THE UNIVERSITY OF CHICAGO

EXPANDING THE SCOPE, STRUCTURE, AND FUNCTION OF SYNTHETIC BIOLOGICS WITH DIELS-ALDER CYCLOADDITIONS AND OTHER STRATEGIES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

JEFFREY EVAN MONTGOMERY

CHICAGO, ILLINOIS

JUNE 2021

Nature exhibits not simply a higher degree, but an altogether different level of complexity.

– Benoit B Mandelbrot

You're only given a little spark of madness. You can't lose it.

– Robin Williams

Table of Contents

List of Figures

List of Tables

List of Schemes

Abstract

Peptides offer distinct advantages as a protein targeting modality for mechanistic probe and drug development. Given their large surface area and the potential for highly tailored design, peptide ligands can have exquisite target-binding affinities and specificities. To counteract conformational and metabolic instabilities, many peptide stabilization strategies have been deployed, in some cases leading to cell-permeable bioactive compounds. A noteworthy peptide stabilization chemistry is ring-closing olefin metathesis, used in the formation of stapled peptides, constrained alpha-helical peptides that serve as protein-protein interaction inhibitors. Herein we report first-in-class cell-active stapled peptide inhibitors of RAB25, a recalcitrant protein target implicated in the pathogenesis of a variety of cancers. This work exemplifies the potential of stabilized peptides to engage challenging targets in relevant biological contexts, however the current repertoire of peptide stabilization chemistries have several limitations, including: the requirement of exogeneous reagents or harsh conditions often incompatible with unprotected peptide, full length proteins, or aqueous conditions; a narrow scope of applicable folds and resulting stabilized structures; limited diversity in stabilizing linker structure, which rigidifies peptide conformation and can itself impact target-binding. To begin addressing these shortcomings, we applied the Diels-Alder reaction as a bioorthogonal chemistry for peptide cyclization. Our studies confirm its suitability for both on-resin and in-solution peptide cyclization, in organic and aqueous solutions

respectively. Cyclization kinetics are rapid and high-yielding across a range of diene and dienophile functional groups and peptide folds, predominantly resulting in cycloadducts with *endo* stereochemistry as ascertained by NMR and X-ray crystallographic studies. Further, Diels-Alder cyclized peptides display enhanced bioactivity, with cycloadduct composition and geometry having differential impacts on target-binding, confirmed by the observation of substantial cycloadduct-protein contacts in a crystal structure of an SRC2-derived Diels-Alder cyclized peptide bound to its target estrogen receptor alpha. Additionally, Diels-Alder peptide cyclization is shown to be compatible with both ring-closing olefin metathesis and disulfide formation, suggesting the broad applicability of this chemistry alongside other stabilization chemistries. Separate work reports on the design and screening of peptide inhibitors targeting a highly conserved coronavirus methyltransferase complex, nsp16/nsp10, central to the infectivity of SARS-CoV-2, the cause of the ongoing COVID-19 global pandemic. Finally, we present a robust, multi-readout cell penetration, compound stability, and membrane disruption assay for peptides and other synthetic biologics. This allows for conclusive reporting of key pharmacologic properties of these promising protein-targeting modalities. Taken together, these works serve to expand the application, function, and study of synthetic biologics and related compounds for probing biology and treating disease.

Acknowledgements

There are so many people who have made this work possible, and I'm beyond grateful to all of them. From the top, thank you Professor Ray Moellering for giving me an opportunity and the freedom to explore such a wide breadth of science in your laboratory. It's hard to imagine a more enriching environment to develop as a scientist. From high level insights to critical details to long term strategic vision, and just plain common sense, I'm grateful for all that I've learned in your lab.

My friends and coworkers in the lab, past and present, have also greatly contributed to the work discussed here, as well as my development as a scientist. Many thanks to: Dr. Somayeh Ahmadiantehrani for her kindness and expert guidance as an editor and Adobe Illustrator whiz; Dr. Jae Won Chang for help and advice, especially in the early days of the lab; Dr. Gang Li for teaching me so many skills, his patience and encouragement, as well as for help with microscopy work found in Chapter 2 and in many things outside the scope of this dissertation; Dr. Thomas Speltz for bouncing ideas, dimer schemes, Karlos, molecular modeling of DAC peptides in Chapter 3, and collaborating on the work in Chapter 5; Dr. David McCutcheon and Anthony Carlos for many helpful discussions and support, as well as for providing benzylguanine precursor for work in Chapter 5; Dr. Shaghayegh Fathi for her encouragement, support, and chemical expertise; Drs. Gihoon Lee and Joey Huang for advice on proteomics; Dr. Sean Shangguan for working out the synthesis of Fmoc-cysteine(hexadiene) found in Chapter

3; John Coukos for help with 2D-NMR and Q-TOF analysis of DAC peptides found in Chapter 3; Justin Donnelly for tremendous contributions from the beginning of DAC peptide formation in Chapter 3 and great insights on most everything; Abigail Estes for assistance with the work in Chapter 5, many other innovations not discussed in this thesis, and taking each other on walks; Zeyu Qiao for sharing many intermediates, precursors, and laughs; Kavya Pillai for her incredible team work and insight; Chris Lee for sharing plasmids and laughs; Anna Savin for melting proteins; Debbie Thomas for continuing work on DAC peptides.

Many additional thanks go out to all the collaborators who have made contributions to this work as well. These include: the lab of Professor Gordon Mills at MD Anderson, especially Dr. Shreya Mitra who did much of the biochemistry and cellular profiling of RAB25-targeting stapled peptides in Chapter 2, as well as Drs. Kang Jeong and Bo Peng for running RNAseq for this study; the lab of Professor Gregory Verdine, and especially Matthew Kolar, for early work synthesizing and screening the stapled peptides found in Chapter 2; the lab of Professor Geoffrey Greene at the University of Chicago, and especially Dr. Sean Fanning for providing protein and his expertise crystalizing DAC peptide bound to its target, as found in Chapter 3; Dr. Joseph Sachleben in the Biomolecular NMR Core at the University of Chicago for expert help with 2D-NMR studies of DAC peptides in Chapter 3; the lab of Professor Karla Satchell at Northwestern University, and especially Dr. Monica Rosas-Lemus for providing protein, expert advice, and support for the work described in Chapter 4.

Special thanks to my thesis defense committee members Professors Aaron Esser-Kahn and Weixin Tang for their helpful comments and insightful discussion. Also many thanks to my candidacy committee members Professors Chuan He, Sergey Kozmin, and Jared Lewis. Separately, I also thank Professor Jared Lewis for being my undergraduate research mentor, giving me the opportunity to pursue my own research from the beginning of my training, and his support and encouragement through the years. I also thank the NIH for financial support through the Chemical Biology Interface Training Grant and the Martha Ann and Joseph A. Chenicek Graduate Research Fund Fellowship for additional funding.

Finally, I wish to thank my family for their unwavering support through my time in college and graduate school, I could not have done any of this without you. I also thank my friends and acquaintances through the years who have supported me, given me advice, and been all-around good hangs.

CHAPTER 1: Synthetic biologics provide new handles for challenging targets

1.1 Introduction

In the early nineteenth century, as a rudimentary molecular understanding of proteins was emerging, the term was coined from the Greek πρωτειος, or *proteios* ("primary" or "of the first rank") [1, 2]. Subsequently and fittingly, proteins have emerged to be central drivers in the majority of biological processes. Proteins are thus the primary targets of most pharmaceuticals to treat disease, as well as the targets of probe molecules to elucidate biological mechanisms [3-6]. Indeed, there are numerous examples throughout human history of protein modulation through the use of exogeneous substances; the use of plant-derived salicylates (predecessors of acetylsalicylic acid, modern day aspirin) for their anti-inflammatory properties being just one example [7]. Therefore, over the last century, a large part of modern research has been dedicated to the design, synthesis, and study of molecules that modulate protein function. These are central goals across the fields of chemical biology, medicinal chemistry, and many of the biological sciences. This focus on the study and targeting of proteins is exemplified by the lofty goal of designing a targeted probe for every one of the ~20,000 proteins in the human body [8]. Probe design (and likewise drug design and development) can be an

exceptionally challenging, even perilous, process [9]. This is due in large part to proteins' structural similarities, their complex interdependencies and context-dependencies, and the physicochemical requirements for adequate probe or drug design.

Protein structure is encoded by its primary amino acid sequence, which in turn dictates function. These functions can include enzymatic catalysis, signaling, and regulation of other proteins and biomolecules via binding interactions for translocation, sequestration, inhibition, and activation. Given the evolutionary history of proteins and families of proteins (*e.g.* gene duplication events or convergent evolution), as well as the fact proteins across all domains of life are composed of almost exclusively the same 20 amino acids, many structural features are often homologous. This necessarily makes selective protein target engagement a challenge, as outlined in later sections. Adding further complexity to this targeting problem, the structures and functions of proteins are context-dependent across many variables including: space (*e.g.* tissue type, cell type, subcellular distribution); time (*e.g.* cell state, protein lifetime, dynamic processes); and environment or proximity (*e.g.* phases of matter, solute composition, binding partners). One of the most critical variables defining protein structure and function are interactions with other proteins [10]. Driven by molecular recognition, these protein-protein interactions (PPIs) are involved in the vast majority of biological processes, and are therefore highly regulated, complex events critical for the proper functioning of biological systems [11]. When dysregulated, PPIs are often drivers of disease states. Much work has been devoted to the discovery and characterization of PPIs [12-14], yet they remain challenging targets for probe and drug development [15]. The following sections outline

the strengths and weaknesses of current protein targeting modalities, as well as recent progress and new opportunities for selective inhibition of PPIs.

1.2 Small molecules

The pharmacopeia has historically been composed of small molecules that ideally elicit effects on target proteins with high potency and specificity. Since prehistory, humanity used compounds found in natural occurring sources to treat disease, so not surprisingly natural product discovery and screening was central to many of the early successes of modern drug discovery [16, 17]. Through this process, it was discovered that small molecules are well-suited for development into drugs due to their size, physicochemical properties, and similarity to biologically active compounds found in nature. Their hydrophobicity makes them readily cell-permeable and often orally bioavailable; their constituent functional groups can be tuned for binding with protein backbones and side chains; they are unlikely to trigger immunogenicity due to their small size. The synthetic tractability of small molecules is another attractive property, continuously enriched by the growing repertoire of chemical transformations that expands the scope of synthesizable matter.

Indeed, the development of medicinal chemistry over the last half-century has led to the refinement of design principles for more selective, targeted small molecule therapeutics. Heuristics like Lipinksi's Rule of Five (*i.e.* five or less hydrogen bond donors; ten or less hydrogen bond acceptors; 500 dalton or less molecular weight; log P below 5) and beyond have helped improve design, identification, and selection of more drug-like small molecules. Theoretically, such data and experience-driven metrics should increase

the chances of success while progressing through the drug discovery and development pipeline [18-20]. With further advances, such as data-driven computational design [21] and increasingly sophisticated high-throughput screening platforms [22, 23], it would seem that any target could yield to small molecule drug discovery given sufficient time and effort.

A noteworthy example is imatinib, sold under the tradename Gleevec. Imatinib targets the BCR-ABL fusion protein, the product of a genetic translocation that results in the formation of the Philadelphia Chromosome and the central driver of chronic myeloid leukemia and other blood cancers [24-26]. Through iterative rounds of medicinal chemistry, an early hit targeting protein kinase C was derivatized to both selectively target the fusion protein's ABL tyrosine kinase active site (**Figure 1.1 A-C**) and improve its drug-like properties, ultimately leading to rapid clinical testing and FDA approval for an otherwise hard to treat disease with poor prognosis [24]. Imatinib is further noteworthy as it was one of the first drugs that brought to bear advances in the understanding of a disease's genetic basis, and was thought to herald a revolution in pharmacogenomic drug discovery at the dawn of the genomics era, in part through the successful design of next generation analogs for treatment-resistant BCR-ABL mutants [27, 28]. While genomics has greatly enhanced our understanding of health and disease, and advances in genetic technologies has enriched many drug discovery platforms, this revolution has so far not come to pass in the post-genomic era. We have a much better understanding of the molecular underpinnings of disease, with ~4000 protein-coding genes implicated in disease progression [29, 30]. Unfortunately, fundamental limitations of small molecules

have held back progress in expanded targeting of these proteins. Like the vast majority of approved small molecule drugs, imatinib and its successors elicit their effect by engaging and inhibiting the activity of a binding site evolved to bind small molecule metabolites.

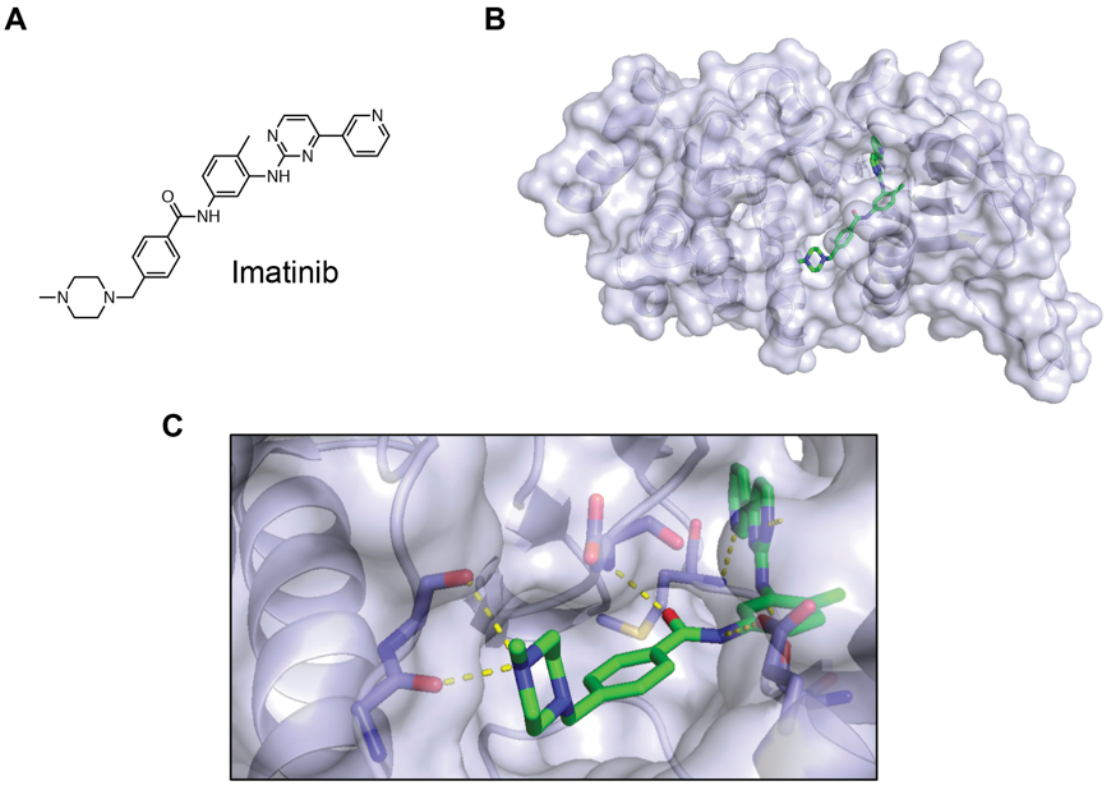


Figure 1.1: Imatinib, a representative small molecule inhibitor of BCR-ABL's tyrosine kinase domain.
(A) Structure of imatinib. **(B)** Imatinib (green, sticks) shown bound in the active site of the ABL tyrosine kinase domain (light blue) (PDB: 3K5V). **(C)** Detailed image of polar binding interactions (yellow dashes) between imatinib and main chain and side chain functional groups of the ABL tyrosine kinase active site.

Only approximately 10% of the proteome is believed to contain binding pockets suitable for engagement by small molecules (**Figure 1.1 C**) [31, 32]. Smoother, large interfaces such as those found at PPIs, are especially challenging targets for small molecules [33-35]. Further, due to the nature of molecular evolution, many proteins within

the same family or that share similar structural or functional features may contain analogous small molecule binding sites. This increases the risk of off-target binding, diminishing therapeutic efficacy and potentially causing unacceptable toxicity [36, 37]. While the potential chemical space occupied by relatively broad definitions of drug-like small molecules is enormous (~10^{60} distinct possible structures), only a very small subset will actually have requisite drug-like properties while also being sufficiently differentiated to specifically engage a target protein [38-40]. Thus in recent years, Eroom's Law – an inversion of Moore's Law, charting the continual decrease in newly approved drugs per billion US dollars spent on pharmaceutical research and development – has been invoked. Additionally, some protein targets or classes of targets have begun to be labeled as "undruggable" following decades of targeting attempts [41].

A notable class of so-called undruggable targets are the many transcription factors for which mounting evidence implicates their dysregulation in an outsized number of cancers [42, 43]. Their purported "undruggability" stems from several key, frustrating features. As opposed to nuclear receptor transcription factors, which are typically ligand activated, many other transcription factors lack the small molecule binding sites or other known pockets required for traditional small molecule drug design [44]. Further, many are unstructured, intrinsically disorder proteins that only take on structure upon engagement with other protein binding partners. This reliance on PPIs and protein-DNA interactions for their functions, means many transcriptions factors are primarily composed of large, smooth surfaces. In addition, the design of small molecule inhibitors targeting the largely positively-charged DNA-binding regions of transcription factors is traditionally a non-

starter as resulting molecules would lack requisite small molecule drug-like properties [45, 46]. Recent advances in targeting transcription factor, PPIs, and other allegedly undruggable targets with small molecules, however, have emerged, primarily mediated through inhibitory allosteric interactions [34, 47, 48]. In addition, covalent small molecule inhibitors and other technologies such as molecular glues and degraders are emerging as more potent methods for hitting recalcitrant target classes [49-51]. While small molecule therapeutics continue to expand in targetable scope and potency, other emerging modalities can greatly complement these potential therapies.

1.3 Biologics

Biologics have emerged in recent decades to seize a large share of the drug market. While newer modalities such as RNA-based therapeutics are growing in prominence, proteins dominate the category, composing over 20% of newly approved drugs in 2019, with antibodies representing the majority and many achieving blockbuster status [52, 53]. Antibodies have many favorable pharmacologic properties and a history of therapeutic use dating back to protective 'serum therapies', despite being considered poor candidates for drugs as recently as the 1980s [54]. Antibodies' antigen-binding surface area is quite large (~1000 $Å^2$ on average), providing for exceptional target-binding specificity and affinity (**Figure 1.2**) [55]. This coverage allows antibodies to inhibit PPIs, an exceptionally challenging target for traditional small molecules [33]. Their capacity for iterative rounds of selection and optimization can confer high binding affinities for protein targets. Given that the majority of therapeutic antibodies are now produced in humanized forms, and mechanisms for maintaining high antibody titers exist in the human body, the

circulating half-life of an antibody can be weeks, prolonging the therapeutic window and reducing the frequency of administration [56, 57]. In addition, the deep study of the post-translational modifications of antibodies, as well as bioconjugation chemistry, provides for increased quality control as well as the ability to modify antibodies for enhanced efficacy, stability, and targeting of therapeutic payloads, as in the emerging class of antibody-drug conjugates [58-60]. Other protein therapeutics, including enzyme replacement therapies and engineered proteins designed to mimic existing signaling factors such as interleukins, are also growing in prominence, sharing or taking advantage of many of the same advances as described above for antibodies.

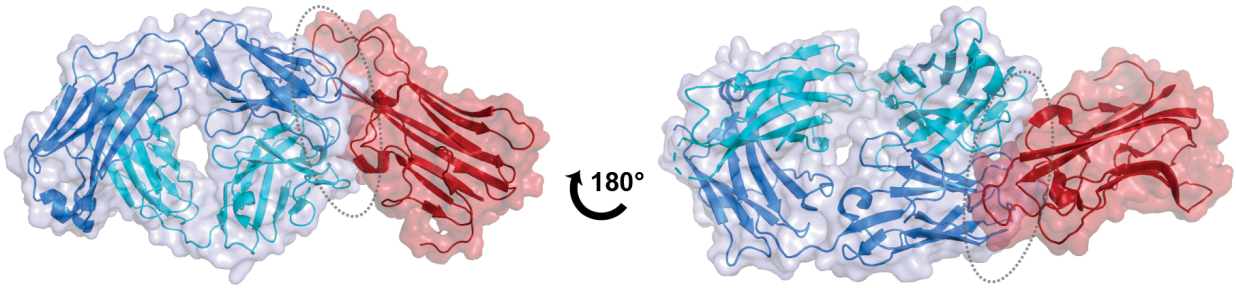


Figure 1.2: Representative structure of the antigen-binding fragment of antibody Adalimumab (blue) bound to its target TNFalpha (red), displaying the large binding interface characteristic of these biomolecular interactions (PDB 3WD5).

Antibodies and other protein therapeutics are not without down sides. Since they are proteins, they cannot survive passage through the gastrointestinal tract and are thus not orally bioavailable; intravenous injection is the most common route of administration to ensure rapid and complete delivery into the blood stream. Their large size (150 kilodaltons) also typically precludes the targeting of intracellular proteins. Therefore only extracellular proteins, ~10% of the proteome, are addressable by antibodies. Further, antibodies may trigger unfavorable immunogenic responses in some patients or with

repeated administration. Long term storage and stability of antibodies and other biologics is another perennial issue, and despite decades of advances in antibody expression system technologies and largescale manufacturing, minor changes in production conditions can lead to wide variance in final products which may have dramatic impacts on efficacy. Taken together, the current arsenal of therapeutic modalities leaves much to be desired. This is profoundly apparent in the fact that currently approved drugs only target ~754 disease-associated proteins, of an estimated 4000 targets with evidence-based involvement with various diseases [29, 61].

1.4 Peptides: In the "Goldilocks zone"?

Peptide drugs can achieve many of the beneficial properties of both small molecules and biologics, while mitigating many of their down sides (**Figure 1.3**) [62]. Synthesis of peptides has been standardized through the advent of solid-phase peptide synthesis (SPPS), a facile method for the routine synthesis of up to ~40-50mer peptides from constituent orthogonally-protected amino acid monomers [63-65]. Advances in SPPS methods and automation, as well as the introduction of native chemical ligation, have provided access to longer sequences, allowing for the full-length chemical synthesis or semi-synthesis of entire proteins [66-68]. Due to the iterative, modular nature of SPPS, natural and unnatural amino acids can be incorporated with minimal optimization of synthetic conditions, resulting in products with high purity. Thus, like small molecules, peptides are synthetically tractable, while also sampling across a large span of chemical space due to their size. This large surface area also contributes to increases in peptide binding affinities and specificities, as seen in biologics. Further, given that peptides are

chemically identical to the proteins they target, fragments designed to structurally mimic a binding partner may achieve comparable affinities to that of the native interaction, and are also relatively benign from a toxicity standpoint. The fact that suitable peptide sequences can be rapidly identified, synthesized, and derivatized for screening is another attractive feature. **Chapter 4** will explore this process toward the development of peptide inhibitors targeting SARS-COV-2. In addition, advances in recent decades have allowed for the high-throughput synthesis and screening of peptides, synthesized either chemically or biosynthetically, the latter providing access to directed evolutionary approaches for high affinity binding [69]. Indeed, pioneering work in phage display and other expression technologies have enhanced humanity's ability to harness the incredible power of molecular evolution to discover peptide and protein ligands with high target binding affinity and other desirable properties [70, 71].

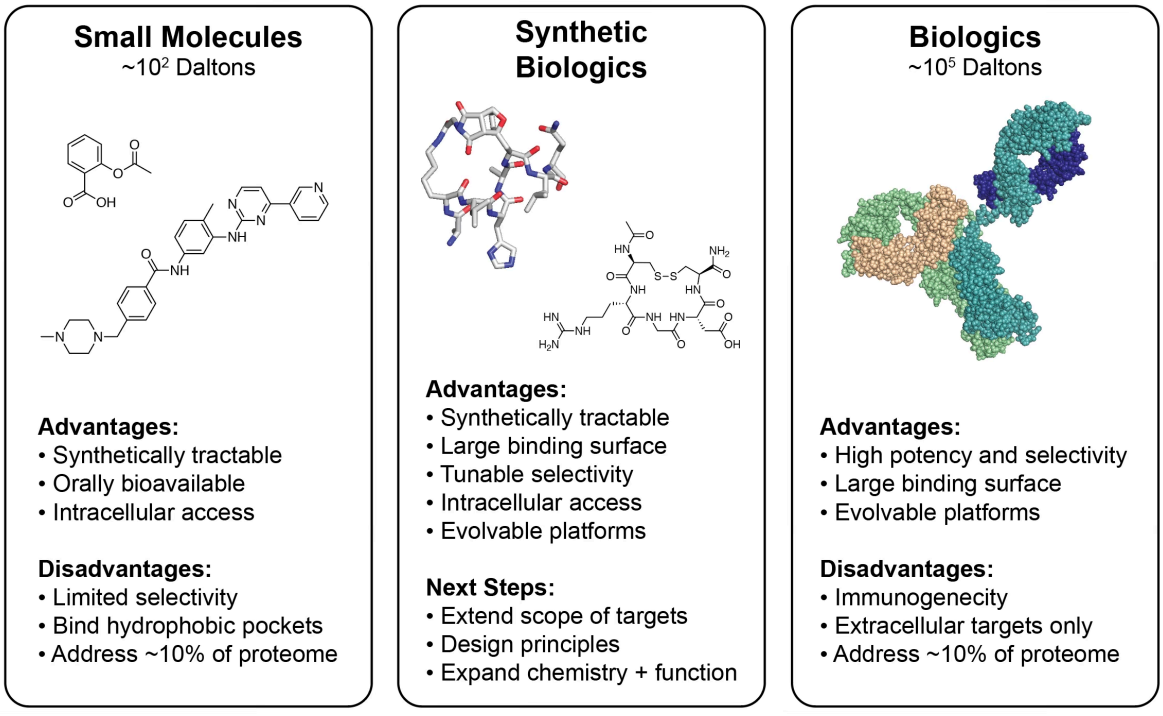


Figure 1.3: Representative structures of small molecules, synthetic biologics, and biologics.

Figure 1.3, continued: Aspirin and imatinib, structures in the left panel, are two examples of small molecule inhibitors with mechanisms of action which target the pockets of protein active sites. Diels-Alder cyclized SRC2 peptide 6a (from PDB 6PIT) and a disulfide-bridged RGD peptide are depicted in the middle panel. An antibody structure is depicted in the right panel, composed of two light chains (wheat and dark blue) and two heavy chains (light green and teal) (PDB 1HZH).

Peptides on their own, however are often not ideal drug candidates. Many linear peptide sequences, if intended to inhibit a protein-protein interaction through mimicking one of the interacting partners' surfaces, will lack the requisite topology for high-affinity binding once removed from the larger context of their source protein. A given peptide's size and proteinaceous nature can also confer many of the same liabilities as biologics. Typically, peptides are not cell-permeable and are susceptible to proteolysis by the wide array of proteases in the human body. Thus, many synthetic analogs of peptide hormones, such as insulin, glucagon-like peptide 1 (GLP-1), and atrial natriuretic peptide (ANP), are not orally bioavailable and must be administered by injection, and since the body tightly regulates the activities of their native counterparts, these synthetic hormones have low circulating half-lives [72, 73]. All of the above largely holds true for the vast majority of linear peptides, as proteases preferentially cleave extended peptide chains [74].

1.5 Peptide macrocyclization

For millions of years, nature has solved the peptide stability problem through the use of macrocyclization. Peptide macrocycles are ubiquitous across the natural world, first emerging as bacterially-produced antibiotics for chemical warfare against competing microbes. These microbial arsenals have served as sources of and inspiration for critical

medicines. Cyclization rigidifies the peptide's structure, reducing its length, flexibility, and ultimately susceptibility to proteolysis. This altered topology can also confer binding advantages through ligand pre-organization, reducing the entropic penalty of binding [75]. Additionally, some macrocyclic peptide drugs, notably cyclosporin A, can be formulated for oral administration, spurring continual research in this area [76-78]. A wide array of stabilization chemistries have been observed in nature, from the common disulfide bond to back-bone cyclized structures and diverse sidechain-sidechain crosslinks, some of which require non-ribosomal peptide synthesis or post-translational modification of ribosomal peptides [79-81]. Taken together, synthetic biologics present an intriguing opportunity for high-affinity, tunable targeting of proteins, especially intracellular protein-protein interactions, both as a therapeutic modality and mechanistic probing strategy.

Inspired by the beneficial properties of peptide macrocyclization, chemists have devised many strategies to cyclize peptides to form various secondary structures, motifs, and ring sizes in order to impart desirable properties [82, 83]. Given that many peptide macrocycles or other derivatives are often synthetic mimics of biological structures containing non-natural functionalities for improved pharmacologic properties, they are sometimes referred as peptidomimetics or synthetic biologics. The observation that macrocyclization can preferentially stabilize a peptide into a privileged high-affinity conformation has driven the study of this phenomena [75], including the recent exploration of so-called "dark conformers" through the use of dominant rotors to confine macrocycles in a tunable manner [84, 85]. As stated above, cyclization can reduce peptide susceptibility to proteolysis, and (from direct experimental observation)

appropriately designed and stored stabilized peptides retain their structure for many years, suggesting long term storage and distribution in a variety of settings would not be as large an issue as with other biologics. Further, cyclization can improve peptide cell permeability by reducing polar content/increasing hydrophobicity, stabilizing conformations that promote internal hydrogen bonding, or otherwise occlude polar functional groups from contributing to the total polar surface area [86, 87]. A review of general stabilization strategies, highlighted macrocyclization chemistries, and synthetic considerations follows.

1.5.1 Stabilization through native peptide and protein functional groups

Various sidechain-to-sidechain, head-to-sidechain, and head-to-tail peptide macrocyclization processes are found throughout the natural world, largely in microbes and plants. The cyclizing step, which may occur between a wide variety of residues or the peptide termini, is predominantly carried out by non-ribosomal peptide synthetases – multicomponent complexes of enzymes that: stitch together proteinogenic and non-proteinogenic amino acids; can chemically modify them; can form branched and/or cyclic peptide products – or through the biosynthesis of ribosomally synthesized and post-translationally modified peptides (*i.e.* RiPPs), an important emerging class of peptide natural products [81, 88-90]. The history of synthetic organic chemistry, and many of its most profound and elegant discoveries, is built in large part on the total synthesis Inspired by these powerful ligands, chemists have extensively explored diverse connection strategies utilizing the canonical amino acids.

A common strategy, very frequently encountered in natural peptides and proteins, begins with the incorporation of one or more pair of cysteines into a target peptide scaffold. These are then oxidized to form disulfide bond(s) (**Figure 1. A**). In many cases, particularly when multiple disulfides are desired, careful synthetic route design and incorporation of orthogonally protected cysteines may be required (**Figure 3.1 B**). This is to ensure accurate, sequential templating of intramolecular disulfides, and often requires a combination of on-resin and in-solution manipulations, all while controlling factors such as reaction redox state and dilution [91-93]. There are documented examples of disulfide-rich peptides or proteins that, through evolution or design, do not require onerous conditions for establishing the required connectivity [94, 95]. This is often not the case, however, so careful compound characterization and screening of conditions is usually necessary.

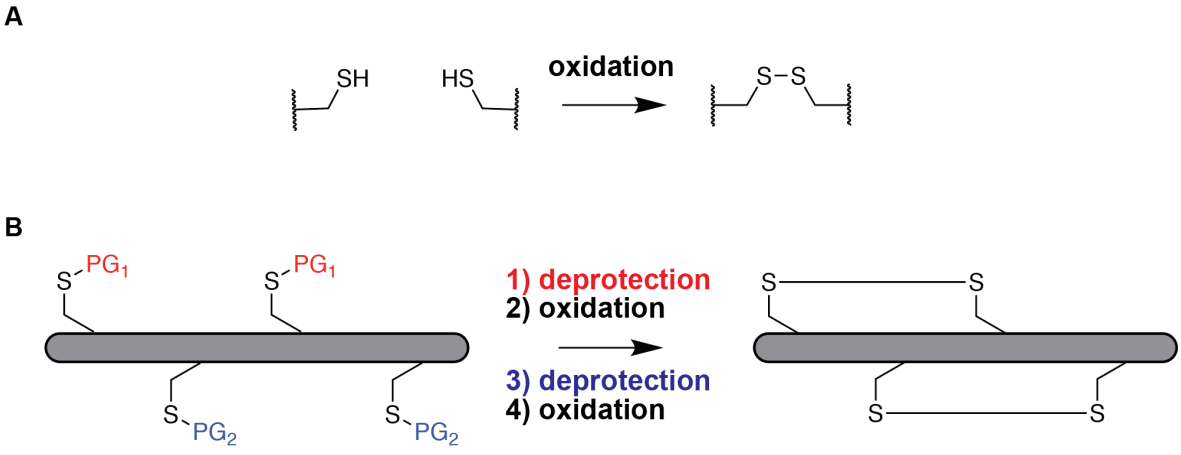


Figure 1.4: Disulfide bonds, a native peptide and protein stabilization strategy. **(A)** Disulfide formation. **(B)** Schematic representation of selective deprotection and oxidation to form multiple disulfide bonds with the desired connectivity.

Another prevalent strategy is amide bond or lactam formation on peptide scaffolds. Lactam formation can be achieved on resin through the selective deprotection of lysine and a carboxylic acid group of an aspartic or glutamic acid residue, allowing for sidechain-to-sidechain cross-linking, typically over short separating distances, using similar coupling chemistries as main chain SPPS [96, 97]. Head-to-tail peptide macrocyclization can similarly be achieved by N-terminal deprotection, cleavage of the fully protected peptide from resin, and subsequent peptide bond formation [98]. As with disulfide formation above, careful design and control of reaction conditions is required to ensure the accurate cross-linked species is produced and to minimize side products.

1.5.2 Stabilization through exogeneous reactive linkers or enzymatic processes

Another popular peptide stabilization strategy takes advantage of the nucleophilic properties of cysteines and lysines through the incorporation of, or reaction with, a bridging component containing requisite electrophilic groups. Double cysteine S-alkylation or arylation with a bridging group provides for increased cross-linking lengths and structural diversity relative to disulfide bond formation; further, peptides stabilizes by these chemistries are not susceptible to reductive linearization [99-102]. Notable examples of these linkers include di-brominated species such as 1,3-bis(bromomethyl)benzene and perfluoro-aryl groups (**Figure 1.5 A**) [103, 104]. In a similar manner, a lysine residue or N-terminal amine may be derivatized with a suitable functional group for reaction with a neighboring cysteine or lysine (**Figure 1.5 B,C**) [105-108]. Such strategies often require careful synthetic planning for on-resin derivatization and subsequent cyclization, judicious selection of by-standing residues, and/or tailored

reagents for selective modification of side chains in solution [109, 110]. Extensions of some of these strategies have been applied to phage-displayed libraries of cyclic peptides or native proteins [69, 111]. Such protein-expression based platforms allow for directed evolution-type strategies for enhanced screening of optimal stabilized peptide or protein structures.

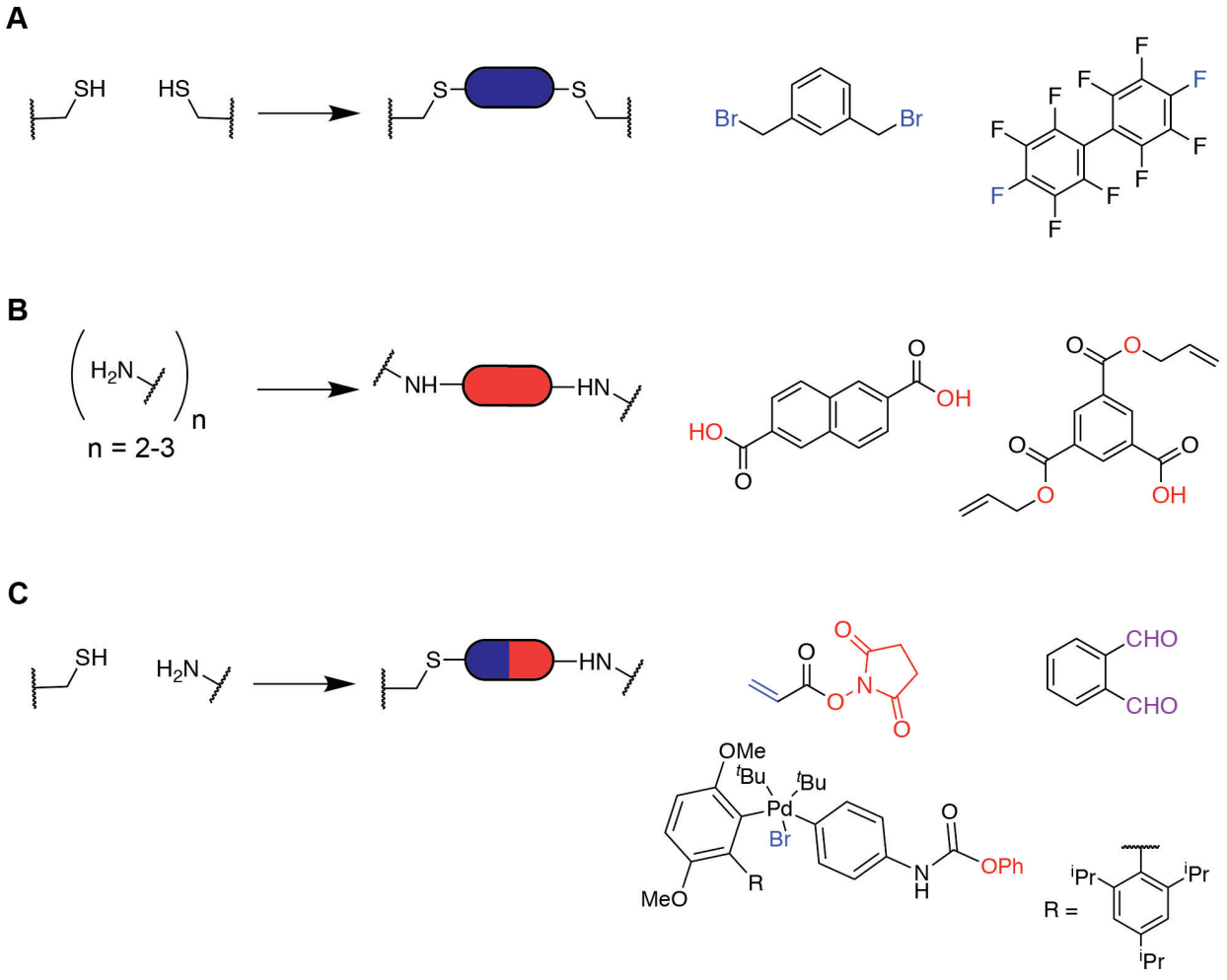


Figure 1.5: Representative examples of cysteine-cysteine (A), lysine-lysine (B), and cysteine-lysine (C) cross-linking chemistries.

In addition, more exotic macrocyclization reactions have emerged in recent years, from enzymatic cyclization, multi-component cross-linking chemistries, photochemical

methods, and cross-links that occur at less frequently utilized residues [112-116]. Key limitations with many of these chemistries include: their incompatibility with layering multiple stabilizing chemistries; requirement of extensive manipulations; requirements for on-resin crosslinking/dilution to avoid side reactions/oligomerization; may require forcing conditions due to impediments of complete reaction (*e.g.* target sequence or cross-link composition); or may necessitate further engineering of reaction conditions or tailored cross-linking reagents for highly specific, robust reactions [109, 117, 118].

1.5.3 Stabilization through non-canonical functional groups

Non-native functionalities may also be incorporated directly during peptide synthesis or through the use of non-canonical amino acid incorporation orthogonal protein expression systems [119-121]. Such reported macrocyclization chemistries include copper-catalyzed azide-alkyne cycloaddition (CuAAC), thiol-ene click chemistry, ring-closing metathesis, as well as a variety of multicomponent reactions [112, 122-126]. The incorporation of non-canonical amino acids enabling metal-catalyzed bond formation is a noteworthy example [127]. The most well-studied of these strategies include ruthenium-catalyzed ring-closing olefin metathesis (RCM) of terminal alkene-containing groups, as well as [3 + 2] Huisgen ligation of alkyne/azide pairs [122, 128-131]. These “stapled” peptides show increased favorable properties including structural stability, binding affinity, cellular uptake, and *in vivo* pharmacokinetics for direct or allosteric targeting of PPIs [132-135]. Notably, compounds in the RCM class have entered clinical trials for various indications [136]. The utility of stapled peptides resides in their structure and function: they are short, constrained, alpha-helical peptides capable of inhibiting PPIs by faithfully

templating and mimicking an alpha helix found at the interface of interaction (**Figure 1.6 A**) [132, 133, 135]. This is accomplished through the incorporation and RCM of one or more pairs of olefin-containing non-natural amino acids through the treatment of the fully-protected peptide on resin with Grubbs I catalyst (**Figure 1.6 B**). First synthesized by Blackwell and Grubbs [137], the Verdine group standardized the design of optimal stapled peptide inhibitors [128, 138]. Central to this was the synthesis and use of (S)-N-Fmoc-2-(4'-pentenyl)alanine (S_5) and (R)-N-Fmoc-2-(7'-octenyl)alanine (R_8), bis-alkylated olefin-containing amino acids. Two S_5 residues (incorporated with *i, i+3* or *i, i+4* spacing), or one S_5 residue and one R_8 residue (incorporated with *i, i+7* spacing), are positioned on the same face of the alpha helix allowing for productive cross metathesis [129, 139, 140]. Residues selected for S_5 or R_8 substitution should of course not be critical contributors to target binding, however we and others have observed through structural study that resulting stabilization crosslinks can contribute favorably to binding interactions [141-143]. In addition to optimal structural mimicry for high-affinity binding and stabilization, stapled peptides also display an increased propensity for cell penetration, in part likely due to the hydrophobic nature of the bridging 'staple' and intramolecular hydrogen bonding characteristic of alpha helices. Amino acid content is also a major contributing factor to cell penetrance and is central to the related study of cell penetrating peptides [144, 145]. **Chapter 2** will present work on the design, synthesis and characterization of first-in-class stapled peptide inhibitors of the GTPase RAB25.

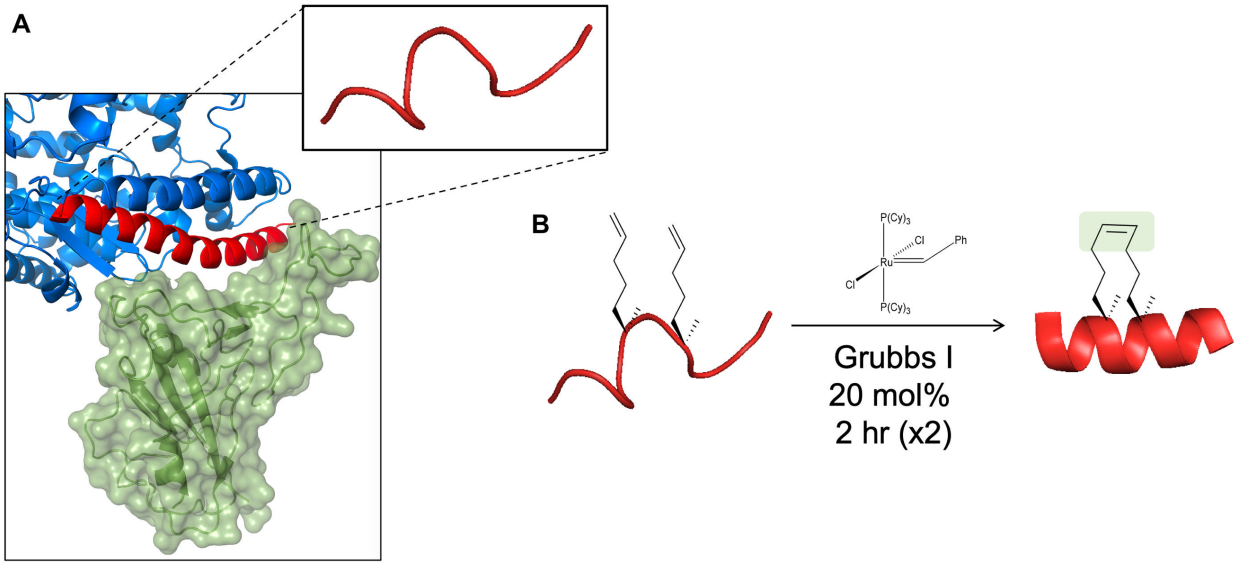


Figure 1.6: Design inspiration, rationale, and chemical stabilization of stapled peptides by RCM.
(A) The H1 alpha helix (red) of the ACE2 receptor (blue) composes the bulk of the interaction surface with the receptor binding domain of the SARS-COV-2 Spike protein (green) (PDB: 6M17). When isolated outside the context of the full-length protein, the H1 peptide, and most linear sequences, are largely unstructured (inset). **(B)** Through the incorporation of bis-alkylated olefin-containing amino acids and cross metathesis of the olefin groups with Grubbs I catalyst, a helix can be stabilized.

While all of the above chemistries can successfully stabilize peptide secondary structure, most macrocyclization transformations require some combination of exogenous reagents, harsh conditions, peptide side-chain protection, and typically must be performed on-resin. They are also often incompatible with diverse structures, other natural and non-natural functionalities, and the aqueous conditions necessary for proper folding of specific peptide and protein scaffolds. Further, they often have limited orthogonality with other stabilization chemistries and may only template a subset of peptide folds. In cases of in-solution macrocyclization chemistries, highly selective or otherwise tailored reagents are required for the modification and cross-linking of target side chains. Additionally, the transition-metal catalyzed reactions mentioned often require

expensive and toxic catalysts that must be removed during purification. Therefore, new strategies to control secondary and tertiary structure, ideally with wide functional group and reaction condition tolerance, would greatly expand the synthetic arsenal for developing peptide-based chemical probes. **Chapter 3** will describe the application of Diels-Alder cycloaddition toward expanding the repertoire of peptide cyclization chemistries.

1.6 Application and structural and functional characterization of synthetic biologics

As interest in peptide probes and therapeutics has grown, researchers have devised numerous methods to best establish their structures and functions. Traditional analytical methods such as liquid chromatography, mass spectrometry, and nuclear magnetic resonance (NMR) are used to assess compound identity, purity, connectivity and structure. X-ray crystallography can be used to further elucidate peptide structure as well as inform design through the study of protein targets' binding interfaces and their interacting partners. Circular dichroism (CD) spectroscopy is widely used to monitor well-defined secondary structures of peptides, especially for the alpha helical content of stapled peptides, which is characterized by distinct minima at 208 and 222 nm (**Figure 1.7**) [146].

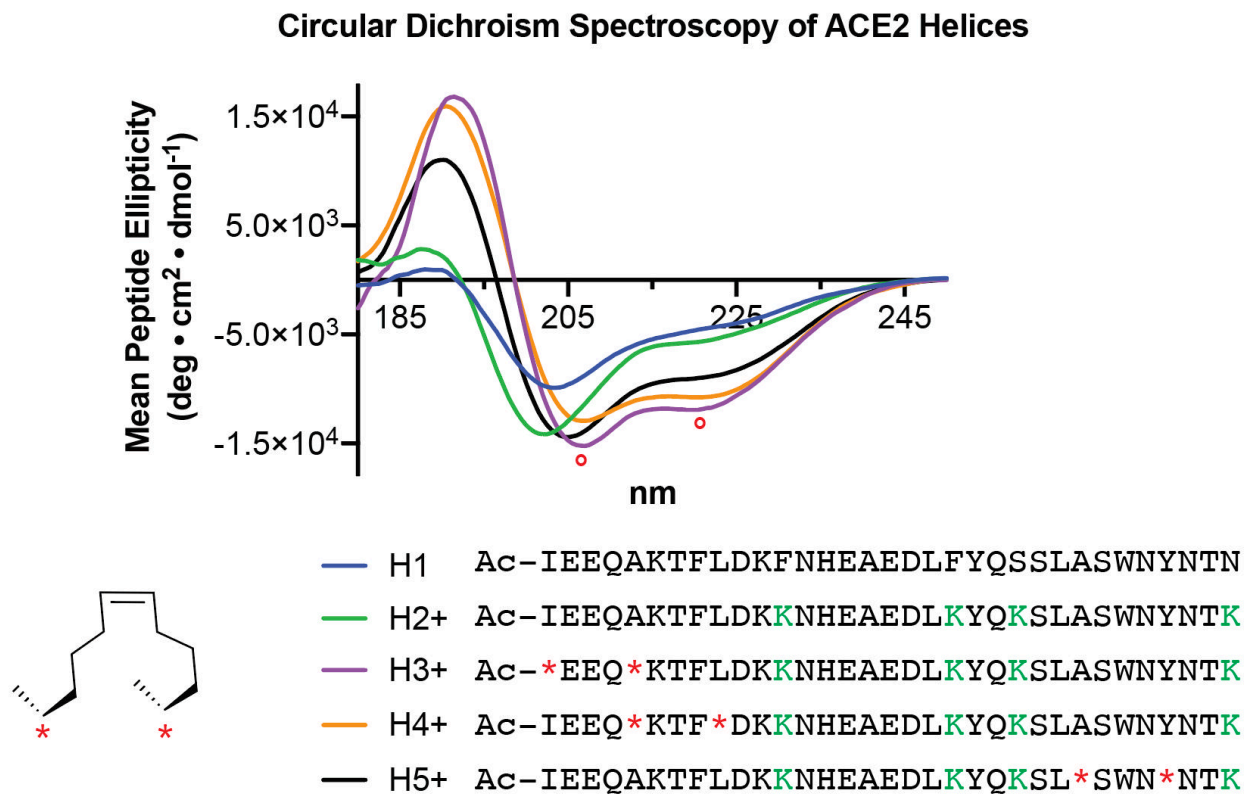


Figure 1.7: Representative data for CD spectroscopic analysis of helical content of linear and stapled peptides.
Linear ACE2 receptor H1 analogs, either H1 wildtype (blue) or solubilized lysine-substituted H2+ (green), display CD spectra characteristic of unstructured random coil. RCM-stabilized variants H3-5+ display spectra characteristic of high alpha helical content, with minima at 208 and 222 nm (denoted by open circles) *, denotes staple-bridging residues.

The ease with which peptides can be functionalized by appending linkers and desired moieties on their N-termini during SPPS has led to the wide adoption of fluorescence and enrichment-based methods for studying peptide function. Peptides labeled with fluorescein isothiocyanate (FITC), for example, can serve both as “tracer” molecules in fluorescence polarization experiments to assess *in vitro* target binding as well as probes to assess cell penetrance through fluorescence imaging [144, 147]. The limitations of this fluorescent cell penetration readout will be discussed in greater depth

in **Chapter 5**, along with the design and application of a new strategy for measuring peptide cell penetrance and stability. Biotinylated peptides can be used in binding assays such as surface plasmon resonance (SPR) and pull-down experiments to detect, and in the former quantify, peptide-protein binding interactions [148, 149]. Additional biochemical analyses such as isothermal calorimetry, gel-shift assays, and thermal shift assays can similarly be used for characterization. Further in the development process, cell- and animal-based studies are employed to study the *in cellulo* and *in vivo* activity of peptide inhibitors. These can range from kit-based viability assays and fluorescent reporter assays, to phenotypic screens such as cell migration assays or changes in transcriptomic or proteomic profiles.

1.7 Concluding remarks

While synthetic biologics show great promise as therapeutics and mechanistic probes, the field is still maturing. Expansions in target scope and deeper understanding of design principles for current strategies such as stapled peptides are critical. I believe increasing the structural and functional diversity of synthetic biologics will be paramount to further advance the field. To accomplish this, new chemistries are necessary to stabilize a broader spectrum of peptide and protein folds, ideally with the versatility to operate in aqueous environments, on large scaffolds, and in concert with other orthogonal stabilization or conjugation chemistries. This will allow for the incorporation of multiple functional features into one molecule for enhanced: stability; binding; tissue-, cell-, or subcellular-specific targeting; and overall potency through combination with other emerging modalities in drug discovery. In addition, advances in analytical, (bio)synthetic,

and high-throughput screening technologies should be brought to bear on the development of next generation synthetic biologics in the same manner that medicinal chemistry has impacted small molecule drug discovery in the last half century.

1.8 Dissertation scope

This dissertation will focus on my work in the Moellering Lab broadly addressing the design, synthesis, characterization, and screening of synthetic biologics. **Chapter 2** presents first-in-class stapled peptide inhibitors of the RAB25 GTPase, a challenging target and context-dependent driver or inhibitor of tumor progression. **Chapter 3** presents the application of Diels-Alder cycloadditions for the stabilization of diverse peptide scaffolds. **Chapter 4** presents recent work toward the targeting of the NSP16 methyltransferase of SARS-COV-2 with peptide inhibitors. **Chapter 5** presents the development of a cell penetration assay for synthetic biologics that reports in an unbiased manner on the cytosolic delivery and stability of cargos.

CHAPTER 2: Design, synthesis, and evaluation of stapled peptide inhibitors of RAB25*

2.1 Introduction

The dense, complex intracellular environment is the setting for the vast majority of biological processes; with the vast majority of these processes carried out by proteins; with all the protein-mediated processes at least tangentially affected by protein-protein interactions (PPIs). With estimates of total mammalian intracellular protein concentration on the order of ~10^2 mg/mL (or ~10^9-10^{10} individual proteins per cell), regulatory binding events mediated by PPIs must have sufficiently high affinity and specificity to faithfully carry out their actions [11, 150]. Multicomponent protein complexes, prevalent binding partner promiscuity, and context-dependent changes in the structure and function of proteins add additional layers of complexity to the study of the “interactome” in healthy and diseased states [13, 35, 151]. Unsurprisingly, potent and specific intracellular PPI inhibitors that could serve as mechanistic probes or drugs are comparatively rare relative to active site-targeting small molecule inhibitors (though the reported specificities of many probes in the literature are suspect) [9].

Taking inspiration from the remarkable molecular recognition properties of proteins, chemists have brought forward a multitude of synthetic peptide-based probes for targeting intracellular PPIs [11, 152, 153]. To accomplish this, probe sequence and structure is often informed by these corresponding properties of the PPI-members of interest. Perhaps the most noteworthy and widely implemented example of this is peptide stapling [140]. Through intramolecular cross-linking of two side chains of a natively-helical peptide sequence, the resulting product can have a templated alpha-helical structure, capable of binding with affinities comparable to or better than the targeted interaction. This methodology, primarily employing ring-closing olefin metathesis (RCM) as the cross-linking chemistry, has been applied for the synthesis of inhibitors of many otherwise challenging PPIs [129, 133, 154-156]. Using this powerful technology, we sought to develop mechanistic probes for a new target: RAB25, a small GTPase with a globular and relatively featureless structure (**Figure 2.1**), had so far proven to be an intractable target for probe or drug development.

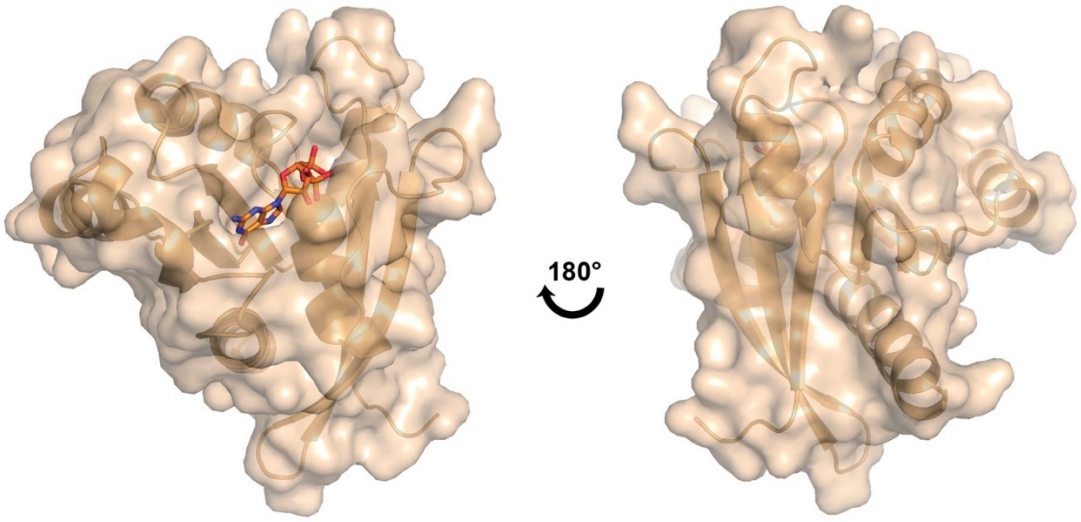


Figure 2.1: X-ray crystal structure of the small GTPase RAB25, bound to GDP (orange sticks) (PDB: 2OIL).

Members of the Ras-oncogene superfamily of small GTPases, RAB proteins are predominantly involved in membrane trafficking through regulation of vesicle formation, targeting, and fusion. RAB11a/b and RAB25, members of the RAB11 subfamily, play roles in early endosome compartmentalization of membrane protein trafficking, localization, and recycling [157, 158]. Several RAB proteins, including RAB25, have been implicated in the progression of multiple cancer types [159-162]. A molecular understanding of RAB25-mediated contributions to these different disease states remains poorly defined. Further, seemingly contradictory reports suggest a tumor suppressor role for RAB25 that is silenced in invasive breast cancers and intestinal cancers [163-166]. Taken together, this mounting body of evidence serves as motivation for the development of RAB25 inhibitors, both for therapeutic use as well as to parse the clearly complex biology surrounding RAB25 function in cancers and other contexts.

The RAB11-family of interacting proteins (Rab11-FIPs, referred to herein as FIPs) have previously been demonstrated as obligate members of RAB11/25 trafficking complexes [167, 168]. Subdivided into Class-I (FIP1, FIP2, and FIP5) and Class-II (FIP3, 4) proteins, FIP proteins have been shown to bind RAB11 and RAB25 through a conserved C-terminal RAB-binding domain (RBD), which appears to make contact with a hydrophobic patch on RAB25 in an extended alpha-helix-turn-3_{10}-helix conformation (**Figure 2.2 A**) [169, 170]. Structural data also show that RAB/FIP complexes exist, at least *in vitro*, as heterotetramers, with extensive RAB-FIP and FIP-FIP contacts (**Figure 2.2 A,B**). Overexpression of dominant-negative mutant FIP proteins, incapable of binding RAB11/25, and shRNA knockdown of *FIP* expression block cargo protein recruitment by

RAB11 and/or RAB25 in cells [171, 172]. We reasoned that development of inhibitors targeting the RAB25-FIP binding interface may allow disruption of RAB25 and/or RAB11 signaling in cells. Here we report the design and synthesis of all-hydrocarbon stapled peptides that exhibit increased structural stability and binding affinity toward RAB25. Several optimized cell permeable stapled peptides disrupt RAB25:FIP complex formation *in vitro* and *in situ*, and oppose the context-specific phenotypes associated with RAB25 function in ovarian and breast cancer cell lines.

2.2 Results

2.2.1 Initial restrained-FIP peptide design, synthesis, and screening

Due to the alpha-helical nature of the FIP-RBD, we hypothesized that it could be a suitable scaffold for stapled peptides targeting RAB25. Study of RBD sequences from FIP1-4 revealed several conserved residues, shown to contact RAB11a and RAB25 in X-ray structures with FIP3 and FIP2, respectively, as well as residues on the other side of the binding interface that could be suitable for stapling sites (**Figure 2.2 B**). A representative panel of stapled peptides (Restrained-FIP Peptides, RFPs) with a single $i, i+4$ staple at each of the four candidate positions were designed from Class-I (FIP1) and Class-II (FIP3) sequences (**Figure 2.2 C**). These initial RFPs, as well as corresponding wildtype peptides, were synthesized with an N-terminal monoethylene glycol linker and a fluoresceinisothiocyanate (FITC) label for downstream fluorescence polarization (FP) analysis of recombinant RAB11a (activated Q70L mutant, herein referred to as RAB11a) and RAB25 binding affinity measurements.

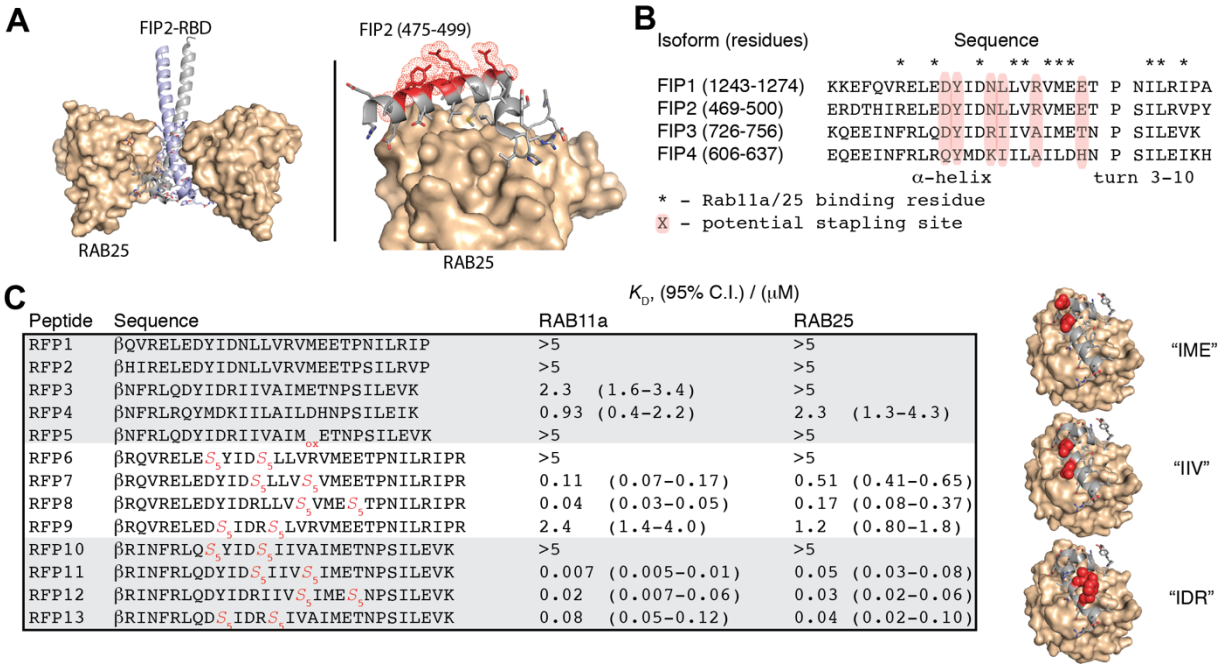


Figure 2.2: Design, synthesis, and RAB11 and RAB25 binding of linear wildtype and RFP stapled peptides.
(A) Crystal structure of the RAB25:FIP2 heterotetramer (left). The C-terminal RAB-binding domain of FIP2 binds as a helix-turn-helix peptide (gray and blue) to a hydrophobic groove on RAB25 (tan surface). Side chains not involved in RAB25 binding are highlighted as red surfaces while others are shown as sticks (right). PDB accession: 3TSO. **(B)** Sequence alignment of the RAB-binding domains from Class-I (FIP1 and 2) and Class-II (FIP3 and 4) FIP proteins. Residue annotations were assigned based on published RAB11a:FIP3 and RAB25:FIP2 crystal structures. **(C)** Table of initial series of FIP1- and FIP3-derived stapled peptides, with apparent binding affinities to both RAB11a and RAB25. Each peptide stapling position is denoted by the residues in FIP3 spanned by the hydrocarbon staple (e.g., "IIV"), highlighted in red on the schematic structures shown (right). The respective RAB11a/25 ligand affinities in the table show the mean apparent K_D with 95% confidence interval from triplicate replicates and application of a sigmoidal curve fit using Prism 5 software.

Unmodified Class-I peptides (FIP1 and 2) showed minimal binding to either RAB11a or RAB25; unmodified Class-II (FIP3 and 4) peptides showed micromolar affinities (**Table 2.1**). The RAB11a-Class-II FIP interactions appeared quite specific, as a molecule identical to **RFP4** besides an oxidized methionine (**RFP5**, an isolable side product of the synthesis) displayed ~5-fold lower binding affinity for RAB11a. In contrast

to the unmodified peptides, the stapled peptides displayed a wide range of affinities for RAB11a and RAB25. In line with the unmodified peptide results, Class-II-derived stapled peptides (**RFP10–13**) generally displayed stronger binding to each protein than Class-I sequences (**RFP6–9**). Several structure-activity relationships (SARs) emerged for each staple position in both classes (**Table 2.1** and **Figure 2.2 C**). Both peptides containing the most N-terminal staple position **(RFP6** and **RFP10**, spanning residues 'YID' in FIP3, thus referred to as the 'YID' staple position) displayed negligible affinity toward RAB11a and RAB25. This is likely caused by a steric clash between the staple and Ala75 and Ala76 in RAB11a and RAB25, respectively (**Figure 2.3 A,B**). Stapled peptides containing the internal staple positions 'IME', 'IIV', and 'IDR' displayed apparent K_D values that were significantly lower (up to ~100-fold) than the corresponding unmodified peptides, highlighting the positive effect of hydrocarbon stapling on target engagement (**Figure 2.3 C**). Finally, peptides from both FIP classes generally displayed higher affinities for RAB11a compared to RAB25, which, to our knowledge, has not been systematically explored in the literature.

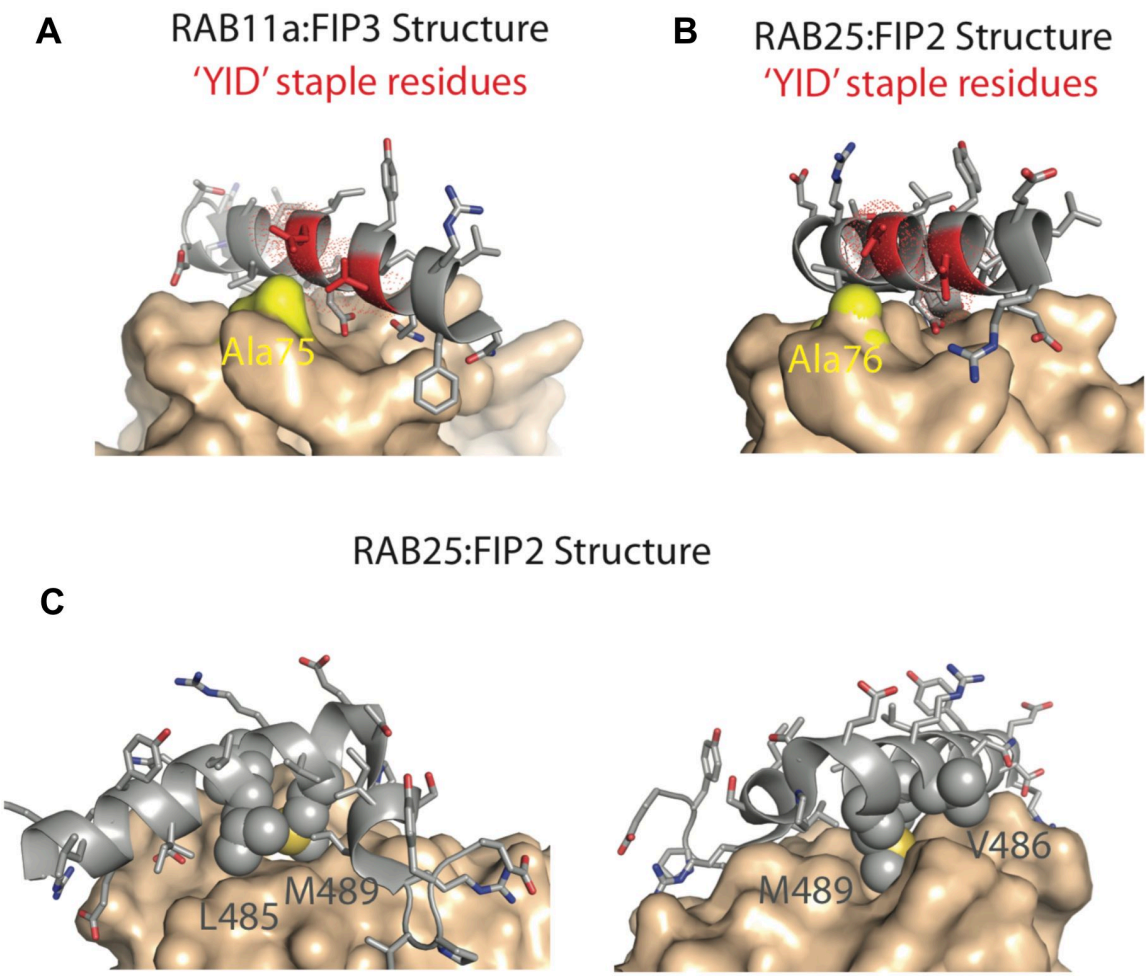


Figure 2.3: Relevant interactions highlighted in RAB11a:FIP3 and RAB25:FIP2 crystal structures.
(A,B) Potential steric clash between the hydrocarbon staple in peptides harboring the 'YID' staple position with Ala75 or Ala76 in RAB11a (PDB 2HV8) and RAB25 (PDB 3TSO), respectively. **(C)** The hydrophobic residues L485, V486 and M489 in FIP2 make critical contacts with the RAB25 binding interface and are the sites of alanine substitution in the negative control peptides **RFP31** and **RFP32**.

2.2.2 Improved stability and binding of optimized RFPs

Several stapled peptides in the initial panel showed improved binding affinity for RAB11a and RAB25 compared to their corresponding wildtype peptides. Therefore, we sought to optimize several of their properties, including: chemical stability, cell permeability, solubility, and isoform-selectivity. Chemical stability was accomplished by

replacing the oxidation-sensitive critical methionine present in wild-type FIP1-3 (*e.g.*, M489 in the RAB25:FIP2 structure; **Figure 2.3 C**) with a norleucine (N_L) isostere or other hydrophobic residues. To improved peptide solubility and cell permeability, we replaced non-binding, acidic residues with neutral-polar or cationic residues, as well as incorporation of additional positively charged residues on the periphery of the stapled peptide to convey a net-positive charge [144]. Candidate stapled peptides were iteratively synthesized and screened for binding against RAB11a and RAB25 by fluorescence polarization. Three optimized peptides from the 'IME' and 'IDR' staple positions, **RFP14**, **RFP24**, and **RFP26**, exhibited tight binding with a preference for RAB25 relative to RAB11a (**Figure 2.4 A**). As negative controls, two inactive stapled peptides were synthesized, **RFP31** and **RFP32**, containing alanine-substitutions of three critical hydrophobic FIP residues involved in RAB11a/25 binding (L485, V486, and M489 in FIP2; **Figure 2.3 C**). Several studies have shown stapled peptide formal charge tends to correlate with active cell uptake [132, 133, 144, 156, 173, 174], and therefore **RFP31** and **RFP32** differed by a single Glu-to-Arg mutation to span the range of formal charges in the active lead compounds (**Figure 2.4 A**). Relative to the active leads, these negative control peptides showed reduced binding affinity (K_D >5 μM) for both RAB11a and RAB25, further supporting the specificity of the binding interaction between active RFP peptides and RAB11a/25 (**Figure 2.4 A**).

Circular dichroism (CD) spectroscopy was used to study the effects hydrocarbon stapling had on the secondary structure of these RFPs. Wildtype FIP3 and FIP4 peptides (**RFP3** and **RFP4**) generated spectra characteristic of β-sheet and partially disordered

peptide folds, respectively (**Figure 2.4 B**). **RFP24**, **26**, **31**, and **32** generated spectra that were more alpha-helical than their parent **RFP3** sequence as measured by CD signal at 222 nm (**Figure 2.4 C**). **RFP14** exhibited classic alpha-helical characteristics, with strong minima at ~208 and 222 nm (**Figure 2.4 C**). To further test the stabilizing effect of hydrocarbon stapling, especially considering that the bioactive conformation of FIP peptides is not expected to be entirely α-helical, we measured the thermal stability of several unmodified and stapled peptides by CD. While the unmodified **RFP4** peptide displayed modest helical character by CD (**Figure 2.4 B**), it exhibited low thermal stability, with a T_m value of 37 °C (**Figure 2.4 D**). All optimized hydrocarbon stabilized RFPs showed significantly higher thermal stability, with T_m values ranging from 52 to 69 °C (**Figure 2.4 D**). Taken together, these data confirm that peptide stapling results in stabilization of FIP peptide secondary structure, which is correlated with significant gains in affinity toward both RAB11a and RAB25.

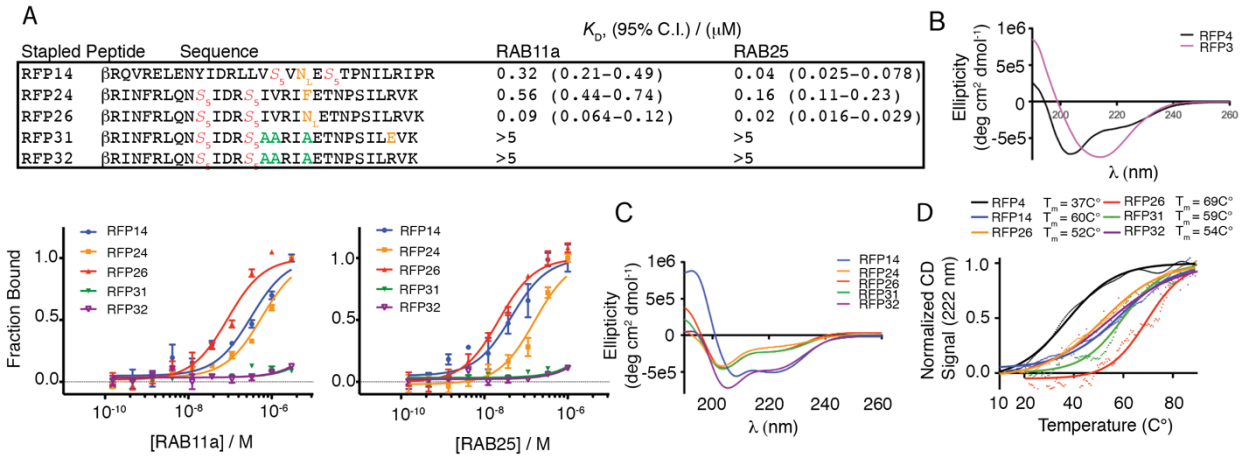


Figure 2.4: Optimized stapled peptides targeting RAB25 show correlated structural stability and high-affinity binding.
(A) Fluorescence polarization binding curves for three lead compounds, **RFP14**, **24**, and **26**, as well as two negative control peptides, **RFP31** and **32**, to both RAB11a and RAB25 (bottom). Sequences and calculated apparent K_D values are shown (top). Residues highlighted in yellow denote potential gain-of-function mutations to RAB-binding positions, while those in green denotes loss-of-function alanine replacement of hydrophobic RAB-binding side chains for the negative controls. **(B)** Circular dichroism (CD) spectra of unmodified peptides **RFP3** and **4**, derived from the RBD of FIP3 and FIP4. **(C)** CD spectra of optimized RFP stapled peptides. **(D)** Thermal denaturation CD curves measuring relative helical content (CD absorbance at 222 nm) of the indicated peptides over a temperature range from 10 to 90 °C. Individual data points at one-degree increments are shown with a sigmoidal curve fit overlay. Binding data points represent the mean ± s.e.m. from triplicate measurements. Affinities in the table show the mean apparent K_D with 95% confidence interval from triplicate replicates and application of a sigmoidal curve fit using Prism 5 software. β, beta-alanine; N_L, norleucine.

2.2.3 Cell permeable RFPs inhibit RAB25:FIP complex formation

We next sought to measure direct RAB25:RFP complex formation as well as RFP antagonism of the RAB25:FIP binding interaction. We developed an ALPHAscreen proximity assay to measure RAB25:FIP complexes in solution [175]. Unmodified or stapled N-terminally biotinylated RFPs were used with recombinant RAB25 proteins harboring affinity tags (*e.g.*, His_6- or GST-), which, when associated in solution, localize acceptor and donor beads in proximity to initiate a sensitive and specific luminescent

signal (**Figure 2.5 A**). Biotinylated **RFP14** (**bio-RFP14**) strongly associated with both His_6-labeled and GST-labeled RAB25 (**Figure 2.5 B**), which varied in saturation point due to differences in binding capacities of the corresponding beads. We next quantified the relative potencies of soluble RFPs and unmodified peptides by applying this assay in a competitive format. While **RFP3** (FIP3, residues 726–756) and FIP3 protein (residues 649–756) showed modest inhibition of RAB25-**bio-RFP14** complex formation at high concentrations (**Figure 2.5 C**), the stapled peptides **RFP14**, **24**, and **26** all displayed very potent inhibition (**Figure 2.5 C**). The negative control mutant **RFP31** exhibited significantly reduced inhibition. Finally, to test whether RFPs inhibit complex formation with full-length endogenous RAB25 and FIP proteins, we performed competitive pull-down assays in lysate from HEY cells expressing HA-RAB25 and His_6-FIP1. Immunoprecipitation of HA-RAB25 effectively co-precipitated FIP1, and this interaction was significantly inhibited by **RFP14**, **RFP24**, and **RFP26** (**Figure 2.5 D**), whereas **RFP31** was less effective. Together, these data confirm that RFP stapled peptides inhibit RAB25:FIP complex formation.

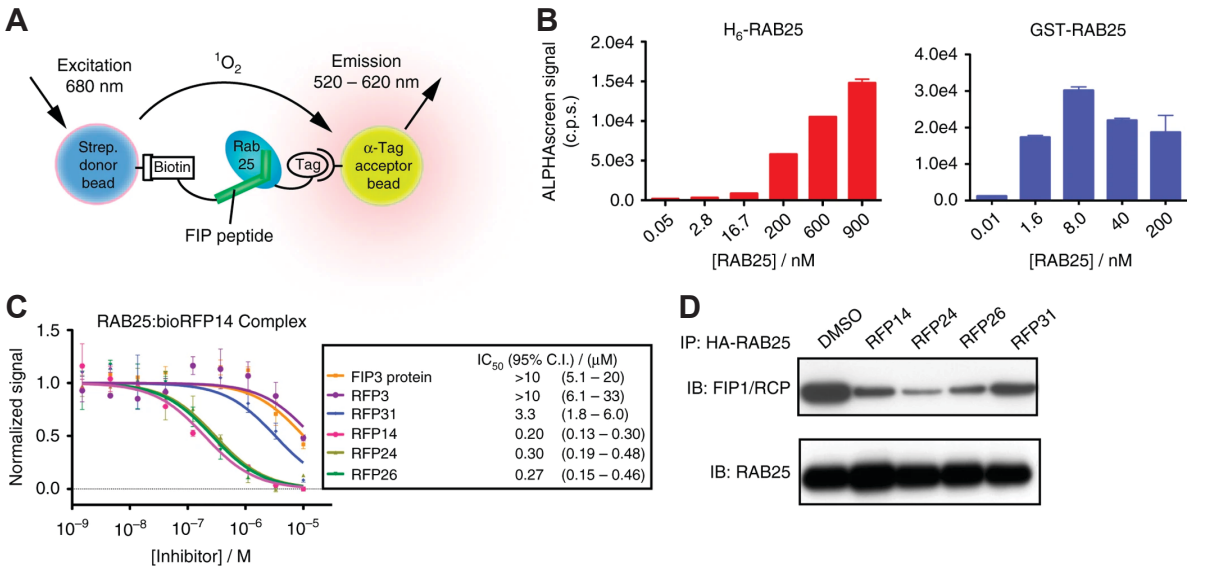


Figure 2.5: RFPs form stable complexes with RAB25 and compete for FIP binding.

Figure 2.5, continued: (A) Schematic of RAB25 ALPHAscreen proximity assay. Incubation of a synthetic, biotinylated-FIP-derived peptide with tagged (His_6- or GST-) RAB25 leads to soluble complex formation and proximal association of streptavidin-coated donor beads with acceptor beads (Ni^{2+}-NTA or anti-GST). Donor bead excitation at 680 nm produces singlet oxygen, which selectively initiates a luminescent cascade in bound acceptor beads. **(B) Bio-RFP14** binds His_6- (left) and GST-RAB25 (right) in a dose-dependent manner, leading to robust ALPHAscreen signal. **(C)** Competitive inhibition of RAB25:**bio-RFP14** complex formation by soluble RFP and FIP peptide ligands. Constant His6-RAB25 (0.2 μM) and bio-RFP14 (0.2 μM) were incubated in the presence of indicated ligand concentrations prior to addition of ALPHAscreen reagents and reading of luminescent signal. **(D)** RFP stapled peptides (1 μM) compete with HA-RAB25 binding and co-immunoprecipitation of FIP1 in HEY RAB25 cell lysate. These results are representative of experiments performed in triplicate. Binding data points represent the mean ± s.e.m. from triplicate measurements. IC_{50} values represent mean with 95% confidence interval from triplicate replicates and application of a sigmoidal curve fit using Prism 5 software.

As a final filter for RFP optimization, we used confocal fluorescence microscopy to evaluate cellular uptake of RFPs in MCF7 breast cancer cells, which express RAB25. Treatment with 5 μM FITC-labeled RFPs for 8 hours revealed that **RFP14**, **24**, **26**, **31**, and **32** each showed cellular uptake and distribution in the cytosol (**Figure 2.6**). While having a formal charge between the other RFP stapled peptides (+ 2) and sharing the 'IDR' structural backbone, **RFP31** displayed only modest cellular uptake with punctate fluorescence distribution relative to other compounds. With a formal charge of + 1, **RFP14** showed significant cellular uptake and distribution throughout cytosol and nucleus. **RFP24**, **26**, and **32**, all derived from the 'IDR' structural backbone, showed high levels of intracellular fluorescence. Importantly, while **RFP31** and **RFP32** showed equivalently reduced binding to RAB11a and RAB25 *in vitro* (**Figure 2.4 A**), **RFP32** showed superior cellular uptake. Thus it was chosen to serve as negative control in cellular studies. Together with the active compounds **RFP14**, **RFP24**, and **RFP26**, which encompass a

range of target binding affinities with roughly equivalent, robust cell penetration, this panel of stapled peptides were ideal candidates to study RAB25-mediated phenotypes in cells.

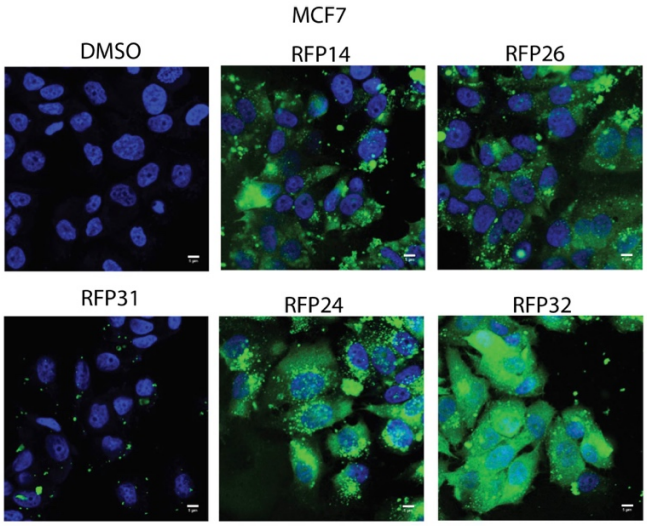


Figure 2.6: FITC-RFP peptides are cell permeable.
Confocal fluorescence imaging of DAPI-stained MCF7 cells treated with FITC-RFP peptides in 10% FBS-containing RPMI (5 μM, 8 hours). **FITC-RFP14**, **24**, and **26** show comparable cellular uptake of peptide, whereas one negative control appears less permeable (**FITC-RFP31**) and the other (**FITC-RFP32**) appears more permeable. DMSO-treated control cells show no background fluorescence under the acquisition settings. Inset white scale bars = 5 μm.

2.2.4 RFP14 inhibits RAB25-dependent phenotypes in cancer cells

With the optimization and validation of our panel of RFP stapled peptides complete, we next sought to establish if RFP-mediated disruption of RAB25-FIP complexes in cells would antagonize RAB25-mediated phenotypes. It has been established in the literature that *RAB25* expression confers either pro-tumorigenic or anti-tumorigenic phenotypes in differential cellular contexts. For instance, HEY ovarian cancer and MCF7 breast cancer cell lines overexpressing *RAB25* display increased cell proliferation, migration, and other malignant phenotypes [159, 164]. Conversely, aggressive MDA-MB231 triple-negative breast cancer cells, which natively exhibit

hypermethylation of the *RAB25* locus [176], display a reduction in oncogenic phenotypes when *RAB25* is overexpressed, including reduced cell migration and proliferation [164, 177]. Motivated by these context-dependent differences in RAB25-associated phenotype, we sought to test RFP effects in different cellular contexts, with or without *RAB25* overexpression, such that we may observe opposite effects across contexts. We reasoned this would increase the likelihood that changes observed would be due to RAB25-dependent activity. To accomplish this, we generated HEY, MCF7, and MDA-MB231 cell lines that stably overexpress *RAB25* (referred to as RAB25-HEY, RAB25-MCF7, and RAB25-MDA-MB231). Endogenous RAB25 protein is expressed at low levels in HEY and MDA-MB231 cells, while MCF7 expresses RAB25 at higher endogenous levels (**Figure 2.7 A**). As previously observed, RAB11a is constitutively expressed in all of these lines (**Figure 2.7 B**).

We first measured RAB25-dependent effects on cell proliferation, finding that *RAB25* overexpression promotes cell growth in HEY cells, as previously observed (**Figure 2.7 C**). For RFP cell-based studies, we used the two most potent compounds, **RFP14** and **RFP26**, as well as the negative control **RFP32**. Treatment of control HEY cells with **RFP14** and **RFP32** had no effect on proliferation at 48 hours. In contrast, RFP14 significantly inhibited RAB25-HEY cell proliferation, but not the negative control RFP32 (**Figure 2.7 C**). **RFP26** inhibited proliferation for both control HEY cells and RAB25-HEY cells. MCF7 cells, which endogenously express RAB25, exhibited similar proliferation in mock and RAB25 overexpression conditions (**Figure 2.7 D**). In this context, **RFP14** and **RFP26** reduced proliferation in both control and RAB25-MCF7 cells, as seen in previous

studies in which shRNA knockdown of endogenous *RAB25* decreased MCF7 proliferation [159, 164]. Finally, *RAB25* overexpression significantly decreased MDA-MB231 cell proliferation relative to mock cells, in agreement with previous data on a tumor suppressive role for RAB25 in this cellular context. **RFP14** treatment slightly increased proliferation in both cell lines, an effect that was more pronounced in *RAB25* overexpressing cells (**Figure 2.7 E**). As seen in HEY cells, **RFP26** decreased MDA-MB231 proliferation in both contexts, deviating from the RAB25-dependent effects observed with **RFP14**. This suggests that **RFP26** is potentially causing non-specific cell toxicity (see **Chapter 5**, **Section 5.2.4**, for confirmation of this effect). These studies established that modulation of RAB25 levels, but not RAB11a, which was constitutively expressed in all cell lines tested, resulted in context-specific and opposing effects on proliferation, all of which were significantly inhibited by **RFP14**.

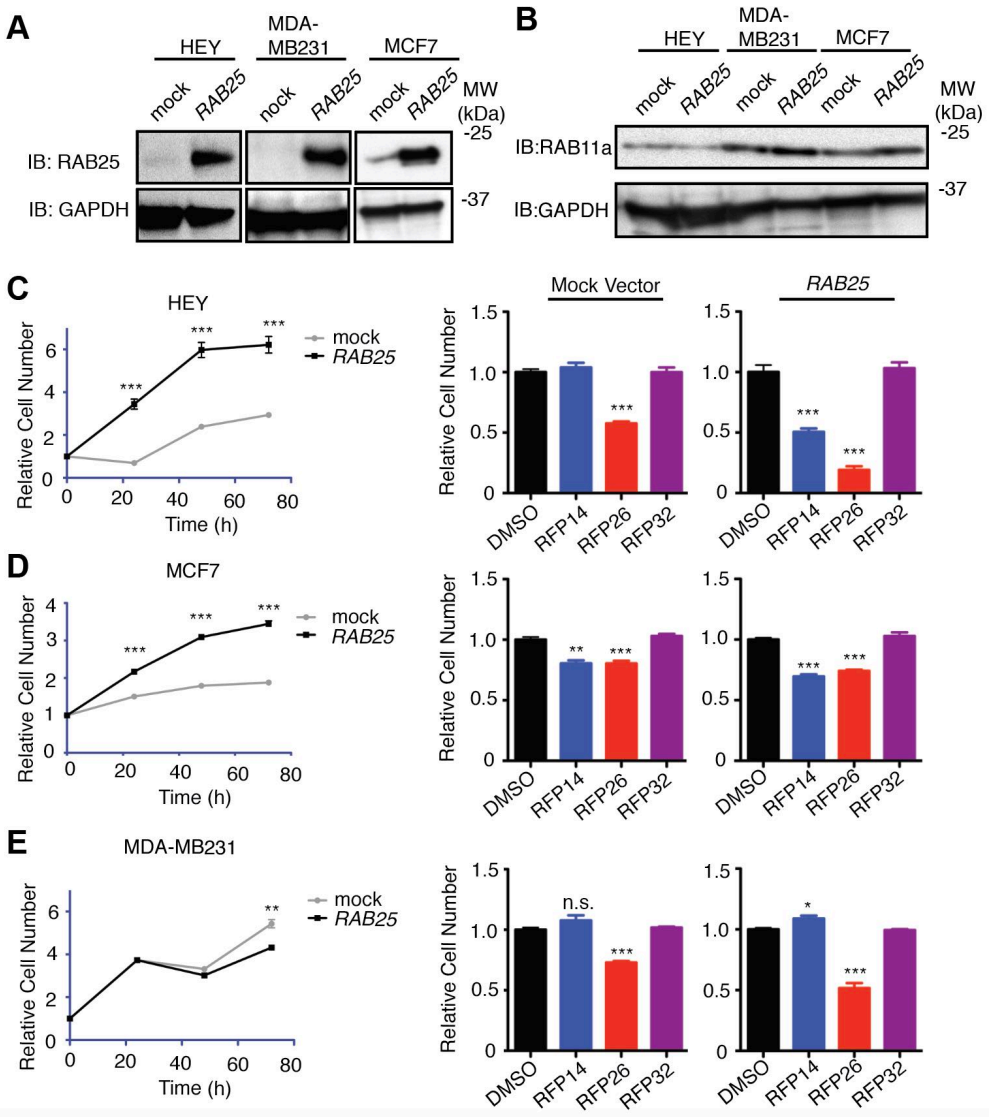


Figure 2.7: RFPs alter RAB25-driven cell proliferation.
(A) Western blot analysis of mock vector and RAB25 overexpressing HEY, MDA-MB231, and MCF7 cells. **(B)** Western blot analysis for endogenous RAB11a in HEY, MDA231, and MCF7 isogenic lines, with or without RAB25 overexpression. **(C)** Left panel compares the viability of untreated HEY cells under low-serum condition in mock or RAB25 expressing lines while right panel shows the effects of treatment with DMSO or **RFP14**, **26**, and **32** (15 μM) on the viability of HEY mock or RAB25 expressing cell lines at 48 hour post-treatment. **(D,E)** Corresponding results from MCF7 isogenics and MDA-MB231 isogenics, respectively. RAB25 expression results in enhancement (HEY, MCF7) or inhibition (MDA-MB231) of cell proliferation, which is reversed by **RFP14**, but not **RFP32**. Images and data shown represent mean ± s.e.m. of triplicate biological replicates. *P < 0.05; **P < 0.01; ***P < 0.001, Student's t-test. NS, not significant.

Cell migration is the most well-studied RAB25-dependent phenotype in cancer. In this context, we again validated opposing RAB25-dependent effects, with significant RAB25-dependent induction of migration in both HEY and MCF7 cells, and RAB25-dependent inhibition of migration in MDA-MB231 cells (**Figure 2.8 A,B**). RFP treatment of control HEY cells (10 μM) did not significantly affect cell migration (**Figure 2.8 C**), likely due to the low endogenous levels of RAB25. Conversely, **RFP14** treatment of *RAB25* overexpressing HEY and MCF7 cells resulted in significant inhibition of cell migration (**Figure 2.8 D,E**). Negative control peptide **RFP32** did not display any inhibitory effect. Notably, **RFP14** treatment of *RAB25* expressing MDA-MB231 cells significantly increased cell migration and wound closure, with no significant effect observed with **RFP32** treatment (**Figure 2.8 F**). In agreement with the previous cell proliferation data, treatment with **RFP26** showed toxicity in all cell lines tested, preventing proper wound healing studies. In combination with the previous data, unexpected effects of **RFP26** on proliferation and migration in mock and *RAB25* overexpressing cells may be a result of a lack of selectivity between RAB11a and RAB25, supported by its high affinity for both isoforms *in vitro* relative to other lead inhibitors **(Figure 2.4 A**), or may be caused by other non-specific cell toxicity. Separately, it should be emphasized that these RAB25-dependent effects were observed in RAB11a-expressing contexts (**Figure 2.7 B**), and thus phenotypic responses stemming from RAB11a inhibition should be apparent and independent of *RAB25* expression. As an additional measure to support RFP-target engagement of RAB25 and/or RAB11a in cells, we applied a cellular thermal shift assay (CETSA) [178] to assess binding of RAB25 and RAB11a. Monitoring endogenous RAB25

and RAB11a in MCF7 proteomes revealed no appreciable T_m shifts with **RFP32** treatment, indicative of no apparent binding of either isoform (**Figure 2.9**). Conversely, **RFP14** treatment resulted in a significant +2.5 °C shift in T_m for RAB25, but not for RAB11a, suggestive of compound target engagement of a significant fraction of RAB25 in cells but not for RAB11a. Taken together, these data indicate the specific effects of **RFP14** on RAB25-dependent signaling likely stem from a combination of higher RAB11a abundance, differential regulation of RAB11a signaling complexes in cells, as well as the tighter binding affinity observed between **RFP14** and RAB25 relative to RAB11a. These biochemical annotations of **RFP14**, and its ability to confer both loss- and gain-of-function phenotypes in distinct contexts, strongly supports its utility as a RAB25-targeting chemical probe.

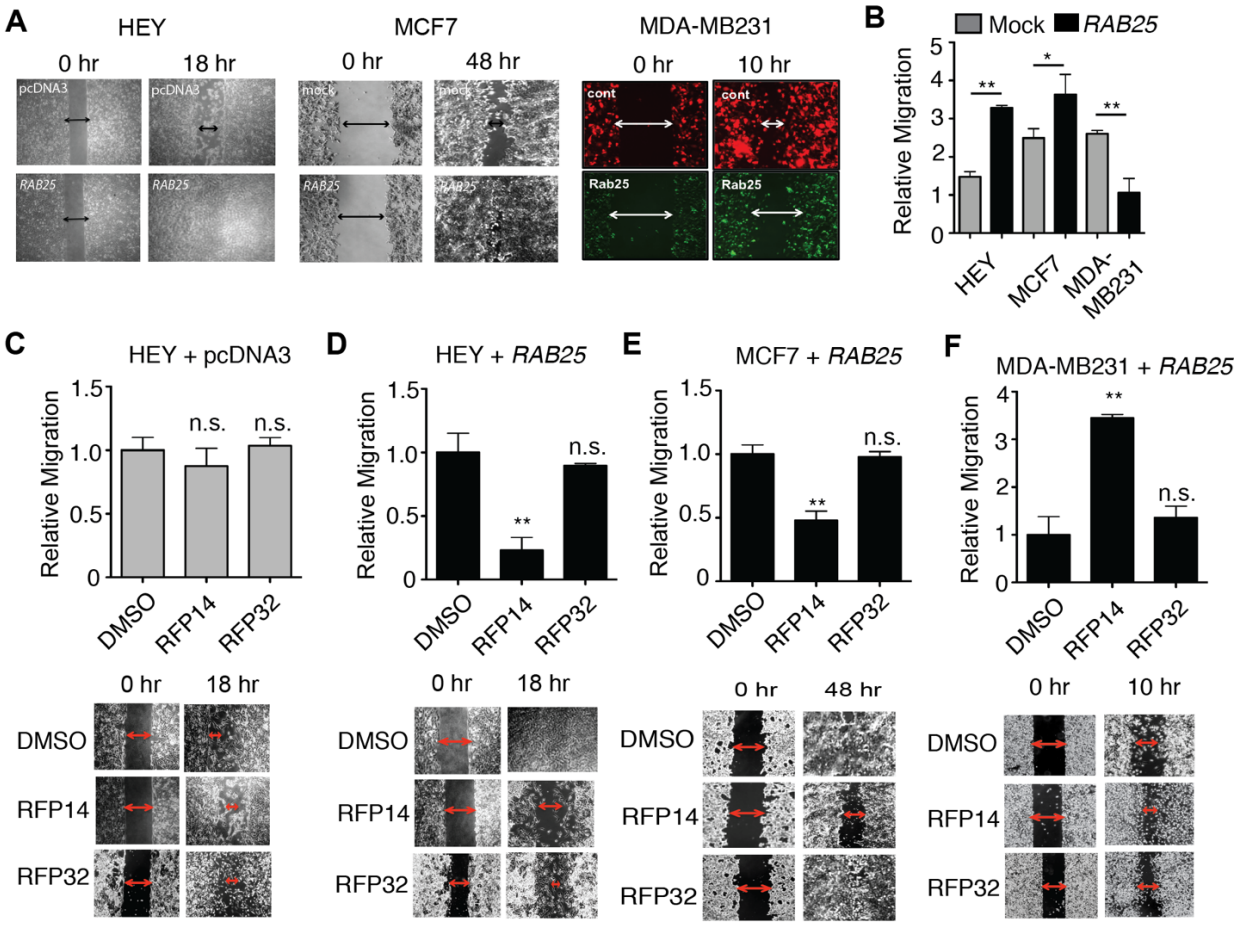


Figure 2.8: RFP14 specifically inhibits RAB25-dependent migration in cancer cells. **(A,B)** Representative time-dependent scratch-wound cell migration assay images and quantification, respectively, of mock vector (**A**, top) and RAB25-expressing (**A**, bottom) HEY, MCF7, and MDA-MB231 cells. RAB25 expression results in significant enhancement (HEY, MCF7) or inhibition (MDA-MB231) of cell migration. These data are representative of triplicate biological replicates. **(C,D) RFP 14** (10 μM) significantly impaired cell migration of RAB25-expressing (**D**), but not vector control (**C**), HEY cells. **(E) RFP14**, but not **RFP32**, significantly impaired cell migration of RAB25-expressing MCF7 cells. **(F)** Treatment of RAB25-expressing MDA-MB231 cells with **RFP14**, conversely, significantly increased cell migration relative to vehicle while the inactive control **RFP32** remained unchanged compared to vehicle control. Representative scratch wound images are shown for **RFP14**, **RFP32**, and vehicle (DMSO) treatment conditions with overlaid arrows highlighting distance between cell fronts. Images and data shown represent mean ± s.e.m. of triplicate biological replicates. *$P<0.05$; **$P<0.01$, Student's t-test. NS, not significant. Inset scale bars for HEY cells in **A** = 250 μm; MCF7 and MDA-MB231 cells in **A** = 100 μm. Inset scale bars for **C-E**, f = 250 μm.

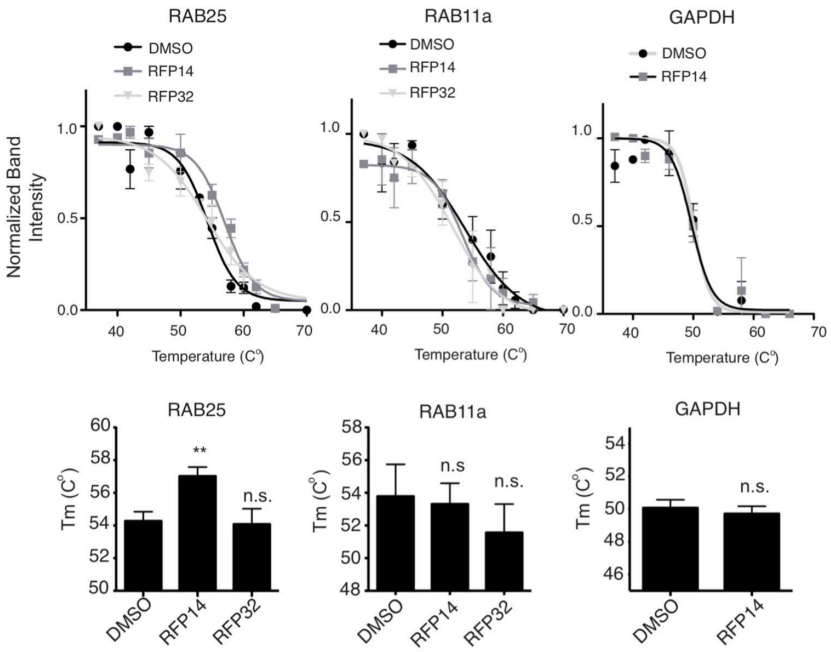


Figure 2.9: Cellular thermal shift assay (CETSA) denaturation curves (top) and Tm values (bottom) from indicated treatments in MCF7 cells.
Lysate samples were incubated with the indicated peptide at 10 μM for 30 min, aliquoted, and processed for CETSA analysis and western blot. The only significant and positive Tm-shift is seen for RAB25 with **RFP14** treatment, which is indicative of target-binding and thermal-stabilization. These results are representative of two or more technical replicates from two or more biological replicates. Data points represent the mean ± s.e.m. from three or more technical replicates. T_m values represent mean with s.e.m from the application of a sigmoidal curve fit using Prism 5 software.

Finally, to compliment the RAB25-dependent phenotypic assays employed here, we also sought to capture **RFP14** treatment effects on global signaling in a RAB25-dependent context. Specifically, we wanted to identify the gene expression changes enacted by RAB25-dependent signaling, and determine whether these program(s) are affected by **RFP14** treatment. To ask this question in a global and unbiased manner, we performed RNAseq on transcripts from pcDNA3-expressing and RAB25-expressing HEY cells. Expression profiling identified a set of 104 and 269 genes that were significantly increased (RAB25_UP gene set) and decreased (RAB25_DOWN gene set) with RAB25

expression, respectively (**Table 2.2**). In parallel, we performed global profiling of transcripts that changed in response to **RFP14** treatment, relative to DMSO, in RAB25-expressing HEY cells. Each RAB25-dependent gene set was queried for enrichment within the expression profile generated by comparing DMSO and RFP14 treated RAB25-HEY cells using GSEA [176], thereby asking whether RAB25-regulated genes are targeted by RFP14 in an unbiased manner and in the specific biological contexts associated with phenotypic changes. Both RAB25-dependent gene sets were significantly enriched in the RFP14 gene expression profile, but in the opposite direction (*i.e.* genes upregulated by RAB25 were downregulated by **RFP14**, **Figure 2.10 A-C**), and leading edge gene ontology analysis identified genes involved in ER and Golgi-mediated vesicular trafficking, small GTPase signaling, transmembrane signaling, GPCR signaling, and cAMP signaling, among other pathways (**Figure 2.11**). Targeted expression analysis with qPCR validated significant effects of **RFP14**, contrasted by minimal effects of the less active control **RFP32**, on the expression of enriched genes from functional categories associated with protein modification and vesicular trafficking (*MGAT1, PRSS23, LY6K*), migration (*AMIGO3*, *PRSS23*), GPCR signaling (*FPR1*, *GPER1*, *LPAR4, P2RY14*), regulation of transcription (*CEBPG*, *FOXI1*), as well as tumorigenesis (*IGF1*, *LPAR4*, *VHL*; **Figure 2.11**; **Table 2.4**). These expression profiles should provide hypotheses to further explore the downstream pathways enacted by RAB25 function and their link to pathogenic phenotypes in certain cancers. Furthermore, these data establish in an unbiased manner that **RFP14** treatment antagonizes the RAB25-dependent gene

expression program in a context where it also abrogates several RAB25-dependent phenotypes.

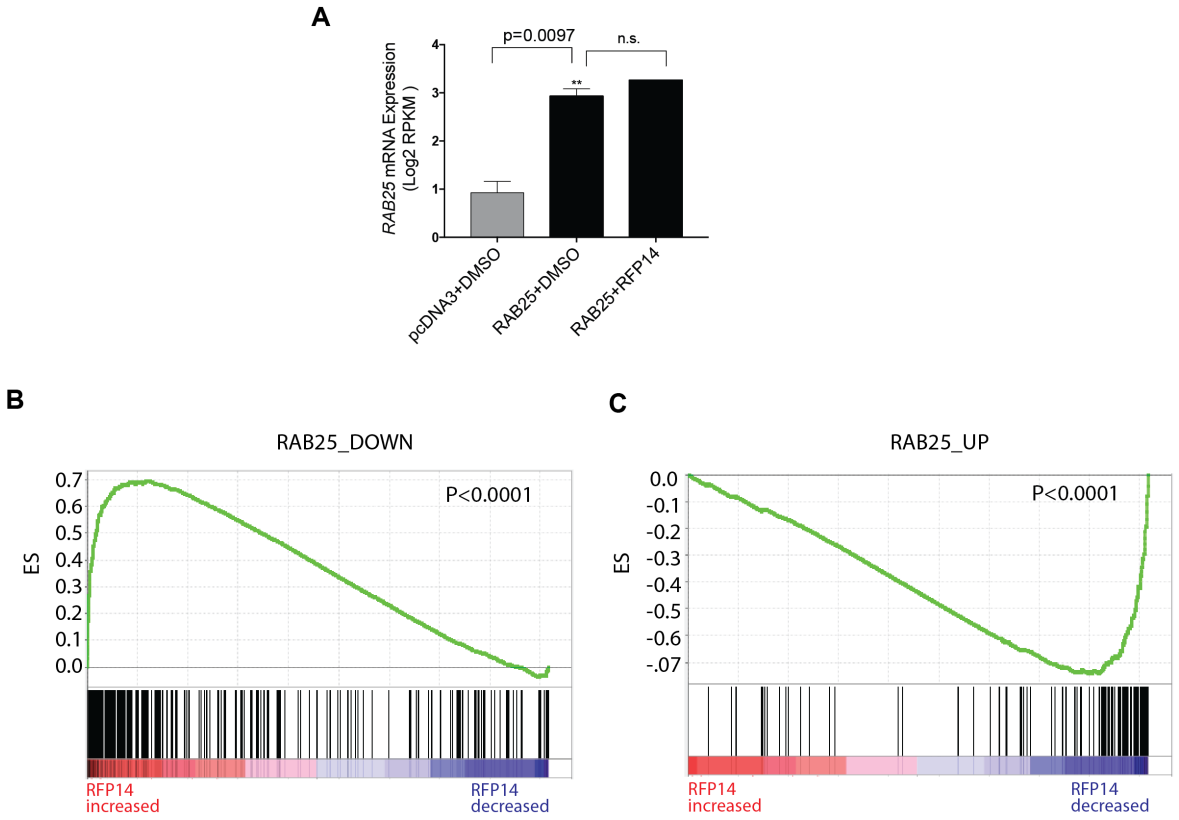


Figure 2.10: RFP14 inhibits RAB25-dependent gene expression in ovarian cancer cells.
(A) Relative RAB25 mRNA levels in pcDNA3-HEY cells treated with DMSO, RAB25-expressing HEY cells treated with DMSO, and RAB25-expressing HEY cells treated with **RFP14** measured by RNA-Seq profiling. **(B,C)** Quantitative GSEA plots of RAB25_UP and RAB25_DOWN gene set enrichment within the **RFP14** vs. DMSO gene expression profile in RAB25-expressing HEY cells. Both RAB25-depenent gene sets were significantly enriched in the **RFP14** vs. DMSO profile as indicated by the GSEA results for RAB25_DOWN (ES = 0.70, NES = 1.35, $P < 0.0001$) and RAB25_UP (ES = –0.74, NES = –1.61, $P < 0.0001$). The data represents two technical replicates from two independent biological replicate experiments. ES, enrichment score; NES, normalized enrichment score; NS, not significant. ***$P < 0.005$, Student's t-test.

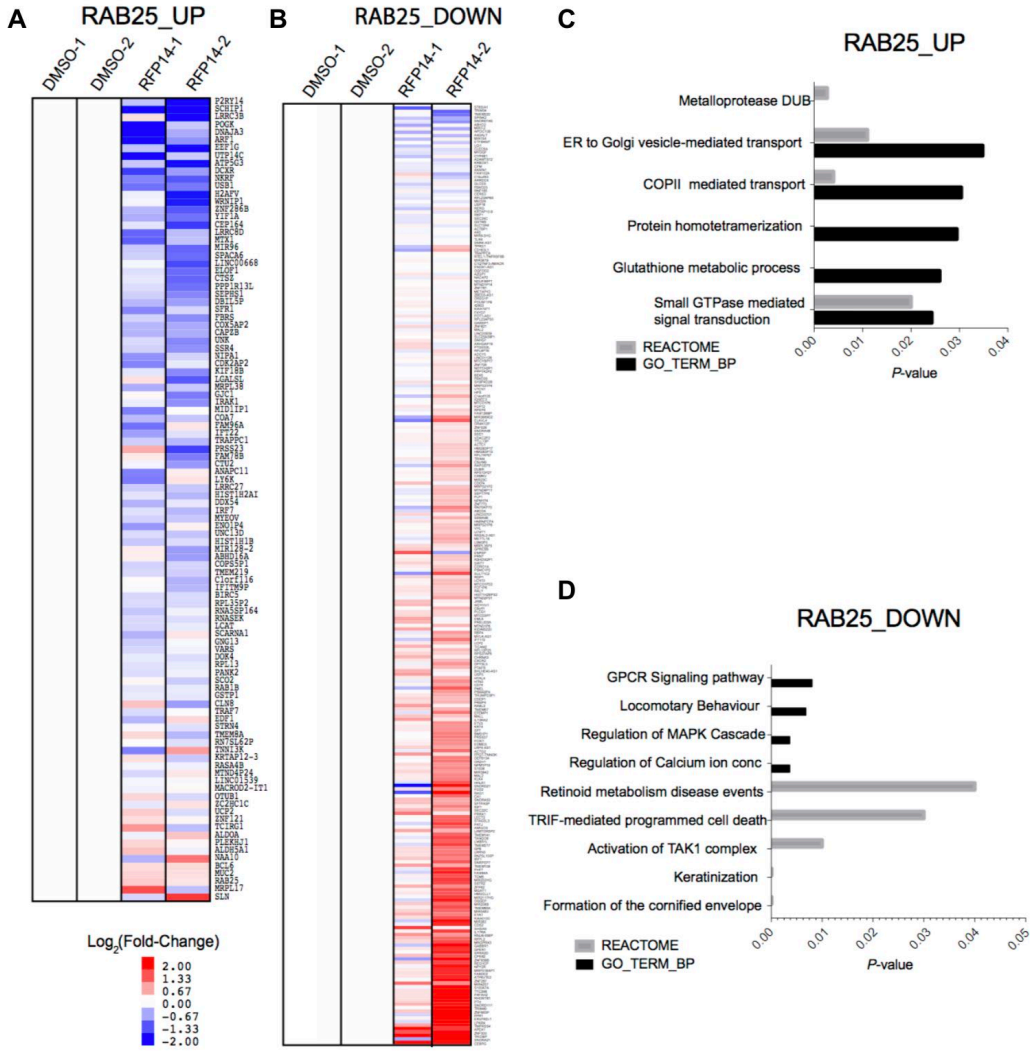


Figure 2.11: RFP14 opposes the global RAB25-dependent gene expression program in HEY ovarian cancer cells.
(A,B) Heat map representations showing the effect of **RFP14**, normalized on a per-gene basis to DMSO, on the expression of each gene in the RAB25_UP (**A**) and RAB25_DOWN (**B**) gene sets in RAB25-expressing HEY cells. Heat maps are derived from the GSEA analyses presented in **Figure 2.10**. **(C,D)** Gene ontology biological process categories (GOTERM_BP, identified by DAVID.

2.3 Discussion

While the critical role of RAB GTPases, and the broader small GTPase superfamily, in cancers and other diseases is increasingly evident, few cell-active probes have been developed targeting this class of proteins, or its broader superfamily [179,

180]. Recently, stabilized peptides targeting the RAB8a:RIP interaction resulted in probes with improved low micromolar affinities and protease resistance, but were not tested for cell activity [174, 181]. Using stapled peptides, we developed cell permeable, stabilized peptides capable of inhibiting RAB25-FIP complex formation in cells. We extensively studied the biochemical interactions of the FIP-RBD of several FIP classes, characterizing their engagement with both RAB11a and RAB25. Critically, we demonstrated the generation of RFP stapled peptides capable of engaging both proteins with high affinity and that these compounds can inhibit RAB25-FIP complex formation. **RFP14**, **RFP24** and **RFP26**, the three lead compounds applied to cell-based studies, all demonstrated sufficient cell permeability and improved binding for RAB25 compared to RAB11a. Such SARs are challenging to derive, design, and assess, therefore we carried forward several active compounds for extensive screening. **RFP14** exhibited preferential binding to RAB25 over RAB11a, and displayed RAB25-dependent effects in functional assays. Context-dependent effects of compound treatment on migration and proliferation phenotypes established **RFP14** antagonizes RAB25 leading to loss- and gain-of-function phenotypes, as well as effects on this protein's global gene expression program. Conversely, **RFP26** demonstrated RAB25-independent effects. Given **RFP26** retained the most binding to RAB11a, these effects may stem from inhibition of RAB11a-mediated signaling or other potential targets and effects. While the activity of **RFP14** was primarily studied in the context of its parent FIP-family protein-associated effects, further biochemical and cellular work is warranted to establish additional effects of this compound

and derivatives on global RAB-GTPase signaling networks, potential crosstalk between pathways, and impacts on cancer phenotypes.

Here we report the first cellular-active probes targeting RAB25, which suggest RAB25-FIP interfaces may serve as a pharmacologic handle for modulating RAB25 signaling. We believe the structure-activity relationships derived here will allow for the design of future stapled peptides or peptidomimetics targeting other RAB-family proteins (**Table 2.1**). Given the oncogenic role assigned to RAB25 in several cancers, **RFP14** represents a an entry point toward therapeutic development targeting RAB25 in cancers. Critically, we believe these first-in-class probes will enable, as established for chemical probes targeting other protein classes [182], future studies into RAB-family biochemistry and signal transduction both in normal and pathologic biological processes. Finally, this work serves as another instance of the development of synthetic biologic probes for targeting historically challenging proteins.

2.4 Tables

Peptide	Sequence	Apparent K_D, (95% C.I.) / (μM) RAB11a	RAB25	T_m at 222 nm	Cell Uptake	Inhibition of RAB25-dependent phenotypes
RFP1	βQVRELEDYIDNLLVRVMEETPNILRIP	>5	>5			
RFP2	βHIRELEDYIDNLLVRVMEETPSILRVP	>5	>5			
RFP3	βNFRLQDYIDRIIVAIMETNPSILEVK	2.3 (1.6-3.4)	>5			
RFP4	βNFRLRQYMDKIILAILDHNPSILEIK	0.93 (0.4-2.2)	2.3 (1.3-4.3)	37	-	-
RFP5	βNFRLQDYIDRIIVAIM$_{ox}$ETNPSILEVK	>5	>5			
RFP6	βRQVRELES_5YIDS_5LLVRVMEETPNILRIPR	>5	>5			
RFP7	βRQVRELEDYIDS_5LLVS_5VMEETPNILRIPR	0.11 (0.07-0.17)	0.51 (0.41-0.65)			
RFP8	βRQVRELEDYIDRLLVS_5VMES_5TPNILRIPR	0.04 (0.03-0.05)	0.17 (0.08-0.37)			
RFP9	βRQVRELEDS_5IDRS_5LVRVMEETPNILRIPR	2.4 (1.4-4.0)	1.2 (0.80-1.8)			
RFP10	βRINFRLQS_5YIDS_5IIVAIMETNPSILEVK	>5	>5			
RFP11	βRINFRLQDYIDS_5IIVS_5IMETNPSILEVK	0.007 (0.005-0.01)	0.05 (0.03-0.08)			
RFP12	βRINFRLQDYIDRIIVS_5IMES_5NPSILEVK	0.02 (0.007-0.06)	0.03 (0.02-0.06)			
RFP13	βRINFRLQDS_5IDRS_5IVAIMETNPSILEVK	0.08 (0.05-0.12)	0.04 (0.02-0.10)			
RFP14	βRQVRELENYIDRLLVS_5VN$_L$ES_5TPNILRIPR	0.32 (0.21-0.49)	0.04 (0.025-0.078)	60	++	+
RFP24	βRINFRLQNS_5IDRS_5IVRIFETNPSILRVK	0.56 (0.44-0.74)	0.16 (0.11-0.23)	52	++	+
RFP26	βRINFRLQNS_5IDRS_5IVRIN$_L$ETNPSILRVK	0.09 (0.064-0.12)	0.02 (0.016-0.029)	69	++	-
RFP31	βRINFRLQNS_5IDRS_5AARIAETNPSILEVK	>5	>5	59	+	-
RFP32	βRINFRLQNS_5IDRS_5AARIAETNPSILRVK	>5	>5	54	++	-

Table 2.1: Stapled RFP peptide structure activity relationships.
Name, sequence, binding affinities to RAB11a and RAB25, structural characteristics and cell- based data for the unmodified and stapled RFP peptides analyzed in this study. All blank positions represent assays that were not performed. Negative (-), positive (+) and double positive (++) signs are qualitative indicators of compound performance in an indicated assay. Relevant biochemical, structural and cell-based parameters for the main compounds utilized in this study are presented here for structure activity relationship analysis; additional quantitative annotations for the given compounds are presented in the main text.

PROBE	NAME	CHANGE
ABHD16A	abhydrolase domain containing 16A(ABHD16A)	UP
ALDH5A1	aldehyde dehydrogenase 5 family member A1(ALDH5A1)	UP
ALDOA	aldolase, fructose-bisphosphate A(ALDOA)	UP
ANAPC11	anaphase promoting complex subunit 11(ANAPC11)	UP
ARF1	ADP ribosylation factor 1(ARF1)	UP
ATP5G3	ATP synthase, H+ transporting, mitochondrial Fo complex subunit C3 (subunit 9)(ATP5G3)	UP
BCL6	B-cell CLL/lymphoma 6(BCL6)	UP
BIRC5	baculoviral IAP repeat containing 5(BIRC5)	UP
C1orf116	chromosome 1 open reading frame 116(C1orf116)	UP
CAPZB	capping actin protein of muscle Z-line beta subunit(CAPZB)	UP
CDK2AP2	cyclin dependent kinase 2 associated protein 2(CDK2AP2)	UP
CEP164	centrosomal protein 164(CEP164)	UP
CLN8	ceroid-lipofuscinosis, neuronal 8(CLN8)	UP
COA7	cytochrome c oxidase assembly factor 7 (putative)(COA7)	UP
COPS5P1	COP9 signalosome subunit 5 pseudogene 1(COPS5P1)	UP
COX5AP2	cytochrome c oxidase subunit 5A pseudogene 2(COX5AP2)	UP
CTSZ	cathepsin Z(CTSZ)	UP
CTU2	cytosolic thiouridylase subunit 2(CTU2)	UP
DBIL5P	diazepam binding inhibitor-like 5, pseudogene(DBIL5P)	UP
DCXR	dicarbonyl and L-xylulose reductase(DCXR)	UP
DDX54	DEAD-box helicase 54(DDX54)	UP
DNAJA3	DnaJ heat shock protein family (Hsp40) member A3(DNAJA3)	UP
DOK4	docking protein 4(DOK4)	UP
EDF1	endothelial differentiation related factor 1(EDF1)	UP
EEF1G	eukaryotic translation elongation factor 1 gamma(EEF1G)	UP
ELOF1	elongation factor 1 homolog(ELOF1)	UP
ENO1P4	enolase 1 pseudogene 4(ENO1P4)	UP
FAM78B	family with sequence similarity 78 member B(FAM78B)	UP
FAM96A	family with sequence similarity 96 member A(FAM96A)	UP
FBRS	fibrosin(FBRS)	UP
GJC1	gap junction protein gamma 1(GJC1)	UP
GNG13	G protein subunit gamma 13(GNG13)	UP
GSTP1	glutathione S-transferase pi 1(GSTP1)	UP
H2AFV	H2A histone family member V(H2AFV)	UP
HIST1H1B	histone cluster 1 H1 family member b(HIST1H1B)	UP
HIST1H2AI	histone cluster 1 H2A family member i(HIST1H2AI)	UP
IFITM9P	interferon induced transmembrane protein 9 pseudogene(IFITM9P)	UP
IFT22	intraflagellar transport 22(IFT22)	UP
IRAK1	interleukin 1 receptor associated kinase 1(IRAK1)	UP
IRF7	interferon regulatory factor 7(IRF7)	UP
KIF18B	kinesin family member 18B(KIF18B)	UP
KRTAP12-3	keratin associated protein 12-3(KRTAP12-3)	UP
LCAT	lecithin-cholesterol acyltransferase(LCAT)	UP
LGALSL	galectin like(LGALSL)	UP
LINC00668	long intergenic non-protein coding RNA 668(LINC00668)	UP
LINC01539	long intergenic non-protein coding RNA 1539(LINC01539)	UP
LRRC27	leucine rich repeat containing 27(LRRC27)	UP
LRRC3B	leucine rich repeat containing 3B(LRRC3B)	UP
LRRC8D	leucine rich repeat containing 8 family member D(LRRC8D)	UP
LY6K	lymphocyte antigen 6 complex, locus K(LY6K)	UP
MACROD2-IT	MACROD2 intronic transcript 1(MACROD2-IT1)	UP

Table 2.2: RAB25_UP and RAB25_DOWN gene sets.

PROBE	NAME	CHANGE
MID1IP1	MID1 interacting protein 1(MID1IP1)	UP
MIR128-2	microRNA 128-2(MIR128-2)	UP
MIR96	microRNA 96(MIR96)	UP
MRPL17	mitochondrial ribosomal protein L17(MRPL17)	UP
MRPL38	mitochondrial ribosomal protein L38(MRPL38)	UP
MTND4P24	mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 4 pseudogene 24(MTND4P24)	UP
MTX1	metaxin 1(MTX1)	UP
MUC2	mucin 2, oligomeric mucus/gel-forming(MUC2)	UP
MYEOV	myeloma overexpressed(MYEOV)	UP
NAA10	N(alpha)-acetyltransferase 10, NatA catalytic subunit(NAA10)	UP
NIPA1	non imprinted in Prader-Willi/Angelman syndrome 1(NIPA1)	UP
NKRF	NFKB repressing factor(NKRF)	UP
OTUB1	OTU deubiquitinase, ubiquitin aldehyde binding 1(OTUB1)	UP
P2RY14	purinergic receptor P2Y14(P2RY14)	UP
PANK2	pantothenate kinase 2(PANK2)	UP
PLEKHJ1	pleckstrin homology domain containing J1(PLEKHJ1)	UP
POGK	pogo transposable element with KRAB domain(POGK)	UP
PPP1R13L	protein phosphatase 1 regulatory subunit 13 like(PPP1R13L)	UP
PRSS23	protease, serine 23(PRSS23)	UP
RAB1B	RAB1B, member RAS oncogene family(RAB1B)	UP
RAB25	RAB25, member RAS oncogene family(RAB25)	UP
RASA4B	RAS p21 protein activator 4B(RASA4B)	UP
RN7SL62P	RNA, 7SL, cytoplasmic 62, pseudogene(RN7SL62P)	UP
RNA5SP164	RNA, 5S ribosomal pseudogene 164(RNA5SP164)	UP
RNASEK	ribonuclease K(RNASEK)	UP
RPL13	ribosomal protein L13(RPL13)	UP
RPL35P2	ribosomal protein L35 pseudogene 2(RPL35P2)	UP
SCARNA1	small Cajal body-specific RNA 1(SCARNA1)	UP
SCHIP1	schwannomin interacting protein 1(SCHIP1)	UP
SCO2	SCO2, cytochrome c oxidase assembly protein(SCO2)	UP
SEPHS1	selenophosphate synthetase 1(SEPHS1)	UP
SFR1	SWI5 dependent homologous recombination repair protein 1(SFR1)	UP
SLN	sarcolipin(SLN)	UP
SPACA6	sperm acrosome associated 6(SPACA6)	UP
SSR4	signal sequence receptor subunit 4(SSR4)	UP
STRN4	striatin 4(STRN4)	UP
TCIRG1	T-cell immune regulator 1, ATPase H+ transporting V0 subunit a3(TCIRG1)	UP
TMEM219	transmembrane protein 219(TMEM219)	UP
TMEM8A	transmembrane protein 8A(TMEM8A)	UP
TNNI3K	TNNI3 interacting kinase(TNNI3K)	UP
TRAF7	TNF receptor associated factor 7(TRAF7)	UP
TRAPPC1	trafficking protein particle complex 1(TRAPPC1)	UP
UCP2	uncoupling protein 2(UCP2)	UP
UNC13D	unc-13 homolog D(UNC13D)	UP
UNK	unkempt family zinc finger(UNK)	UP
USB1	U6 snRNA biogenesis phosphodiesterase 1(USB1)	UP
UTP14C	UTP14, small subunit processome component homolog C (S. cerevisiae)(UTP14C)	UP
VARS	valyl-tRNA synthetase(VARS)	UP
WRNIP1	Werner helicase interacting protein 1(WRNIP1)	UP
YIF1A	Yip1 interacting factor homolog A, membrane trafficking protein(YIF1A)	UP
ZC2HC1C	zinc finger C2HC-type containing 1C(ZC2HC1C)	UP
ZNF121	zinc finger protein 121(ZNF121)	UP
ZNF286B	zinc finger protein 286B(ZNF286B)	UP

Table 2.2, continued.

PROBE	NAME	CHANGE	PROBE	NAME	CHANGE
A4GALT	alpha 1,4-galactosyltransferase(A4GALT)	DOWN	FAM102A	family with sequence similarity 102 member A(FAM102A)	DOWN
ABCD4	ATP binding cassette subfamily D member 4(ABCD4)	DOWN	FAM136BP	family with sequence similarity 136 member B, pseudogen	DOWN
ABHD2	abhydrolase domain containing 2(ABHD2)	DOWN	FAM84A	family with sequence similarity 84 member A(FAM84A)	DOWN
ACTBP1	actin, beta pseudogene 1(ACTBP1)	DOWN	FAXDC2	fatty acid hydroxylase domain containing 2(FAXDC2)	DOWN
ACTC1	actin, alpha, cardiac muscle 1(ACTC1)	DOWN	FBXO25	F-box protein 25(FBXO25)	DOWN
ACTG2	actin, gamma 2, smooth muscle, enteric(ACTG2)	DOWN	FBXO36	F-box protein 36(FBXO36)	DOWN
ADAMTS12	ADAM metallopeptidase with thrombospondin type 1 moti	DOWN	FCF1P6	FCF1 pseudogene 6(FCF1P6)	DOWN
ADCY5	adenylate cyclase 5(ADCY5)	DOWN	FGD2	FYVE, RhoGEF and PH domain containing 2(FGD2)	DOWN
AK5	adenylate kinase 5(AK5)	DOWN	FGF12	fibroblast growth factor 12(FGF12)	DOWN
AKAIN1	A-kinase anchor inhibitor 1(AKAIN1)	DOWN	FOXI1	forkhead box I1(FOXI1)	DOWN
AMIGO3	adhesion molecule with Ig like domain 3(AMIGO3)	DOWN	FPGT-TNNI3K	FPGT-TNNI3K readthrough(FPGT-TNNI3K)	DOWN
APEX1	apurinic/apyrimidinic endodeoxyribonuclease 1(APEX1)	DOWN	FPR1	formyl peptide receptor 1(FPR1)	DOWN
ARHGAP19	Rho GTPase activating protein 19(ARHGAP19)	DOWN	FXYD7	FXYD domain containing ion transport regulator 7(FXYD7)	DOWN
ARRDC4	arrestin domain containing 4(ARRDC4)	DOWN	GABBR1	gamma-aminobutyric acid type B receptor subunit 1(GABB	DOWN
ATP6V1E2	ATPase H+ transporting V1 subunit E2(ATP6V1E2)	DOWN	GARSP1	glycyl-tRNA synthetase pseudogene 1(GARSP1)	DOWN
AZGP1	alpha-2-glycoprotein 1, zinc-binding(AZGP1)	DOWN	GLOD5	glyoxalase domain containing 5(GLOD5)	DOWN
BEX5	brain expressed X-linked 5(BEX5)	DOWN	GP6	glycoprotein VI platelet(GP6)	DOWN
BHLHE40-AS1	BHLHE40 antisense RNA 1(BHLHE40-AS1)	DOWN	GPER1	G protein-coupled estrogen receptor 1(GPER1)	DOWN
BMS1P1	Ribosome Biogenesis Factor Pseudogene	DOWN	GPRC5B	G protein-coupled receptor class C group 5 member B(GPR	DOWN
C14orf105	chromosome 14 open reading frame 105(C14orf105)	DOWN	GSTM3	glutathione S-transferase mu 3(GSTM3)	DOWN
C19orf45	chromosome 19 open reading frame 45(C19orf45)	DOWN	HHLA1	HERV-H LTR-associating 1(HHLA1)	DOWN
C1QTNF3-AMACR	C1QTNF3-AMACR readthrough (NMD candidate)(C1QTNF3	DOWN	HIST1H2BPS2	histone cluster 1 H2B pseudogene 2(HIST1H2BPS2)	DOWN
C5orf49	chromosome 5 open reading frame 49(C5orf49)	DOWN	HMGB3P13	high mobility group box 3 pseudogene 13(HMGB3P13)	DOWN
C6orf1	chromosome 6 open reading frame 1(C6orf1)	DOWN	HMGB3P17	high mobility group box 3 pseudogene 17(HMGB3P17)	DOWN
CA1	carbonic anhydrase 1(CA1)	DOWN	HMGCLL1	3-hydroxymethyl-3-methylglutaryl-CoA lyase like 1(HMGCL	DOWN
CAMKV	CaM kinase like vesicle associated(CAMKV)	DOWN	HNRNPCP4	heterogeneous nuclear ribonucleoprotein C pseudogene 4(	DOWN
CD163L1	CD163 molecule like 1(CD163L1)	DOWN	HPX	hemopexin(HPX)	DOWN
CD74	CD74 molecule(CD74)	DOWN	HTN3	histatin 3(HTN3)	DOWN
CDCP1	CUB domain containing protein 1(CDCP1)	DOWN	HYAL4	hyaluronoglucosaminidase 4(HYAL4)	DOWN
CDON	cell adhesion associated, oncogene regulated(CDON)	DOWN	IFT172	intraflagellar transport 172(IFT172)	DOWN
CDS2	CDP-diacylglycerol synthase 2(CDS2)	DOWN	IGF1	insulin like growth factor 1(IGF1)	DOWN
CEBPG	CCAAT/enhancer binding protein gamma(CEBPG)	DOWN	IL13RA2	interleukin 13 receptor subunit alpha 2(IL13RA2)	DOWN
CERS3	ceramide synthase 3(CERS3)	DOWN	IL17RA	interleukin 17 receptor A(IL17RA)	DOWN
CHRNA3	cholinergic receptor nicotinic alpha 3 subunit(CHRNA3)	DOWN	IQSEC3	IQ motif and Sec7 domain 3(IQSEC3)	DOWN
CLEC5A	C-type lectin domain family 5 member A(CLEC5A)	DOWN	IRF1	interferon regulatory factor 1(IRF1)	DOWN
CORO1A	coronin 1A(CORO1A)	DOWN	JAML	junction adhesion molecule like(JAML)	DOWN
CPEB2	cytoplasmic polyadenylation element binding protein 2(CP	DOWN	KIAA0100	KIAA0100(KIAA0100)	DOWN
CPM	carboxypeptidase M(CPM)	DOWN	KIAA1671	KIAA1671(KIAA1671)	DOWN
CXCR2	C-X-C motif chemokine receptor 2(CXCR2)	DOWN	KIDINS220	kinase D-interacting substrate 220kDa(KIDINS220)	DOWN
CYP4B1	cytochrome P450 family 4 subfamily B member 1(CYP4B1)	DOWN	KLK4	kallikrein related peptidase 4(KLK4)	DOWN
DEFB134	defensin beta 134(DEFB134)	DOWN	KRBOX1	KRAB box domain containing 1(KRBOX1)	DOWN
DPYSL5	dihydropyrimidinase like 5(DPYSL5)	DOWN	KRT4	keratin 4(KRT4)	DOWN
DUBR	DPPA2 upstream binding RNA(DUBR)	DOWN	KRTAP10-9	keratin associated protein 10-9(KRTAP10-9)	DOWN
EFEMP1	EGF containing fibulin like extracellular matrix protein 1(EF	DOWN	LAMTOR3P2	late endosomal/lysosomal adaptor, MAPK and MTOR activ	DOWN
ELAVL4	ELAV like RNA binding protein 4(ELAVL4)	DOWN	LCMT1	leucine carboxyl methyltransferase 1(LCMT1)	DOWN
EML6	echinoderm microtubule associated protein like 6(EML6)	DOWN	LCN10	lipocalin 10(LCN10)	DOWN
ENOX1-AS1	ENOX1 antisense RNA 1(ENOX1-AS1)	DOWN	LCP2	lymphocyte cytosolic protein 2(LCP2)	DOWN
ENPEP	glutamyl aminopeptidase(ENPEP)	DOWN	LECT2	leukocyte cell derived chemotaxin 2(LECT2)	DOWN
EOMES	eomesodermin(EOMES)	DOWN	LGI1	leucine rich glioma inactivated 1(LGI1)	DOWN
ERVFRD-1	endogenous retrovirus group FRD member 1(ERVFRD-1)	DOWN	LINC00701	long intergenic non-protein coding RNA 701(LINC00701)	DOWN
ETFBKMT	electron transfer flavoprotein beta subunit lysine methyltr	DOWN	LINC00939	long intergenic non-protein coding RNA 939(LINC00939)	DOWN
ETV3	ETS variant 3(ETV3)	DOWN	LINC01128	long intergenic non-protein coding RNA 1128(LINC01128)	DOWN
EYA1	EYA transcriptional coactivator and phosphatase 1(EYA1)	DOWN	LMBR1L	limb development membrane protein 1 like(LMBR1L)	DOWN

Table 2.2, continued.

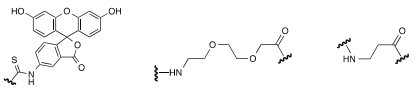
PROBE	NAME	CHANGE	PROBE	NAME	CHANGE
LPAR4	lysophosphatidic acid receptor 4(LPAR4)	DOWN	PKMP4	pyruvate kinase, muscle pseudogene 4(PKMP4)	DOWN
LRP4-AS1	LRP4 antisense RNA 1(LRP4-AS1)	DOWN	PLCG1	phospholipase C gamma 1(PLCG1)	DOWN
LRRN3	leucine rich repeat neuronal 3(LRRN3)	DOWN	PLP1	proteolipid protein 1(PLP1)	DOWN
LSM3P5	LSM3 homolog, U6 small nuclear RNA and mRNA degradati	DOWN	PMEL	premelanosome protein(PMEL)	DOWN
MAL2	mal, T-cell differentiation protein 2 (gene/pseudogene)(MA	DOWN	POT1-AS1	POT1 antisense RNA 1(POT1-AS1)	DOWN
MALL	mal, T-cell differentiation protein like(MALL)	DOWN	POU5F1P6	POU class 5 homeobox 1 pseudogene 6(POU5F1P6)	DOWN
MED28	mediator complex subunit 28(MED28)	DOWN	PPP1R2P2	protein phosphatase 1 regulatory inhibitor subunit 2 pseuc	DOWN
METAP1D	methionyl aminopeptidase type 1D, mitochondrial(METAP	DOWN	PRELID3A	PRELI domain containing 3A(PRELID3A)	DOWN
METTL16	methyltransferase like 16(METTL16)	DOWN	PRNT	prion protein (testis specific)(PRNT)	DOWN
MGAT1	mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylgluco	DOWN	PRRX1	paired related homeobox 1(PRRX1)	DOWN
MIR154	microRNA 154(MIR154)	DOWN	PRSS37	protease, serine 37(PRSS37)	DOWN
MIR202HG	MIR202 host gene(MIR202HG)	DOWN	PSMA6P4	proteasome subunit alpha 6 pseudogene 4(PSMA6P4)	DOWN
MIR208B	microRNA 208b(MIR208B)	DOWN	PSMC1P2	proteasome 26S subunit, ATPase 1 pseudogene 2(PSMC1P2	DOWN
MIR2117HG	MIR2117 host gene(MIR2117HG)	DOWN	PTAFR	platelet activating factor receptor(PTAFR)	DOWN
MIR23C	microRNA 23c(MIR23C)	DOWN	PTGES3L	prostaglandin E synthase 3 like(PTGES3L)	DOWN
MIR362	microRNA 362(MIR362)	DOWN	PTH	parathyroid hormone(PTH)	DOWN
MIR3679	microRNA 3679(MIR3679)	DOWN	R3HDM2P1	R3H domain containing 2 pseudogene 1(R3HDM2P1)	DOWN
MIR3689D2	microRNA 3689d-2(MIR3689D2)	DOWN	RABL3	RAB, member of RAS oncogene family like 3(RABL3)	DOWN
MIR3942	microRNA 3942(MIR3942)	DOWN	RAG1	recombination activating 1(RAG1)	DOWN
MIR4257	microRNA 4257(MIR4257)	DOWN	RALY	RALY heterogeneous nuclear ribonucleoprotein(RALY)	DOWN
MIR548V	microRNA 548v(MIR548V)	DOWN	RAPGEF5	Rap guanine nucleotide exchange factor 5(RAPGEF5)	DOWN
MIR7-2	microRNA 7-2(MIR7-2)	DOWN	RASAL2-AS1	RASAL2 antisense RNA 1(RASAL2-AS1)	DOWN
MIR9-3HG	MIR9-3 host gene(MIR9-3HG)	DOWN	RBP1	retinol binding protein 1(RBP1)	DOWN
MRGPRX3	MAS related GPR family member X3(MRGPRX3)	DOWN	RBP4	retinol binding protein 4(RBP4)	DOWN
MRPL35P3	mitochondrial ribosomal protein L35 pseudogene 3(MRPL3	DOWN	REG1CP	regenerating family member 1 gamma, pseudogene(REG1(	DOWN
MRPS18AP1	mitochondrial ribosomal protein S18A pseudogene 1(MRP	DOWN	RERG	RAS like estrogen regulated growth inhibitor(RERG)	DOWN
MRPS21P2	mitochondrial ribosomal protein S21 pseudogene 2(MRPS2	DOWN	RFPL2	ret finger protein like 2(RFPL2)	DOWN
MRPS21P8	mitochondrial ribosomal protein S21 pseudogene 8(MRPS2	DOWN	RGP1	RGP1 homolog, RAB6A GEF complex partner 1(RGP1)	DOWN
MRPS31P4	mitochondrial ribosomal protein S31 pseudogene 4(MRPS3	DOWN	RHOBTB1	Rho related BTB domain containing 1(RHOBTB1)	DOWN
MTCO1P22	mitochondrially encoded cytochrome c oxidase I pseudoge	DOWN	RN7SKP70	RNA, 7SK small nuclear pseudogene 70(RN7SKP70)	DOWN
MTCO1P6	mitochondrially encoded cytochrome c oxidase I pseudoge	DOWN	RN7SL100P	RNA, 7SL, cytoplasmic 100, pseudogene(RN7SL100P)	DOWN
MTCO3P7	mitochondrially encoded cytochrome c oxidase III pseudog	DOWN	RNF165	ring finger protein 165(RNF165)	DOWN
MTCYBP17	mitochondrially encoded cytochrome b pseudogene 17(MT	DOWN	RNU6-656P	RNA, U6 small nuclear 656, pseudogene(RNU6-656P)	DOWN
MTND1P14	mitochondrially encoded NADH:ubiquinone oxidoreductase	DOWN	RPEP4	ribulose-5-phosphate-3-epimerase pseudogene 4(RPEP4)	DOWN
MTND1P6	mitochondrially encoded NADH:ubiquinone oxidoreductase	DOWN	RPL12P25	ribosomal protein L12 pseudogene 25(RPL12P25)	DOWN
MTND2P31	mitochondrially encoded NADH:ubiquinone oxidoreductase	DOWN	RPL23AP30	ribosomal protein L23a pseudogene 30(RPL23AP30)	DOWN
MTND6P11	mitochondrially encoded NADH:ubiquinone oxidoreductase	DOWN	RPL23AP66	ribosomal protein L23a pseudogene 66(RPL23AP66)	DOWN
MYDGF	myeloid derived growth factor(MYDGF)	DOWN	RPL7AP57	ribosomal protein L7a pseudogene 57(RPL7AP57)	DOWN
MYLK-AS1	MYLK antisense RNA 1(MYLK-AS1)	DOWN	RPL9P16	ribosomal protein L9 pseudogene 16(RPL9P16)	DOWN
NACAP2	nascent polypeptide associated complex alpha subunit pse	DOWN	RPS10P27	ribosomal protein S10 pseudogene 27(RPS10P27)	DOWN
NDUFB4P1	NADH:ubiquinone oxidoreductase subunit B4 pseudogene	DOWN	RPS27AP6	ribosomal protein S27a pseudogene 6(RPS27AP6)	DOWN
NOTCH2P1	notch 2 pseudogene 1(NOTCH2P1)	DOWN	RTEL1-TNFRSF	RTEL1-TNFRSF6B readthrough (NMD candidate)(RTEL1-TNF	DOWN
NPM1P19	nucleophosmin 1 pseudogene 19(NPM1P19)	DOWN	S100A7A	S100 calcium binding protein A7A(S100A7A)	DOWN
NPM1P4	nucleophosmin 1 pseudogene 4(NPM1P4)	DOWN	S100B	S100 calcium binding protein B(S100B)	DOWN
NPY2R	neuropeptide Y receptor Y2(NPY2R)	DOWN	SDC1	syndecan 1(SDC1)	DOWN
OGFOD2	2-oxoglutarate and iron dependent oxygenase domain con	DOWN	SEC22C	SEC22 homolog C, vesicle trafficking protein(SEC22C)	DOWN
OR2G1P	olfactory receptor family 2 subfamily G member 1 pseudog	DOWN	SEC24C	SEC24 homolog C, COPII coat complex component(SEC24C)	DOWN
OR2H1	olfactory receptor family 2 subfamily H member 1(OR2H1)	DOWN	SEMA4B	semaphorin 4B(SEMA4B)	DOWN
OR4K12P	olfactory receptor family 4 subfamily K member 12 pseudo	DOWN	SEPT7P9	septin 7 pseudogene 9(SEPT7P9)	DOWN
OSGEP	O-sialoglycoprotein endopeptidase(OSGEP)	DOWN	SFTPA3P	surfactant protein A3, pseudogene(SFTPA3P)	DOWN
PAFAH2	platelet activating factor acetylhydrolase 2(PAFAH2)	DOWN	SH3PXD2B	SH3 and PX domains 2B(SH3PXD2B)	DOWN
PATJ	PATJ, crumbs cell polarity complex component(PATJ)	DOWN	SHISA9	shisa family member 9(SHISA9)	DOWN
PHF7	PHD finger protein 7(PHF7)	DOWN	SIRT7	sirtuin 7(SIRT7)	DOWN

Table 2.2, continued.

PROBE	NAME	CHANGE
SLC13A4	solute carrier family 13 member 4(SLC13A4)	DOWN
SLC25A39P1	solute carrier family 25 member 39 pseudogene 1(SLC25A3	DOWN
SNHG7	small nucleolar RNA host gene 7(SNHG7)	DOWN
SNORA21	small nucleolar RNA, H/ACA box 21(SNORA21)	DOWN
SNORA48	small nucleolar RNA, H/ACA box 48(SNORA48)	DOWN
SNORA53	small nucleolar RNA, H/ACA box 53(SNORA53)	DOWN
SNORD111	small nucleolar RNA, C/D box 111(SNORD111)	DOWN
SNORD14E	small nucleolar RNA, C/D box 14E(SNORD14E)	DOWN
SNORD21	small nucleolar RNA, C/D box 21(SNORD21)	DOWN
SNRK-AS1	SNRK antisense RNA 1(SNRK-AS1)	DOWN
SNRPEP7	small nuclear ribonucleoprotein polypeptide E pseudogene	DOWN
SP7	Sp7 transcription factor(SP7)	DOWN
SPINK2	serine peptidase inhibitor, Kazal type 2(SPINK2)	DOWN
SPRR2D	small proline rich protein 2D(SPRR2D)	DOWN
SSTR2	somatostatin receptor 2(SSTR2)	DOWN
ST8SIA1	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase	DOWN
STAG3L3	stromal antigen 3-like 3 (pseudogene)(STAG3L3)	DOWN
SULT1C2	sulfotransferase family 1C member 2(SULT1C2)	DOWN
TANGO6	transport and golgi organization 6 homolog(TANGO6)	DOWN
TGM5	transglutaminase 5(TGM5)	DOWN
THUMPD3P1	THUMP domain containing 3 pseudogene 1(THUMPD3P1)	DOWN
TICAM2	toll like receptor adaptor molecule 2(TICAM2)	DOWN
TLR4	toll like receptor 4(TLR4)	DOWN
TMEM141	transmembrane protein 141(TMEM141)	DOWN
TMEM158	transmembrane protein 158 (gene/pseudogene)(TMEM15	DOWN
TMEM217	transmembrane protein 217(TMEM217)	DOWN
TMEM220	transmembrane protein 220(TMEM220)	DOWN
TMEM67	transmembrane protein 67(TMEM67)	DOWN
TMEM86A	transmembrane protein 86A(TMEM86A)	DOWN
TMPRSS4	transmembrane protease, serine 4(TMPRSS4)	DOWN
TPRG1	tumor protein p63 regulated 1(TPRG1)	DOWN
TRAPPC9	trafficking protein particle complex 9(TRAPPC9)	DOWN
TRIM34	tripartite motif containing 34(TRIM34)	DOWN
TRIM4	tripartite motif containing 4(TRIM4)	DOWN
TRIM49	tripartite motif containing 49(TRIM49)	DOWN
TRIOBP	TRIO and F-actin binding protein(TRIOBP)	DOWN
TTC39B	tetratricopeptide repeat domain 39B(TTC39B)	DOWN
TTLL13P	tubulin tyrosine ligase like 13, pseudogene(TTLL13P)	DOWN
USP18	ubiquitin specific peptidase 18(USP18)	DOWN
USP5	ubiquitin specific peptidase 5(USP5)	DOWN
VDAC2P2	voltage dependent anion channel 2 pseudogene 2(VDAC2P	DOWN
VHL	von Hippel-Lindau tumor suppressor(VHL)	DOWN
VTCN1	V-set domain containing T cell activation inhibitor 1(VTCN1	DOWN
WDYHV1	WDYHV motif containing 1(WDYHV1)	DOWN
WFDC10B	WAP four-disulfide core domain 10B(WFDC10B)	DOWN
ZBED3-AS1	ZBED3 antisense RNA 1(ZBED3-AS1)	DOWN
ZFP62	ZFP62 zinc finger protein(ZFP62)	DOWN
ZNF287	zinc finger protein 287(ZNF287)	DOWN
ZNF300	zinc finger protein 300(ZNF300)	DOWN
ZNF528	zinc finger protein 528(ZNF528)	DOWN
ZNF658B	zinc finger protein 658B (pseudogene)(ZNF658B)	DOWN
ZNF663P	zinc finger protein 663, pseudogene(ZNF663P)	DOWN
ZNF726	zinc finger protein 726(ZNF726)	DOWN
ZNF773	zinc finger protein 773(ZNF773)	DOWN
ZNF781	zinc finger protein 781(ZNF781)	DOWN
ZNF821	zinc finger protein 821(ZNF821)	DOWN

Table 2.2, continued: Gene symbols and descriptions for genes identified as significantly increased (red) and decreased (blue) in RAB25-expressing vs. pcDNA3 control HEY ovarian cancer cells.

Peptide	Sequence	Retention Time (min)	m/z expected / (m/z found)
fRFP1	FITC-(PEG)-KKEFQVRELEDYIDNLLVRVMEETPNILRIPA	8.0	1469.4, 1102.3 /(1469.3, 1102.2)
fRFP2	FITC-(PEG)-ERDTHIRELEDYIDNLLVRVMEETPSILRVPR	8.0	1483.7, 1113.0 /(1483.6, 1112.9)
fRFP3	FITC-(PEG)-KQEEINFRLQDYIDRIIVAIMETNPSILEVK	7.8	1418.3, 1064.0 /(1418.2, 1064.1)
fRFP4	FITC-(PEG)-EQEEINFRLRQYMDKIILAILDHNPSILEIKH	7.5	1485.4, 1114.3 /(1485.3, 1114.3)
fRFP5	FITC-(PEG)-KQEEINFRLQDYIDRIIVAIM$_{ox}$ETNPSILEVK	7.1	1423.6, 1068.0 /(1423.6, 1068.0)
fRFP6	FITC-(PEG)-βRQVRELES_5YIDS_5LLVRVMEETPNILRIPR	9.1	1423.6, 1068.0 /(1423.6, 1068.0)
fRFP7	FITC-(PEG)-βRQVRELEDYIDS_5LLVS_5VMEETPNILRIPR	9.1	1389.3, 1042.2 /(1389.2, 1042.2)
fRFP8	FITC-(PEG)-βRQVRELEDYIDRLLVS_5VMES_5TPNILRIPR	9.6	1398.3, 1049.0 /(1398.2, 1049.0)
fRFP9	FITC-(PEG)-βRQVRELEDS_5IDRS_5LVRVMEETPNILRIPR	8.6	1401.3, 1051.2 /(1401.3, 1051.2)
fRFP10	FITC-(PEG)-βRINFRLQS_5YIDS_5IVAIMETNPSILEVK	8.3	1343.6, 1007.9 /(1343.3, 1007.9)
fRFP11	FITC-(PEG)-βRINFRLQDYIDS_5IIVS_5IMETNPSILEVK	8.6	1330.5, 998.2 /(1330.0, 998.2)
fRFP12	FITC-(PEG)-βRINFRLQDYIDRIIVS_5IMES_5NPSILEVK	8.4	1348.9, 1011.9 /(1348.3, 1012.2)
fRFP13	FITC-(PEG)-βRINFRLQDS_5IDRS_5IVAIMETNPSILEVK	8.0	1351.6, 1014.0 /(1351.3, 1013.9)
bRFP14	biot-PEG-WβRQVRELENYIDRLLVS_5VN$_L$ES_5TPNILRIPR	9.7	1400.0, 1050.3 /(1399.5, 1050.1)
bRFP24	biot-PEG-WβRINFRLQNS_5IDRS_5IVRIFETNPSILRVK	7.7	1364.3, 1023.5 /(1363.9, 1023.2)
bRFP26	biot-PEG-WβRINFRLQNS_5IDRS_5IVRIN$_L$ETNPSILRVK	7.7	1353.0, 1015.0 /(1352.6, 1014.6)
bRFP31	biot-PEG-WβRINFRLQNS_5IDRS_5AARIAETNPSILEVK	7.2	1306.5, 980.2 /(1306.0, 979.8)
fRFP14	FITC-PEG-βRQVRELENYIDRLLVS_5VN$_L$ES_5TPNILRIPR	15.9*	1392.0, 1044.2 /(1391.9, 1044.1)
fRFP24	FITC-PEG-βRINFRLQNS_5IDRS_5IVRIFETNPSILRVK	15.8*	1356.6, 1017.7 /(1356.3, 1017.4)
fRFP26	FITC-PEG-βRINFRLQNS_5IDRS_5IVRIN$_L$ETNPSILRVK	15.6*	1344.6, 1008.7 /(1344.9, 1008.8)
fRFP31	FITC-PEG-βRINFRLQNS_5IDRS_5AARIAETNPSILEVK	15.0*	1298.2, 973.9 /(1298.5, 974.3)
fRFP32	FITC-PEG-βRINFRLQNS_5IDRS_5AARIAETNPSILRVK	14.8*	1308.9, 981.9 /(1307.5, 980.7)
RFP4	W-βEQEEINFRLRQYMDKIILAILDHNPSILEIKH	13.9*	1393.3, 1045.2 /(1392.4, 1044.7)
RFP14	W-βRQVRELENYIDRLLVS_5VN$_L$ES_5TPNILRIPR	14.5*	1276.2, 957.4 /(1276.1, 957.3)
RFP24	W-βRINFRLQNS_5IDRS_5IVRIFETNPSILRVK	14.3*	1240.5, 930.6 /(1240.1, 930.4)
RFP26	W-βRINFRLQNS_5IDRS_5IVRIN$_L$ETNPSILRVK	14.0*	1229.2, 922.1 / (1229.0, 922.0)
RFP31	W-βRINFRLQNS_5IDRS_5AARIAETNPSILEVK	13.4*	1182.7, 887.3 / (1182.4, 887.0)
RFP32	W-βRINFRLQNS_5IDRS_5AARIAETNPSILRVK	13.2*	1192.3, 894.5 / (1191.8, 894.0)



FITC- -PEG- -beta-alanine-

Table 2.3: Name, sequence and characteristic LC-MS data for peptides used in this study.

Residues highlighted in yellow denote potential gain-of-function mutations, those in green denote loss-of-function alanine replacement for negative controls, and red denotes S_5 amino acid residues incorporated for peptide stapling. Chemical structures for linkers and fluorescent tags are shown. The first series of RFP peptides (RFP1-RFP13) were analyzed by LC-MS using a C18 reverse-phase column (Agilent, 2.1 x 150 mm, pore size 80 Å, particle size 3.5 µm); Buffer A (H_2O/0.1% TFA) and Buffer B (ACN/0.1% TFA); and a 15 minute method with the following gradient (flow rate 0.5 mL/min): 10-100% buffer B over 10 min, 100% buffer B for 2 min, 100- 10% buffer B over 1 min, and 10% buffer B over 2 min. Optimized RFP peptides (denoted by * on retention time) were analyzed by LC-MS using a C18 reverse-phase column (Phenomenex, 5.0 x 50 mm, pore size 110 Å, particle size 5µm); Buffer A (5/95/0.1% ACN/H_2O/TFA) and Buffer B (95:5:0.1% ACN/H_2O/TFA); and a 20 minute method with the following gradient (flow rate 0.5 mL/min): 0% buffer B over 3 min, 0- 65% buffer B over 15 min, 65-100% buffer B over 1 min; 100-0% buffer B over 1 min.

Gene name	ID	Forward Primer	Reverse Primer
AMIGO3	NM_198722.2	caagaacggcctctacttgc	aggccaagcatacgtactcg
CEBPG	NM_001252296.1	catggatcgaaacagtgacg	gcttccaaccgttcattctc
CXCR2	NM_001168298.1	cgctccgtcactgatgtcta	aaatccagccattcaccttg
FOXI1	NM_012188.4	cctctagcacagcctccttg	gaggctccatccaagatgtc
FPR1	NM_001193306.1	tattgccaccaagatccaca	gatatggggaccagcagaga
GPER1	NM_001039966.1	agggacaagctgaggctgta	gctgaacctcacatccgact
IGF1	NM_000618.4	tctcttctacctggcgctgt	cacgaactgaagagcatcca
LPAR4	NM_001278000.1	accttgtgccttgcaactct	ggctttgtggtcaaaggtgt
LY6K	NM_001160354.1	aaggaggtgcaaatggacag	cctgtgtgttgaatgctgct
MGAT1	NM_001114617.1	ccagctggacctgtcttacc	cggtcattggtcctcacttt
P2RY14	NM_001081455.1	cctccagatgaatcctgctc	agagctgggcacgtaaaaga
PRSS23	NM_001293178.1	caggtgggcatctacatcct	gtccttcccaaaaatgctga
RAB25	NM_020387.3	ccctcctggtgtttgaccta	tttgttacccacgagcatga
DNAJA3	NM_001135110.2	aggacaagccaagcagaaaa	tgcaccctgaacgtaatgaa
VHL	NM_000551.3	aggtcacctttggctcttca	ttggcaaaaataggctgtcc
ATCB	NM_001101.3	ctcttccagccttccttcct	agcactgtgttggcgtacag
18S		TaqMan	Assay ID: Hs99999901_s1

Table 2.4: Primer sequences used for qPCR analysis.

2.5 Methods

Cell lines and reagents.

HEY, MCF7, MDA-MB231 cells lines were originally obtained from ATCC. The MDA-MB231 and MCF7 lines were transduced by Precision Lenti ORF, RAB25 construct or control (Thermo Scientific, Open Biosystem) followed by Blasticidin selection (30 µg/ml) up to 7 days. Subsequently limited serial dilution was performed and single cells were allowed to expand to separate colonies for genetically homogenous population. Two independent clones were obtained for each constructs and target gene expression was evaluated with western blotting. Additionally, a sequence confirmed hemagglutinin epitope-tagged *RAB25* (Invitrogen, Carlsbad, CA) expression vector on pcDNA 3.1 (referred to as pcDNA3) was used to generate stable HEY isogenic cell line.

Fmoc-protected natural amino acid precursors, activating reagents (HCTU) and solid phase support for chemical synthesis of stapled peptides were obtained from Novabiochem (EMD). Non-natural S_5 ((*S*)-N-Fmoc-2-(4′-pentenyl) alanine) amino acid was obtained from Anaspec Inc. Solvents and other chemical reagents were obtained from Sigma-Aldrich and used as received unless otherwise noted. HEY, MDA-MB231, OVCAR3 and MCF7 cells were from ATCC and were not STR profiled. Cell lines have been tested for mycoplasma contamination. HEY and MDA-MB231 cell lines were maintained in RPMI media supplemented with 10% fetal calf serum, 1% penicillin/streptomycin and 2 mM L-glutamine. OVCAR3 and MCF7 cell lines were maintained in RPMI-1640 media supplemented with 10% fetal calf serum and 1% penicillin/streptomycin.

Stapled peptide synthesis and purification.

Unmodified and hydrocarbon stapled peptides were synthesized by Fmoc-based solid phase peptide synthesis and purified by reverse-phase HPLC with a C18 column (Agilent, Palo Alto, CA) as previously reported [138]. The first series of RFP peptides (RFP1-RFP13) were analyzed by LC-MS using a C18 reverse-phase column (Agilent, 2.1 × 150 mm, pore size 80 Å, particle size 3.5 μM); Buffer A (H_2O/0.1% TFA) and Buffer B (ACN/0.1% TFA); and a 15 min method with the following gradient (flow rate 0.5 mL/min): 10–100% buffer B over 10 min, 100% buffer B for 2 min, 100–10% buffer B over 1 min, and 10% buffer B over 2 min. Optimized RFP peptides (denoted by * on retention time in **Table 2.3**) were analyzed by LC-MS using a C18 reverse-phase column

(Phenomenex, 5.0 × 50 mm, pore size 110 Å, particle size 5 μm); Buffer A (5/95/0.1% ACN/H_2O/TFA) and Buffer B (95:5:0.1% ACN/H_2O/TFA); and a 20 min method with the following gradient (flow rate 0.5 mL/min): 0% buffer B over 3 min, 0–65% buffer B over 15 min, 65–100% buffer B over 1 min; 100–0% buffer B over 1 min. Purified peptides were lyophilized, quantified by A_{280}, dissolved in DMSO as 10 mM stocks and stored at –20 °C.

Recombinant protein constructs, expression, and purification.

Recombinant RAB and FIP proteins were expressed in BL21 cells using the following vectors: GST-RAB25 (1–180) in pET41 Ek/LIC; His_6-RAB25 (1–180) in pET28; GST-FIP3 (649–756) in pGEX 6P-1; GST-RAB11a (1–173, Q70L mutant) in pGEX5x-1. Transformed cells were incubated at 37 °C to an OD_{600} = 0.8, at which time induction was initiated with 1 mM IPTG for an additional 12 h at 20 °C. Cells were pelleted by centrifugation and lysed by sonication in lysis buffer (20 mM Tris pH 8.0, containing 150 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT, 2% glycerol and protease inhibitors (EDTA-free, Roche)). GST-tagged proteins were purified by binding to glutathione 4B sepharose resin (GE Lifesciences) at 4 °C, washed three times in lysis buffer, followed by elution with 20 mM reduced glutathione in lysis buffer. His-tagged RAB25 was expressed and lysed as above, prior to loading onto Ni-NTA resin (Qiagen), washed with lysis buffer containing 10 mM imidazole and eluted in lysis buffer containing 250 mM imidazole. All proteins were further purified by FPLC gel filtration (Superdex 75, GE Lifesciences) and dialyzed into appropriate assay buffers for downstream analysis.

Fluorescence polarization assays.

Fluorescence polarization assays were performed with for GST-RAB25 and GST-RAB11a to quantify binding to FITC-labeled unmodified and stapled RFP peptides. Initial fluorescence polarization measurements for RFP1–RFP13 were made by incubating FITC-RFP peptides (25 nM) with threefold dilutions of a given protein in FP buffer, consisting of 20 mM Tris pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT, 10 μM GTP and 2% glycerol. Dilutions and incubations were performed in 96-well, black flat-bottom plates (Nunc) to a total volume of 100 μL, and equilibrated at room temperature for 30 min. Polarization was measured on a Spetramax-M5 multi-label plate reader (Molecular Devices) with $\lambda_{ex} = 485$ nm and $\lambda_{em} = 525$ nm. Polarization values were determined using the equation: $P = (V - H)/(V + H)$, where P denotes polarization, V denotes vertical emission intensity and H represents horizontal emission intensity. Fraction bound was calculated using the equilibrium maximum polarization value obtained for a given protein among ligands, to which background-subtracted fluorescence polarization values were internally normalized. Fluorescence polarization measurements for FITC-RFP14, 24, 26, 31, and 32 were performed as detailed above with the exception that 10 nM FITC-ligand was used. Binding curves, apparent K_d values and 95% confidence intervals were determined generated using Prism 5 graphing software by fitting data to sigmoidal binding curve (4-parameter) according to the equation: $Y = \text{bottom} + (\text{top} - \text{bottom})/(1 + 10^{(\text{LogEC}_{50} - X) \times \text{HillSlope}})$ as detailed on the Prism 5 website.

ALPHAscreen assays.

ALPHAscreen assays were performed using alpha buffer, consisting of FP Buffer supplemented with 0.1% BSA (w/v) in white, flat-bottom 384-well ALPHAscreen plates (Perkin Elmer). For bimolecular association assays in **Figure 2.5 B**, biotinylated RFP peptide and tagged RAB25 protein was diluted to 2× the indicated final concentrations in 20 µL of buffer on the ALPHAscreen plates and incubated for 30 min at room temperature. Following incubation, 10 µL of 4× ALPHAscreen donor beads (10 µg/mL final concentration, Streptavidin-linked, Perkin Elmer) was added and incubated for 15 min at room temperature prior to final addition of 10 µL of 4× ALPHAscreen acceptor beads (Ni^{2+}-NTA-linked or anti-GST, 10 µg/mL final concentration, Perkin Elmer) and incubated for an additional 30 min. Proximity-based luminescence was measured on an ALPHAscreen-capable EnVision plate reader (Perkin Elmer) according to the manufacturers settings and protocols. Peptide competition assays were performed as above with the indicated concentrations of soluble peptide or proteins incubated with biotin-RFP14 and H_6-RAB25 (0.2 µM each in all wells) prior to addition of donor and acceptor beads. Binding curves, IC_{50} values and 95% confidence intervals were determined generated using Prism 5 graphing software by fitting data to sigmoidal inhibition curves (4-parameter) according to the equation: $Y = \text{bottom} + (\text{top} - \text{bottom})/(1 + 10^{(\text{LogIC}_{50} - X) \times \text{HillSlope}})$
as detailed on the Prism 5 website.

Circular dichroism spectroscopy.

Circular dichroism spectroscopy experiments were performed on a Jasco J-170 using a quartz cuvette (path length: 0.1 cm). Peptides were dissolved to 50 µM in deionized water and CD measurements were recorded at one nm increments between 190 and 260 nm, at room temperature. Thermal denaturation experiments were performed by recording CD absorbance at 222 nm at one-degree increments from 10 to 90 °C using a thermostat-controlled cuvette chamber. Raw denaturation data were fit to sigmoidal curves with Prizm 5 software, and the half-maximal temperature was reported as the peptide T_m.

Cellular pull-down assays.

HEY-pcDNA and HEY HA-RAB25 cells were transfected with pcDNA3-His B-FIP1 (kind gift from Dr. Jim Norman) [183] using Lipofectamine 2000 transfection reagent following manufacturers' protocol (Invitrogen). Forty-eight hours post transfection, the cells were collected using the HA-Lysis buffer provided in the Sigma HA-IP kit (IP0010-1KT Sigma-Aldrich, St. Louis, MO**)** and pretreated with either DMSO or indicated concentrations of the stapled RFP compounds for 1 hour at 4 °C. Subsequently, ~600 µg of total protein was loaded in each IP column subjected to immunoprecipitation using agarose beads coupled to anti-HA monoclonal antibody. Unbound proteins were removed by extensive washing and specifically associated proteins resolved by SDS-PAGE and western blotting for detection of RAB25-FIP1 interaction under various treatment conditions. Normal IgG was used as negative control.

Western Blots.

Western blotting was performed using primary antibodies for RAB25 (Cell Signaling 4314; 1:500), RAB11a (Cell Signaling 2413; 1:1000) and FIP1 (Gift from Dr Jim Norman, 1:1000) [183], GAPDH (Santa Cruz; 1:1000), and secondary antibodies, anti-rabbit or anti-mouse immunoglobulin G (IgG) horseradish peroxidase-linked secondary antibody (Cell Signaling Technology; 1:2000).

Confocal fluorescent microscopy for peptide cellular penetration.

MCF7 cells were grown in 8-well chamber slides (Ibidi) and treated with FITC-labeled RFPs (5 μM) in 10% FBS-containing RPMI for 8 h at 37 °C. Cells were washed 5× in PBS, fixed in 2% paraformaldehyde/PBS and stained with DAPI according to the manufacturer's protocol (Sigma). A coverslip was mounted onto the slide and confocal fluorescence microscopy performed with an Olympus DSU spinning disk confocal microscope (Olympus Corporation of the Americas, Center Valley, PA) with a Hamamatsu model C9100 EM-CCD camera (Hamamatsu Photonics, Skokie, IL) run by SlideBook v5.0 software (Intelligent Imaging Innovations, Denver, CO). Post-acquisition processing (multi-channel overlay) was performed using ImageJ software (NIH). Results are representative of images taken from three fields across the same well in at least two biological replicates.

Cell proliferation studies.

The day before treatment, control and stable RAB25 overexpressing HEY, MDA-MB231, and MCF7 cells (8×10^3 cells) were seeded in 96-well plates in culture medium containing 5% FBS overnight. The next day cells were washed twice with PBS and treated with DMSO or RFP peptides (15 µM) in 0.5% FBS containing medium for 48 h. Cell viability was determined using 8% PrestoBlue (Life Technologies, Federick MD), a resazurin-based solution that functions as a cell viability indicator, which was read (excitation wavelength 530 nm; emission wavelength 604 nm) using a TECAN microplate reader at 0 h and 48 h. For untreated cells, the reagent was added at 24, 48, and 72 h post plating (in 0.5% FBS containing RPMI medium) and absorbance was recorded at each timepoint.

Cell migration assays.

Cells were seeded at a density of 5000–7000 cells for HEY and MDA-MB231 or 15,000 cells for MCF7 on each side of an Ibidi culture insert (Ibidi, Munich, Germany), with a 500 µm separation between each side of the well, and allowed to grow for 24 h to attain a compact monolayer. Cells were pretreated with, or without, RFP peptides (10 µM) or DMSO control for 12 h before removal of the insert, and following removal of the insert cells were incubated in RFP-containing media or DMSO control for the duration of wound closure. Cells were photographed using a Life Technologies EVOS XL Core imaging system 4× or a Nikon DS camera connected to a Nikon Eclipse Ti microscope objective

at insert removal (0 h) and when wound closure was complete under control conditions. Experiments were performed in duplicate and a minimum of four fields of the injury were photographed and used for quantification.

Cellular thermal shift assay (CETSA).

Following sonication and centrifugation of MCF7 cells in Mammalian RAB lysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT, 0.1% TritonX-100, and protease inhibitors (EDTA-free, Roche)), bulk samples of lysate (diluted to 1 mg/mL) were incubated with 10 μM of the indicated peptide (or DMSO) for 10 min. Lysates were then aliquoted into 200 μL PCR tubes, heated for 3 min at a temperature point between 37 and 70 °C, incubated an additional 3 min at room temperature, centrifuged (15,000×*g* for 20 min), and supernatants were removed and combined with SDS-PAGE loading buffer for western blot analysis, and then analyzed by western blot (blotting for either RAB25, 11a, or GAPDH as negative control) and downstream ImageJ quantification. A sigmoidal curve was fit to the resulting data from each target protein for each of the three sets of treatment conditions, from which T_m values were calculated.

RNA isolation cDNA, library preparation and capture.

For the conditions listed, RAB25-HEY cells (10 cm plate, 70–80% confluent) were treated with RFP14 (10 µM) or the equivalent amount of DMSO in serum free media overnight, followed by stimulation with 5% FBS for 8 h. pcDNA3-HEY cells were treated with DMSO overnight and stimulated with 5% FBS for 8 h in parallel. Individual biological replicates

performed on different days were used for RNAseq studies. Samples were snap frozen in liquid nitrogen at the time of collection and total RNA was isolated from each condition by Norgen Total RNA Purification Kit (Norgen Biotek Corp, ON Canada), quantified by Picogreen (Invitrogen), and quality was assessed using a 2200 Tapestation (Agilent). RNA from each biological replicate (500 ng) was converted to double-stranded cDNA using Ovation RNA-Seq System V2 kit (Nugen), and cDNA was sheared by sonication with the following conditions: Peak Incident Power = 175, Duty Cycle = 20%, Intensity = 5, Cycles per burst = 200, for a time = 120 s using Covaris E220 sonicator (Covaris). To ensure the proper fragment size, samples were checked on a TapeStation using the DNA High Sensitivity kit (Agilent). The sheared DNA proceeded to library preparation using KAPA library prep hyper kit (KAPA) following the "with beads" manufacturer protocol. Briefly, this protocol consists of three enzymatic reactions for end repair, A-tailing and adaptor ligation, followed by barcode insertion by PCR using KAPA HiFi polymerase (6 cycles). PCR primers were removed by using 1.8× volume of Agencourt AMPure PCR Purification kit (Agencourt Bioscience Corporation). At the end of library preparation, samples were analyzed by TapeStation to verify correct fragment size and the absence of extra bands. Samples were quantified using KAPA qPCR quantification kit. Equimolar amounts of DNA were pooled for capture (2–6 samples per pool). We used whole exome biotin labeled probes from Roche Nimblegen (V3) and followed manufacture's protocol for the capture step. Briefly, DNA was pooled (2–6 samples), dried, resuspended with capture reagents and probes, and incubated at 47 °C on a thermocycler with a heated lid (57 °C) for 64–74 h. Targeted regions were recovered using streptavidin beads,

streptavidin-biotin-probe-target complexes were washed, and another round of PCR amplification was performed according to manufacturer's protocol. The quality of each captured sample was analyzed by TapeStation using the DNA High Sensitivity kit, and the enrichment was accessed by qPCR using specific primers designed by Roche Nimblegen. The minimum cutoff for the enrichment was 50-fold.

RNA sequencing and data analysis.

The captured libraries were sequenced on a HiSeq 2000 (Illumina Inc., San Diego, CA, USA) on a version 3 TruSeq paired-end flowcell at a cluster density between 700 and 1000 K clusters/mm^2 according to manufacturer's instructions. Sequencing was performed on a HiSeq 2000 for 2×100 paired-end reads with a 7 nt read for indexes using Cycle Sequencing v3 reagents (Illumina). The resulting BCL files containing the sequence data were converted into ".fastq.gz" files and individual libraries within the samples were demultiplexed using CASAVA 1.8.2 with no mismatches. All regions were covered by >20 reads. Data analysis was performed using a comprehensive in-house RNA-Seq pipeline (IPCT, Bioinformatics and Computational Biology, M.D. Anderson). STAR was used to align paired-end reads to the hg19 version of the reference genome, featureCount was used to obtain expression counts of genes and exons, and Cufflinks was used to estimate gene expression (FPKM). Genetic variants were called using GATK unified genotyper. Fusions were detected using STAR and filtered using Oncofuse. Quality of raw and aligned reads was assessed using FastQC and Qualimap.

FPKM values of six technical samples (pcDNA3-HEY + DMSO, RAB25-HEY + DMSO, RAB25-HEY + RFP14, each from individual biological replicates) were normalized according to total FPKM values of 34,560 genes. Genes with zero expression in one or more samples compared to FPKM values of the rest of the 22,912 genes were eliminated. To define the set of transcripts that were consistently and differentially expressed as a result of RAB25 expression, a mean fold-change of > 2.14 was applied (to control the number of genes between 100 and 500, which is optimal for GSEA analysis) between pcDNA-HEY and RAB25-HEY RNAseq replicates. This resulted in gene sets of 104 transcripts that were upregulated by RAB25 (named RAB25_UP) and 269 transcripts that were downregulated by RAB25 (named RAB25_DOWN). Gene set enrichment analysis (GSEA) was performed using the GSEA portal (www.broad.mit.edu/GSEA/) in the following way. Each gene set was queried for enrichment within the expression profile generated by comparing DMSO and RFP14 treated RAB25-HEY cells, thereby asking whether RAB25-regulated genes are specifically regulated by RFP14 in an unbiased manner. GSEA was performed with the following parameters: probe set collapse = false; phenotype = RFP14 vs. DMSO; permutation: sample, permutations = 1000. Gene set size: $15 < n < 500$. Heat maps showing the per-gene effect of RFP14 on the RAB25-regulated genes (each in the RAB25_UP and RAB25_DOWN gene sets) were generated using Cluster 3.0, and visualized using Tree View software. Samples are plotted in columns and the genes in rows.

qRT-PCR and statistical analysis.

HEY ovarian cancer cells with stable ectopic expression of RAB25 were plated and allowed to reach 70% confluency before treating with DMSO, RFP14, or RFP32 (10 µM or equivalent DMSO) in serum free media for 12 h, followed by stimulation with 5% FBS for 8 h. Total cellular RNA was then isolated from cells using Thermo RNA Isolation kit (Thermo Fisher Scientific, MA, Cat #AM1560), according to the manufacturer's instructions. Two micrograms of RNA per sample was reverse transcribed to generate complementary DNA, using reverse transcriptase Thermo RT kit (Cat #4368814). qRT-PCR was performed using a 7500 Fast Real-time PCR System (Applied Biosystems Inc., CA) with SYBR Green Power master mix (Thermo Cat #4309155). The primers used for the ACTB and 18 S (internal control) and other listed genes were designed using the 'Primer 3 Output' software from cDNA sequences found in the NCBI Gene Database (Nucleotide). Their specificity was confirmed using a BLAST analysis against the genomic NCBI database. qRT-PCR reaction conditions were: 50 °C for 2 min; 95 °C for 2 min; 40 × (95 °C for 15 s; 60 °C for 15 s; 72 °C for 1 min). The relative quantification of gene expression between the treatments and control cells was calculated by the $2^{-\Delta\Delta Ct}$ approximation method. Primers used for qPCR analysis are listed below.

CHAPTER 3: Diels-Alder Cyclized Peptides*

*Portions of this chapter were adapted with permission from:

Montgomery, et al. J Am Chem Soc, 2019. 141(41): 16374-81.

3.1 Introduction

Peptides present a powerful starting point for the development of bioactive ligands. Protein–protein interactions (PPIs) tend to be mediated by short, linear peptide sequences that adopt diverse secondary structures, including helix, sheet, loop, and turn motifs. Deregulation of PPIs underpins many diseases, either through mutations, aberrant protein expression, or other contextual changes [151, 184, 185]. Removal from their native protein context typically abrogates peptide structural rigidity, leading to reductions in binding affinity and selectivity, as well as metabolic stability. Therefore, numerous chemical strategies have been developed to synthetically rigidify peptides, potentially stabilizing secondary structures and augmenting pharmacologic utility, notably by imbuing the cyclized peptide with increased target-binding affinity, cell-permeability, and bioavailability [82, 83, 85, 86, 186]. Most of these peptide stabilizing chemistries capitalize on alkylation or acylation reactions on the naturally occurring nucleophilic residues cysteine and lysine; head-to-tail loop closure; or incorporation of unnatural residues for metal-catalyzed cross-linking. These approaches, and others, have been applied to stabilize many different sequences and structures [96, 103, 107, 131, 187, 188]. Despite

incredible progress in synthetic methods for peptide stabilization, however, current strategies have various limitations. These can include harsh or forcing conditions; the requirement of on-resin and/or fully-protected peptide macrocyclization; limited suitability across target sequences or folds; incompatibility with aqueous conditions or other stabilization chemistries; use of large excess or highly-tailored cross-linking reagents. Therefore, new chemistries to control peptide secondary and tertiary structure, ideally with wide functional group and reaction condition tolerance, would greatly expand the synthetic arsenal for developing peptide-based chemical probes.

A notable carbon–carbon bond forming reaction that has enjoyed widespread synthetic application is the Diels–Alder [4 + 2] cycloaddition (**Figure 3.1 A**). First reported in 1928, this reaction is employed by nature and chemists alike to construct complex molecules [189-191]. A pericyclic reaction, it occurs between a diene and dienophile, driven by orbital overlap (**Figure 3.1 B**). Generations of synthetic chemists have experimented with tuning Diels-Alder reactant electronic properties and reaction conditions to explore increasingly diverse applications of this chemistry. Several aspects of this transformation make it especially attractive for synthesis, including a wide range of diene-dienophile pairs, high regio- and chemoselectivity in many applications, simultaneous introduction of multiple stereocenters, and compatibility with a range of reaction conditions. Indeed, the power and versatility of this transformation is apparent in the wide variety of natural products and other pharmacologically relevant molecules that rely on Diels-Alder chemistry for the construction of their central carbon frameworks and precise installation of stereocenters (**Figure 3.1 C**). Further, pioneering studies

demonstrated that aqueous conditions are not only tolerated, but can also enhance reaction rates by driving reactant proximity [192].

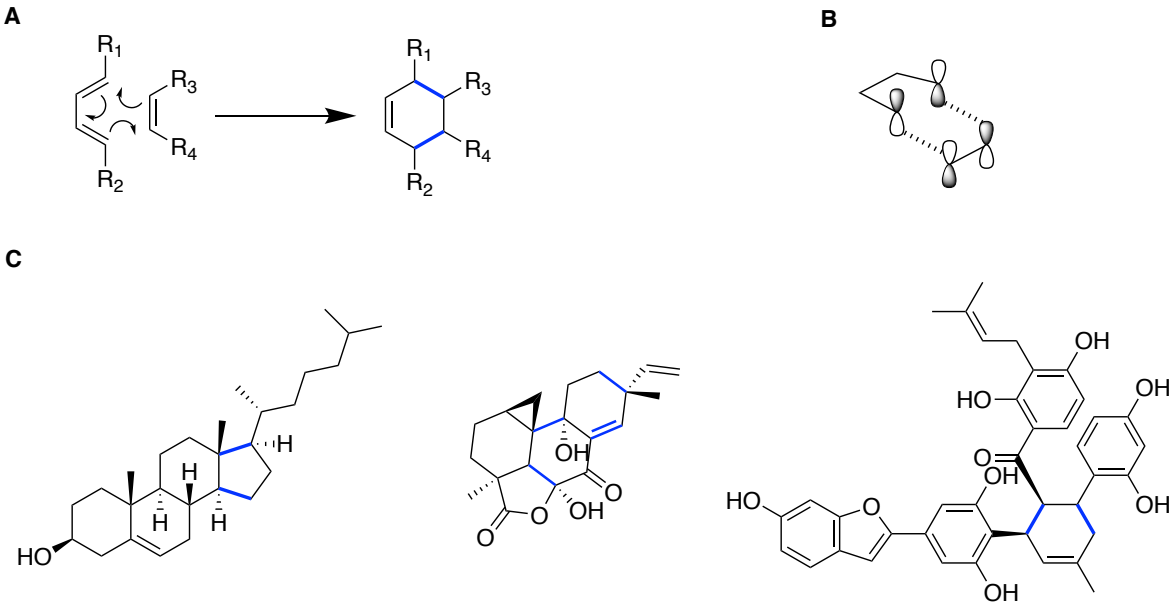


Figure 3.1: Overview of Diels-Alder chemistry.
(A) Concerted, pericyclic Diels-Alder reaction mechanism, with newly formed carbon-carbon bonds in blue. **(B)** Depiction of molecular orbital overlap seen in the forward demand reaction. **(C)** Representative synthetic and biosynthetic natural products, with resulting bonds formed by Diels-Alder reaction in blue [193-195].

While many of the above qualities overlap with those desired for selective modification of biomolecules, Diels-Alder has only seen limited use for the intermolecular ligation of biomolecules; the chemistry has not been applied to intramolecular, site-specific stabilization of peptides or proteins [196-199]. Here we report the application of Diels-Alder carbon-carbon bond forming reactions for the macrocyclization of diverse turn, loop and alpha-helical peptide motifs.

3.2 Results

We initiated our exploration of Diels-Alder cycloadditions for peptide macrocyclization by designing and testing a panel of turn and loop-type peptides to

investigate the synthetic parameters for efficient cyclization, followed by compound characterization and preliminary screening of properties. Key learnings were then applied toward the design of Diels-Alder stabilized alpha helical peptides for subsequent biophysical and biochemical study to elucidate effects of resulting cycloadducts on peptide structure and function. Subsequent studies have worked toward expanding the scope of peptide scaffolds and compatible orthogonal chemistries.

3.2.1 Initial Diels-Alder cyclization of turn and loop peptides

We began our study of Diels–Alder cyclized (DAC) peptides with turn and loop motifs, including an antimigratory RGD peptide, analogs of which have undergone clinical testing in oncology [200-202]. Given that there is wide variance in structure and reactivity of diene and dienophile functional groups, and presumably we chose a moderately activated (diene 2,4-hexadiene) and the well-characterized maleimido dienophile [203]. A selection of peptides of various length and sequence were synthesized containing orthogonally protected cysteine (tBuS-) and lysine (Mmt-) side chains [204], with the former invariably in the *i*-position, and the latter in the i+4, i+5, and i+7 positions (**Table 3.1**).

On-resin reductive cysteine deprotection and hexadiene alkylation with 1-bromo-2,4-hexadiene proceeded quantitatively as determined by LC-MS (**Figure 3.2 A-B**; **Scheme 3.1 A**). The corresponding diene-containing amino acid, Fmoc-Cys(2,4-hexadiene)-OH, was also synthesized and found to be compatible with direct incorporation by SPPS (**Figure 3.2 A**; **Scheme 3.1 B**; **Section 3.5**). On-resin acidic lysine deprotection and acylation with N-maleimidoglycine of an RGD model peptide resulted in

formation of two peaks as observed by LC-MS analysis (**1** and **1a**) with the same mass and ~1 minute retention time shift, suggesting formation of a cyclic, earlier eluting product on the C18 column (**Figure 3.2 A,B**). This product appeared rapidly, with an approximate 2:1 ratio of **1a** to **1** within 5 minutes of dienophile introduction. Extended on-resin incubation resulted in near-complete conversion to **1a** (16 hours, **Figure 3.2 B**). Kinetic reaction monitoring confirmed rapid conversion of later-eluting **1** to **1a**, as well as minor product **1b**, with ~90% formation within 1 hour (**Figure 3.2 C**). This pattern was observed across various sequences and diene-dienophile spacings tested (**Table 3.1**), with high yields of ~85–95% conversion to earlier-eluting, presumably cyclized species. In each scaffold studied, additional earlier-eluting minor isomeric species (≤10% relative to major product) were observed eluting ~4 minutes prior to linear precursor (**Figures 3.2 B** and **3.3**).

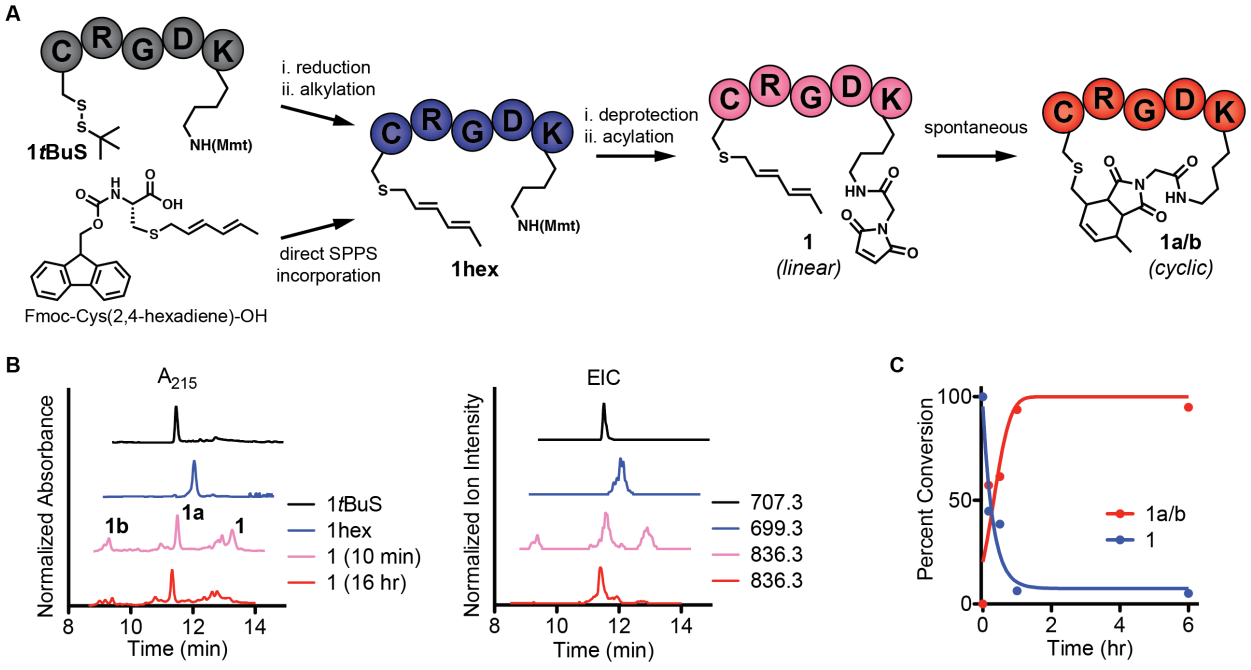
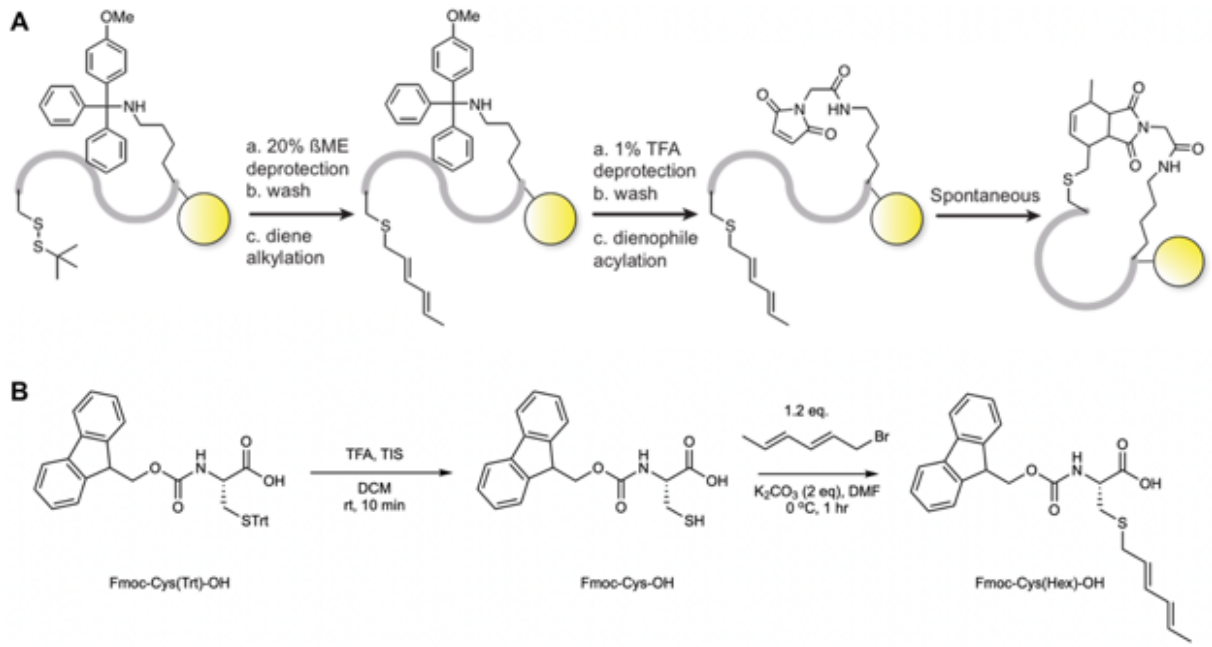


Figure 3.2: Synthesis and reaction monitoring of Diels–Alder cyclized RGD peptide.

Figure 3.2, continued: (A) Schematic of DAC RGD peptide functionalization. **(B)** Absorbance (215 nm) and extracted ion chromatogram (EIC) LC–MS traces of crude peptide following the indicated functionalization and reaction times of the RGD compound 1 scaffold (black, protected Cys precursor peptide; blue, hexadiene-functionalized peptide; pink, 10 minutes post-maleimide functionalization; red, 16 hours post-maleimide functionalization. **(C)** Representative reaction time course quantifying conversion of 1 (blue) to cyclized 1a/1b (red) isomers.



Scheme 3.1: Synthetic schemes for (A) on-resin Diels-Alder functionalization and cyclization and (B) synthesis of Fmoc-cysteine(2,4-hexadiene)-OH.

A limited screen of reaction conditions revealed that mild heating on resin in DMSO was an optimal condition to maximize cyclic conversion and also increased the ratio of major to minor isomeric products observed (*vide infra*, **Table 3.2**). This is likely due to disaggregation of the protected peptide on resin [205, 206], which may allow for preferential formation of kinetic cycloadduct product(s).

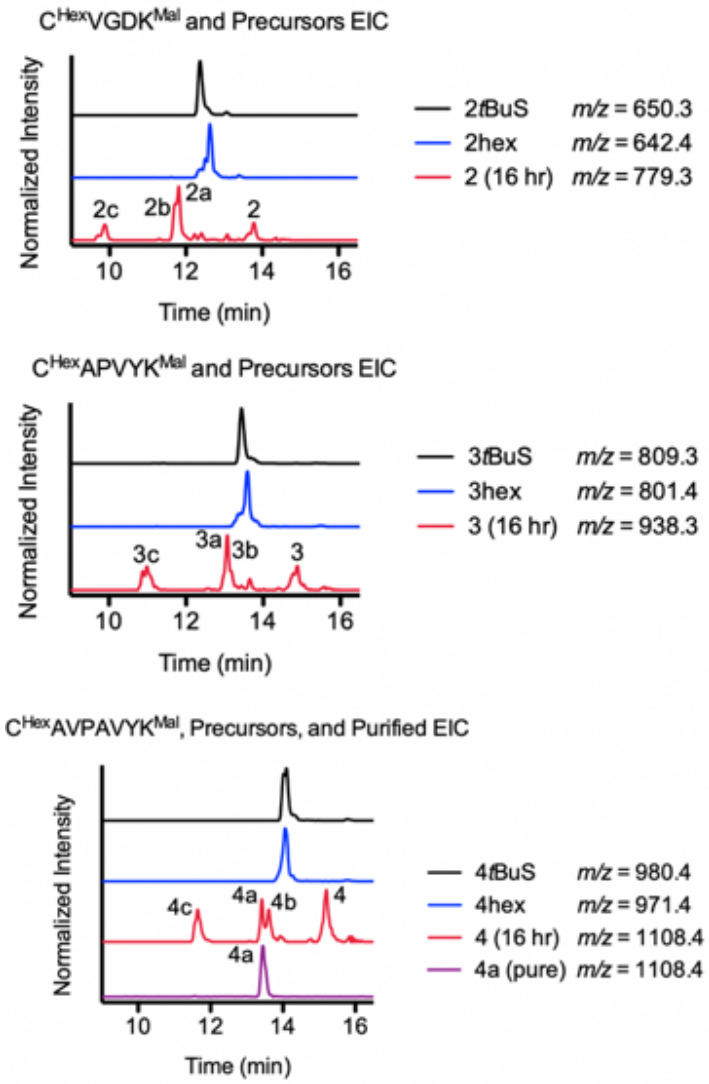


Figure 3.3: EIC LC-MS traces of Diels-Alder cyclized peptides and precursors. Traces of crude protected cysteine precursor peptides (black), hexadiene-alkylated intermediates (blue), maleimido-functionalized peptides (red) following acylation and purified *endo* stereoisomer **4a** (purple). Each compound number represents the non-cyclized (**X**) and cyclized isomers (**Xa**, **Xb**…) with lettering reported in order of relative intensity of cyclic products as determined by separation on LC-MS.

We also tested other less reactive dienophiles, including acryloyl and crotonyl groups, and found they react with hexadiene in yields comparable to maleimide (**Figure 3.4**). Notably, these viable alternative dienophiles are compatible with direct incorporation during SPPS [207]. They also offer increased diversity of cyclic structures, and, due to

their asymmetric nature, should preferentially form at most two, rather than four, distinct isomeric cycloadduct products.

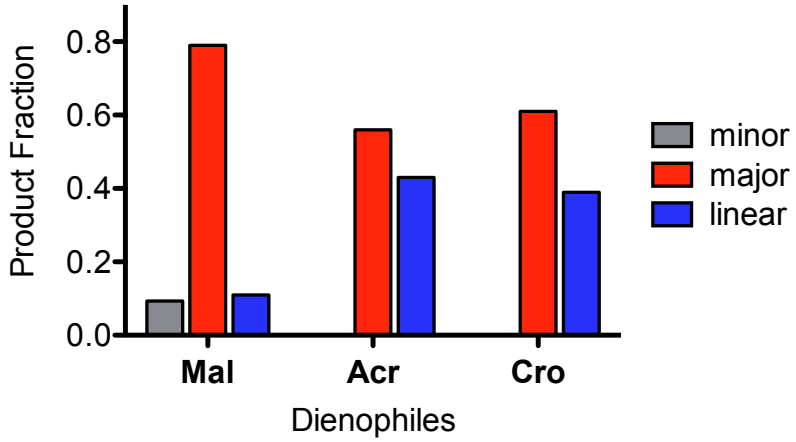


Figure 3.4: Comparison of alternate dienophile reactivities for Diels-Alder cyclization of peptides.
Quantification of crude product profiles following on-resin lysine acylation of **3hex** with either *N*-maleimidoglycine (Mal), acrylic acid (Acr) or crotonic acid (Cro) and overnight incubation.

Similar reaction profiles were observed for peptides produced by on-resin diene incorporation or direct incorporation during SPPS using of Fmoc-cysteine(2,4-hexadiene)-OH, as well as upon incubation in aqueous solution following peptide cleavage (**Figure 3.5**). Taken together, these data confirm the compatibility of Diels–Alder chemistry for peptide cyclization, as well as with diverse organic and aqueous chemical environments.

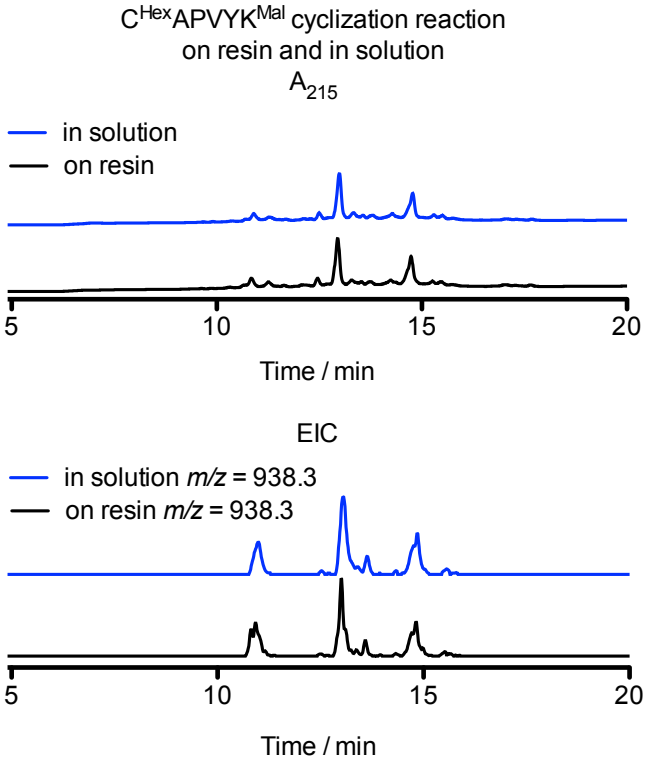
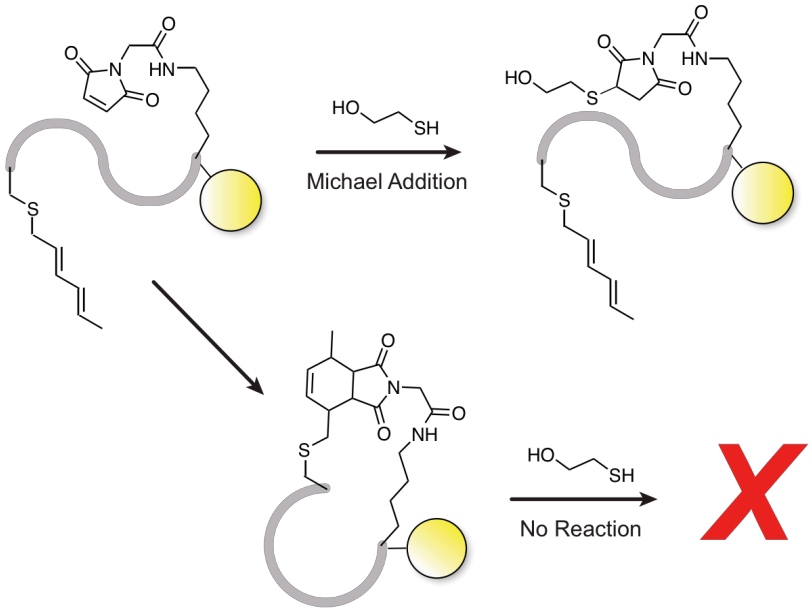


Figure 3.5: Comparison of Diels-Alder peptide cyclization in aqueous solution or protected on-resin in organic solvent.
215 nm absorbance and EIC LC-MS traces of Diels-Alder cyclization of compound **3** following 16 hr incubation on-resin in DMF (black) or in aqueous solution (blue) after cleavage (10% ACN/H_2O, 0.1% TFA).

3.2.2 Diels–Alder cycloadditions form chemically stable peptide macrocycles with enhanced protease resistance and biological activity

To confirm that **1** and **1a** were non-cyclized and Diels-Alder cyclized products, respectively, we capitalized on the presence of the maleimide dienophile in the non-cyclized precursor peptide. On-resin incubation of the **1** and **1a** crude mixture with excess 2-mercaptoethanol (βME), which should only react via Michael addition with the linear maleimide-containing peptide, resulted in complete consumption of **1**, yielding a species with the mass of the βME conjugate addition product (**Scheme 3.2**; **Figure 3.6**).

Conversely, **1a** levels remained constant, consistent with the loss of the α,β-unsaturated electrophile after Diels–Alder cyclization. Similarly, prolonged exposure of purified **1a** to βME in aqueous solution did not result in the observation of an addition product, suggesting formation of a stable DAC-RGD peptide macrocycle (**Figure 3.7 A**).



Scheme 3.2: Schematic of on-resin trapping of linear diene and dieneophile functionalized peptide via Michael adduct formation with 2-mercaptoethanol (βME).

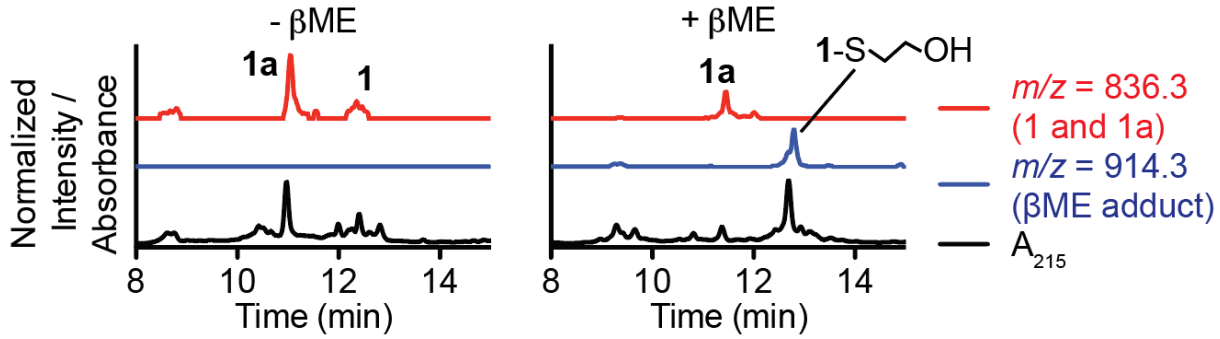


Figure 3.6: LC-MS analysis of on-resin trapping of linear diene and dienophile functionalized peptide.
EIC and 215 nm absorbance LC-MS traces of crude **1** and **1a/1b** (red, *m*/z = 914.3). with or without βME treatment, showing conversion of **1** to the corresponding βME-trapped species (blue, *m*/z = 914.3).

The RGD motif has been the subject of considerable investigation and therapeutic development as an antimigratory agent targeting cell-surface integrins in aggressive cancers [208]. As seen with other peptides, *in vivo* use of linear RGD analogs shows limited efficacy, primarily due to cleavage by circulating proteases; head-to-tail cyclized analogs demonstrate increased metabolic stability and *in vivo* efficacy [209]. To test if DAC peptides similarly display differential protease susceptibility, we exposed DAC peptide **1a** and its linear, wildtype analog **1wt** to *in vitro* trypsinolysis screening. It was observed that 1wt rapidly degraded within 2 hours, whereas the cyclic peptide **1a** showed an extended half-life, with 30% remaining intact after 4 hours (**Figure 3.7 B, top**). To determine if this was a more general effect applicable to other sequences and proteases, we exposed the major DAC isomer of *i, i+7* spaced peptide **4a** and its linear analog **4wt** to chymotrypsin. As seen with the RGD scaffold, the linear species **4wt** was degraded more rapidly than DAC **4a** (**Figure 3.7 B, bottom**). Correspondingly, the more metabolically stable cyclized RGD analog **1a** significantly inhibited wound closure, an integrin-mediated phenotype, relative to linear control peptide **1wt** (**Figure 3.7 C, D**). Taken together, these results confirm that Diels-Adler cyclization can affect the structure and biological function of loop peptide ligands.

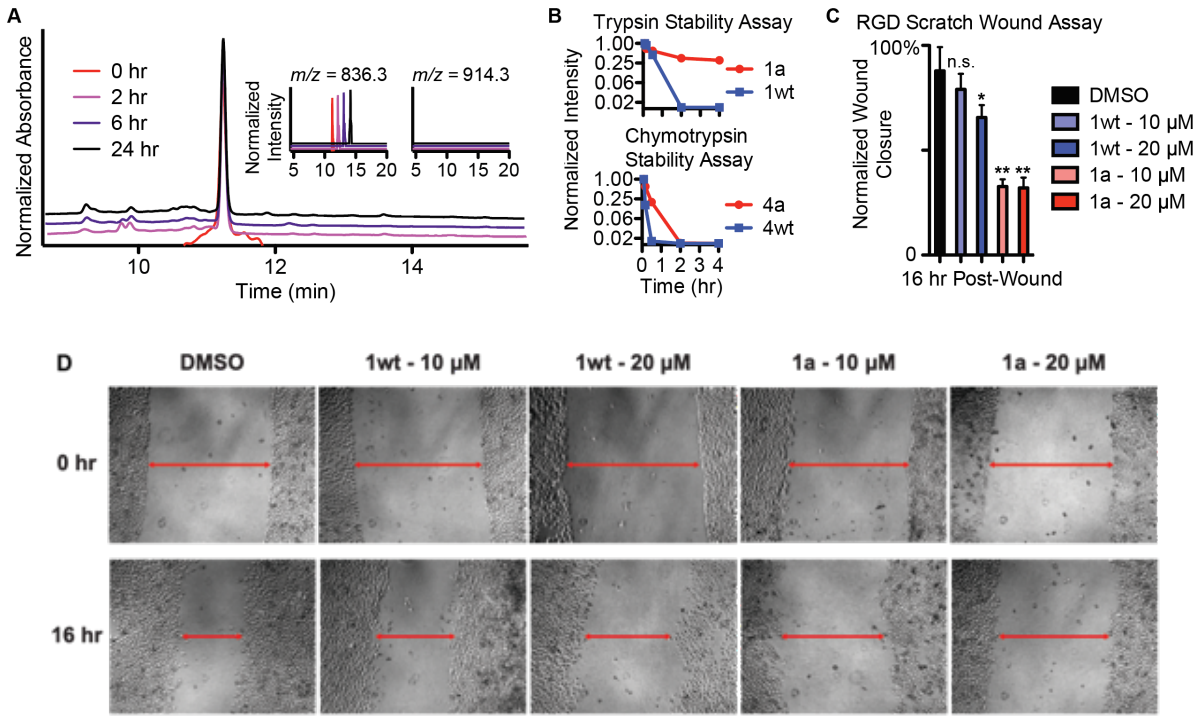


Figure 3.7: Chemical and biochemical stability and biological efficacy of Diels-Alder cyclized RGD peptide.
(A) Extended time course tracking of HPLC-purified 1a incubated with βME in aqueous buffer, verifying cycloadduct stability (βME-trapped species m/z = 914.3; not observed). **(B)** In vitro trypsin and chymotrypsin resistance assays for cyclic 1a versus linear analog 1wt, and cyclic 4a and linear analog 4wt, respectively. **(C)** Cell migration model scratch wound assay comparing reduction in wound closure by RGD peptides 1wt and 1a at 10 and 20 μM relative to DMSO control. Data represent mean ± s.d. distance measurements from images of 2 biological replicates, 5 distance measurements each. n.s. = not significant; * = $P < 0.05$; ** = $P < 0.0005$; Student's t test. **(D)** Representative images from 0 and 16 hr time points of scratch wound assay with HeLa cells. Treatment conditions included DMSO control, wildtype and DAC RGD peptides 1wt and 1a. Red line represents a single measure of wound distance. 5 measurements were taken per view across two biological replicates of each condition. Migration distance was calculated by subtracting 16 hr distance from 0 hr distance, averaging, and normalizing to DMSO migration distance.

3.2.3 Diels–Alder cycloadducts preferentially form with *endo* stereochemistry and stabilize peptide conformation

To characterize the putative Diels–Alder cycloadduct and resulting effects on peptide conformation, DAC **4a** and precursor peptides **4wt** and **4hex** were subjected to

a series of NMR experiments. ^{1}H NMR of **4a** confirmed the presence of two vinylic protons, absent in 4wt, and distinct from those of the diene in **4hex** (**Figure 3.8 A**) or α,β-unsaturated maleimido protons. The observed doublet of triplets pattern of the vinylic protons in **4a** results from small ^{3}J and ^{4}J coupling constants of the unsaturated system, indicative of near-90° dihedral angles between neighboring vinyl and allyl protons. TOCSY experiments permitted complete assignment of cyclized and non-cyclized peptides (**Section 3.5**). Direct through-bond coupling was observed between **4a** vinylic protons and all other protons within the putative cycloadduct (**Figure 3.8 B** and **Section 3.5**), supporting the formation of a fused, bicyclic adduct. The combination of TOCSY and NOESY experiments further resolved the connectivity of the cycloadduct in **4a**, revealing extensive through-space interactions between its protons (**Figures 3.8 B** and **Section 3.5**). Most notably, interactions between H3/H6 allylic protons and H4/H5 succinimidyl bridgehead protons suggest the cycloadduct of **4a** is the endo stereoisomer (**Figure 3.7 B** and **Section 3.5**).

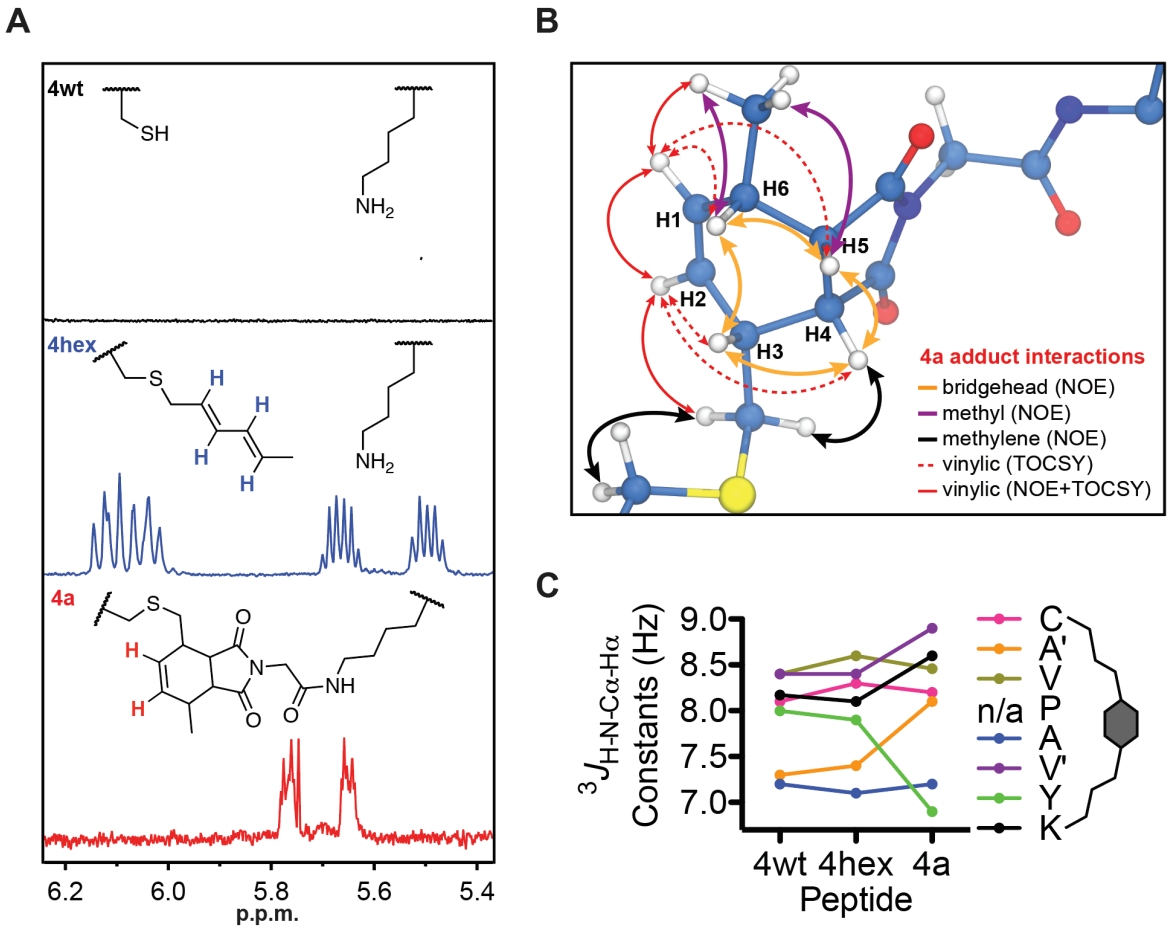


Figure 3.8: NMR spectroscopy characterization DAC peptide 4a and linear precursors.
(A) ^{1}H NMR spectra of **4wt** (black), **4hex** (blue), and **4a** (red) in the vinylic (~6–5 ppm) region. **(B)** Model of **4a** cycloadduct depicting through-bond and through-space interactions between cycloadduct protons as determined by TOCSY and NOESY (orange: bridgehead interactions; purple: methyl-group interactions; black: Cys-side chain interactions; red: vinylic interactions). **(C)** $^{3}J_{\text{H-N-Cα-Hα}}$ coupling constant values for each residue of the three peptides studied.

NMR results further revealed substantial perturbation to the cyclic **4a** peptide structure relative to linear analogs. Following hexadiene incorporation, only modest changes in **4hex** chemical shifts were observed for the cysteine and its nearby residues, with virtually no change in ^{3}J-coupling constants between backbone amide and alpha protons (**Figure 3.8 C** and **Section 3.5**). Conversely, **4a** displays marked changes in

chemical shifts and backbone ^{3}J-coupling constants for several residues, most notably the tyrosine and alanine adjacent to the cyclizing residues (**Figure 3.8 C** and **Section 3.5**). These alterations likely result from both cyclization-induced limits on degrees of freedom, as well as emergent intramolecular interactions such as hydrogen-bonding between amide and carbonyl oxygen atoms found on the cycloadduct and peptide.

To further investigate the structure of this *i, i+7* DAC peptide, we carried out molecular dynamics (MD) studies of **4wt** and **4a**. Briefly, models were built in MOE and run in the NAMD2 simulation package (see **3.6 Methods** for additional details) [210, 211]. Trajectory analysis using VMD revealed linear **4wt** displayed no persistent structure, while cyclized **4a** displayed a highly stabilized extended loop structure over 50 ns simulations (**Figure 3.9 A**) [212]. RMSD measurements of the respective backbones confirm a substantial decrease in **4a** conformational flexibility relative to linear **4wt** (**Figure 3.9 A, B**; average backbone RMSD = 1.07 and 2.69 Å, respectively), which also translated to reduced motion in the side chain dynamics of **4a** (**Figure 3.9 B**).

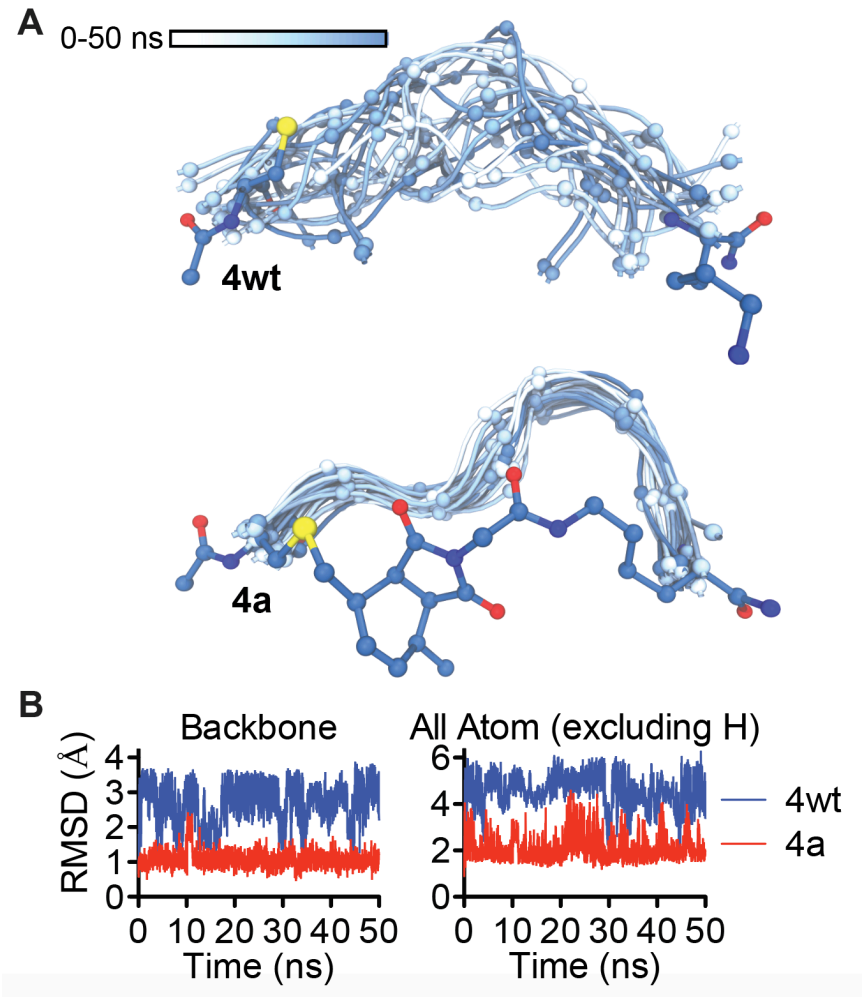


Figure 3.9: Molecular dynamics simulations of Diels Alder cyclized peptide 4a and linear analog 4wt.
(A) Peptide backbone alignment over 50 ns simulations for 4wt and 4a. **(B)** RMSD values for 4wt and 4a peptide backbones and all atoms (excluding hydrogens).

3.2.4 Diels–Alder peptide cycloadditions operate in aqueous solution and stabilize alpha helices

Helical peptides represent the dominant structural feature that mediates biomolecular interactions [11, 213]. Therefore, we sought to apply Diels–Alder cyclization strategies to stabilize these privileged peptide conformations from diverse peptide sequences. We first tested whether a conformationally restricted and less reactive furan

diene could cyclize on a p53-derived helical sequence [132]. Like Fmoc-Cys(2,4-hexadiene)-OH, we found that Fmoc-protected (2-furanyl)alanine (denoted as A_{Fur}) could be directly incorporated during SPPS (**Figure 3.10 A,B**). Notably, Diels–Alder cyclization between A_{Fur} and K_{Mal} at previously established *i, i+7* positions in the p53-derived sequence proceeded slowly on solid support. This initial effort resulted in a dominant linear species that could be isolated by HPLC (**Figure 3.10 A,B**). Intriguingly, incubation of the purified linear peptide **5** in phosphate-buffered saline, however, resulted in the appearance of earlier-eluting cyclized species; a minor amount of peptide dimers was also observed (**Figure 3.10 B, Table 3.3**). As seen with the DAC-RGD scaffold, Michael-addition trapping experiments confirmed that the putatively cyclized compounds **5a**, **5b**, and **5c** did not contain a reactive maleimide, contrasting with the isolated linear peptide **5** (**Figure 3.10 C**). Circular dichroism (CD) spectroscopy of the linear control p53 sequence and the three isolable DAC species confirmed increased α-helical character for the dominant DAC peptide **5a** relative to precursor peptide and **5b** and **5c** minor products (**Figure 3.10 D**). These data are reminiscent of other studies that suggest stereochemical yield is correlated with global stability and/or pre-organization of the peptide fold [214, 215]. The observations made here with the p53 scaffold suggest the aqueous chemical environment may aid in the selection of favorable adducts that promote global helix stabilization.

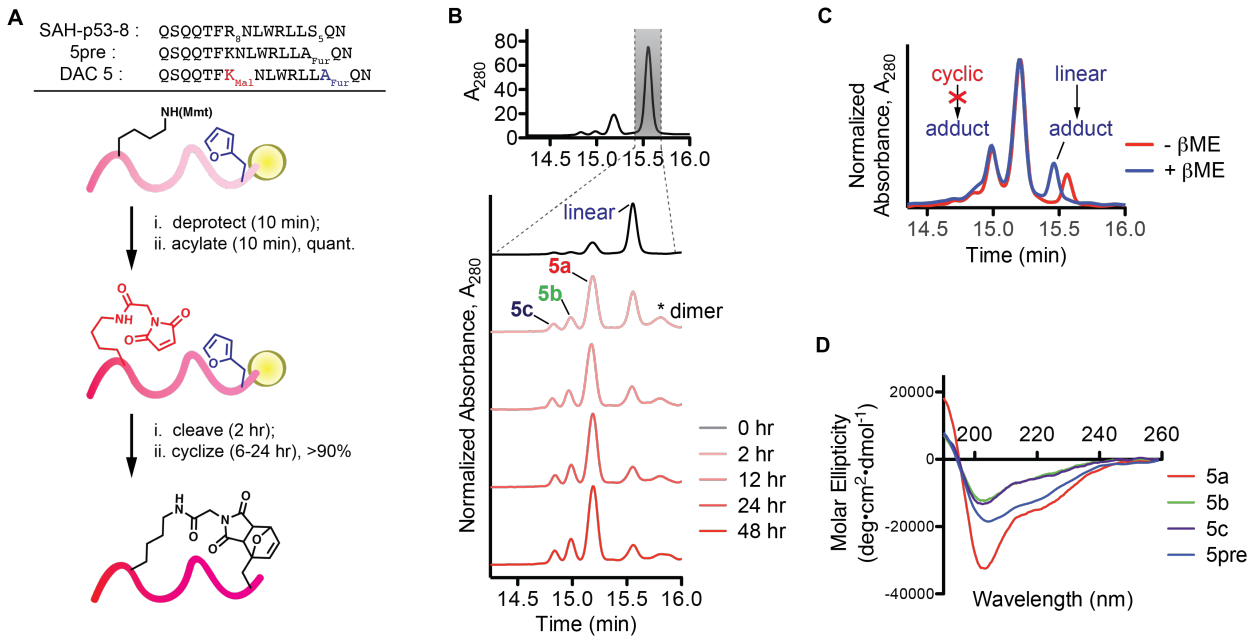


Figure 3.10: Synthesis, stability, and secondary structural characterization of Diels–Alder cyclized p53 peptide.
(A) Schematic of DAC-p53 peptide synthesis. **(B)** HPLC trace and resulting aqueous solution reaction time course of conversion of 5 to cyclized isomers 5a–c. **(C)** Incubation of DAC-p53 peptide 5 mixture with βME in aqueous buffer, verifying Michael adduction of linear precursor 5 and stability of cyclized isomers. **(D)** CD spectra of linear analog 5wt and 5a–c.

3.2.5 Diels–Alder cycloadduct geometry, alone or in tandem with other chemistries, differentially affects peptide helicity and activity

To further explore the potential to stabilize diverse helical conformations, we synthesized a series of Diels–Alder stabilized, alpha helical peptides derived from the steroid receptor coactivator-2 protein (SRC2). A short helical stretch within this protein contains a conserved LXXLL motif that engages a hydrophobic groove in estrogen receptor-alpha (ER alpha; **Figure 3.11 A**) [142, 216]. We synthesized several *i, i+4* "DAC-

stapled" peptides, which were tolerant of direct 2-furanylalanine incorporation during SPPS. Unlike the p53-derived sequence, both SRC2 peptides **6** and **7** cyclized on resin rapidly and in high yield following maleimide incorporation (**Figure 3.11 B**, **Table 3.3**). We characterized the two isolable cyclized species **6a** and **6b**, and the one dominant cyclic compound **7a** by CD spectroscopy. The dominant species **6a** and **7a** were significantly more helical than either the unmodified wildtype SRC2 peptide or minor **6b** compound (**Figure 3.11 C**).

Previous attempts to stabilize the SRC2 LXXLL motif have shown that hydrophobic residues that engage ER alpha can be replaced by a hydrocarbon staple without significant loss in binding affinity [217]. Therefore, we investigated the potential of synthesizing a double-stapled SRC2 ligand, containing both a hydrocarbon staple installed by ring-closing metathesis of olefin containing amino acids in concert with Diels–Alder cyclization of a diene–dienophile pair (**Figure 3.11 A,D**). Direct incorporation of olefin-containing S_5 amino acids and furan yielded a clean peptide precursor, **8pre**, that was readily cyclized by Grubbs-I catalyst (**8rcm**; **Figure 3.11 D,E**). Subsequent introduction of the maleimide dienophile resulted in immediate appearance of an earlier-eluting tandem stapled peptide, which was the stoichiometric product after mild heating on-resin for 4 hours (**Figure 3.11 E**, **Table 3.3**). CD characterization of this bicyclic peptide **8a** revealed classic alpha-helical minima at 208 and 222 nm (**Figure 3.11 F**). ^{1}H NMR analysis confirmed the presence of Diels–Alder cycloadduct vinylic protons and a single allylic proton in all major DAC SRC2 peptide products, and, in the case of **8a**, peaks corresponding to the RCM alkene product were observed (**Section 3.5**).

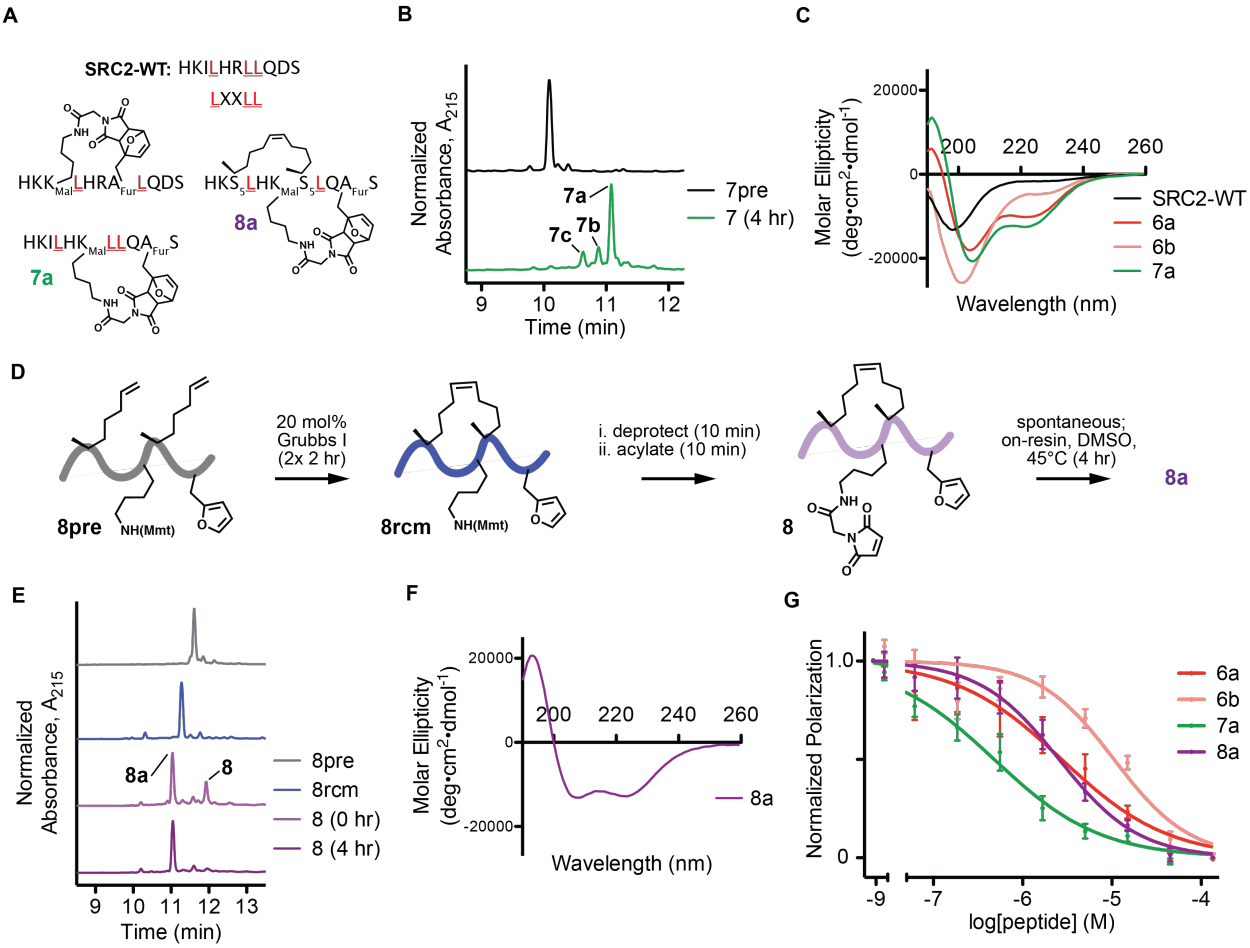


Figure 3.11. Synthesis, secondary structural characterization, and activity of Diels–Alder cyclized and double-stapled SRC2 peptides.
(A) Sequence of SRC2 wildtype peptide used as scaffold with conserved LXXLL motif; staple placement and structures of DAC-SRC2 peptides 6a,b and 7a and double-stapled 8a. **(B)** HPLC traces of precursor peptide used and resulting cyclized product profile of compound 7 following 4 h on-resin heating in DMSO. **(C)** CD spectra of isolated DAC-SRC2 peptides compared to unmodified wildtype sequence. **(D)** Synthetic scheme for layering ring-closing metathesis and Diels–Alder cyclization for the synthesis of double-stapled SRC2 peptide 8a. **(E)** HPLC traces of precursor peptide 8pre, ring-closing metathesis product 8rcm, Diels–Alder functionalized peptide 8 and fully-DAC-SRC2 double-stapled peptide 8a. **(F)** CD spectra of helical compound 8a. **(G)** Normalized FP assay results showing a dose range of DAC-SRC2 peptides effectively competing off fluorescein-labeled wildtype SRC peptide from wildtype ER alpha. Each data point represents measurements from two independent experiments run in triplicate, ± s.e.m. Dose–response curves fit with a four-parameter log(inhibitor) vs response equation in Prism 5 graphing software.

To probe the structure activity relationships of these DAC-SRC2 peptides, we performed competitive fluorescence polarization binding assays with ER alpha. Both the single DAC-stapled **7a** and tandem stapled **8a** peptides competed a fluorescein-linked SRC2 peptide off of ER alpha with IC_{50} values of approximately 0.5 and 2.5 μM respectively (**Figure 3.11 G, Table 3.4**). Intriguingly, the dominant, more helical isomer **6a** demonstrated 4-fold improved activity compared to the less helical, minor product **6b**. These results echo those observed for the p53 sequence, where stereochemical yield is correlated with overall peptide fold stabilization and, in the case of SRC2, improved biochemical potency.

3.2.6 X-ray crystal structure of a DAC stapled SRC2 peptide–ERα complex demonstrates cycloadduct stereochemistry and contribution to target binding

To determine the structure of the Diels–Alder adduct in **6a** and its potential role in promoting the interaction with ER alpha, we solved an X-ray structure of **6a** and estradiol co-crystallized with residues 300–500 of the ER alpha ligand binding domain (LBD) at 2.25 Å resolution (**Figure 3.12 A, Table 3.5**). The Y537S mutant of ERα was used, as it stabilizes the agonist conformation of the receptor, aiding in crystal formation. An $F_o–F_c$ difference map, using wild-type SRC2-bound ER alpha, confirmed the unambiguous presence of estradiol bound in the core of the LBD, and DAC **6a** peptide bound in the canonical activating function 2 (AF2) cleft of ER alpha (**Figure 3.12 A** and **Figure 3.13 A,B**). The Diels–Alder adduct was highly ordered, permitting unequivocal confirmation that the more stabilizing and dominant stereochemical product in the context of SRC2 is that of the *endo* isomer (**Figure 3.12 B**). The *endo* bicycle presents a convex hydrophobic

surface that directly contacts ER alpha residues Val355, Ile358, and Leu539 that form the hydrophobic shelf adjacent to the cleft that binds the LXXLL motif (**Figure 3.12 C**). Residues in **6a** derived from SRC2 Leu690 and Leu694 retained their canonical contacts and were deeply engaged with a network of hydrophobic residues in the AF2 groove (**Figure 3.12 C**, bottom). Additional orienting contacts are mediated by His691 to the edge opposite the Diels–Alder adduct, as well as a conserved hydrogen bond between Glu542 of ER alpha that caps the N-terminus of the peptide helix (**Figure 3.13 C**). At the C-terminus, the last ordered residue present in the DAC–peptide structure, Gln695, folds back on the peptide to satisfy a hydrogen bond to the *i - 2* amide carbonyl, effectively capping the helix (**Figure 3.13 C**). Taken together, these structural data confirm that the unique Diels–Alder chemical and stereochemical composition can stabilize an active helical conformation and directly contribute to target engagement, as evidenced here by forming a molecular “clasp” around the core and edge of the ER alpha AF2 cleft.

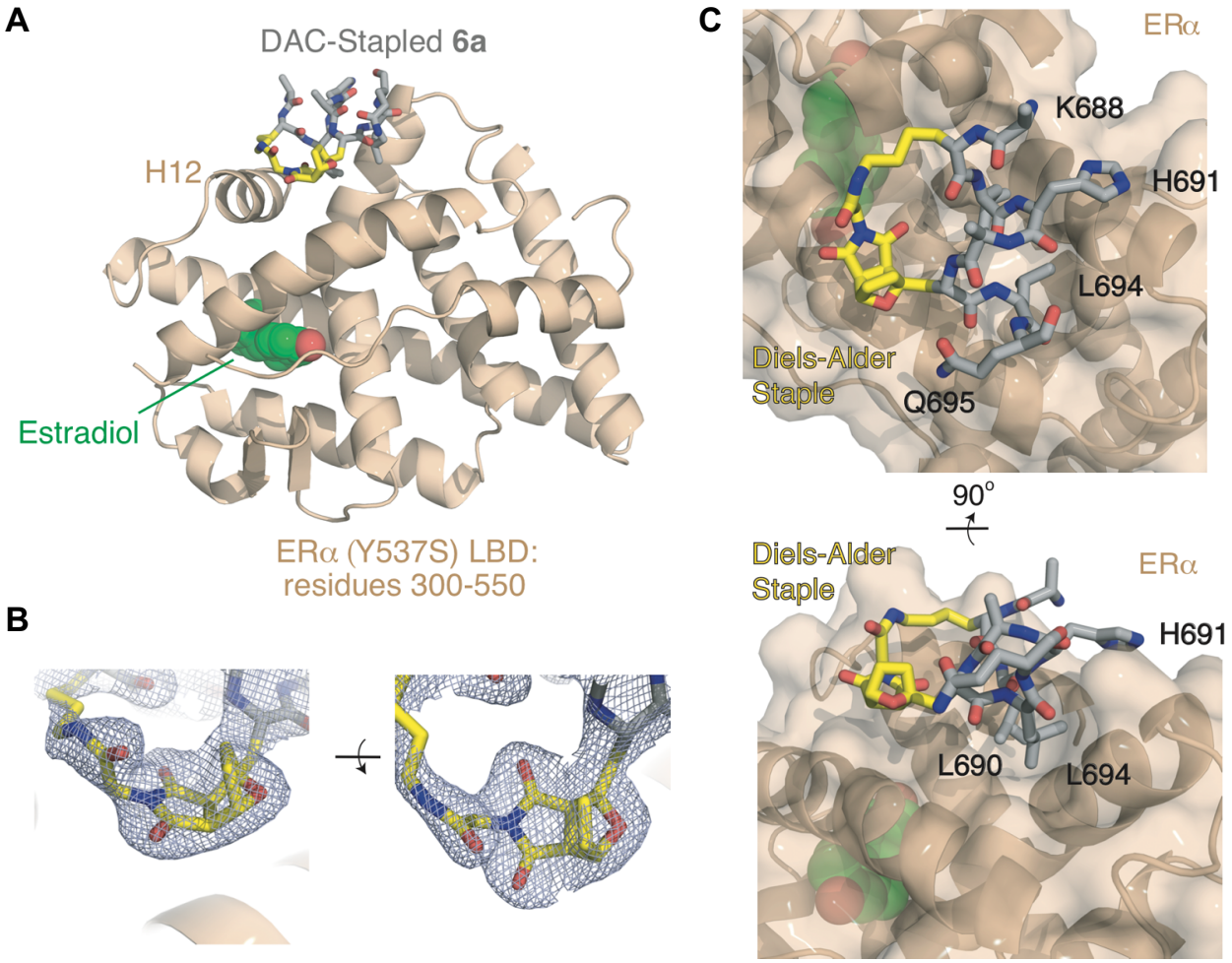


Figure 3.12: X-ray crystal structure of Diels–Alder cyclized SRC2 peptide 6a bound to ER alpha.
(A) LBD of ER alpha Y537S mutant protein bound with estradiol and 6a located in the AF2 cleft (PDB 6PIT). **(B)** Two views of the structure of the DAC-SRC2 peptide 6a cycloadduct (C = yellow, N = blue, O = red) with *endo* stereochemistry overlaid with electron density map (gray grid) from resolved crystallographic data (2mFo-DFc map contoured to 1σ). **(C)** Dorsal (top) and axial (bottom) view of 6a bound in the AF2 binding pocket. Relevant residues numbered according to SRC2 sequence are highlighted, as well as the Diels–Alder cyclization moiety in yellow. Flexible portions of the structure with low electron density are omitted.

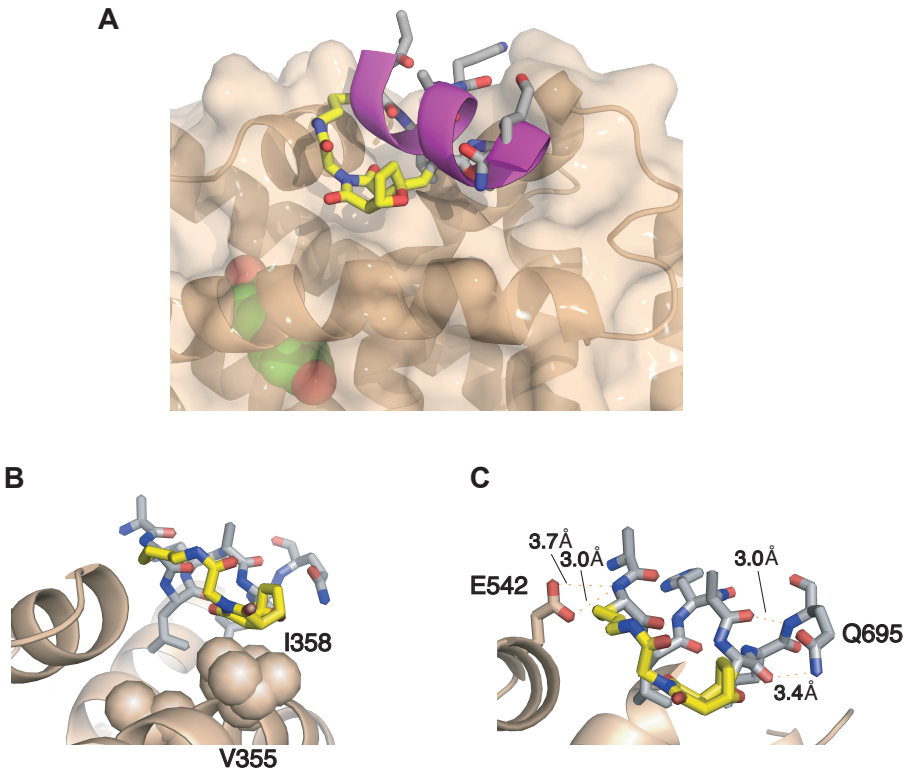


Figure 3.13: Detail of X-ray crystal structure results of Diels-Alder cyclized SRC2 peptide 6a bound to ER alpha.
(A) Overlay of x-ray crystal structure results from this study show strong similarity to those of Transcriptional Intermediary Factor 2 peptide (purple) bound to estrogen receptor alpha (PDB 1GWR) [218]. **(B)** Interactions between Diels-Alder cycloadduct and hydrophobic shelf adjacent to AF2 cleft. **(C)** Hydrogen bonds formed between protein and peptide and intramolecular hydrogen bonding caps C-terminal peptide residues *via* Gln695.

3.2.7 Toward the development of minimal DAC SRC2 ligands, more reactive Diels-Alder cross-linking groups, and extensions of DAC peptide and protein design

From the crystallographic data above, we observed that the external residues of the DAC SRC2 peptide **6a** (*i.e.* those residues more distal to the central stabilizing cross-link) were more flexible, resulting in much lower electron densities in those regions (**Figure 3.14 A,B**, green residues in **6a**). This suggested these residues do not appreciably participate in ER alpha binding, motivating us to explore minimal DAC SRC2 structures. As a prototypical example to explore the effects of different Diels-Alder

cycloadducts on peptide structure and activity, we synthesized **miniSRC-1**, cross-linked with the hexadiene and maleimido reactive pair (**Figure 3.14 B**). In this peptide, the exogeneous residues are removed and the positions of the cross-linking residues are N-terminally shifted, to retain the full SRC LXXLL motif. Synthesis was carried out as previously described, with the terminal Diels-Alder cyclizing step screened both in aqueous solution and in DMSO on-resin with mild heating. In agreement with previous results, the on-resin reaction proceeded more rapidly, resulting in the enrichment of the major DAC product, eluting at ~7.6 minutes as observed by LC-MS (**Figure 3.14 C**).

While this work is still ongoing, there are several noteworthy features of DAC peptide design it is intended to elucidate. First, by honing down an alpha helical peptide structure to its helical bioactive core, it should better allow interrogation of the influence of cycloadducts on structure and activity. These can include peptide physicochemical properties and secondary structure as informed by LC-MS retention time and CD spectroscopy; binding affinity as informed by the FP assay or other biochemical screens; cell penetration and efficacy in cell culture; and more granular interrogation of peptide binding through X-ray crystallographic studies. Toward this end, we will explore different reactive pairs, different pair orientations, and potentially different cycloadduct stereochemistries. Further, given the increased hydrophobic character of **miniSRC-1** and derivatives, these peptides may have more favorable cell uptake properties (potentially allowing for passive diffusion into cells), as has been observed for advanced stapled peptide analogs currently undergoing clinical testing [134]. Differential effects on cell

uptake due to the cycloadduct present are another intriguing area we intend to investigate with this series.

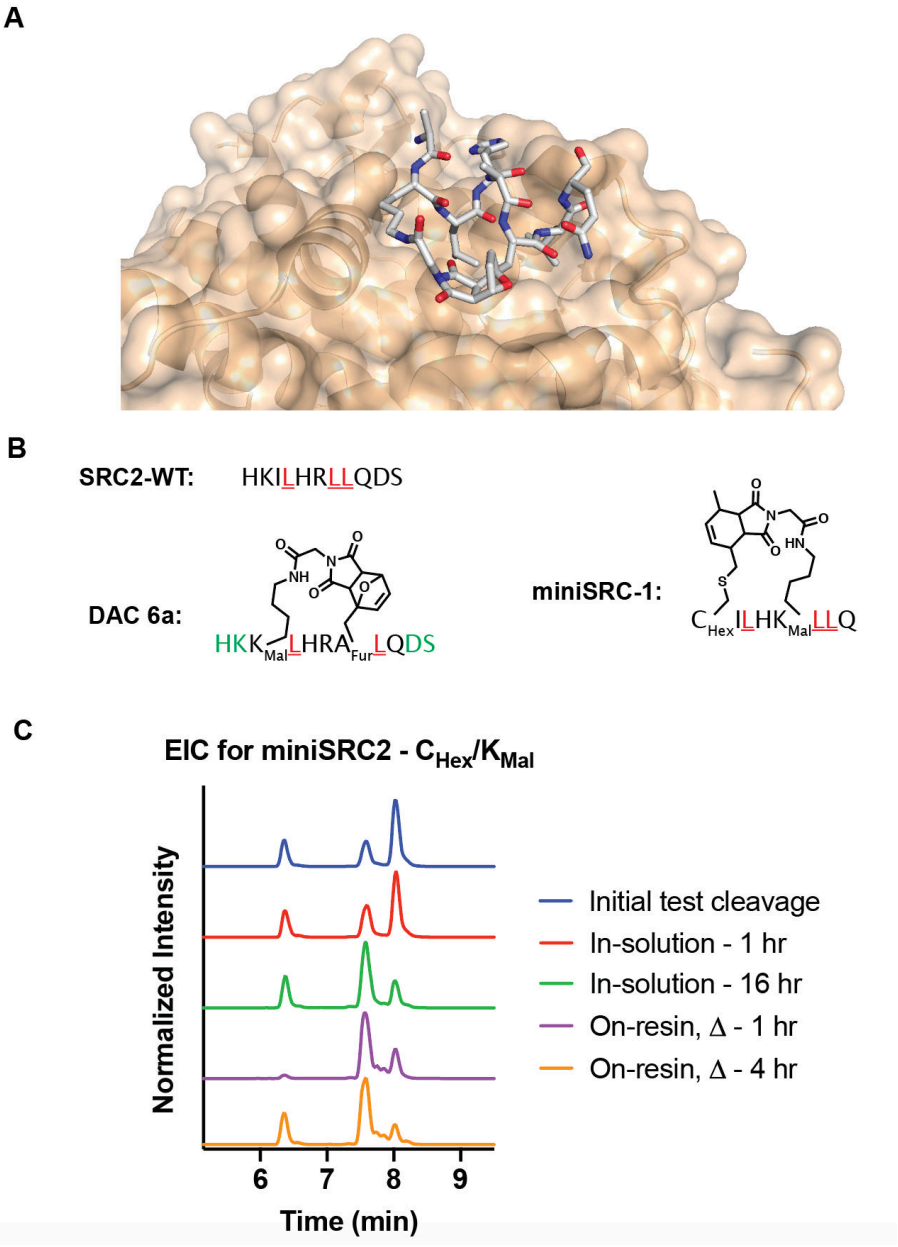


Figure 3.14: Minimal DAC SRC2 peptide ligand design and synthesis.
(A) Detail of the X-ray crystal structure, showing the central, stabilized **6a** peptide residues bound to ERalpha (PDB 6PIT). **(B)** Sequences of progenitor peptide structures and the prototype miniSRC-1 (red, underlined: Leu residues of the LXXLL motif; green: flexible, low electron-density regions of the peptide). (C) Extracted ion chromatograph data

Beyond the Diels-Alder reactive pairs discussed previously, and the sidechain-to-sidechain type cyclization design for the miniSRC series, we have begun exploring alternatives. A noteworthy highly reactive diene is cyclopentadiene, with its heightened reactive derived from its diene being fixed in the *cis* orientation. While cyclopentadiene derivatives have been explored for their utility for bioconjugation chemistry [199], we sought to use its rapid reaction kinetics with maleimides and other dienophiles to drive formation of DAC peptides. We synthesized a cyclopentadiene-containing linker ((cyclopentadienyl)propanoic acid (**cpdp-COOH**)) as well as an NHS-activated analog for on-resin functionalization (**Figure 3.15 A**). We began testing N-terminal and lysine acylation with cyclopentadiene and maleimide linkers on several model peptides to form N-terminal stabilization motifs. This is analogous to similar applications of these motifs for N-terminal templating of alpha helical peptides, as reported for a variety of hydrogen bond surrogate stabilization strategies [219]. Initial results for a model RGD peptide (**Figure 3.15 B**), with final lysine acylation by **cpdp-NHS**, indicate a high degrees of cyclization, which is suggested by no appearance of a Michael adduct species following on-resin βME-treatment for trapping of free maleimide (**Figure 3.15 C,D**). The resulting DAC peptides were produced in an approximately 2:1 isomeric ratio of two closely eluting species, suggesting they are both *endo* adducts (inferred from our previous experience with Diels-Alder cycloadduct formation) of different regioisomers of the substituted cyclopentadiene, given the cpdp-linker undergoes 1,5 hydride [199, 220]. Unfortunately, extensive side product formation was observed across a variety of reaction conditions tested (*e.g.*, peptide bond formation with **cpdp-COOH**; acylation with activated **cpdp-**

NHS, various solvent screens and reaction times). While these side products are as of yet uncharacterized, they likely correspond to cyclopentadiene group decomposition during peptide cleavage; subsequent side products resulting from the deprotection of side chain protecting groups by cyclopentadiene; and/or Diels-Alder mediated peptide oligomers.

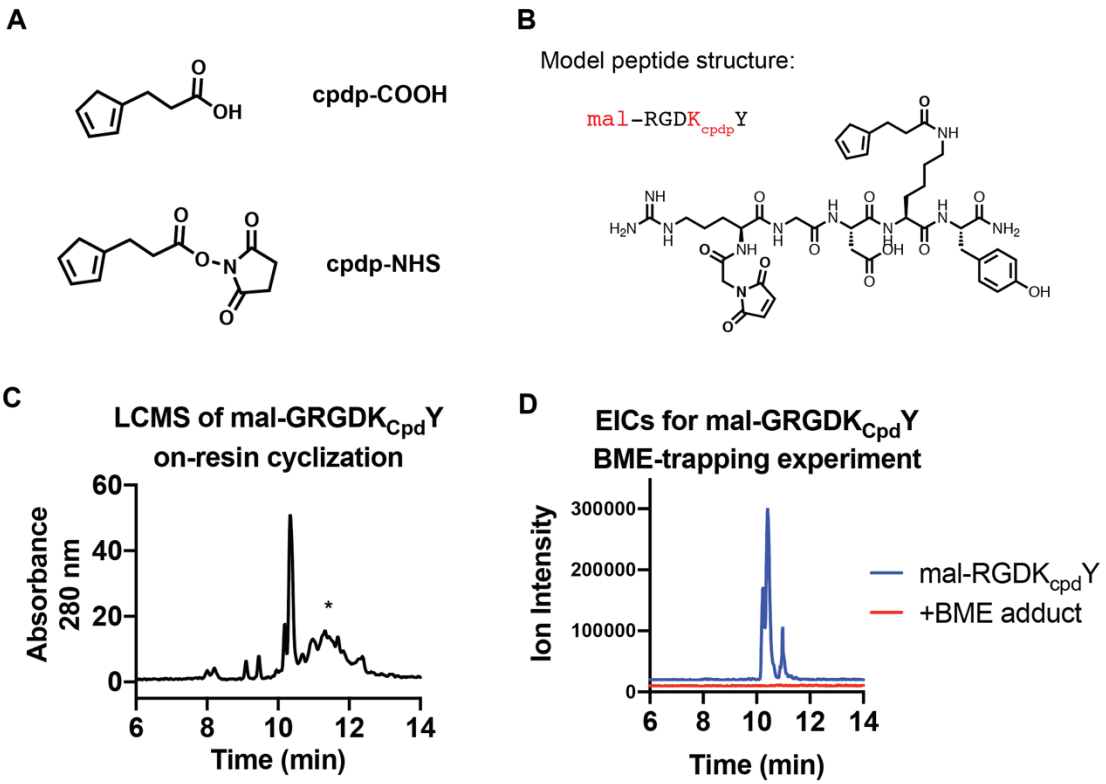


Figure 3.15: Structures of cyclopentadiene linkers and synthesized model peptide. **(A)** Structures of cyclopentadiene linkers. **(B)** Model RGD peptide used for initial cpdp-corporation. **(C)** Representative LC-MS trace following N-terminal functionalization with maleimide and lysine acylation with **cpdp-NHS**. *, side products. **(D)** Extracted ion chromatograms corresponding to fully-functionalized RGD peptide (blue) and Michael adduct product (red; not observed).

To attempt to reduce side product formation while still working toward the study of the conformational effects of cycloadduct structure and position, we turned to a simple

poly-alanine model system previously shown to template helices at reduced temperature [221]. The same cyclization strategy was used as above, with N-terminal maleimide functionalization and lysine acylation with **cpdp-NHS**. The position of the lysine group was varied, allowing for positional and spatial scanning of resulting formed adducts for effects on helicity (**Figure 3.16 A**). This small panel of Diels-Alder stabilized poly-alanine helix (daPAH) peptides was synthesized and heated for 4 hours on-resin in DMSO. While side product formation was still observed, the resulting LC-MS profiles indicated a variety of isomeric cyclic species were formed for each derivative (**Figure 3.16 B**). Perhaps not surprisingly, **daPAH1**, in which the stabilizing groups are only separated by a single residue, had a near equivalent yield for each cyclic species. This suggests reduced selectivity of adducts formed due to the high proximity of the two highly-reactive Diels-Alder reagents. Attempts were made to isolate at least one **daPAH1** product, however the material was of insufficient quantity and purity to make downstream characterization possible. **daPAH2** and **daPAH3** displayed similar cyclic product profiles, with two major later-eluting products dominating. The latest-elution (*i.e.* fourth) peak of each was isolable by HPLC. Intriguingly, **daPAH4** resulted in predominantly one major cyclic isomer, eluting earlier relative to the major products of **daPAH2** and **3**. This cyclic isomer was isolated and characterized by CD spectroscopy with the others. While each daPAH series member displayed marked alpha helicity, denoted by absorbance minima at ~208 and 222 nm, the minima at 222 nm was lowest for daPAH4 (**Figure 3.16 C**). This is especially noteworthy given both the difference in cyclic product profiles and reactive group separation for **daPAH4** relative to the other 2 peptides studied. This again echoes the results for DAC

5 and **6** above, in which the more helical DAC product corresponded to the higher yielding product, however further study is required to ascertain the helicity of the other cyclic species not analyzed here. In addition, we are curious if the **daPAH4** peptide N-terminal helix-templating motif's spacing and orientation would prove to be generalizable to other helical scaffolds.

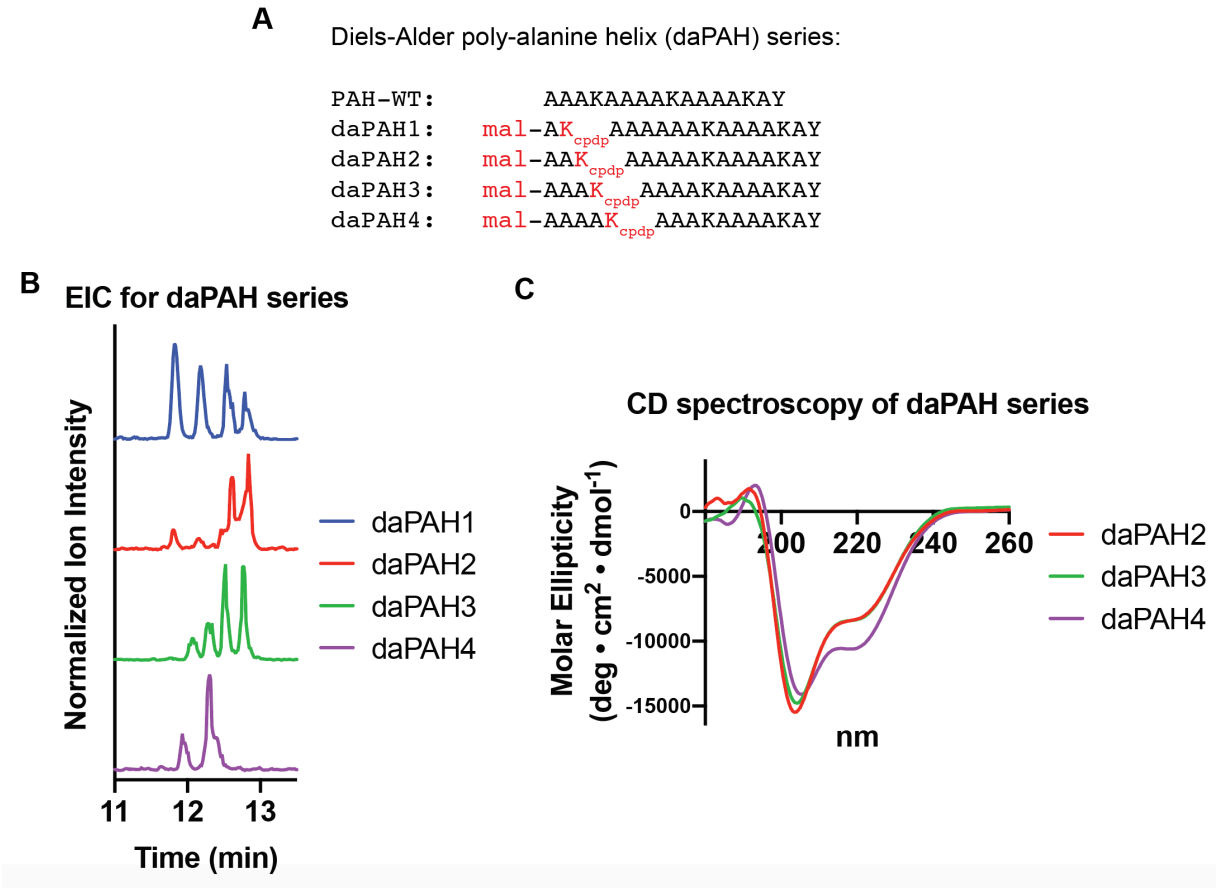


Figure 3.16: Diels-Alder poly-alanine helix peptides, design, cyclic profile, and structural characterization.
(A) Sequences of daPAH peptides synthesized. **(B)** Extracted ion chromatograms of each member of the daPAH peptide series, depicting different cyclic isomeric profiles. **(C)** CD spectroscopy results for **daPAH2**, **3**, and **4**, with the predominant **daPAH4** cyclic product displaying the most alpha helical character.

While these results represent just the beginning of explorations of the effects of Diels-Alder cycloadduct structure, positioning, and stereoisomers on peptide conformation and activity, they warrant further enabling technology development. First, an optimal cyclopentadiene linker or incorporation conditions must be determined. One potential possibility is the synthesis of a protected form of cyclopentadiene for either direct incorporation during SPPS or post-SPPS functionalization. Furan-protected maleimide derivatives have been developed for similar applications [222], suggesting the combination of these strategies for the direct incorporation of orthogonally-protected highly reactive Diels-Alder pairs for rapid synthesis and formation of DAC peptides. Additional compounds in the above series should also be synthesized to better elucidate the effects of various cycloadduct features on resulting peptide structure and function. Key learnings and apparently favorable cycloadduct structures should then be applied to diverse peptide scaffold to explore the generality of the observed effects. A longer term goal, stemming from the **miniSRC** work, is the development of a robust, helical DAC peptide scaffold for expansion of peptide libraries, to both search for enhanced binders of ER alpha as well as alternative targets.

In contrast to the above designs toward DAC peptide miniaturization, we have also begun the design of Diels-Alder stabilized and/or ligated full-length proteins. To realize these goals, we will implement non-canonical amino acid (ncAA) incorporation [120, 121]. Unlike many stabilization chemistries, the Diels-Alder reaction can occur spontaneously and – through judicious selection and design of diene- and dienophile-containing ncAAs – selectively. Cyclopentadienyl-functionalized lysine derivatives have been utilized for

ncAA incorporation and have been found to homodimerize when positioned in high proximity [223]. Further, acryloyl and norbornene-functionalized ncAAs have also been demonstrated [224]. Taken together, this suggests the potential for both homo- and heterodimerization strategies for Diels-Alder stabilized or ligated proteins. Beyond the ability to selectively cross-link proteins (potentially in tandem with other stabilization or bioconjugation chemistries) and many of the same benefits described above for DAC peptides, this provides the basis for genetically-encoded screening. Thus, the application of Diels-Alder chemistry with orthogonal protein expression would greatly enhance our abilities to stabilize a wide variety of peptide and protein structures.

3.2.8 DAC atrial natriuretic peptide demonstrates Diels-Alder peptide stabilization is compatible with disulfide formation

To further expand the scope of Diels-Alder cycloadditions for peptide stabilization, we turned our attention to atrial natriuretic peptide (ANP). ANP, first characterized in 1984, is a human hormone that functions as a diuretic, natriuretic, and vasodilator, primarily through engagement with natriuretic peptide receptor A (NPRA), a cell-surface receptor with guanyl cyclase activity (**Figure 3.17**) [225, 226]. While a synthetic version of the hormone has seen some clinical utility for the treatment of heart failure in Japan, recent investigations suggest ANP may reduce inflammatory reactions during surgery and immunotherapy treatment (in part by reducing the release of cytokines) as well as potentially playing a preventative role toward cancer metastasis [227-229]. While ANP is cyclized via a disulfide linkage, it has a very short circulating half-life (<5 minutes). This has resulted in the investigation of extending ANP stability for improved pharmacologic

utility [230]. Further, the recent discovery of post-translational modifications of ANP, and their effects on ligand binding and stability, suggest that additional chemical modification may enhance or otherwise alter its pharmacologic properties [231].

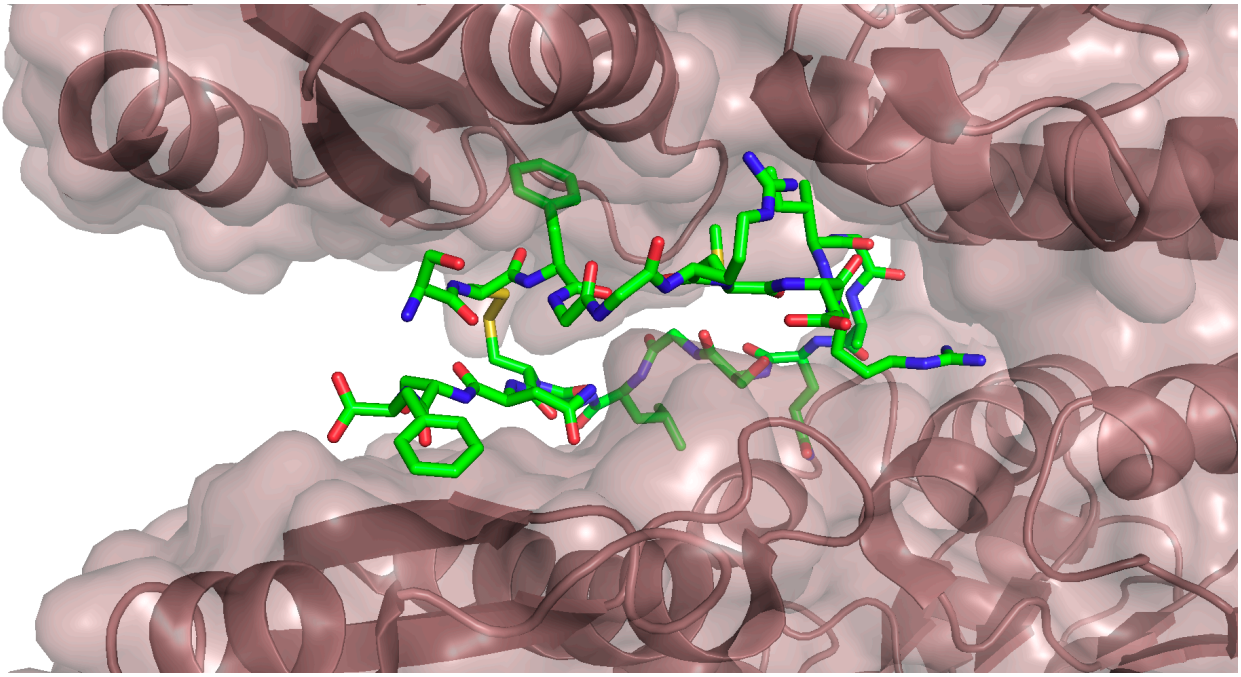
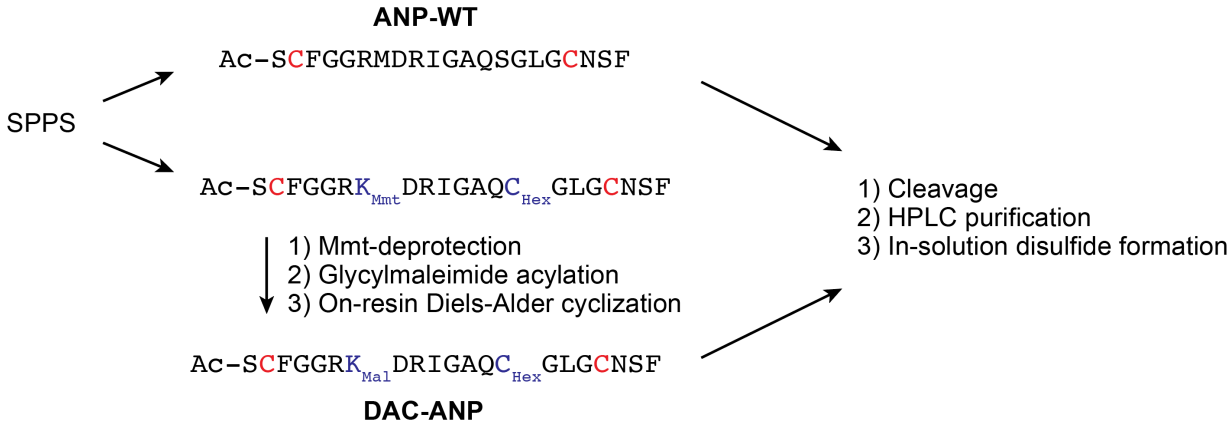


Figure 3.17: ANP (green, sticks) bound to its receptor, ANPR (PDB 1YK0).

ANP is an intriguing target, not only for the potential enhancements Diels-Alder stabilization may confer to peptide stability and activity, but also it allows us to assess the compatibility of DAC peptide chemistry with the formation of disulfide bonds. As noted previously, many approaches for peptide stabilization are not compatible with each other. Further, some strategies, such as disulfide formation, require careful material handling and control of reaction conditions and/or potentially harsh or forcing conditions for proper and complete intramolecular cross-linking. Thus, they may not be compatible with some desired peptide or protein folds or other stabilization chemistries.

We chose to synthesize a 21mer ANP sequence, comprising all of the residues which make contact with its receptors, including the two cysteine residues which form the disulfide bridge [232]. Precursor peptides were synthesized by standard Fmoc-SPPS. The wildtype species (**ANP-WT**) was cleaved, HPLC purified, and lyophilized prior to oxidative formation of the disulfide bond (**Scheme 3.3**). In the DAC peptide analog of ANP (**DAC-ANP**), lysine(Mmt) and cysteine(1,4-hexadiene) were incorporated at residues 7 and 14, respectfully. Following on-resin selective lysine deprotection and acylation with *N*-maleimidoglycine, Diels-Alder cyclization was driven to ~70% completion (as assessed by test cleavage and LC-MS analysis) by heating the resin at 45°C in DMSO for 4 hours. Only a trace amount of an earlier-eluting product (presumably cyclized with the *exo* cycloadduct) was observed. Similar to the wildtype analog, **DAC-ANP** was cleaved, HPLC purified, and lyophilized prior to oxidative formation of the disulfide bond (**Scheme 3.3**).



Scheme 3.3: Synthetic scheme for ANP-WT and DAC-ANP precursor synthesis, purification, and oxidation.
Residues in red form the disulfide bridge; blue residues are functionalized with Diels-Alder pair *N*-maleimidoglycine and hexadiene, for on-resin cyclization.

Oxidation of the purified ANP peptides was carried out under dilute conditions (~0.1 mM), in ammonium acetate buffer, at pH 8.5 and room temperature. Following a 1 hour incubation, samples were removed for analysis by LC-MS. In the case of the wildtype peptide **ANP-WT**, ~15% conversion to the putatively oxidized species was observed, as monitored by the appearance of a new peak in the LC trace, with an earlier retention time relative to the purified, fully-reduced precursor peptide (**Figure 3.18 A**). Beyond the shift in retention time, an anticipated 2 dalton decrease in the mass spectra of the putatively oxidized species was observed. This conversion is evident in the corresponding aligned peaks of the extracted ion chromatograms (EICs) for the calculated major ions of reduced and oxidized **ANP-WT** (calculated $[m+2H]^{2+}$: 1102.5 and 1101.5, respectively) (**Figure 3.18 B**; **Table 3.6**). Similar chromatographic trends were observed by LC-MS analysis of **DAC-ANP** following 1 hour incubation, however a shoulder can be observed at the front of the initial HPLC-purified reduced DAC-ANP peak, with a corresponding peak observed in the EIC for the oxidized species, indicating that oxidation was already occurring, even in the mildly acidic LC-MS loading buffer. Due to it already being cross-linked with a Diels-Alder cycloadduct, a minimal shift in retention time is observed for the oxidized species at 1 hour by absorbance (here primarily apparent only as peak-broadening) and EIC (**Figure 3.18 C,D**; **Table 3.6**). Extended incubation of **ANP-WT** peptide under oxidizing conditions resulted in moderate conversion to the disulfide cross-linked target; a trace amount of reduced peptide remained, along with other side products, which appeared to correspond to oxidation of cysteine or methionine residues. Conversely, bicyclic DAC-

ANP disulfide-crosslinked peptide was formed as the major product and remained stable over the extended incubation.

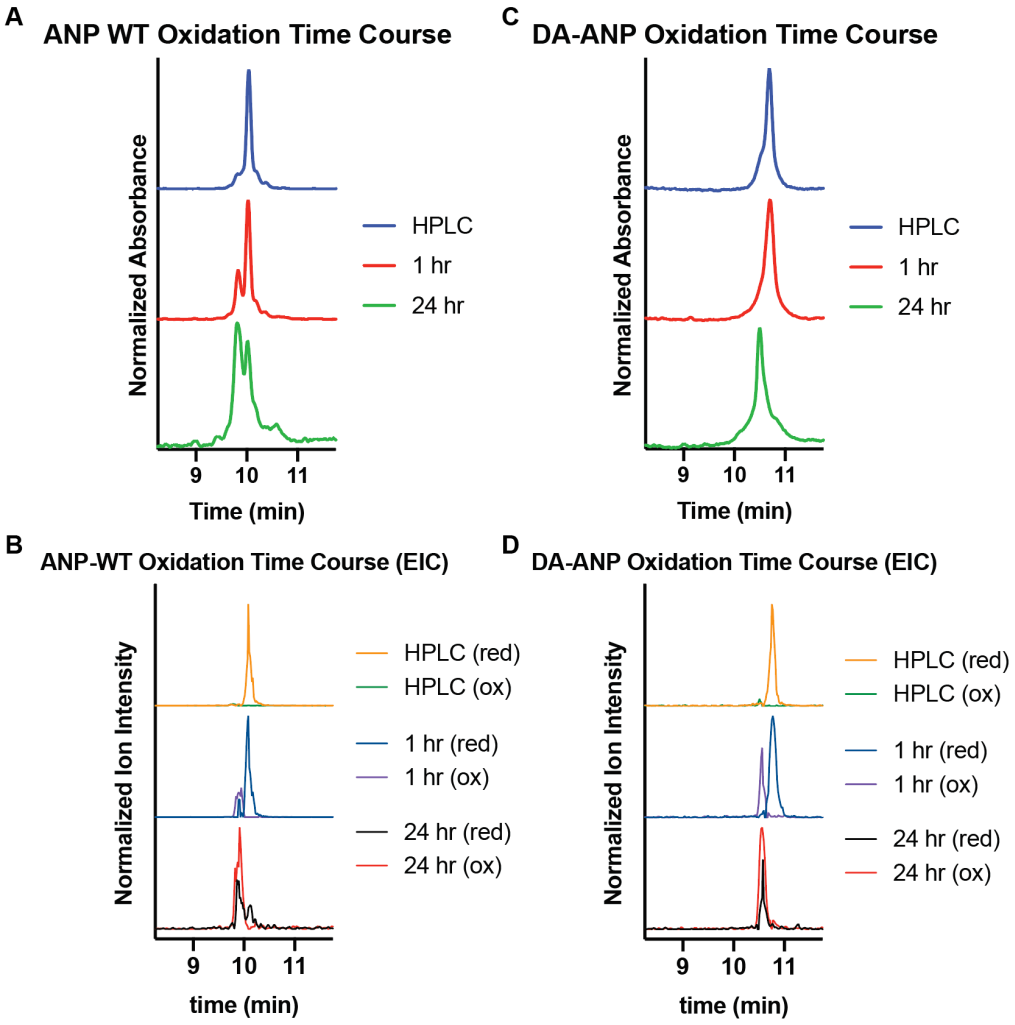


Figure 3.18: LC-MS analysis of ANP-WT and DAC-ANP oxidation time course. **(A)** HPLC traces of initially fully-reduced **ANP-WT** peptide (blue); 1 hour oxidation time point displaying partial **ANP-WT** oxidation (red); 24 hour oxidation time point displaying increased oxidation of **ANP-WT** and considerable side product formation (green). **(B)** extracted ion chromatogram (EIC) LC–MS traces of reduced ($[M+2H]^{2+}$ = 1102.9 ± 0.5) and oxidized ($[M+2H]^{2+}$ = 1101.4 ± 0.5) **ANP-WT** at corresponding time points (oxidized species elutes earlier on column). **(C)** HPLC traces of reduced **DAC-ANP** peptide (blue); 1 hour oxidation time point displaying partial **DAC-ANP** oxidation (red; apparent through peak broadening); 24 hour oxidation time point displaying near-complete oxidation of DAC-ANP and overall compound stability (green). **(D)** extracted ion chromatogram (EIC) LC–MS traces of reduced ($[M+2H]^{2+}$ = 1217.8 ± 0.5) and oxidized ($[M+2H]^{2+}$ = 1216.9 ± 0.5) **DAC-ANP** at corresponding time points (corresponding shifts in retention time and mass most apparent in middle EIC traces).

Taken together, these results indicate that DAC peptide chemistry is compatible with disulfide cross-linking. Further, in the context of this ANP peptide scaffold and under these relatively mild oxidizing conditions, the bicyclic **DAC-ANP** disulfide peptide oxidized more readily and with no side reactions, relative to the **ANP-WT** peptide. Intriguingly, on-resin Diels-Alder cross-linking templates the subsequently cleaved and deprotected peptide to promote a favorable conformation for proper disulfide formation. This suggests a potential strategy in which multiple disulfides (or other cross-links) could be formed with the desired residue-to-residue pairing by first cyclizing the peptide with Diels-Alder chemistry.

Future studies of this first **DAC-ANP** peptide should go toward investigation of its chemical and conformational structure relative to **ANP-WT** and confirming the presence of the anticipated *endo* cycloadduct, as ascertained by NMR. Screening for differential rates of proteolysis will confirm the hypothesis that the bicyclic **DAC-ANP** is more rigidified and therefore will have a longer circulating half-life relative to the wildtype peptide. Screening for receptor engagement using reporter cell lines or commercially available kits to quantify generation of cGMP in cells could be used to compare compound efficacy [230]. Following validation of these screening assays, a broader synthetic campaign could then be initiated, testing alternate Diels-Alder cross-linked positions, reactive pairs, and orientations, with or without the disulfide bond, or with other stabilization chemistries. If multiple DAC peptide stereoisomers were produced in sufficient yield, they should also be screened for stereochemical induction of favorable peptide conformations or cycloadduct contributions for enhanced stability or receptor binding. Further, should

suitable lead compounds emerge, they could be applied to cell culture model systems of metastasis and cell migration (*e.g.* scratch wound assays) and potentially in vivo cancer and/or cytokine release syndrome models [227, 229].

3.3 Discussion

Backbone- and sidechain-directed peptide macrocyclization can markedly improve pharmacologic properties for chemical probes or therapeutics. Our studies establish several emergent and optimal characteristics that Diels-Alder cycloadditions can add to the chemical arsenal used for the macrocyclization and stabilization of peptides. First, we demonstrated the suitability of a range of reactive diene–dienophile pairs for the cyclization of diverse peptide sequences in both organic and aqueous reaction environments, on-resin or in-solution. This opens up many possibilities, including the tailoring of macrocyclization motifs to both improve peptide structure through rigidification and corresponding physicochemical properties and target binding. This potential is emphasized by the observation that the preferred *endo* DAC isomer of the SRC2 peptide not only better templates an alpha helix relative to the presumably *exo* minor product, but also directly contributes to ERalpha binding. Specifically, the convex hydrophobic surface of the adduct presented by the *endo* isomer rests on a hydrophobic surface adjacent to the AF2 pocket. Further exploration of this cycloadduct-target interaction, and others, are warranted to ascertain if alternate Diels–Alder cycloadducts could be developed to contribute selective polar and non-polar contacts to binding surfaces. Another noteworthy feature of Diels-Alder chemistry, as applied to peptide macrocyclization, is potential access to unique adduct structures that could convey privileged peptide conformations,

analogous to the desirable small molecule scaffold achieved through the use of the Diels-Alder reaction in organic synthesis. Given that the majority of stabilization chemistries result in predominantly hydrophobic cross-links, the Diels-Alder reaction can offer a wide variety of different adducts with correspondingly different physicochemical properties. Indeed, it is under-appreciated that while added hydrophobicity can improve target binding affinity, this may come at the expense of increased off-target interactions. The range of possible Diels–Alder cycloadducts may enable design of and selection for macrocycles that temper these contributions, or even increase the specificity of DAC peptide ligands. Structural studies of resulting Diels-Alder stabilized loop and helical peptides confirmed the formation of the expected *endo* Diels–Alder cycloadduct. Furthermore, we established that mild heating on-resin in organic solvents can drive stereoselective formation of the *endo* isomer across diverse peptides. On the basis of the evidence here, we expect that this will be general for many peptides, however future studies are warranted to determine if there are conditions that will bias toward the formation of the *exo* isomer if desired.

Another noteworthy feature of this DAC peptide strategy is its essentially "reagentless" nature. That is, following installation of the reactive diene and dienophile pairs, macrocyclization occurs spontaneously, in contrast to many other cross-linking approaches that require harsh, forcing conditions, metal catalysts, or other exogenous reagents. These requirements limit reaction conditions and the range of suitable peptide scaffolds, face the challenge of cross-reactivity with other functional groups present within peptides, and are further hampered by practical considerations of cost and reagent

removal during synthesis and purification. Diels–Alder cycloaddition on peptides, by contrast, is primarily driven by proximity, chemical environment, and reaction conditions. This suggests cyclization or intramolecular cross-linking could be performed on native protein folds, or aid in proper folding, alone or in concert with other chemistries. As was seen for DAC-p53 peptide **5**, an aqueous reaction environment promoted spontaneous cyclization and preferentially formed the more helical macrocyclic isomer **5a** helicity. Indeed, this implies the context specific contributions of different Diels–Alder adducts and even stereochemical products of unique reactive pairs could be screened for inherent stabilization and extrinsic effects on the activity of stabilized peptides.

Another attractive feature offered by new approaches to the chemical modification and stabilization of peptides is their application to unique and larger peptide structures, including those containing multiple secondary structural elements. The work presented here suggest current and future Diels-Alder cyclization pairs can contribute to these goals in two ways. First, the proper folding of larger peptide and proteomimetic structures may be required for optimal templating and formation of cross-links or other chemical modifications. Thus, such chemistries would need to robustly and selectively operate in the aqueous environments necessary for folding to occur. Our demonstrations for the suitability of Diels-Alder peptide stabilization in these environments presents it as an attractive option for cyclization, in contrast to other chemistries that cannot operate on unprotected peptides and proteins, or in aqueous environments. Further, the stabilization of larger, more complex peptide structures may require multiple stabilizing elements, however the use of multiple compatible chemistries, without excessive engineering or

manipulations, remains a challenge. To begin to address these challenges, herein we demonstrated both facile double stapling of a helical SRC2 peptide, contained RCM and Diels-Alder reactive residues, as well as the tandem bicyclization of an ANP peptide scaffold via Diels-Alder cycloaddition and disulfide bond formation. Future explorations of peptide and protein stabilization with Diels–Alder chemistry will explore its application to diverse secondary and tertiary structures, incorporation with other natural and non-natural stabilizing elements, and the structural and functional study of specific reactive pairs for improved pharmacologic properties. Further, we have initiated work toward the formation of non-canonical amino acid incorporation of Diels-Alder compatible amino acids for the cyclization and ligation of full-length proteins via Diels-Alder chemistry. In conclusion, we believe the future application of Diels–Alder chemistry will provide diverse opportunities for the construction of enhanced peptide macrocycle and protein domains.

3.4 Tables

sequence	compound(s)	% conv.	product ratio*	retention time (min)	m/z (calc./obs.)
CRGDK	**1wt**	-	-	10.1	619.3 / 619.3
C_{tBuS}RGDK	**1*t*BuS**	-	-	10.8	707.3 / 707.3
C_{Hex}RGDK	**1hex**	quant.	-	11.3	699.4 / 699.3
C_{Hex}RGDK$_{Mal}$	**1, 1a, 1b**	95	>50:1	12.4, 11.2, 9.1	836.4 / 836.3
C_{tBuS}VGDK	**2*t*BuS**	-	-	12.2	650.8 / 650.3
C_{Hex}VGDK	**2hex**	quant.	-	12.5	641.8 / 642.4
C_{Hex}VGDK$_{Mal}$	**2, 2a, 2b, 2c**	92	9:1	13.8, 11.9, 11.8, 9.9	778.9 / 779.3
C_{tBuS}APVYK	**3*t*BuS**	-	-	13.6	810.1 / 809.3
C_{Hex}APVYK	**3hex**	quant.	-	13.7	802.0 / 801.4
C_{Hex}APVYK$_{Mal}$	**3, 3a, 3b, 3c**	89	8:1	14.7, 13.1, 13.2, 11.0	939.1 / 938.3
CAVPAVYK	**4wt**	-	-	13.1	891.5 / 891.3
C_{tBuS}AVPAVYK	**4*t*BuS**	-	-	14.0	980.3 / 980.4
C_{Hex}AVPAVYK	**4hex**	quant.	-	14.0	972.2 / 971.4
C_{Hex}AVPAVYK$_{Mal}$	**4, 4a, b, 4c**	85	9:1	15.3, 13.7, 13.8, 11.8	1109.3 / 1108.4

Table 3.1: Diels-Alder cyclized loop peptides and precursors.
Compound list and data for turn and loop motif DAC peptides and their precursors synthesized for this study. * Reaction conditions: 2-6 hr heating at 45°C, on-resin, in DMSO.

Solvent	temp. (°C)	% conv.*	major:minor
control	n/a	77.4	12.3
H_2O	25	81.9	11.8
DMF	25	85.3	9.9
H_2O	45	89	15.2
DMF	45	88.2	23.5
DMSO	45	92.2	50.2

Table 3.2: Results from a screen of on-resin cyclization reaction conditions for compound 1.
Increased conversion and major product enrichment is seen for many compounds with extended heating. *****Reactions were carried out for 2 hours followed by standard cleavage from resin.

sequence	compound(s)	% conv.	product ratio[a]	retention time (min)	m/z (calc./ obs.)
QSQQTFKNLWRLLA$_{Fur}$QN	5pre	-	-	15.3	1041.5/ 1041.3
QSQQTFK$_{Mal}$NLWRLLA$_{Fur}$QN	5, 5a, 5b, 5c	88	6:1	15.6, 15.2, 15.0, 14.8	1110.1/ 1101.3
HKILHRLLQDS	SRC2-WT	-	-	11.1	700.9/ 701.0
fitc-b-SLTERHKILHRLLQE	fitc-SRC1-BoxII	-	-	12.3	778.1/ 778.2
HKKLHRA$_{Fur}$LQDS	6pre	-	-	8.1	720.4/ 720.5
HKK$_{Mal}$LHRA$_{Fur}$LQDS	6, 6a, 6b	95	4:1	8.9, 8.7, 8.2	788.9/ 788.8
HKILHKLLQA$_{Fur}$DS	7pre	-	-	10.2	697.9/ 698.0
HKILHK$_{Mal}$LLQA$_{Fur}$DS	7, 7a, 7b, 7c	95	7:1	11.4, 11.1, 10.9, 10.6	766.4/ 766.5
HKS$_5$LHKS$_5$LQA$_{Fur}$DS	8pre	-	-	11.6	723.9/ 724.0
HKS$_5$*LHKS$_5$*LQA$_{Fur}$DS	8rcm	>95	-	11.2	709.9/ 710.0
HKILHK$_{Mal}$LLQA$_{Fur}$DS	8, 8a	>95	-	11.9, 11.0	778.4/ 778.5

Table 3.3: Compound list and data for helix motif DAC peptides, their precursors, and related analogs synthesized for this study.
[a] Reaction conditions: 2-6 hours heating at 45 °C, on resin, in DMSO. Ratio reported as the DAC peptide major product:minor product(s). * Metathesis cross-linked residues.

μM (±s.e.m.)	**6a**	**6b**	**7a**	**8a**
IC_{50}	2.74 (0.095)	10.6 (0.160)	0.48 (0.004)	2.56 (0.100)
K_i	1.19 (0.041)	4.61 (0.068)	0.21 (0.001)	1.12 (0.045)

Table 3.4: Calculated IC50 and Ki values from competitive FP assay of DAC-SRC2 peptides targeting ER alpha.

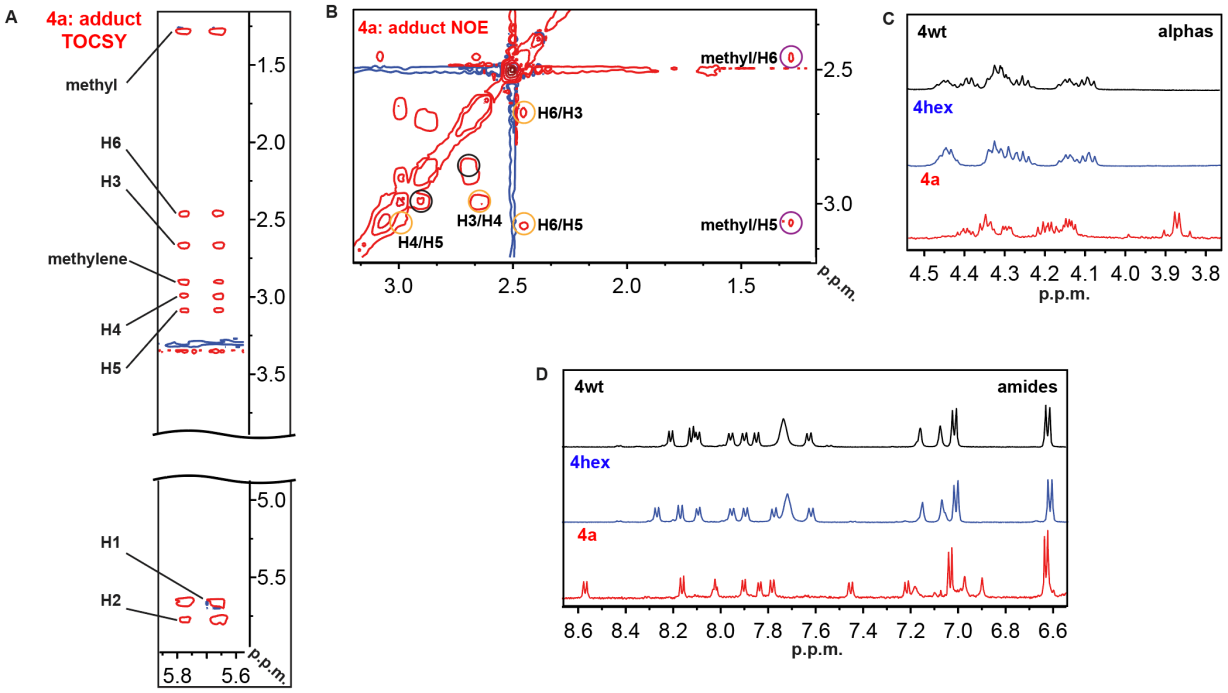
	Y537S ERα LBD-E2-6a
Data Collection	
Wavelength (Å)	0.97
Resolution range (Å)	43.5 - 2.25 (2.28 - 2.25)
Space group	P 1 2_1 1
Unit cell	
***a, b, c* (Å)**	55.534, 71.03, 58.082
α, β, γ (°)	90, 108.68, 90
Total reflections	21,894
Unique reflections	6,634
Redundancy	3.3
Completeness (%)	99.5 (97.9)
Mean I/sigma(I)	13.5 (1.5)
Wilson B-factor	29.78
R_{pim} (%)	13.6 (48.7)
$CC_{1/2}$	0.516
CC*	0.825
Refinement	
Reflections used for R-free:	
R-work	0.2182 (0.2758)
R-free	0.2471 (0.2619)
Number of non-hydrogen atoms	3721
macromolecules	3454
ligands	114
water	153
Protein residues	443
RMS(bonds)	0.005
RMS(angles)	0.9
Ramachandran favored (%)	98.8
Ramachandran allowed (%)	1.2
Ramachandran outliers (%)	0
Clashscore	2.25
Average B-factor	37.4
macromolecules	36.9
ligands	52.1
solvent	37.6

Table 3.5: Data collection and refinement statistics for co-crystal structure of Y537S ER alpha LBD, estradiol, and DAC-SRC2 peptide 6a (PDB: 6PIT).

sequence	redox state	% conv.[a]	retention time (min)	m/z (calc./ obs.)
SCFGGRMDRIGAQSGLGCNSF	reduced	-	10.1	1102.5/1102.9
SCFGGRMDRIGAQSGLGCNSF	oxidized	65	9.9	1101.5/1101.4
SCFGGRK$_{Mal}$DRIGAQC$_{Hex}$GLGCNSF	reduced	70	10.5	1217.5/1217.8
SCFGGRK$_{Mal}$DRIGAQC$_{Hex}$GLGCNSF	oxidized	89	10.6	1216.5/1216.9

Table 3.6: Compound list and data for ANP-WT and DAC-ANP peptides.
[a] R% conversion approximated from area under the curve calculations of 215 nm absorbance product peak relative to total area; DAC-ANP reduced (third table entry) calculated from resulting LC-MS profile of test cleavage after 4 hours heating at 45°C on resin in DMSO; ANP-WT and DAC-ANP oxidized (second and fourth table entry, respectively) calculated from resulting LC-MS profile following 24 hours room temperature incubation in buffer.

3.5 Spectroscopic data



Representative NMR spectroscopic data for DAC peptide 4a and linear precursors. **(A)** TOCSY plot resolving through-bond interactions between **4a** vinylic protons and all protons within the cycloadduct and participating exocyclic groups. **(B)** NOESY plot with corresponding through-space interacting groups circled by colors corresponding to those in **Figure 3.8 B**. (C,D) ^{1}H-NMR spectra for backbone alpha **(C)** and amide **(D)** protons within **4wt**, **4hex** and **4a**.

Fmoc-Cys(2,4-hexadiene)-OH

^{1}H-NMR (500 MHz, $CDCl_3$-d_1) δ 7.75 (d, 2H, Ar CH), 7.60 (t, 2H, Ar CH), 7.39 (t, 2H, Ar CH), 7.31 (t, 2H, Ar CH), 6.33 (br d, 1H, NH), 5.97–6.08 (m, 2H, C=CH), 5.62–5.68 (m, 1H, C=CH), 5.42-5.47 (m, 1H, C=CH), 4.60 (m, 1H, α-H), 4.41 (d, 2H, Fmoc CH 2), 4.23 (t, 1H, Fmoc CH), 3.16 (d, 2H, β-H), 2.98 (d, 2H, S-CH 2), 1.70 (d, 3H, CH 3).

4wt (Ac-CAVPAVYK)

^{1}H-NMR (500 MHz, DMSO-d_6) δ 9.19 (s, 1H), 8.20 (d, J = 7.3 Hz, 1H), 8.12 (d, J = 8.1, 1H), 8.09 (d, J = 7.2 Hz, 1H), 7.95 (d, J = 8.0 Hz, 1H), 7.89 (d, J = 8.17 Hz, 1H), 7.84 (d, J = 8.4 Hz, 1H), 7.73 (s, 3H), 7.62 (d, J = 8.4 Hz, 1H), 7.16 (s, 1H), 7.07 (s, 1H), 7.01 (d, J = 8.4 Hz, 2H), 6.62 (dd, J = 8.3 Hz, 2H), 4.44 (q, 1H), 4.38 (q, 1H), 4.31 (m, 3H - overlap), 4.25 (m, 1H), 4.14 (q, 1H), 4.09 (q, 1H), 3.68 (d, 1H), 3.56 (d, 1H), 2.91 (dd, J = 14.3, 4.7

Hz, 1H), 2.74 (m, 3H - overlap), 2.66 (m, 2H), 2.38 (t, *J* = 8.5 Hz, 1H), 2.01 (m, 1H), 1.94 (m, 1H), 1.9 (m, 1H), 1.87 (s, 3H), 1.86 (m, 1H), 1.80 (m, 2H), 1.66 (m, 1H), 1.51 (m, *J* = 8.2 Hz, 3H), 1.27 (m, 2H), 1.19 (m, 6H - overlap), 0.91 (d, 3H), 0.87 (d, 3H) , 0.75 (d, 6H).

4hex (Ac-C^{hex}AVPAVYK)
^{1}H-NMR (500 MHz, DMSO-d_6) δ 9.19 (s, 1H), 8.26 (d, *J* = 7.4 Hz, 1H), 8.16 (d, *J* = 8.3 Hz, 1H), 8.09 (d, *J* = 7.1 Hz, 1H), 7.95 (d, *J* = 7.9 Hz, 1H), 7.89 (d, *J* = 8.1 Hz, 1H), 7.77 (d, *J* = 8.4 Hz, 1H), 7.72 (s, 3H), 7.62 (d, *J* = 8.6 Hz, 1H), 7.15 (s, 1H), 7.07 (s, 1H), 7.01 (d, 2H), 6.62 (d, 2H), 6.12 (dd, *J* = 14.7, 10.6 Hz, 1H), 6.04 (dd, *J* = 14.9, 10.5, 1H), 5.66 (dq, *J* = 13.7, 6.7 Hz, 1H), 5.49 (dt, *J* = 14.8, 7.5 Hz, 1H), 4.44 (m, 2H - overlap), 4.33 (m, 3H - overlap), 4.25 (q, 1H), 4.14 (q, 1H), 4.09 (q, 1H), 3.70 (m, 1H), 3.56 (m, 1H), 3.19 (dd, *J* = 7.6, 2.6 Hz, 2H), 2.92 (d, 1H), 2.75 (m, 2H), 2.70 (m, 2H), 2.55 (d, 1H), 2.01 (dt, 1H), 1.91 (m, 3H – overlap), 1.85 (s, 3H), 1.81 (m, 2H - overlap), 1.71 (d, 3H), 1.66 (m, 1H), 1.51 (m, 3H - overlap), 1.27 (m, 2H), 1.19 (t, 6H - overlap), 0.91 (d, 3H), 0.86 (d, 3H), 0.75 (d, 6H).

4a (Ac-C^{Hex}AVPAVYK^{Mal} – *endo* adduct stereochemistry)
^{1}H-NMR (600 MHz, DMSO-d_6) δ 9.14 (s, 1H), 8.58 (d, *J* = 8.1 Hz, 1H), 8.17 (d, *J* = 8.2 Hz, 1H), 8.03 (t, *J* = 5.6 Hz, 1H), 7.91 (d, *J* = 7.2 Hz, 1H), 7.84 (d, *J* = 6.9 Hz, 1H), 7.79 (d, *J* = 8.6 Hz, 1H), 7.46 (d, *J* = 8.9 Hz, 1H), 7.22 (d, J = 8.46 Hz, 1H), 7.04 (d, *J* = 8.5 Hz, 2H), 6.97 (s, 1H), 6.90 (s, 1H), 6.63 (d, *J* = 8.4 Hz, 2H), 5.77 (dt, *J* = 9.1, 3.1, 2.9 Hz, 1H), 5.66 (dt, *J* = 9.1, 3.1, 3.0 Hz, 1H), 4.40 (m, 1H), 4.35 (m, 2H - overlap), 4.30 (dd, 1H), 4.2 (m, 2H - overlap), 4.14 (m, 2H - overlap), 3.87 (d, *J* = 6.7 Hz, 2H), 3.63 (m, 1H), 3.55 (m, 1H), 3.07 (m, 3H - overlap), 2.99 (m, 2H - overlap), 2.90-2.85 (m, 2H - overlap), 2.70-2.63 (m, 3H - overlap), (2.47 – overlapped by DMSO signal), 2.08-1.96 (m, 4H - overlap), 1.95 (m, 1H), 1.84 (s, 4H - overlap), 1.66 (m, 1H), 1.46 (m, 2H - overlap), 1.32-1.23 (m, 8H - overlap), 1.18 (m, 1H), 1.09 (d, 3H), 0.96 (d, 3H), 0.88 (d, 3H), 0.75 (d, 3H), 0.70 (d, 3H).

DAC helical peptide ^{1}H-NMR measurements for vinylic protons of cycloadduct:

1a: 5.73 (dt, *J* = 9.2, 3.1, 3.0 Hz, 1H), 5.66 (dt, *J* = 9.1, 3.3, 3.0 Hz, 1H).

2a: 5.75 (dt, *J* = 9.1, 3.1, 3.1 Hz, 1H), 5.66 (dt, *J* = 9.1, 3.3, 3.1 Hz, 1H).

3a: 5.72 (dt, *J* = 9.3, 3.1, 3.0 Hz, 1H), 5.65 (dt, *J* = 9.2, 3.3, 3.0 Hz, 1H).

5a: 6.34 (d, *J* = 5.6 Hz, 1H), 6.21 (d, *J* = 5.8 Hz, 1H), 5.28 (d, *J* = 5.6 Hz, 1H).

6a: 6.36 (d, *J* = 5.8 Hz, 1H), 6.19 (d, *J* = 5.6 Hz, 1H), 5.24 (d, *J* = 5.6 Hz, 1H).

7a: 6.40 (d, *J* = 5.8 Hz, 1H), 6.21 (d, *J* = 5.8 Hz, 1H), 5.25 (d, *J* = 5.7 Hz, 1H).

8a: 6.48 (d, *J* = 5.8 Hz, 1H), 6.22 (d, *J* = 5.8 Hz, 1H), 5.43 (dt, *J* = 9.7, 6.5 Hz, 1H), 5.31 (m - 1H, overlap), 5.28 (d, *J* = 5.8 Hz, 1H).

cpdp-COOH
^{1}H NMR (500 MHz, DMSO-d_6) δ 10.59 (br.s., 1 H), 6.48-6.11 (m, 3 H), 2.97 (d, 1 H), 3.01 (d, 1 H), 3.01 - 2.91 (m, 2 H), 2.86 - 2.63 (m, 2 H) ppm.

cpdp-NHS
^{1}H NMR (500 MHz, DMSO-d_6) δ 6.44-6.03 (m, 3 H), 2.99 (d, 1 H), 2.94 (d, 1 H), 2.90 - 2.75 (m, 8 H) ppm.

3.6 Methods

DAC peptide synthesis, purification and analysis.

Peptides were synthesized by Fmoc-based solid phase peptide synthesis (SPPS), ring-closing metathesized (where appropriate) and purified by reverse-phase HPLC with a C18 column (Agilent, Palo Alto, CA) as previously reported [138]. Fmoc-Cys(2,4-hexadiene)-OH and Fmoc-(2-furanyl)alanine-OH were incorporated in a similar manner on resin. Peptides were analyzed by LC-MS using a C18 reverse-phase column (Phenomenex, 5.0 × 50 mm, pore size 110 Å, particle size 5 μm); Buffer A (5/95/0.1% ACN/H2O/TFA) and Buffer B (95:5:0.1% ACN/H2O/TFA); and a 20 min method with the following gradient (flow rate 0.5 mL/min): 0% buffer B over 3 min, 0–65% buffer B over 15 min, 65–100% buffer B over 1 min; 100–0% buffer B over 1 min. Purified peptides were lyophilized in tared low-retention microcentrifuge tubes, quantified by mass and UV

absorbance where relevant, and either brought forward for structural characterization or dissolved in DMSO as 10 mM stocks and stored at -20 °C for later use.

***N*-(9-fluorenylmethoxycarbonyl)-L-cysteine.**

To a suspension of 1.76 g *N*-(9-fluorenylmethoxycarbonyl)-*S*-trityl-L-cysteine (3 mmol) in 50 mL dichloromethane was added 3 mL triisopropylsilane (14.6 mmol), followed by 10 mL trifluoroacetic acid (131 mmol). The resulting mixture was stirred at room temperature for 10 minutes until it turned colorless. The solvent was removed under reduced pressure and the residual mixture was triturated by hexanes, yielding 1.03 g white solid (quantitative), which was directly used in the following step without further purification.

***N*-(9-fluorenylmethoxycarbonyl)-S-(*trans*,*trans*-2,4-hexadienyl)-L-cysteine.**

A solution of 1.03 g *N*-(9-fluorenylmethoxycarbonyl)-L-cysteine (3 mmol) and 0.83 g potassium carbonate (6 mmol) in 50 mL *N*,*N*-dimethylformamide was protected under nitrogen and cooled to 0 °C. 0.58 g *trans*,*trans*-2,4-hexadienyl bromide (3.6 mmol) was added, and the reaction mixture was stirred for 1 hour at 0 °C under nitrogen. After the reaction was complete as monitored by TLC, 50 mL water was added and the mixture was acidified to pH 2 with 1 N HCl. The mixture was extracted with ethyl acetate (50 mL × 3). The organic layer was then combined, washed with 5% LiCl (25 mL × 4), brine (50 mL × 2) and dried over sodium sulfate. The solvent was removed under reduced pressure,

and the crude mixture was purified by flash chromatography to yield 0.77 g off-white solid (**1c**, 61%) with R_f = 0.15 (9:1 dichloromethane/methanol).

DAC peptide functionalization.

DAC peptide functionalization was typically carried out on 30 μmol MBHA rink amide (0.86 mmol/g) or NOVAPEG (0.31 mmol/g) resins following SPPS and N-terminal acetylation. When necessary, Cys-StBu deprotection was achieved by bubbling with N_2 in 3 mL 20/10/70% βME/DIPEA/DMF, 2x 2 hour. Following copious washing with DMF and DCM, Cys alkylation was carried out to completion with 19 mg bromohexadiene (4 eq; 120 μmol) (synthesized as previously reported [233]) and 41 μL DIPEA (8 eq; 240 μmol) in 3 mL DMF, 4 hour at room temperature, as monitored by LC-MS analysis. Alternatively, Fmoc-Cys(2,4-hexadiene)-OH Following washing with DMF and DCM, Lys-Mmt deprotection was achieved by bubbling with N_2 in 3 mL 1% TFA/DCM, 5x 2 min. Following copious washing with DCM and DMF, Lys acylation was carried out by first pre-activating 4.65 mg *N*-maleimidoglycine (3 eq; 90 μmol) (synthesized as previously reported [234]) with 37 mg HCTU (3 eq; 90 μmol) for 10 min in 0.5 mL DMF. The solution was then added to the resin with free amine-containing peptide with 1 mL DMF and 16 μL DIPEA (3 eq; 90 μol) and bubbled with N_2 for 5-10 min. Following copious washing and drying of the resin and peptide cleavage, complete maleimide incorporation and macrocyclization was analyzed by LC-MS. Similar procedures were followed for the incorporation of acryloyl and crotonoyl functional groups onto free Lys on resin.

DAC peptide 1 cyclization time-course.

Cys-alkylated **1hex** was lysine-functionalized with *N*-maleimidoglycine dienophile according to the above protocol. Resin was then dried. A sample was reserved for analysis at 4°C and remaining resin split into 4 samples. These samples were brought up in DMF and agitated at 1,000 rpm at room temperature for 30 min, 1 hr, 2 hr, and 6 hr. At each time point, solvent was removed by filtration, resin dried, peptide cleaved, and product profile analyzed and quantified by LC-MS.

Optimization of DAC peptide reaction conditions.

On resin reaction condition screens were carried out immediately after *N*-maleimidoglycine acylation of RGD peptide **1hex** as described above. Briefly, dry resin (~2-5 μmol) was split between microcentrifuge tubes and 500 μL of the indicated solvents (**Table 3.2**) were added and incubated at room temperature or 45 °C for 2 hr. Following peptide cleavage, each reaction profile was analyzed by LC-MS. All future on-resin cyclization reactions were carried out in DMSO and heated at 45°C for 2-6 hr as empirically determined by DAC peptide conversion profiles as observed by LC-MS.

DAC peptide maleimide trapping and cyclization experiment.

Cys-alkylated **1hex** was lysine-functionalized with *N*-maleimidoglycine dienophile according to the above protocol. Resin was then split and either incubated in DMF or 20% βME/DMF for 2 hr. Following copious washing, solvent was removed by filtration, resin dried, peptide cleaved, and product profile analyzed by LC-MS.

DAC peptide cycloadduct chemical stability assay.

Purified, lyophilized peptide **1a** (1 eq) or a mixture of cyclic and linear peptide **5** (1 eq) were dissolved in PBS, pH 8.0. βME was added (20 eq) and the solution incubated at 37 °C, 1,000 rpm. 2 hr, 6 hr, and 24 hr time-points were frozen and lyophilized (a single 2 hr time point was taken for **5**). Lyophilized samples were dissolved in 0.1% TFA/H_2O and analyzed by LC-MS.

DAC protease stability assays.

DAC peptide protease degradation kinetics were determined with Thermo Pierce™ MS-Grade Trypsin or Chymotrypsin Protease in PBS or TBS with 5 mM DTT and 2 mM $CaCl_2$. Protease (1 eq) was added to buffer for a final assay concentration of 10 μg/mL trypsin or 0.5 μg/mL chymotrypsin. Each respective reaction mixture was then added to either **1wt** and **1a** (2000 eq) or **4wt** and **4a** (10,000 eq). Samples were then incubated at 37 °C, 1,000 rpm. 5 min, 10 min, 30 min, 2 hr, and 4 hr time points were quenched in an equal volume of 2% TFA/H_2O on ice. After 5 min, samples were centrifuged for 2 min at 17,000 rcf to precipitate quenched trypsin. Supernatant solutions were then analyzed and intact peptide quantified by LC-MS or Q-TOF.

DAC scratch wound assay.

100,000 HeLa cells were seeded in each well of a 24 wellplate in 1 mL RPMI (GIBCO) supplemented with 10% FBS. Cells were incubated at 37 °C, 5% CO_2 and

reached ~70% confluency within 24 hr. Each well's monolayer of cells was scratched with a sterile 200 μL pipette tip, media and debris was aspirated, wells washed with PBS and then fresh, RPMI was added, containing either DMSO, 10 μM or 20 μM compounds **1wt** or **1a**. Brightfield microscopy images of each well's wound were taken immediately after wounding (time = 0 hr) and following overnight incubation with compounds (time = 16 hr). Two biological replicates were performed for each condition and at least 5 wound distance measures were taken from each image, averaged and normalized relative to the distance of wound closure in the DMSO control.

Peptide NMR spectroscopy.

Peptides were dissolved in 400 μL *d*6-DMSO to ~5 mM for NMR experiments. All experiments were performed on a Bruker AVANCE II+ 500 MHz NMR (**1a**, **2a**, **3a**, **4wt**, **4hex**, **6a**, **7a**, **8a**) or Bruker AVANCE IIIHD 600 MHz NMR (**4a**) using standard acquisition parameters with empirically determined 90-degree pulse widths and in-house methods for solvent suppression for **4a** spectra. Data were processed and plotted in Bruker Topspin 3.5 and MestReNova 11.0.2.

Molecular modeling.

A structural model of **4a** was built using the Molecular Operating Environment molecular modeling package [210]. In accordance with NMR analysis, the *endo* configuration of the Diels-Alder adduct was enforced and backbone torsion angles were manually adjusted according to measured J-coupling constants [235]. Missing atom

names for the covalent linker were assigned and the final structure was exported using the PDB format. Custom topology files for the covalent linker were used to reconstruct the molecular system in CHARMM format (PSF+PDB) using the PSF_{GEN} plugin for VMD [212]. A custom patching protocol was used to prepare the cyclic peptide *in silico* and the N-terminal acetyl and C-terminal amidate patches were used to match the synthesized chemical structure. The SOLVATE plugin was used to generate a 15 Å TIP3P water box and the AUTOIONIZE plugin was used to bring the salt concentration to 0.1 M NaCl. The solvated system totaled ~9.3K atoms. Aside from lacking the covalent linker, the linear peptide was constructed using the same method.

The NAMD 2.12 software package was used for molecular dynamics simulations. The peptide and ions were described using the CHARMM36m forcefield. The required parameters for the covalent linker were assigned by analogy from the CHARM36m and CGENFF parameter sets. All simulations were performed at a target temperature of 298K and a pressure of 1 atm maintained by a Nosé-Hoover thermostat and Langevin piston. Periodic boundary conditions were applied, full system periodic electrostatics were computed using the particle mesh Ewald (PME) method with a grid spacing set to > 1/Å, and non-bonded interactions were treated using an exponential switching function with a switch distance of 10 Å and cutoff of 12 Å. The solvated and ionized systems were energy minimized (20 ps) and velocities were introduced with a harmonic constraint scaling factor of 2 set for all non-hydrogen atoms of the peptides. Following 1 ns of equilibration the harmonic constraints were removed and production simulations were run for an additional 50 ns.

The simulation trajectory files were down-sampled to 10 ps/frame for analysis. Simulation trajectories were aligned to the peptide backbone of the first frame of the respective production run and RMSD was found for the backbone or non-hydrogen atoms using the RMSD Trajectory Tool in VMD. The output RMSD values were plotted using GraphPad PRISM 7. An overlay of the peptide backbone at 250 ps intervals was generated for the image displayed in **Figure 3.7 D**.

DAC peptide 5 in-solution cyclization time-course.

Furan-containing peptide **5pre** was lysine-functionalized with *N*-maleimidoglycine dienophile, resin was dried and then peptide cleaved according to the above protocols. The resulting crude peptide was then subjected to HPLC purification and the predominant linear isomer **5** was isolated and lyophilized overnight. The linear species was then dissolved in PBS and incubated at room temperature. Time points were removed at 0, 2, 12, 24 and 48 hr, at which points the product profile was analyzed by LC-MS.

Circular dichroism (CD) spectroscopy.

CD spectroscopy experiments were performed on a Jasco J-170 using a quartz cuvette (path length: 0.1 cm). Peptides were dissolved to 50 μM in 50 mM phosphate buffer (pH 7.4) and CD measurements were recorded at 0.5 nm increments between 190 and 260 nm, at room temperature.

Competitive fluorescence polarization (FP) assay.

FP experiments were performed in black 96 well plates (Corning) on a Synergy Neo plate reader (BioTek). 3-fold dilution series of DAC peptides **6a**, **6b**, **7a** and **8a** were plated in 60 μL FP buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.01% NP-40), and further diluted by the addition to each well of 30 μL FP buffer containing 750 nM wildtype ERalpha wildtype protein, 3.75 μM estradiol (E2), and 10 nM FITC-SRC1-BoxII peptide (final volume 90 μL; final concentrations: 250 nM ERalpha, 1.25 μM E2, and 3.3 nM tracer peptide). Plates were incubated at 4 °C for 1 hr, imaged using standard FP assay settings at 30 °C, and data analyzed as previously described [236].

Protein Expression and Purification.

Wildtype and Y537S 6×His-TEV-ERalpha (residues 300-550) in pET21a(+) was expressed in *E.coli* BL21(DE3) as described [237]. Cells were resuspended in a lysis buffer comprised of 20 mM HEPES pH 8.0, 500 mM NaCl, 40 mM imidazole pH 8.0, 5% glycerol, 15 mM betamercaptoethanol (BME), Sigma-FAST protease inhibitor cocktail, and lysed using sonication. Cellular debris was removed by centrifugation at 22,000 *g* for 30 minutes at 4 °C. The soluble fraction was placed onto a column with Ni-NTA resin that was pre-equilibrated with 20 mM HEPES pH 8.0, 500 mM NaCl, 40 mM imidazole pH 8.0, 15 mM BME, and 5% glycerol (wash buffer). The column was washed with 10 column volumes of wash buffer and protein was eluted with buffer plus 500 mM imidazole pH 8.0. The His-tag was removed by incubation with 1:15 w/w ratio of 6x-His-Tev protease:ERalpha LBD overnight in 4 L of wash buffer at 4 °C then passed over Ni-NTA

resin to capture the His-tag and TEV protease. Protein was concentrated to 5 mL using a spin concentrator at 4 °C then further purified on a Superdex 200 HiLoad 200 16/600 size exclusion column equilibrated with 20 mM pH 8.0, 150 mM NaCl, 15 mM BME, and 5% glycerol. A single peak was observed on the chromatogram and fractions corresponding to that peak for pooled and concentrated to 10 mg/mL. Aliquots were flash frozen and stored at -80°C.

Co-Crystallization and X-Ray Data Collection of Y537S ERalpha LBD in Complex with Estradiol and 6a.

Purified Y537S ERalpha LBD was incubated with 1 mM estradiol (E2) at 1.025 mM **6a** overnight at 4 °C then centrifuged at 20,000 *g* for 30 minutes at 4 °C to remove insoluble ligand. Co-crystals were generated using hanging drop vapor diffusion where 1 μL of protein complex was mixed with 1 μL well solution. Clear rectangular crystals emerged in 20 mM Tris pH 8.0, 15% PEG 3,350, 200 mM $MgCl_2$ after 1 week at room temperature. Crystals were cryo-protected in paratone-N. Diffraction data was collected at the Advanced Photon Source, Argonne National Laboratories, Argonne, Illinois, at the SBC 19-BM beamline (0.97 Å). Data were indexed, scaled and merged using HKL-3000 [238]. Molecular replacement was performed in Phenix using PDB: 6CBZ as the search model with ligand and peptide removed [239]. The model was refined using iterative rounds of Phenix Refine and manual inspection with Coot [240]. Ligand constraints for **6a** were generated using Elbow [241]. Poorly resolved atoms/residues were not included in the final model. The structure was deposited in the Protein Data Bank with the accession

code 6PIT. All x-ray crystal structure images were generated using Pymol. **Table 3.5** shows data collection and refinement statistics. Figure 3.11 B shows the electron density map for the adduct of **6a** that corresponds to protein chain A.

(Cyclopentadienyl)propanoic acid (**cpdp-COOH**):

Sodium cyclopentadienide (2.4 M solution in THF, 15 mL, 30 mmol, 1.2 eq) was added to 50 mL THF at -78 °C. Dropwise addition of ethyl 3-bromopropionate (3.2 mL, 25 mmol, 1 equiv) was carried out over ~15 minutes and the reaction was then stirred at -78 °C for 2 hours, allowed to warm to room temperature, and then stirred for an additional 2 hours. 3 mL water and 3 g silica gel were added to quench the reaction, stirring for 5 minutes. After filtering the reaction through a plug of silica gel with 30 mL DCM, the solvent was removed, yielding **cpdp-COOEt**, a yellow oil carried forward to the next reaction without further purification. A solution of **cpdp-COOEt** (~15 mmol) was dissolved in 10 mL EtOH and cooled to 0 °C. 10 mL of 4 M NaOH solution was added dropwise with stirring, and the reaction was carried out at 0 °C for 1 hour and then let. Upon full deprotection (as monitored by TLC), the reaction was poured into a separatory funnel and 2 M HCl was added until reaching pH 3.0. The aqueous solution was extracted with DCM (3 x 25 mL). Combined organic layers were washed with brine (2 x 20 mL), dried over Na_2SO_4, filtered, the solvent removed to yield **cpdp-COOH** as a light yellow solid stored at -20 °C to prevent dimerization.

2,5-Dioxopyrrolidin-1-yl(cyclopenta-1,3-dienyl)propanoate (**cpdp-NHS**):

Cpdp-COOH (1.25 g, 8.9 mmol, 1 equiv) was dissolved in a round bottom flask in 10 mL THF. To the flask was added *N*-hydroxysuccinimide (1.5 g, 13.0 mmol, 1.5 equiv), N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (2.1 g, 10.7 mmol, 1.25 equiv) and DCM (15 mL). The reaction was stirred for 1.5 hours. The resulting mixture was filtered through a silica plug with 30 mL DCM. The solution was evaporated and the resulting residue purified by flash column chromatography (Hexane:EtOAc, 2:1→1:1) to yield **cpdp-NHS** as an off-white powder. The product was observed to be stable for several months with proper storage at -20 °C. Rf (Hexane:EtOAc, 2:1): 0.3.

ANP-WT and DAC-ANP synthesis and oxidation.

SPPS, Diels-Alder functionalization, HPLC purification, and lyophilization were carried out as described above. The lyophilized, fully-reduced peptides were dissolved to ~0.1 mM in 100 mM ammonium acetate buffer, pH 8.5. The solutions were rotated at room temperature and samples analyzed at 0, 1, and 24 hours by LC-MS to monitor progress of disulfide formation.

CHAPTER 4: Targeting the SARS-CoV-2 nsp16/nsp10 methyltransferase complex with peptide inhibitors

4.1 Introduction

4.1.1 The COVID-19 pandemic and SARS-CoV-2 nsp16/nsp10 as a potential pan-viral drug target

COVID-19, a severe respiratory infection officially declared a global pandemic in March 2020, is caused by the novel severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) [242]. To date, there have been nearly 140 million documented infections worldwide, with a direct death toll of nearly 3 million, with likely millions more due to disruptions in access to healthcare and associated effects on human well-being and mental health (**Figure 4.1**) [243]. In what must only be described as an unprecedented feat carried out by the biopharmaceutical industry, healthcare workers, scientists, and regulators, multiple vaccines have been brought forward in record time to combat the spread of the virus [244-246]. The efficacy of these vaccines defy all preceding expectations at the onset of the pandemic, and has greatly contributed to easing the burden on our global society. The predominant mechanism of action of these vaccines is the triggering of an immune response in the patient against the Spike proteins found on

the surface of the virus [247, 248]. During infections, however, viruses have a short generation time over which many copies are produced. Some of these copies contain mutations that could potentially convey a selective advantage against host immunity, either derived from previous infections or vaccination. Indeed, several emerging SARS-CoV-2 variants, with mutations present in the Spike protein and higher rates of transmissibility, have been screened against immunized patient sera for neutralizing activity. The B.1.351 variant was shown to have significantly reduced neutralization relative to the progenitor strain and other variants [249, 250]. While these results are preliminary and will require real world validation within immunized populations, these will certainly not be the last SARS-CoV-2 variants, and this will certainly not be the last virus that threatens our globalized community.

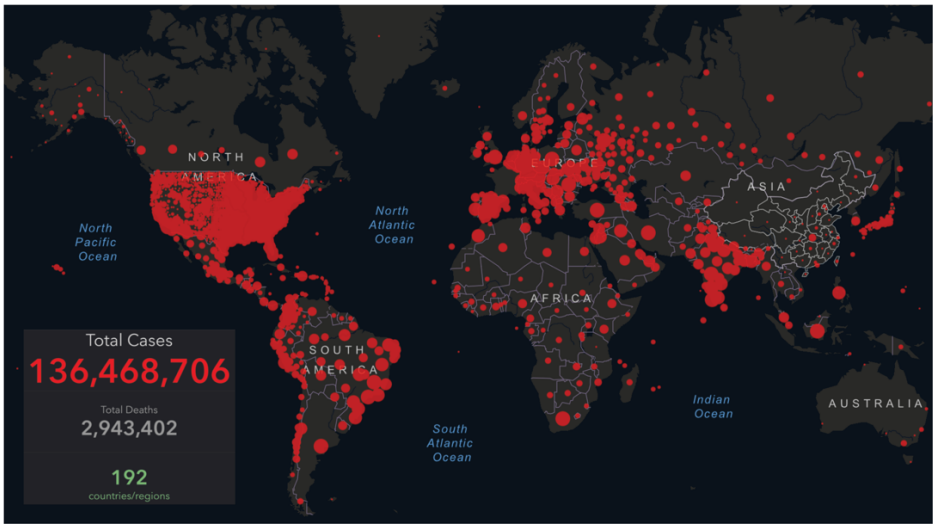


Figure 4.1: World map, case count, and death toll of the COVID-19 pandemic, as of April 12, 2021. Taken from Johns Hopkins Coronavirus Resource Center.

While standards of care for COVID19 infection have markedly improved over the course of the pandemic, few anti-viral treatment options have emerged as “magic bullets”. Antibody therapies have demonstrated promise as prophylactics and can reduce infection

severity [251], however face challenges in widespread distribution and administration. Further, antibodies, like the vaccines previously mentioned, reduce viral infectivity by targeting the Spike protein, which binds the cell surface angiotensin-converting enzyme 2 (ACE2) receptor to gain entry into host cells and is amenable to mutations that improve infectivity, transmissibility, and avoidance of immune detection [252]. Therefore, alternate systems within viral biology should be investigated for suitable therapeutic strategies. One such system is the RNA methyltransferase (MTase) system of SARS-CoV-2. It is in part composed of a non-structural protein (nsp) complex, nsp16/nsp10, with 2'-O-MTase activity (**Figure 4.2 A**). The nsp16 component, a 7-methylguanine-triphosphate-adenosine-specific, S-adenosyl-methionine (SAM)-dependent 2'-O-MTase, carries out the 2'-O-methylation. It is activated by a protein-protein interaction (PPI) with the small accessory protein nsp10, which is expressed at ~10-fold relative to nsp16 in infected cells, and contains two zinc finger domains (**Figure 4.2 B**) [253-256]. The nsp16/nsp10-mediated 2'-O-methylation of coronavirus RNA is essential for preventing recognition by host to evade immune responses that are triggered by viral mRNAs [253]. Normally, the human immune system detects free viral RNA, triggering the production of interferon and subsequent immune signaling, however these methylation events block this response [257, 258]. Inhibition of the 2'-O-MTase activity of SARS-CoV-2 presents a potential therapeutic strategy to both prevent infection via prophylaxis as well as reduce the severity of infection, by promoting viral clearance by the host immune system.

Given the criticality of nsp16/nsp10 complex activity for viral fitness and its bimolecular nature, mutations – and thus the potential for concomitant drug resistance –

are rare and often deleterious [259]. Further, many coronaviruses have a high degree of sequence homology for these two proteins, suggesting that inhibitors of nsp16/nsp10 complex formation in the context of SARS-CoV-2 may be applicable to other viruses (**Figure 4.3**) [260]. Notably, in the last 18 years, the world has experienced three major coronavirus outbreaks (SARS, Middle Eastern respiratory syndrome (MERS), and COVID-19), so a strategy targeting this highly-conserved viral feature would be an advantageous addition to the therapeutic arsenal.

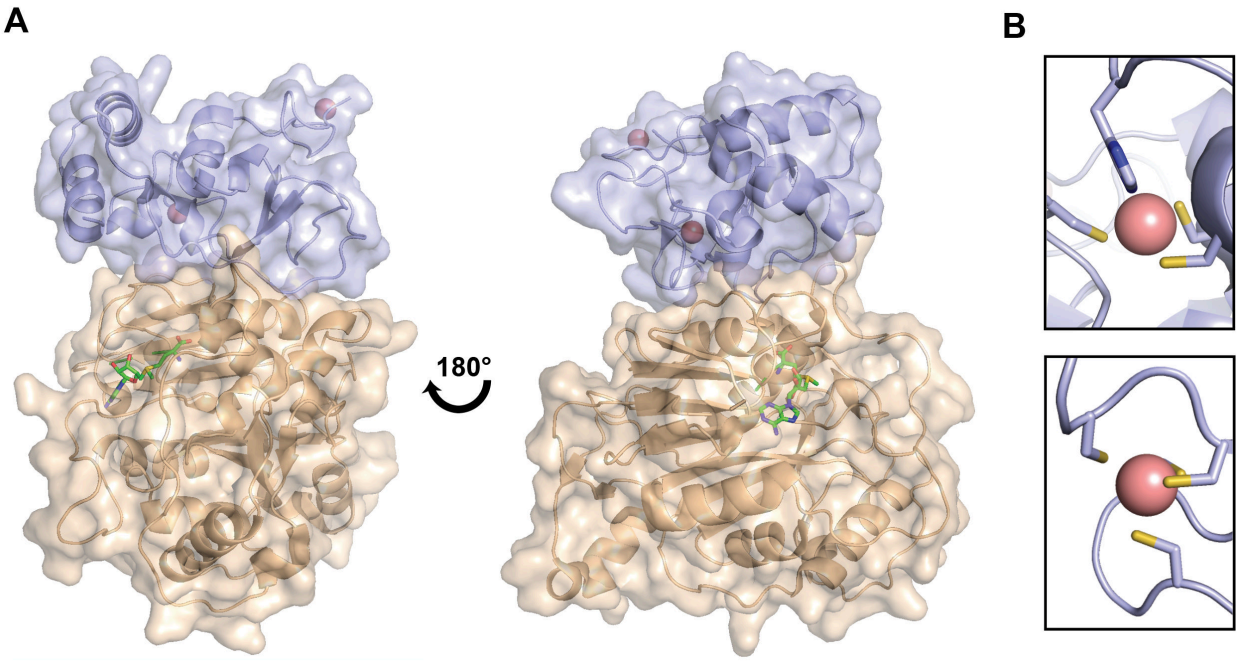
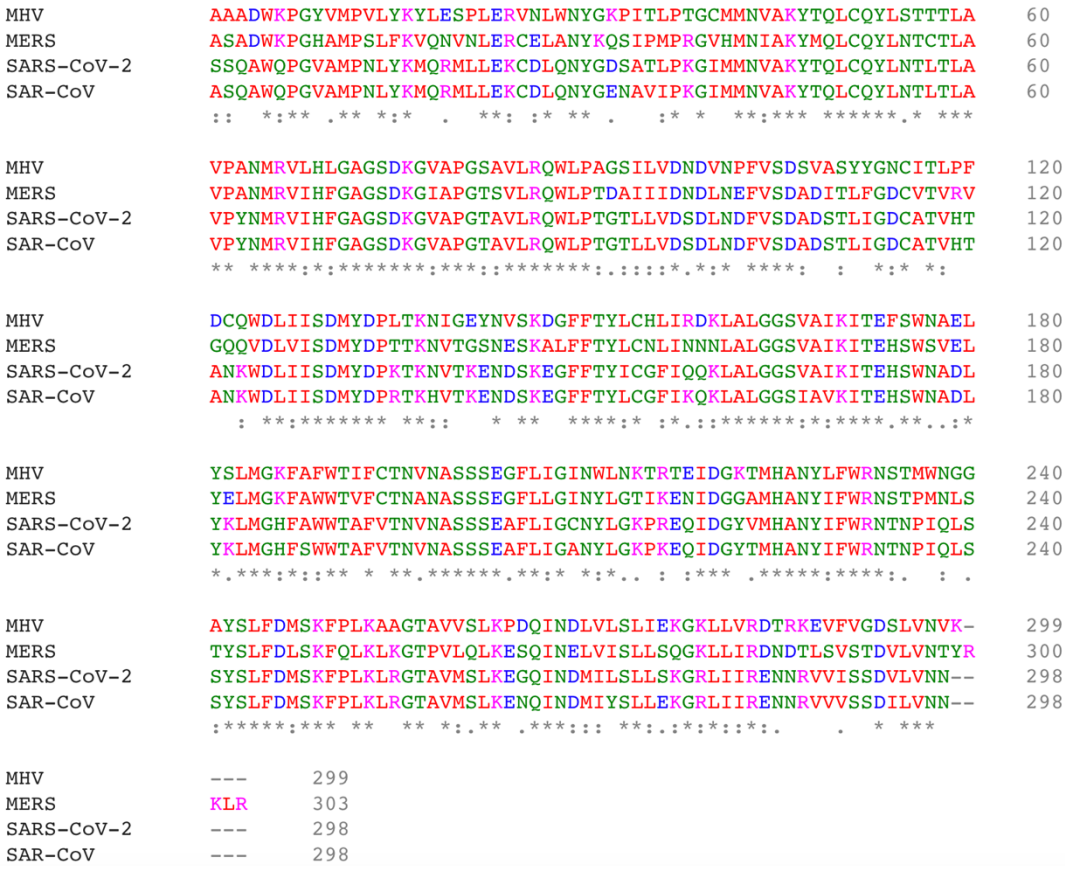


Figure 4.2: Structure of nsp16/nsp10 complex.
(A) Views of a crystal structure of nsp16 (wheat) bound to SAM (green sticks), in complex with nsp10 (light blue) with coordinating zinc atoms (rose spheres) (PDB: 3R24). **(B)** N-terminal zinc finger coordination complex (top) and C-terminal zinc finger coordination complex (bottom; zinc finger site distal to nsp16 in (**A**)).

NSP16:



NSP10:

Figure 4.3: Sequence alignments of nsp16 and nsp10 from select coronaviruses. Derived from MHV (accession number: AY910861), MERS (JX869059), SARS-CoV (AY278741), and SARS-CoV-2 (NC_045512). Aligned with CLUSTAL Omega. Red, hydrophobic; blue, acidic; magenta, basic; green, polar. ***** , fully conserved; **:** , strongly similar residues; **.** , weakly similar residues.

4.1.2 Targeting nsp16/nsp10 with peptide inhibitors

Previous work has validated the therapeutic potential of targeting the MTase complex nsp16/nsp10. In studies of SARS-CoV (the virus responsible for the 2003 SARS pandemic) and the related model coronavirus murine hepatitis virus (MHV), inhibition or inactivation of nsp16 was shown to be critical for viral infection [261, 262]. Notably, a 29 amino acid peptide probe, **P29**, mimicking part of the nsp16-binding interface of nsp10, was shown to inhibit 2'-O-MTase activity *in vitro* [261]. Further, a cell penetrating peptide (CPP) fusion, **TP29**, synthesized to enhance cell entry, was demonstrated to inhibit viral replication in cells as well as mouse models infected with MHV, with 100% of control mice dying by 7 days post-infection and a 100% survival rate among peptide-treated mice.

While these previous studies validate the potential of drugging the nsp16/nsp10 complex with peptide-based inhibitors, there is virtually no structure activity relationship (SAR) data for these compounds. We have thus devised a multi-pronged approach to test, characterize, and design improved nsp16 inhibitors. First, this involves the synthesis of a SARS-CoV-2-derived nsp10-mimetic peptide similar to the one previously reported, as well as complementary probes, for downstream analysis. Next, biophysical and biochemical screens were developed to study the structure and confirm target-binding of this peptide. Further, NMR and x-ray crystallography structural studies, as well as molecular modeling, will be implemented to study target engagement and compare peptide structure and binding to that of its parent nsp10 protein. Iterative synthesis and compound screening, informed by structural and modeling studies, were then carried out to develop improved peptide ligands. In addition, peptide variants, designed for enhanced

cell uptake and stability were synthesized for efficacy screening in biochemical and cell-based assays.

4.2 Results

4.2.1 Design, synthesis, and screening of an initial SARS-CoV-2 nsp16-targeting peptide inhibitor

We began work toward the development of peptide inhibitors targeting nsp16 by designing a 29 amino acid peptide, **COV21**. It comprises the sequence of the N-terminal zinc finger domain of the nsp10 protein (residues 4321-4349), seen in high resolution crystal structures to compose a large part of the binding interface with nsp16 (**Figure 4.4 A,B** and **Table 4.1**) [260]. Due to the presence of many Cys residues in the sequence, special care was taken during peptide cleavage, purification, analysis, and storage to ensure the peptide remained fully reduced prior to downstream characterization (see **Methods**). Preliminary stability and solubility studies confirmed the peptide was readily soluble in water and buffers used for various analyses, including LC-MS (**Figure 4.4 C**). Prolonged incubation in water resulted in the gradual oxidation and intramolecular cross-linking of the peptide, however brief exposure to 1 mM tris(2-carboxyethyl)phosphine (TCEP; a reductive additive present in the buffers used for downstream characterization) kept the peptide fully-reduced for days during storage at 4°C (data not shown). Circular dichroism (CD) spectroscopy, performed on apo **COV21** and following addition of zinc, demonstrated pronounced conformational changes evident in the resulting spectra of the zinc-bound peptide (**Figure 4.4 D**), similar to those seen in previous studies of short, zinc

finger peptide domains [263, 264]. NMR studies of the effects of zinc-binding on **COV21**, as well as co-crystallization studies with nsp16, are ongoing.

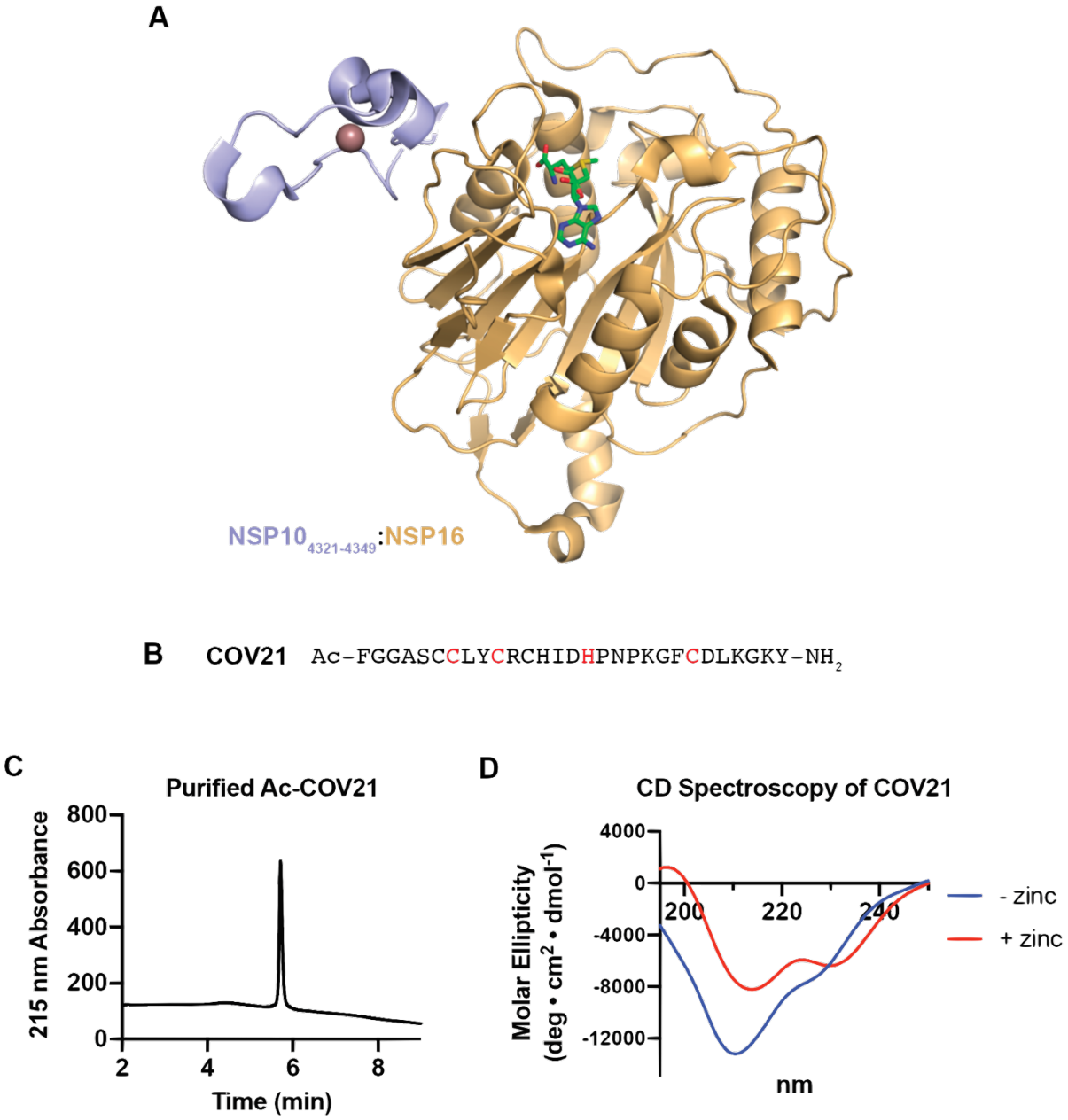


Figure 4.4: Design and preliminary characterization of COV21 peptide.
(A) Crystal structure of nsp16/nsp10, displaying nsp16 (wheat) with SAM bound (green sticks) and a portion of the nsp10 structure corresponding to the **COV21** peptide (blue) with zinc atom bound (rose sphere) (PDB 3R24). **(B)** Sequence of **COV21**, with putative zinc-coordinating residues in red. **(C)** HPLC trace of purified **COV21**. **(D)** CD spectra of **COV21** with or without addition of zinc.

Next, we sought to establish a rapid, sensitive, and high-throughput assay for assessing the nsp16-binding properties of **COV21** and later derivatives. Fortunately, our collaborators in the Satchell Lab at Northwestern University are presently the only lab in the world capable of expressing stable, recombinant nsp16 alone [260]. Typically, nsp16 is co-expressed with its stabilizing binding partner nsp10, limiting targeted study. Unfortunately, nsp16 alone is not stable at high concentrations and expression yields are relatively low. Therefore, we designed a 384-well plate format fluorescence polarization (FP) assay to limit required input protein while maximizing assay sensitivity.

To run direct binding assays against a dilution series of nsp16 to ascertain compound binding affinities, we synthesized a fluoresceinisothiocyanate (FITC) labeled **COV21** analog, **f-COV21**, separated by an aminohexanoic acid (ahx) linker (**Table 4.1**). We first compared **f-COV21**-nsp16 binding with or without the presence of zinc (with the source being $ZnSO_4$ throughout). FP results indicated similar binding profiles, both conditions displaying a ~150 nM K_D for **f-COV21**-nsp16 binding (**Figure 4.5 A**). Given the low concentration of fluorescent tracer used in FP assays (2.5 nM herein), and the observation that zinc is present in deionized water at ~10 nM [265], we reasoned that there was sufficient zinc in the FP buffer to coordinate the zinc finger domain. Future analysis by FP included addition of 10 nM zinc (unless otherwise noted) to ensure batch and run consistency.

We next attempted to confirm the binding specificity of **f-COV21** at the same predicted interface of nsp16/nsp10 binding. A nsp10 dose-dependent competitive format

FP assay was performed against 2.5 nM **f-COV21** pre-incubated with a sub-saturating 0.8 μM nsp16. Initial results displayed inconsistent **f-COV21** binding profiles, in which polarization rapidly fell off as nsp10 concentration surpassed the concentration of zinc present in the assay (data not shown). This suggested that nsp10, with its two zinc fingers not fully coordinated with two zinc atoms, competed with **f-COV21** zinc-binding, abrogating its structure and target binding. To mitigate this assay interference, we added zinc to 100 μM in the competitive assay. While measured binding was now found to be consistent, incomplete competition was observed at maximum nsp10 concentration, with a consistent low-level increase in polarization observed at intermediate concentrations of nsp10 (**Figure 4.5 B**). Given previous evidence of non-covalent dimerization and potentially higher-order oligomerization of nsp10 protein from SARS-CoV and SARS-CoV-2 in X-ray crystal structures and solution studies [266-268], we screened **f-COV21** for direct binding to nsp10, with 20 or 100 μM zinc. Consistent with the **f-COV21**-nsp10-dimerization hypothesis, direct **f-COV21**-nsp10 binding was observed in both conditions, with an apparent K_D of 204 nM (**Figure 4.5 C**). Further, in the 20 μM zinc condition, a decrease in binding was observed with increasing nsp10 concentration, consistent with the apparent nsp10 competition for zinc-binding observed in the preliminary competitive FP results.

Next, to asses a functional readout of peptide activity, our collaborators screened the **COV21** peptide in an *in vitro* MTase activity assay. Preliminary treatment of 50 μM peptide (with or without zinc) against the nsp16/nsp10 complex (0.2 / 1.6 μM) showed 75% and 25% inhibition with and without zinc, respectively (**Figure 4.5 D**). Taken

together, these data establish **COV21** as a suitable ligand and starting point for further development of nsp16-targeting inhibitors. Additionally, evidence suggests the potential for dual-targeting of nsp10. Isothermal calorimetry (ITC) studies are ongoing to confirm these results and arrive at more precise values for **COV21** binding properties.

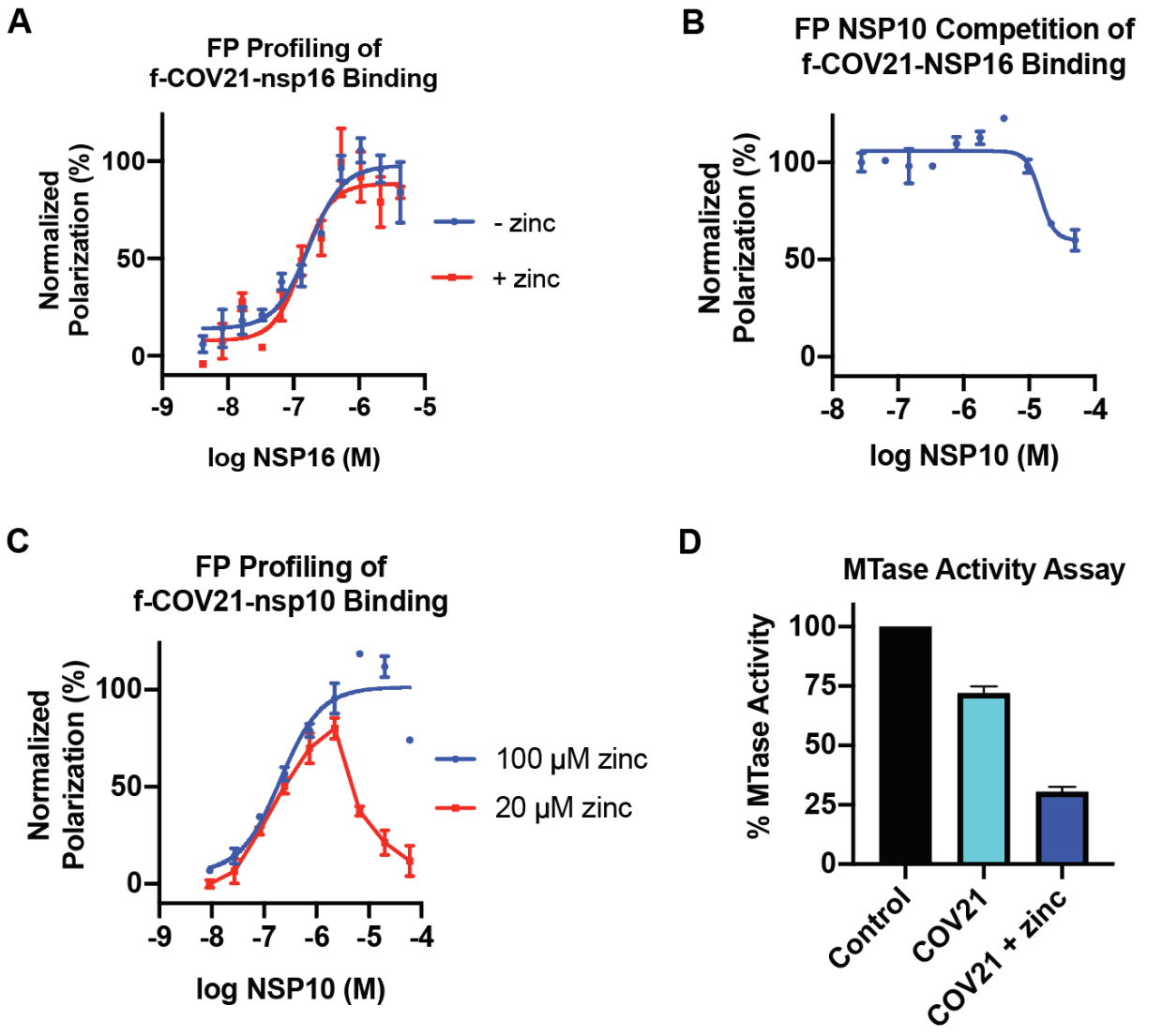


Figure 4.5: Fluorescence polarization and MTase inhibition data for peptides with nsp16/nsp10 complex members.
(A) FP results for nsp16 dose-dependent **f-COV21** binding. **(B)** Competitive FP results showing incomplete competition by nsp10 of the **f-COV21**-nsp16 complex, suggestive of multiple potential binding events. **(C)** FP results for nsp10 dose-dependent **f-COV21** binding. **(D)** Initial results for **COV21** inhibition of *in vitro* nsp16/nsp10 MTase activity.

4.2.2 Improved inhibitor design informed by structure and molecular modeling

We next sought to modify the initial **COV21** peptide structure to improve binding as informed by structural and modeling data. Attempts are ongoing to co-crystallize **COV21** with nsp16 or to soak it in to existing nsp16 crystals. These structures would greatly inform design as we do not yet know if the peptide occupies the same nsp16-interaction surface as its parent protein nsp10. Further, we do not know if **COV21** has a comparable structure or nsp16-binding pose as the corresponding domain of nsp10. Preliminary molecular dynamics simulations suggest that **COV21** stably interacts with nsp16 in an analogous fashion as nsp10 (**Figure 4.6 A**, red structures). However, as the trajectory progresses, the peptide appears to take on two other distinct, stable binding poses with different interfaces (**Figure 4.6 A**, yellow and blue structures). At each stage, the simulated nsp16 protein undergoes only minimal fluctuations, most notably in the C-terminal region of the protein (**Figure 4.6 A**, bottom right corner). While the relevance and insights of these preliminary simulations are unclear, they suggest the potential for **COV21** to engage nsp16 in multiple distinct binding poses. Further molecular dynamics simulations are ongoing, including some with subsequently designed mutants, described below.

Additional interrogation of existing nsp16/nsp10 structural data revealed a potential handle for augmenting **COV21** to improve binding. A hydrophobic cleft (**Figure 4.6 B**, grey spheres, indicating hydrophobic residues), proximal to the putative binding interface of the N-terminus of **COV21**, is typically engaged by a complementary hydrophobic coil of nsp10 in many crystal structures. In fact, we had attempted to engage

this cleft with a hexamer peptide derived from the nsp10 sequence, but did not observe binding by FP (data not shown). We wondered if the ahx linker or FITC fluorophore of the **f-COV21** tracer engaged this cleft during FP binding assays. This may also explain the large excess of unlabeled **COV21** peptide required for inhibition in the *in vitro* MTase activity assay. Regardless, we posited that the addition of a hydrophobic residue at an N-terminal position of **COV21** may enhance peptide binding affinity (**Figure 4.6 B**, red spheres indicating positions of hydrophobic mutations).

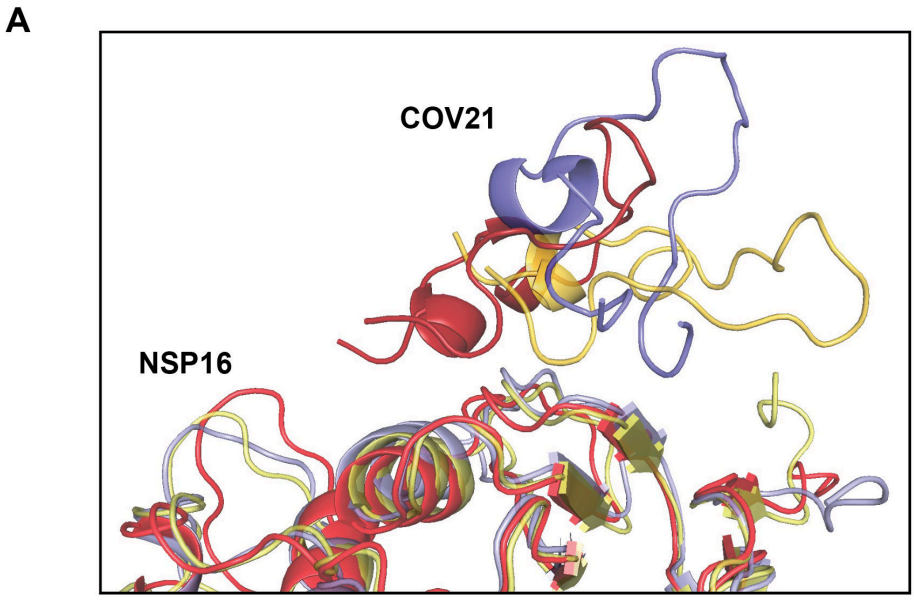


Figure 4.6: Insights from molecular modeling and structural studies.
(A) Preliminary results from molecular dynamics simulations of **COV21** and nsp16. Red, early nsp10-like **COV21** pose; yellow, trajectory taken from middle of simulation; blue, trajectory taken from late in the simulation.

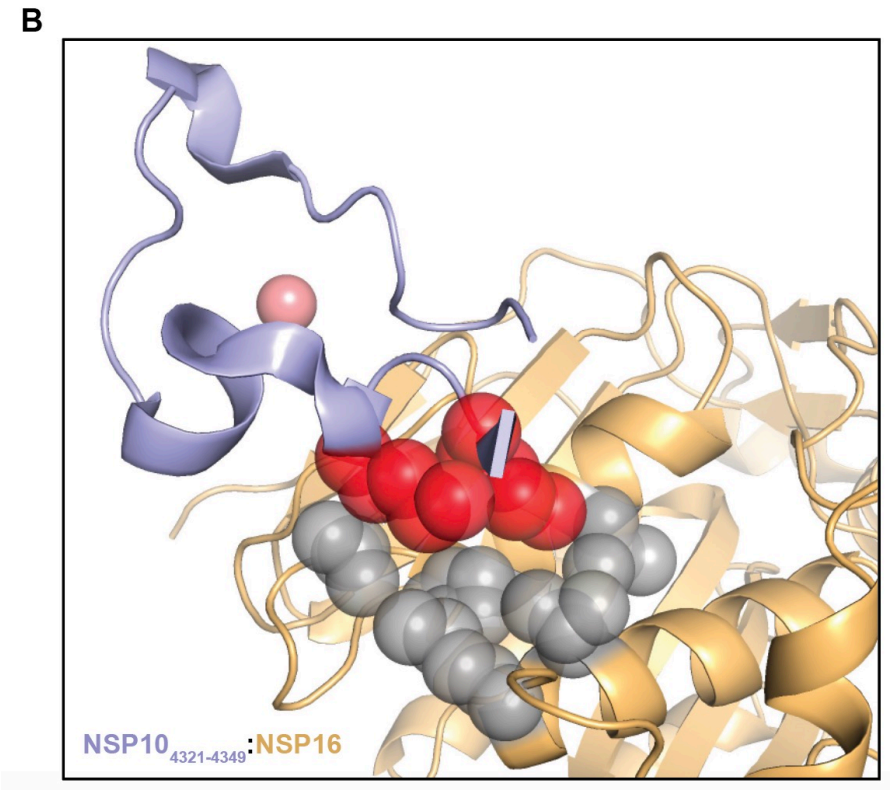


Figure 4.6 (continued):
(B) Detail of crystal structure of nsp10/nsp16 complex (blue/wheat) (PDB 3R24). Grey spheres depicting residues of hydrophobic cleft. Red spheres depicting potential hydrophobic mutation positions.

A small panel of **COV21** mutant tracer peptides was synthesized. These contained a mutated hydrophobic amino acid at either the second or fifth residue (**Figure 4.7 A**). A previously reported double-Ala mutant (**COVdm**), in which two key binding residues at the C-terminus were mutated to alanine, was also synthesized [261]. Following purification and analysis, the mutants were screened against nsp16 in the FP binding assay. Every hydrophobic mutant displayed increased affinity for nsp16 relative to **f-COV21**, with the highest affinity mutant, **f-COV24**, displaying an apparent K_D of 73 nM (**Figure 4.7 B**). Intriguingly, each hydrophobic mutant also displayed higher levels of relative polarization compared to **f-COV21**, while the double-Ala mutant **f-COVdm**

displayed reduced binding and levels of maximum polarization (**Figure 4.7 B**). This suggests that even the conservative introduction of additional hydrophobic bulk at the N-terminus may bias the peptide toward a binding interaction that reduces degrees of freedom of the adjacent ahx-linked FITC moiety. Given the targeting of the proximal nsp16 hydrophobic cleft by these mutations, it is conceivable that the addition of these interactions locks the ahx linker and/or tracer into this cleft. Such an effect could result in the observed increase in measured polarization, beyond that of the protein-bound tracer alone, as well as potentially increasing the quantum yield of the fluorophore [269, 270]. These results strengthen the notion that **COV21** and derivative peptides may take on differential binding poses with nsp16 relative to that seen in existing nsp16/nsp10 complex crystal structures. This is perhaps unsurprising, given the absence of the majority of nsp10, which no doubt influences the structure of the N-terminal zinc finger domain. Taken together, this first iteration of mutant **COV2X** peptides presents one viable design strategy for improving peptide engagement of nsp16, and contributes more generally to our understanding of the nature of this binding interaction. Work toward probing mutant binding of nsp10, effects on MTase activity, and further study of their binding interactions by ITC are ongoing.

A

f-COV21	FITC-ahx-FGGASCCLYCRCHIDHPNPKGFCDLKGKY-NH_2
f-COVdm	FITC-ahx-FGGASCCLYCRCHIDHPNPKGFCDLAGKA-NH_2
f-COV22	FITC-ahx-FGGAN_LCCLYCRCHIDHPNPKGFCDLKGKY-NH_2
f-COV23	FITC-ahx-FGGAFCCLYCRCHIDHPNPKGFCDLKGKY-NH_2
f-COV24	FITC-ahx-FGGAWCCLYCRCHIDHPNPKGFCDLKGKY-NH_2
f-COV25	FITC-ahx-FN_LGASCCLYCRCHIDHPNPKGFCDLKGKY-NH_2

B

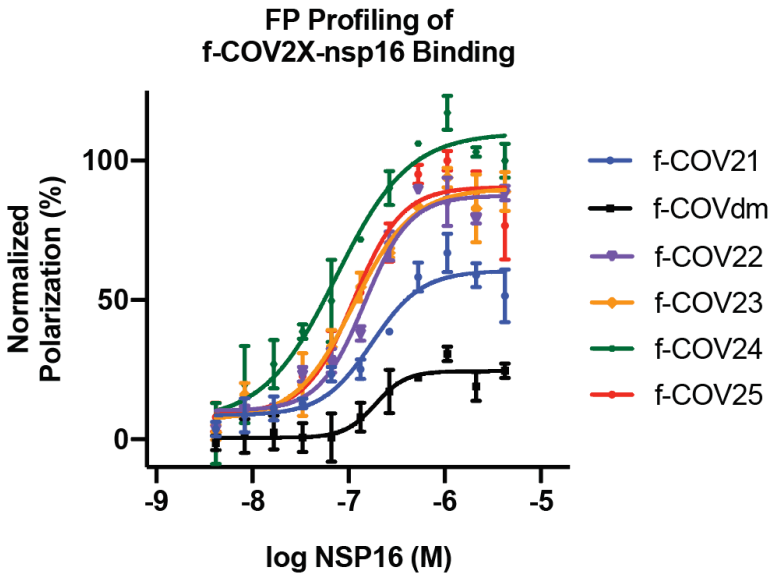


Figure 4.7: Fluorescence polarization (FP) of f-COV2X mutants and nsp16.
(A) FP results for nsp16 dose-dependent **f-COV21** binding. Red, putative zinc-coordinating residues; blue, putative gain-of-function mutations; grey, putative loss-of-function mutations; ahx, aminohexanoic acid; N_L, norleucine. **(B)** FP results for nsp16 dose-dependent binding of **f-COV21**, **f-COVdm**, and **f-COV2X** mutants. Dose-response curves fit through the application of a sigmoidal curve fit using Prism 8 software.

4.2.3 Current work and future directions

While *in vitro* study of these nsp16/nsp10 complex-targeting peptides is still on-going, a first round of probes for cell-based assays have been synthesized. These are composed of a fusion of **COV21** linked by ahx at the N-terminus to either a linear or

stapled CPP, penetratin (**PEN**) or stabilized alpha helix of penetratin (**SAH-PEN**), as well as the corresponding CPPs as controls (**Table 4.1**). These CPPs have been shown to have increased cell uptake relative to **TAT**, the CPP incorporated into the **TP29** peptide, used in the original cell and mouse studies, to target SARS-CoV nsp16/nsp10 [144, 261, 271]. While sufficient amounts of the fusion peptides (**PEN-COV21** and **SAH-PEN-COV21**) were synthesized for ongoing cell-based infectivity assays, additional work is going toward the development of a convergent synthetic route to improve yield and decrease synthesis time (*e.g.,* automated synthesis of **COV21** presently takes ~40 hours to complete, with an additional 16-36 hours required for CPP incorporation).

Alongside ongoing work toward the design and synthesis of further improved mutant peptides, a near term goal is the synthesis of stabilized variants of **COV21**. Several chemical crosslinking strategies are being pursued, including Diels-Alder chemistry (as described in **Chapter 3**), thiol alkylation, and the use of heterobifunctional crosslinkers. Two sites of stabilization are being investigated: one distal to the binding interface, and another mediated by cyclization at the N- and C-termini. This latter strategy is informed by structural and modeling studies in which the termini are invariably seen in close proximity, often induced by intramolecular hydrogen-bonding. Further, an introduced long hydrophobic linker between the termini may be able to engage the nsp16 hydrophobic cleft, likely central to the observed improvements in **COV2X** mutant-binding affinities. Beyond stabilizing the peptide for reduced proteolysis, such a crosslink may obviate the need of zinc-coordination for peptide activity. This would greatly ease the requirements of peptide synthesis, preparation, and storage. In addition, to the best of

our knowledge, this would represent the first instance of a stabilized zinc finger domain without requiring the presence of zinc. Finally, this chemically-stabilized scaffold would likely be more amenable to the incorporation of mutations that could improve binding, pharmacologic properties, and potentially allow for further diversification.

4.3. Discussion

The COVID-19 pandemic is a still on-going global crisis that has massively impacted the way of life for all of humanity. While incredible advancements have been made in preventative and therapeutic intervention for this virus, they are not without their limitations, and this will certainly not be the last coronavirus that threatens human health. Here we have proposed the use of peptide inhibitors to block the activity of a PPI-mediated driver of SARS-CoV-2 infectivity. The highly conserved nature of the nsp16/nsp10 protein complex members suggests targeting this interaction would have a durable effect, applicable to other current and future coronaviruses. Our preliminary data indicate the **COV21** peptide and its derivatives engage nsp16, and potentially nsp10, and can effectively inhibit MTase complex activity. Current and future work is going toward further improving peptide inhibitor design, enhancing cell uptake, stabilization strategies, and confirming compound efficacy in cell-based assays and eventually in *in vivo* models. We believe that suitable peptide inhibitors will offer complementary strategies for studying viral biology and preventing and reducing the severity of viral infections for this current outbreak and future ones.

4.4 Tables

Compound	Sequence	m/z (calc./ obs.)
COV21	Ac–FGGASCCLYCRCHIDHPNPKGFCDLKGKY	1091.8/ 1092.2
f-COV21	FITC–ahx–FGGASCCLYCRCHIDHPNPKGFCDLKGKY	1245.2/ 1246.1
f-COVdm	FITC–ahx–FGGASCCLYCRCHIDHPNPKGFCDLAGKA	1195.5/ 1196.4
f-COV22	FITC–ahx–FGGAN_LCCLYCRCHIDHPNPKGFCDLKGKY	1253.9/ 1254.8
f-COV23	FITC–ahx–FGGAFCCLYCRCHIDHPNPKGFCDLKGKY	1265.2/ 1265.9
f-COV24	FITC–ahx–FGGAWCCLYCRCHIDHPNPKGFCDLKGKY	1278.2/ 1278.8
f-COV25	FITC–ahx–FN_LGASCCLYCRCHIDHPNPKGFCDLKGKY	1263.9/ 1264.5
PEN-COV21	Ac–RQIKIWFQNRRN$_L$KWKK–ahx–**COV21**	1399.7/ 1400.5
SAHPEN-COV21	Ac–RQIKIWFQ*RRN$_L$*WKK–ahx–**COV21**	1401.7/ 1402.4
PEN	Ac–RQIKIWFQNRRN$_L$KWKK	757.1/ 757.4
SAHPEN	Ac–RQIKIWFQ*RRN$_L$*WKK	759.8/ 760.1

Table 4.1: nsp10-mimetic compounds synthesized for this study, their sequences, and MS results.
Red, putative zinc-coordinating residues; blue, putative gain-of-function mutations; grey, putative loss-of-function mutations; FITC, fluoresceinisothiocyanate; ahx, aminohexanoic acid; N_L, norleucine; *, sites of S_5 stapling residues.

4.5 Methods

Peptide synthesis.

Unmodified and hydrocarbon stapled peptides were synthesized by Fmoc-based solid phase peptide synthesis and purified by reverse-phase HPLC with a C18 column

(Agilent, Palo Alto, CA) as previously reported, with modifications [138]. The peptide cleavage cocktail contained trifluoroacetic acid (TFA), triisopropylsilane (TIS), water, and ethanedithiol (EDT) in a ratio of 91:3:3:3 to ensure Cys remained reduced. Peptide cleavage was carried out for at least 3.5 hours at room temperature. Purified peptide fractions were concentrated *in vacuo*, re-dissolved in 0.1% TFA/H_2O, quantified by A_{280}, aliquoted, lyophilized, and stored at -20 °C.

Circular dichroism spectroscopy.

Circular dichroism spectroscopy experiments were performed on a Jasco J-170 using a quartz cuvette (path length: 0.1 cm). Peptides were dissolved to 50 µM in 10 mM tris(hydroxymethyl)aminomethane HCl (Tris-HCl) buffer solution, pH 8.5, with or without the addition of 150 μM $ZnSO_4$, and CD measurements were recorded at one nm increments between 190 and 260 nm, at room temperature.

Fluorescence polarization experiments.

FP experiments were performed in black 384 well plates (Corning) on a Synergy Neo plate reader (BioTek). For direct binding assays, a 2-fold dilution series of protein was plated in 10 μL FP buffer (10 mM Tris-HCl, pH 8.5, 150 mM NaCl, 5% glycerol, 1 mM TCEP, 1 mM $MgCl_2$), and further diluted by the addition to each well of 5 μL FP buffer containing 7.5 nM tracer peptide and 30 nM $ZnSO_4$ (unless otherwise noted). Final assay volume: 15 μL; final concentrations: 2.5 nM tracer, 10 nM zinc). For competitive binding assays, a 2-fold dilution series of competing protein (*i.e.* nsp10) was plated in 10 μL FP

buffer, and further diluted by the addition to each well of 5 μL FP buffer containing 7.5 nM tracer peptide, 2.4 μM nsp16, and 30 nM $ZnSO_4$. Final assay volume: 15 μL; final concentrations: 2.5 nM tracer, 0.8 μM nsp16, 10 nM zinc). Plates were incubated at room temperature for 1.5 hours, imaged using standard FP assay settings at 25°C, and data analyzed using Prism 8 graphing software.

CHAPTER 5: Development of cell-penetration assays for assessing cytosolic access and stability of synthetic biologics

5.1. Introduction

5.1.1 Cell penetrating peptide design and related considerations

Given the primary role proteins serve in biological processes, and the frequency with which protein-protein interaction (PPI) deregulation contributes to disease, inhibiting these targets is a central challenge in the fields of chemical biology and drug discovery [13, 35, 151, 213]. Several factors make PPIs especially recalcitrant targets. Firstly, the majority of disease-associated PPI targets occur intracellularly, excluding otherwise well-suited biologics such as antibodies from the list of potential targeting modalities. Secondly, small molecules, the consummate modality with good potential for cell-permeability, are not well-suited for targeting the large interfaces of PPIs [33]. Moreover, many PPIs are multicomponent processes and/or have high binding affinities, necessitating highly potent and specific inhibitors for effective targeting.

In recent years, synthetic biologics have been demonstrated to have a balance of favorable properties of both small molecules and biologics [11, 129]. This has resulted in increased study and advancement of the field. A key property is their potential to

penetrate cells to engage otherwise challenging intracellular targets. This phenomenon has been widely studied among a growing class of cell penetrating peptides (CPPs), with goals of improving delivery of diverse cargos into cells, as well as determining the rules that govern this behavior [87, 145]. CPPs are sequences that are found in nature or synthetically designed to have increased cell uptake. While the mechanisms are still not fully understood, consensus is emerging that cellular peptide uptake is mediated through any one of a number of process [87]. Some peptides passively diffuse across the cell membrane, analogous to small molecule. Like most cell-permeable small molecule drugs, these peptides may comply with Lipinski's Rule of Five, but not always [18]. N-alpha methylation of backbone amides has been used by nature and peptide chemists to reduce hydrogen bond donor count in CPPs. As seen with the orally bioavailable drug cyclosporin A, some non-Rule of Five compliant peptides can take on conformations that promote intramolecular hydrogen bonding to reduce polar surface area [77, 86]. Alternatively, in the same way as some small molecules, a subset of CPPs bind to proteins on the cell surface that actively transport them into cells.

Typically, however, when one encounters CPPs, the canonical description is that of a short peptide with a highly basic sequence. It then gains entry into cells by first associating with the outer surface of the phospholipid membrane, eventually triggering endocytosis, and then undergoing endosomal escape into the cytosol for downstream effects. While the precise mechanisms of this pathway of entry are not fully understood, it is an area of intense research interest [87, 144, 173]. Derived from the transactivator of transcription protein found in HIV, the TAT peptide was one of the first reported CPPs

that catalyzed extensive study in this area, with subsequent reports of ~2000 CPPs so far [272-274]. While many studies use TAT as a conjugate for transporting cargo into cells, early studies of TAT and its parent protein often used a “scrape-loading” method in which cells were mechanically removed from their dish, transiently disrupting cell membranes, to ensure “quantitative and reproducible” transactivation activity [275, 276]. Similarly, cell density, culturing conditions, and growth time dependent effects have been noted to have major impacts in protein transduction studies [277]. While TAT is certainly cell-penetrant, it is a linear, largely unstructured peptide with a highly basic and proteolytically susceptible sequence [278]. It is not nearly as robust a CPP compared to some linear peptides such as Antennapedia-derived penetratin or synthetically optimized cyclic CPPs, which generally demonstrate higher levels of cell penetrance relative to TAT [144, 271, 279]. Such findings have moved the field away from “first generation” linear wildtype CPPs, instead favoring cyclic, side chain or backbone modified CPPs for enhanced penetrance and stability [87]. While this is an impactful trend, indisputable characterization of cell penetrance and intracellular stability of CPPs is challenging to establish, hampering the development and selection of the most optimal designs.

5.1.2 Understanding and measuring cell penetrance: The state of the art

Many intracellularly-targeted synthetic biologic peptide probes incorporate design principles of CPPs, have their structure and target-binding properties extensively studied *in vitro*, and are then often screened in cell culture for cell penetrance and specific target-dependent effects. While well-designed phenotypic screens will be context-dependent and vary from target to target, assessment of cell penetrance is somewhat more

standardized. Typically, fluorophore-labeled analogs of the investigated synthetic biologics are synthesized, applied to cells, and then imaged by fluorescence microscopy. There are many ways these experiments can be run, with corresponding variations in conditions that may potentially impact results: cell growth state and time, presence of serum in media, or post-treatment washing steps. Therefore, assessment of cell penetration is often based on largely qualitative differences in fluorescent intensities and possibly localization. Fluorescence correlation microscopy and other techniques, however, can improve limits of detection, quantitation, and determination of localization [173]. Additionally, in some carefully designed and performed studies, the endosomal uptake mechanism can be inferred: intracellular fluorescent puncta (*i.e.* endosomes) are observed at early time points and gradually fade as fluorescent signal diffuses throughout the cytosol (*i.e.* putative endosomal escape by the peptide) [280, 281]. In addition, significant effort has gone toward increasing the robustness and throughput of these fluorescence readouts, allowing for broader screening of compounds and conditions affecting penetrance [144, 279].

While this cell penetration assay modality has become a (perhaps not gold) standard across many fields, it still leaves room for improvement. It is often unclear if the fluorophore-conjugated peptide is actually in the cytosol, adhering to the cell membrane, or localized to other compartments. Also: are we observing the full-length fluorophore-conjugate evenly distributed throughout the cytosol, the fluorophore alone, or truncations or other alterations to the conjugate? This is a critical question, as a truncated or modified peptide would likely have reduced target affinity and specificity. From a discovery

standpoint, if a still-potent peptide has been proteolyzed or modified during cell entry, this could help inform potential future designs.

A variety of alternative methods have been devised to address these shortcomings. Some use reporter molecule-cargo conjugates that selectively readout upon engagement with an intracellular sensor protein, to unequivocally determine intracellular or subcellular localization [282, 283]. These typically do not report on the stability or identity of the cargo following cytosolic access. Others have used mass spectrometry (MS) based approaches to determine the status of CPPs or other cargos following cell treatment. In some cases, a cargo enrichment strategy is used, increasing sensitivity, however this is not able to directly report on the localization of cargo [284]. In other cases, subcellular fractionation is used to directly assess specific compartment localization [102]. These approaches, however, often only measure for the presence of the full-length peptide or cargo, and don't capture modified forms. Further, they require sensitive and time-consuming processing, which limits throughput. A general criticism of many of the above methods is the requirement for relatively high treatment concentrations to observe evidence of uptake, making it challenging to assess the lower limits of penetration and running the risk of disrupting cell membranes (an established off-target feature of some CPPs, as well as amphipathic peptides more generally).

5.1.3 A multi-format synthetic biologic uptake and stability assay

Given the critical nature of assessing cellular uptake and stability for determining the efficacy of synthetic biologics, we sought to develop an alternative, multi-format assay utilizing the widely studied O^6-benzylguanine (BG), SNAP-tag (SNAP) chemically induced

dimerization (CID) system [285]. An engineered O^6-alkylguanine-DNA alkyltransferase, SNAP selectively reacts with BG derivatives, allowing for precise labeling, proximity induction, and other applications for SNAP-fusion proteins in live cells and organisms [286-288]. To apply this enzymatic labeling strategy, we first functionalized a panel of CPP/synthetic biologics with BG and confirmed their reactivity with recombinant SNAP protein. We then designed various FLAG-tagged SNAP plasmids (for cytosolic or other subcellular-compartment localization). These were then stably-expressed in mammalian cells. With the entire system thus in place, SNAP-FLAG cells can then be treated with BG-peptides; upon reaching the specific intracellular compartment occupied by SNAP, the protein will then covalently label itself with the peptide. This labeling event can then be confirmed and quantified by either gel, western blot, or MS-based readouts (**Figure 5.1**). Notably, readout by western blot or MS would allow for qualitative or quantitative reporting of the BG-peptide probe's stability, as well as for the potential to identify intracellular degradation or modification products.

In addition to screening for the cell penetration of the initial BG-peptide panel, we observed that this system could be implemented in a reverse format to monitor for membrane disruption by unlabeled peptides. This could serve as a complement to enzymatic assays for membrane disruption, such as for the presence of lactate dehydrogenase (LDH) in cell media (*i.e.* the LDH assay). We synthesized a BG- and fluorescein isothiocyanate (FITC)-functionalized non-cell permeable peptide, composed of the charged residues Glu-Glu-Lys-Glu (**EEKE**). Only upon membrane disruption by co-treatment of an unlabeled peptide can the **EEKE** probe gain intracellular access to

covalently label SNAP protein for downstream detection (**Figure 5.2**). Taken together, this BG-SNAP system offers a suite of applications for improved study of synthetic biologics and related molecular probes.

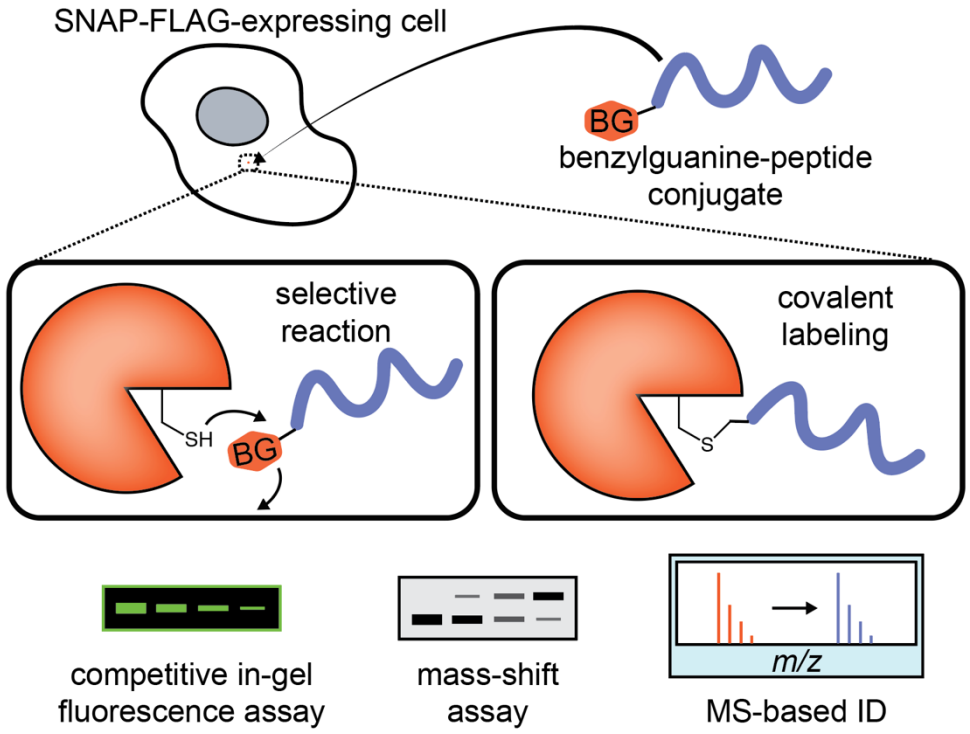


Figure 5.1: Schematic depictions of the BG-SNAP cell penetration assay and potential readouts.

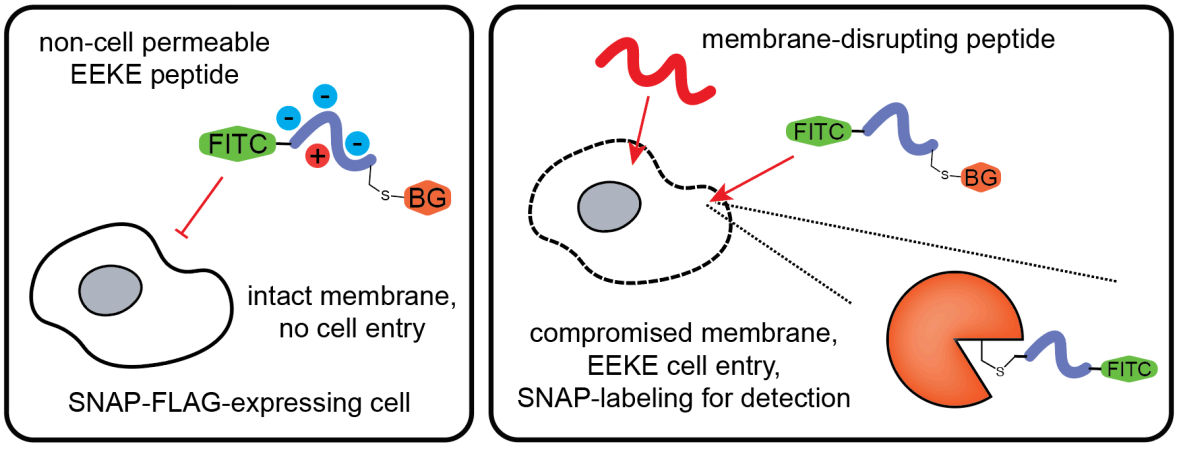


Figure 5.2: Schematic depictions of the BG-SNAP membrane disruption assay.

5.2 Results

5.2.1 Design and synthesis of BG-peptides and related probes

We selected the canonical CPP TAT and two well-studied stapled peptide inhibitors for our initial panel of BG-peptides. These inhibitors were SAHM1 and StAx35R, highly stabilized and cell-penetrant inhibitors of the NOTCH transcription factor complex and the WNT/beta-catenin signaling pathway, respectively ([133, 156]). A panel of related unlabeled peptides as well as the **EEKE** peptide were also synthesized (**Table 5.1**). We chose an in-solution thiol-maleimide bioconjugation strategy for the BG-functionalization of these peptides. This strategy necessitated the mutation of an internal **SAHM1** Cys residue to Ser. Following standard SPPS and RCM of the stapled variants, an acetylated Cys and beta-Ala were installed on the N-termini of all future BG-conjugated peptides prior to cleavage and HPLC-purification. In the case of the **EEKE** peptide, a C-terminal Lys(Mmt) residue was selectively deprotected and then acylated with FITC prior to purification. Unlabeled peptides were synthesized and acetylated prior to purification. Purified free thiol containing peptides were conjugated in phosphate buffer solution to 4-maleimidobutyrl benzylguanine (**BG-maleimide**) and then subsequently HPLC purified (**Figure 5.3 A,B**).

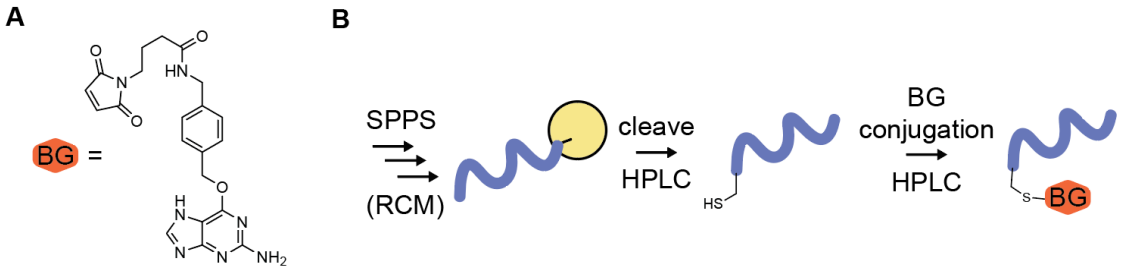


Figure 5.3: Structure of BG-maleimide linker (A), and schematic of synthesis and conjugation strategy for BG-peptides (B).

5.2.2 BG-SNAP covalent labeling system *in vitro* validation

To confirm that the BG-functionalized peptides could serve as viable substrates for SNAP protein, we first tested *in vitro* labeling of purified, recombinant SNAP protein and **BG-TAT** peptide (**Figure 5.4 A**). SNAP-peptide labeling was apparent by a clear ~2 kDa mass shift increase, indicating covalent labeling of SNAP with TAT peptide. Encouraged by this, we generated a HeLa cell line that stably-expresses a FLAG-labeled, cytosolically-localized SNAP protein for detection by western blot. To confirm the mammalian-expressed SNAP-FLAG protein would similarly react with our synthesized BG-peptides, we lysed cells in the presence of protease inhibitors, then treated lysate with a dose range of each peptide at 37°C for 30 minutes. As anticipated, concentration dependent mass shifts, corresponding to SNAP-peptide covalent dimers, were observed by anti-FLAG western blot (**Figure 5.4 B**). **BG-SAHM1** and **BG-StAx35R** fully labeled SNAP protein in lysate by 1 μM (**Figure 5.4 B**, condition **C**), as indicated by a single, higher molecular weight band. Intriguingly, multiple bands are observed with increasing concentrations of **BG-TAT** peptide, which presumably correspond to differentially degraded peptide products covalently linked to SNAP-FLAG. Even though the lysate was treated with protease inhibitors prior to peptide introduction, a sufficient amount of proteases must have remained active to degrade the linear peptide, while leaving the stapled peptides relatively untouched. This demonstrates that TAT is much more sensitive to proteolysis relative to stapled peptides. Further, this observation indicates degraded peptides can still label SNAP and that a simple, western blot-based mass shift assay can also qualitatively report on peptide stability.

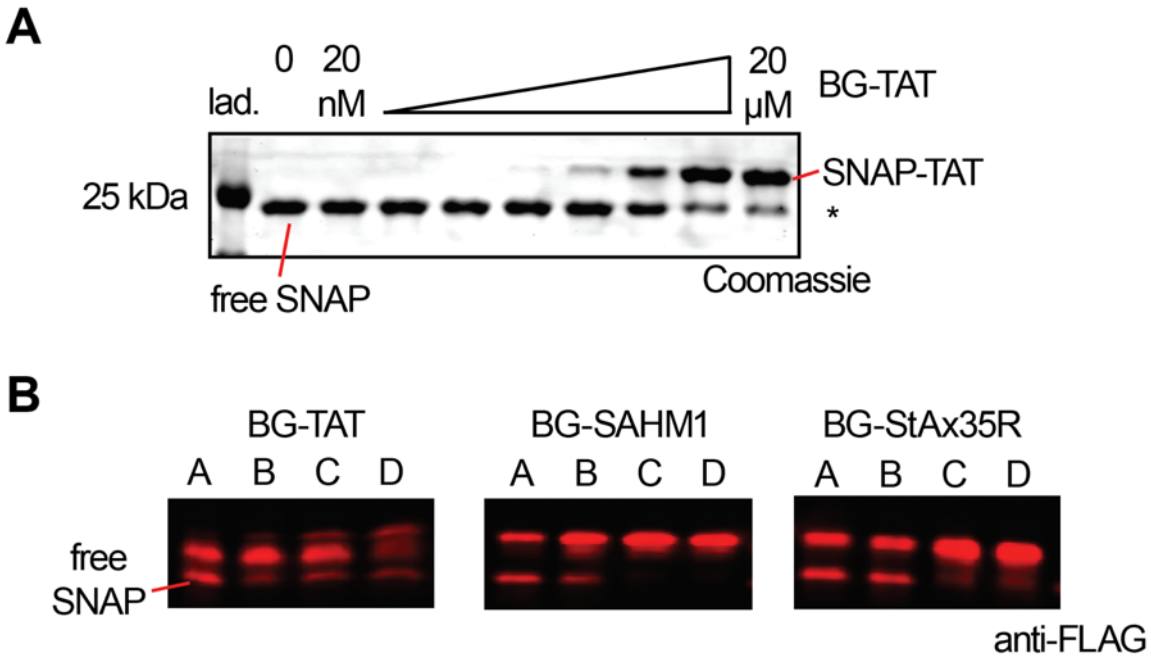


Figure 5.4: Representative *in vitro* data confirming the viability of the BG-SNAP system.
(A) SDS-PAGE results of recombinantly-expressed SNAP protein at 2 μM labeled across a dose range of **BG-TAT** peptide for 30 minutes in PBS, 1 mM DTT. Labeling efficiency was comparable to BG-functionalized small molecules. * Indicates fraction of unreacted SNAP protein; additional testing indicated this was likely due to active site Cys oxidation during batch purification. **(B)** Anti-FLAG western blot results of stable HeLa+SNAP-FLAG lysate labeled across dose ranges of **BG-TAT**, **BG-SAHM1**, and **BG-StAx35R** peptides for 30 minutes, immediately preceded by cell lysis in PBS with protease inhibitors (A = 0.05 μM; B = 0.2 μM; C = 1 μM; D = 10 μM).

5.2.3 Cell penetration and stability profiling with the BG-SNAP covalent labeling system

Next we turned to direct measures of cell penetration and intracellular stability of our peptide probes. Cell-based experiments were performed in 24 well plates, with consistent seeding of cells at 10,000 per well and BG-peptide treatment after 12-20 hours recovery, to ensure reproducibility across experiments. Following treatment, cells were washed and then lysed in-well, with RIPA buffer containing protease inhibitors and

additional additives depending on the purpose of the experiment. To allow for quantification of SNAP labeling, which corresponds with BG-peptide cytosolic access, a high concentration of **BG-FITC** probe was included in the lysis buffer. During lysis, the BG-FITC labels all un-reacted SNAP protein, allowing for both a quantitative competition-based readout as well as preventing the occurrence of SNAP-labeling by BG-peptides adhering to the cell surface or trapped in other compartments (*e.g.* exosomes). One untreated well was lysed in the presence of 20 μM BG-peptide to serve as a positive control and a marker for complete, full-length peptide-labeled SNAP protein. Additional controls were run with lysis buffer containing high concentrations of iodoacetamide and EDTA, known to rapidly inhibit SNAP protein activity, to ensure the fidelity of this system; mass shift results were comparable across these various conditions.

First, we assessed the dose and time dependent cellular uptake of **BG-TAT** peptide by western blot (**Figure 5.5 A**). As indicated by a dose-dependent decrease in FITC signal, **BG-TAT** exhibited dose-dependent cellular uptake and labeling of SNAP peptide. In general agreement with the previous lysate results, SNAP appears to be predominantly labeled with degraded TAT peptide, however only a small mass-shifted band can be observed at higher concentrations of BG-TAT treatment. Also echoing the previous lysate results, the positive control well, with BG-TAT co-treated during lysis, shows a higher molecular weight SNAP-FLAG band, presumably representing full length TAT labeling, however a variety of intermediate bands are also evident, suggesting a variety of degradation products are formed. Similarly, the time course experiment shows no apparent mass-shifted bands, but the gradual competing off of BG-FITC labeling over

extended treatment times (**Figure 5.5 A**). These collective results, indicative of TAT peptide's low metabolic stability, are perhaps not surprising, given its largely unstructured nature and the presence of an Arg residue near the N-terminal position of the BG moiety. Cleavage at this position by one of the many trypsin-like proteases found in, secreted by, and on the surface of human cells could occur at any point during the cell uptake process, resulting in labeled SNAP protein without sufficient differential mass to be detected by a band shift. Regardless, the FITC competition format of this assay allowed for quantification of uptake, with an apparent IC_{50} of ~10.2 μM and an apparent ET_{50} of ~2.2 hours at 10 μM. An emerging caveat in these data is that the measured labeling events likely represent the effects of a diverse mixture of degradation products.

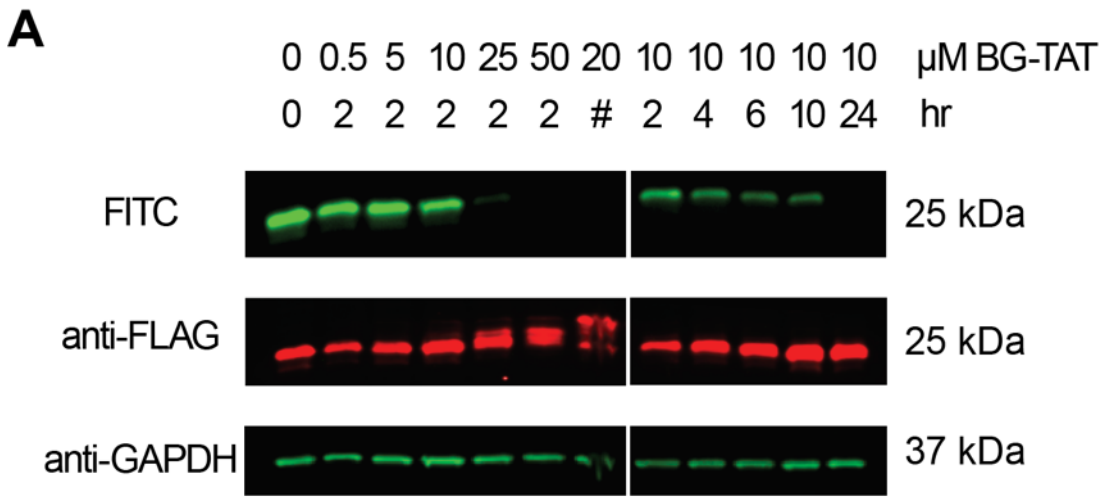


Figure 5.5: Assessment of BG-TAT dose and time dependent cell penetration and stability.
(A) FITC, anti-FLAG, and anti-GAPDH control western blot results of stable HeLa+SNAP-FLAG cells treated with either a dose range of **BG-TAT** for 2 hours, or a time course at 10 μM. All cells were lysed with 20 μM BG-FITC, except lane marked #, which indicates lysis with 20 μM **BG-TAT** as a positive control for full labeling of peptide.

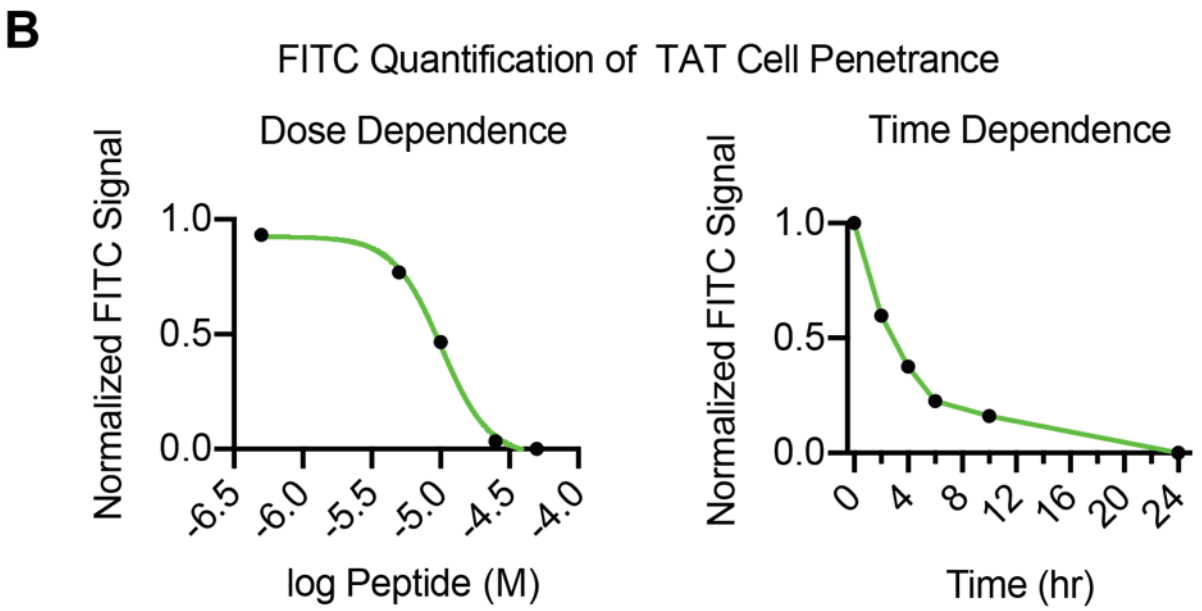


Figure 5.5 (continued):
(B) Quantification of FITC signal from results in **(A)**. Curve calculated from the application of a sigmoidal curve fit using Prism 8 software (left).

We next applied the stapled peptides **BG-SAHM1** and **BG-StAx35R** to dose dependent screening of cellular uptake with this assay. As seen with **BG-TAT**, treatment over 2 hours with either peptide resulted in decreased FITC signal, indicating dose-dependent cellular uptake and labeling of SNAP peptide (**Figure 5.6 A,B**). Full length mass-shifted peptide-SNAP-FLAG bands are only apparent with 25-50 μM treatment, as observed by anti-FLAG western blot, confirmed by the positive control lane of **BG-SAHM1** co-treatment during lysis lane (**Figure 5.6 A**, indicated by *****). Curiously, while **BG-SAHM1** fully labeled cytosolic SNAP-FLAG by 25 μM, ~20-30% of the protein remains unlabeled by **BG-StAx35R**, even at high treatment concentrations. This may represent mechanistic differences in the uptake, endosomal escape, or stability of these two stapled peptides.

FITC competition allowed for quantification of uptake, with an apparent IC_{50} of ~8.8 μM for **BG-SAHM1** and ~12.4 μM for **BG-StAx35R** at 2 hour treatment (**Figure 5.6 C**).

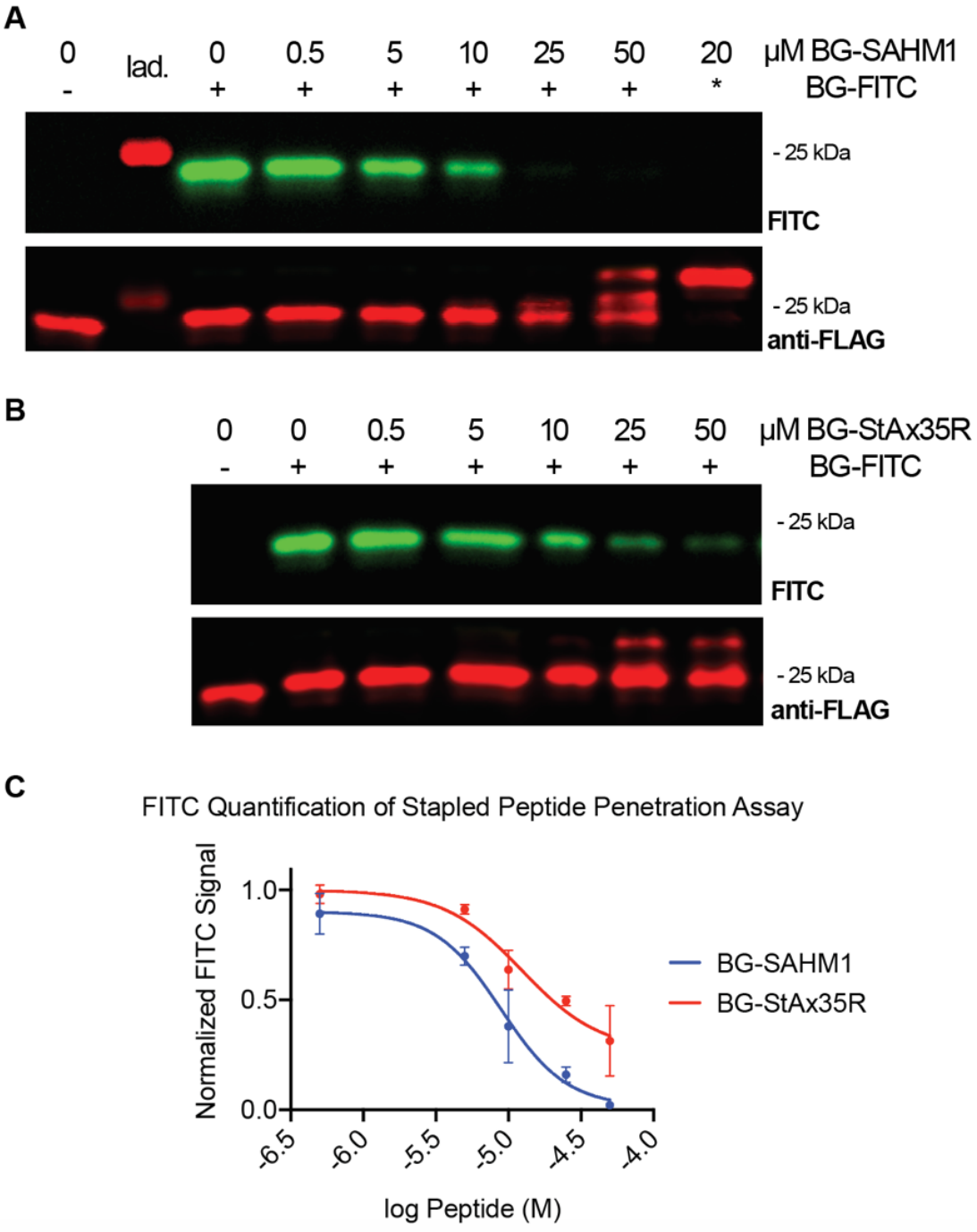


Figure 5.6: Assessment of BG-SAHM1 and BG-StAx35R dose dependent cell penetration and stability.

Figure 5.6, continued: (A) FITC and anti-FLAG western blot results of stable HeLa+SNAP-FLAG cells treated with a dose range of **BG-SAHM1** for 2 hours. All cells were lysed with 20 μM BG-FITC, except for left most negative control lane and lane marked *****, which indicates lysis with 20 μM **BG-SAHM1** as a positive control for full labeling with peptide. **(B)** FITC and anti-FLAG western blot results of stable HeLa+SNAP-FLAG cells treated with a dose range of **BG-StAx35R** for 2 hours. All cells were lysed with 20 μM BG-FITC, except for left most negative control lane. **(C)** Quantification of FITC signal from results in **(A,B)**. Curves calculated from the application of a sigmoidal curve fit using Prism 8 software.

Taken together with the previous **BG-TAT** results, a few questions and key learnings emerges from these data. Firstly, given previous reports of instability of thiol-maleimide conjugates in reducing environments [289], we wondered if the BG group detached from the peptides during treatment and uptake. The less-hindered free BG group may then preferentially react with the SNAP-FLAG protein, producing a decrease in fluorescence without a concomitant mass shift. Given the presence of free glutathione in the media and cellular environment, we would expect a retro Michael-type reaction to occur with subsequent thiol exchange. To test this, we incubated BG-peptides in PBS with a range of glutathione concentrations, then analyzed samples by LC-MS. Our results indicated the BG-peptide conjugates remained intact, ruling out this possibility (data not shown).

Alternatively, as previously mentioned with respect to the TAT sequence, both stapled peptides have Arg residues near the N-termini. Proteolysis occurring at these positions would result in a cleavage product, containing BG and 1 to 2 basic residues, which could preferentially enter cells and/or escape endosomes, resulting in rapid labeling of SNAP protein. Indeed, we developed a LC-MS-based proteolysis mapping assay to

study these effects. In a time course experiment, we exposed acetylated **StAx35R** to conditioned cell culture media (*i.e.* media in which cells had been grown for over 24 hours, presumably containing excreted proteases), simulating the treatment conditions of the uptake assay. By 24 hours, the majority of the peptide had been degraded as observed by LC-MS (**Figure 5.7**). Presumably the effects observed in this experiment would be enhanced by direct peptide treatment on cells, given the presence of cell-surface proteases, as well as the presence of proteases within endosomes and the cytosol upon uptake. This experiment further validates cell culture and treatment condition dependent differences noted across the CPP and protein transduction literature, emphasizing the importance of judicious selection of controls and reporting of experimental conditions to ensure reproducibility. In addition, taken alongside the preliminary results of the BG-SNAP cell penetration assay, it is clear that peptide stapling does not guarantee peptide stability, and further emphasizes the importance of capturing high-resolution data of the potential degradation products that do reach the cytosol. As noted previously, these products will have differential properties and on and off target effects relative to the full-length probes, which would have been screened against purified proteins *in vitro* prior to cell-based testing.

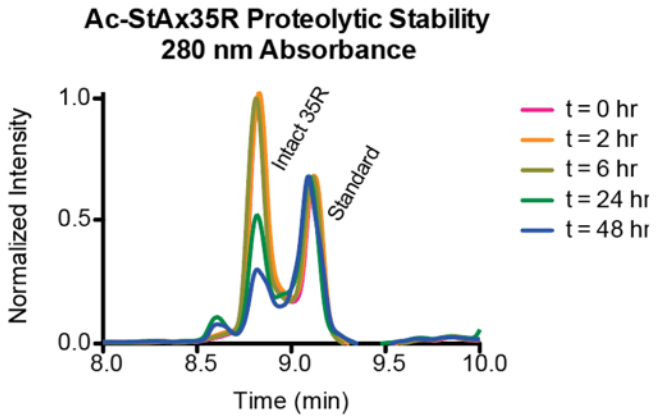


Figure 5.7: StAx35R shows proteolytic instability over time.
Over time, decreasing levels of the intact stapled peptide were observed by monitoring absorbance at 280 nm (signals normalized to a spiked-in analytical standard, eluting at ~9.1 minutes).

5.2.4 Using the BG-SNAP covalent-labeling system to assess membrane disruption

Another application of this BG-SNAP intracellular covalent-labeling system is for the assessment of membrane disruption by unlabeled peptide probes. This is an established problem in the field, as highly basic or amphipathic peptides, as well as some small molecules, can cause membrane disruption and thus non-specific toxicity in cell-based screening assays [290, 291]. Assays that measure extracellular release of intracellular contents, such as enzymes, are often used to capture these events. One prominent example is the lactate dehydrogenase assay. Existing assays typically only have endpoint readouts and may not be able to report on transient membrane disruption events occurring during treatment. Thus we sought to develop an alternative format of the cell penetration assay. This membrane disruption assay utilizes the non-membrane permeable **EEKE** peptide probe (conjugated with both BG and FITC) such that, when co-treated with unlabeled membrane-disrupting peptides, it could then gain entry into cells

and label the cytosolically localized SNAP-FLAG protein for fluorescent-based detection by gel (**Figure 5.2**).

We first confirmed the predicted impermeability of **EEKE** by exposing HeLa+SNAP-FLAG cells to extended incubations with the probe, as well as across a range of concentrations, using the same experimental design and processing as performed for the standard cell penetration assay (**Figure 5.8**). As anticipated, the highly-charged peptide could not access the cytosol on its own, with labeling only observed during co-treatment with lysis buffer (**Figure 5.8**, lane labeled with *).

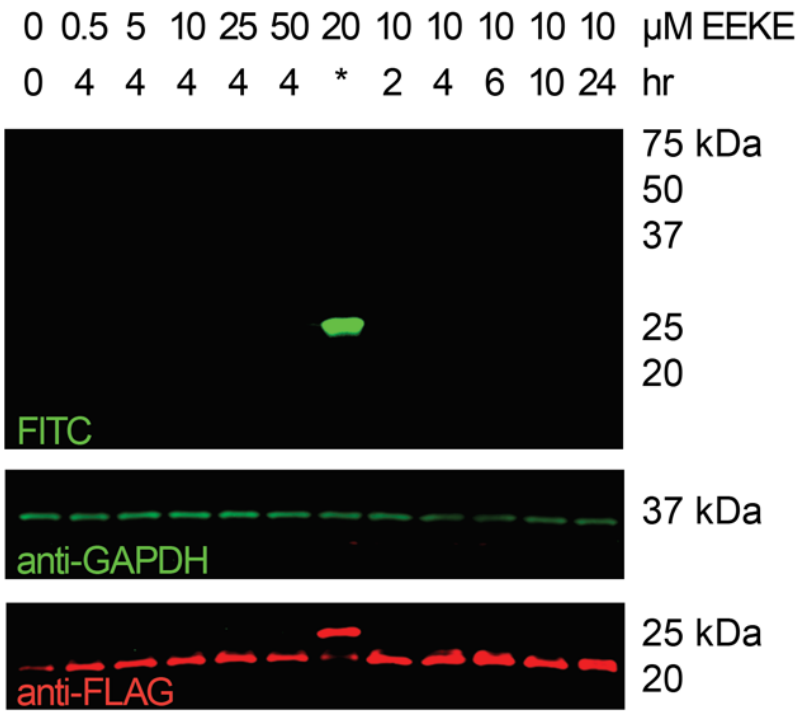


Figure 5.8: Validating EEKE peptide membrane impermeability.
HeLa+SNAP-FLAG cells were treated with **EEKE** peptide across a dose range for 4 hours and a time course at 10 μM. No **EEKE** peptide SNAP-labeling occurred across these conditions, except for in the lane labeled *, indicating co-treatment of 20 μM EEKE during cell lysis (positive control). Additional anti-GAPDH and anti-FLAG control blotting confirms consistent loading, lysis, and collection of sample, as well as an **EEKE**-labeling induced mass shift of SNAP detected by anti-FLAG blotting.

We then screened a panel of acetylated stapled peptide inhibitors in this assay to assess their potential for membrane disruption, in some cases testing with or without the presence of 10% fetal bovine serum (FBS) in culture media. **SAHM1** only displayed minor evidence of membrane disruption at the high concentration of 100 μM when no FBS was present (**Figure 5.9 A,B**). **StAx35R**, on the other hand, displayed minor membrane disruption at higher concentration in FBS-containing media, which was exacerbated in media that did not contain FBS (**Figure 5.9 A,B**). We also screened stapled peptide RAB25 inhibitors **RFP24** and **RFP26**, notable for having identical sequences, except for a conservative Phe (**RFP24**) to norleucine (**RFP26**) mutation (**Table 5.1**). Intriguingly, **RFP24** only displayed minor membrane disruption at 100 μM, whereas **RFP26** displayed increasing membrane disruption across tested doses (**Figure 5.9 A,B**). To counter-screen and confirm the validity of the assay's results, we subjected the same panel of peptides to the LDH assay, with 100 μM treatment in the presence of 10% FBS (**Figure 5.9 C**). The results from this comparison were largely in agreement for **SAHM1** and **StAx35R**, however **RFP24** and **RFP26** both displayed much higher levels of membrane disruption as assessed by LDH assay. These latter differences may be due to subtle variation in treatment and assay conditions (*e.g.* the 96 well format of the LDH assay, or due to co-treatment with **EEKE** peptide). Alternatively, they may reflect signal amplification due to the enzymatic nature of the endpoint LDH assay readout, or the fact that **EEKE**-mediated signal only accumulates over the course of treatment as membrane disruption gradually occurs. Regardless, these phenomena and comparative validations necessitate further study.

Taken together, these preliminary results validate the potential suitability of this alternate format BG-SNAP system assay for assessing membrane disruption. Further, these data emphasize the importance of careful selection of screening conditions, as well as the highly context- and sequence-dependent nature of peptide probe effects, which are often hard to predict, challenging to screen for, and largely poorly understood.

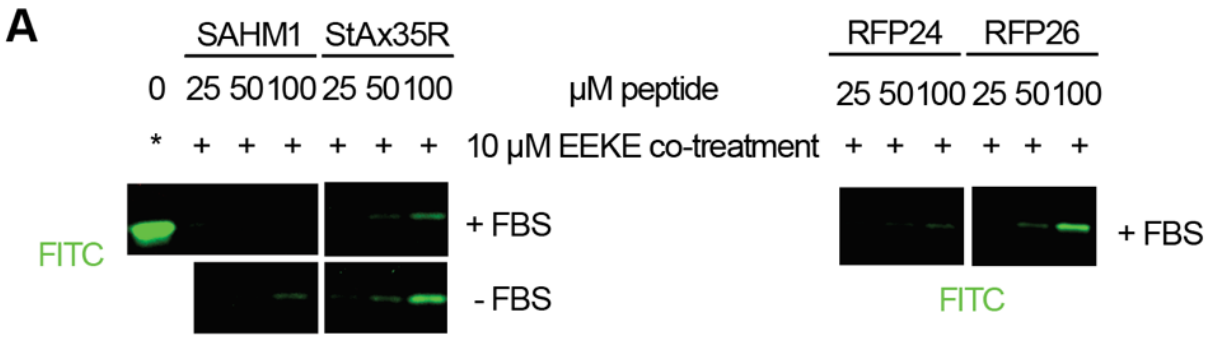


Figure 5.9: Screening unlabeled peptides for membrane disruption with co-treatment of EEKE peptide and comparison to LDH assay results.
(A) Representative fluorescent gels from co-treatment of HeLa+SNAP-FLAG cells with 10 μM **EEKE** peptide across a dose range of unlabeled peptides for 2 hours. Membrane disruption by **SAHM1** and **StAx35R** was also assessed in media lacking fetal bovine serum (FBS). * indicates co-treatment of 10 μM EEKE during cell lysis, to serve as a positive control for internal normalization.

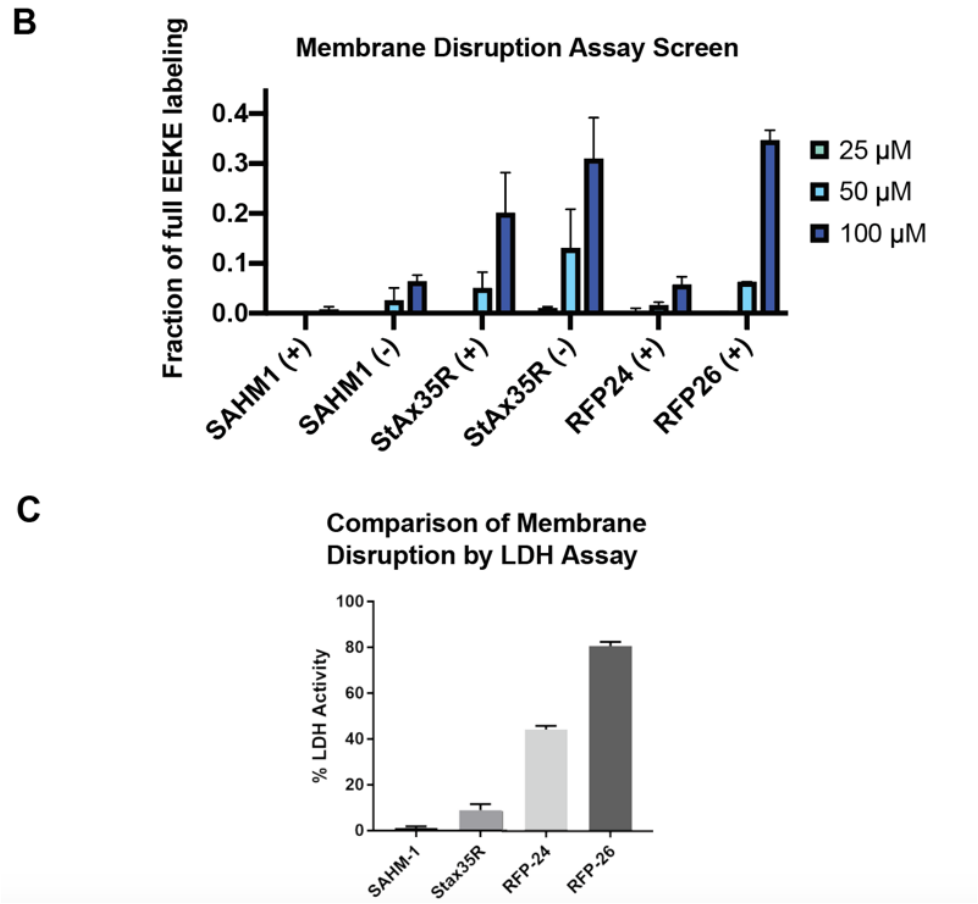


Figure 5.9 (continued):
(B) Quantification of membrane disruption as measured by band fluorescence intensity normalized to positive control lane *. **(+)** indicates 10% FBS, **(-)** indicates no FBS. Data from 2 biological replicates, with 2 technical replicates from each. **(C)** LDH assay results from corresponding conditions with 100 μM peptide treatment, normalized to SDS-treated positive control wells. 2 technical replicates each.

5.3 Discussion

Future work in developing the BG-SNAP cell penetration assay will take on several different tracts. While it demonstrates increased utility over existing approaches, by revealing both cytosolic localization of peptides and qualitative assessment of their stability throughout cell uptake, additional design features must be considered. First, N-terminal BG-functionalization, which was primarily chosen out of synthetic ease, will likely

not be ideal for most peptide probes. Several different BG-functionalization locations will likely be warranted to maximally report on the different degradation species formed during cell uptake. In the case of SAHM1, for instance, conjugation of BG via the thiol-maleimide strategy could occur at the naturally occurring Cys residue (mutated to Ser in these experiments) that is bridged by the stapling groups (**Table 5.1**). This would allow for capture and study of the central, presumably most stabilized region of the peptide probe, which would likely retain the majority of its specific bioactivity. A more generalized strategy, already explored for a variety of other applications [292], involves functionalization of the stapling cross-link itself. This could be accomplished through using different stabilization chemistries or alternatively functionalized amino acids for late-stage incorporation of BG. Any of these alternative BG-functionalized probes would first need to be tested by *in vitro* SNAP labeling, to ensure no steric effects would inhibit potential readout.

As stated in the introduction of this chapter, further expansion of this assay's utility would involve coupling it with an MS-based readout. In this regime, following BG-peptide treatment of cells and subsequent lysis, the FLAG-tagged SNAP protein would be enriched by immunoprecipitation. Protein-labeled peptide identification could then be accomplished by subjecting samples to analysis by MALDI readout. The single ionization state of the analytes could then be readily used to unequivocally assign the identities of the peptide degradation products that ultimately achieved cytosolic access. This would present an incredibly powerful addition to the suite of tools used to study intracellularly targeted peptide and protein probes, especially given the perhaps unsurprising finding

that peptide probes, even when chemically stabilized, are still susceptible to proteolysis. In preclinical studies of peptide and protein drug candidates, *in vivo* compound circulation is typically assessed by measuring the presence of a radio-labeled version of the ligand, which does not report on the presence or stability of the full length compound [134, 293]. As previously addressed, any degradation products that are formed during treatment would likely have differential activities relative to their precursor.

The ability to determine and mitigate degradation of these probes is an emerging goal in the study of synthetic biologics. Beyond stabilization by chemical cross-linking, residues or sequences with high degrees of proteolytic liability have been replaced with D-amino acids or residues with otherwise modified sidechains or backbones [152, 294, 295]. To most effectively design these optimal synthetic biologics, however, robust analytical techniques that provide the most relevant and high-resolution data will need to be adopted. We believe that the BG-SNAP covalent labeling system, described herein, for measuring cell penetration of synthetic biologics and other cargos, offers an attractive facile alternative to other approaches. Its ability to also screen for membrane disruption using the same cell line and probe reagents can further aid in the triaging in early-stage compound discovery and development.

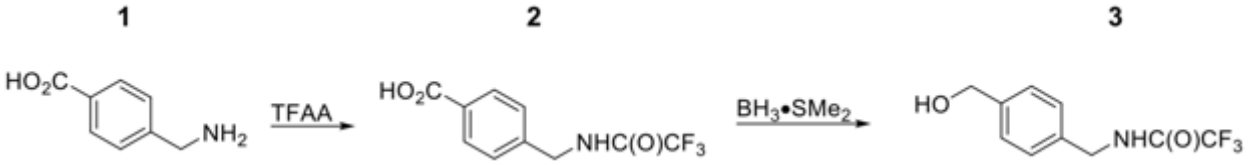
5.4 Tables

Compound	Sequence
BG-TAT	BG–C–β–GRKKRRQRRRPQ
BG-SAHM1	BG–C–β–ERLRRRI*LSR*HHST
BG-StAx35R	BG–C–β–RRWPR*ILD*HVRRVWR
EEKE	BG–C–β–EEKEK$_{FITC}$
Ac-SAHM1	Ac–ERLRRRI*LCR*HHST
Ac-StAx35R	Ac–RRWPR*ILD*HVRRVWR
Ac-RFP24	Ac–RINFRLQN*IDR*IVRIFETNPSILRVK
Ac-RFP26	Ac–RINFRLQN*IDR*IVRIN$_L$ETNPSILRVK

Table 5.1: Compounds synthesized for BG-SNAP assay.
β, beta-alanine; *, S_5; N_L, norleuecine; blue, relevant mutations.

5.5 Methods

Chemical synthesis of 4-(aminomethyl)benzyl guanine (7)

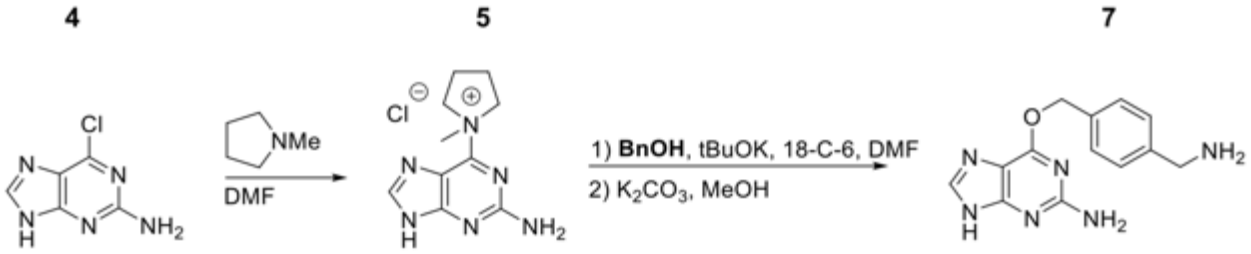


4-[[(2,2,2-trifluoroacetyl)amino]methyl]benzoic acid (2).

Solid 4-(aminomethyl)benzoic acid (1) (15.1 g, 100 mmol) was dissolved in TFAA (42 mL) cooled to 0 °C. Once dissolved, the ice bath was removed and the reaction was allowed to stir at room temperature until starting material was consumed, ~2 hr. Upon completion, the reaction was quenched with H2O (100 mL) and precipitate collected via vacuum filtration. The product was dried under suction then collected to afford benzoic acid (2) (24.0 g 97%) as a white solid. 1H NMR (500 MHz, DMSO-d6) δ 7.93 (d, J = 8.3 Hz, 2H), 7.39 (d, J = 8.3 Hz, 2H), 4.47 (d, J = 6.0 Hz, 2H). 13C NMR (125 MHz, DMSO-d6) δ 167.12, 156.58 (q, J = 36.3 Hz), 142.50, 129.89, 129.65, 127.44, 116.05 (q, J = 228.1 Hz), 42.41.

2,2,2-trifluoro-N-[[4-(hydroxymethyl)phenyl]methyl]acetamide (3).

Borane dimethylsulfide (13.8 mL, 145 mmol) was added dropwise to an anhydrous THF (483 mL) solution of 4-[[(2,2,2-trifluoroacetyl)amino]methyl]benzoic acid (2) (11.9 g, 48.3 mmol) while maintaining an internal temperature of 0 °C. After complete addition the ice bath was removed and the mixture was stirred at room temperature overnight. The reaction was quenched with MeOH (100 mL) and stirred at room temperature for an additional 1 hr. The volatiles were removed and the residue taken up in EtOAc. Impurities were removed by successive washes with 1M NaOH, H_2O, and brine. The organics were then dried over Na_2SO_4, filtered, and concentrated in vacuo. Subsequent column chromatographic purification of the residue, eluting with 20:1 (CH_2Cl_2:MeOH), provided amide product (3) (10.4 g, 92%) as a white solid. ^{1}H NMR (500 MHz, CDCl3-d) δ 7.35 (d, J = 8.2 Hz, 1H), 7.27 (d, J = 8.0 Hz, 1H), 4.67 (s, 1H), 4.50 (d, J = 5.8 Hz, 1H). 13C NMR (125 MHz, CDCl3-d) δ 157.32 (q, J = 36.7 Hz), 141.11, 135.33, 128.29, 127.67, 115.97 (q, J = 287.8 Hz), 64.90, 43.75.



6-(1-methylpyrrolidin-1-ium-1-yl)-7H-purin-2-amine chloride (5).

Neat N-methylpyrollidine (4) (7.80 mL, 73.7 mmol) was added to an anhydrous DMF (144 mL) solution of 6-chloro-7H-purin-2-amine (5.00 g, 29.5 mmol) and stirred at 40°C overnight. The resultant chloride salt (5) (5.45 g, 73%) was collected via vacuum filtration, dried under suction, and used without further purification.

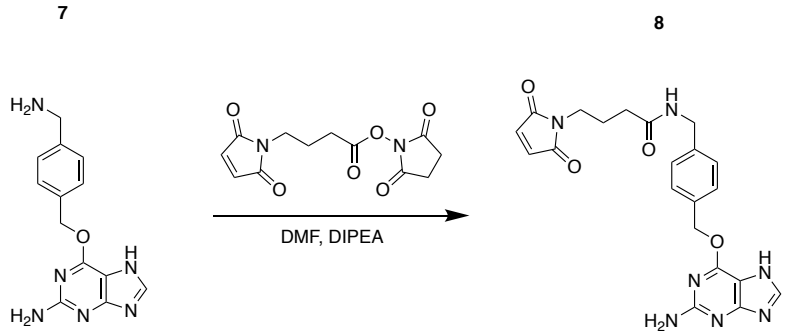
N-[[4-[(2-amino-7H-purin-6-yl)oxymethyl]phenyl]methyl]-2,2,2-trifluoroacetamide (6).

An oven dried round-bottom flask containing 6-(1-methylpyrrolidin-1-ium-1-yl)-7H-purin- 2-amine chloride (5) (6.90 g, 27.1 mmol), 2,2,2-trifluoro-N-[[4-(hydroxymethyl)phenyl]methyl] acetamide (3) (12.6 g, 54.2 mmol), potassium tert-butoxide (12.2 g, 108 mmol), and 18-c-6 (1.07 g, 4.07 mmol) in 54 mL DMF was stirred for 6 hrs at 50 °C. Upon completion the solvent was evaporated and the crude residue absorbed to silica. Purification was achieved using column chromatography eluting with a gradient of MeOH in CH_2Cl_2 (2-10 %) to give acetamide product (6) (8.94 g, 90%) as a white solid.

6-[[4-(aminomethyl)phenyl]methoxy]-7H-purin-2-amine (7).

N-[[4-[(2-amino-7H-purin-6-yl)oxymethyl]phenyl]methyl]-2,2,2-trifluoroacetamide (6) (2.59 g, 7.06 mmol) was added to a suspension of K_2CO_3 (4.84 g, 35.0 mmol) in 21 mL of MeOH:H_2O (20:1) and stirred vigorously overnight at 50°C. Upon consumption of the starting material the mixture was filtered through a pad of celite, washing with MeOH.

The filtrate was concentrated in vacuo and the residue taken up in 10 mL H_2O. While cooling, the pH was adjusted to ~7 with HCl. The resultant precipitate was isolated via suction filtration and washed with cold water to yield amine product (7) (1.77 g, 93%) as a white solid. 13C NMR (12d MHz, DMSO-d6) δ 159.65, 157.73, 143.72, 140.45, 135.00, 134.67, 128.47, 127.14, 126.98, 66.66, 45.27.



maleimidobutyryl-[4-(aminomethyl)phenyl]methoxy]-7H-purin-2-amine (8) (BG-maleimide).

6-[[4-(aminomethyl)phenyl]methoxy]-7H-purin-2-amine (7) (50 mg, 185 μmol) was dissolved in 2 mL DMF and DIPEA (65 μL, 370 μmol). N-γ-maleimidobutyryl-oxysuccinimide ester (56 mg, 200 μmol) was added to solution and it was stirred at room temperature for 2 hours until consumption of starting material. The crude reaction mixture was diluted in MeOH and purified by HPLC.

Stapled peptide synthesis and purification.

Unmodified and hydrocarbon stapled peptides were synthesized by Fmoc-based solid phase peptide synthesis and purified by reverse-phase HPLC with a C18 column

(Agilent, Palo Alto, CA) as previously reported [138]. In the case of the EEKE peptide, prior to cleavage and purification, the lysine(Mmt) was deprotected and acylated with 5 eq FITC and 10 eq DIPEA in DMF, in the dark overnight. Peptides were analyzed by LC-MS using a C18 reverse-phase column (Phenomenex, 5.0 × 50 mm, pore size 110 Å, particle size 5 μm); Buffer A (5/95/0.1% ACN/H_2O/TFA) and Buffer B (95:5:0.1% ACN/H_2O/TFA); and a 20 min method with the following gradient (flow rate 0.5 mL/min): 0% buffer B over 3 min, 0–65% buffer B over 15 min, 65–100% buffer B over 1 min; 100–0% buffer B over 1 min. Purified peptides were lyophilized, quantified by A_{280}, dissolved in DMSO as 10 mM stocks and stored at –20 °C; free thiol containing peptides were similarly quantified where appropriate, lyophilized, and stored at -20 °C for less 5 days prior to conjugation reactions.

BG-peptide conjugation reaction.

1.2 equiv BG-maleimide (**8**) was dissolved in 25% ACN/phosphate buffer, pH 7.4 and then added to 1 equiv lyophilized thiol-containing peptide. Reaction solution pH was checked and adjusted to 7.4 if necessary. The reaction was rotated at room temperature for 2 hours, diluted in 50% ACN/H_2O, and purified by reverse-phase HPLC. Purified BG-peptides were lyophilized, quantified by A_{280}, dissolved in DMSO as 10 mM stocks and stored at -20 °C.

Cell penetration and membrane stability assay.

HeLa+SNAP-FLAG cells were plated in a 24 well plate at 10^5/well in 500 μL RPMI and grown overnight (12-20 hours). BG-peptide (or unlabeled peptide and **EEKE** peptide) DMSO stocks were diluted to the appropriate concentration in media. Media was aspirate from wells and peptide treatment media was added. The plate was incubated at 37°C for the indicated time(s). At the end of treatment, media was aspirated and cells were carefully washed with ice-cold PBS. In-well lysis was then performed, covered, on ice, for 20-30 min. Lysis buffer: 75 μL RIPA buffer, with fresh protease inhibitors cocktail added, 20 μM **BG-FITC** (where appropriate). A no-**BG-FITC** negative control and BG-peptide/EEKE peptide positive control were included (where appropriate). Following lysis, lysates were carefully pipetted up-and-down 3-5x to homogenize sample. Gel loading samples were prepared by adding 4x loading buffer and then subjected to standard SDS-PAGE and western blotting procedures.

Conditioned-media peptide stability assay.

Conditioned media was collected from HeLa cells grown for at least 24 hours. A DMSO stock of **StAx35R** peptide was diluted to 100 μM in the conditioned media and incubated at 37°C. Aliquots taken at different time points during the incubation were stored at -20°C then thawed on ice at the end of the experiment. Following this, protein and other media content were then precipitated with ice-cold acetonitrile (ACN) and then pelleted out of solution. The supernatant was then collected and lyophilized. Lyophilized samples were

dissolved in 10% ACN/H_2O/0.1% TFA and prepared for liquid chromatography-mass spectrometry (LC-MS) analysis as described above.

References

1. Hartley, H., *Origin of the word 'protein'.* Nature, 1951. **168**(4267): p. 244.

2. Moiola, M., M.G. Memeo, and P. Quadrelli, *Stapled Peptides-A Useful Improvement for Peptide-Based Drugs.* Molecules, 2019. **24**(20).

3. Rang, H.P. and R.G. Hill, *Choosing the target*, in *Drug Discovery and Development*, R.G. Hill and H.P. Rang, Editors. 2013, Churchill Livingstone. p. 63-76.

4. Schenone, M., et al., *Target identification and mechanism of action in chemical biology and drug discovery.* Nat Chem Biol, 2013. **9**(4): p. 232-40.

5. Hughes, J.P., et al., *Principles of early drug discovery.* Br J Pharmacol, 2011. **162**(6): p. 1239-49.

6. Gerry, C.J. and S.L. Schreiber, *Chemical probes and drug leads from advances in synthetic planning and methodology.* Nat Rev Drug Discov, 2018. **17**(5): p. 333-352.

7. Lévesque, H. and O. Lafont, *L'aspirine à travers les siècles: Rappel historique.* La Revue de Médecine Interne, 2000. **21**: p. S8-S17.

8. Mullard, A., *A probe for every protein.* Nat Rev Drug Discov, 2019. **18**(10): p. 733-736.

9. Arrowsmith, C.H., et al., *The promise and peril of chemical probes.* Nat Chem Biol, 2015. **11**(8): p. 536-41.

10. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 13-20.

11. Wang, H., et al., *Peptide-based inhibitors of protein–protein interactions: biophysical, structural and cellular consequences of introducing a constraint.* Chemical Science, 2021.

12. Stelzl, U., et al., *A human protein-protein interaction network: a resource for annotating the proteome.* Cell, 2005. **122**(6): p. 957-68.

13. Luck, K., et al., *A reference map of the human binary protein interactome.* Nature, 2020. **580**(7803): p. 402-408.

14. Cho, N.H., et al., OpenCell: proteome-scale endogeneous tagging enables the cartography of human cellular organization, bioRxiv, 2021. 03.29.437450.

15. Scott, D.E., et al., *Small molecules, big targets: drug discovery faces the protein-protein interaction challenge.* Nat Rev Drug Discov, 2016. **15**(8): p. 533-50.

16. Harvey, A.L., R. Edrada-Ebel, and R.J. Quinn, *The re-emergence of natural products for drug discovery in the genomics era.* Nat Rev Drug Discov, 2015. **14**(2): p. 111-29.

17. Dias, D.A., S. Urban, and U. Roessner, *A historical overview of natural products in drug discovery.* Metabolites, 2012. **2**(2): p. 303-36.

18. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25. 1.* Advanced Drug Delivery Reviews, 2001. **46**(1-3): p. 3-26.

19. Lovering, F., J. Bikker, and C. Humblet, *Escape from flatland: increasing saturation as an approach to improving clinical success.* J Med Chem, 2009. **52**(21): p. 6752-6.

20. Shultz, M.D., *Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs.* J Med Chem, 2019. **62**(4): p. 1701-1714.

21. Bender, A. and I. Cortes-Ciriano, *Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet.* Drug Discov Today, 2021. **26**(2): p. 511-524.

22. Macarron, R., et al., *Impact of high-throughput screening in biomedical research.* Nat Rev Drug Discov, 2011. **10**(3): p. 188-95.

23. Mennen, S.M., et al., *The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future.* Organic Process Research & Development, 2019. **23**(6): p. 1213-1242.

24. Capdeville, R., et al., *Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug.* Nat Rev Drug Discov, 2002. **1**(7): p. 493-502.

25. Nowell, P.C., *Discovery of the Philadelphia chromosome: a personal perspective.* J Clin Invest, 2007. **117**(8): p. 2033-5.

26. Lugo, T.G., et al., *Tyrosine kinase activity and transformation potency of bcr-abl oncogene products.* Science, 1990. **247**(4946): p. 1079-82.

27. Mehr, I.J., *Preparing for the revolution--pharmacogenomics and the clinical lab.* Pharmacogenomics, 2000. **1**(1): p. 1-4.

28. Vener, C., et al., *First-line imatinib vs second- and third-generation TKIs for chronic-phase CML: a systematic review and meta-analysis.* Blood Adv, 2020. **4**(12): p. 2723-2735.

29. Wishart, D.S., et al., *DrugBank: a comprehensive resource for in silico drug discovery and exploration.* Nucleic Acids Res, 2006. **34**(Database issue): p. D668-72.

30. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.

31. Hopkins, A.L. and C.R. Groom, *The druggable genome.* Nat Rev Drug Discov, 2002. **1**(9): p. 727-30.

32. Dixon, S.J. and B.R. Stockwell, *Identifying druggable disease-modifying gene products.* Curr Opin Chem Biol, 2009. **13**(5-6): p. 549-55.

33. Arkin, M.R. and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing towards the dream.* Nat Rev Drug Discov, 2004. **3**(4): p. 301-17.

34. Arkin, M.R., Y. Tang, and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing toward the reality.* Chem Biol, 2014. **21**(9): p. 1102-14.

35. Stumpf, M.P., et al., *Estimating the size of the human interactome.* Proc Natl Acad Sci U S A, 2008. **105**(19): p. 6959-64.

36. Garcia-Serna, R. and J. Mestres, *Anticipating drug side effects by comparative pharmacology.* Expert Opin Drug Metab Toxicol, 2010. **6**(10): p. 1253-63.

37. Bleicher, K.H., et al., *Hit and lead generation: beyond high-throughput screening.* Nat Rev Drug Discov, 2003. **2**(5): p. 369-78.

38. Brown, J.B. and Y. Okuno, *Systems biology and systems chemistry: new directions for drug discovery.* Chem Biol, 2012. **19**(1): p. 23-8.

39. Bon, R.S. and H. Waldmann, *Bioactivity-guided navigation of chemical space.* Acc Chem Res, 2010. **43**(8): p. 1103-14.

40. Reymond, J.L., *The chemical space project.* Acc Chem Res, 2015. **48**(3): p. 722-30.

41. Scannell, J.W., et al., *Diagnosing the decline in pharmaceutical R&D efficiency.* Nat Rev Drug Discov, 2012. **11**(3): p. 191-200.

42. Darnell, J.E., Jr., *Transcription factors as targets for cancer therapy.* Nat Rev Cancer, 2002. **2**(10): p. 740-9.

43. Lee, T.I. and R.A. Young, *Transcriptional regulation and its misregulation in disease.* Cell, 2013. **152**(6): p. 1237-51.

44. Chen, T., *Nuclear receptor drug discovery.* Curr Opin Chem Biol, 2008. **12**(4): p. 418-26.

45. Hagenbuchner, J. and M.J. Ausserlechner, *Targeting transcription factors by small compounds--Current strategies and future implications.* Biochem Pharmacol, 2016. **107**: p. 1-13.

46. Bushweller, J.H., *Targeting transcription factors in cancer - from undruggable to reality.* Nat Rev Cancer, 2019. **19**(11): p. 611-624.

47. Morelli, X., R. Bourgeas, and P. Roche, *Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I).* Curr Opin Chem Biol, 2011. **15**(4): p. 475-81.

48. Mullard, A., *Phosphatases start shedding their stigma of undruggability.* Nat Rev Drug Discov, 2018. **17**(12): p. 847-849.

49. Lu, W., et al., *Fragment-based covalent ligand discovery.* RSC Chemical Biology, 2021.

50. Faust, T.B., et al., *Small-Molecule Approaches to Targeted Protein Degradation.* Annual Review of Cancer Biology, 2021. **5**(1): p. 181-201.

51. Burslem, G.M. and C.M. Crews, *Small-Molecule Modulation of Protein Homeostasis.* Chem Rev, 2017. **117**(17): p. 11269-11301.

52. Mullard, A., *2019 FDA drug approvals.* Nat Rev Drug Discov, 2020. **19**(2): p. 79-84.

53. Schuhmacher, A., O. Gassmann, and M. Hinder, *Changing R&D models in research-based pharmaceutical companies.* J Transl Med, 2016. **14**(1): p. 105.

54. Marks, L., *The birth pangs of monoclonal antibody therapeutics: the failure and legacy of Centoxin.* MAbs, 2012. **4**(3): p. 403-12.

55. Ramaraj, T., et al., *Antigen-antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures.* Biochim Biophys Acta, 2012. **1824**(3): p. 520-32.

56. Bailly, M., et al., *Predicting Antibody Developability Profiles Through Early Stage Discovery Screening.* MAbs, 2020. **12**(1): p. 1743053.

57. Hansel, T.T., et al., *The safety and side effects of monoclonal antibodies.* Nat Rev Drug Discov, 2010. **9**(4): p. 325-38.

58. Drago, J.Z., S. Modi, and S. Chandarlapaty, *Unlocking the potential of antibody-drug conjugates for cancer therapy.* Nat Rev Clin Oncol, 2021.

59. Xu, X., et al., *Quantitation and modeling of post-translational modifications in a therapeutic monoclonal antibody from single- and multiple-dose monkey pharmacokinetic studies using mass spectrometry.* PLoS One, 2019. **14**(10): p. e0223899.

60. Saunders, K.O., *Conceptual Approaches to Modulating Antibody Effector Functions and Circulation Half-Life.* Front Immunol, 2019. **10**: p. 1296.

61. *The druggable proteome,* The Human Protein Atlas, 2021. Retrieved from www.proteinatlas.org, April 27, 2021.

62. Zorzi, A., K. Deyle, and C. Heinis, *Cyclic peptide therapeutics: past, present and future.* Curr Opin Chem Biol, 2017. **38**: p. 24-29.

63. Palomo, J.M., *Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides.* RSC Adv., 2014. **4**(62): p. 32658-32672.

64. Merrifield, R.B., *Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide.* Journal of the American Chemical Society, 1963. **85**(14): p. 2149-2154.

65. Jaradat, D.M.M., *Thirteen decades of peptide synthesis: key developments in solid phase peptide synthesis and amide bond formation utilized in peptide ligation.* Amino Acids, 2018. **50**(1): p. 39-68.

66. Dawson, P.E., et al., *Synthesis of proteins by native chemical ligation.* Science, 1994. **266**(5186): p. 776-9.

67. Cistrone, P.A., et al., *Native Chemical Ligation of Peptides and Proteins.* Curr Protoc Chem Biol, 2019. **11**(1): p. e61.

68. Hartrampf, N., et al., *Synthesis of proteins by automated flow chemistry.* Science, 2020. **368**(6494): p. 980-987.

69. Owens, A.E., et al., *MOrPH-PhD: An Integrated Phage Display Platform for the Discovery of Functional Genetically Encoded Peptide Macrocycles.* ACS Cent Sci, 2020. **6**(3): p. 368-381.

70. Scott, J.K. and G.P. Smith, *Searching for peptide ligands with an epitope library.* Science, 1990. **249**(4967): p. 386-90.

71. Wang, T., et al., *Continuous directed evolution of proteins with improved soluble expression.* Nat Chem Biol, 2018. **14**(10): p. 972-980.

72. Lau, J.L. and M.K. Dunn, *Therapeutic peptides: Historical perspectives, current development trends, and future directions.* Bioorg Med Chem, 2018. **26**(10): p. 2700-2707.

73. Tan, A.C., et al., *Atrial natriuretic peptide. An overview of clinical pharmacology and pharmacokinetics.* Clin Pharmacokinet, 1993. **24**(1): p. 28-45.

74. Madala, P.K., et al., *Update 1 of: Proteases universally recognize beta strands in their active sites.* Chem Rev, 2010. **110**(6): p. PR1-31.

75. Kessler, H., *Conformation and Biological Activity of Cyclic Peptides.* Angewandte Chemie International Edition in English, 1982. **21**(7): p. 512-523.

76. Gang, D., D.W. Kim, and H.S. Park, *Cyclic Peptides: Promising Scaffolds for Biopharmaceuticals.* Genes (Basel), 2018. **9**(11).

77. Naylor, M.R., et al., *Cyclic peptide natural products chart the frontier of oral bioavailability in the pursuit of undruggable targets.* Curr Opin Chem Biol, 2017. **38**: p. 141-147.

78. Drucker, D.J., *Advances in oral peptide therapeutics.* Nat Rev Drug Discov, 2020. **19**(4): p. 277-289.

79. Bockus, A.T., C.M. McEwen, and R.S. Lokey, *Form and function in cyclic peptide natural products: a pharmacokinetic perspective.* Curr Top Med Chem, 2013. **13**(7): p. 821-36.

80. Blanco, M.J., *Building upon Nature's Framework: Overview of Key Strategies Toward Increasing Drug-Like Properties of Natural Product Cyclopeptides and Macrocycles.* Methods Mol Biol, 2019. **2001**: p. 203-233.

81. Arnison, P.G., et al., *Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature.* Nat Prod Rep, 2013. **30**(1): p. 108-60.

82. White, C.J. and A.K. Yudin, *Contemporary strategies for peptide macrocyclization.* Nat Chem, 2011. **3**(7): p. 509-24.

83. Yudin, A.K., *Macrocycles: lessons from the distant past, recent developments, and future directions.* Chem Sci, 2015. **6**(1): p. 30-49.

84. Diaz, D.B., et al., *Illuminating the dark conformational space of macrocycles using dominant rotors.* Nat Chem, 2021. **13**(3): p. 218-225.

85. Jwad, R., D. Weissberger, and L. Hunter, *Strategies for Fine-Tuning the Conformations of Cyclic Peptides.* Chem Rev, 2020. **120**(17): p. 9743-9789.

86. Rezai, T., et al., *Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers.* J Am Chem Soc, 2006. **128**(8): p. 2510-1.

87. Dougherty, P.G., A. Sahni, and D. Pei, *Understanding Cell Penetration of Cyclic Peptides.* Chem Rev, 2019. **119**(17): p. 10241-10287.

88. Kohli, R.M. and C.T. Walsh, *Enzymology of acyl chain macrocyclization in natural product biosynthesis.* Chem Commun (Camb), 2003(3): p. 297-307.

89. Sussmuth, R.D. and A. Mainz, *Nonribosomal Peptide Synthesis-Principles and Prospects.* Angew Chem Int Ed Engl, 2017. **56**(14): p. 3770-3821.

90. Montalban-Lopez, M., et al., *New developments in RiPP discovery, enzymology and engineering.* Nat Prod Rep, 2021. **38**(1): p. 130-239.

91. Wang, C.K. and D.J. Craik, *Designing macrocyclic disulfide-rich peptides for biotechnological applications.* Nat Chem Biol, 2018. **14**(5): p. 417-427.

92. Gongora-Benitez, M., J. Tulla-Puche, and F. Albericio, *Multifaceted roles of disulfide bonds. Peptides as therapeutics.* Chem Rev, 2014. **114**(2): p. 901-26.

93. Mochizuki, M., et al., *Regioselective formation of multiple disulfide bonds with the aid of postsynthetic S-tritylation.* Org Lett, 2015. **17**(9): p. 2202-5.

94. Gongora-Benitez, M., et al., *Optimized Fmoc solid-phase synthesis of the cysteine-rich peptide linaclotide.* Biopolymers, 2011. **96**(1): p. 69-80.

95. Chevalier, A., et al., *Massively parallel de novo protein design for targeted therapeutics.* Nature, 2017. **550**(7674): p. 74-79.

96. Hill, T.A., et al., *Constraining cyclic peptides to mimic protein structure motifs.* Angew Chem Int Ed Engl, 2014. **53**(48): p. 13020-41.

97. Harrison, R.S., et al., *Downsizing human, bacterial, and viral proteins to short water-stable alpha helices that maintain biological potency.* Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11686-91.

98. Shinbara, K., et al., *Methodologies for Backbone Macrocyclic Peptide Synthesis Compatible With Screening Technologies.* Front Chem, 2020. **8**: p. 447.

99. de Araujo, A.D., H.T. Nguyen, and D.P. Fairlie, *Late-Stage Hydrocarbon Conjugation and Cyclisation in Synthetic Peptides and Proteins.* Chembiochem, 2021.

100. Boutureira, O. and G.J. Bernardes, *Advances in chemical protein modification.* Chem Rev, 2015. **115**(5): p. 2174-95.

101. Zhang, C., et al., *Arylation Chemistry for Bioconjugation.* Angew Chem Int Ed Engl, 2019. **58**(15): p. 4810-4839.

102. Wendt, M., et al., *Bicyclic beta-sheet mimetics that target the transcriptional coactivator beta-catenin and inhibit Wnt signaling.* Angew Chem Int Ed Engl, 2021.

103. Peraro, L., et al., *Diversity-Oriented Stapling Yields Intrinsically Cell-Penetrant Inducers of Autophagy.* J Am Chem Soc, 2017. **139**(23): p. 7792-7802.

104. Spokoyny, A.M., et al., *A perfluoroaryl-cysteine S(N)Ar chemistry approach to unprotected peptide stapling.* J Am Chem Soc, 2013. **135**(16): p. 5946-9.

105. Sutherland, B.P., et al., *On-Resin Macrocyclization of Peptides Using Vinyl Sulfonamides as a Thiol-Michael "Click" Acceptor.* Bioconjug Chem, 2018. **29**(12): p. 3987-3992.

106. Silva, M., et al., *Efficient Amino-Sulfhydryl Stapling on Peptides and Proteins Using Bifunctional NHS-Activated Acrylamides.* Angew Chem Int Ed Engl, 2021.

107. Derda, R. and M.R. Jafari, *Synthetic Cross-linking of Peptides: Molecular Linchpins for Peptide Cyclization.* Protein Pept Lett, 2018. **25**(12): p. 1051-1075.

108. Lian, W., et al., *Screening bicyclic peptide libraries for protein-protein interaction inhibitors: discovery of a tumor necrosis factor-alpha antagonist.* J Am Chem Soc, 2013. **135**(32): p. 11990-5.

109. Kubota, K., et al., *Palladium Oxidative Addition Complexes for Peptide and Protein Cross-linking.* J Am Chem Soc, 2018. **140**(8): p. 3128-3133.

110. Zhang, Y., et al., *Chemoselective Peptide Cyclization and Bicyclization Directly on Unprotected Peptides.* J Am Chem Soc, 2019. **141**(31): p. 12274-12279.

111. Pelay-Gimeno, M., et al., *In Situ Cyclization of Native Proteins: Structure-Based Design of a Bicyclic Enzyme.* Angew Chem Int Ed Engl, 2018. **57**(35): p. 11164-11170.

112. Reguera, L. and D.G. Rivera, *Multicomponent Reaction Toolbox for Peptide Macrocyclization and Stapling.* Chem Rev, 2019. **119**(17): p. 9836-9860.

113. Nguyen, G.K., et al., *Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis.* Nat Chem Biol, 2014. **10**(9): p. 732-8.

114. Albert, L. and O. Vazquez, *Photoswitchable peptides for spatiotemporal control of biological functions.* Chem Commun (Camb), 2019. **55**(69): p. 10192-10213.

115. Madden, M.M., et al., *Facile synthesis of stapled, structurally reinforced peptide helices via a photoinduced intramolecular 1,3-dipolar cycloaddition reaction.* Chem Commun (Camb), 2009(37): p. 5588-90.

116. Makwana, K.M. and R. Mahalakshmi, *Trp-Trp Cross-Linking: A Structure-Reactivity Relationship in the Formation and Design of Hyperstable Peptide beta-Hairpin and alpha-Helix Scaffolds.* Org Lett, 2015. **17**(10): p. 2498-501.

117. Stephanopoulos, N. and M.B. Francis, *Choosing an effective protein bioconjugation strategy.* Nat Chem Biol, 2011. **7**(12): p. 876-84.

118. Koniev, O. and A. Wagner, *Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation.* Chem Soc Rev, 2015. **44**(15): p. 5495-551.

119. Nolan, M.D. and E.M. Scanlan, *Applications of Thiol-Ene Chemistry for Peptide Science.* Front Chem, 2020. **8**: p. 583272.

120. Xuan, W., S. Shao, and P.G. Schultz, *Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains.* Angew Chem Int Ed Engl, 2017. **56**(18): p. 5096-5100.

121. Young, D.D. and P.G. Schultz, *Playing with the Molecules of Life.* ACS Chem Biol, 2018. **13**(4): p. 854-870.

122. Jagasia, R., et al., *Peptide cyclization and cyclodimerization by Cu(I)-mediated azide-alkyne cycloaddition.* J Org Chem, 2009. **74**(8): p. 2964-74.

123. Ingale, S. and P.E. Dawson, *On resin side-chain cyclization of complex peptides using CuAAC.* Org Lett, 2011. **13**(11): p. 2822-5.

124. Lundquist, J.T.t. and J.C. Pelletier, *A new tri-orthogonal strategy for peptide cyclization.* Org Lett, 2002. **4**(19): p. 3219-21.

125. Aimetti, A.A., et al., *On-resin peptide macrocyclization using thiol-ene click chemistry.* Chem Commun (Camb), 2010. **46**(23): p. 4061-3.

126. Malins, L.R., et al., *Peptide Macrocyclization Inspired by Non-Ribosomal Imine Natural Products.* J Am Chem Soc, 2017. **139**(14): p. 5233-5241.

127. Rivera, D.G., et al., *Peptide macrocyclization by transition metal catalysis.* Chem Soc Rev, 2020. **49**(7): p. 2039-2059.

128. Schafmeister, C.E., J. Po, and G.L. Verdine, *An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides.* Journal of the American Chemical Society, 2000. **122**(24): p. 5891-5892.

129. Verdine, G.L. and G.J. Hilinski, *Stapled peptides for intracellular drug targets.* Methods Enzymol, 2012. **503**: p. 3-33.

130. Li, H., R. Aneja, and I. Chaiken, *Click chemistry in peptide-based drug design.* Molecules, 2013. **18**(8): p. 9797-817.

131. Tan, Y.S., D.P. Lane, and C.S. Verma, *Stapled peptide design: principles and roles of computation.* Drug Discov Today, 2016. **21**(10): p. 1642-1653.

132. Bernal, F., et al., *Reactivation of the p53 tumor suppressor pathway by a stapled p53 peptide.* J Am Chem Soc, 2007. **129**(9): p. 2456-7.

133. Moellering, R.E., et al., *Direct inhibition of the NOTCH transcription factor complex.* Nature, 2009. **462**(7270): p. 182-8.

134. Chang, Y.S., et al., *Stapled alpha-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy.* Proc Natl Acad Sci U S A, 2013. **110**(36): p. E3445-54.

135. Mitra, S., et al., *Stapled peptide inhibitors of RAB25 target context-specific phenotypes in cancer.* Nat Commun, 2017. **8**(1): p. 660.

136. Meric-Bernstam, F., et al., *Phase I trial of a novel stapled peptide ALRN-6924 disrupting MDMX- and MDM2-mediated inhibition of WT p53 in patients with solid tumors and lymphomas.* Journal of Clinical Oncology, 2017. **35**(15_suppl): p. 2505-2505.

137. Blackwell, H.E. and R.H. Grubbs, *Highly Efficient Synthesis of Covalently Cross-Linked Peptide Helices by Ring-Closing Metathesis.* Angew Chem Int Ed Engl, 1998. **37**(23): p. 3281-3284.

138. Kim, Y.W., T.N. Grossmann, and G.L. Verdine, *Synthesis of all-hydrocarbon stapled alpha-helical peptides by ring-closing olefin metathesis.* Nat. Protoc., 2011. **6**.

139. Bird, G.H., et al., *Hydrocarbon double-stapling remedies the proteolytic instability of a lengthy peptide therapeutic.* Proc Natl Acad Sci U S A, 2010. **107**(32): p. 14093-8.

140. Walensky, L.D. and G.H. Bird, *Hydrocarbon-stapled peptides: principles, practice, and progress.* J Med Chem, 2014. **57**(15): p. 6275-88.

141. Baek, S., et al., *Structure of the stapled p53 peptide bound to Mdm2.* J Am Chem Soc, 2012. **134**(1): p. 103-6.

142. Speltz, T.E., et al., *Stapled Peptides with gamma-Methylated Hydrocarbon Chains for the Estrogen Receptor/Coactivator Interaction.* Angew Chem Int Ed Engl, 2016. **55**(13): p. 4252-5.

143. Montgomery, J.E., et al., *Versatile Peptide Macrocyclization with Diels-Alder Cycloadditions.* J Am Chem Soc, 2019. **141**(41): p. 16374-16381.

144. Chu, Q., et al., *Towards understanding cell penetration by stapled peptides.* MedChemComm, 2015. **6**(1): p. 111-119.

145. Derakhshankhah, H. and S. Jafari, *Cell penetrating peptides: A concise review with emphasis on biomedical applications.* Biomed Pharmacother, 2018. **108**: p. 1090-1096.

146. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876-90.

147. Moerke, N.J., *Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding.* Curr Protoc Chem Biol, 2009. **1**(1): p. 1-15.

148. Rich, R.L. and D.G. Myszka, *Grading the commercial optical biosensor literature-Class of 2008: 'The Mighty Binders'.* J Mol Recognit, 2010. **23**(1): p. 1-64.

149. Rich, R.L., et al., *A global benchmark study using affinity-based biosensors.* Anal Biochem, 2009. **386**(2): p. 194-216.

150. Kuhn, T., et al., *Protein diffusion in mammalian cell cytoplasm.* PLoS One, 2011. **6**(8): p. e22962.

151. Cheng, F., et al., *Comprehensive characterization of protein-protein interactions perturbed by disease mutations.* Nat Genet, 2021. **53**(3): p. 342-353.

152. Horne, W.S. and T.N. Grossmann, *Proteomimetics as protein-inspired scaffolds with defined tertiary folding patterns.* Nat Chem, 2020. **12**(4): p. 331-337.

153. Bionda, N. and R. Fasan, *Peptidomimetics of α-Helical and β-Strand Protein Binding Epitopes.* Small Molecule Medicinal Chemistry, 2015.

154. Walensky, L.D., *Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix.* Science, 2004. **305**.

155. Bernal, F., et al., *Reactivation of the p53 tumor suppressor pathway by a stapled p53 peptide.* J. Am. Chem. Soc., 2007. **129**.

156. Grossmann, T.N., et al., *Inhibition of oncogenic Wnt signaling through direct targeting of beta-catenin.* Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17942-7.

157. Casanova, J.E., *Association of Rab25 and Rab11a with the apical recycling system of polarized Madin-Darby canine kidney cells.* Mol. Biol. Cell, 1999. **10**.

158. Wang, X., et al., *Regulation of vesicle trafficking in madin-darby canine kidney cells by Rab11a and Rab25.* J. Biol. Chem., 2000. **275**.

159. Cheng, K.W., *The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers.* Nat. Med., 2004. **10**.

160. Cheng, K.W., et al., *Emerging role of RAB GTPases in cancer and human disease.* Cancer Res., 2005. **65**.

161. Wheeler, D.B., et al., *Identification of an oncogenic RAB protein.* Science, 2015. **350**.

162. He, H., *Identification and characterization of nine novel human small GTPases showing variable expressions in liver cancer tissues.* Gene Exp., 2002. **10**.

163. Cheng, K.W., *Rab25 increases cellular ATP and glycogen stores protecting cancer cells from bioenergetic stress.* EMBO Mol. Med., 2012. **4**.

164. Mitra, S., *Rab25 acts as an oncogene in luminal B breast cancer and is causally associated with Snail driven EMT.* Oncotarget, 2016. **7**.

165. Goldenring, J.R. and K.T. Nam, *Rab25 as a tumour suppressor in colon carcinogenesis.* Br. J. Cancer, 2011. **104**.

166. Nam, K.T., *Loss of Rab25 promotes the development of intestinal neoplasia in mice and is associated with human colorectal adenocarcinomas.* J. Clin. Invest., 2010. **120**.

167. Hales, C.M., *Identification and characterization of a family of Rab11-interacting proteins.* J. Biol. Chem., 2001. **276**.

168. Lindsay, A.J. and M.W. McCaffrey, *Rab11-FIP2 functions in transferrin recycling and associates with endosomal membranes via its COOH-terminal domain.* J. Biol. Chem., 2002. **277**.

169. Wallace, D.M., et al., *Rab11-FIP4 interacts with Rab11 in a GTP-dependent manner and its overexpression condenses the Rab11 positive compartment in HeLa cells.* Biochem. Biophys. Res. Commun., 2002. **299**.

170. Lall, P., *Structural and functional analysis of FIP2 binding to the endosome-localised Rab25 GTPase.* Biochim. Biophys. Acta, 2013. **1834**.

171. Caswell, P.T., *Rab-coupling protein coordinates recycling of alpha5beta1 integrin and EGFR1 to promote cell migration in 3D microenvironments.* J. Cell Biol., 2008. **183**.

172. Muller, P.A., *Mutant p53 drives invasion by promoting integrin recycling.* Cell, 2009. **139**.

173. LaRochelle, J.R., et al., *Fluorescence correlation spectroscopy reveals highly efficient cytosolic delivery of certain penta-arg proteins and stapled peptides.* J. Am. Chem. Soc., 2015. **137**.

174. Cromm, P.M., *Protease-resistant and cell-permeable double-stapled peptides targeting the Rab8a GTPase.* ACS Chem. Biol., 2016. **11**.

175. Ullman, E.F., *Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence.* Proc. Natl Acad. Sci. USA, 1994. **91**.

176. Subramanian, A., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc. Natl Acad. Sci. USA, 2005. **102**.

177. Cheng, J.M., *Tumor suppressor function of Rab25 in triple-negative breast cancer.* Int. J. Cancer, 2010. **126**.

178. Martinez Molina, D., *Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay.* Science, 2013. **341**.

179. Lu, S., et al., *Drugging Ras GTPase: a comprehensive mechanistic and signaling structural view.* Chem Soc Rev, 2016. **45**(18): p. 4929-52.

180. Gray, J.L., F. von Delft, and P.E. Brennan, *Targeting the Small GTPase Superfamily through Their Regulatory Proteins.* Angew Chem Int Ed Engl, 2020. **59**(16): p. 6342-6366.

181. Spiegel, J., *Direct targeting of Rab-GTPase-effector interactions.* Angew. Chem. Int. Ed. Engl., 2014. **53**.

182. Moellering, R.E. and B.F. Cravatt, *How chemoproteomics can enable drug discovery and development.* Chem. Biol., 2012. **19**.

183. Rainero, E., *Diacylglycerol kinase alpha controls RCP-dependent integrin trafficking to promote invasive migration.* J. Cell Biol., 2012. **196**.

184. Merritt, H.I., N. Sawyer, and P.S. Arora, *Bent Into Shape: Folded Peptides to Mimic Protein Structure and Modulate Protein Function.* Pept Sci (Hoboken), 2020. **112**(1).

185. Gavenonis, J., et al., *Comprehensive analysis of loops at protein-protein interfaces for macrocycle design.* Nat Chem Biol, 2014. **10**(9): p. 716-22.

186. Bock, J.E., J. Gavenonis, and J.A. Kritzer, *Getting in shape: controlling peptide bioactivity and bioavailability using conformational constraints.* ACS Chem Biol, 2013. **8**(3): p. 488-499.

187. Marti-Centelles, V., et al., *Macrocyclization Reactions: The Importance of Conformational, Configurational, and Template-Induced Preorganization.* Chem Rev, 2015. **115**(16): p. 8736-834.

188. Rojas, A.J., et al., *Divergent unprotected peptide macrocyclisation by palladium-mediated cysteine arylation.* Chem Sci, 2017. **8**(6): p. 4257-4263.

189. Diels, O. and K. Alder, *Synthesen in der hydroaromatischen Reihe.* Justus Liebig's Annalen der Chemie, 1928. **460**(1): p. 98-122.

190. Corey, E.J., *Catalytic enantioselective Diels--Alder reactions: methods, mechanistic fundamentals, pathways, and applications.* Angew Chem Int Ed Engl, 2002. **41**(10): p. 1650-67.

191. Oikawa, H., *Nature's Strategy for Catalyzing Diels-Alder Reaction.* Cell Chem Biol, 2016. **23**(4): p. 429-30.

192. Rideout, D.C. and R. Breslow, *Hydrophobic acceleration of Diels-Alder reactions.* Journal of the American Chemical Society, 2002. **102**(26): p. 7816-7817.

193. Woodward, R.B., et al., *The Total Synthesis of Steroids1.* Journal of the American Chemical Society, 2002. **74**(17): p. 4223-4251.

194. Chu-Moyer, M.Y., S.J. Danishefsky, and G.K. Schulte, *Total Synthesis of (.+-.)-Myrocin C.* Journal of the American Chemical Society, 2002. **116**(25): p. 11213-11228.

195. Gao, L., et al., *FAD-dependent enzyme-catalysed intermolecular [4+2] cycloaddition in natural product biosynthesis.* Nat Chem, 2020. **12**(7): p. 620-628.

196. Yousaf, M.N. and M. Mrksich, *Diels–Alder Reaction for the Selective Immobilization of Protein to Electroactive Self-Assembled Monolayers.* Journal of the American Chemical Society, 1999. **121**(17): p. 4286-4287.

197. de Araujo, A.D., et al., *Diels-Alder ligation and surface immobilization of proteins.* Angew Chem Int Ed Engl, 2005. **45**(2): p. 296-301.

198. Borsenberger, V. and S. Howorka, *Diene-modified nucleotides for the Diels-Alder-mediated functional tagging of DNA.* Nucleic Acids Res, 2009. **37**(5): p. 1477-85.

199. St Amant, A.H., et al., *Tuning the Diels-Alder Reaction for Bioconjugation to Maleimide Drug-Linkers.* Bioconjug Chem, 2018. **29**(7): p. 2406-2414.

200. Pierschbacher, M.D. and E. Ruoslahti, *Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule.* Nature, 1984. **309**(5963): p. 30-3.

201. Ruoslahti, E., *RGD and other recognition sequences for integrins.* Annu Rev Cell Dev Biol, 1996. **12**: p. 697-715.

202. Ley, K., et al., *Integrin-based therapeutics: biological basis, clinical use and new drugs.* Nat Rev Drug Discov, 2016. **15**(3): p. 173-83.

203. Renault, K., et al., *Covalent Modification of Biomolecules through Maleimide-Based Labeling Strategies.* Bioconjug Chem, 2018. **29**(8): p. 2497-2513.

204. Tong, X. and A. Hong, *A Highly Efficient Method for Synthesis of Fmoc-Lysine(Mmt)-OH.* Lebl M., Houghten R.A. (eds) Peptides: The Wave of the Future. American Peptide Symposia, vol 7. , 2001: p. 73-74.

205. Hyde, C., T. Johnson, and R.C. Sheppard, *Internal aggregation during solid phase peptide synthesis. Dimethyl sulfoxide as a powerful dissociating solvent.* Journal of the Chemical Society, Chemical Communications, 1992(21).

206. Miranda, L.P. and P.F. Alewood, *Accelerated chemical synthesis of peptides and small proteins.* Proc Natl Acad Sci U S A, 1999. **96**(4): p. 1181-6.

207. Huhtiniemi, T., et al., *N(epsilon)-Modified lysine containing inhibitors for SIRT1 and SIRT2.* Bioorg Med Chem, 2010. **18**(15): p. 5616-25.

208. Danhier, F., A. Le Breton, and V. Preat, *RGD-based strategies to target alpha(v) beta(3) integrin in cancer therapy and diagnosis.* Mol Pharm, 2012. **9**(11): p. 2961-73.

209. Kumagai, H., et al., *Effect of cyclic RGD peptide on cell adhesion and tumor metastasis.* Biochem Biophys Res Commun, 1991. **177**(1): p. 74-82.

210. *Molecular Operating Environment (MOE)*, 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2021.

211. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

212. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

213. Bullock, B.N., A.L. Jochim, and P.S. Arora, *Assessing helical protein interfaces for inhibitor design.* J Am Chem Soc, 2011. **133**(36): p. 14220-3.

214. Kim, Y.W. and G.L. Verdine, *Stereochemical effects of all-hydrocarbon tethers in i,i+4 stapled peptides.* Bioorg Med Chem Lett, 2009. **19**(9): p. 2533-6.

215. Mangold, S.L., D.J. O'Leary, and R.H. Grubbs, *Z-Selective olefin metathesis on peptides: investigation of side-chain influence, preorganization, and guidelines in substrate selection.* J Am Chem Soc, 2014. **136**(35): p. 12469-78.

216. Phillips, C., et al., *Design and structure of stapled peptides binding to estrogen receptors.* J Am Chem Soc, 2011. **133**(25): p. 9696-9.

217. Speltz, T.E., et al., *A "cross-stitched" peptide with improved helicity and proteolytic stability.* Org Biomol Chem, 2018. **16**(20): p. 3702-3706.

218. Warnmark, A., et al., *Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha.* J Biol Chem, 2002. **277**(24): p. 21862-8.

219. Patgiri, A., A.L. Jochim, and P.S. Arora, *A hydrogen bond surrogate approach for stabilization of short peptide sequences in alpha-helical conformation.* Acc Chem Res, 2008. **41**(10): p. 1289-300.

220. Mironov, V.A., E.V. Sobolev, and A.N. Elizarova, *Some general characteristic properties of substituted cyclopentadienes.* Tetrahedron, 1963. **19**(12): p. 1939-1958.

221. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides.* Proc Natl Acad Sci U S A, 1989. **86**(14): p. 5286-90.

222. Koehler, K.C., et al., *Development of a Maleimide Amino Acid for Use as a Tool for Peptide Conjugation and Modification.* International Journal of Peptide Research and Therapeutics, 2013. **19**(3): p. 265-274.

223. St Amant, A.H., et al., *A Diene-Containing Noncanonical Amino Acid Enables Dual Functionality in Proteins: Rapid Diels-Alder Reaction with Maleimide or*

Proximity-Based Dimerization. Angew Chem Int Ed Engl, 2019. **58**(25): p. 8489-8493.

224. Lee, K.J., D. Kang, and H.S. Park, *Site-Specific Labeling of Proteins Using Unnatural Amino Acids.* Mol Cells, 2019. **42**(5): p. 386-396.

225. Kangawa, K. and H. Matsuo, *Purification and complete amino acid sequence of alpha-human atrial natriuretic polypeptide (alpha-hANP).* Biochem Biophys Res Commun, 1984. **118**(1): p. 131-9.

226. Zhang, J., Z. Zhao, and J. Wang, *Natriuretic peptide receptor A as a novel target for cancer.* World J Surg Oncol, 2014. **12**: p. 174.

227. Nojiri, T., et al., *Atrial natriuretic peptide prevents cancer metastasis through vascular endothelial cells.* Proc Natl Acad Sci U S A, 2015. **112**(13): p. 4086-91.

228. Nojiri, T., et al., *Effect of low-dose human atrial natriuretic peptide on postoperative atrial fibrillation in patients undergoing pulmonary resection for lung cancer: a double-blind, placebo-controlled study.* J Thorac Cardiovasc Surg, 2012. **143**(2): p. 488-94.

229. Staedtke, V., et al., *Disruption of a self-amplifying catecholamine loop reduces cytokine release syndrome.* Nature, 2018. **564**(7735): p. 273-277.

230. Mezo, A.R., et al., *Atrial natriuretic peptide-Fc, ANP-Fc, fusion proteins: semisynthesis, in vitro activity and pharmacokinetics in rats.* Bioconjug Chem, 2012. **23**(3): p. 518-26.

231. Hansen, L.H., et al., *Discovery of O-glycans on atrial natriuretic peptide (ANP) that affect both its proteolytic degradation and potency at its cognate receptor.* J Biol Chem, 2019. **294**(34): p. 12567-12578.

232. He, X.L., A. Dukkipati, and K.C. Garcia, *Structural determinants of natriuretic peptide receptor specificity and degeneracy.* J Mol Biol, 2006. **361**(4): p. 698-714.

233. Vayer, M., et al., *Acid-catalysed intramolecular addition of beta-ketoesters to 1,3-dienes.* Org Biomol Chem, 2017. **15**(3): p. 584-588.

234. Allen, V.C., et al., *A simple network of synthetic replicators can perform the logical OR operation.* Org Lett, 2010. **12**(9): p. 1920-3.

235. Ramachandran, G.N., R. Chandrasekaran, and K.D. Kopple, *Variation of the NH-C alpha-H coupling constant with dihedral angle in the NMR spectra of peptides.* Biopolymers, 1971. **10**(11): p. 2113-31.

236. Fuchs, S., et al., *Proline primed helix length as a modulator of the nuclear receptor-coactivator interaction.* J Am Chem Soc, 2013. **135**(11): p. 4364-71.

237. Fanning, S.W., et al., *Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation.* Elife, 2016. **5**.

238. Minor, W., et al., *HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 8): p. 859-66.

239. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

240. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

241. Nettles, K.W., et al., *NFkappaB selectivity of estrogen receptor ligands revealed by comparative crystallographic analyses.* Nat Chem Biol, 2008. **4**(4): p. 241-7.

242. Keni, R., et al., *COVID-19: Emergence, Spread, Possible Treatments, and Global Burden.* Frontiers in Public Health, 2020. **8**(216).

243. Center, J.H.C.R., *Worldwide cases.* Johns Hopkins University & Medicine, 2021.

244. Baden, L.R., et al., *Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine.* N Engl J Med, 2021. **384**(5): p. 403-416.

245. Voysey, M., et al., *Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK.* The Lancet, 2021. **397**(10269): p. 99-111.

246. Polack, F.P., et al., *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.* N Engl J Med, 2020. **383**(27): p. 2603-2615.

247. Wrapp, D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.* Science, 2020. **367**(6483): p. 1260-1263.

248. Dai, L. and G.F. Gao, *Viral targets for vaccines against COVID-19.* Nat Rev Immunol, 2021. **21**(2): p. 73-82.

249. Wu, K., et al., *Serum Neutralizing Activity Elicited by mRNA-1273 Vaccine.* New England Journal of Medicine, 2021.

250. Liu, Y., et al., *Neutralizing Activity of BNT162b2-Elicited Serum.* New England Journal of Medicine, 2021.

251. Baum, A., et al., *Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies.* Science, 2020. **369**(6506): p. 1014-1018.

252. Li, Q., et al., *The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity.* Cell, 2020. **182**(5): p. 1284-1294 e9.

253. von Grotthuss, M., L.S. Wyrwicz, and L. Rychlewski, *mRNA cap-1 methyltransferase in the SARS genome.* Cell, 2003. **113**(6): p. 701-2.

254. Decroly, E., et al., *Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex.* PLoS Pathog, 2011. **7**(5): p. e1002059.

255. Bouvet, M., et al., *Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes.* J Biol Chem, 2014. **289**(37): p. 25783-96.

256. Chen, Y. and D. Guo, *Molecular mechanisms of coronavirus RNA capping and methylation.* Virol Sin, 2016. **31**(1): p. 3-11.

257. Blanco-Melo, D., et al., *Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19.* Cell, 2020. **181**(5): p. 1036-1045.e9.

258. Lazear, H.M., J.W. Schoggins, and M.S. Diamond, *Shared and Distinct Functions of Type I and Type III Interferons.* Immunity, 2019. **50**(4): p. 907-923.

259. Smith, E.C., et al., *Mutations in coronavirus nonstructural protein 10 decrease virus replication fidelity.* J Virol, 2015. **89**(12): p. 6418-26.

260. Rosas-Lemus, M., et al., *High-resolution structures of the SARS-CoV-2 2′-O-methyltransferase reveal strategies for structure-based inhibitor design.* Science Signaling, 2020. **13**(651): p. eabe1202.

261. Wang, Y., et al., *Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide In Vitro and In Vivo To Reduce Replication and Pathogenesis.* J Virol, 2015. **89**(16): p. 8416-27.

262. Ke, M., et al., *Short peptides derived from the interaction domain of SARS coronavirus nonstructural protein nsp10 can suppress the 2'-O-methyltransferase activity of nsp10/nsp16 complex.* Virus Res, 2012. **167**(2): p. 322-8.

263. Frankel, A.D., J.M. Berg, and C.O. Pabo, *Metal-dependent folding of a single zinc finger from transcription factor IIIA.* Proc Natl Acad Sci U S A, 1987. **84**(14): p. 4841-5.

264. Lombardo, C.M., et al., *Design and Structure Determination of a Composite Zinc Finger Containing a Nonpeptide Foldamer Helical Domain.* J Am Chem Soc, 2019. **141**(6): p. 2516-2525.

265. Cornelison, G.L. and S.J. Mihic, *Contaminating levels of zinc found in commonly-used labware and buffers affect glycine receptor currents.* Brain Res Bull, 2014. **100**: p. 1-5.

266. Joseph, J.S., et al., *Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs.* J Virol, 2006. **80**(16): p. 7894-901.

267. Su, D., et al., *Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10.* J Virol, 2006. **80**(16): p. 7902-8.

268. Rogstam, A., et al., *Crystal Structure of Non-Structural Protein 10 from Severe Acute Respiratory Syndrome Coronavirus-2.* Int J Mol Sci, 2020. **21**(19).

269. Lundblad, J.R., M. Laurance, and R.H. Goodman, *Fluorescence polarization analysis of protein-DNA and protein-protein interactions.* Mol Endocrinol, 1996. **10**(6): p. 607-12.

270. Lea, W.A. and A. Simeonov, *Fluorescence polarization assays in small molecule screening.* Expert Opin Drug Discov, 2011. **6**(1): p. 17-32.

271. Patel, S.G., et al., *Cell-penetrating peptide sequence and modification dependent uptake and subcellular distribution of green florescent protein in different cell lines.* Scientific Reports, 2019. **9**(1): p. 6298.

272. Milletti, F., *Cell-penetrating peptides: classes, origin, and current landscape.* Drug Discov Today, 2012. **17**(15-16): p. 850-60.

273. Jones, A.T. and E.J. Sayers, *Cell entry of cell penetrating peptides: tales of tails wagging dogs.* J Control Release, 2012. **161**(2): p. 582-91.

274. van den Berg, A. and S.F. Dowdy, *Protein transduction domain delivery of therapeutic macromolecules.* Curr Opin Biotechnol, 2011. **22**(6): p. 888-93.

275. Frankel, A.D. and C.O. Pabo, *Cellular uptake of the tat protein from human immunodeficiency virus.* Cell, 1988. **55**(6): p. 1189-1193.

276. Green, M. and P.M. Loewenstein, *Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein.* Cell, 1988. **55**(6): p. 1179-1188.

277. Chauhan, A., et al., *The taming of the cell penetrating domain of the HIV Tat: myths and realities.* Journal of controlled release : official journal of the Controlled Release Society, 2007. **117**(2): p. 148-162.

278. Fominaya, J., J. Bravo, and A. Rebollo, *Strategies to stabilize cell penetrating peptides for in vivo applications.* Ther Deliv, 2015. **6**(10): p. 1171-94.

279. Mueller, J., et al., *Comparison of cellular uptake using 22 CPPs in 4 different cell lines.* Bioconjug Chem, 2008. **19**(12): p. 2363-74.

280. Steinauer, A., et al., *HOPS-dependent endosomal fusion required for efficient cytosolic delivery of therapeutic peptides and small proteins.* Proc Natl Acad Sci U S A, 2019. **116**(2): p. 512-521.

281. Sahni, A., Z. Qian, and D. Pei, *Cell-Penetrating Peptides Escape the Endosome by Inducing Vesicle Budding and Collapse.* ACS Chem Biol, 2020. **15**(9): p. 2485-2492.

282. Peraro, L., et al., *Cell Penetration Profiling Using the Chloroalkane Penetration Assay.* J Am Chem Soc, 2018. **140**(36): p. 11360-11369.

283. Verdurmen, W.P.R., M. Mazlami, and A. Pluckthun, *A Biotin Ligase-Based Assay for the Quantification of the Cytosolic Delivery of Therapeutic Proteins.* Methods Mol Biol, 2017. **1575**: p. 223-236.

284. Burlina, F., et al., *A direct approach to quantification of the cellular uptake of cell-penetrating peptides using MALDI-TOF mass spectrometry.* Nat Protoc, 2006. **1**(1): p. 200-5.

285. Stanton, B.Z., E.J. Chory, and G.R. Crabtree, *Chemically induced proximity in biology and medicine.* Science, 2018. **359**(6380).

286. Keppler, A., et al., *A general method for the covalent labeling of fusion proteins with small molecules in vivo.* Nat Biotechnol, 2003. **21**(1): p. 86-9.

287. Juillerat, A., et al., *Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo.* Chemistry & Biology, 2003. **10**(4): p. 313-317.

288. Yang, G., et al., *Genetic targeting of chemical indicators in vivo.* Nat Methods, 2015. **12**(2): p. 137-9.

289. Wu, H., et al., *Manipulation of Glutathione-Mediated Degradation of Thiol-Maleimide Conjugates.* Bioconjug Chem, 2018. **29**(11): p. 3595-3605.

290. Gajski, G. and V. Garaj-Vrhovac, *Melittin: a lytic peptide with anticancer properties.* Environ Toxicol Pharmacol, 2013. **36**(2): p. 697-705.

291. Sato, H. and J.B. Feix, *Peptide-membrane interactions and mechanisms of membrane destruction by amphipathic alpha-helical antimicrobial peptides.* Biochim Biophys Acta, 2006. **1758**(9): p. 1245-56.

292. Vasco, A.V., et al., *A Multicomponent Stapling Approach to Exocyclic Functionalized Helical Peptides: Adding Lipids, Sugars, PEGs, Labels, and Handles to the Lactam Bridge.* Bioconjug Chem, 2019. **30**(1): p. 253-259.

293. Beaulieu, M.E., et al., *Intrinsic cell-penetrating activity propels Omomyc from proof of concept to viable anti-MYC therapy.* Sci Transl Med, 2019. **11**(484).

294. Cheng, R.P., S.H. Gellman, and W.F. DeGrado, *beta-Peptides: from structure to function.* Chem Rev, 2001. **101**(10): p. 3219-32.

295. Yamashita, H., et al., *Development of helix-stabilized cell-penetrating peptides containing cationic alpha,alpha-disubstituted amino acids as helical promoters.* Bioorg Med Chem, 2017. **25**(6): p. 1846-1851.