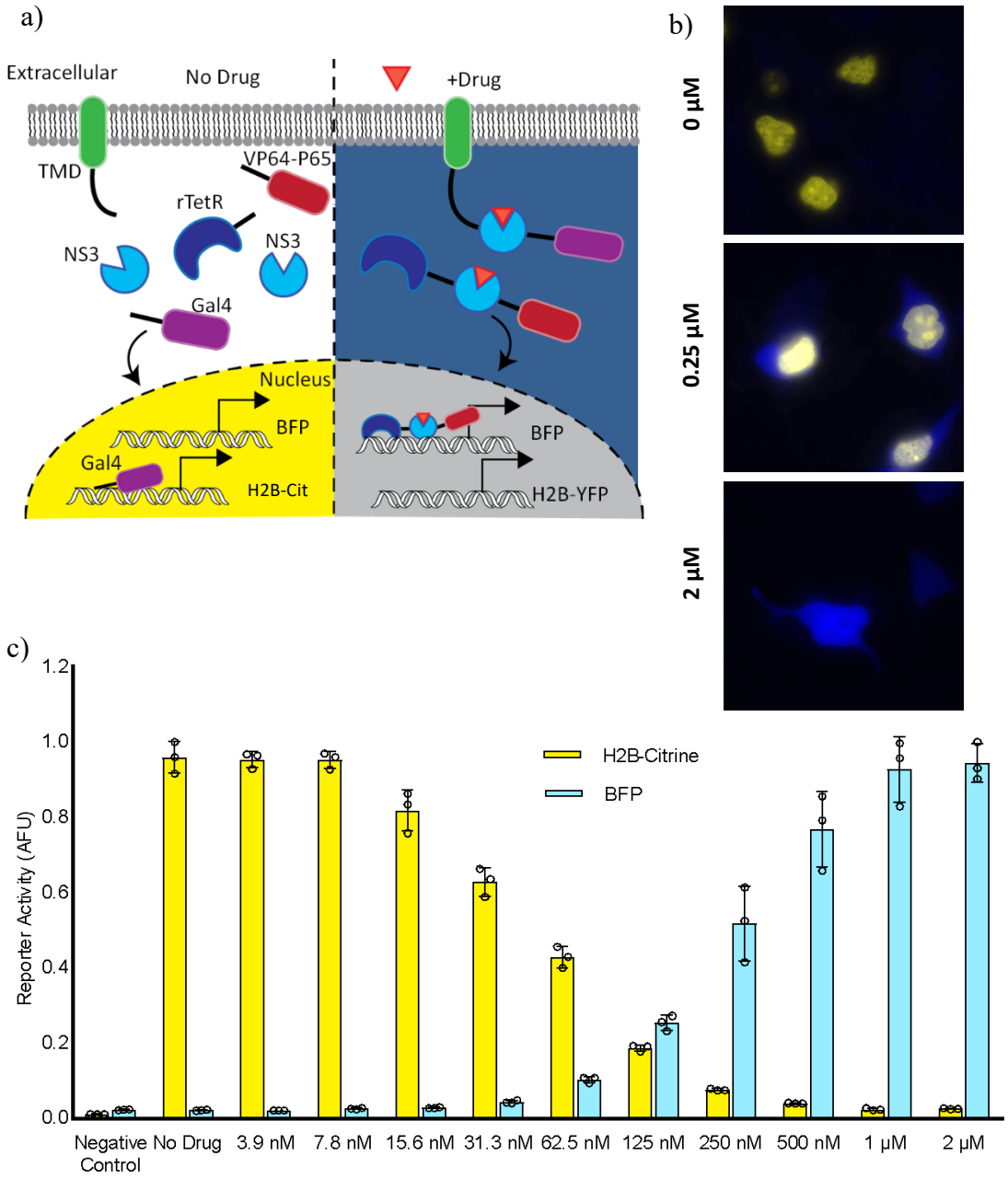


Figure 6: Co-expression of TMD-NS3-Gal4$_{min}$ and rTetR-NS3-VP64-p65
a) Schematic of co-expressed "Turn-on" and "Turn-off" systems. b) Fluorescence imaging of HEK 293FT reporter cells transiently transfected with TMD-NS3-Gal4$_{min}$ and rTetR-NS3-VP64-p65 as well as UAS-H2B-Citrine and TRE-BFP. Images taken 24 hours after addition of BILN-2061. 200ng/mL of dox was applied to all samples to turn rTetR on. **c)** Fluorescence emission from HEK293FT cells transiently transfected with the two constructs and two reporters from **(b)**. Cells were treated with 100 ng/mL doxycycline and denoted concentrations of BILN-2061. Negative controls correspond to background activation of each reporter in cells containing mismatched TF. Values represent geometric mean fluorescence ± s.d.; n = 3 biologically independent samples based on geometric mean fluorescence determined by flow cytometry.

2.5 Discussion

The applications described above show that the LInC strategy can be used to engineer drug sensitivity to control both gene activity and protein localization through the use of the destabilized protein as a dissolvable linker. We have demonstrated that this strategy is modular with respect to membrane linker. This strategy is applicable to any gene with an appropriately TF-binding dependent promoter and we additionally anticipate that the tethering scheme can work with any protein, a strategy that will be employed in further chapters.

A significant advantage of the method is the availability of highly selective NS3 inhibitors, of which some are FDA approved and this continues to be an area of active development [126, 127]. Finally, in addition to its potential applications in basic biology investigations, the LInC approach may also serve as a powerful strategy for regulating therapeutic cells in vivo with safe and clinically approved drugs.

Other groups have additionally made important advancements with NS3 as a gene activator [128], circuit control element [129], part of a protein dimerizer system [125], and dynamic controller of protein display [130] indicating that this continues to be an active area of research. NS3 has joined FRB/FKBP and TEV protease as a modular and easy to implement protein engineering domain, and this has expanded the invention space for designing protein-based modulators and control systems [131,132].

Chapter 3: Inteins as a mechanism of dynamic control of protein function in mammalian cells

3.1 Inteins in their native context and synthetic applications

Inteins are small peptide splicing elements, following the naming conventions of introns and exons, which mediate the change in coding sequence from DNA to RNA [133]. However, where introns change the coding sequence, inteins mediate the final form of a protein by post translational sequence changes via excision and splicing of amino acid chains. They have been labeled “selfish” gene elements after the pattern of transposons and may have originated as parasitic gene elements, yet many have been selected for host advantage and make useful contributions to host fitness [134].

Inteins were discovered in 1987 by a group comparing the sequence of Neurospora vacuolar ATPase to plant ATPases; they found an exonuclease domain that was not conserved among other known ATPases [135]. These exonuclease domains were matched with yeast ATPase sequences and found to be removed upon translation, forming two protein components when run on an SDS-PAGE gel [136,137]. This protein contained the first of many single chain inteins to be discovered. Single chain inteins have all components encoded in a single amino acid chain that processes itself into two proteins: the intein concatenates the extein domains with a native peptide bond, and the inner piece, the intein itself, often has other protein components included, such as kinase activities or domains that regulate transcription (**Figure 7a**) [91]. Inteins also exist as split inteins, where the two protein extein components are translated as two different protein chains and get concatenated post translationally (**Figure 7b**). These two naturally

occurring types of inteins have found their niche in biotechnological applications. Single chain inteins can regulate the equimolar production of two proteins for metabolic processes with sensitive kinetics or for accurate quantification of protein production [138,139]. Split inteins have been used to overcome the size constraints of gene delivery mechanisms by splitting up large therapeutic proteins, as well as driving multiplexed gene expression by allowing resistance genes to be split among multiple gene expression cassettes [140-144].

Inteins are viewed as powerful biological tools due to their autonomy; they require neither cofactors nor external enzymes and have naturalized themselves to a variety of biological environments, including varied temperature and pH [145-147]. Intein splicing is based on neutrophilic attack by the C-terminal residue against a linear ester intermediate formed by a neutrophilic N-terminal residue adjacent to the target splice site [148]. The key residues are the neutrophilic N and C terminus of the inteins, usually cysteine but sometimes serine or threonine, and this reaction is stabilized by surrounding intein and extein residues [134]. Functionally, this means that many inteins will require a specific amino acid sequence similar to their native context in order to splice, however a number of inteins have been found that are promiscuous in splicing activity, either by nature or design [149-151]. This enables nigh scarless splicing, with only a residual cysteine or serine residue at the splice site [152,153].

All of these factors draw our interest in light of making protein tools that are orthogonal, fast, precise, and are easy to implement in varied contexts. Inteins are naturally orthogonal to each other, to the point that even inteins within the same family

can be unable to splice each other [154]. There is a large body of fast and precise inteins to draw from, many of which are derived from archaea, synthetic, and even cryptic sources, making them prime targets to be employed in therapeutics and technologies [155]. They have been employed safely in mammalian systems [141,156,157]. Intein autonomy makes them low-impact and simple to add into a system, and in varied contexts the small extein sequences can be added or scarless inteins can be used to create a system whereby our protein tool can be concatenated to novel proteins with very little sequence change. This makes inteins plug-and-play protein design friendly, and we make use of these factors in the construction of dynamic protein tools.

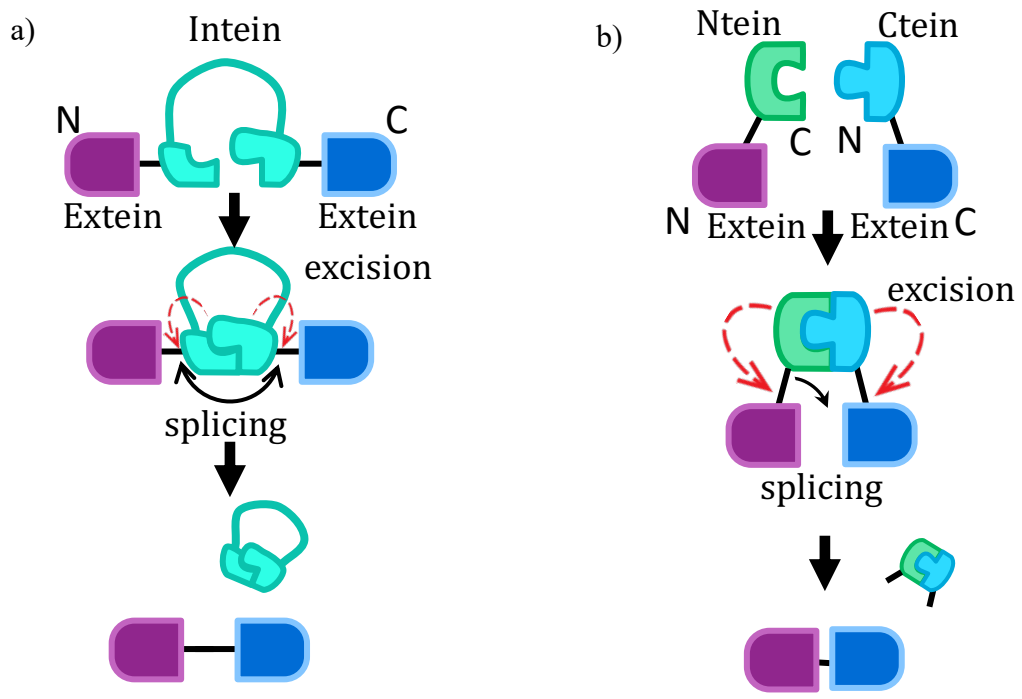


Figure 7: Single chain and split inteins
Generalized intein activity in both **(a)** single chain conformation and **(b)** naturally split. A native peptide bond is formed with no necessary co-factors or ATP.

3.2: Control of protein activity and degradation using an intein-circularization scheme

3.2.1: Background and Motivation

Control of protein degradation has emerged as a powerful and sensitive strategy for control of protein activity. Tools for regulating DNA and RNA expression of proteins are important and valuable, but ultimately the phenotype of a cell is defined by the active proteins in the system. Proteins that are constructed and then degraded do not have a significant impact on the cellular environment. A native example of this is PINK1 in the mitochondria, which is continually constructed and degraded; it is only when this process breaks down that the PINK1 kinase is active and signals the destruction of the mitochondria [158]. A number of protein regulation strategies have been developed to take advantage of this. A common design strategy is to add a small peptide degron tag or otherwise target a protein of interest for destruction rapidly upon synthesis in a dynamic manner.

One example is the commercialized Shield-1 system discovered by the Wandless lab [159]. A destabilized version of FKBP12 protein targets itself and any attached protein components to be rapidly degraded. Shield-1 is a small molecule, membrane permeable drug that stabilizes the FKBP12 domain, thereby preserving the attached protein. Systems of this sort are fast and reversible through the continuous production and degradation of protein. Proteins are stabilized immediately upon interaction of the small molecule drug with the tag, driving up concentration of protein at the time of drug introduction. Another common degradation platform is through the N-end rule.

Mammalian cells regulate protein lifespan through the amino acid residues at the N terminus, which are acylated and ubiquitinated [160]. Steric shielding of the N terminal residues in a dynamic manner such as with the Cog-1 and Cog-X shield protein described by the Varshavsky lab is another method of controlling which proteins are targeted by endogenous systems for degradation [161].

We wished to take advantage of the tight control found in rapid protein clearance in conjunction with the uncommon stability of circularized proteins. Circularized proteins exist in nature, especially in bacteria and plants, for proteins that need particular stability like bacteriocins, which are biological attacks amongst bacteria, and plant self-defense peptides [162,163]. They have also been a popular platform for developing stable peptides for library development and, more recently, bioactive peptides for therapies due to their stability at high temperatures and resistance to proteolytic degradation [164,165].

Inteins are a common strategy for creating cyclized peptides, as the splicing machinery of a split intein lends itself well to the formation of a bond between the N-terminus and C-terminus of a protein. The SICLOPPS (split intein circular ligation of peptides and proteins) method takes advantage of split inteins to construct stable libraries of peptides that can be used in a number of manners, including screening for peptide inhibitors of enzymatic reactions [166]. In an effort to control the numbers of peptides for protection of their producer cells and the quality of the peptides produced, the Tavassoli group added a C-terminal degron to their constructs [152]. This targeted their proteins for degradation until the split intein was able to splice the protein into a cyclized peptide, effectively getting rid of any proteins missing active splicing components and

increasing the overall quality of the preserved peptides. This system was also found to have a significant reductive effect on toxicity, making it an important addition to the library production scheme.

Here we sought to make a similar system where a degron would be under intein control. However, in this case we wished to use intein cyclization to occlude the degron such that the protein would be transiently stable, allowing the degron to first delineate protein quality, and secondly drive protein clearance upon input of a signal. This would create a switch-like system whereby the target protein would be persistent within the cell until a signal leads to swift clearance through degradation.

3.2.2: Design Strategy

We chose a fluorescent protein as the target protein in our system for ease of identification and quantification. Our design process followed the SICLOPPS method in placing the intein halves at each terminus of the protein but we decided to use the split intein gp41-1, due to it being, at the time, one of the fastest known naturally split inteins [69]. This intein was first identified in a metagenomic sample of ocean water, as a component of the gp41 helicase gene of an unknown source [167]. Gp41-1 is a small intein, 9.8 kDa for the N terminus intein half (Ntein) and 4.9 kDa for the C terminus component (Ctein), and this taken in conjunction with its fast splicing rate, has made it a target for researchers trying to determine the essential components of an active intein [168]. It has been used as the template for taking only the necessary structures from naturally single-chain inteins and converting them into split inteins [169]. The structure of this protein construct can be seen in **Figure 8a**.

We first wished to confirm gp41-1 activity in mammalian cells. We concatenated the halves of a minimal Gal4 transcription factor (TF) to each half of the intein, separating the activating domain of the TF from the DNA binding domain of the TF (**Figure 8b**). In conditions where the intein is active, the halves should electrostatically attract each other and mediate a splicing reaction between their exteins forming an active TF with both a DNA binding domain and an activating domain. We used a UAS driven fluorescent reporter, H2B-Citrine, as a measure of active TF reconjugation and measured by flow cytometry intein activity together with activating domains of different strength (**Figure 8c**). The TF has activity based on the strength of the activating domain when both halves are expressed, and this activity can be turned off when a non-splicing C1A mutation is made to the Ntein [92].

To make a splicing dependent degron, we placed a ubiquitin-independent degron between the C-terminus of the protein and the C-terminal Ntein (**Figure 8d**). We use mouse ornithine decarboxylase C-terminus (residues 422-461), also called a PEST domain (proline, glutamic acid, serine, and threonine rich polypeptide region), as our degron [122]. This PEST tag has also been successfully implemented with GFP to make the two-hour half-life GFP for studies of transient phenomena, and has been shown to be ubiquitin independent in action [170-173]. Ubiquitin independent proteosome degradation occurs when two design constraints are met: a) specific association with the proteosome and 2) a destabilized domain [174,175]. The mouse ornithine PEST meets these conditions, however interestingly, split inteins are also destabilized prior to association with their partner and therefore should contribute to the degradation scheme

[176]. We hypothesized that the steric constraints caused by cyclization would decrease the disorder of the PEST such that as a cyclized protein it would no longer be degraded, creating a semi-stable protein only when successful splicing could occur.

Additionally, we placed a cut site for the well characterized protease from tobacco etch virus (TEV-P) between the GFP and the C-terminal Ntein [177]. TEV-P is popular for synthetic applications due to the specificity of the protease target, and is commonly employed to remove purification tags or otherwise process recombinant proteins [178-180]. We anticipated that protease cleavage would convert the stable GFP to an unstabilized phenotype.

3.2.3: Additional Methods

Please see Section 2.3.2 for basic methods used in these assays. Gp41-1 sequence was published in Wang, et al, 2018 [181], humanized using IDT mammalian optimization protocol (Integrated DNA Technologies), and synthesized as a long gene fragment (Integrated DNA Technologies). The final sequence was Ntein: TGTCTGGACCTGAAGACTCAGGTCCAGACACCTCAGGGAATGAAGGAGATTTCCAATATACAAGTAGGAGACTTGGTGTTGAGCAACACTGGCTATAACGAAGTTCTCAACGTATTTCCCAAATCAAAGAAAAAGTCCTATAAAATTACACTTGAAGACGGCAAGGAAATAATATGCTCCGAGGAGCACTTGTTCCCAACCCAAACTGGTGAGATGAACATTAGTGGTGGGTTGAAGGAGGGAATGTGTTTGTATGTCAAAGAG and Ctein: ATGATGCTGAAGAAGA TTCTGAAAATAGAAGAGCTGGATGAACGGGAGCTGATCGACATCGAGGTCTCAGGGAACCACCTGTTCTACGCGAACGATATACTTACGCACAACTCA. For best

splicing the contextual SSDV amino acids were added directly after the active serine on the Ctein. Cloning was performed using standard Gibson cloning methods. The TEV protease was a gift from Xiaokun Shu (Addgene plasmid #64276) and the protease cutsite used was ENLYFQG [182,177]. The citrine reporter and Gal4 as well as VP64 and p65 domains were the same as reported in Chapter 2, and were a gift from Michael Elowitz (Caltech). The alpha-helix transmembrane domain was also the same used in Chapter 2.

HEK293FT cells (ThermoFisher) were transiently transfected using Lipofectamine 3000 reagent (ThermoFisher) at a ratio of 50 ng total DNA per 500,000 cells unless reported otherwise. Cells were plated 125,000 cells per 96 well. Cells were transfected in suspension, plated and incubated for 24 hours before analysis.

Primary antibodies used: anti-GFP rabbit polyclonal (Sigma Aldrich G1544, 1:200 for immunofluorescence, 1:3000 dilution for western blot) and anti-Gapdh mouse (ThermoFisher MA5-15738, 1:3000 dilution). Secondary antibodies used: goat anti-rabbit AlexFluor 647 (Sigma Aldrich SAB4600184, 1:200 dilution), anti-rabbit HRP (BioRad 170-6515, 1:3000 dilution), and anti-mouse HRP conjugate (Cell Signalling; 7076; 1:3000 dilution).

Fluorescence images were taken in FluoroBrite DMEM (ThermoFisher) with cells seeded 120,000 cells per well of 8-well optically clear plastic plates with ibiTreat (ibidi). Images were taken using a GFP or Cy5 filter cube, as appropriate. The image files were processed using ImageJ image analysis package Fiji and were contrasted uniformly across experiments.

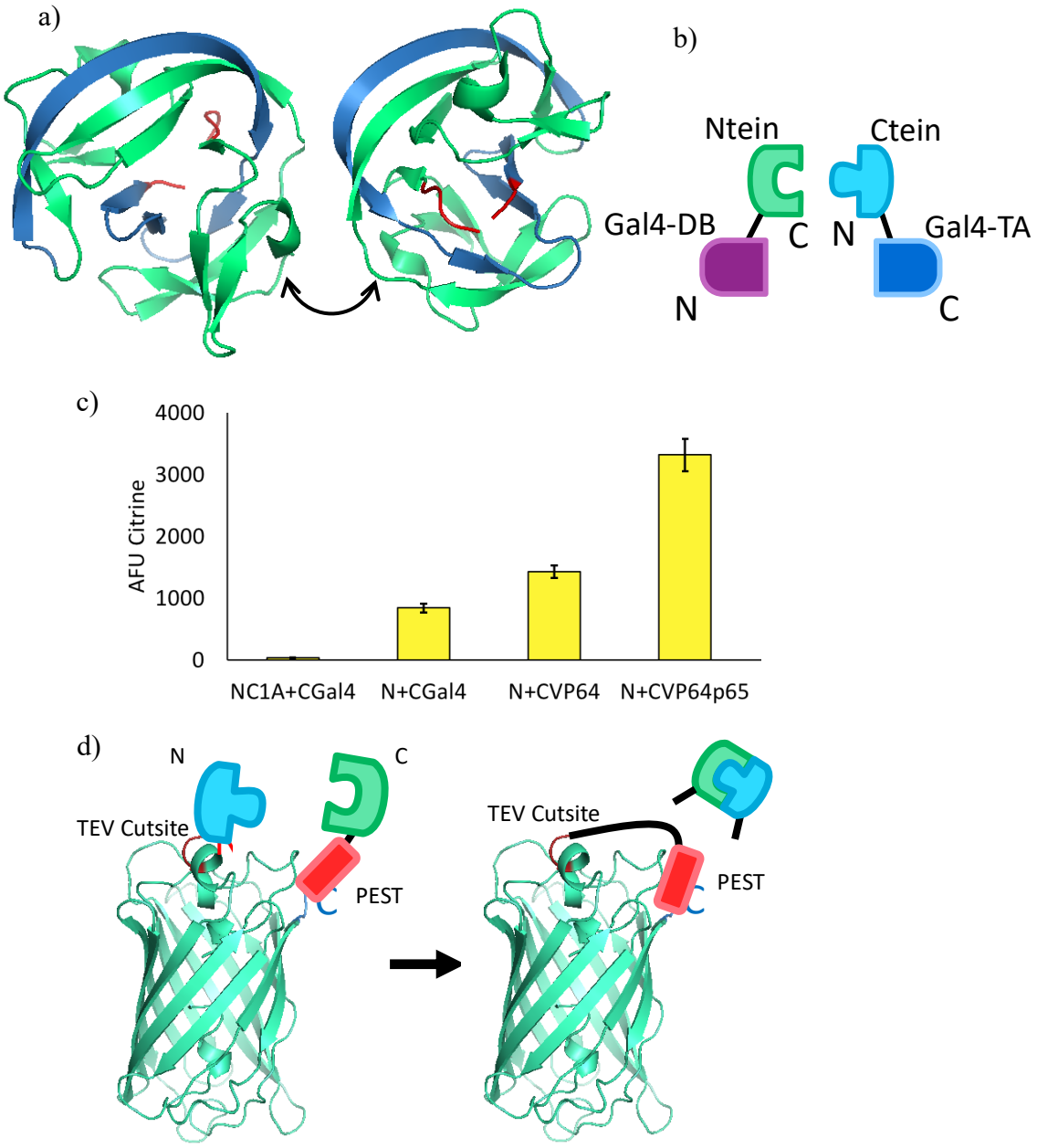


Figure 8: Gp41-1 intein structure and synthetic context
a) Gp41-1 protein structure. The Ntein is depicted in green and the Ctein is depicted in blue, with the active site marked in red. **b)** Schematic for our intein testing platform; successful splicing will result in a full Gal4min which can drive a reporter gene. **c)** HEK239FT cells were transiently transfected with DNA for Nteins and Cteins with varying splicing ability and TA strength along with a UAS driven H2B-Citrine reporter. Reporter output was measured by flow cytometry and reported in geometric mean fluorescence ± s.d.. **d)** Schematic for a cyclizable GFP with a degron and protease cut site.

3.2.4: Results

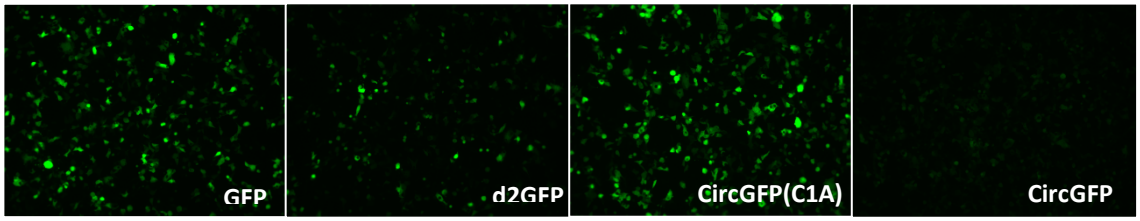
We expressed our cyclizable, PEST-tagged GFP in HEK293FT cells along with GFP, GFP-PEST (also known as d2GFP, established to have a 2 hour half-life), and a splicing incompetent GFP-PEST control with a C1A mutation in the Ntein. By fluorescence microscopy, we see a bright signal from GFP and a dimmer signal from d2GFP due to the high turnover rate as expected (**Figure 9a**). While we had previously seen the C1A mutation lead to an inactive TF, we find that the cleavage activity preserved by this mutation allows the protein to recover the longevity of stable GFP [92]. However, where we expected the cyclized form to be stable, we instead saw a lack of fluorescence.

We immunostained these cells for GFP to determine if the GFP-based construct was being produced (**Figure 9b**). We saw a very strong signal from these cells, indicating that GFP was being produced and retained for the cyclized GFP at a rate much higher than the other GFP constructs. Western blot indicated the expected weights for GFP, d2GFP, and cyclized GFP with C1A, however there was a series of bands for cyclized GFP corresponding to 1x, 2x, 3x, up to 7x and possibly up to 10x the weight of the cyclic product (which would include the linker and the PEST domain but not the inteins) (**Figure 10a**). The band strength of the blot indicates that the molecules are more likely to splice with another protein than remain a singlet cyclized protein. We wondered if this was driven by the overproduction of product by the HEK293FT cells, which could be ameliorated by reducing the amount of plasmid transfected into the producer cells and we would expect to see the band weights change as more cells were producing the construct

via a single transfected plasmid over cells transfected with two or more. We titrated the DNA used for transfection, but saw no change in band wight (**Figure 10a**). In contrast, when we expressed TEV-P along with the construct, we were able to recover singlet GFP and this was retained at a higher rate than the mutated C1A construct with GFP-equivalent stability (**Figure 10b**).

We tested whether this band pattern of polymerization was due to the concentration of construct at the production site and attempted to alter the environment of production by tagging the protein with a transmembrane tether that would be released upon splicing. This would target the ribosomes to the endoplasmic reticulum before full translation of the protein and these proteins would be actively trafficked to the plasma membrane, away from the production site, upon completion. However, this did not change the band pattern (**Figure 10c**).

a)



b)

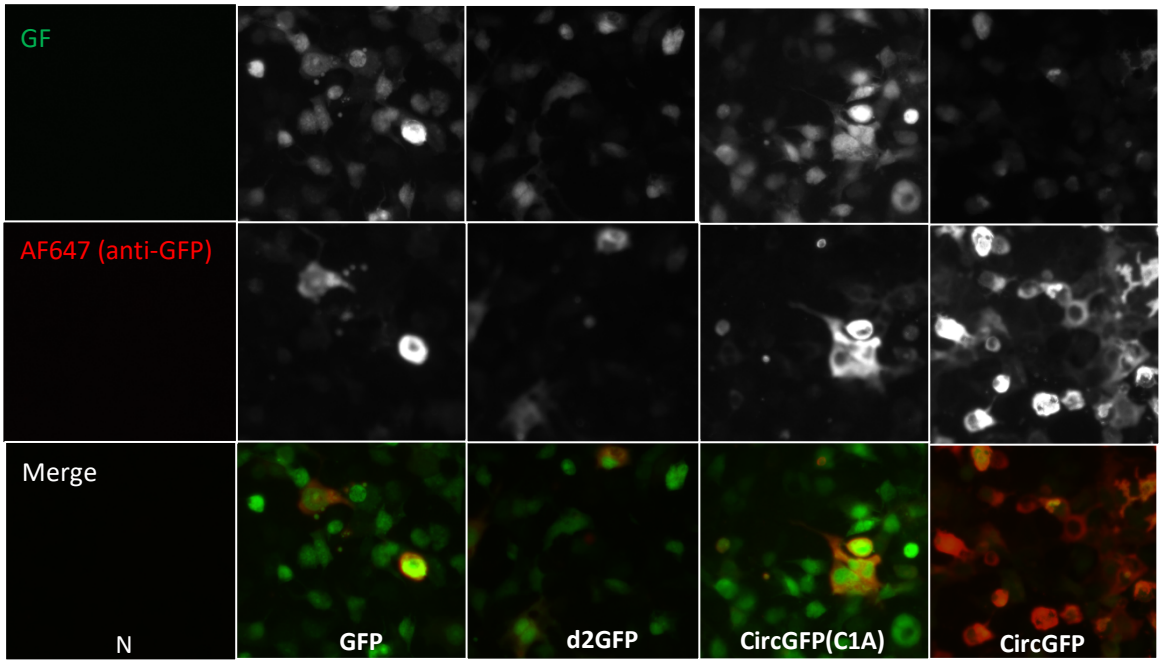


Figure 9: Expression of intein-GFP construct in HEK293FT cells
a) Fluorescence images of transiently transfected HEK293FT cells expressing GFP, PEST tagged d2GFP, the cyclizable construct with a C1A mutation, and the circularizable construct with active inteins. **b)** Transiently transfected cells with the above constructs and a non-transfected control were fixed in paraformaldehyde and immunostained for GFP using a rabbit polyclonal antibody (Sigma Aldrich G1544) and a goat anti-rabbit-AlexaFluor 647 conjugated secondary (Sigma Aldrich SAB4600184) along with an examination of retained fluorescence. It can be seen that while the cyclizable GFP does not fluoresce, it is still well expressed.

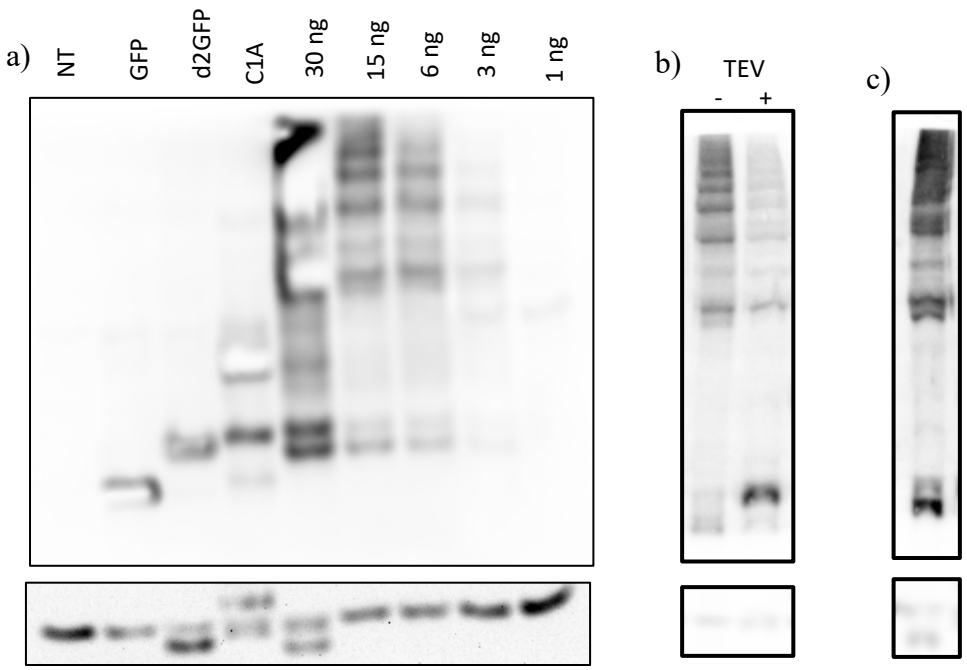


Figure 10: Western blots of cyclizable GFP
a) HEK293FT cells were transfected with 20 ng of control constructs (GFP, d2GFP, and C1A mutated cyclizable GFP) as well as cells transiently transfected with titrated cyclizable GFP DNA. A polyclonal rabbit anti-GFP (Sigma Aldrich G1544) was used to identify GFP, expected at 26.8 kDa, d2GFP at 31.3 kDa, and non-splicing C1A GFP with the cleaved Ctein at 42.1 kDa. Singlet cyclized GFP is expected at 32.3 kDa, but bands are also seen at ~65 kDa, 95 kDa, and 130 kDa, corresponding to 2x, 3x, and 4x. Higher weight bands corresponding to greater polymerization are evident. Gapdh was evaluated as a loading control using a mouse anti-Gapdh antibody (Thermo Fisher MA5-15738). **b)** HEK293FT cells expressing the cyclizable GFP with and without TEV protease show enrichment at the 32.3 kDa band. **c)** Cyclized GFP with an added TMD was transiently transfected into HEK 293FT cells, but this did not significantly change the band pattern of the polymerization reaction.

3.2.5: Unanswered Questions

We find it striking that a fast-splicing intein placed at nearby termini of a stable protein would be more likely to splice with an adjacent protein. This may indicate a significant amount of time during the GFP maturation process that the termini are sterically constrained away from each other, or at least an amount of time longer than the time required for the intein to associate and splice with another protein. This could be tested by altering the cargo protein and or the intein to compare resulting band patterns.

It would be of interest to produce this construct in a cell free system where the concentration of protein produced can be carefully titrated. This would help determine the concentrations necessary to push the reaction to longer polymers or shorter triads and diads and could give functional data about the environment within the producer cell if the distribution can be replicated artificially. It is also possible that this band pattern is due to the physical aspects of the system including time the inteins are sterically constrained from each other with respect to the association and splicing time. The splicing time is known, and this could help elucidate the intermediate stages of GFP maturation.

We find that this project has opened up more questions than answers about the use of inteins in synthetic contexts. We would be very interested to see advances in the realm of studying transient protein dynamics to elucidate the cause of the interactions here.

3.3 Engineering inteins with inducible activity

3.3.1 Background and motivation

Inteins are already utilized as powerful tools in synthetic contexts, however tools are more effective the more precise control the user has over them. While there have been efforts to improve the splicing speed and efficiency of known split inteins [148,150], there are also advantages to having inteins that only splice in specific contexts. This system would be very similar to controlled dimerization; a platform that has enabled drug sensitive control of gene expression [96,183], light controlled enzymatic activity [184,185], and other powerful processes. However, since inteins are autonomous splicing units, they also confer unique switch-like behavior, which we anticipate can make the system more sensitive to signals. Additionally, the intein domains are removed from the protein product, making the system much more flexible and plug-and-play friendly. There is much less need to go through the lengthy and costly iterative design process to determine best split sites [186,187]. Engineers have worked hard to develop systems with control schemes that are easy to implement in laboratory or therapeutic settings and confer precise spatial and temporal control. We suggest that implementing this design process on inteins would result in invaluable protein tools.

Inteins have evolved environmental sensitivity in a variety of contexts, not just pH and temperature as previously mentioned but also cases of DNA damage and high reactive oxygen species [188-190]. In the native context, these splicing elements often act as sensors for environmental stressors [189-191] despite their origin as selfish genetic elements [134]. It has been suggested that these could be synthetically implemented as

environmental sensors as well, such as for reactive oxygen species [192]. However, changing the environmental context in the case of an experiment or a therapy as a measure of control over the activity of a protein isn't always simple or possible, and increasing concentration of reactive oxygen species or DNA damage is clearly not orthogonal to their host. Indeed, neither is changing temperature or pH. To this end, it is necessary to engineer novel function into inteins in order to produce the desired tools.

3.3.2 Small-molecule drug induced intein splicing

3.3.2.1 Drug inducible inteins

Small molecule drug systems are popular in both discovery and therapeutic platforms for their ease of use, temporal control, and built-in gradation of signal by control of dosage. A small molecule drug driven splicing tool would be very valuable for control of protein activity. A separated, inactive protein could be turned on by a drug-driven splicing reaction at the time drug is added and an intein-based split system could be used to inactivate proteins without the lengthy library screening and testing necessary to find small molecule or peptide inhibitors [193,194]. It's also advantageous over small molecule protein heterodimers like rapamycin and gibberellin that require fusion proteins, as splicing removes the associated machinery, leaving a small to inconsequential scar [195,196]. Implementation of heterodimerizes within new proteins requires lengthy rounds of testing and checking, but this can be reduced with intein splicing [186,197]. Of note, the other difference when using inteins is that splicing is irreversible. This can be an advantage when switch-like behavior is desirable, as well as a disadvantage, when temporal sensitivity to drug removal is required. We suggest that this

can be additionally addressed and tuned via the use of degrons [198].

A few labs have taken approaches to this problem. The first small molecule induced intein splicing system was by HD Mootz in the Muir lab, who artificially split a single-chain intein from yeast, Sce Vma intein [199]. Since this intein was adapted to a constrained length between the Ntein and Ctein, it did not have the extensive charged association domains commonly found in naturally split inteins, which made the splicing slow unless it was artificially brought into proximity with the other half [176]. Mootz et al used FKBP and FRB rapamycin induced dimerizers. This system was successfully employed in HEK293 mammalian cells and in flies, suggesting general applicability of the method [200,201].

The second approach, by D Buskirk in David Liu's lab, uses an estrogen receptor to fix the two sides of the intein conjugation machinery of a single chain intein in an open configuration until it binds 4-hydroxytamoxifen, an estrogen analog [202]. Upon binding, the protein changes configuration to allow the intein domains to associate and splice. This system uses steric separation of protein domains to inactivate the associated protein extein. Since it employs an estrogen receptor, it is not orthogonal to mammalian systems, so while the first round of engineering in yeast yielded good results, this had to be revisited to adapt it to mammalian systems [203].

Neither of these systems received much employment in the scientific community, despite being established more than fifteen years ago and having development work suggesting broad applicability to a number of biological contexts. This suggests that there are still needs not addressed by these modifications.

3.3.2.2 Design

We wished to employ our NS3 HCV protease and drug inhibitor system to create a drug inducible intein. Working off of the structure of the Mootz, et al construct, we first developed a testable prototype that would function as an AND-gate [199]. It would require both rapamycin and NS3 inhibitor in order to splice by using the proximity dependent intein, VMA, and employing NS3 protease domains as the linker between the intein and the FKBP and FRB association domains (**Figure 11**)**.** We intended to switch the FKBP and FRB domains for constitutive protein association domains once the technology was confirmed to work under the previously established framework.

We acknowledged that an AND gate reduces the anticipated number of spliced proteins, so we employed the splicing machinery within a Cre recombinase. Cre recombinase targets loxP sites and only two active molecules within a single cell are required to mediate genetic changes, making it a sensitive system for testing intein splicing [204]. We used a selected pool of cells expressing a DsRed to GFP Cre stoplight cassette, which contains loxP sites that excise the sequence for the DsRed and a terminator, allowing the cell to produce GFP [205]. We used a previously published intein recombination site in Cre, at amino acid 232 [206]. However, despite confirming activity of the Cre recombinase via intein splicing at that site using other inteins, we could never recover VMA activity, including when reducing the structure to that of Mootz et al, 2002. These results indicated that the VMA intein was not very suitable for use in mammalian cells, and this was corroborated by a 2019 paper from the lab that failed to see activity from the VMA-rapamycin system [207].

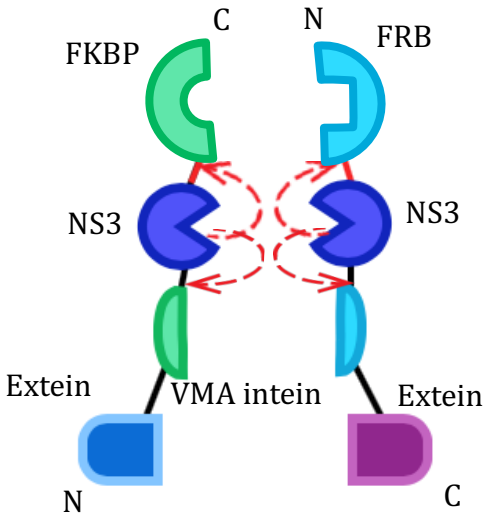


Figure 11: Drug inducible intein
Schematic for the use of a LiNC in the creation of a drug inducible intein. This inducible intein requires induction from both a NS3 inhibitor to stabilize the VMA intein connection to the dimerizers and rapamycin to bring FKBP and FRB together.

3.3.2.3 Outlook on drug-inducible splicing

The established methods for drug inducible splicing clearly don't fit the engineering needs of the scientific community. One technology utilizes the estrogen receptor as a scaffold, and this has endogenous interactions in mammalian systems. The other technology utilizes an intein with irreproducible splicing activity.

The 2019 paper from the Muir lab describes the construction of a new proximity-induced intein system based on the well-established Npu intein, and show that it can be brought together with rapamycin, making a drug inducible intein system. However, the proximity-induced activity is via a caging mechanism, where the cages have a loose electrostatic attraction that blocks the naturally split intein from finding its pair, but will bind the pair when closely associated as with rapamycin dimerization [208]. This structure is unsuitable for the protease drug-inhibited cleavage system due to size and association constraints, and these factors may limit the usefulness to the scientific community as well. There remains a need for an easily implemented small molecule sensitive splicing system.

We hypothesize that one can be constructed in the future using one of many small molecule drug hybridizers using the schema published by Mootz et al., in 2002 [199]. The need, in this case, is for a structured proximity induced intein that is active in the target system, such as mammalian systems. We additionally hypothesize that this can be constructed using either directed evolution for splicing activity only in the presence of the small molecule dimerizer in systems such as the eVOLVR [82], or using rational protein editing to reduce spontaneous intein association, either by reduction of the intein

association domain or by altering the electrostatic attraction such that it does not spontaneously associate. The first strategy retains the drawback of non-mammalian contexts used for directed evolution. There would remain a need to corroborate proximity splicing in mammalian cells. The second strategy is labor intensive, but would be facilitated by the decreasing cost of gene construction, reducing the time needed in house to test a library of rationally designed constructs. Despite advancement in cloning technology, it is still a lengthy and sometimes obtuse prospect to edit protein structure in a traceless manner. Both of these strategies would require careful editing to retain enough association to allow the intein components to align for splicing but not so much that the intein has background splicing activity in the absence of a dimerizer, and this specific titration would be difficult to measure in alternate biological systems. We hope that future endeavors will design with orthogonality and plug-and-play implementation in mind.

3.3.3 Light activated inteins

3.3.3.1 Motivation

Light activated systems are particularly useful in laboratory contexts where the targets can be made easily accessible to light due to the precise temporal control, same as drug inducible systems, with the added advantage of spatial control. With the use of advanced optics, single cells can be targeted [208]. These systems have found fewer purposes in therapeutics for the obvious reason that humans are large, opaque structures, yet some have still been approved for skin related diseases such as cancer [209] as well as for treating light accessible areas like the airways [210]. There remain therapeutic

systems where light-inducible activity is welcome, and so a need for orthogonal development.

There have been a few approaches to a light-induced intein splicing system. The earliest dates to 1995, where the Noren lab probed plant-derived intein activity by inserting a non-canonical photo-caged amino acid at the active serine site, thereby proving that it was necessary for the splicing activity [211]. This approach was repeated by a number of labs with inteins in vitro and in vivo using non-canonical amino acid incorporation tools [212-215]. Other strategies have included one from the Muir lab, using photodimerizers from *Arabidopsis thaliana* in conjunction with the proximity induced VMA intein [216]. Additionally, Light, Oxygen, Voltage (LOV) domains have been employed as steric hindrance to splicing in the absence of light, however these did not generate splicing with a great dynamic range, the second reporting a dynamic range of ~2.5x [217,218]. There remains space for an effective, easily implemented light-controlled intein system.

3.3.3.2 Design

For the construction of our light-inducible intein, we wished to use the photocleavable protein PhoCl [219]. PhoCl was engineered from a photoconvertible green to red protein to spontaneously dissociate. The chromophore still converts, however during this process the chromophore is released from the beta-barrel and the protein domain conjugated to it detaches from the complex, forming a light destabilizing linker. We wished to place this dissociation domain in between an intein and a caging peptide that blocks the sites essential for intein pair capture. This strategy is based on

Intein Zymogens by Gramespacher, et al., which employ these cages as a tool for identifying viral protease activity by proteolytic cleavage that thereby allows for a splicing dependent signal [220]. This paper reported that the efficiently splicing gp41-1 intein was a poor choice for this system, as it was more likely to escape the cage and splice non-specifically, so we chose a slightly less efficient intein from the same source called gp41-8 [92]. We tested this system using a minimal Gal4 TF to determine splicing activity (**Figure 12a**).

3.3.3.3 Results

We expressed our PhoCl-conjugated caged-inteins in HEK293FT mammalian cells to test the dynamics of the system. First, we wished to see if we could see the photo conversion under a fluorescence microscope. Under the 545 nm lamp for imaging red fluorophores, we first saw no signal (**Figure 12b**). We observe a strong signal under the 470 nm lamp as expected for unconverted PhoCl and after returning to the 545 nm light, we see signal for converted peptides from the light used to visualize the unconverted protein. We use this process for photoconversion moving forward.

We expressed our PhoCl inteins together with each other and with uncaged gp41-8 inteins in series in HEK293FT cells with a co-transfected mTurqoise2 UAS-driven reporter; we designated PhoCl inteins as Cag-Ctein or Cag-Ntein and uncaged inteins as Un-Ctein and Un-Ntein (**Figure 12c**) [221]. Half of these samples were hit with uncaging light for five minutes; PhoCl was observed to uncage at 500s so we expected this would be sufficient for our samples [219]. At best, we observe a dynamic range of 1.36 with a caged Ntein and uncaged Ctein, which falls short of our target of 2.5.

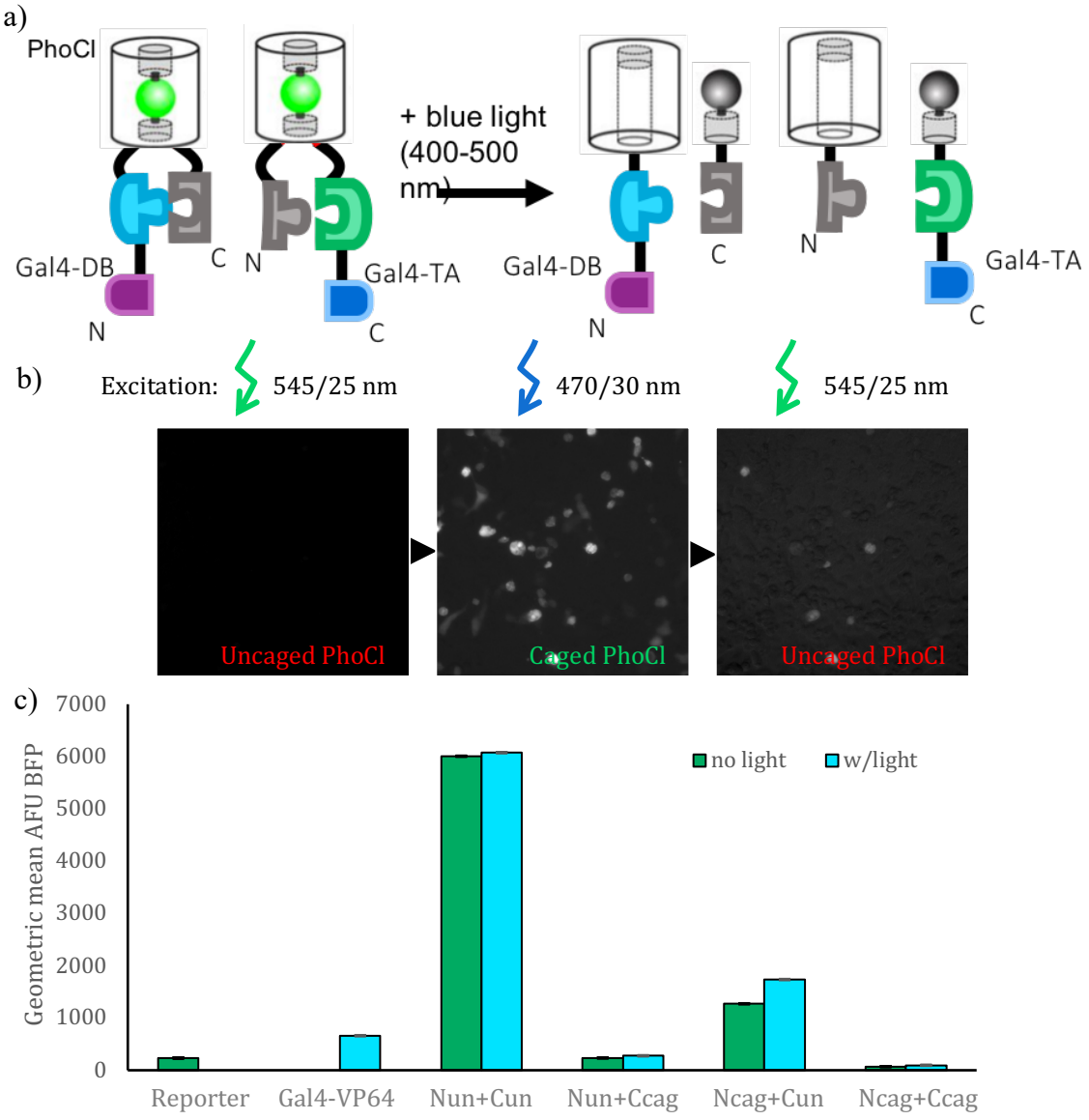


Figure 12: A light inducible intein
a) Schematic for the use of a light-activated cleavable peptide in the creation of a light-inducible intein. Inteins are caged until photocleavage releases the inhibiting domain. **b)** Fluorescence images of a sample of cells before, during, and after photo-cleavage, visible through the formation of a red fluorescent molecule before the photocleaved domains diffuse away from each other. **c)** HEk293FT cells were transiently transfected with a normal intein or photocaged intein half along with a UAS driven BFP reporter 24 hours prior to exposure to 470 nm light. The cells were analyzed by flow cytometry 24 hours post activation and compared to a reporter only control and constitutively expressed Gal4-VP64. BFP fluorescence is reported in geometric mean fluorescence ± s.d..

3.3.3.4 Discussion

We were able to see some light activated splicing dependence in our PhoCl caged intein construct, however it is clear that more development of the system would be necessary to fully control it via light. We saw very little activity using both caged inteins, even when activated via light. When the PhoCl-caged inteins were expressed with a non-caged intein, activity was recovered, however the greater portion of this activity was light independent. The greatest dynamic range between un-activated and light activated inteins was seen in the case of caged Ntein and uncaged Ctein, however we observed a dynamic range of 1.36, which falls short of the previously published work which established a dynamic range of 2.5 [218].

We were not able to reach our goal of a greater dynamic range than the leading photoactivated intein competitor. An important area of exploration would be improvement of the tools used for photoactivation, including stronger and more specific light activators and dark rooms to reduce background light activation. The unactivated samples were protected from light as much as possible but needed to be worked with in communal spaces and grown in a communal incubator. However, a system this difficult to implement does appear to have drawbacks for use in a laboratory setting, let alone expansion into therapeutics, so it is possible that more success might be found using a system with a further-UV shifted excitation peak such that ambient light contamination would not have as strong of an effect.

It would be of interest to determine if the dynamic range could be improved by directed evolution or a library of mutants screened for distinct caged and uncaged

phenotypes, similar to the process used to elucidate PhoCl. We acknowledge that more work needs to be done to make this a useful tool for downstream biotechnological applications.

3.3.4 Additional Methods

The VMA intein was reported in Mootz, et al 2002 [199]. We humanized the sequence using IDT expression optimization software and had it synthesized as a long gene fragment in sequence with NS3 and cutsites (Integrated DNA Technologies). These were cloned using standard Gibson cloning techniques into pcDNA3 backbone in conjunction with Cre recombinase (a gift from Connie Cepko, Addgene plasmid #13775) and FRB and FKBP domains (a gift from Peter Varnai, Addgene plasmid #40897) [222,223]. HEK293FT cells with a clonally integrated Cre recombinase stoplight using the clone production methods reported in Chapter 2 (a gift from Niels Geijsen, Addgene plasmid #62732) were transfected with 25 ng of Ctein and 25 ng of Ntein plasmid.

PhoCl was a gift from Robert Campbell (Addgene plasmid #87693) and the caged gp41-8 sequence was pulled from Gramespacher et al. 2017 [224,220]. The gp41-8 sequence was optimized for human expression using IDT's expression optimization tools (Integrated DNA Technologies) and constructed as a gene fragment by Twist (Twist Bioscience). The reporter plasmid was constructed in house and the mTurqoise2 sequence (BFP reporter) was a gift from Dorus Gadella (Addgene #36204) [225].

PhoCl constructs and controls were transiently transfected using Lipofectamine 3000 reagent (ThermoFisher) at a ratio of 50 ng DNA per 500,000 cells. This was broken

down into 15 ng of each half of the TF and 20 ng of UAS reporter. 15 ng of control Gal4-VP64 was used to approximate the amount of expected full TF produced from each half. PhoCl constructs were activated 24 hours after transfection, and 24 hours after analyzed using fluorescence imaging or flow cytometry. Activation was performed using a 470/30 excitation filter on a Zeiss Axio Observer with a X-Cite 120 lamp (Lumen Dynamics) for 5 minutes. Non-activated controls were kept on a different plate to protect from light.

Cells analyzed by flow cytometry were gated for living cells by scatter detection. Additionally, they were transfected with a constitutive mCherry transfection marker (total breakdown 10 ng mCherry, 10 ng UAS-mTurq2, 15 ng Gal4-DBD, and 15 ng VP64-TA) which was an mCherry tagged H2B gifted by Robert Benezera (Addgene plasmid #20972) [226]. Cells were gated to the top 1% red fluorescence of untransfected cells, collected those expressing both mCherry and red fluorescing PhoCl during uncaging. The geometric mean measured reporter fluorescence levels were reported in arbitrary fluorescence units (AFU). All flow cytometry data was collected using three biologically independent samples and the error bars represent the standard deviation.

3.3.5 Closing remarks

We tested multiple systems for enabling spatial and temporal control over intein splicing. Despite working with published constructs, we were not able to implement drug activated splicing with the VMA intein, and this intein was found and corroborated to be inactive in modern mammalian cellular systems [207]. We hypothesize that the limiting factor in the production of a small-molecule drug-induced intein splicing system is the

engineering of a small proximity-induced intein. We saw a trend toward light activated splicing with our PhoCl caged, light sensitive intein constructs, despite not seeing very much activity when both constructs were light dependently caged. However, the best reported dynamic range of 1.36 does not encourage use of this system over previously reported systems with a reported dynamic range of 2.5 [218].

We remain interested in the developmental space for inducible intein constructs. Inteins are fairly new tools in the protein engineering space. They were only discovered in 1988 [135], giving them a little more than 30 years of development time, whereas rapamycin was discovered in 1972 [227]. They were hindered at the outset by inefficient splicing and slow splicing rates [137,145]. However, despite this, these autonomous protein editing domains were recognized as powerful protein tools and implemented synthetically to remove selection domains from isolated proteins [228]. This drove interest in the protein elements, which in turn drove research into native systems and protein engineering of existing domains. We hope to see continued engineering of these systems moving forward.

There are existing naturally inducible protein splicing domains. These are inducible via conditions that are not useful for therapeutics, but we look forward to seeing them implemented as environmental sensors [190,192]. Additionally, these may be an area of interest for engineering drug or light inducible activity, and further research into these protein elements may reveal more strategies to test. From the active patent space [228], we see that inteins remain an area of active interest and development, and anticipate what tools may come out of the field soon.

Chapter 4: Intein-gated delivery of natively-folded proteins via extracellular vesicles

4.1: Background and motivation

Extracellular vesicles (EVs) have emerged as promising novel therapeutic cargo carriers and powerful biotechnological tools. EVs are endogenous cellular systems for inter-cellular transport and communication. They were originally thought to be a method of excreting waste or extraneous components from a cell, by packaging these unwanted elements in lipid envelopes and expelling them into the extracellular space [229]. It has since been confirmed that these small lipid capsules can be taken up by other cells, allowing cells to exchange signalling molecules as well as membrane components and even nucleic acids [230]. This makes EVs an invaluable tool for transporting synthetic cellular components into target cells for therapy and engineering.

EVs have been tested for application to a number of therapeutic contexts. They are of interest due to their cell-similar membrane composition which makes them less targeted by the immune system and longer lasting in the circulatory system than synthetic lipid nanoparticles like liposomes [231]. EVs have been used to deliver small molecule drugs, encapsulated during vesicle construction via modified cells or after production by diffusion [232]. They have also been key delivery candidates for endogenously folded proteins including membrane proteins. Since EVs are formed from endogenous lipid bilayers, membrane proteins can be natively constructed and then extruded onto the EV with no loss of conformation due to dissociation with the bilayer. Therapeutic delivery of membrane proteins has some existing challenges but it has successfully delivered targeted therapeutic membrane proteins such as membrane-bound TNF-related apoptosis

inducing ligands (TRAIL) to tumors in vivo [233]. These tools can also be used to deliver therapeutic nucleic acids, such as EGFR siRNAs to inhibit tumor growth [234].
EVs are also of interest for bioengineering. Many biological engineering tools like base editors and recombinases are protein constructs, which are easily constructed and encapsulated in EVs. The advantage of EVs as delivery mechanisms for these protein constructs over more traditional mechanisms like electroporation, transfection, or viral transduction is the direct, transient application of the protein. Traditional mechanisms induce production of the therapeutic protein via addition of a DNA template. This means that the cell produces the protein on its own, potentially producing vast quantities. Direct protein delivery limits the amount of protein in the cell, and this reduces off-target effects, cell load, and immune response [235]. EVs have been used to deliver recombinases, Cas-9 ribonucleoproteins, and other base editors in cellular engineering contexts with reduced off-target effects [236, 237, 238].

Endogenous EVs are a powerful tool for therapeutics and engineering, however they also have some persisting challenges. Most methods for collecting endogenous EVs collect a heterogenous population including both microvesicles, formed from pinching off the plasma membrane, and exosomes, formed within endosomes and then released into the extracellular space, and there are difficulties in isolation of desirable populations and scale up [239]. A strategy that addresses this is EV production via viral envelope proteins (VEPs). A number of VEPs have vesicle forming properties, so when expressed by the producing cell can induce production of a high quantity of more homogenous EVs [235, 240, 241]. In particular, the VEP vesicular stomatitis virus glycoprotein (VSV-G)

has been of interest due to its familiarity as an engineering tool. It is used in the construction of lentivirus to expand the tropism of the HIV gene insertion complex due to the way it targets the low density lipid receptor (LDL-R) and family, a conserved family of receptors in multicellular organisms [242, 243].

The use of VSV-G as a vesicle inducing protein also has other advantages, due to its fusogenic properties. VSV-G mediates viral infection not only by targeting receptors, but by causing the fusion of the viral envelope with the target cell's plasma membrane through conformation changes [244]. Non-fusogenic EVs can be impeded from delivering their payload due to endosomal entrapment, where they are endocytosed and the endosome is directly fused to a lysosome instead of allowing the vesicle to fuse with the membrane and release their cargo. VSV-G on the surface of EVs mediates vesicle uptake by direct fusion, making it a more robust system for getting payloads into target cells [245].

VSV-G mediated EVs (VEVs) address a lot of challenges faced when engineering EVs to deliver synthetic cargos, including inducing a greater production of more homogenous EVs for ease of isolation and characterization, and evading endosomal entrapment to improve delivery of the cargo to the target. However, the broad tropism afforded by VEVs is a new challenge to overcome when there is a desire to target certain cells or tissues. There are many methods for doing this, including using other fusogenic receptors with different targeting moieties, or cargos that would only activate in target cells. We focused on a robust method that could be implemented with a variety of cargos and cell targets with minimal changes. This would make it fast and easy to implement in

a variety of contexts. To do this, we developed a system using inteins as a membrane tether and biological lock for the cargo protein.

4.2 Inteins as a translation-dependent membrane tether

We first defined the intein lock system to be implemented in cargo-carrying VEVs. Taking inspiration from the NS3 tethered-cargo system we first replaced the NS3 with an intein to confirm whether this system could, in a translationally dependent way, release a cargo (**Figure 13a**). We chose the Gp41-1intein pair for its efficiency and fast splicing [246]. In this system, we acknowledge that there are two possible configurations: the Ntein could be the linker or lock, mediating splicing between the Ctein-key and extein and the membrane tether, or the Ctein could be the linker, mediating splicing between the cargo and the Ntein-key and extein (**Figure 13b**).

We tested both configurations using the transcription factor (TF) Gal4 as the cargo and GFP as the extein stabilizer for the non-membrane-bound intein halves. These protein constructs were transiently transfected into a HEK293FT clonal line with an integrated UAS driven H2B-mCherry reporter. There is a clear difference in free TF between the systems transfected with the stable singular component versus with the intein key to release the cargo. This is visible via the fluorescent reporter in the nucleus by fluorescence microscopy as well as when evaluated by flow cytometry (**Figure 14a, b**). This system was also tested with other membrane tethers, including a myristoylation-palmitoylation (MP) tag and the signalling sequence from Tom20, which directs the protein construct to be tethered to the mitochondrial membrane (**Figure 14c, d**).

Of note, we observed some leakiness in the cases where only the membrane bound cargo was transfected. This is especially evident with a Cre recombinase cargo. Cre recombinase is a gene editor originating from bacteriophage P1 that mediates DNA splicing at loxP sites, and only two free recombinases need to make it into the nucleus to make a genetic change. The Cre cargo was tested by transient transfection into a Cre stoplight reporter cell line; these cells express a red fluorescent reporter with a terminator which is between two loxP sites, and when excised, allows a green fluorescent reporter to be expressed (**Figure 15a**) [247]. The cells were transfected with both the Ntein lock and the Ctein lock and evaluated by fluorescence microscopy (**Figure 15b**). The cells expressing the Cre cargo construct in isolation display significant GFP production, indicating that there is some small amount of background release of cargo; we expect this is due to degradation of the construct or overburdening of the myristoylation system and could be reduced by decreasing the amount of construct produced. Transient transfection leads to heightened production of a construct over an integrated gene due to single cells picking up multiple copies of the gene, which can be avoided by transfecting with decreased concentrations of DNA. As a proof of concept, we transfected varied amounts of cargo construct DNA and found that not only were we able to greatly reduce background, a greater dynamic range was found with both TF and Cre cargo systems (**Figure 15c, d**).

We have shown that inteins work as a dynamic biological lock and key mechanism. These directly transfected systems display some background cargo release,

however we have shown that this is reduced when protein production is reduced. We hypothesized that the low protein concentration delivered via EVs will similarly address this leakiness.

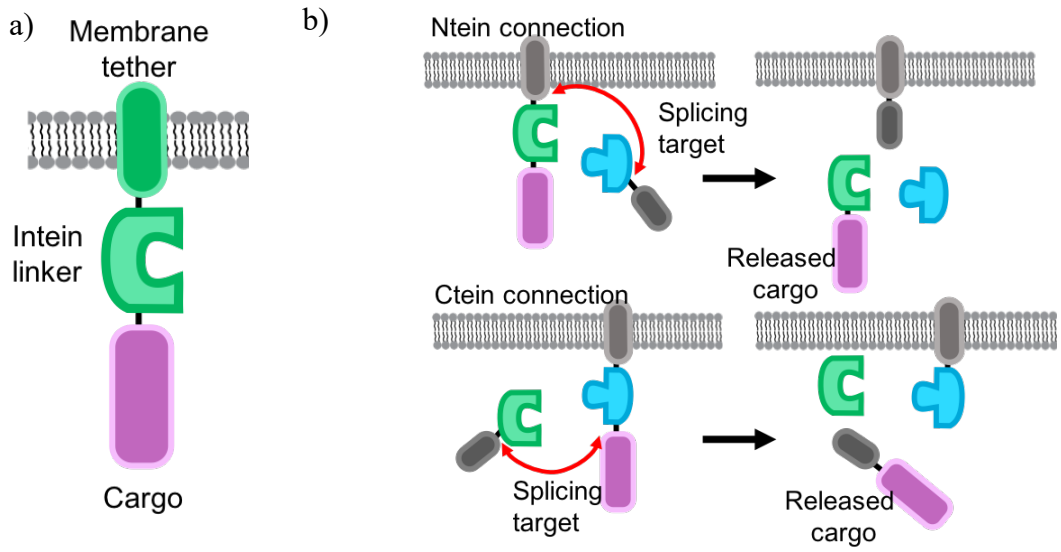


Figure 13: An intein based dynamic membrane tether
a) Generalized design of the cargo tethered system employed in chapter 2; a cargo is concatenated to a generic plasma membrane tether via a dynamic linker, in this case, an intein. **b)** The possible configurations of an intein linker include a Ntein connection where the splicing forms a peptide bond between the Ctein extein and the membrane linker, releasing the cargo along with the associated Ntein; and a Ctein connection where the splicing occurs between the Ntein extein and the cargo, releasing it from the membrane but tagging it with an associated protein domain.

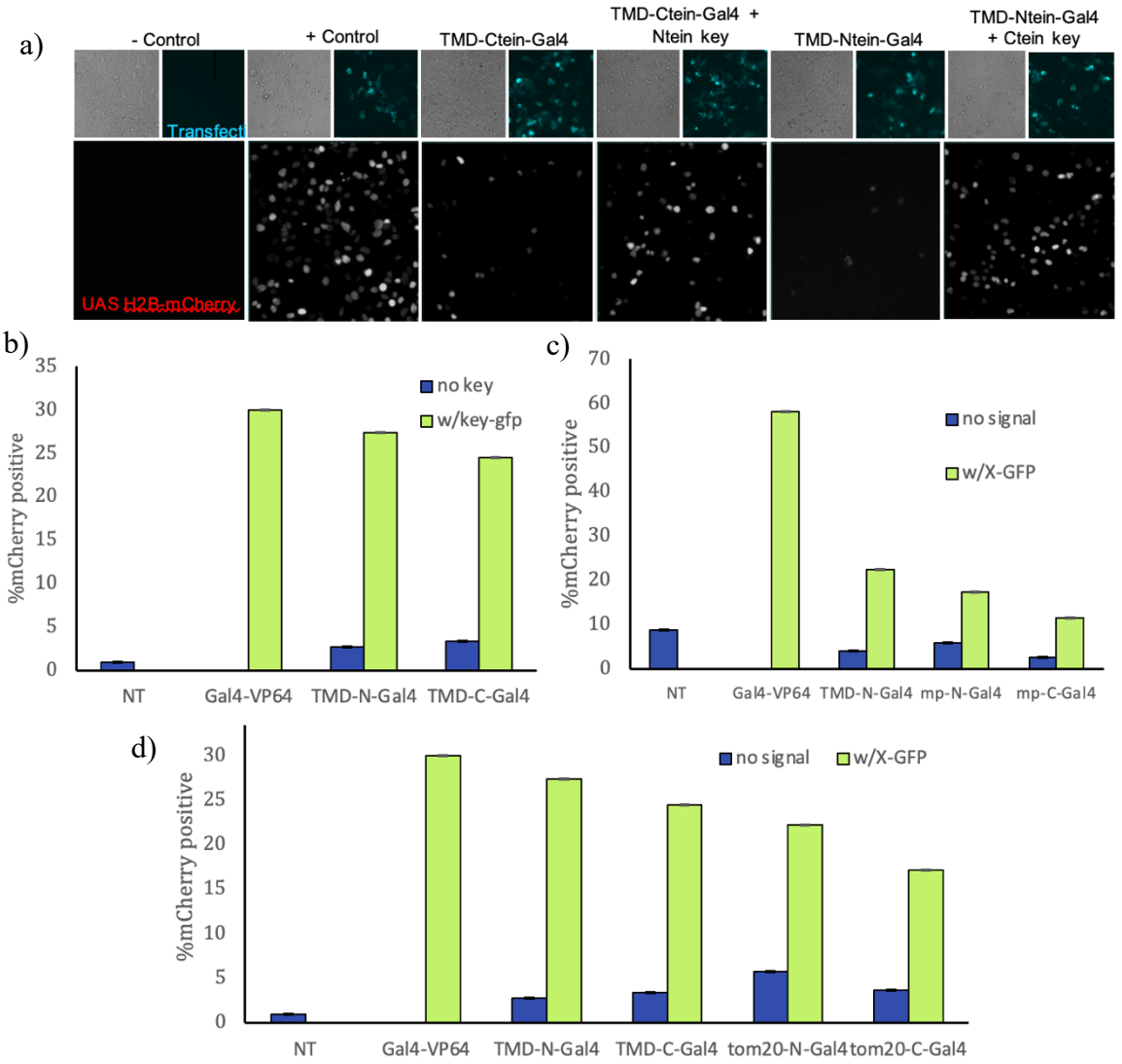


Figure 14: Evaluation of key-dependent cargo release from a membrane tether **a)** Fluorescence images of a UAS driven H2B-mCherry reporter clonal line of HEK293FT with transiently transfected 20 ng of no DNA, constitutively expressed Gal4, and a Ctein and Ntein connected TF cargo, with and without the respective intein key. Brightfield images confirm consistent cell density and BFP was used as a transfection control. We observe some leakage in the case where the cargo cassette is expressed without the key. **b)** Cells treated in the same manner as in **(a)** were analyzed by flow cytometry with gating for live cells and BFP transfection control. The remaining population was gated for mCherry positive cells by the top 1% of WT cells and the percentages of cells that fall into this category are reported. We use the same process with the MP bound cassette **(c)** and Tom20 bound cassette **(d)**.

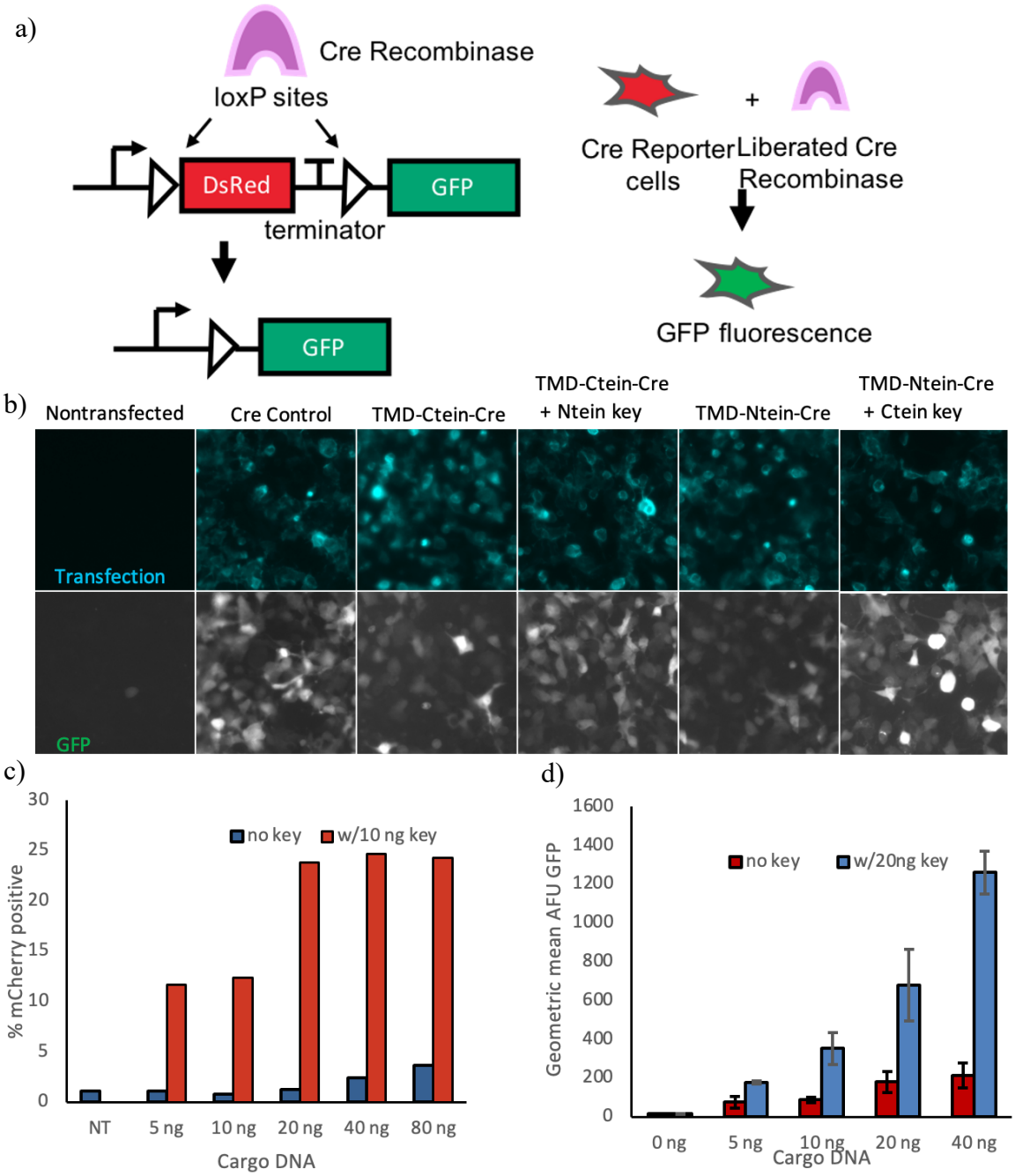


Figure 15: Implementation of the intein-membrane tether to deliver Cre recombinase. **a)** Schematic for Cre stoplight recombination. **b)** Fluorescence images of a Cre stoplight reporter line of HEK293FT with transiently transfected intein tethered Cre cassettes, imaged 48 hours post transfection. These are compared to a naïve and constitutively expressed Cre control. **c)** A UAS driven H2B-mCherry reporter transfected with various concentrations of TMD-Ntein-Gal4$_{min}$ recovers near nontransfected levels of background. **d)** Cre stoplight cells transfected with various concentrations of TMD-Ntein-Cre shows reduced recombination background.

4.3 Reduced tropism of intein-locked VEVs

4.3.1 Design strategy

We next wished to develop a robust lock-and-key mechanism for delivery with VEVs (**Figure 16a**). We used a MP tag as a membrane linker for the system, since the myristoylation residue in particular has been found to be a robust membrane tether for efficient incorporation of proteins into EVs [248, 249]. We compared MP tagged construct with TMD for incorporation into VEVs by western blot and confirmed that using MP led to an enrichment of cargo in the isolated VEVs (**Figure 16b**). We employed the Ntein of gp41-1 as the lock, as it consistently displayed greater dynamic range and less leakiness in all previous tests over a Ctein lock. The cargo is a placeholder for any protein that is inactive at the plasma membrane but can become active once freed. In the ensuing studies, this will be modeled by Cre recombinase and tested in Cre stoplight reporter cell lines. When this construct comes into contact with the other half of the intein, the gp41-1 Ctein key, the intein will drive the formation of a peptide bond between the stabilizing components adjacent to the Ctein (we use BFP as a stabilizer and fluorescent tag) and the membrane tether, releasing the cargo into the cytoplasm of the cell (**Figure 16c**).

The full membrane tether-Ntein-cargo complex will be expressed by the producer cells in conjunction with VSV-G and incorporated into the VEVs (**Figure 16d**). The Ctein-key is induced in target cells, and induction of this complex confers susceptibility to the VEV cargo. We developed a pool isolated by selection via puromycin of HEK293FT Cre stoplight reporter cells that also clonally expressed the Ctein-BFP key.

Pool expression of the BFP tagged construct was confirmed by flow cytometry (**Figure 16e**). These engineered VEVs have limited tropism due not by limiting fusogenic targets but via restricted release of the cargo from the membrane (**Figure 16f**).

To produce large amounts of Cre-charged VEVs, we transiently transfected HEK293FT cells with both VSV-G and the MP-Ntein-Cre construct. The resulting media was collected and concentrated by centrifugation. We used western blot of both the producer cells and the concentrated media to confirm production and isolation of VEVs (**Figure 16g**). We further confirmed that each of the components was necessary for VEV production by performing the collection protocol on HEK293FT producer cells transfected with VSV-G and MP-Ntein-Cre in isolation. While some groups had found that a myristoylation sequence effectively targets proteins to endogenous exosomes, we found no evidence of Cre export in the VSV-G-negative control [239, 248].

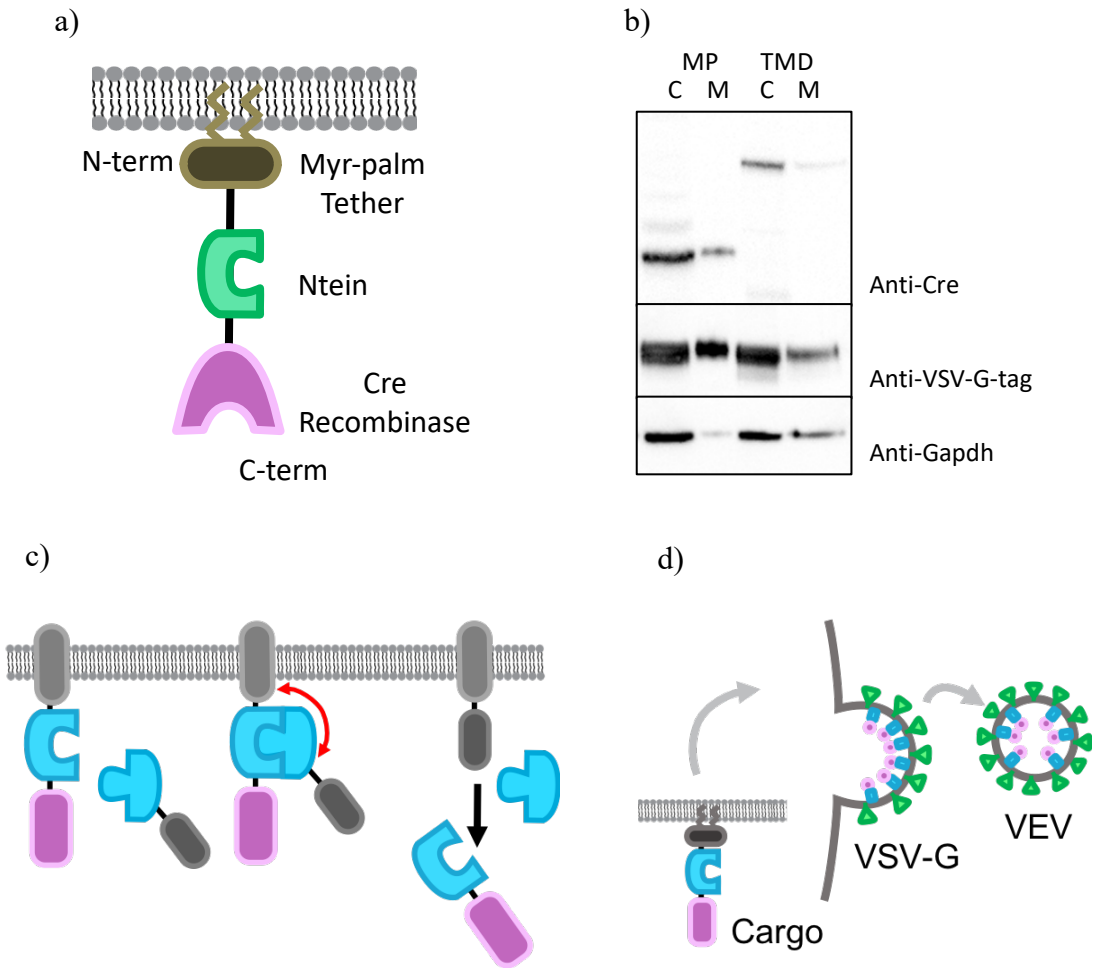


Figure 16: An intein lock for VEVs
a) Schematic of MP-Ntein-Cargo system for integration into VEVs. b) Comparison of Cre integration into VEVs by MP and TMD tethers. Cells were transfected with constitutively expressed VSV-G and cargo cassette. 24 hours post transfection, the media was exchanged for OPTI-MEM, a synthetic, reduced serum media and the cells were allowed to express VEVs into this media for 48 hours. The media was collected and filtered through a 0.45 uM syringe filter and spun at 8000xg overnight at 4ºC. The pellet was resuspended at a ratio of 30 uL for 2mL original media and stored at -80ºC. This process was used for all subsequent collections of EVs for analysis or testing. 10 ul of cell lysate and 3 ul of resuspended media pellet were run on the gel for western blotting. The blot was probed with rabbit anti-Cre (EMD Millipore 69050-3), rabbit anti-VSV-G-tag (abcam, ab183497), and rat anti-Gapdh-HRP (BioLegend 607904). These antibodies and this order of antibodies is used throughout this chapter. c) Intein key activity to release the cargo. d) Cargo incorporation into VEVs.

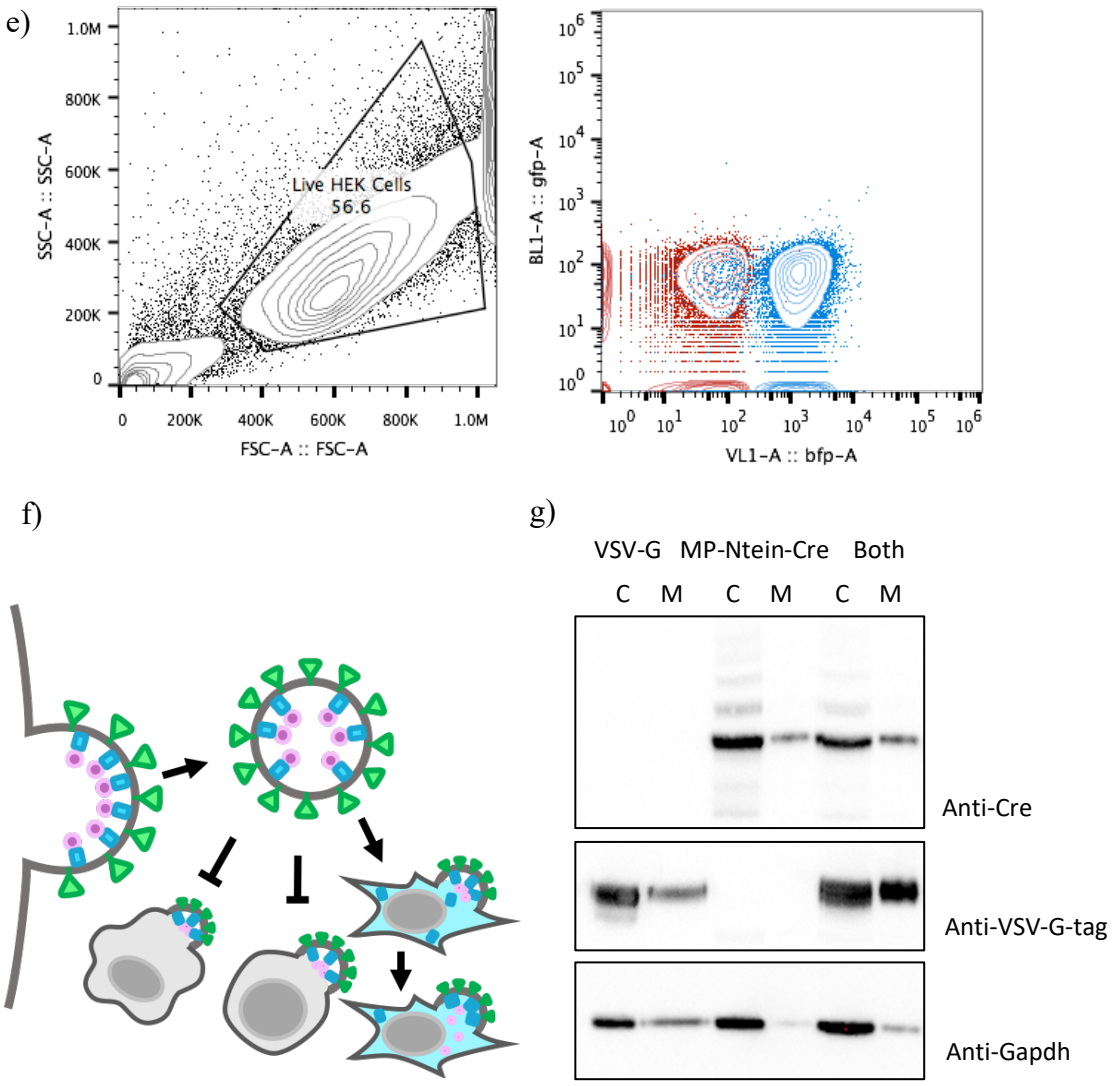


Figure 16: An intein lock for VEVs, cont.
e) Gating for Cre stoplight reporter line and Cre stoplight + Ctein-BFP line. GFP positive cells were determined by gating for the top 0.1% fluorescence in the GFP channel of NT cells. f) Depiction of locking mechanism. Of note, the construct will still be delivered to all cells that can be infected via VSV-G, however only cells with the intein key will be able to "unlock" the cargo from the membrane such that it can be activated within the cell. g) Confirmation of components necessary for collecting bioactive Cre recombinase. Both the MP-Ntein-Cre cassette and VSV-G are required to get a signal from Cre stoplight reporters.

4.3.2 Activity

The activity of the intein locked VEVs was tested in a co-culture of Cre stoplight reporter cells with and without the Ctein-BFP-key (**Figure 17a**). These are compared to a non-treated sample of cells, which display no recombination via GFP fluorescence, and a sample treated with Cre Gesicles from Tarkarabio. Takara's Cre Gesicles are VEVs loaded with recombinase by a chemical dimerizer that drives increased concentration at the membrane, which functions as a positive control. Only the Ntein-locked VEVs display significantly more recombination in the Ctein+ cells than the Ctein- cells. Separate cultures of Ctein+ and Ctein- cells were treated with VEVs to find that the Ntein-lock has a background release of Cre recombinase in 0.5% of cells ± 0.15% with a dynamic range of 36x (**Figure 17b**).

To confirm that the recombinase activity seen is due to intein-key release, we developed a set of low splicing variations. The first variation was a direct conjugate of the MP tag and Cre, which has no splicing capabilities. We confirmed by western blot that this is still incorporated into the VEVs (**Figure 17c**). When Ctein+ and Ctein- cells are treated with these, there is a small amount of background splicing due to inefficient degradation of the cargo, but this effect is insensitive to Ctein production (**Figure 17d**). The second contains a C1A mutation of the Ntein, which had been found by Carvajal-Vallejos et al. to be inhibitive of splicing activity by driving a cleavage of the ctein-extein bond but leaving the Ntein fragment intact [246]. It was also employed to stabilize a full intein fusion by Beyer et al. for determining the crystal structure [250]. However, we found it to trigger a small amount of Ctein dependent release of cargo, indicating that

while the full splicing of the exteins may not occur, there was still some release of cargo from the membrane tether. The Nostoc punctiforme DnaE intein, among others, is also not completely inhibited by a C1A mutation, and this is ameliorated by mutating another residue at the active site (N137A) [251]. Mutating the analogous residue in gp41-1 would require mutating the Ctein, so we instead elected to mutate other residues at the active site, leading to a set of mutations that were able to inhibit intein dependent cargo release: C1A, L2G, D3A. These variations in splicing activity leading to variations in cargo delivery indicate that cargo delivery is intein splicing mediated except in the rare case of degradation escape by the cargo (**Figure 17d**).

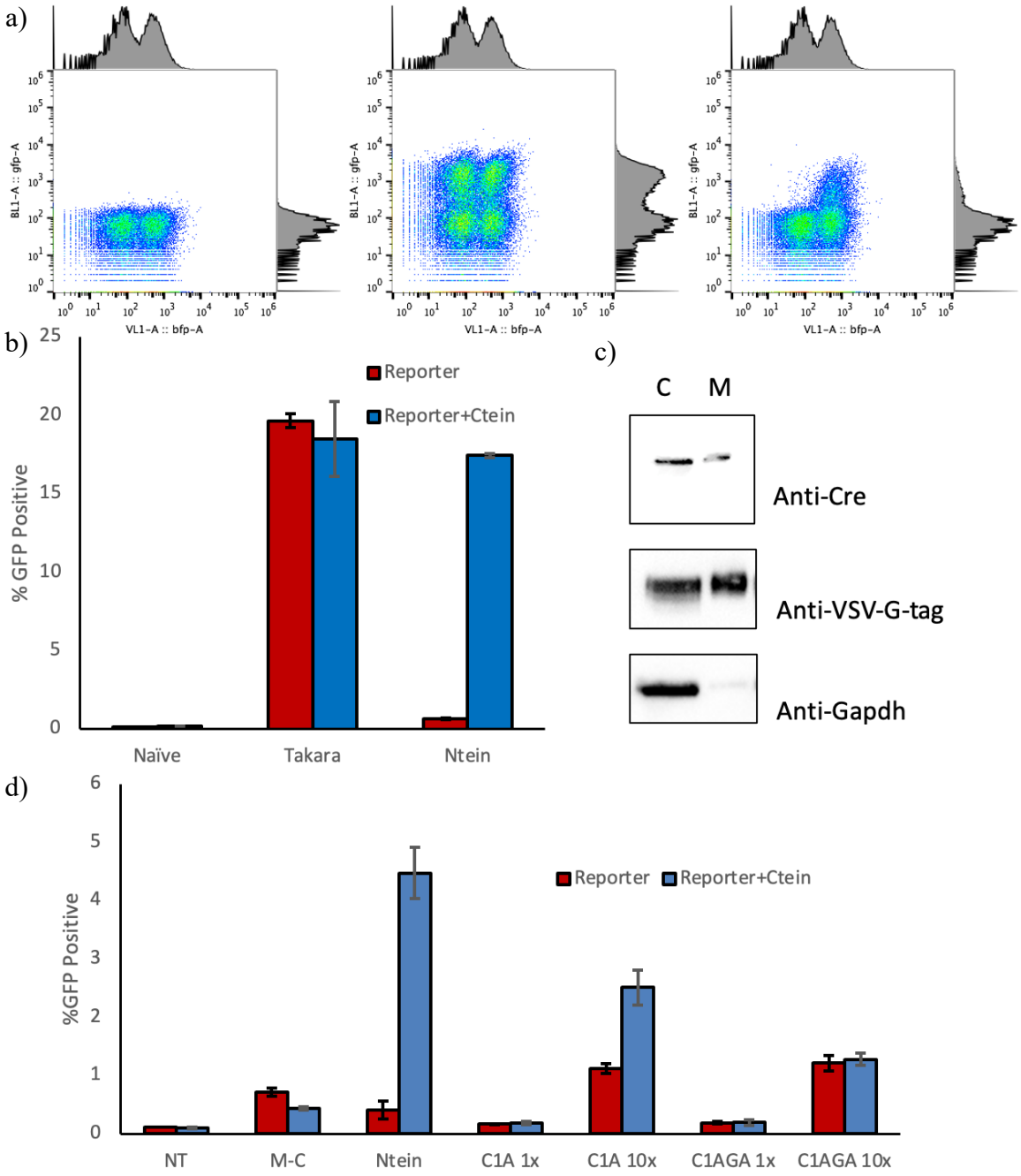


Figure 17: Reduced tropism with the Ntein lock.
a) Cre stoplight reporter with and without Ctein-BFP are displayed with blue fluorescence along the x-axis and green fluorescence along the y-axis. NT samples of both lines display similar green fluorescence, and similar sensitivity to a commercial preparation of Cre recombinase carrying VEVs (Takara 631449). Only the cells expressing Ctein-BFP are able to switch with the addition of our preparation of intein-locked VEVs. b) % of stoplight reporter cells GFP positive with no VEV, commercial VEV, and intein-locked VEV. c) Blot to confirm the packaging of Cre recombinase into VEVs using MP-Cre. d) Activity of slow splicing mutants.

4.3.3 Quantification

We wished to quantify the Cre recombinase carried by our fusogenic vesicles. To this end, we obtained a stock Cre recombinase solution from New England Biolabs (NEB M0298L), however the sample was measured according to editing activity rather than the concentration of Cre. To adjust for this, the concentration of Cre was compared to the standardized 0.3 mg/ml BSA added to the sample. A known amount of sample (diluted 1:2 from 5 ul to 0.4 ul) was run per well on an SDS-Page gel in loading buffer (**Figure 18a**). The gel was stained with GelCode Blue reagent (ThermoFisher 24590; stained for 1 hr, washed in DI H2O for 2 hours) to evaluate protein concentrations. The pixel intensity of the bands for each of the BSA and two bands for Cre recombinase was evaluated by ImageJ's Gel submenu. These values were related to the known concentration of BSA by the following formula: $x = \frac{\frac{0.3mg}{\mu L} * Volume}{Intensity}$ where x is the intensity multiplier that was then used to determine the concentrations of the two Cre recombinase bands at each volume measure. The resulting concentrations were 0.26 mg/uL for the top band and 0.38 mg/mL for the bottom, for a total of 0.64 mg/uL Cre recombinase in the NEB sample.

A small amount of this sample was run alongside western blots as a loading control since a standardized amount of antibody should bind to each Cre molecule regardless of molecular weight or conjugated protein components. We use this to evaluate the concentration of Cre recombinase in both a standardized sample of Takara gesicles as well as our VEVs using both the MP and TMD samples (**Figure 18b**). We obtain a concentration of 0.011 mg/mL Cre recombinase for our MP gesicles, or about 1:30 of the concentration of Cre recombinase from NEB, and for the TMD we obtain a

value of about 0.0013 mg/mL Cre recombinase, meaning that only about a tenth of the Cre recombinase is concentrated from the media for this system.

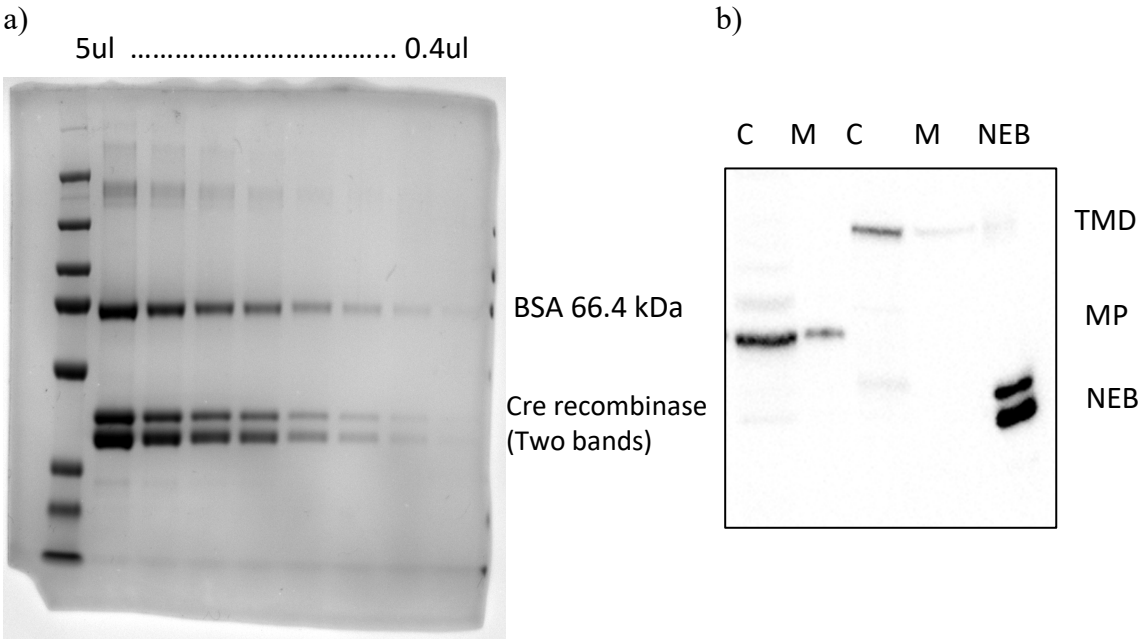


Figure 18: Quantification of encapsulated Cre recombinase
a) GelCode Blue stained SDS-Page gel of NEB Cre recombinase sample. The sample was titrated in 1:2 fractions and mixed with LDS loading buffer to run on the gel. The starting amount was 5 uL with a known 0.3 mg/mL concentration of BSA (66.4 kDa), and Cre recombinase was expected to run at about 38 kDa (though this will change depending on the purification methods used, and we see that there are two bands for the protein in this gel). **b)** A western blot stained with anti-Cre rabbit (EMD Millipore 69050-3) picks up both Cre recombinase bands from NEB (0.5 uL). These were compared to the bands from the media fractions (M) of the MP and TMD using ImageJ gel analysis software to estimate the amount of Cre contained in 3 uL of VEVs.

4.4 NS3 protease as an alternative cargo release mechanism

In Chapter 1, we described a membrane tether system where a cargo was liberated by cis-protease cleavage. Functionally, it operates in a very similar way to the intein-lock system, where the cargo is collected at the membrane until a trigger (the removal of the protease inhibitor) allows the cargo to be released from the plasma membrane and make its way to the inner components of the cell. We therefore assumed that this could be delivered by EVs in a similar manner to the intein-locked cargo. The VEVs could be produced in inhibitor-rich media, allowing the cargo to be collected at the membrane of the cell for packaging into VEVs. The inhibiting drug would be diluted during the resuspension step after collecting and spinning down the VEVs, and further diluted into the media of the target sample. This would allow the cis-protease to activate, cleaving the cargo free into the lumen of the vesicle such that it would be released upon fusion of the VEV with the membrane of a target cell. This system does not target the VEVs or cargos to particular cells in a sample the way the intein-lock does, however it also does not require the target cells to contain a translated component to release the cargo, making it less specific but more flexible.

We generated NS3 VEVs with and without 5 μM of the NS3 inhibitor grazoprevir. The VEVs were western blotted to determine loading of Cre recombinase in relation to MP-Cre as determined above (**Figure 19a**). We tested these on Cre reporter cells at 1X and 10X the usual concentration of VEV solution due to the expected low concentration of Cre recombinase (**Figure 19b**). While this strategy delivers recombinase irrespective of expression of Ctein, it only encapsulates a fraction of the Cre recombinase

as compared to the Ntein version and has only a fraction of the activity.

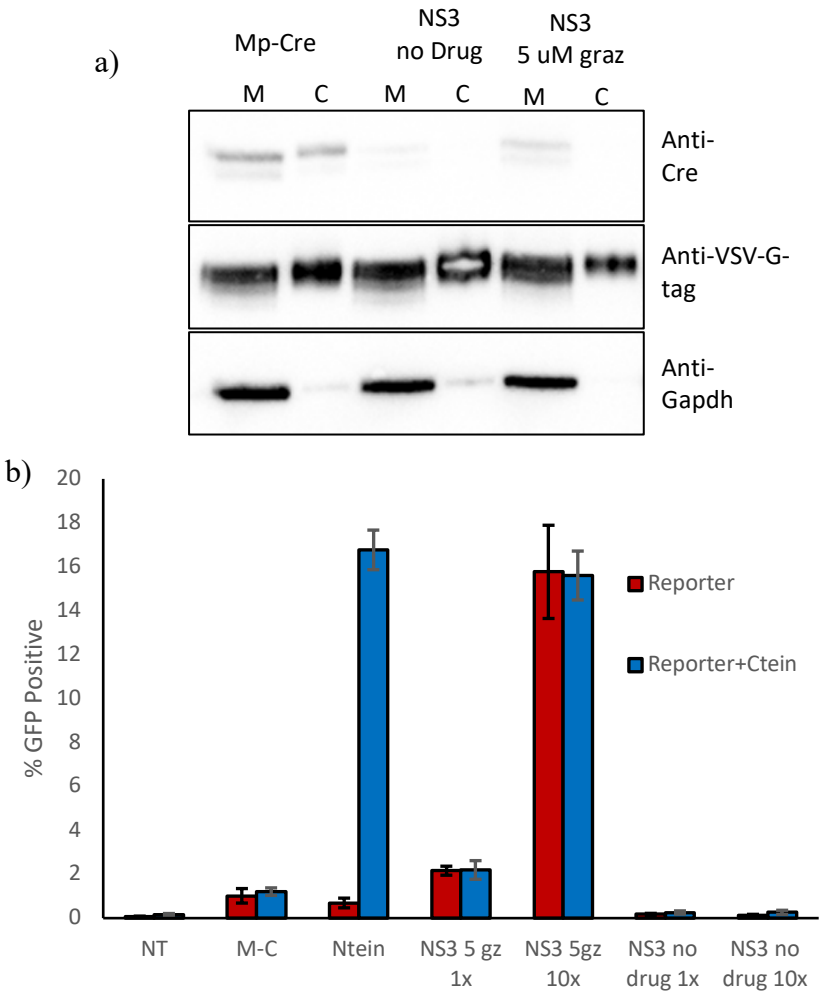


Figure 19: Using anti-viral drugs to load VEVs
a) Western blot of cell fractions and media fractions from HEK293FT cells transfected with VSV-G and MP-Cre, MP-NS3-Cre, and MP-NS3-Cre with 5 uM grazoprevir. We do not see a significant fraction of Cre recombinase in the NS3 mediated cargo loading system. b) Analysis of Cre stoplight reporter cells treated with these VEV fractions for 48 hours show significant Cre activity, though 10X the concentration of the cell fraction is required to reach the activity of the MP-Ntein-Cre system. This activity is not dependent on expression of Ctien-BFP, as we expect. No activity was seen in the case of VEVs prepped without NS3 inhibitor.

4.5 Methods for construction and validation of VEVs

We used second generation lentiviral components in the construction of our VEV system. This included fusogenic VSV-G, a gift from Bob Weinberg (Addgene #8454) [252]. Packaging plasmid psPAX2 was a gift from Didier Trono (Addgene #12260). The sequence of gp41-1 was reported in section 3.2.3., the Cre recombinase used was reported in section 3.3.4, the membrane binding domains and minimal Gal4 were reported in Chapter 2 and standard Gibson cloning processes were used in the construction of the membrane linked constructs. All constructs were confirmed via sanger sequencing (Eton Bioscience). A clonal line of HEK293FT expressing UAS driven H2B-mCherry was produced using the methods described in Chapter 2, and the plasmid was produced in house from sequences from Michael Elowitz (Caltech). Additionally, the Cre stoplight clonal line reported in Chapter 3.3.4 had a constitutive Ctein(gp41-1)-BFP clonally integrated using the same methods for VEV experiments.

Production of VEVs was performed in low passage HEK293FT cells without any other integrated reporters or constructs (ThermoFisher). Cells were transiently transfected with 800 ng VSV-G plasmid and 550 ng Cre construct per million cells, 24 hours incubated on bovine plasma fibronectin (Product #F1141, Sigma-Aldrich) and confluent in a 6 well plate. Cells were incubated for 24 hours in transfection mixture before removal and refresh of media using OptiMEM (Gibco). 48 hours post media refresh, the media was collected and concentrated overnight by centrifugation at 8,000 x g and 4°C in a temperature-controlled microcentrifuge (Thermo Scientific). The supernatant was removed by dumping and the pellet (not always visible) was eluted in 60 uL OptiMEM.

See Takara Bioscience gesicle production manual for additional details [253]. Multiple isolations were collected and mixed to create a consistent concentration of VEVs for downstream applications. This was stored at -80°C unless used the same day. This solution was used at 1 uL per 96 well unless reported otherwise. These were compared to a commercial prep of VEVs Cre Recombinase Gesicles (Takara 631449).

For experiments where intein constructs were transiently expressed in HEK293T cells, 20 ng of each construct was used per 120,000 cells transfected unless reported otherwise. 30,000 cells were plated per 96 well. Additionally, 10 ng of a Lifeact-BFP transcription marker was co-transfected, a gift from Michael Davidson (Addgene plasmid #54602) [254]. These cells were transfected in suspension, plated and incubated for 24 h prior to analysis via imaging or flow cytometry. In the case of Cre recombinase expression, the cells were incubated for 48 h prior to analysis to allow for recombination and expression of GFP.

Images were taken by epifluorescence microscopy in optically clear plastic bottomed 8-wells (ibidi). Fluorescence images were contrasted uniformly across each experiment and analyzed with the Image-J based image analysis package Fiji.

Flow cytometry was carried out on an Attune NxT (ThermoFisher). Cells were gated for live cells via scatter. Additionally, cells were gated for transfection using the top 1% BFP signal of non-transfected cells. UAS driven H2B-mCherry were gated for expression of mCherry via the top 1% of non-transfected clonal reporters, and the percentage falling into this category of BFP producing live cells were reported. Cre stoplight reporter cells were analyzed for GFP production and the geometric mean AFU

was reported or gated for GFP production via the top 1% of naïve cells and the percent active was reported.

Primary antibodies used: anti-VSV-G-tag (abcam, ab183497, diluted 1:10,000), anti-Cre (EMD Milipore 69050-3, diluted 1:3000), and rat anti-Gapdh HRP conjugate (BioLegend 607904). Western blots were performed in this order with stripping via Restore PLUS Western Blot Stripping Buffer (ThermoFisher) between stains, and distinct signals and bands were recovered with every stain. The following secondary antibody was used: anti-rabbit HRP conjugate (BioRad; 170-6515; 1:3000 dilution). GelCode Blue stain was used to detect protein concentration (ThermoFisher).

4.6 Discussion

In this chapter, we have demonstrated an intein-based mechanism for restricting tropism of VEVs, an otherwise widely trophic tool for delivering cargo to cellular systems. We first demonstrated inteins as a robust system for translationally limiting release of diverse cargo proteins from a variety of membrane tethers. We further showed that the use of the MP tethering moiety allowed these cargos to be packaged into VEVs, were deliverable to specified targets in a heterogeneous population of cells, and the cargo release was dependent on intein activity. We found that the Cre recombinase cargo was packaged into VEVs at a concentration of 11ng/μL, which was a fraction of the packaging efficiency of commercial Cre recombinase carrying fusogenic vesicles from Takarabio. We also demonstrate that the NS3 protease based LInC system can be employed as a drug dependent cargo release mechanism.

We have shown that this system is modular in the cargo and fusogenic envelope

protein components. We expect that the flexibility of this system extends beyond the packaging of proteins to include ribonucleoproteins such as Cas9 and even nucleic acids through the use of protein sequences such as HUH tags which form a DNA-protein covalent bond [255]. We anticipate that new cargos would not require extensive testing to achieve targeted delivery as no change to the delivery mechanism would be required for different cargos or different targets. This makes this system truly plug-and-play for implementation in a variety of contexts.

We hypothesize that it would be efficiently incorporated by a number of micro-vesicle forming viral envelope proteins, and therefore incorporated in many virus-like particles (VLPs). We expect this flexibility due to the consistent and significant expression of MP conjugated constructs at the plasma membrane, the site of viral genesis for many viruses. In addition to increasing the flexibility of this system, the ability to use it with different VLPs would allow the implementation of an AND gate, through restrictive tropism both of the viral entry and the key intein expression. This could be invaluable if the application requires precise delivery of the cargo protein.

The Liu group recently released an alternate method for restricting the tropism of VEVs using an engineered VSV-G that retains some fusogenic activity, however it is no longer able to mediate the plasma membrane targeting on its own [256]. They employ other receptors and targeting moieties instead, and show that the VEV is still able to deliver cargo but only in the targeted cell types. This additional strategy for restricting VEV entry shows how essential targeting can be in the delivery of therapeutic proteins. Their strategy is valuable because it requires no change to the native context. However, it

does not allow for targeting of specifically engineered cells out of all the cells in a native environment. Fortunately, this is an alternate type of VEV that our intein-lock system can be employed in, allowing researchers to take advantage of both targeting mechanisms while still retaining the powerful fusogenic activity of VSV-G.

Not only is this system modular, it retains orthogonality in mammalian contexts. There is little concern for endogenous overlap with the intein release mechanism. While the gp41-1 intein was found in a metagenomic sample of ocean water, the sequence is unique and has not been mapped to any known genome [257]. It has been found to be orthogonal to a body of known split inteins, suggesting that endogenous inteins or similar structures would not be cross-reactive [258]. This suggests that the system would be suitable for use in therapeutic protein delivery as well as in the laboratory.

Since there is such a body of small and efficient peptide splicers, it follows that a number of orthogonal VEVs could be developed and delivered at the same time, triggering release only in the target cells expressing the correct intein key. This would allow for multiplexed cargo delivery using a single VEV containing multiple locked cargos, and we hope to see others develop this technology further.

The main drawback of the system is the requirement of expression of a recombinant protein for release of the cargo in the target cell. This makes it difficult to employ in traditional therapeutic contexts, where the endogenous cells would not be expressing the necessary Ctein key. However this system is appropriate and powerful in that it can target specifically engineered cells, even when there are endogenous cells that are extremely similar. This would allow it to target CAR-T cells used in cancer therapy,

or cells that have been delivered therapeutic proteins in an earlier VLP via introduction of the intein key at that time [9,259].

In conclusion, we have developed a flexible and novel control system for delivery with EVs. The strength of the system is in its modularity. The cargo protein is extremely modular, allowing the user to charge VEVs with any protein or protein-conjugated structure desired. Exchanging fusogenic envelope proteins allows for changes in targeting. Changing the lock-and-key inteins allows for multiplexed delivery. We hope that the plug-and-play nature, as well as the robust and specific cargo delivery will make this a useful biological tool delivery system for a wide variety of contexts.

5. Conclusions and Future Perspectives

In this work, we have demonstrated the use of HCV protease NS3, light sensitive domain PhoCl, and small inteins gp41-1 and gp41-8 as protein modulation tools through their governance of protein excision and splicing. Protease NS3 was used to confer drug quenchable activity to a TF, and thereby install drug regulated gene expression in mammalian cells. Intein activity was used to change the association dynamics of GFP and indicates an area for further study, as well as being tested in drug inducible systems using NS3 and rapamycin as well as light inducible systems using PhoCl. Finally, gp41-1 intein was used as a biological lock and key mechanism for the orthogonal release of proteins from viral capsid packages called VEVs. Throughout this work, we have focused on the use of protein domains orthogonal to mammalian biology such that these systems could be trivially implemented in therapeutics. We have also focused on autonomous protein domains that can be used in a plug-and-play manner to decrease the need for downstream development time. There are many indications that these strategies will form the basis for next generation protein engineering development.

NS3 HCV protease is a powerful tool to confer drug dependent activities to protein systems. This is highlighted by the recent work producing a protein dimerizer system using NS3 as a docking protein for multiple inhibitory drugs, including danaprevir and grazoprevir [125]. This group then evolved protein binding elements specific for NS3 bound to these specific inhibitory drugs, and this system is independent of NS3 activity, in fact they use a catalytically dead S139A mutation. Other systems have focused on the protease activity of NS3, such as with CHOMP, a system of multi-protein

control using a series of proteases and protease inhibitors to confer complex system regulation [129]. The key factor that enables all of these developments is the body of varied inhibitor compounds available for NS3, and the numbers continue to grow [260]. This allows researchers the choice in compound, such that if a particular drug is unsuitable for their application, another one can be chosen. This flexibility is likely to give rise to multiple implemented protein-based technologies where NS3 confers drug inducible activity in the near future.

Inteins are powerful tools for controlling protein structure, and through structure, function. This was clearly illustrated in the gp41-1 circularization scheme where intein splicing-dependent polymerization destroyed GFP fluorescence despite significant expression. This brings up new questions about the dynamics of translation, GFP maturation, and intein splicing activity, that would be exciting to explore in the future using tools such as cell-free systems [261]. We also tested inteins in drug inducible as well as light inducible schemes, and while these systems were not very robust, it explores the experimental space available for the creation of inducible protein splicing systems. Existing reactive inteins from native sources may be able to be repurposed to detect ecological conditions such as pH, temperature, and concentration of reactive oxygen species due to their existing features and significant autonomy [190]. Additionally, there seems to be a need for a proximity induced splicing intein that is active in mammalian systems. We believe that once this need is filled, production of drug induced, light induced, and other induction systems for protein splicing would be inconsequential. We hope to see successful efforts on this endeavor in the future.

Finally, we employed inteins as a biological lock to release packaged proteins in a highly specific manner. This is complementary to the methods through which inteins have been employed in protein engineering the most; to splice and release target proteins after isolation in a traceless manner, removing all protein isolation tags [228]. In this case the inteins are also splicing and releasing proteins, and the target cell must produce the intein release mechanism in order to benefit from the packaged protein. By similar principles, inteins are robust systems of regulation of protein release in unfamiliar contexts, and this context was found to be as easily taken advantage of in mammalian cells as in a test tube. This system can work with any VLPs, though it is most useful in a broadly trophic platform like VSV-G mediated fusogens. New advances in fusogenic VLPs have improved targeting by inactivating portions of the VSV-G complex, and we see advantages for using our platform in conjunction with this for specific targeting both via outer moieties and via inner release mechanism [256]. VLPs hold extraordinary promise for delivery of therapeutic proteins in a landscape where protein therapies are becoming more common [1]. We hope that this tool and others like it will assist in implementing VLPs in therapeutic contexts.

This work focuses on orthogonal and plug-and-play design elements, and finds these areas to be points of active development and research. This indicates that other researchers are working with these design features in mind. We hope that these tools can be effectively and efficiently implemented in therapeutic contexts in ways that improve the lives of patients.

BIBLIOGRAPHY

1. Usmani SS, Bedi G, Samuel JS, et al. THPdb: Database of FDA-approved peptide and protein therapeutics. PLoS One. 2017;12(7):e0181748. Published 2017 Jul 31. doi:10.1371/journal.pone.0181748

2. Benedini L. Advanced protein drugs and formulations [published online ahead of print, 2021 Dec 9]. Current Protein & Peptide Science. 2021;10.2174/1389203722666211210115040. doi:10.2174/1389203722666211210115040

3. Mushtaq N, Redpath MB, Luzio JP, Taylor PW. Prevention and cure of systemic Escherichia coli K1 infection by modification of the bacterial phenotype. Antimicrobial Agents and Chemotherapy. 2004;48(5):1503-1508. doi:10.1128/AAC.48.5.1503-1508.2004

4. van den Broek BTA, van Doorn J, Hegeman CV, et al. Hurdles in treating Hurler disease: potential routes to achieve a "real" cure. Blood Advances. 2020;4(12): 2837-2849. doi:10.1182/bloodadvances.2020001708

5. Luthra R, Kaur S, Bhandari K. Applications of CRISPR as a potential therapeutic. Life Sciences. 2021;284:119908. doi:10.1016/j.lfs.2021.119908

6. Marston HD, Paules CI, Fauci AS. Monoclonal Antibodies for Emerging Infectious Diseases-Borrowing from History. New England Journal of Medicine. 2018;378(16):1469-1472. doi:10.1056/NEJMp1802256

7. Hwang YC, Lu RM, Su SC, et al. Monoclonal antibodies for COVID-19 therapy and SARS-CoV-2 detection. Journal of Biomed Science. 2022;29(1):1. Published 2022 Jan 4. doi:10.1186/s12929-021-00784-w

8. RECOVERY Collaborative Group. Casirivimab and imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet. 2022;399(10325):665-676. doi:10.1016/S0140-6736(22)00163-5

9. Fischer A. FDA approves car-T cell therapy to treat adults with certain types of large B-cell lymphoma. U.S. Food and Drug Administration. Published October 18, 2017. https://www.fda.gov/news-events/press-announcements/fda-approves-car-t-cell-therapy-treat-adults-certain-types-large-b-cell-lymphoma.

10. June CH, Sadelain M. Chimeric Antigen Receptor Therapy. New England Journal of Medicine. 2018;379(1):64-73. doi:10.1056/NEJMra1706169

11. Bai XC, Yan C, Yang G, et al. An atomic structure of human γ-secretase. Nature. 2015;525(7568):212-217. doi:10.1038/nature14892

12. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2

13. Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373(6557):871-876. doi:10.1126/science.abj8754

14. Hildebrand PW, Rose AS, Tiemann JKS. Bringing Molecular Dynamics Simulation Data into View. Trends in Biochemical Sciences. 2019;44(11):902-913. doi:10.1016/j.tibs.2019.06.004

15. Frank J, Ourmazd A. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. Methods. 2016;100:61-67. doi:10.1016/j.ymeth.2016.02.007

16. Murata M, Narahara S, Kawano T, et al. Design and Function of Engineered Protein Nanocages as a Drug Delivery System for Targeting Pancreatic Cancer Cells via Neuropilin-1. Molecular Pharmaceutics. 2015;12(5):1422-1430. doi:10.1021/mp5007129

17. Dou J, Vorobieva AA, Sheffler W, et al. De novo design of a fluorescence-activating β-barrel. Nature. 2018;561(7724):485-491. doi:10.1038/s41586-018-0509-0

18. Schuster J, Koulov A, Mahler HC, et al. In Vivo Stability of Therapeutic Proteins [published correction appears in Pharm Res. 2020 Mar 12;37(3):68]. Pharmaceutical Research. 2020;37(2):23. Published 2020 Jan 3. doi:10.1007/s11095-019-2689-1

19. Gilbert W, Müller-Hill B. The lac operator is DNA. Proceedings of the National Academy of Sciences of the United States of America. 1967;58(6):2415-2421. doi:10.1073/pnas.58.6.2415

20. Dickson RC, Abelson J, Barnes WM, Reznikoff WS. Genetic regulation: the Lac control region. Science. 1975;187(4171):27-35. doi:10.1126/science.1088926

21. Marians KJ, Wu R, Stawinski J, Hozumi T, Narang SA. Cloned synthetic lac operator DNA is biologically active. Nature. 1976;263(5580):744-748. doi:10.1038/263744a0

22. Heyneker HL, Shine J, Goodman HM, et al. Synthetic lac operator DNA is functional in vivo. Nature. 1976;263(5580):748-752. doi:10.1038/263748a0

23. Polisky B, Bishop RJ, Gelfand DH. A plasmid cloning vehicle allowing regulated expression of eukaryotic DNA in bacteria Proceedings of the National Academy of Sciences of the United States of America. 1976;73(11):3900-3904. doi:10.1073/pnas.73.11.3900

24. Itakura K, Hirose T, Crea R, et al. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. Science. 1977;198(4321):1056-1063. doi:10.1126/science.412251

25. Smith GE, Summers MD, Fraser MJ. Production of human beta interferon in insect cells infected with a baculovirus expression vector. Molecular and Cellular Biology. 1983;3(12):2156-2165. doi:10.1128/mcb.3.12.2156-2165.1983

26. Emery VC, Bishop DH. The development of multiple expression vectors for high level synthesis of eukaryotic proteins: expression of LCMV-N and AcNPV polyhedrin protein by a recombinant baculovirus. Protein Engineering. 1987;1(4): 359-366. doi:10.1093/protein/1.4.359

27. Galloway JA. Insulin treatment for the early 80s: facts and questions about old and new insulins and their usage. Diabetes Care. 1980;3(5):615-622. doi:10.2337/diacare.3.5.615

28. Lenzi RN, Altevogt BM, Gostin LO, Committee on the Independent Review and Assessment of the Activities of the NIH Recombinant DNA Advisory Committee; Board on Health Sciences Policy; Institute of Medicine, eds. Oversight and Review of Clinical Gene Transfer Protocols: Assessing the Role of the Recombinant DNA Advisory Committee. Washington (DC): National Academies Press (US); March 27, 2014.

29. Slaff J, Jacobson D, Tillman CR, Curington C, Toskes P. Protease-specific suppression of pancreatic exocrine secretion. Gastroenterology. 1984;87(1):44-52.

30. Salomon F, Cuneo RC, Hesp R, Sönksen PH. The effects of treatment with recombinant human growth hormone on body composition and metabolism in adults with growth hormone deficiency. New England Journal of Medicine. 1989;321(26):1797-1803. doi:10.1056/NEJM198912283212605

31. Rosado JL, Solomons NW, Lisker R, Bourges H. Enzyme replacement therapy for primary adult lactase deficiency. Effective reduction of lactose malabsorption and milk intolerance by direct addition of beta-galactosidase to milk at mealtime. Gastroenterology. 1984;87(5):1072-1082.

32. Brown A, Hughes M, Tenner S, Banks PA. Does pancreatic enzyme supplementation reduce pain in patients with chronic pancreatitis: a meta-analysis. American Journal of Gastroenterology. 1997;92(11):2032-2035.

33. Leader B, Baca QJ, Golan DE. Protein therapeutics: a summary and pharmacological classification. Nature Reviews. Drug Discovery. 2008;7(1):21-39. doi:10.1038/nrd2399

34. Veronese FM, Mero A. The impact of PEGylation on biological therapies. BioDrugs. 2008;22(5):315-329. doi:10.2165/00063030-200822050-00004

35. Dimitrov DS. Therapeutic proteins. Methods in Molecular Biology. 2012;899:1-26. doi:10.1007/978-1-61779-921-1_1

36. Albers GW, Bates VE, Clark WM, Bell R, Verro P, Hamilton SA. Intravenous tissue-type plasminogen activator for treatment of acute stroke: the Standard Treatment with Alteplase to Reverse Stroke (STARS) study. JAMA: The Journal of the American Medical Association. 2000;283(9):1145-1150. doi:10.1001/jama.283.9.1145

37. Assessment of the Safety and Efficacy of a New Thrombolytic Regimen (ASSENT)-3 Investigators. Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: the ASSENT-3 randomised trial in acute myocardial infarction. Lancet. 2001;358(9282):605-613. doi:10.1016/S0140-6736(01)05775-0

38. Hutchings MI, Truman AW, Wilkinson B. Antibiotics: past, present and future. Current Opinion in Microbiology. 2019;51:72-80. doi:10.1016/j.mib.2019.10.008

39. FDA approves asciminib. US Food and Drug Association. Published October 29, 2021. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-asciminib-philadelphia-chromosome-positive-chronic-myeloid-leukemia

40. Blair HA. Sotorasib: First Approval [published correction appears in Drugs. 2021 Nov;81(16):1947]. Drugs. 2021;81(13):1573-1579. doi:10.1007/s40265-021-01574-2

41. Coronavirus (COVID-19) Update: FDA Authorizes First Oral Antiviral for Treatment of COVID-19. US Food and Drug Association. Published December 22, 2021. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19

42. Hammond J, Leister-Tebbe H, Gardner A, et al. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19 [published online ahead of print, 2022 Feb 16]. New England Journal of Medicine. 2022;NEJMoa2118542. doi:10.1056/NEJMoa2118542

43. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 2001;46(1-3):3-26. doi:10.1016/s0169-409x(00)00129-0

44. Shultz MD. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. Journal of Medicinal Chemistry. 2019;62(4):1701-1714. doi:10.1021/acs.jmedchem.8b00686

45. National Center for Biotechnology Information. PubChem Compound Summary for CID 118984375, Insulin human. Accessed Apr. 10, 2022. https://pubchem.ncbi.nlm.nih.gov/compound/Insulin-human.

46. Czuba E, Diop M, Mura C, et al. Oral insulin delivery, the challenge to increase insulin bioavailability: Influence of surface charge in nanoparticle system. International Journal of Pharmaceutics. 2018;542(1-2):47-55. doi:10.1016/j.ijpharm.2018.02.045

47. Kotelchuck D, Scheraga HA. The influence of short-range interactions on protein onformation. II. A model for predicting the alpha-helical regions of proteins. Proceedings of the National Academy of Sciences of the United States of America. 1969;62(1):14-21. doi:10.1073/pnas.62.1.14

48. Di L, Kerns EH. Stability challenges in drug discovery. Chemistry & Biodiversity. 2009;6(11):1875-1886. doi:10.1002/cbdv.200900061

49. Carter PJ. Introduction to current and future protein therapeutics: a protein engineering perspective. Experimental Cell Research. 2011;317(9):1261-1269. doi:10.1016/j.yexcr.2011.02.013

50. Butreddy A, Janga KY, Ajjarapu S, Sarabu S, Dudhipala N. Instability of therapeutic proteins - An overview of stresses, stabilization mechanisms and analytical techniques involved in lyophilized proteins. International Journal of Biological Macromolecules. 2021;167:309-325.

doi:10.1016/j.ijbiomac.2020.11.188

51. Chemistry Nobel Honors CRISPR, an "Essential" Tool for Cancer. Cancer Discovery. 2020 Dec;10(12):1782-1783. doi: 10.1158/2159-8290.CD-NB2020-090. Epub 2020 Oct 9. PMID: 33037027.

52. Papasavva P, Kleanthous M, Lederer CW. Rare Opportunities: CRISPR/Cas-Based Therapy Development for Rare Genetic Diseases. Molecular Diagnosis & Therapy. 2019;23(2):201-222. doi:10.1007/s40291-019-00392-3

53. Vertex and CRISPR therapeutics present new data in 22 patients with greater than 3 months follow-up post-treatment with investigational CRISPR/cas9 gene-editing therapy, CTX001™ at European Hematology Association annual meeting. Vertex Pharmaceuticals. Published June 11, 2021. https://investors.vrtx.com/news-releases/news-release-details/vertex-and-crispr-therapeutics-present-new-data-22-patients.

54. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology. 2013;31(9): 822-826. doi:10.1038/nbt.2623

55. Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Research. 2014;24(1): 132-141. doi:10.1101/gr.162339.113

56. Guilinger JP, Pattanayak V, Reyon D, et al. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. Nature Methods. 2014;11(4):429-435. doi:10.1038/nmeth.2845

57. Naeem M, Majeed S, Hoque MZ, Ahmad I. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. Cells. 2020;9(7):1608. Published 2020 Jul 2. doi:10.3390/cells9071608

58. Wüst RCI, Houtkooper RH, Auwerx J. Confounding factors from inducible systems for spatiotemporal gene expression regulation. Journal of Cell Biology. 2020;219(7):e202003031. doi:10.1083/jcb.202003031

59. Kolos JM, Voll AM, Bauder M, Hausch F. FKBP Ligands-Where We Are and Where to Go? Frontiers in Pharmacology. 2018;9:1425. Published 2018 Dec 5. doi:10.3389/fphar.2018.01425

60. Brudno JN, Kochenderfer JN. Toxicities of chimeric antigen receptor T cells: recognition and management. Blood. 2016;127(26):3321-3330.

doi:10.1182/blood-2016-04-703751

61. Zetsche B, Volz SE, Zhang F. A split-Cas9 architecture for inducible genome editing and transcription modulation. Nature Biotechnology. 2015;33(2):139-142. doi:10.1038/nbt.3149

62. Nguyen DP, Miyaoka Y, Gilbert LA, et al. Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. Nature Communications. 2016;7:12009. Published 2016 Jul 1. doi:10.1038/ncomms12009

63. Li J, Kim SG, Blenis J. Rapamycin: one drug, many effects. Cell Metabolism. 2014;19(3):373-379. doi:10.1016/j.cmet.2014.01.001

64. Rivera VM, Berk L, Clackson T. Dimerizer-mediated regulation of gene expression in vivo. Cold Spring Harbor Protocols. 2012;2012(7):821-824. Published 2012 Jul 1. doi:10.1101/pdb.prot070144

65. Voß S, Klewer L, Wu YW. Chemically induced dimerization: reversible and spatiotemporal control of protein function in cells. Current Opinion in Chemical Biology. 2015;28:194-201. doi:10.1016/j.cbpa.2015.09.003

66. Ahler E, Sullivan WJ, Cass A, et al. Doxycycline alters metabolism and proliferation of human cell lines. PLoS One. 2013;8(5):e64561. Published 2013 May 31. doi:10.1371/journal.pone.0064561

67. Freiss G, Rochefort H, Vignon F. Mechanisms of 4-hydroxytamoxifen anti-growth factor activity in breast cancer cells: alterations of growth factor receptor binding sites and tyrosine kinase activity. Biochemical and Biophysical Research Communications. 1990;173(3):919-926. doi:10.1016/s0006-291x(05)80873-3

68. Bayle JH, Grimley JS, Stankunas K, Gestwicki JE, Wandless TJ, Crabtree GR. Rapamycin analogs with differential binding specificity permit orthogonal control of protein activity. Chemistry & Biology. 2006;13(1):99-107. doi:10.1016/j.chembiol.2005.10.017

69. Schreiber KH, Arriola Apelo SI, Yu D, et al. A novel rapamycin analog is highly selective for mTORC1 in vivo. Nature Communications. 2019;10(1):3194. Published 2019 Jul 19. doi:10.1038/s41467-019-11174-0

70. Peterson, A. FY22 Budget Request: National Institutes of Health. Aip.org. American Institute of Physics. Published 2022. https://www.aip.org/fyi/2021/fy22-budget-request-national-institutes-health

71. Azoulay P, Li D, Zivin JSG, Sampat BN. Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules. Review of Economic Studies. 2019;86(1):117-152. doi:10.1093/restud/rdy034

72. Turi JB. The Most Active Biotech Venture Investors in 2021. Crunchbase News. Published Jan 2022. https://news.crunchbase.com/news/most-active-biotech-investors-in-2021-ra-capital-orbimed.

73. Spikevax and Moderna Covid-19 vaccine. U.S. Food and Drug Administration. Updated March 29, 2022. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/spikevax-and-moderna-covid-19-vaccine.

74. Michelson A. How the coronavirus vaccine development compares to other shots in history. Business Insider. Published November 19, 2020. https://www.businessinsider.com/how-coronavirus-vaccine-development-compares-to-other-shots-in-history-2020-11.

75. Swint-Kruse L. Using Evolution to Guide Protein Engineering: The Devil IS in the Details. Biophysical Journal. 2016;111(1):10-18. doi:10.1016/j.bpj.2016.05.030

76. Köhler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature. 1975;256(5517):495-497. doi:10.1038/256495a0

77. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. Nature Biotechnology. 1997;15(6):553-557. doi:10.1038/nbt0697-553

78. Lee YJ, Jeong KJ. Challenges to production of antibodies in bacteria and yeast. Journal of Bioscience and Bioengineering. 2015;120(5):483-490. doi:10.1016/j.jbiosc.2015.03.009

79. Mahdavi SZB, Oroojalian F, Eyvazi S, et al. An overview on display systems (phage, bacterial, and yeast display) for production of anticancer antibodies; advantages and disadvantages [published online ahead of print, 2022 Mar 24]. International Journal of Biological Macromolecules. 2022;208:421-442. doi:10.1016/j.ijbiomac.2022.03.113

80. Wang L, Jackson WC, Steinbach PA, Tsien RY. Evolution of new nonantibody proteins via iterative somatic hypermutation. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(48):16745-16749. doi:10.1073/pnas.0407752101

81. Warszawski S, Borenstein Katz A, Lipsh R, et al. Correction: Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. PLoS Computational Biology. 2020;16(10):e1008382. Published 2020 Oct 21. doi:10.1371/journal.pcbi.1008382

82. Wong BG, Mancuso CP, Kiriakov S, Bashor CJ, Khalil AS. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nature Biotechnology. 2018;36(7):614-623. doi:10.1038/nbt.4151

83. Heins ZJ, Mancuso CP, Kiriakov S, Wong BG, Bashor CJ, Khalil AS. Designing Automated, High-throughput, Continuous Cell Growth Experiments Using eVOLVER. Journal of Visualized Experiments: JoVE. 2019;(147): 10.3791/59652. Published 2019 May 19. doi:10.3791/59652

84. Wang Y, Xue P, Cao M, Yu T, Lane ST, Zhao H. Directed Evolution: Methodologies and Applications. Chemical Reviews. 2021;121(20):12384-12444. doi:10.1021/acs.chemrev.1c00260

85. Goldstein B, King N. The Future of Cell Biology: Emerging Model Organisms. Trends in Cell Biology. 2016;26(11):818-824. doi:10.1016/j.tcb.2016.08.005

86. Kwong AD, Kim JL, Rao G, Lipovsek D, Raybuck SA. Hepatitis C virus NS3/4A protease. Antiviral Research. 1999;41(1):67-84.

87. Özen A, Prachanronarong K, Matthew AN, Soumana DI, Schiffer CA. Resistance outside the substrate envelope: hepatitis C NS3/4A protease inhibitors. Critical Reviews in Biochemistry and Molecular Biology. 2019;54(1):11-26. doi:10.1080/10409238.2019.1568962

88. FDA approves treatment for adults and children with all genotypes of hepatitis C and compensated cirrhosis that shortens duration of treatment to eight weeks. US Food and Drug Administration. Published September 26, 2019. https://www.fda.gov/news-events/press-announcements/fda-approves-treatment-adults-and-children-all-genotypes-hepatitis-c-and-compensated-cirrhosis

89. Nanda A, Nasker SS, Mehra A, Panda S, Nayak S. Inteins in Science: Evolution to Application. Microorganisms. 2020;8(12):2004. Published 2020 Dec 16. doi:10.3390/microorganisms8122004

90. Oeemig JS, Aranko AS, Djupsjöbacka J, Heinämäki K, Iwaï H. Solution structure of DnaE intein from Nostoc punctiforme: structural basis for the design of a new split intein suitable for site-specific chemical modification. FEBS Letters. 2009; 583(9):1451-1456. doi:10.1016/j.febslet.2009.03.058

91. Ramsden R, Arms L, Davis TN, Muller EG. An intein with genetically selectable markers provides a new approach to internally label proteins with GFP. BMC Biotechnology. 2011;11:71. Published 2011 Jun 27. doi:10.1186/1472-6750-11-71

92. Carvajal-Vallejos P, Pallissé R, Mootz HD, Schmidt SR. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. Journal of Biological Chemistry. 2012;287(34):28686-28696. doi:10.1074/jbc.M112.372680

93. Herrmann IK, Wood MJA, Fuhrmann G. Extracellular vesicles as a next-generation drug delivery platform. Nature Nanotechnology. 2021;16(7):748-759. doi:10.1038/s41565-021-00931-2

94. van Dongen HM, Masoumi N, Witwer KW, Pegtel DM. Extracellular Vesicles Exploit Viral Entry Routes for Cargo Delivery. Microbiology and Molecular Biology Reviews. 2016;80(2):369-386. Published 2016 Mar 2. doi:10.1128/MMBR.00063-15

95. Lienert F, Lohmueller JJ, Garg A, Silver PA. Synthetic biology in mammalian cells: next generation research tools and therapeutics. Nature Reviews. Molecular Cell Biology. 2014;15(2):95-107. doi:10.1038/nrm3738

96. Gossen M, Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(12):5547-5551. doi:10.1073/pnas.89.12.5547

97. Spencer DM, Wandless TJ, Schreiber SL, Crabtree GR. Controlling Signal Transduction with Synthetic Ligands. Science. 1993;262(5136):1019-1024. doi:10.1126/science.7694365

98. Goreshnik I, Maly DJ. A Small Molecule-Regulated Guanine Nucleotide Exchange Factor. Journal of the American Chemical Society. 2009;132(3):938-940. doi:10.1021/ja907886v

99. Davis KM, Pattanayak V, Thompson DB, Zuris JA, Liu DR. Small molecule–triggered Cas9 protein with improved genome-editing specificity. Nature Chemical Biology. 2015;11(5):316-318. doi:10.1038/nchembio.1793

100. Rose JC, Stephany JJ, Valente WJ, et al. Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. Nature Methods. 2017;14(9):891-896. doi:10.1038/nmeth.4368

101. Wu CY, Roybal KT, Puchner EM, Onuffer J, Lim WA. Remote control of therapeutic T cells through a small molecule–gated chimeric receptor. Science. 2015;350(6258). doi:10.1126/science.aab4077

102. Meganck RM, Baric RS. Developing therapeutic approaches for twenty-first-century emerging infectious viral diseases. Nature Medicine. 2021;27(3):401-410. doi:10.1038/s41591-021-01282-0

103. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018;359(6382):1350-1355. doi:10.1126/science.aar4060

104. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. Nature Reviews. Microbiology. 2020;18(3):125-138. doi:10.1038/s41579-019-0311-5

105. Majerová T, Novotný P. Precursors of Viral Proteases as Distinct Drug Targets. Viruses. 2021;13(10):1981. Published 2021 Oct 2. doi:10.3390/v13101981

106. Danino YM, Even D, Ideses D, Juven-Gershon T. The core promoter: At the heart of gene expression. Biochimica et Biophysica Acta. 2015;1849(8):1116-1131. doi:10.1016/j.bbagrm.2015.04.003

107. Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. Journal of Virology. 1993;67(7):3835-3844. doi:10.1128/jvi.67.7.3835-3844.1993

108. Lin MZ, Glenn JS, Tsien RY. A drug-controllable tag for visualizing newly synthesized proteins in cells and whole animals. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(22):7744-7749. doi:10.1073/pnas.0803060105

109. Palida SF, Butko MT, Ngo JT, et al. PKM , But Not PKC , Is Rapidly Synthesized and Degraded at the Neuronal Synapse. Journal of Neuroscience. 2015;35(20):7736-7749. doi:10.1523/jneurosci.0004-15.2015

110. Chung HK, Jacobs CL, Huo Y, et al. Tunable and reversible drug control of protein production via a self-excising degron. Nature Chemical Biology. 2015;11(9):713-720. doi:10.1038/nchembio.1869

111. Pu J, Chronis I, Ahn D, Dickinson BC. A Panel of Protease-Responsive RNA Polymerases Respond to Biochemical Signals by Production of Defined RNA Outputs in Live Cells. Journal of the American Chemical Society. 2015;137(51):

15996-15999. doi:10.1021/jacs.5b10290

112. Bell AM, Wagner JL, Barber KE, Stover KR. Elbasvir/Grazoprevir: A Review of the Latest Agent in the Fight against Hepatitis C. International Journal of Hepatology. 2016;2016:1-8. doi:10.1155/2016/3852126

113. Siebel C, Lendahl U. Notch Signaling in Development, Tissue Homeostasis, and Disease. Physiological Reviews. 2017;97(4):1235-1294. doi:10.1152/physrev.00005.2017

114. Kopan R, Ilagan MX. The canonical Notch signaling pathway: unfolding the activation mechanism. Cell. 2009;137(2):216-233. doi:10.1016/j.cell.2009.03.045

115. Ryan JA. Cell Cloning by Serial Dilution in 96 Well Plates Protocol. Corning Incorporated Life Sciences. Published August, 2008. https://www.corning.com/catalog/cls/documents/protocols/Single_cell_cloning_protocol.pdf.

116. McPhee F, Sheaffer AK, Friborg J, et al. Preclinical Profile and Characterization of the Hepatitis C Virus NS3 Protease Inhibitor Asunaprevir (BMS-650032). Antimicrobial Agents and Chemotherapy. 2012;56(10):5387-5396. doi:10.1128/AAC.01186-12

117. Lamarre D, Anderson PC, Bailey M, et al. An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus [published correction appears in Nature. 2003 Nov 20;246]. Nature. 2003;426(6963):186-189. doi:10.1038/nature02099

118. Seiwert SD, Andrews SW, Jiang Y, et al. Preclinical characteristics of the hepatitis C virus NS3/4A protease inhibitor ITMN-191 (R7227). Antimicrobial Agents and Chemotherapy. 2008;52(12):4432-4441. doi:10.1128/AAC.00699-08

119. Summa V, Ludmerer SW, McCauley JA, et al. MK-5172, a selective inhibitor of hepatitis C virus NS3/4a protease with broad activity across genotypes and resistant variants [published correction appears in Antimicrob Agents Chemother. 2014 Aug;58(8):4995. Huang, Qian [added]]. Antimicrobial Agents and Chemotherapy. 2012;56(8):4161-4167. doi:10.1128/AAC.00324-12

120. Center for Drug Evaluation and Research. Pharmacology/Toxicology NDA Review and Evaluation (Application Number 208,261). U.S. Food & Drug Administration. Published May 28. 2015. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2016/208261Orig1s000PharmR.pdf

121. Ng TI, Tripathi R, Reisch T, et al. In VitroAntiviral Activity and Resistance Profile of the Next-Generation Hepatitis C Virus NS3/4A Protease Inhibitor Glecaprevir. Antimicrobial Agents and Chemotherapy. 2017;62(1):e01620-17. Published 2017 Dec 21. doi:10.1128/AAC.01620-17

122. Loetscher P, Pratt G, Rechsteiner M. The C terminus of mouse ornithine decarboxylase confers rapid degradation on dihydrofolate reductase. Support for the pest hypothesis. Journal of Biological Chemistry. 1991;266(17):11213-11220.

123. Li X, Zhao X, Fang Y, et al. Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. Journal of Biological Chemistry. 1998; 273(52):34970-34975. doi:10.1074/jbc.273.52.34970

124. Rajagopalan R, Misialek S, Stevens SK, et al. Inhibition and binding kinetics of the hepatitis C virus NS3 protease inhibitor ITMN-191 reveals tight binding and slow dissociative behavior. Biochemistry. 2009;48(11):2559-2568. doi:10.1021/bi900038p

125. Foight GW, Wang Z, Wei CT, et al. Multi-input chemical control of protein dimerization for programming graded cellular responses. Nature Biotechnology. 2019;37(10):1209-1216. doi:10.1038/s41587-019-0242-8

126. Bakulin I. NS3 protease inhibitors for treatment of chronic hepatitis C: Efficacy and safety. World Journal of Hepatology. 2014;6(5):326. doi:10.4254/wjh.v6.i5.326

127. Meewan I, Zhang X, Roy S, et al. Discovery of New Inhibitors of Hepatitis C Virus NS3/4A Protease and Its D168A Mutant. ACS Omega. 2019;4(16):16999-17008. doi:10.1021/acsomega.9b02491

128. Jacobs CL, Badiee RK, Lin MZ. StaPLs: versatile genetically encoded modules for engineering drug-inducible proteins. Nature Methods. 2018;15(7):523-526. doi:10.1038/s41592-018-0041-z

129. Gao XJ, Chong LS, Kim MS, Elowitz MB. Programmable protein circuits in living cells. Science. 2018;361(6408):1252-1258. doi:10.1126/science.aat5062

130. Vlahos AE, Kang J, Aldrete CA, et al. Protease-controlled secretion and display of intercellular signals. Nature Communications. 2022;13(1):912. Published 2022 Feb 17. doi:10.1038/s41467-022-28623-y

131. Dyer RP, Weiss GA. Making the cut with protease engineering. Cell Chemical Biology. 2022;29(2):177-190. doi:10.1016/j.chembiol.2021.12.001

132. Fink T, Jerala R. Designed protease-based signaling networks [published online ahead of print, 2022 Apr 13]. Current Opinion in Chemical Biology. 2022;68: 102146. doi:10.1016/j.cbpa.2022.102146

133. Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M. Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. Nucleic Acids Research. 1994 Apr 11;22(7): 1125-1127. doi: 10.1093/nar/22.7.1125

134. Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E. Inteins: Structure, Function, and Evolution. Annual Review of Microbiology. 2002;56(1): 263-287. doi:10.1146/annurev.micro.56.012302.160741

135. Bowman EJ, Tenney K, Bowman BJ. Isolation of genes encoding the Neurospora vacuolar ATPase. Analysis of vma-1 encoding the 67-kDa subunit reveals homology to other ATPases. Journal of Biological Chemistry. 1988 Oct 5; 263(28):13994-4001

136. Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae. Journal of Biological Chemistry. 1990 Apr 25;265(12):6726-33.

137. Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science. 1990;250(4981):651-657. doi:10.1126/science.2146742

138. Zhang B, Rapolu M, Liang Z, Han Z, Williams PG, Su WW. A Dual-Intein Autoprocessing Domain that Directs Synchronized Protein Co-Expression in Both Prokaryotes and Eukaryotes. Scientific Reports. 2015;5(1). doi:10.1038/srep08541

139. Ramsden R, Arms L, Davis TN, Muller EG. An intein with genetically selectable markers provides a new approach to internally label proteins with GFP. BMC Biotechnology. 2011;11(1). doi:10.1186/1472-6750-11-71

140. Li J, Sun W, Wang B, Xiao X, Liu XQ. ProteinTrans-Splicing as a Means for Viral Vector-MediatedIn VivoGene Therapy. Human Gene Therapy. 2008;19(9): 958-964. doi:10.1089/hum.2008.009

141. Tornabene P, Trapani I, Minopoli R, et al. Intein-mediated protein trans-splicing expands adeno-associated virus transfer capacity in the retina. Science

Translational Medicine. 2019;11(492). doi:10.1126/scitranslmed.aav4523

142. Chen L, Pradhan S, Evans TC. Herbicide resistance from a divided EPSPS protein: the split Synechocystis DnaE intein as an in vivo affinity domain. Gene. 2001;263(1-2):39-48. doi:10.1016/s0378-1119(00)00568-0

143. Jillette N, Du M, Zhu JJ, Cardoz P, Cheng AW. Split selectable markers. Nature Communications. 2019;10(1). doi:10.1038/s41467-019-12891-2

144. Palanisamy N, Degen A, Morath A, et al. Author Correction: Split intein-mediated selection of cells containing two plasmids using a single antibiotic. Nature Communications. 2020;11(1):276. Published 2020 Jan 9. doi:10.1038/s41467-019-13716-y

145. Mathys S, Evans TC, Chute IC, et al. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene. 1999; 231(1-2):1-13. doi:10.1016/s0378-1119(99)00103-1

146. Tan G, Chen M, Foote C, Tan C. Temperature-Sensitive Mutations Made Easy: Generating Conditional Mutations by Using Temperature-Sensitive Inteins That Function Within Different Temperature Ranges. Genetics. 2009;183(1):13-22. doi:10.1534/genetics.109.104794

147. Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M. A genetic system yields self-cleaving inteins for bioseparations. Nature Biotechnology. 1999;17(9):889-892. doi:10.1038/12879

148. Paulus H. Protein Splicing and Related Forms of Protein Autoprocessing. Annual Review of Biochemistry. 2000;69(1):447-496. doi:10.1146/annurev.biochem.69.1.447

149. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. Modulation of intein activity by its neighboring extein substrates. Proceedings of the National Academy of Sciences of the United States of America 2009;106(27):11005-11010. doi:10.1073/pnas.0904366106

150. Cheriyan M, Pedamallu CS, Tori K, Perler F. Faster Protein Splicing with the Nostoc punctiforme DnaE Intein Using Non-native Extein Residues. Journal of Biological Chemistry. 2013;288(9):6202-6211. doi:10.1074/jbc.m112.433094

151. Stevens AJ, Sekar G, Shah NH, Mostafavi AZ, Cowburn D, Muir TW. A promiscuous split intein with expanded protein engineering applications. Proceedings of the National Academy of Sciences of the United States of

America. 2017;114(32):8538-8543. doi:10.1073/pnas.1701083114

152. Townend JE, Tavassoli A. Traceless Production of Cyclic Peptide Libraries in E. coli. ACS Chemical Biology. 2016;11(6):1624-1630. doi:10.1021/acschembio.6b00095

153. Bhagawati M, Terhorst TME, Füsser F, et al. A mesophilic cysteine-less split intein for protein trans-splicing applications under oxidizing conditions. Proceedings of the National Academy of Sciences of the United States of America. 2019;116(44):22164–22172. https://doi.org/10.1073/pnas.1909825116

154. Carvajal-Vallejos P, Pallissé R, Mootz HD, Schmidt SR. Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources. Journal of Biological Chemistry. 2012;287(34):28686-28696. doi:10.1074/jbc.m112.372680

155. Pinto F, Thornton EL, Wang B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. Nature Communications. 2020;11(1). doi:10.1038/s41467-020-15272-2

156. Lim CKW, Gapinske M, Brooks AK, et al. Treatment of a Mouse Model of ALS by In Vivo Base Editing. Molecular Therapy. 2020;28(4):1177-1189. doi:10.1016/j.ymthe.2020.01.005

157. Purde V, Kudryashova E, Heisler DB, Shakya R, Kudryashov DS. Intein-mediated cytoplasmic reconstitution of a split toxin enables selective cell ablation in mixed populations and tumor xenografts. Proceedings of the National Academy of Sciences of the United States of America. 2020;117(36):22090-22100. doi:10.1073/pnas.2006603117

158. Heo JM, Rutter J. Ubiquitin-dependent mitochondrial protein degradation. International Journal of Biochemistry & Cell Biology. 2011;43(10):1422-1426. doi:10.1016/j.biocel.2011.06.002

159. Banaszynski LA, Chen L chun, Maynard-Smith LA, Ooi AGL, Wandless TJ. A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. Cell. 2006;126(5):995-1004. doi:10.1016/j.cell.2006.07.025

160. Dougan DA, Micevski D, Truscott KN. The N-end rule pathway: From recognition by N-recognins, to destruction by AAA+proteases. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2012;1823(1):83-91. doi:10.1016/j.bbamcr.2011.07.002

161. Shemorry A, Hwang CS, Varshavsky A. Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. Molecular Cell. 2013;50(4):540-551. doi:10.1016/j.molcel.2013.03.018

162. Conlan BF, Gillon AD, Craik DJ, Anderson MA. Circular proteins and mechanisms of cyclization. Biopolymers. 2010;94(5):573-583. doi:10.1002/bip.21422

163. Craik DJ, Daly NL, Bond T, Waine C. Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. Journal of Molecular Biology. 1999;294(5):1327-1336. doi:10.1006/jmbi.1999.3383

164. Satoh T, Li S, Friedman TM, Wiaderkiewicz R, Korngold R, Huang Z. Synthetic Peptides Derived from the Fourth Domain of CD4 Antagonize CD4 Function and Inhibit T Cell Activation. Biochemical and Biophysical Research Communications. 1996;224(2):438-443. doi:10.1006/bbrc.1996.1045

165. Poth AG, Chan LY, Craik DJ. Cyclotides as grafting frameworks for protein engineering and drug design applications. Biopolymers. 2013;100(5):480-491. doi:10.1002/bip.22284

166. Tavassoli A, Benkovic SJ. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. Nature Protocols. 2007;2(5):1126-1133. doi:10.1038/nprot.2007.152

167. Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrokovski S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. Nucleic Acids Research. 2009;37(8):2560-2573. doi:10.1093/nar/gkp095

168. Aranko AS, Wlodawer A, Iwaï H. Nature's recipe for splitting inteins. Protein Engineering, Design & Selection. 2014;27(8):263-271. doi:10.1093/protein/gzu028

169. Beyer HM, Mikula KM, Li M, Wlodawer A, Iwaï H. The crystal structure of the naturally split gp41-1 intein guides the engineering of orthogonal split inteins fromcis-splicing inteins. The FEBS Journal. 2019;287(9):1886-1898. doi:10.1111/febs.15113

170. Corish P, Tyler-Smith C. Attenuation of green fluorescent protein half-life in mammalian cells. Protein Engineering, Design and Selection. 1999;12(12):1035-1040. doi:10.1093/protein/12.12.1035

171. Li X, Zhao X, Fang Y, et al. Generation of destabilized green fluorescent protein as a transcription reporter. Journal of Biological Chemistry. 1998;273(52):34970-34975. doi:10.1074/jbc.273.52.34970

172. Bercovich Z, Rosenberg-Hasson Y, Ciechanover A, Kahana C. Degradation of ornithine decarboxylase in reticulocyte lysate is ATP-dependent but ubiquitin-independent. Journal of Biological Chemistry. 1989;264(27):15949-15952.

173. Murakami Y, Matsufuji S, Kameji T, et al. Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. Nature. 1992;360(6404):597-599. doi:10.1038/360597a0

174. Tsvetkov P, Reuven N, Shaul Y. Ubiquitin-independent p53 proteasomal degradation. Cell Death & Differentiation. 2009;17(1):103-108. doi:10.1038/cdd.2009.67

175. Erales J, Coffino P. Ubiquitin-independent proteasomal degradation. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2014;1843(1):216-221. doi:10.1016/j.bbamcr.2013.05.008

176. Shah NH, Eryilmaz E, Cowburn D, Muir TW. Naturally Split Inteins Assemble through a “Capture and Collapse” Mechanism. Journal of the American Chemical Society. 2013;135(49):18673-18681. doi:10.1021/ja4104364

177. Carrington JC, Dougherty WG. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. Proceedings of the National Academy of Sciences of the United States of America. 1988;85(10):3391-3395. doi:10.1073/pnas.85.10.3391

178. Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. Release of Proteins and Peptides from Fusion Proteins Using a Recombinant Plant Virus Proteinase. Analytical Biochemistry. 1994;216(2):413-417. doi:10.1006/abio.1994.1060

179. Kapust RB, Waugh DS. Controlled Intracellular Processing of Fusion Proteins by TEV Protease. Protein Expression and Purification. 2000;19(2):312-318. doi:10.1006/prep.2000.1251

180. Nallamsetty, S., Kapust, R. B., Tözsér, J., Cherry, S., Tropea, J. E., Copeland, T. D., & Waugh, D. S. (2004). Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. Protein Expression and Purification, 38(1), 108–115. https://doi.org/10.1016

181. Wang H, Liu J, Yuet KP, Hill AJ, Sternberg PW. Split cGAL, an intersectional strategy using a split intein for refined spatiotemporal transgene control in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(15):3900-3905. doi:10.1073/pnas.1720063115

182. To TL, Piggott BJ, Makhijani K, Yu D, Jan YN, Shu X. Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(11):3338-3343. doi:10.1073/pnas.1502857112

183. Donahue PS, Draut JW, Muldoon JJ, Edelstein HI, Bagheri N, Leonard JN. The COMET toolkit for composing customizable genetic programs in mammalian cells. Nature Communications. 2020;11(1):779. Published 2020 Feb 7. doi:10.1038/s41467-019-14147-5

184. Kennedy MJ, Hughes RM, Peteya LA, Schwartz JW, Ehlers MD, Tucker CL. Rapid blue-light-mediated induction of protein interactions in living cells. Nature Methods. 2010;7(12):973-975. doi:10.1038/nmeth.1524

185. Meador K, Wysoczynski CL, Norris AJ, Aoto J, Bruchas MR, Tucker CL. Achieving tight control of a photoactivatable Cre recombinase gene switch: new design strategies and functional characterization in mammalian cells and rodent. Nucleic Acids Research. 2019;47(17):e97. doi:10.1093/nar/gkz585

186. Weinberg BH, Pham NTH, Caraballo LD, et al. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. Nature Biotechnology. 2017;35(5):453-462. doi:10.1038/nbt.3805

187. Diaz S, Li L, Wang K, Liu X. Expression and purification of functional recombinant CUL2•RBX1 from E. coli. Scientific Reports. 2021;11(1):11224. Published 2021 May 27. doi:10.1038/s41598-021-90770-x

188. Lennon CW, Stanger M, Belfort M. Protein splicing of a recombinase intein induced by ssDNA and DNA damage. Genes & Development. 2016;30(24):2663-2668. doi:10.1101/gad.289280.116

189. Topilina NI, Green CM, Jayachandran P, et al. SufB intein of Mycobacterium tuberculosis as a sensor for oxidative and nitrosative stresses. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(33): 10348-10353. doi:10.1073/pnas.1512777112

190. Belfort M. Mobile self-splicing introns and inteins as environmental sensors. Current Opinion in Microbiology. 2017;38:51-58. doi:10.1016/j.mib.2017.04.003

191. Kelley DS, Lennon CW, Li Z, et al. Mycobacterial DnaB helicase intein as oxidative stress sensor. Nature Communications. 2018;9(1):4363. Published 2018 Oct 19. doi:10.1038/s41467-018-06554-x

192. Groitl B, Jakob U. Thiol-based redox switches. Biochimica et Biophysica Acta. 2014;1844(8):1335-1343. doi:10.1016/j.bbapap.2014.03.007

193. Gao G, Liu X, Pavlovsky A, Viola RE. Identification of Selective Enzyme Inhibitors by Fragment Library Screening. Journal of Biomolecular Screening. 2010;15(9):1042-1050. doi:10.1177/1087057110381383

194. Goodnow RA, Dumelin CE, Keefe AD. DNA-encoded chemistry: enabling the deeper sampling of chemical space. Nature Reviews Drug Discovery. 2016;16(2): 131-147. doi:10.1038/nrd.2016.213

195. Rivera VM, Clackson T, Natesan S, et al. A humanized system for pharmacologic control of gene expression. Nature Medicine. 1996;2(9):1028-1032. doi:10.1038/nm0996-1028

196. Miyamoto T, DeRose R, Suarez A, et al. Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. Nature Chemical Biology. 2012;8(5): 465-470. doi:10.1038/nchembio.922

197. Wang P, Chen T, Sakurai K, et al. Intersectional Cre Driver Lines Generated Using Split-Intein Mediated Split-Cre Reconstitution. Scientific Reports. 2012; 2(1). doi:10.1038/srep00497

198. Chassin H, Müller M, Tigges M, Scheller L, Lang M, Fussenegger M. A modular degron library for synthetic circuits in mammalian cells. Nature Communications. 2019;10(1). doi:10.1038/s41467-019-09974-5

199. Mootz HD, Muir TW. Protein Splicing Triggered by a Small Molecule. Journal of the American Chemical Society. 2002;124(31):9044-9045. doi:10.1021/ja026769o

200. Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW. Conditional Protein Splicing: A New Tool to Control Protein Structure and Function in Vitro and in Vivo. Journal of the American Chemical Society. 2003;125(35):10561-10569. doi:10.1021/ja0362813

201. Schwartz EC, Saez L, Young MW, Muir TW. Post-translational enzyme activation in an animal via optimized conditional protein splicing. Nature Chemical Biology. 2006;3(1):50-54. doi:10.1038/nchembio832

202. Buskirk AR, Ong YC, Gartner ZJ, Liu DR. Directed evolution of ligand dependence: Small-molecule-activated protein splicing. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(29): 10505-10510. doi:10.1073/pnas.0402762101

203. Peck SH, Chen I, Liu DR. Directed Evolution of a Small-Molecule-Triggered Intein with Improved Splicing Properties in Mammalian Cells. Chemistry & Biology. 2011;18(5):619-630. doi:10.1016/j.chembiol.2011.02.014

204. Sternberg N, Hamilton D. Bacteriophage P1 site-specific recombination. Journal of Molecular Biology. 1981;150(4):467-486. doi:10.1016/0022-2836(81)90375-2

205. Yang YS, Hughes TE. Cre Stoplight: A Red/Green Fluorescent Reporter of Cre Recombinase Expression in Living Cells. BioTechniques. 2001;31(5):1036-1041. doi:10.2144/01315st03

206. Ge J, Wang L, Yang C, et al. Intein-mediated Cre protein assembly for transgene excision in hybrid progeny of transgenic Arabidopsis. Plant Cell Reports. 2016; 35(10):2045-2053. doi:10.1007/s00299-016-2015-x

207. Gramespacher JA, Burton AJ, Guerra LF, Muir TW. Proximity Induced Splicing Utilizing Caged Split Inteins. Journal of the American Chemical Society. 2019;141(35):13708-13712. doi:10.1021/jacs.9b05721

208. Zhang D, Jin S, Piao X, Devaraj NK. Multiplexed Photoactivation of mRNA with Single-Cell Resolution. ACS Chemical Biology. 2020;15(7):1773-1779. doi:10.1021/acschembio.0c00205

209. Serini SM, Cannizzaro MV, Dattola A, et al. The efficacy and tolerability of 5-aminolevulinic acid 5% thermosetting gel photodynamic therapy (PDT) in the treatment of mild-to-moderate acne vulgaris. A two-center, prospective assessor-blinded, proof-of-concept study. Journal of Cosmetic Dermatology. 2019;18(1): 156–162. https://doi.org/10.1111/jocd.12670

210. Gunaydin G, Gedik ME, Ayan S. Photodynamic Therapy for the Treatment and Diagnosis of Cancer–A Review of the Current Clinical Status. Frontiers in Chemistry. 2021;9. doi:10.3389/fchem.2021.686303

211. Cook SN, Jack WE, Xiong X, et al. Photochemically Initiated Protein Splicing. Angewandte Chemie International Edition in English. 1995;34(15):1629-1630.

doi:10.1002/anie.199516291

212. Berrade L, Kwon Y, Camarero JA. Photomodulation of Protein Trans-Splicing Through Backbone Photocaging of the DnaE Split Intein. ChemBioChem. 2010;11(10):1368-1372. doi:10.1002/cbic.201000157

213. Binschik J, Zettler J, Mootz HD. Photocontrol of Protein Activity Mediated by the Cleavage Reaction of a Split Intein. Angewandte Chemie. International Edition. 2011;50(14):3249-3252. doi:10.1002/anie.201007078

214. Ren W, Ji A, Ai H wang. Light Activation of Protein Splicing with a Photocaged Fast Intein. Journal of the American Chemical Society. 2015;137(6):2155-2158. doi:10.1021/ja508597d

215. Böcker JK, Dörner W, Mootz HD. Light-control of the ultra-fast Gp41-1 split intein with preserved stability of a genetically encoded photo-caged amino acid in bacterial cells. Chemical Communications. 2019;55(9):1287-1290. doi:10.1039/c8cc09204d

216. Tyszkiewicz AB, Muir TW. Activation of protein splicing with light in yeast. Nature Methods. 2008;5(4):303-305. doi:10.1038/nmeth.1189

217. Wong S, Mosabbir AA, Truong K. An Engineered Split Intein for Photoactivated Protein Trans-Splicing. Isalan M, ed. PLoS ONE. 2015;10(8):e0135965. doi:10.1371/journal.pone.0135965

218. Jones DC, Mistry IN, Tavassoli A. Post-translational control of protein function with light using a LOV-intein fusion protein. Molecular BioSystems. 2016;12(4): 1388-1393. doi:10.1039/c6mb00007j

219. Zhang W, Lohman AW, Zhuravlova Y, et al. Optogenetic control with a photocleavable protein, PhoCl. Nature Methods. 2017;14(4):391-394. doi:10.1038/nmeth.4222

220. Gramespacher JA, Stevens AJ, Nguyen DP, Chin JW, Muir TW. Intein Zymogens: Conditional Assembly and Splicing of Split Inteins via Targeted Proteolysis. Journal of the American Chemical Society. 2017;139(24):8074-8077. doi:10.1021/jacs.7b02618

221. Goedhart J, von Stetten D, Noirclerc-Savoye M, et al. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nature Communications. 2012;3(1). doi:10.1038/ncomms1738

222. Matsuda T, Cepko CL. Controlled expression of transgenes introduced by in vivo electroporation. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(3):1027-1032. doi:10.1073/pnas.0610155104

223. Tóth DJ, Tóth JT, Gulyás G, et al. Acute depletion of plasma membrane phosphatidylinositol 4,5-bisphosphate impairs specific steps in endocytosis of the G-protein-coupled receptor. Journal of Cell Science. 2012;125(Pt 9):2185-2197. doi:10.1242/jcs.097279

224. Zhang W, Lohman AW, Zhuravlova Y, et al. Optogenetic control with a photocleavable protein, PhoCl. Nature Methods. 2017;14(4):391-394. doi:10.1038/nmeth.4222

225. Goedhart J, von Stetten D, Noirclerc-Savoye M, et al. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nature Communications. 2012;3:751. Published 2012 Mar 20. doi:10.1038/ncomms1738

226. Nam HS, Benezra R. High levels of Id1 expression define B1 type adult neural stem cells. Cell Stem Cell. 2009;5(5):515-526. doi:10.1016/j.stem.2009.08.017domain. ACS Synthetic Biology. 2015;4(6):689-699. doi:10.1021/sb500322u

227. Seto B. Rapamycin and mTOR: a serendipitous discovery and implications for breast cancer. Clinical and Translational Medicine. 2012;1(1):29. Published 2012 Nov 15. doi:10.1186/2001-1326-1-29

228. Prabhala SV, Gierach I, Wood DW. The Evolution of Intein-Based Affinity Methods as Reflected in 30 years of Patent History. Frontiers in Molecular Biosciences. 2022;9:857566. Published 2022 Apr 8. doi:10.3389/fmolb.2022.857566

229. Johnstone RM, Adam M, Hammond JR, Orr L, Turbide C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). Journal of Biological Chemistry. 1987;262(19): 9412-9420.

230. van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. Nature Reviews. Molecular Cell Biology. 2018;19(4):213-228. doi:10.1038/nrm.2017.125

231. Yang Y, Hong Y, Cho E, Kim GB, Kim IS. Extracellular vesicles as a platform for membrane-associated therapeutic protein delivery. Journal of Extracellular Vesicles. 2018;7(1):1440131. doi:10.1080/20013078.2018.1440131

232. Herrmann IK, Wood MJA, Fuhrmann G. Extracellular vesicles as a next-generation drug delivery platform. Nature Nanotechnology. 2021;16(7):748-759. doi:10.1038/s41565-021-00931-2

233. Rivoltini L, Chiodoni C, Squarcina P, et al. TNF-Related Apoptosis-Inducing Ligand (TRAIL)–Armed Exosomes Deliver Proapoptotic Signals to Tumor Site. Clinical Cancer Research. 2016;22(14):3499-3512. doi:10.1158/1078-0432.ccr-15-2170

234. Ohno S ichiro, Takanashi M, Sudo K, et al. Systemically Injected Exosomes Targeted to EGFR Deliver Antitumor MicroRNA to Breast Cancer Cells. Molecular Therapy. 2013;21(1):185-191. doi:10.1038/mt.2012.180

235. Lyu P, Wang L, Lu B. Virus-Like Particle Mediated CRISPR/Cas9 Delivery for Efficient and Safe Genome Editing. Life. 2020;10(12):366. doi:10.3390/life10120366

236. Voelkel C, Galla M, Maetzig T, et al. Protein transduction from retroviral Gag precursors. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(17):7805-7810. doi:10.1073/pnas.0914517107

237. Montagna C, Petris G, Casini A, et al. VSV-G-Enveloped Vesicles for Traceless Delivery of CRISPR-Cas9. Molecular Therapy - Nucleic Acids. 2018;12:453-462. doi:10.1016/j.omtn.2018.05.010

238. Yao X, Lyu P, Yoo K, et al. Engineered extracellular vesicles as versatile ribonucleoprotein delivery vehicles for efficient and safe CRISPR genome editing. Journal of Extracellular Vesicles. 2021;10(5). doi:10.1002/jev2.12076

239. Shen B, Wu N, Yang JM, Gould SJ. Protein Targeting to Exosomes/Microvesicles by Plasma Membrane Anchors. Journal of Biological Chemistry. 2011;286(16):14383-14395. doi:10.1074/jbc.m110.208660

240. Mangeot PE, Dollet S, Girard M, et al. Protein Transfer Into Human Cells by VSV-G-induced Nanovesicles. Molecular Therapy. 2011;19(9):1656-1666. doi:10.1038/mt.2011.138

241. Zhang X, Xu Q, Zi Z, et al. Programmable Extracellular Vesicles for Macromolecule Delivery and Genome Modifications. Developmental Cell. 2020;55(6):784-801.e9. doi:10.1016/j.devcel.2020.11.007

242. Finkelshtein D, Werman A, Novick D, Barak S, Rubinstein M. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. Proceedings of the National Academy of Sciences of the United States of

America. 2013;110(18):7306-7311. doi:10.1073/pnas.1214441110

243. Beffert U, Stolt PC, Herz J. Functions of lipoprotein receptors in neurons. Journal of Lipid Research. 2004;45(3):403-409. doi:10.1194/jlr.r300017-jlr200

244. Kim IS, Jenni S, Stanifer ML, et al. Mechanism of membrane fusion induced by vesicular stomatitis virus G protein. Proceedings of the National Academy of Sciences of the United States of America. 2016;114(1). doi:10.1073/pnas.1618883114

245. Fu A, Tang R, Hardie J, Farkas ME, Rotello VM. Promises and Pitfalls of Intracellular Delivery of Proteins. Bioconjugate Chemistry. 2014;25(9):1602-1608. doi:10.1021/bc500320j

246. Carvajal-Vallejos P, Pallissé R, Mootz HD, Schmidt SR. Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources. Journal of Biological Chemistry. 2012;287(34):28686-28696. doi:10.1074/jbc.m112.372680

247. Yang YS, Hughes TE. Cre Stoplight: A Red/Green Fluorescent Reporter of Cre Recombinase Expression in Living Cells. BioTechniques. 2001;31(5):1036-1041. doi:10.2144/01315st03

248. Fang Y, Wu N, Gan X, Yan W, Morrell JC, Gould SJ. Higher-Order Oligomerization Targets Plasma Membrane Proteins and HIV Gag to Exosomes. Walter P, ed. PLoS Biology. 2007;5(6):e158. doi:10.1371/journal.pbio.0050158

249. Shen B, Wu N, Yang JM, Gould SJ. Protein Targeting to Exosomes/Microvesicles by Plasma Membrane Anchors. Journal of Biological Chemistry. 2011;286(16):14383-14395. doi:10.1074/jbc.m110.208660

250. Beyer HM, Mikula KM, Li M, Wlodawer A, Iwaï H. The crystal structure of the naturally split gp41-1 intein guides the engineering of orthogonal split inteins from cis-splicing inteins. FEBS Journal. 2019;287(9):1886-1898. doi:10.1111/febs.15113

251. Shah NH, Eryilmaz E, Cowburn D, Muir TW. Extein Residues Play an Intimate Role in the Rate-Limiting Step of Protein Trans-Splicing. Journal of the American Chemical Society. 2013;135(15):5839-5847. doi:10.1021/ja401015p

252. Stewart SA, Dykxhoorn DM, Palliser D, et al. Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA. 2003;9(4):493-501. doi:10.1261/rna.2192803

253. S Subach OM, Cranfill PJ, Davidson MW, Verkhusha VV. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. PLoS One. 2011;6(12):e28674. doi:10.1371/journal.pone.0028674

254. Takara Bio USA. Guide-it CRISPR/Cas9 Gesicle Production System User Manual. https://www.takarabio.com/documents/User%20Manual/Guide/Guide-it%20CRISPR-Cas9%20Gesicle%20Production%20System%20User%20Manual_053017.pdf

255. Everett BA, Litzau LA, Tompkins K, et al. Crystal structure of the Wheat dwarf virus Rep domain. Acta Crystallographica. Section F, Structural Biology Communications. 2019;75(Pt 12):744-749. doi:10.1107/S2053230X19015796

256. Banskota S, Raguram A, Suh S, et al. Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. Cell. 2022;185(2):250-265.e16. doi:10.1016/j.cell.2021.12.021

257. Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrokovski S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. Nucleic Acids Research. 2009;37(8):2560-2573. doi:10.1093/nar/gkp095

258. Pinto F, Thornton EL, Wang B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. Nature Communications. 2020;11(1). doi:10.1038/s41467-020-15272-2

259. Auclair D, Ketteridge D, Oates S, Hopwood JJ, Byers S. An overview of intra-articular therapy for mucopolysaccharidosis VI. Journal of Pediatric Rehabilitation Medicine. 2010;3(1):3-6. doi:10.3233/PRM-2010-0101

260. Matthew AN, Zephyr J, Nageswara Rao D, et al. Avoiding Drug Resistance by Substrate Envelope-Guided Design: Toward Potent and Robust HCV NS3/4A Protease Inhibitors. mBio. 2020;11(2):e00172-20. Published 2020 Mar 31. doi:10.1128/mBio.00172-20

261. Laohakunakorn N. Cell-Free Systems: A Proving Ground for Rational Biodesign. Frontiers in Bioengineering and Biotechnology. 2020;8:788. Published 2020 Jul 24. doi:10.3389/fbioe.2020.00788

CURRICULUM VITAE