2017

Optochemical Tools For Protein Dimerization In Living Cells

Chanat Aonbangkhen

*University of Pennsylvania*, chanataon@gmail.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the [Cell Biology Commons](https://repository.upenn.edu/edissertations), and the [Organic Chemistry Commons](https://repository.upenn.edu/edissertations)

**Recommended Citation**
https://repository.upenn.edu/edissertations/2738

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/2738
For more information, please contact repository@pobox.upenn.edu.
Optochemical Tools For Protein Dimerization In Living Cells

Abstract
Fundamental biological processes including cell division, migration, and death, are driven by protein interactions. Regulation of protein localization is one of the mechanisms cells utilize to control cellular events with high spatial and temporal precision. Therefore, several techniques have been developed to provide control of protein interactions and localization. A number of elegant approaches employ naturally light-responsive proteins, also known as optogenetics, to reversibly induce protein-protein binding interactions with subcellular precision. However, the application of these light-inducible protein systems to various intracellular locations beyond the plasma membrane has been limited. Moreover, to achieve sustained interactions in some applications, most of these optogenetic systems require continuous illumination, increasing the risk of phototoxicity. Another robust and widely utilized technique to control protein interactions via small molecules is the chemically-induced dimerization (CID) of proteins; the most classic example of this technique being rapamycin-induced dimerization. However, the lack of spatiotemporal control and reversibility in this system has necessitated the development of new dimerizers in the past two decades. By combining light-inducible features with the CID technique, we have created a novel platform to rapidly and reversibly induce protein dimerization using light with high specificity in living cells. This is accomplished with subcellular spatiotemporal resolution using a series of novel, cell-permeable, photoactivatable, and photocleavable chemical dimerizers. The modular design of our system has allowed us to tailor the properties of our molecules for studying various protein functions and biological pathways inside living cells. Furthermore, we demonstrate the utility of our system by applying it to manipulate dynamic biological events including organelle transport and spindle assembly checkpoint. This work establishes a foundation for optogenetic control over protein function and highlights the advantages of a hybrid chemical and genetic approach. We envision our tools to be readily adapted to experimentally probe complex signaling
networks and other cellular processes that depend upon spatiotemporal regulation of protein localization on biologically-relevant timescales.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Chemistry

First Advisor
David M. Chenoweth

Keywords
Cell signaling, Chemical dimerizers, Optogenetics, Photocaged dimerizers, Photocleavable dimerizers, Protein dimerization

Subject Categories
Cell Biology | Chemistry | Organic Chemistry

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2738
OPTOCHEMICAL TOOLS FOR PROTEIN DIMERIZATION IN LIVING CELLS

Chanat Aonbangkhen

A DISSERTATION

in

Chemistry

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

David M. Chenoweth, Associate Professor of Chemistry

Graduate Group Chairperson

Gary A. Molander, Hirschmann-Makineni Professor of Chemistry

Dissertation Committee:

Jeffrey D. Winkler, Merriam Professor of Chemistry

William P. Dailey, Associate Professor of Chemistry

Virgil Percec, P. Roy Vagelos Professor of Chemistry

Michael A. Lampson, Associate Professor of Biology and Graduate Chair in Biology
DEDICATION

To my family and teachers…
ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my PhD advisor, Prof. David Chenoweth, for the opportunity to join his lab and work on a variety of projects, especially the protein dimerization project that set the stage for my future career in the field of chemical biology. I still remember the day when he came to ask me in the lab whether I was interested in working with Ed, who initiated the project with him. Without his broad interests and generosity to open the door for this project, I would not have been involved in all of this TMP-Halo journey. I would also like to extend my appreciation to Prof. Michael Lampson, who truly was my second PhD advisor on the biology part of this project. He was always nice, calm and caring about what I was doing to ensure that I was on the right track. I am absolutely thankful for his kindness and letting me learn biology skills and experiments in his lab, especially the cell culture and confocal imaging, which are very useful skills that can be applied to a broad scope of research. Both of their tremendous support and continued guidance throughout my time in grad school and research led me to the success of obtaining my PhD and the advancement of my career as a scientist. They have shaped me significantly into who I am today. In addition, I would like to thank my undergrad research advisor, Dr. Pitak Chuawong, who was also very supportive and inspired me to enter the chemical biology world by stepping out of my comfort zone (pure organic chemistry).

I would also like to thank my committee members: Prof. Jeffrey Winkler and Prof. William Dailey for their guidance and helpful discussions not only about research but also on general topics such as how to survive in grad school. These discussions were really crucial to international students including myself, who was new to everything here at Penn. I learned a lot as a TA for Prof. Dailey how to deal with pre-med students. I thank them for their support in my applications for a postdoc position as well. Moreover, I would like to thank Prof. Madeleine Joullie and Prof. E. James Petersson for our fantastic joint group meetings in the past 4 years.

I give special thanks to all the past and present members of the Chenoweth and the Lampson labs. Without all of them, I would not have come this far. I totally enjoyed interacting, hanging out, having conversations and discussions about science and non-science stuff, having lunch, dinner, and coffee, with everyone. I had so much fun experiences in and outside of the labs mainly due to these amazing people. I would like to mention everyone but due to limited space, I would highlight some people whom I closely worked with. First, Dr. Edward Ballister, thanks for his friendship and mentorship in guiding and helping me learn all the aspects of the project. Our first paper highly motivated me to continue doing science every day. He was probably the only
person who truly understood how it’s like to work in both chemistry and biology labs simultaneously. I hope we could collaborate again in the future to do more great science. Next, I’d like to thank Alyssa Mayo, for her help with the cell culture, microscope, and fixing cells, as well as Dr. Abram Calderon, for his help with cloning, making plasmids, and stable cell lines. I also enjoyed hearing about his PhD life in Germany. A huge thank to Dr. Huaiying Zhang, my second bio-partner. I really appreciated her hard work to push our optogenetic paper to be published. We went from V9 to V19 even after it was accepted with minor revisions. It’s very impressive how much energy and enthusiasm she had for science even though she was super busy taking care of her newborn baby (super mom). I also thank Dr. Takashi Akera, who always smiled no matter what happened, had faith in my dimerizers, and got CTH to work in oocytes. He had magic hands for cloning and had no need for heat-shock (according to Alper). I very much enjoyed our discussions and also cheesesteak with Yuriko and Chizuru! I would like to give a big hug and thanks to Dr. Stephanie Barros, my best American friend and deskmate at Penn. She was also my English teacher when I was taking coursework. Thanks for countless numbers of essay proof-readings. I would also like to thank Alex Kasznel, Sam Melton, and Dan Wu, too, for being such great friends. I really appreciated their time and effort in carefully reading my papers and the whole thesis, correcting errors, and giving me feedback. Particularly, Dan, who was my chemistry partner. He was really an optimistic person and supportive friend, who always encouraged me in everything, especially going for Postdoc at Harvard. I will miss you, buddy! All other friends and collaborators in both Chemistry and Biology Departments were also vital to me since they made labs a great place and joyful environment for work in the past 5 years. I will definitely miss talking to all of them.

Additionally, my Thai friends in Philadelphia and at Rutgers played a significant part of my PhD journey. P’s Ae, Kla, Zan, Amm, Nook, Pick, Noang, Nok, Nui, Sara, NaamTaan, Wan, Tee, Gift, Park, Gluay, Jiab, etc; they made Penn and Philly feel like home ever since the first day I arrived with warm welcoming. Although I’d never met them before, they sincerely helped me with everything from day 0 (preparing to come to Penn) until my last day (moving out). My life would have been much more difficult without their companionship and substantial support.

I am so humbled and very fortunate to be surrounded by terrific people who made my 5-year PhD at Penn the most memorable experience of my life. Words could not simply express how much I overwhelmingly felt appreciated and extremely thankful for knowing everyone whom I have interacted with (both covalently and non-covalently). Last but not least, I would like to thank the most, my family, my mom, dad, sister, and Blacky, who have been everything to me. I love you all so much and am looking forward to getting together again soon.
ABSTRACT

OPTOCHEMICAL TOOLS FOR PROTEIN DIMERIZATION IN LIVING CELLS

Chanat Aonbangkhen
David M. Chenoweth

Fundamental biological processes including cell division, migration, and death, are driven by protein interactions. Regulation of protein localization is one of the mechanisms cells utilize to control cellular events with high spatial and temporal precision. Therefore, several techniques have been developed to provide control of protein interactions and localization. A number of elegant approaches employ naturally light-responsive proteins, also known as optogenetics, to reversibly induce protein–protein binding interactions with subcellular precision. However, the application of these light-inducible protein systems to various intracellular locations beyond the plasma membrane has been limited. Moreover, to achieve sustained interactions in some applications, most of these optogenetic systems require continuous illumination, increasing the risk of phototoxicity. Another robust and widely utilized technique to control protein interactions via small molecules is the chemically-induced dimerization (CID) of proteins; the most classic example of this technique being rapamycin-induced dimerization. However, the lack of spatiotemporal control and reversibility in this system has necessitated the development of new dimerizers in the past two decades. By combining light-inducible features with the CID technique, we have created a novel
platform to rapidly and reversibly induce protein dimerization using light with high specificity in living cells. This is accomplished with subcellular spatiotemporal resolution using a series of novel, cell-permeable, photoactivatable, and photocleavable chemical dimerizers. The modular design of our system has allowed us to tailor the properties of our molecules for studying various protein functions and biological pathways inside living cells. Furthermore, we demonstrate the utility of our system by applying it to manipulate dynamic biological events including organelle transport and spindle assembly checkpoint. This work establishes a foundation for optogenetic control over protein function and highlights the advantages of a hybrid chemical and genetic approach. We envision our tools to be readily adapted to experimentally probe complex signaling networks and other cellular processes that depend upon spatiotemporal regulation of protein localization on biologically-relevant timescales.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS..............................................................................................................iv  
ABSTRACT.....................................................................................................................................vi  
LIST OF TABLES..........................................................................................................................xi  
LIST OF SCHEMES.........................................................................................................................xii  
LIST OF ILLUSTRATIONS..............................................................................................................xiii  
ABBREVIATIONS..............................................................................................................................xv  

Chapter 1 Research Overview and Literature Review.................................................................1  
Introduction to Proteins and Protein-Protein Interactions .........................................................2  
Controlling Protein-Protein Interactions via Chemically-Induced Dimerization (CID)..............7  
Applications of CID.......................................................................................................................13  
Chemical Dimerizers Developed to Date.....................................................................................20  
Light-Controlled Tools for Biology ............................................................................................38  
The Hook Effect............................................................................................................................48  
The Development of Our Novel System.....................................................................................50  
References........................................................................................................................................53  

Chapter 2 Rational Design of a Novel Light-Inducible Dimerization System.........................64  
The Modular Design.......................................................................................................................66  
Synthesis of the First Generation Photocaged Dimerizer...........................................................71
The NVOC Photolabile Protecting Group ................................................................. 76
The Linker or Spacer ................................................................................................. 78
The Non-caged Ligand ............................................................................................... 80
The Caged Ligand ..................................................................................................... 83
Biological Results and Discussion ............................................................................ 85
References .................................................................................................................. 92

**Chapter 3 Development of a New Photocaged and Photocleavable Dimerizers** ...... 93
Introduction to a New Photocaged Dimerizer and the Biological Results ................. 95
The Photocleavage Mechanism of DEACM ............................................................. 96
Introduction to a New Photocleavable Dimerizer and the Biological Results .......... 102
Synthesis of the New Photocaged Dimerizer CTH ................................................. 107
Synthesis of the New Photocleavable Dimerizer TNH ............................................. 110
References ................................................................................................................ 112

**Chapter 4 Development of a New Photocaged-Photocleavable Dimerizer** ....... 113
Introduction to a New Light-Controlled ON-OFF Protein Dimerization System .... 115
The Rational Design ................................................................................................. 117
Synthesis of the New Dimerizer ............................................................................... 120
Biological Results and Discussion .......................................................................... 127
Conclusions .............................................................................................................. 143
Concluding Remarks and Future Directions ......................................................... 144
References........................................................................................................................................148

Chapter 5 Materials, Methods, Synthesis, and Characterization..................................................150

General Information for Chemistry..................................................................................................151

Methods for Chemistry (Synthesis)..................................................................................................155

Methods for Biological Experiments...............................................................................................187

References........................................................................................................................................194

APPENDIX (NMR Spectra)..............................................................................................................196
LIST OF TABLES

Table 1.1 A list of Chemical Inducers of Protein Heterodimerization Developed...........25 in the Order of Years Published
LIST OF SCHEMES

Scheme 1.1 The First Synthetic Small Molecule Homodimerizer and Its Derivatives..11

Scheme 2.1 Synthesis of TH and NTH.................................................................73

Scheme 2.2 The Photolysis Mechanism of NVOC Photocage.................................77

Scheme 3.1 The Photolysis Mechanism of DEACM..............................................99

Scheme 3.2 Synthesis of the Coumarin-caged CTH..............................................109

Scheme 3.3 Synthesis of the photocleavable TNH...............................................111

Scheme 4.1 Retro-synthesis of CTNH................................................................124

Scheme 4.2 Synthesis of the CTNH..................................................................125
LIST OF ILLUSTRATIONS

Figure 1.1 An Example of a Modularized Interactive Map of E2F1.................................6
Figure 1.2 The Basic Principles of Chemically Inducible Dimerization (CID). .................9
Figure 1.3 Inducible Protein Dimerization to Control a Biological Process. .......................17
Figure 1.4 Chemical Heterodimerizers with Sequential Binding Mechanisms..................22
Figure 1.6 The plant hormones Absciscic Acid and Gibberellin 3.................................33
Figure 1.7 Reversible Photoisomerization of Azobenzene...........................................39
Figure 1.8 Photocaged Chemical Dimerizers..................................................................46
Figure 1.9 The Prozone or Hook Effect and Explanation...............................................49
Figure 2.1 Chemical Structures of New Dimerizers.......................................................69
Figure 2.2 Design of the Photocaged Dimerizer NTH...................................................70
Figure 2.3 Ultraviolet (UV)–Visible Absorbance Spectra of Dimerizers...........................75
Figure 2.4 A Ligand-bound eDHFR-Haloenzyme Ternary Complex Model.......................79
Figure 2.5 A Haloligand-bound Haloenzyme Docking Model.........................................81
Figure 2.6 The Mechanism of Haloligand Labeling in the Halotag System.......................82
Figure 2.7 A TMP-bound eDHFR Docking Model.......................................................84
Figure 2.8 NTH Enters Living Cells................................................................................88
Figure 2.9 Light-induced Dimerization at Centromeres..................................................89
Figure 2.10 No-dimerizer and 488 nm Excitation Negative Controls. .............................90
Figure 2.11 Dimerization at Mitochondria, Centrosomes and Kinetochores.....................91
Figure 3.1 The New Dimerizers.....................................................................................97
Figure 3.2 New Dimerizers Based on a Modular Design and the Concepts..................98
Figure 3.3 Cell-permeability and Cytotoxicity of CTH and TNH..........................100
Figure 3.4 Light-induced Protein Dimerization at Mitochondria with CTH.............101
Figure 3.5 Light-Induced Reversal of Dimerization at Mitochondria with TNH.........104
Figure 3.6 Control of Peroxisome Transport with CTH........................................105
Figure 3.7 Control of Peroxisome Transport with TNH.................................106
Figure 4.1 The Light-controlled ON-OFF Switch for Protein Dimerization.............116
Figure 4.2 UV-Vis Absorption Spectra of CTNH.................................................119
Figure 4.3 Chemical Structures of CTNH Derivatives Synthesized in This Study......123
Figure 4.4 Characterization of DEACM-caged TMP........................................126
Figure 4.5 Determination of Concentration and Incubation Time for CTNH............128
Figure 4.6 AlamarBlue Cell Viability Assay......................................................129
Figure 4.7 Light-induced Protein Dimerization and Reversal at Mitochondria.........134
Figure 4.8 Light-induced Dimerization and the Reversal at Centromeres and.........135
                             the Plasma Membrane
Figure 4.9 Controlling Peroxisome Transport with Light...................................138
Figure 4.10 Activating and Silencing the Spindle Assembly Checkpoint...............141
Figure 4.11 Chemical Structures of Dimerizers in Progress...............................147
ABBREVIATIONS

CDI  1,1’-Carbonyldiimidazole
DBU  2,3,4,6,7,8,9,10-Octahydropyrimidol[1,2-a]azepine
DCM  Dichloromethane
DIEA or DIPEA  N,N-Diisopropylethylamine or Hünig’s base
DMAP  4-(Dimethylamino)pyridine
DMF  N,N-Dimethylformamide
DMSO  Dimethyl sulfoxide
EtOAc  Ethylacetate
EtOH  Ethanol
HATU  1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxid hexafluorophosphate or N-[(Dimethylamino)-1H-1,2,3-triazolo-[4,5-b]pyridin-1-ylmethylene]-N-methylmethanaminium hexafluorophosphate N-oxide
MeOH  Methanol
NaBH₄  Sodium borohydride
NaH  Sodium hydride
NVOC  6-Nitroveratryloxycarbonyl or 4,5-Dimethoxy-2-nitrobenzyl carbamate
Overnight  16 hours
RT  Room temperature (25°C)
TEA  Triethylamine
TFA  Trifluoroacetic acid
CHAPTER 1

RESEARCH OVERVIEW AND LITERATURE REVIEW
INTRODUCTION

Why are we interested in protein-protein interactions?

Cells are a complex mixture of organic and inorganic contents including nucleotides, proteins, lipids, carbohydrates, and other chemicals. These components must synchronously and synergistically interact in a highly precise and specific manner to constitute proper cell function. The fundamental process of life begins with the central dogma; genes encode the information needed to be transcribed and translated into proteins, which are the key players within the cell, carrying out the duties specified by the genes.

Proteins are large biomolecules composed of amino acid residues assembled to form one or more long chains. Proteins have diverse functions within organisms including catalyzing metabolic reactions, DNA replication, responding to stimuli, and transporting molecules from one location to another. The sequence of amino acids directly influence how protein folding occurs. The folded structure of a protein determines its activity. Some proteins are capable of binding small-molecule substrates and other proteins. However, in some cases proteins will bind specifically to other copies of itself to form fibrils. This process often occurs often in structural proteins that consist of globular monomers that self-assemble to form rigid fibers. In all living organisms, protein-protein interactions (PPIs) are fundamental driving forces in biological processes and essential for cellular activity such maintaining homeostasis. PPIs play crucial roles in controlling enzymatic activity, regulating progression through the cell cycle, inducing
conformational changes of proteins, and allowing the formation of large protein complexes that carry out biological functions and construct highly complex signaling networks\(^4\). Therefore, studying these dynamic interactions between specific proteins is crucial in understanding important aspects of cellular function, and ultimately the properties that distinguish between cell types that functionally constitute the whole organism. Localization of proteins is one of the mechanisms exploited by cells in the dynamic regulation of numerous biological signaling networks\(^5\). The precise control of protein localization in time and space is vital to cells because mislocalization of a protein can prevent it from performing a proper function and impact the entire cellular network\(^6\).

**Cellular localization of proteins**

Proteins are synthesized in the cytoplasm but they shuttle throughout the cell. Many events occur following the protein synthesis including proteolysis, post-translational modification and protein folding. Briefly, proteolysis is the removal of N-terminal, C-terminal or internal amino-acid residues. Post-translational modification serves as important cell signaling components such as phosphorylation, glycosylation, lipidation, and carbonylation. Protein folding is usually required for the natural function of the protein. Polypeptide chains often fold to form native secondary and tertiary structures.

Processed proteins are typically transported to a specific location of the cell determined by their role and function at a given time throughout the cell cycle, but how proteins are targeted to specific organelles or cellular structures remains unclear. A
common technique for assessing protein localization is fusing the protein of interest to a "reporter" such as green fluorescent protein (GFP) and subsequently expressing it in a cell. Fluorescence microscopy can be used to visualize the fusion protein.

**Roles of proteins in cell signaling and ligand binding**

Biological processes can involve hundreds of proteins, especially cellular responses and signal transduction. Some proteins, such as insulin, are extracellular proteins that are synthesized and transmitted to other cells in distant tissues as a signaling molecule. Others are located in the membrane and act as protein receptors, whose main function is to bind external stimuli and send a signal to the cell. The biochemical responses may be enzymatic activity or a conformational change detected by other proteins within the cell. For example, transmembrane proteins containing internal channels serve as ligand transport proteins that allow polar or charged molecules and ions, which are difficult to diffuse through the hydrophobic core of the membrane, enter or exit the cell.

As an example of a highly complex regulatory network, the interactive map of E2F1 is illustrated (Figure 1.1). E2F1 is a transcription factor, whose deregulation has been involved in cancer progression, invasiveness, metastasis, and resistance to anti-cancer drugs. This is simply one example of a remarkable protein family that interacts with many genes, proteins, and other transcription factors through a vast array of mechanisms that link essential cancer-related cellular processes such as proliferation, apoptosis, and differentiation. Although this protein and pathway have been extensively
studied, a number of other proteins and signaling networks remain unclear. The crosstalk between the entire cell signaling network in both normal cells and cancerous cells is still poorly understood\textsuperscript{10}. One way to dissect this complicated interaction network is to perturb the system at the biologically-relevant scales. This prompts scientists to develop a variety of research tools and techniques to study protein-protein interactions both \textit{in vitro} and \textit{in vivo} using molecular, genetic, or chemical approaches\textsuperscript{11-12}. In recent years, progress has been made towards understanding and elucidating protein-protein interactions in living systems. Thus, many different techniques have been invented to experimentally control protein-protein interactions in living organisms\textsuperscript{13,14}. One of the most powerful techniques relying on the use of small molecules is chemically-induced dimerization (CID) of proteins and this is our main research focus\textsuperscript{15}. 
Figure 1.1 | An example of a modularized map of E2F1 in tumor progression and metastasis. This interaction map is reproduced from Khan, et al. Nature Communications 8, 198 (2017).
Controlling protein-protein interactions to probe their biological function using chemical inducers of protein dimerization in living cells

As mentioned previously, protein-protein interactions play a vital role in cellular processes ranging from cell division, migration, to death. To probe their function, a powerful technique is to perturb or conditionally control protein interactions with timescales comparable to those that actually occur in living cells. Such timescales would allow for the observation of changes in the cell and the formation of hypotheses and assumptions regarding a given protein’s interaction network.

Chemically-induced dimerization (CID) technique\textsuperscript{15-16} harnesses the naturally-occurring property of proteins that bind small-molecules to control protein-protein interactions using ligands. This technique has opened up new avenues to study complex biological networks, and has led to the development of a new field of chemical biology in which the use of small molecules is employed to elucidate biology\textsuperscript{17}. The principle of the CID technique is conceptually straight-forward: two proteins that require a small molecule in order to interact with one another are brought together upon addition of a ligand that can simultaneously bind both proteins. The spatial proximity induced by the ligand can sway the two proteins to interact in some way, resulting in biological effects (Figure 1.2). When two copies of the same protein are dimerized or oligomerized, the process is called “homodimerization” (Figure 1.2A). Conversely, if two different proteins interact, the term “heterodimerization” is employed (Figure 1.2B). Throughout my thesis, I will frequently use the term “dimerization”, which could be ambiguous, but in most cases, it will refer to “heterodimerization” of different proteins as this is the focus
of our research. In the past three decades, scientists have created new dimerization platforms as research tools; a well-known example being rapamycin-induced dimerization system\textsuperscript{18,19}. With the power of recombinant DNA technology, proteins of interest can be genetically tagged with these dimerizing domains, enabling the study of protein function and related biological processes (Figure 1.2C).
Figure 1.2 | Schematic diagrams illustrating the basic principles of chemically-inducible dimerization (CID). Dimerization is triggered by the addition of a small molecule. (A) Homodimerization of two copies of the same protein. (B) Heterodimerization of two different proteins. (C) CID can be employed as general technique to induce protein–protein interactions by proximity effects, and consequently activate biological phenomena. Using protein engineering technique, any proteins of interest to be dimerized can be genetically fused to a pair of chemically-dimerizable proteins (dimerization domains, DDs).
Brief history of chemically-induced dimerization (CID)

In 1993, a seminal work\textsuperscript{16} led by Schreiber and Cabtree reported the novel synthetic dimeric ligands, FK1012 derivatives, based on a natural product, FK506 (Scheme 1.1)\textsuperscript{16}. These synthetic ligands were used to induce homodimerization of a protein, artificially triggering T-cell receptor (TCR) signaling pathway in the absence of any native TCR ligand. The oligomerization of the intracellular signaling domain of TCR lacking its transmembrane and extracellular domains is elegantly demonstrated as the first example. This work was initially inspired by the studies on the mechanism of action of immunosuppressants, FK506, rapamycin, and cyclosporin A (CsA). They showed that these synthetic ligands acted as natural ligands to dimerize the protein target, calcineurin for CsA-cyclophilin and FK506-FKBP complexes. As a result, the dimerization blocked the translocation of the cytosolic component of the nuclear factor of activated T cells (NF-AT), and ultimately suppressing the activation of the genes necessary for T-cell proliferation. Similarly, rapamycin binds the cytosolic protein FK-binding protein 12 (FKBP12) but the rapamycin-FKBP12 complex inhibits the mammalian target of rapamycin (mTOR) pathway by directly binding to mTOR Complex 1 (mTORC1)\textsuperscript{18}. Although the seminal work from Schreiber and Cabtree\textsuperscript{16} showed homodimerization or oligomerization of a protein, this first report on using a synthetic ligand as a general technique to conditionally control protein interactions could potentially be applied to almost any protein of interest. This has also paved the way to the development of the CID technique until now and has emerged as an important tool in biology\textsuperscript{19}. 
Scheme 1.1 | The first synthetic small molecule homodimerizer FK1012 and its derivatives.

(A) FK1012-A was first synthesized by linking two copies of FK506\textsuperscript{16}, the ligand for FKBP (FK506 binding protein). FK1012 E/Z-3 was later synthesized via a single-step olefin metathesis using Grubb’s catalyst, which can be reduced to obtain FK1012H\textsubscript{2}\textsuperscript{20}. FK1012 bind wild-type (wt) FKBP, abundantly expressed in many eukaryotes. All FK1012 derivatives have been tested and shown to effectively homodimerize FKBP. (B) AP1510 is built from a fully synthetic FKBP ligand (SLF) that is structurally simpler than FK506. (C) AP1903 is added a bulky group to SLF, which now SLF’ binds specifically to FKBP’ (F36V mutant), homodimerizes the mutant FKBP but not bind endogenous FKBP. (D) AP20187, a commercially available homodimerizer with improved pharmacological properties. AP20187 is currently the most widely used mutant FKBP homodimerizer.
Applications of chemically-induced dimerization (CID)

Beyond the previously mentioned application of CID, numerous biological processes have been explored using this technique. Briefly, presented below is an overview of how CID can be exploited to direct protein interactions in several scenarios. The applications have been categorized into two primary groups: dimerization to control specific biochemical reactions or cell signaling pathways and dimerization to control protein translocation, the primary focus of our research.

1.1 Dimerization to control specific biochemical reaction and cell signaling pathways

Controlling protein-protein interactions to induce a specific biochemical activity without directly altering cellular localization of targeted proteins was the first application of inducible dimerization in the original work from Schreiber and Crabtree\textsuperscript{16} (Figure 1.3A). This technique established a general principle of homodimerization and heterodimerization of proteins as a key step to activate signal transduction in many different cellular pathways\textsuperscript{15,16,18,19,21} including autophagy, apoptosis, ER stress, cell proliferation, cell migration, intracellular cargo transport, regulated cell cycle, embryo developmental, stem cell differentiation, gene expression, protein folding and ER quality control, protein degradation, and neurodegenerative disease development and progression. Many signaling pathways and networks have been systematically studied and dissected in broader contexts leading to new discoveries and treatments of diseases. This technique can even be used for therapeutic applications in cancer treatments such as the first-ever approved gene therapy (CAR-T)\textsuperscript{22-24} and the promising cancer-targeted
protein degradation\textsuperscript{25-27}. Very recently, chimeric antigen T cell receptors (CAR-T) immunotherapy has been approved by the US Food and Drug Administration (FDA) for the use against acute lymphoblastic leukaemia (ALL) in children whose cancer is resistant to most drugs available\textsuperscript{23,24}. The principle behind this technique is T cells are drawn from a patient and engineered receptors specific for the patient's particular cancer are expressed onto the T cells by transferring the coding sequence utilizing retroviral vectors. These modified T-cells are then reintroduced into the patient, they specifically recognize cancerous cells, and began killing them effectively. Unlike antibodies against tumor-associated antigens, these CAR-T cells, are not quickly cleared by the body; fortunately, prolonged exposure to CAR-Ts is necessary for good clinical outcome\textsuperscript{28}. However, a significant side-effect is cytokine storm release due to CAR-T cell cytotoxicity against normal cells that express the same receptor, as cancer even with low abundance can cause multi-organ failure and death. A promising way to circumvent this problem is introducing an inducible suicide gene activation with a drug to achieve a remote control, simply called an “OFF-switch” of T-cell activity, which is under clinical investigation\textsuperscript{28}. CID has also been adopted for a similar purpose using a rapamycin-inducible dimerization by designing and incorporating an “ON-switch” CARs that enable small molecule-induced assembly of the split receptors (inactive forms)\textsuperscript{29}. T cell functions are only activated in the presence of a rapamycin analogue that dimerizes the receptor to regain its function while still retaining antigen specificity\textsuperscript{29}. This allows physicians to fine-tune the timing and dosage of T cell activity, thereby mitigating toxicity\textsuperscript{29}. 
Another biochemical activity that has been successfully made inducible using CID technique is split proteins\textsuperscript{21,30}. Exploiting the technique of protein fragment complementation, two separated non-functional parts of a protein can be reconstituted to restore the full activity of the intact form. Previously mentioned signaling responses involving enzymatic reactions i.e. kinases and caspases, which are naturally activated by dimerization and oligomerization. In contrast, inteins and proteases are normally active as monomers, but when they are split into two separated fragments, they are no longer active\textsuperscript{31,32}. Proteins with distinct functional domains, e.g. glycosyltransferases have been shown to be non-functional when split into a golgi-targeting domain and a catalytic domain\textsuperscript{33}. However, addition of a small-molecule capable of inducing dimerization of these domains allows the control of protein glycosylation\textsuperscript{33}. Engineering split proteins with a single-domain globular structures can be challenging because the absence of essential amino acids can dramatically affect the protein folding. However, some success has been achieved as shown with β-lactamase, β-galactosidase, DHFR, GFP, firefly luciferase, ubiquitin, dnaE intein and TEV protease\textsuperscript{30}.

DnaE intein and TEV protease are worth emphasizing because they are proteins that can catalyze ligation and cleavage of protein substrates, respectively. Upon dimerization of the split DnaE intein fragments, of which each is fused to a split protein of interest, protein ligation and splicing occurs. The protein of interest is reformed, exiled from the intein domains, resulting in traceless dissociation from the dimerization domains. This could be advantageous when a protein of interest is sensitive to protein fusion\textsuperscript{32,34}. A major drawback is that the process is inherently irreversible. Likewise, TEV
is a highly specific cysteine protease from Tobacco Etch Virus (TEV), recognizing and
cleaving a sequence that can be genetically incorporated into a protein of interest. The
activity of split TEV can then be reconstituted using CID domains$^{35,36}$. 
Figure 1.3 | Inducible protein dimerization to control a biological process. (A) Dimerization to induce a specific biochemical reaction without changing localization of proteins. An intrinsically localized protein of interest (POI) or different target proteins (POI1 and POI2) can be fused to dimerization domains (DDs). Addition of a ligand dimerizer activates a signaling pathway. (B) A pathway triggered by interactions between an intrinsically localized protein (POI1) and a cytoplasmic, freely diffusible protein (POI2) can be activated by recruiting POI2 to a target in a cellular compartment, increasing the local concentration of a protein. (C) A pathway, which is driven
by localization of a freely diffusible protein (POI2) at a subcellular location or triggered by interactions with an intrinsically localized protein (POI1), can be silenced by recruiting the diffusive protein to another location, where it cannot interact with its partner or perform its function.

1.2. Dimerization to control protein translocation

Inducible protein translocation is a technique using chemically-induced dimerization to change localization of a protein of interest\(^\text{18}\). In the previous applications, biological effects are induced by increasing the effective concentration of two proteins, without necessarily altering their localization. In contrast, CID has been used to effectively control the localization of proteins by recruiting a protein of interest to the desired location (Figure 1.3B). This can be achieved by fusing a dimerization domain to a protein that is known to localize to the desired location and a protein of interest fused to another dimerization domain. Although the protein of interest may be diffused throughout the cell or naturally reside in a different compartment, the protein of interest must travel outside its native location to interact with the recruiter. The equilibrium will eventually shift toward the low energy state to form a dimerization complex at the desired location, depleting the protein of interest from its original site (Figure 1.3B). CID has been successfully applied in recruiting proteins to different cellular compartments such as the plasma membrane, mitochondria, peroxisomes, the nucleus, specific genomic loci\(^\text{18}\). Recently, our lab has demonstrated that the recruitment of a cytoplasmic protein to centrosomes, centromeres, and kinetochores (Chapter 2). CID offers several advantages over genetically fusing a protein of interest directly to a localizable motif, primarily temporal and, in some cases, spatial control. Spatiotemporal control is important for
studying protein interactions and localization on subcellular levels with time resolution similar to natural processes.

Inducible protein translocation can be used to turn on or off a biological process by either recruiting a protein to or removing it from the location in which it is normally active (Figure 1.3B and C). Many cellular events are tightly regulated via the access of signaling proteins to a specific location such as the plasma membrane, to interact with other proteins. Conditionally controlling the access of proteins has been used to activate or inactivate biological processes with the plasma membrane being a common target for inducible protein translocation. The first example of inducible protein relocalization to turn on a pathway is the recruitment of Src kinase to the plasma membrane, which was shown to be sufficient to activate Src signaling cascades\textsuperscript{16}. Heterotrimeric G proteins\textsuperscript{37,38} and the kinases and phosphatases involved in phospholipid signaling\textsuperscript{39-41} have been extensively studied using CID as a protein translocation tool. My colleague, Dr. Ballister exploited this technique to temporally induce the localization of a spindle assembly checkpoint protein, Mad1, at kinetochores\textsuperscript{42}. He elegantly showed that recruitment of Mad1 to kinetochores is sufficient to reactivate the mitotic checkpoint even after it has been silenced through the normal cell cycle, demonstrating the power of CID as a protein translocation technique\textsuperscript{42}.

Inducible translocation, on the other hand, can also be used to inhibit a protein function by sequestering it in a location that it is inactive (Figure 1.3C). The idea is similar to the nuclear import or export process in which proteins are only active in the designated location of the cell. A target protein can be sequestered from its functional site
through inducible dimerization using nuclear import or export as an anchor for a dimerization domain\textsuperscript{43-46}. This translocation could potentially be useful to probe protein function with a technique called “knock-sideways”\textsuperscript{46-49}.

**Chemical dimerizers developed to date**

Numerous applications of CID, particularly rapamycin-inducible systems, have proven to be tremendously useful for research in biomedical science, and have been crucial in significantly advancing the field over the past decades. Since the introduction of FK1012 as the first chemical inducer of protein homodimerization, a variety of different homodimerizers have been developed and utilized for biological applications. Some are natural product-based homodimerizers such as cyclosporin-A (CsA\textsubscript{2})\textsuperscript{50-51}, Coumermycin\textsuperscript{52-53}; others are fully synthetic such as a series of symmetric homobifunctional O\textsuperscript{6}-benzylguanine (BG) molecules (termed CoDis)\textsuperscript{54} and xCrAsH\textsuperscript{55}. As mentioned previously, the focus of this work is on heterodimerization as opposed to homodimerization, as reflected in my effort to develop a new series of heterodimerizers.

**Chemical inducers of protein heterodimerization**

For chemical inducers of protein heterodimerization, rapamycin and its analogs are undoubtedly the most widely used chemical heterodimerizers (Figure 1.4A)\textsuperscript{18}. Rapamycin (Sirolimus) is a macrolide compound, first isolated from the bacterium *Streptomyces hygroscopicus*, found on Rapa Nui (Easter Island). It was first reported as an antifungal drug but soon later was discovered that rapamycin also has potent
immunosuppressive and anti-proliferative properties. The mechanism of action is as follows: rapamycin first binds FKBP, forming a rapamycin-FKBP complex to which the kinase Target of Rapamycin (TOR in yeast, mTOR in mammals) can only bind. Rapamycin sequentially binds FKBP and mTOR (or a small fragment of mTOR called FRB) (Figure 1.4B) with nanomolar affinity. This sequential binding mechanism is highly advantageous as it prevents the “hook effect”, when two separate dimerizer molecules simultaneously saturate the binding sites of each protein, blocking the formation of ternary complex. For a deeper illustration of this effect, please refer to the final section of this chapter. As rapamycin rapidly passes the cell membrane and cooperatively binds its target proteins FKBP and FRB with high affinity, it quickly became a popular platform and was used as a general tool for heterodimerization of FKBP and FRB fusion proteins since 1996. Since FRB only binds to the pre-bound FKBP-rapamycin complex, high concentrations of rapamycin can be added to cells in order to drive the rapid intracellular flux and dimerization. However, there are still four major limitations to rapamycin. First, its interference with mTOR signaling as a potent inhibitor of mTOR, producing significant off-target effects. Second, rapamycin can competitively bind with endogenous FKBP, preventing the dimerization with FKBP-FRB tagged protein of interest. Third, spatial control over dimerization on a single-cell and subcellular levels cannot be achieved using rapamycin system. Lastly, studying the effect after reversal of the dimerization on biologically-relevant short timescales is nearly impossible due to the inherently tight binding ternary complex. Due to rapamycin being one of the first chemical heterodimerizers developed more than two decades ago,
rapamycin-based systems have been in the spotlight, leading to extensive modifications and improvements, and resulting in the establishment of the CID field.

Figure 1.4 | Chemical heterodimerizers with sequential binding mechanisms. (A) chemical structures of FK506, Rapamycin, and AP21967. Binding surfaces are highlighted. (B) Rapamycin first form a complex with FKBP, which then binds FRB (or mTOR). All three components contact each other in this ternary complex. (C) Bump-and-hole technique is used for the design of rapamycin analogs (AP21967) that is specific for mutant FRB.
To address the first problem with off-target effects, rapamycin analogs have been developed using the “bump and hole” technique. A bulky group (bump) is added to the rapamycin structure on the FRB binding surface to create an analog that binds selectively to a mutant FRB with a larger rapamycin binding pocket (hole). The “bump” prevents the rapamycin analog from binding endogenous mTOR (Figure 1.4C)\textsuperscript{66}. The most widely used rapamycin analog (rapalog) (AP21967, Figure 1.4A), is commercially available. Unfortunately, the second problem with competitive binding to endogenous FKBP, has not been widely addressed. No rapalogs that are selective for a mutant FKBP have been reported. Thus, high levels of exogenous FKBP and FRB fusion protein expression are necessary to achieve high efficiency of the dimerization. This limitation can also be effectively overcome by deleting or depleting endogenous FKBP\textsuperscript{42} although this can create another problem with a knock-out or knock-down effects in some cases. Finally, the reversibility and lack of spatial control over dimerization issues of rapamycin-based systems have been addressed by a variety of techniques such as using an analog with a weaker binding affinity (a synthetic ligand of FKBP or SLF), which can be displaced by the addition of rapamycin\textsuperscript{18}. Alternatively, the development of completely new dimerizers to overcome these issues has been an active area of research of many laboratories, including our lab.

**Development of new chemical dimerizers**

The profound limitations of rapamycin and promising future of this technique as an emerging tool for biologists has led to new avenues of research in chemical biology.
With collaboration between synthetic chemists and biologists, these issues can be overcome by the development of new molecules that fulfill the requirements of being “a good dimerizer” for mammalian cell studies.

A “good dimerizer” has several distinct characteristics. An ideal chemical dimerizer would rapidly enter cells and bind to its target proteins with high specificity, forming a stable ternary complex for a reasonable length of time. The dimerizer cannot be cytotoxic to cells or interfere with other biological pathways. One strategy to increase the “bioorthogonality” of dimerizers is to use systems from other species such as bacteria or plants as they do not have targets in mammalian cells.

For this review, specific examples (summarized in Table 1.1) from the literature to illustrate have been selected to illustrate a variety of the new dimerizer developments and their applications.
Table 1.1 | A list of Chemical inducers of protein heterodimerization developed in the order of years published

<table>
<thead>
<tr>
<th>No.</th>
<th>Name/Acronym</th>
<th>Proteins dimerized</th>
<th>Applications</th>
<th>Protein size (kDa)</th>
<th>Corresponding PIs</th>
<th>Year reported/Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FK506</td>
<td>FKBP (3 copies)/ calcineurin A (CnA)</td>
<td>Transcriptional activation domains in reinitiation</td>
<td>12/58</td>
<td>Stuart Schreiber &amp; Gerald Cabtree</td>
<td>1996 [64]</td>
</tr>
<tr>
<td>2</td>
<td>FK506- cyclosporin A (FK-CsA)</td>
<td>FKBP (3 copies) / cyclophilin</td>
<td>Inducible Fas recruitment to the plasma membrane, nuclear localization, and transcriptional activation</td>
<td>12/21</td>
<td>Stuart Schreiber &amp; Gerald Cabtree</td>
<td>1996 [67]</td>
</tr>
<tr>
<td>3</td>
<td>FK506- Dexamethasone (Dex)</td>
<td>FKBP/ glucocorticoid receptor (GR)</td>
<td>The first three hybrid system screening</td>
<td>12/87</td>
<td>Jun Liu</td>
<td>1996 [68]</td>
</tr>
<tr>
<td>4</td>
<td>Rapamycin</td>
<td>FKBP (3 copies) / FRB</td>
<td>Inducible gene expression in mice</td>
<td>12/11</td>
<td>Michael Gilman</td>
<td>1996 [65]</td>
</tr>
<tr>
<td>5</td>
<td>Rapalogs</td>
<td>FKBP (3 copies) / FRB with specific mutations</td>
<td>Inducible gene expression and protein translocation</td>
<td>12/11</td>
<td>Stuart Schreiber</td>
<td>1997 [69]</td>
</tr>
<tr>
<td>6</td>
<td>Mtx-Dex</td>
<td><em>E. Coli</em> DHFR/ glucocorticoid receptor (GR)</td>
<td>Yeast three-hybrid assay, activating transcription of a lacZ reporter gene and β-galactosidase activity</td>
<td>18/30</td>
<td>Virginia Cornish</td>
<td>2000 [70]</td>
</tr>
<tr>
<td>7</td>
<td>FK506-FKderivative</td>
<td>FKBP/ mCAB</td>
<td>Selection of compensatory receptors</td>
<td>12/25</td>
<td>Stuart Schreiber</td>
<td>2002 [71]</td>
</tr>
<tr>
<td>No.</td>
<td>Name/Acronym</td>
<td>Proteins dimerized</td>
<td>Applications</td>
<td>Protein size (kDa)</td>
<td>Corresponding PIs</td>
<td>Year reported/ [Ref.]</td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>-------------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>8</td>
<td>Mtx- O&lt;sub&gt;6&lt;/sub&gt;-benzylguanine (BG)</td>
<td><em>E. Coli</em> DHFR/ mutant O6-alkylguanine-DNA alkyltransferase (hAGT, SNAP-tag)</td>
<td>Transcriptional activation in a yeast three hybrid assay</td>
<td>18/20</td>
<td>Kai Johnsson</td>
<td>2003 [72]</td>
</tr>
<tr>
<td>9</td>
<td>TMP-Dex</td>
<td><em>E. Coli</em> DHFR/ glucocorticoid receptor</td>
<td>Yeast three-hybrid assay, activating transcription of a lacZ reporter gene and β-galactosidase activity</td>
<td>18/30</td>
<td>Virginia Cornish</td>
<td>2007 [73]</td>
</tr>
<tr>
<td>10</td>
<td>TMP-SLF</td>
<td><em>E. Coli</em> DHFR / FKBP (3 copies)</td>
<td>Conditional Glycosylation in Eukaryotic Cells</td>
<td>18/12</td>
<td>Carolyn Bertozzi &amp; Virginia Cornish</td>
<td>2008 [33]</td>
</tr>
<tr>
<td>11</td>
<td>Abcisic acid (ABA)</td>
<td>PYLC&lt;sub&gt;S&lt;/sub&gt; / ABIC&lt;sub&gt;S&lt;/sub&gt;</td>
<td>Regulate transcription, signal transduction</td>
<td>20/33</td>
<td>Gerald Crabtree</td>
<td>2011 [74]</td>
</tr>
<tr>
<td>12</td>
<td>acetoxyethyl gibberellin (GA&lt;sub&gt;3&lt;/sub&gt;-AM)</td>
<td>gibberellin insensitive dwarf1 (GID1) / gibberellin insensitive (GAI)</td>
<td>Combine with rapamycin system for logic gated membrane ruffling</td>
<td>38/10</td>
<td>Takanari Inuoe</td>
<td>2012 [75]</td>
</tr>
<tr>
<td>13</td>
<td>Fusicoccin (FC)</td>
<td>T14-3-3cΔC-M2 / PMA2-CT52M1</td>
<td>Activation of NF-κB signalling</td>
<td>27/6.5</td>
<td>Christian Ottmann</td>
<td>2013 [76]</td>
</tr>
<tr>
<td>14</td>
<td>Halo-SNAP (HaXS8)</td>
<td>HaloTag / SNAPTag</td>
<td>Protein targeting to cytoskeleton, the plasma membrane, and lysosomes. the PI3K/mTOR pathway, and multiplexed protein complex formation with the rapamycin dimerization system</td>
<td>30/20</td>
<td>Matthias Wymann</td>
<td>2013 [77]</td>
</tr>
<tr>
<td>No.</td>
<td>Name/Acronym</td>
<td>Proteins dimerized</td>
<td>Applications</td>
<td>Protein size (kDa)</td>
<td>Corresponding PIs</td>
<td>Year reported/ [Ref.]</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>15</td>
<td>SNAP-SLF (rCD1) SNAPTag / FKBP</td>
<td></td>
<td>Induce and stop phosphatidylinositol 3-kinase (PI3K) activity, allowing the quantitative measurement of the turnover of phosphatidylinositol 3,4,5-trisphosphate (PIP3) and its downstream effectors</td>
<td>20/20</td>
<td>Carsten Schultz</td>
<td>2014 [78]</td>
</tr>
<tr>
<td>16</td>
<td>TMP-SLF’ eDHFR- FKBP’</td>
<td></td>
<td>Reversibly target a constitutively active Rac1mutant to the plasma membrane in live cells, which led to rapid and reversible formation of lamellipodia.</td>
<td>20/18</td>
<td>Yao-wen Wu</td>
<td>2014 [79]</td>
</tr>
<tr>
<td>17</td>
<td>TMP-Halo (TH) eDHFR/ HaloTag</td>
<td></td>
<td>A novel dimerizer that can be efficiently photocaged for light-induced recruitment of cytosolic mCherry protein to many different cellular structures including centromeres, kinetochores, mitochondria, and centrosomes.</td>
<td>18/30</td>
<td>Michael Lampson &amp; David Chenoweth</td>
<td>2014 [80] This work!</td>
</tr>
</tbody>
</table>

This work!
Figure 1.5 | Chemical structures of heterodimerizers reported to date, corresponding to Table 1.1
The first generalizable chemical inducer for protein heterodimerization is FK506 (Figure 1.5A), developed in 1996. FK506 by itself could theoretically be used as a heterodimerizer, inducing FKBP and calcineurin A (CnA) dimerization. Since then, FK506 has been conjugated with other small molecules such as Cyclosporin-A to generate FK-CsA (Figure 1.5A), the first artificial semi-synthetic heterodimerizer. FK-CsA dimerizes FKBP and Cyclophilin A fusion proteins. By coupling FK506 to a glucocorticoid hormone analog, dexamethasone, Dex-FK506 (Figure 1.5B) was synthesized and used to induce dimerization of wild-type FKBP and the full-length glucocorticoid receptor (GR) fusion proteins. FK506 was used again by Schrieber’s lab to produce FK506-FK506 derivative, dimerizing FKBP and mCAB (fusion of CnA340-394 and CnB3-170 with specific mutations). However, FK506-based dimerizers still competitively bind endogenous FKBP12 limiting their efficiency and applicability.

Rapamycin and its analogs were introduced in 1996 and 1997, and have shown to be more promising than FK506 although they still have some serious limitations as previously mentioned. A few years after FK506 and rapamycin dimerizers were reported, Virginia Cornish’s lab has pioneered the development of new systems, aiming to overcome some limitations of existing dimerizers. Several dimerizers, composed of ligands for steroid receptors (hormone binding domain of GR) and dihydrofolate reductases (DHFRs) were produced from either commercially available materials or simple synthetic strategies. The first dimerizer from her lab, developed in 2000, was Dex-Mtx (Figure 1.5C) which combines dexamethasone with methotrexate (Mtx), a dihydrofolate reductase (DHFR) inhibitor that binds both DHFRs from bacteria and
eukaryotes with low selectivity. The Cornish lab later in 2007 developed another dimerizer, Dex-TMP (Figure 1.5E), which the Mtx ligand was replaced by trimethoprim (TMP), an inhibitor 1000-fold more selective toward bacterial vs mammalian DHFRs. Dex-TMP solves the issue observed with Mtx, namely the non-selective binding to mammalian DHFR, but it may still have some undesirable off-target effects from dexamethasone inhibitory activity. Historically, Dex has been prescribed as a corticosteroid drug for a variety of treatments including rheumatic problems, skin diseases, severe allergies, asthma, chronic obstructive lung disease, croup, brain swelling, and tuberculosis. Corticosteroids or steroid hormones are involved in a wide range of physiological processes including stress response, immune response, regulation of inflammation, carbohydrate metabolism, protein catabolism, blood electrolyte levels, and behavior in humans. Some severe adverse effects may occur.

Back in 2003, Kai Johnsson’s lab reported a new heterodimerizer, utilizing methotrexate (Mtx) and his newly developed covalent protein labeling system called SNAP-tag. He combined the SNAP-tag ligand, O\(^6\)-benzylguanine (BG) with Mtx, creating the first half-covalent dimerizer, BG-Mtx (Figure 1.5D). This molecule forms a covalent attachment with the SNAP-tag protein, an engineered human O\(^6\)-alkylguanine-DNA alkyltransferase (hAGT), but forms a non-covalent interaction with DHFR. Although these modular dimerizers have been successfully applied to induce protein dimerization in living cells, the concentration of the dimerizer must be carefully controlled. Excess dimerizer molecules can cause non-cooperative or non-productive binding, preventing the ternary complex formation (the hook effect). One of the
advantages of a half-covalent dimerizer like Mtx-BG is the unreacted or excess dimerizer can be washed out, improving the dimerization efficiency. However, the wash-out effectiveness must still be carefully examined for every different dimerization system and experiment. In theory, this modular half-covalent design is far superior compared to completely non-covalent systems, prompting us to adopt this strategy in our new dimerizer design.

In 2008 Carolyn Bertozzi’s lab, in collaboration with the Cornish lab, reported TMP-SLF (Figure 1.5F), a dimerizer containing a synthetic ligand of FKBP (SLF) combined with TMP\textsuperscript{33}. SLF is an analog of FK506 that lacks the ability to inhibit calcineurin but still binds tightly to FKBP, meaning competitive binding to endogenous FKBP is still an issue.

Notably, in 2011 and 2012, two completely new chemical dimerizers based on plant hormones and their natural receptors, Abscisic acid (ABA) and Gibberellin (GA) or specifically Gibberellic acid (GA\textsubscript{3}) were reported (Figure 1.5G)\textsuperscript{74,75}. These plant hormones exhibit a similar sequential binding mode to rapamycin in that they initially bind the first receptor and the resulting binary complex then binds the second protein, preventing the hook effect. The difference is in the initial binding event, as the plant hormones induce a dramatic conformational change of the protein receptor in order to bind the second target (Figure 1.6B)\textsuperscript{82,83}. ABA dimerizes its receptor, pyrabactin resistance(PYR)/PYR1-like (PYL) and a protein phosphatase PP2C, whereas GA\textsubscript{3} binds the receptor gibberellin insensitive dwarf 1 (GID1) and another protein called gibberellin insensitive (GAI). Both ABA and GA\textsubscript{3} are non-toxic at concentrations used and were
successfully applied to control signal transduction and protein translocation in living mammalian cells\textsuperscript{34,75}. However, optimization of the systems was required as both molecules contain a carboxylic acid group that is deprotonated and negatively charged at physiological pH; this characteristic is known to hamper cellular permeability, necessitating high micromolar concentrations of ABA, and a long incubation period (~30 min). Cleverly, Miyamoto and coworkers tackled this limitation by masking the carboxylic acid group of GA\textsubscript{3} with an ester protecting group (acetoxymethyl, or AM) to yield a neutral prodrug dimerizer, GA\textsubscript{3}-AM that is more readily cell-permeable (Figure 1.6A). Once passing through the cell membrane, GA\textsubscript{3}-AM is promptly cleaved by

\textbf{Figure 1.6} | The plant hormones Abscisic Acid (ABA) and Gibberellin 3 (Gibberellic Acid 3 or GA\textsubscript{3}). (A) GA\textsubscript{3}-AM is a prodrug form of GA\textsubscript{3} in which the carboxylic acid group is masked as acetoxymethyl ester. This modification significantly improves cellular permeability because once inside the cell, the masking group is rapidly removed by promiscuous esterases, yielding GA\textsubscript{3}. (B) These plant hormones confer a sequential binding mechanism of protein dimerization, similar to that of rapamycin, but it differs in the structural basis. Like rapamycin, ABA and GA do not bind the secondary receptor by themselves, but rather initially bind the first hormone receptor, and
then this complex binds a secondary receptor. In contrast to FK506 or rapamycin, ABA and GA binding induces structural rearrangements of the primary receptors, allowing the binding to the secondary receptor. Thus, these plant hormones do not have a direct contact to their second receptor.

promiscuous intracellular esterases (Figure 1.6A), allowing maximal GA₃-induced dimerization within 60 seconds. The slow cellular uptake of ABA could be significantly improved by the same prodrug strategy, but it has never been reported. Unfortunately, none of the new dimerizers reported so far have shown faster kinetics of dimerization compared to rapamycin-based systems (~10 seconds)⁴². Although ABA and GA₃-AM are not yet commercially available, these new dimerizers are promising for future applications due to their cooperative binding mechanism, similar to rapamycin but have no targets in mammalian cells. Moreover, due to the orthogonality of GA₃ to previously developed systems it has been used in conjunction with rapamycin to construct AND and OR logic gates using cell morphology changes (membrane ruffling) as the phenotypic output signal. This has opened up new possibilities in the areas of CID and synthetic biology. Unfortunately, some limitations of these molecules are the lack of spatially-controlled induced dimerization and reversibility on biologically-relevant timescales.

In 2013 Fusicoccin (FC), a natural product produced by the fungus Fusicoccum amygdali, was reported as a chemical dimerizer (Figure 1.5G). It dimerizes an engineered 14-3-3 protein and the C-terminus (CT) of the proton ATPase fusion protein. Although this dimerizer was reversible and claimed to exhibit a cooperative dimerization event⁷⁶,⁸⁴, it did not receive much attention. Perhaps because 14-3-3 proteins are a family
of regulatory molecules conserved in eukaryotes, binding to functionally diverse signaling proteins, this relatively new and complex system requires additional characterization and optimization for increased adoption. Nonetheless, this orthogonal system expands the CID toolbox and is complementary to existing systems.

In the same year, Matthias Wymann’s lab reported a novel CID system employing two covalent ligands from protein labeling methods, Halotag and SNAP-tag (Figure 1.5H). A series of covalent heterodimerizers were developed as a general form, Halo-X-Snap (Figure 1.5H), where X refers to various linkers. They demonstrated that the linker can contribute to cell permeability and overall efficiency of dimerizer. The most optimal molecule (HaXS8, Figure 1.5H) showed the highest level of dimerization (65% of total protein) with the fastest kinetics (>15 minutes). Nonproductive binding was observed when excess dimerizer was used. Slow kinetics, lack of spatial control and irreversibility are the major limitations. The same group later inserted a photocleavable linker into this dimerizer (MeNV-HaXS) and demonstrated a light-controlled release of a protein from individual vesicles. Using light to reverse the dimerization with spatiotemporal control on a subcellular scale is a powerful solution to solve several limitations of previous dimerizers. Thus, this has inspired us to develop our own photocleavable dimerizers (Chapter 3 and Chapter 4).

2014 held significant advancements for the CID field, three new heterodimerizers were developed and reported including our first generation dimerizer (Chapter 2). Carsten Schultz’s lab published a new dimerizer rCD1 (Figure 1.5I), comprising of BG (SNAP-tag ligand) and SLF, and used it with FK506 as a way to reverse dimerization. There
seemed to be a common problem with the BG ligand-containing dimerizers, which is slow kinetics of either cellular uptake or the labeling reaction with the SNAP protein (>30 min to achieve the maximal dimerization after addition). Interestingly, upon addition of FK506 the dimerization could be rapidly reversed within a minute. A few months later, Yao-Wen Wu’s lab reported another dimerizer consisting of SLF’ (a synthetic ligand of F36V mutant of FKBP (referred to as FKBP’), and TMP (Figure 1.5J). SLF’ binds the FKBP’ with sub-nanomolar affinity and displays a 1000-fold selectivity for this mutant over the wild-type FKBP, thus minimizing the off-target effects and endogenous FKBP interference. In addition to the advantages of using TMP as mentioned previously, the non-covalent nature of the interaction allows dimerization to be reversible with competitive binding of the original TMP drug. This reversibility with TMP can be multiple times before rapamycin can be added to permanently reverse the dimerization. The group used this dimerizer to reversibly induce lamellipodia formation in living cells. Beyond this, they considered the applicability of the dimerizer for double relocation of proteins by adding rapamycin, rerouting a protein from the membrane to mitochondria with fast kinetics. This dimerizer highlights the advance of CID in terms of cycled reversibility that was previously impossible with dimerizers such as GA₃ and rCD1. Some drawbacks, however, include the lack of spatial control over dimerization and reversal when using rapamycin, which is probably not the best option due to its off-target effects. Optimal concentrations of TMP-SLF’ need to be determined for a given experiment due to its non-cooperative binding mechanism. Overall, TMP-SLF’ will likely become one of the most promising dimerizers for future applications in biology.
communities once it is commercially available or more readily accessible. The complex synthesis of SLF’ is a significant barrier, preventing it from becoming popular. In addition, we also reported another TMP-based dimerizer (chapter 2) that is coupled to Halotag system in the same year (Figure 1.5K)\textsuperscript{80}.

In summary, no dimerizer is perfect as each possesses varying strengths and weaknesses depending on the application. Some major limitations of these dimerizers, or the CID technique, in general, are that most of them are irreversible on the biologically relevant timescales or they lack spatial control over the reversibility on single-cell and subcellular levels. This limits their potential to dissect highly dynamic and reversible biological events or complex cell signaling networks, with comparable precision analogous to natural processes.

To combat these issues, a powerful approach is the use of light as an external stimulus to activate small molecules or protein–protein interactions. Light can be applied to a specific region within a cell through well-developed optical microscopy techniques. With its “on-demand” feature, optical control of molecules enables high spatiotemporal precision. A promising strategy to create light-inducible CID systems is photocaging\textsuperscript{86,87}. A molecule is masked, or “caged” by a photoremovable protecting group resulting in the silence of the molecule’s function\textsuperscript{88}. Illumination removes the cage and restores its function. Photocaging biomolecules has been around for decades with widely used biomolecules including signaling messengers such as caged-glutamate, caged-adenosine-triphosphate (ATP), and caged amino acids such as lysine. Optically controlling protein function by blocking the reactive amino acid residues with a photolabile protecting group
can prevent the protein from performing catalytic event or interacting with other proteins. Photocaged CID of proteins have been reported to allow light-triggered dimerization with spatiotemporal control; we are particularly interested in exploiting this strategy in our research.

Prior to the discussion of photocaged dimerizers, the relevant techniques that also harness light as an external stimulus to induce protein-protein interactions or dimerization should be explored\(^{89-91}\). These techniques include chemical photoswitches\(^{92,93}\) and naturally light-sensitive proteins, also known as optogenetics\(^{94,95}\). Each technique offers a distinct set of advantages and disadvantages outlined below.

**Light-controlled tools for biology**

Light has been extensively employed as an external stimulus to activate biological phenomena\(^ {89}\). The ability to regulate proteins or signaling molecules, which bind and activate proteins function, with light provides high spatial and temporal control. There are three techniques harnessing the power light to manipulate biological processes; photoswitches, optogenetics, and photocaged dimerizers. This work will primarily focus on photocaged chemical dimerizers, but also briefly discuss photoswitches and optogenetics.

**Chemical photoswitches**

Photoswitches represent a class of molecules that can be switched between at least two distinct thermodynamically stable forms in response to the application of light in a process called photoisomerization. A popular example is azobenzene, which is
composed of two phenyl rings connected by a N=N double bond (Figure 1.7). It is the simplest structure of an aryl azo compound, a term often used to refer to a class of molecules that share the same azobenzene core structure with various functional groups extending from the phenyl rings.

![Figure 1.7 | Reversible photoisomerization of azobenzene](image)

The photoisomerization of trans and cis isomers of azobenzene and its derivatives can be induced by particular wavelengths of light. The ultraviolet light (UV) absorption corresponds to the of π-π* (S2 state) transition energy gap for trans-to-cis conversion, and blue light absorption corresponds to that of the n-π* (S1 state) transition, for cis-to-trans isomerization. The cis isomer is less stable due to its distorted configuration that is less delocalized than the trans configuration. Thus, cis-azobenzene may thermally relax in the dark, to the trans isomer although this process is slow at room temperature. Dirk Trauner’s lab pioneered research in the area of photopharmacology\textsuperscript{93}, where he uses synthetic photoswitches to control biological events. Applications in biology of most synthetic photoswitches have been focused on controlling voltage-gated and ligand-gated ion channels, resulting in artificial photoreceptors that can be used to optically control neurons with exceptional temporal and spatial precision. The technique has been shown to work well in animals and is potentially applicable to vision restoration\textsuperscript{96,97}, as well as
other sensations. The combination of synthetic photoswitches and receptor proteins significantly contributes to the field of optogenetics and adds a new functional dimension to chemical genetics, and as such they coined the term “optochemical genetics”\textsuperscript{93}. Using this technique, light-induced protein–protein interactions can be switched on and off “on-demand”. Although repeated cycles of activation are desired in some applications, intense and continuous illumination may be necessary to drive the equilibrium toward the activated conformation. Consequently, it may not be an ideal technique to be used for applications in which sustained and long-term dimerization is required.

**Optogenetics**

Optogenetics is a technique that uses a combination of optical and genetic technique to control and monitor the activities of living cells that have been genetically modified to express light-responsive proteins such as ion channels. Initially, after an approach involving microbial opsin genes was introduced in 2005, optogenetics has been widely applied as a research tool to precisely measure the effects dependent upon the manipulation of neurons in a spatiotemporal and specific manner provided by light. In 2010, optogenetics was selected to be the "Method of the Year"\textsuperscript{98,99} across all fields of science and engineering by *Nature Methods*, an interdisciplinary research journal. Optogenetics was also highlighted in the article on "Breakthroughs of the Decade"\textsuperscript{100} in the academic research journal *Science*. The optogenetic toolkit has proven to be an increasingly important technique in the field of neuroscience. This technique has been applied to probe a variety of proteins and cellular functions, extending its scope outside
of neurons. Manipulation of contractile forces relevant to cell migration, cell division, and wound healing have been demonstrated using optogenetic approaches\textsuperscript{101}. The field is rapidly growing and there is a potential to utilize optogenetics in cell and developmental biology to dissect signaling pathways by controlling protein localization, post-translational modification, and GTP loading\textsuperscript{95}.

Recently, an emerging application of this technique is reversible and spatiotemporal control of protein–protein interactions via naturally photosensitive domains in a similar to CID. Since 2009, there have been four common systems reported for light-inducible dimerization: PhyB-Pif (phytochrome B-phytochrome interacting factor)\textsuperscript{102}, FKF1-GI (FKF1 and GIGANTEA proteins from \textit{Arabidopsis thaliana})\textsuperscript{103}, Cry2-CIB1 (cryptochrome 2 and its interacting partner CIB1)\textsuperscript{104} and TULIPs (tunable light-inducible dimerization tags) based of LOV2 domain of \textit{Avena sativa} phototropin 1 (AsLOV2) and an engineered PDZ domain, ePDZ\textsuperscript{105}. Conceptually, all four systems work in the same way: a photosensitive protein (which may or may not contain a photosensitive cofactor) undergoes a conformational change in response to light, which then allows dimerization with a second (non-photosensitive) protein. The dimerization can be reversed in the dark or upon exposure of a different wavelength of light. Depending on the system, the dissociation kinetics range from seconds to hours. The detailed characteristics of these systems have been reviewed elsewhere\textsuperscript{94,95}. Like photoswitches, the main feature of this technique is the repeatability of reversible activation-inactivation of protein dimerization. Thus, it is especially suitable for controlling biological processes requiring multiple ON/OFF cycles. However, the
reversible control may rely on relaxation kinetics in the dark; one has to carefully consider what system is most suitable for a particular application. The equilibrium of light and dark states may result in “leakiness” of the system where dimerization occurs in the dark\textsuperscript{106,107}. Although only Phy/Pif can behave as a rapid on-demand photoswitch that is sensitive to red light (650 nm) and far-red light (750 nm), Phy protein itself is large (908 amino acids) compared to typical protein tags like GFP (238 amino acids). This bulky tag may be difficult to incorporate or disruptive when fused to proteins of interest. Technical challenges such as gene cloning or \textit{in vitro} transcription should be considered when using this system. Recently, beyond the application at the plasma membrane in Levskaya’s paper\textsuperscript{102}, Phy/Pif dimerization was tested in budding yeast\textsuperscript{46}. Twenty different Phy-tagged fusion proteins that were supposed to localize to several distinct compartments were constructed and expressed, but more than half failed to localize properly. Notably, all three kinetochore targets failed\textsuperscript{46}, indicating that Phy may possess some technical challenges to be adopted as a general tool.

**Photocaged chemical dimerizers**

Photocaged compounds are light-sensitive molecules encapsulated in an inactive form by blocking an important functional part of the molecule with a photolabile organic protecting group. Illumination liberates the caged molecule, permitting targeted perturbation of a biological process via protein-ligand interactions. Uncaging technology and fluorescence microscopy are two complementary experimental techniques, the
former allows photocontrol while the latter allows observation of biomolecules (such as proteins) correlating to cellular function. Caged ligands have been applied to a great extent in neuroscience; caged glutamate, one of the most popular caged molecules, has been very useful for unraveling neural systems. Two-photon cages have enabled the stimulation of single synapses in animals with spatial and temporal precision.

However, some disadvantages associated with caged ligands is that uncaging is an irreversible process making it difficult, if not impossible, to turn off the light-activated event. Therefore, it is more suitable for studying long-termed effects such as a sustained receptor activation or inhibition. In addition, the by-products produced by uncaging such as the released protecting group can be toxic to cells. Thermal hydrolysis (in the dark) of caged compounds can result in the background release of the ligand. For example, neurotransmitters can be problematic as even low concentrations of the neurotransmitter can exhibit a signal and contribute to off-target effects. Some of these shortcomings can be overcome with the photochromic ligand (PCL) or photoswitchable molecule approach as previously described. The ligand carries a photoswitchable moiety that can toggle between two states, active and inactive forms. The efficacy of the ligand only changes upon illumination, triggering the desired biological effect with light in a reversible fashion. Interestingly, the ligand could be an agonist in one form and an antagonist in the other. However, to create a general platform applicable in studying diverse proteins, photoswitches are not ideal and difficult for biologists to adopt as they require investment in the synthesis of new ligands for a specific protein. Hence, CID and optogenetics are more appealing in this case. Furthermore, we are interested in studying protein
localization, which usually requires prolonged induction of protein-protein interaction, but not repeated ON-OFF cycles, making photocaged CID particularly attractive to us.

For protein heterodimerizers, one of the most investigated molecules is rapamycin, which has been caged with many different strategies. For example, a commonly used ortho-nitrobenzyl photocage, rapamycin dimer, and biotin conjugated molecule. All have been caged at the C-40 hydroxyl group, which is the most chemically accessible based on the crystal structures, and this hydroxyl group forms hydrogen bonds with FKBP.

Photocaged rapamycin was reported using an analog of rapamycin (pRap) and an engineered rapamycin binding domain (iFKBP) (Figure 1.8A). Selective acylation of the C-40 hydroxyl group with nitro-piperonyloxy carbonyl N-hydroxysuccinimide carbonate (NPOCNHS) was utilized as a photolabile protecting group. Inoue’s lab created a biotin-linked rapamycin with a photocleavable linker inserted between both ligands (cRb-A, Figure 1.8A), a novel photocaging method to spatiotemporally control the activity of rapamycin. The rapamycin-biotin molecule was bound to avidin, creating a large complex preventing rapamycin from entering the cell. The photocaged rapamycin derivative induced rapid dimerization of FKBP and FRB upon UV illumination. With this system the spatially confined UV irradiation achieved local activation of Rac, a member of small GTPases at a sub-region of the plasma membrane. Although this technique offers a powerful approach to studies of dynamic intracellular signaling events, it is clearly limited to applications at the plasma membrane due to the cell permeability of the dimerizer. Recently, Deiters’ lab reported another derivative, a rapamycin dimer with a
photocleavable linker (dRap, Figure 1.8A) developed to optically induce FKBP12-FRB dimerization only when exposed to light, and successfully applied it to control kinase, protease, and recombinase function.
Figure 1.8 | Photocaged chemical dimerizers. (A) Photocaged rapamycin structures with different approaches. (B) Abscisic acid (ABA) derivatives caged with 4,5-Dimethoxy-2-Nitrobenzyl (DMNB) and (7-Diethylaminocoumarin-4-yl)methyl (DEACM) caging group. (C) Gibberellic acid (GA₃) derivatives, GA₃₁, GA₃₂, and GA₃₃, caged with 2-(4,5-dimethoxy-2-nitrophe-nyl)propyl (DMNPP), (2′-(4′-bis((2-methoxyethoxy)ethyl)amino)-4-nitro-[1,1′-biphenyl]-3-yl)propan-1-ol) (EANBP), and π-extended 2-(o-nitro-phenyl)propyl caging group, respectively.
Other caged dimerizers have been also reported such as ABA and GA₃, plant hormone-based systems. Abscisic acid (ABA) was chemically modified with a photocaging group (DMNB and DEACM, Figure 1.8B) allowing light-induced protein dimerization of ABA receptors, ABI and PYL¹¹². They successfully showed dose-dependent light regulation of cellular processes including transcription, protein translocation, signal transduction, and cytoskeletal remodeling, without the need to perform extensive protein engineering. Caged ABA can be easily prepared from several synthesis steps. Photocaged GA₃ derivatives (Figure 1.8C) were introduced as a novel one- and two-photon light-activated dimerization system for GID1-GAI fusion proteins¹¹³. GA₃-2 and GA₃-3 are sensitive to wavelength at 800 nm due to the EANBP photocage. A newly developed compound reported in this paper, a π-extended 2-(o-nitrophenyl)propyl caging group, displays red shifted absorption. Interestingly, both ABA and GA₃ were caged on the carboxylic moiety to form an ester bond which can potentially be cleaved by intracellular esterases. The uncaging reaction in the dark should be carefully tested in cells to confirm that uncaging event is not from esterases but exclusively occurs by illumination. Nonetheless, these systems can be used orthogonally with other light-controlled CID systems.
The Hook effect (prozone effect)

The prozone effect or the high-dose hook effect is a known phenomenon in assays involving three-component or multivalent binding events, e.g. immunoassays\textsuperscript{114}. It was discovered that increasing the concentration of antibody does not necessarily increase the binding of antigen. On the other hand, after a certain concentration of an antibody, the amount of the antibody-antigen complexes decreases, resulting in lower signal output as shown in a plot of complex formation or signal readout vs concentration (Figure 1.9A). The region of the overly high antibody concentration was named “prozone”\textsuperscript{115,116}. The prozone effect results from saturation of the proteins bound to a “linker or bridging” molecules, preventing each other for binding partners and, as a result, reducing the amount of fully formed ternary complexes.

In CID, a ligand or a dimerizer serves as a bridging molecule between its receptors. The hook effect is only observed with dimerizers that do not possess cooperative binding mechanism such as artificial dimerizers composed of two distinct ligands joined together by a linker component. The effect is caused by two different dimerizer molecules separately reacting with two different protein receptor molecules, inhibiting ternary complex formation (Figure 1.9B). The resulting non-productive protein-ligand complexes can significantly diminish protein interaction or dimerization efficiency. This is well explained in a recent review from Yimon Aye’s lab\textsuperscript{117} and also observed in PROTACs strategy from Craig Crews’s lab\textsuperscript{118}. Furthermore, David Spiegel’s lab built a mathematical model to comprehend the phenomenon\textsuperscript{119}. 
Figure 1.9 | The prozone or hook effect and explanation. (A) A plot of signal readout as a function of concentration of an antibody. (B) Schematic illustration of a bifunctional ligand or a dimerizer that can simultaneously bind two different protein domains (green: dimerization domain 1, DD1; orange: dimerization domain 2, DD2). An optimal concentration of the dimerizer efficiently induces dimerization of the two proteins. In contrast, excess dimerizer molecules can saturate both binding sites, preventing the formation of ternary complex.
Motivation to develop our novel system

In addition to our work (Chapter 2-4), there are only three photocaged dimerizers reported to date, rapamycin, ABA, and GA₃. As an immunosuppressive drug, rapamycin has inherited properties that can interfere with the mammalian mTOR signaling pathway, it can competitively bind to endogenous FKBP, reducing the efficiency of the exogenous FKBP and FRB dimerization. Photocaged rapamycin derivatives made with MeNPOC caging group at C-40 is not totally sufficient to block the interactions of rapamycin and FKBP, resulting in leakiness prior to light-activation. Photocaged rapamycin conjugated with biotin is not cell-permeable, thus it is limited to only plasma membrane targeting. Although photocaged ABA and GA₃, plant-based systems, do not have endogenous targets in mammalian cells, they are not efficiently cell-permeable and in some cases, higher concentrations are required, leading to undesired cytotoxicity. Notably, it has been reported that these molecules increased cellular acidity, which may activate off-target effects.

Additionally, none of the above mentioned photocaged dimerizers are covalent dimerizers, thus they can diffuse away from the location where they were photoactivated after uncaging; i.e. they lack localized target manipulation. Moreover, all the photocaged dimerizers reported to date are not based on a modular design resulting in difficulties in making structural modifications for fine tuning properties tailored for different purposes or improving the system for broader applications in both in cultured cells and other popular model organisms such as C. elegans, Drosophila, zebrafish, Xenopus, and mice.
These limitations of existing photoactivatable CID systems have prompted us to develop a new photocaged dimerizer possessing the following properties: modular design for future modifications, high spatiotemporal control, specificity, and cell-permeability, sequential binding mechanism to avoid the hook effect, no toxicity, and completely inert prior to uncaging and reversible. To accomplish this goal, we have successfully synthesized a novel photocaged dimerizer composed of a non-caged covalent ligand and a photocaged non-covalent ligand jointed by a simple linker. Our first generation photocaged dimerizer exhibits excellent light-inducible dimerization properties that will be discussed in Chapter 2. Following this, we developed a second-generation photocaged dimerizer more sensitive to light and longer wavelengths as compared to the first generation (Chapter 3). The lack of spatial control for the reversal of dimerization and the slow reversibility of our previous dimerizers led us to develop a photocleavable dimerizer based on the same platform (Chapter 3). This photocleavable dimerizer exhibits rapid light-induced reversal of dimerization with high spatiotemporal precision, suitable for controlling dynamic reversible biological processes. We can employ these tools to probe protein interactions with either gain or loss of function depending on the choice of the chemical dimerizer. However, to study the effect of both protein dimerization and reversal in the same cell with spatiotemporal resolution, a protein dimerizer able to be both turned on and off using light is required. Taking advantage of our modular design, we combined both photocage and photocleavable moieties that are sensitive to two different wavelengths into a single-molecule dimerizer. This allows for photouncaging and photocleavage using wavelength-selective strategy; this novel reversible optical
dimerizer is discussed in Chapter 4. An important feature of our system is the ability to wash off unbound free dimerizer; thus, avoiding the hook effect. Finally, the power of our modular design facilitates the future development of new dimerizers tailored for a variety of applications.
REFERENCES


CHAPTER 2
RATIONAL DESIGN OF A NOVEL LIGHT-INDUCIBLE SYSTEM FOR PROTEIN
DIMERIZATION IN LIVING CELLS
AND
SYNTHESIS OF A PHOTOCAGED DIMERIZER

The content in this chapter is reprinted from the following reference:


Contributions: E.R.B., M.A.L. and D.M.C. designed the experiments. E.R.B. initiated dimerizer synthesis and conducted all the biological experiments. C.A. conducted synthesis and chemical characterization. A.M.M. contributed to cloning. E.R.B., M.A.L. and D.M.C. analyzed the data and wrote the manuscript.
Summary

Our first generation photocaged chemical dimerizer consists of four components: a photolabile protecting group, a photocaged ligand, a flexible linker, and a non-caged ligand. For the non-caged ligand, we chose HaloTag system, a covalent ligand/receptor pair. For the photocaged ligand, we chose the trimethoprim (TMP) that selectively binds to a bacterial dihydrofolate reductase (eDHFR). We use 6-nitroveratryloxycarbonyl (NVOC) as a photocage on TMP bound via carbamate linkage. The dimerizer enters living cells and covalently binds to the HaloTag protein. Upon illumination, the NVOC is released, unveiling the native TMP. The two orthogonal ligands in this system bind specifically to their respective cognate receptors with high affinity and the heterodimerization of proteins fused to the receptors is rapidly induced. In addition, we achieve light-induced recruitment of a cytosolic protein to individual centromeres, kinetochores, mitochondria and centrosomes in human cells, demonstrating that our system is applicable to many cellular locations.
**The molecular design rationale**

Most chemical inducers of protein heterodimerization initially developed early two decades ago were based on natural products that bind to two receptors sequentially. Examples of these molecules include rapamycin, ABA, and gibberellin. Disrupting parts of these dimerizers can impact the binding properties significantly, making it difficult to modify the native structures for future applications. Therefore, a modular design of dimerizers pioneered by the Cornish lab in 2000\(^1\), offered flexibility and advantages such as the ease of structural modification without affecting the ligand binding properties, switching ligands for further improvement, and facilitating the future development of new dimerizers with additional properties (Chapter 3 and 4). We envisioned that a photocaged semi-covalent ligand would be the best choice for us to overcome the limitations of existing dimerizers (Chapter 1). Hence, we selected trimethoprim (TMP) and HaloTag system based on previous literature precedent and novelty. The HaloTag system is referred to both the ligand and receptor protein. For clarity, we will call the haloalkane ligand as “Haloligand” or just “Halo” (for short in some cases), and its receptor protein is called “Haloenzyme”. Although both TMP and Haloligand have been successfully incorporated in separate dimerizing systems in the past, the combination of TMP and Haloligand together had not been reported at the time we conducted our research. By changing a ligand, or even one atom in a molecule, the properties of a CID system can change dramatically.

To test whether we could actually use this pair of ligands for protein dimerization in living cells, we began by synthesizing the uncaged TMP-Halo, (TH, **Figure 2.1**). For
the first proof-of-principle experiment, we used HeLa cells co-expressing GFP-Haloenzyme targeted to centromeres by fusing to centromere protein B (CENP-B) and cytosolic mCherry-eDHFR. Using confocal fluorescence microscopy for live cell imaging, if TH could induce protein dimerization in cells, we would observe the colocalization of GFP and mCherry, more specifically, we would see the localization of mCherry at centromeres after the addition of TH. As expected, TH successfully induced dimerization between Haloenzyme and eDHFR fusion proteins (data obtained by our collaborator, the Lampson lab), indicating that TH could be used as a new efficient dimerization tool in living system. This has led us to develop the first generation of photocaged TH, named NVOC-TMP-Halo (NTH, Figure 2.1). After initial testing of all photocaged dimerizers synthesized, we found no difference between NTH derivatives. Therefore, we decided to focus on NTH (2) for biological studies.

Our first photocaged dimerizer (2, Figure 2.2A) that is half-covalent on the non-caged ligand due to the advantages I have discussed previously. The covalent attachment to any protein in the cell serves as an “anchor” for the recruitment of another protein to the local target upon uncaging. The sequential binding mode also prevents the hook effect as the excess dimerizer can be washed out.

We can divide our dimerizer design into four modules (2, Figure 2.2A): a photolabile protecting group (purple), a photocaged ligand (red), a flexible linker (black), and a non-caged ligand (green). For the non-caged ligand and the ligand to be caged, we chose a covalent Haloligand, and non-covalent TMP, respectively. TMP selectively binds to a bacterial dihydrofolate reductase (eDHFR) over mammalian DHFR with ~1000-fold
difference. The 6-nitroveratryloxy carbonyl (NVOC) was chosen as our first photocage on TMP connected via carbamate linkage. Haloenzyme and eDHFR are globular proteins of 294 and 158 amino acids, respectively, comparable to GFP (239 amino acids). Both Haloenzyme and eDHFR have been successfully fused to the amino (N) and carboxy (C) termini of a variety of proteins, making them a promising combination. The detail of each component will be discussed later in this chapter.
Figure 2.1 | Chemical structures of new dimerizers. The non-caged TMP-Halo, TH (1), the two regioisomers of the NVOC caged TMP-Halo, NTH (2 and 3), and a longer linker version of NTH (4)
**Figure 2.2 | Design of photocaged dimerizer NTH.** (A-B) A chemical structure and schematic diagrams of selected NTH (2) and its two receptors: E. coli DHFR (eDHFR) and the Halotag protein (Haloenzyme). (B) Schematic of light-induced protein dimerization in living cells via NTH. The cell-permeable photocaged dimerizer enters cells and irreversibly reacts with the Haloenzyme. Any unreacted dimerizer can be removed by washout. The photocage prevents binding with eDHFR before illumination, and is removed by irradiation with 385–405nm light, allowing eDHFR to bind the uncaged TMP group, thus dimerizing eDHFR and Haloenzyme.
Synthesis of the first generation photocaged dimerizers

Trimethoprim (TMP) can be caged at two different positions of the exocyclic amines in the pyrimidine ring. The two isomers of NTH were isolated (Figure 2.1). In addition, two different lengths of linker or spacer between TMP and Haloligand were synthesized and tested (data not shown). The first linker for Haloligand (Scheme 2.1) was synthesized from a commercially available starting material, 2-(2-aminoethoxy)-ethanol, which was Boc protected in quantitative yield. The hydroxy group of the resulting Boc-protected compound 19 was deprotonated using sodium hydride (NaH) and then reacted with the commercially available 1-chloro-6-iodohexane via alkylation to obtain the protected Haloligand, compound 20. The Haloligand was deprotected in 1:1 solution of trifluoroacetic acid (TFA) and dichloromethane (DCM) at room temperature (rt) for 2 h, and then extended with succinic anhydride under basic conditions yielding the Haloligand with a terminal carboxy functional group, compound 21. The alkyl linker for TMP was synthesized in two simple steps from a commercially available starting compound via Appel reaction with iodine, triphenylphosphine (PPh$_3$), and imidazole to obtain compound 23. TMP ligand was functionalized to obtain compound 24 by first selective demethylation of the methoxy group at the para- position to the methylene bridge substitution. The desired mono-demethylated TMP 24, was then connected to the alkyl linker 23, giving rise to the TMP-NBoc 25, as an intermediate for making both uncaged TH and the photocaged NTH. TH was synthesized in the final step by taking 25 to deprotect the Boc group on the linker and directly couple it to the Haloligand with carboxylic group via standard HATU reagent coupling condition using N,N-
Diisopropylethylamine, or Hünig's base. The photocaged TMP was synthesized from compound 25, reacting with the commercially available 4,5-Dimethoxy-2-nitrobenzyl chloroformate, under basic conditions in DCM at room temperature for 16 h. This results in two regioisomers of the carbamate-caged TMP approximately 1:1 ratio, compound 26 and 27, because the two exocyclic amines in the pyrimidine ring exhibit similar nucleophilic reactivity. These isomers run closely on thin-layer chromatography (TLC) in 10% MeOH:DCM eluent with compound 26 showing a slightly higher R_f. We isolated both isomers, characterized, and confirmed them by ^1H, ^15N nuclear magnetic resonance (NMR) and high-resolution mass spectrometry (HRMS). The key characterization is using two-dimension (2D) ^1H-^15N NMR to observe the correlation between the carbamate proton (NH) which shows up ~10 ppm, with the nitrogens in the pyrimidine ring of TMP. Both isomers were separately carried on constructing the final products NTH, compound 2 and 3. Detailed synthetic procedures and characterization can be found in Chapter 5 and Appendix. Our initial testing of both isomers for protein dimerization showed no apparent difference in the light-induced protein dimerization in CENP-B assay (obtained by the Lampson lab), so we decided to focus on compound 2 for other biological experiments. Moreover, we have also synthesized a version of NTH that has a longer linker between TMP and Haloligand (4) with an increase of ethylene glycol linker by two units (6 atoms). The preliminary testing of light-induced recruitment of a cytosolic protein to centromeres showed no significant difference, compared to compound 2, under the same condition (unpublished data). We therefore decided to continue with the shorter linker, which is more economically efficient and easier to synthesize.
Scheme 2.1 | Synthesis of TH (1), and NTH (2 and 3).
A UV-Vis absorption spectrum of NTH (2) shows the absorbance of NVOC moiety in the NTH spans up to ~420 nm with the maxima ~365 nm. This indicates that we could use 405 nm laser, a targetable laser available in our lab to uncage the NTH and compatible with common wavelengths used for imaging GFP and mCherry, 488 and 594 nm, respectively, which will not uncage the NTH (Figure 2.3).
Figure 2.3 | Ultraviolet (UV)–visible absorbance spectra of dimerizers. Chemical structures and normalized UV–visible spectra (260–600 nm) of NTH (2, in 90% phosphate-buffered saline (PBS) pH 7.4/10% dimethyl sulphoxide), compared to the non-caged analogue TH (1, in 75% PBS pH 7.4/25% DMSO). Light sources used for uncaging (387/11, 405 nm) or for imaging (488, 594 nm) are indicated.
The NVOC photolabile protecting group

We initially selected the ortho-nitrobenzyl derivative: ortho-nitroveratryloxy-carbonyl group (purple colored, Figure 2.2) for various reasons. Firstly, this class of photocaging group has been extensively studied and the mechanism of photolysis is well understood, allowing us to tune its photo-physical properties. Secondly, it is one of the most popular class of photocages employed in the caging of various biomolecules used in living cells and organisms. This was indicative that this particular structure would be compatible with our biological system. Thirdly, the wavelength of light used for uncaging is long enough (>350 nm) that it is not harmful to living organisms. The quantum efficiency is high enough so that the uncaging is near quantitative so only a short pulse of UV light is required for uncaging, minimizing phototoxicity. Finally, the byproduct generated after uncaging reaction is not toxic to living cells. NVOC has improved photoproperties compared to the simplest form of nitrobenzyl derivative (a, Scheme 2.2). The dimethoxy substitutions red-shift absorption and increase solubility of the molecule, and the rate of photolysis in aqueous solution is reasonably fast. The mechanism of photolysis is given (Scheme 2.2). Upon photo excitation, the N-O bond undergoes cleavage, forming a radical species in triplet state. Next, the abstraction of the benzylic proton by the excited nitro group is followed by the formation of the aci-nitro ring and rearrangement to the nitroso derivative. The rapid decarboxylation (CO₂) releases the amine, giving rise to the uncaged compound (Scheme 2.1; TH, Figure 2.1).
Scheme 2.2 | The mechanism of photolysis of NVOC photocage.
The linker or spacer

A protein dimerizer can be created by linking any two different known ligands with a chemical spacer. The linker must be added to the position of the ligands that do not interfere with the binding to protein receptors. Some common chemical linkers are alkanes or polyethylene glycol (PEG) derivatives. Pure hydrocarbon linkers may confer some solubility issues, therefore PEG linkers are more commonly used due to its hydrophilic nature. In our initial molecular designs, we employ a mix of both alkyl and ethylene glycol linker due to the convenience of the synthesis. We have since then improved and synthesized a version of NTH with a longer PEG linker (4, unpublished). With the linker length being the only derivation in our compounds, we have found no explicit difference in light-induced recruitment of cytosolic protein to centromeres (2). Hence, we decided to continue with the more easily synthesized linker. Moreover, to visualize the space between Haloenzyme and eDHFR in a ternary complex, we generated a superimposed model by aligning our TMP and Haloligand into reported ligand-bound crystal structures (Figure 2.4). We found that two ethylene glycol units was necessary for Haloligand binding while a few carbons linked to TMP is sufficient, confirming that the linker length in NTH (2) is optimal.
**Figure 2.4 | A TMP-Halo ligand bound eDHFR-Haloenzyme ternary complex model.** Energy minimization and total charges of the ligand structures were performed by using PM6 calculation. A TMP-Halo based dimerizer structure was energy minimized using MM2 calculation in Chem3D 15.0 to obtain the fully extended structure of the linker. The superimposed dimerizer-bound ternary protein complex is obtained by using Coot program is used to manually match the coordinates of the dimerizer aligned to the individual docked TMP and Haloligand-bound proteins. The spacer is indicated by the black box, and in this particular model, it was measured to be ~35 Å using UCSF Chimera program. The ternary complex structure shows no steric crash between eDHFR and HaloTag proteins and there is extra space for linker flexibility.
The non-caged ligand

For the non-caged ligand, we chose HaloTag system. HaloTag system is derived from a bacterial enzyme, haloalkane dehalogenase, and designed to covalently react with a synthetic ligand of choice. A common ligand is a chlorohexane derivative (Figure 2.5). The enzyme can be fused to a protein of interest on either N- or C-terminus and the recombinant vectors, which are commercially available from Promega. This system offers rapid covalent protein labeling that is bioorthogonal in human cells, and commonly used for visualization of the subcellular localization, immobilization, or capture of a protein of interest. The HaloTag is a genetically modified active site hydrolase, which specifically binds the reactive chloroalkane ligand and has a fast rate of irreversible ligand binding (Figure 2.6). Under physiological conditions, nucleophilic attack of the aspartate residue in the active site to the ligand displaces the halogen, resulting in the formation of a covalent alkyl-enzyme intermediate. In the wild-type hydrolase, this intermediate would be hydrolyzed by an amino acid residue, leading to regeneration of the enzyme. However, in the modified haloalkane dehalogenase (HaloTag), the mutation prevents the reaction intermediate from hydrolysis causing the intermediate to stably attach as a covalent adduct (Figure 2.6). This system has also been successfully incorporated in a chemical inducer for protein heterodimerization.
Figure 2.5 | A haloligand-bound haloenzyme docking model. (A) Depicted is a cartoon of a Haloligand with an ethylene glycol linker (green stick representation with oxygen atoms shown in red) bound in the active site of a haloenzyme (light blue-ribbon diagram, PDB: 1BN7). (B) A close-up look in the active site with all amino acid residues involved in the catalysis including key residues, aspartate and histidine. This model was created by a default-parameter setting in www.dockingserver.com, and the ligand was structurally optimized by Chem3D.
Figure 2.6 | The mechanism of Haloligand labeling in the HaloTag system.
The mutated haloalkane dehalogenase compared to the wild-type counterpart.
The caged ligand

Trimethoprim (TMP) is chosen to be caged in our system due to many reasons. First, TMP is an FDA-approved dihydrofolate reductase (DHFR) inhibitor, which is highly selective for bacterial DHFR with ~1 nM affinity, >1,000 fold stronger than mammalian DHFR\(^2\) so we envisioned that it should be cell-permeable and safe for uses in living systems. Second, TMP and Escherichia coli DHFR (eDHFR) pair has been well validated as a bioorthogonal ligand/receptor pair in eukaryotic cells\(^6\). Third, TMP is structurally simple and can be derivatized by easy and preceded procedures. Adding a linker to the para position of the trimethoxyphenyl ring of TMP has been shown that it is not interfering with eDHFR binding (Figure 2.7)\(^6\). The 2,4-diaminopyrimidine ring of TMP is buried in the tight eDHFR substrate pocket attracted by hydrogen bonds as a key interaction, so we speculated that the addition of a bulky photocage to this group would effectively block this hydrogen bonding and the DHFR binding (Figure 2.7). A photocaged TMP had not been reported, therefore we caged the diaminopyrimidine ring of TMP with 6-nitroveratryl carbamate (NVOC). Upon illumination, NVOC rapidly decarboxylates allowing photolysis to occur, releasing the native amine of the TMP substrate. We joined this photocaged TMP with a Haloligand with a simple chemical spacer, yielding NVOC-TMP-Halo (NTH, Figure 2.1).
Figure 2.7 | A TMP-bound eDHFR docking model. (A) Depicted is a cartoon of a TMP molecule (pale yellow stick representation with oxygen atoms shown in red; nitrogen, blue; and hydrogen, white) bound in the active site of eDHFR (light blue-ribbon diagram, PDB: 1RG7). (B) A close-up look in the active site with all amino acid residues involved. This model was created by a default-parameter setting in www.dockingserver.com, and the ligand was structurally optimized by Chem3D.
Biological results and discussion

To test whether NTH is cell-permeable and optimize the efficiency of protein dimerization, Dr. Ballister developed an assay for detecting the occupancy of the NTH called a dye-blocking assay. We created a cell line expressing Haloenzyme and CENP-B and optimized the concentration and incubation times of NTH. To perform the assay, a Haloligand attached to a fluorescent dye (Halotag-OregonGreen) was added to the cells after the addition and incubation of NTH. Unoccupied Haloenzyme active sites are labelled with this green fluorescent dye, while any occupied sites block this reaction. Cells untreated with NTH serve as a control normalized to 100% of green fluorescent signal, indicating saturation of the Haloenzyme sites by Halotag-OregonGreen. Relative fluorescent signal compared to the control was plotted as a function of NTH concentration and incubation time (Figure 2.8). We found that 20 µM NTH incubated for 1 h was the optimal condition for dimerization at centromeres. Using these conditions, we determined the kinetics of recruitment of cytosolic proteins to centromeres. Upon a short pulse of UV illumination to the cells pre-incubated with NTH, mCherry-eDHFR was rapidly recruited to centromeres with a t½ of ~15 s (Fig. 2.9). No such effect was observed in control experiments when cells were repeatedly exposed to 488 nm light or in the absence of NTH (Figure 2.10). Next, we used a pointable 405-nm laser to achieve rapid and spatially defined recruitment to an individual centromere, indicating controlled protein dimerization with high spatiotemporal resolution in living cells (Figure 2.9). We also show that the NTH-induced dimerization can be reversed by adding the free TMP ligand to competitively bind to eDHFR. Furthermore, to apply our new tool to other
locations in the cell, we constructed Haloenzyme–GFP fusion proteins anchored to kinetochores, centrosomes and mitochondria (Figure 2.11).

Kinetochores are protein complexes, which assemble at centromeres and serve as attachment points for spindle microtubules and signaling hubs for mitotic checkpoint. We fused Haloenzyme-GFP to the kinetochore protein Nuf2, which localizes to the outer kinetochore, proximal to the microtubule binding site. Centrosomes are the primary micro-tubule organizing centers of animal cells. The replication and positioning of these organelles are tightly regulated and linked to polarized cell division and cancer. We target Haloenzyme to centrosomes by fusing it to centrosome localizing domain from AKAP9. Centromeres, kinetochores and centrosomes are not membrane-bound, so we chose mitochondria as a representative for a membrane-bound organelle. Mitochondria are structurally unique organelles involved in oxidative metabolism and apoptotic signaling among other critical cellular pathways. Dysfunction of mitochondria often leads to mitophagy which is strongly associated with neurodegenerative diseases and cancer. We fused haloenzyme to the C-terminal domain of the Listeria monocytogenes ActA protein which localizes to the mitochondrial outer membrane and is exposed to the cytosol. All three of these Haloenzyme–GFP-fused proteins were expressed and localized properly, and we successfully performed light-induced recruitment of mCherry–eDHFR to individual structures in all of these locations using short pulses (10–100 ms) of 405 nm laser illumination (Figure 2.11). Together, these results demonstrate that NTH functions robustly at a diverse set of organelles (mitochondria, kinetochores and centrosomes, DNA binding protein, and centromeres.) To test the limitation of precision of our system,
we were able to successfully recruit an individual kinetochore within a pair of sister kinetochores separated by ~1 mm (Figure 2.11). This highlights the power of our system for spatially and temporally precise manipulation of protein localization in living human cells.
Figure 2.8 | NTH enters living cells. The extent of NTH reaction with CENP-B-Haloenzyme as a function of NTH concentration or treatment time was measured using a dye-blocking assay. Cells expressing CENP-B-Haloenzyme (without GFP) were incubated with NTH, washed, treated with 100nM Halotag-Oregon Green (Halo-OreGrn) for 20 min, then washed again before imaging. Decreased Halo-OreGrn indicates NTH occupancy of CENP-B-Haloenzyme sites. Each Halo-OreGrn image is displayed at two brightness levels to aid visualization. Within each row, all the images are displayed using identical brightness levels. **(A,B)** Cells were treated with 1, 5, 10 or 20 µM NTH for 1 h, then treated with Halo-OreGrn as described above. Treatment with 10 µM NTH for 1 h is sufficient to block ~90% of Halo-OreGrn binding. **(C,D)** Cells were treated with 20 µM NTH for 15, 30 or 60 min, or left untreated as control, then treated with Halo-OreGrn as described above. Treatment with 20 µM NTH for 30 min is sufficient to block ~90% of Halo-OreGrn binding. Images **(A,C)** are maximum-intensity projections of representative cells from each condition. Average Halo-OreGrn intensity at centromeres was quantified for each condition **(B,D)**. Error bars represent s.d. (n=15 fields for each data point, multiple cells per field). a.u., arbitrary unit.
Figure 2.9 | Light-induced dimerization at centromeres. Cells expressing CENP-B–GFP-Haloenzyme and mCherry–eDHFR were treated with 20 µM NTH for 1 h, then washed before imaging. (A) Cell-wide recruitment of mCherry–eDHFR to centromere-localized CENP-B–GFP-Haloenzyme in response to a 2-s pulse of 387(±5.5) nm light. (B) Average mCherry–eDHFR centromere intensity at time points before and after uncaging, error bars represent s.d. (n.10 cells). (C) A single centromere (indicated by arrowhead in inset) was irradiated with a 405-nm laser to induce mCherry–eDHFR recruitment. Insets show boxed regions in GFP (top row) and mCherry (middle row) and colour-merge (bottom row) from indicated time points. Scale bars, 5 or 1 mm in insets. a.u., arbitrary unit.
**Figure 2.10 | No-dimerizer and 488 nm excitation negative controls.** (A,B) Cells were expressing CENP-B-GFP-Haloenzyme and mCherry eDHFR. (A) Cells without any small molecule treatment were exposed to a 2-s pulse of 387(±5.5) nm light. No change in mCherry-eDHFR localization was observed, indicating that UV light does not cause mCherry-eDHFR recruitment to centromeres in the absence of NTH. (B) Cells treated with 20 μM NTH for 1 hour were imaged hundreds of times in the GFP channel, using maximum 488 nm laser intensity. GFP and mCherry images after 100 s and 250 s of cumulative 488 nm laser exposure are shown. No change in mCherry-eDHFR localization is detectable, demonstrating that NTH is functionally insensitive to 488 nm light.
**Figure 2.11 | Dimerization at individual mitochondria, centrosomes and kinetochores.** Cells expressing mCherry–eDHFR and Haloenzyme–GFP-anchor domain fusion proteins specific for (A) mitochondria (ActA), (B) centrosomes (AKAP9) or (C) kinetochores (Nuf2) were incubated with 20 µM NTH for 1 h, then washed before imaging. Cells were imaged before and after targeted laser illumination, as indicated. Individual structures in these cells (indicated by arrowheads in insets) were targeted with a 10–100 ms pulse from a 405-nm laser immediately before the ‘post’ image. Insets show boxed regions in GFP (top row) and mCherry (middle row) and color-merge (bottom row) from indicated time points. GFP is locally photobleached by the 405-nm uncaging pulse. Gaussian smoothing with a radius of 1 pixel was applied to mCherry images in C. Scale bars, 5 or 1 µm in insets.
REFERENCES


CHAPTER 3
DEVELOPMENT OF NEW PHOTOCAGED AND PHOTOCLEAVABLE DIMERIZERS

The content in this chapter is reprinted from the following reference;


Contributions: D.M.C. and M.A.L. designed research. H.Z. and E.V.T. designed and conducted kinetochore experiments which are not included in this thesis. C.A. synthesized and characterized the dimerizers, conducted mitochondria and peroxisome experiments. E.R.B. contributed to the design of dimerizers and checkpoint experiments. H.Z wrote the manuscript and C.A., D.M.C. and M.A.L. edited the manuscript.
Summary

Optogenetic techniques have emerged as important tools to probe dynamic cellular processes such as organelle transport, cell signaling and polarity. These techniques have the potential to provide both spatiotemporal control and molecular specificity by using light and genetically encoded protein tags, respectively. We have previously reported a photocaged chemical dimerizer NTH, that can recruit tagged proteins from the cytosol to multiple cellular structures (Chapter 2). Using this molecule, dimerization can be reversed by the addition of excess TMP as a competitor, but a major limitation of this is that reversal is slow and lacks spatial control. One feature of our CID system is the modular design of the dimerizers, which facilitates the development of new molecules on the same platform. Here we exploit this design by developing two new dimerizers. One can be uncaged using less light and longer wavelengths, while the other allows reversal of dimerization so that proteins can be recruited to and subsequently released from cellular structures using light. We use these dimerizers to control peroxisome transport as a proof-of-principle. Furthermore, we apply these dimerizers to optogenetically control kinetochore function, which will not be discussed in this thesis, but more information can be found in our published article referenced above.
Novel optogenetic tools

Our previously reported dimerizer, NTH$^1$ (2, Chapter 2, Figure 3.1A), composed of four important modules: 6-nitroveratryl oxycarbonyl (NVOC) photocage (purple) to prevent untargeted dimerization; trimethoprim (TMP, red), which non-covalently binds to *Escherichia coli* dihydrofolate reductase (eDHFR); Haloligand (green) that covalently binds to a bacterial alkyldehalogenase enzyme (referred to as Haloenzyme; and a linker (black). Upon Illumination, the photocage is removed, allowing dimerization of eDHFR-tagged proteins with Halo-tagged proteins. Based on this modular design, we developed two new dimerizers (5 and 6, Figure 3.1A) that offer additional properties: increased sensitivity to light (Figure 3.1B; Figure 3.2) and rapid light-induced reversal of dimerization, respectively (Figure 3.2).

Development of a new photocaged dimerizer

We replaced the NVOC with a [7-(diethylamino)-coumarin-4-yl]methyl (DEACM) photocage, which is more sensitive to light and can be uncaged at longer wavelengths (Figure 3.1A). The synthesis of this molecule is discussed in the next section of this chapter; details and characterization can be found in Chapter 5. This new molecule, DEACM-TMP-Haloligand (CTH) (5, Figure 3.1A), enters living cells, as shown in cells expressing Haloenzyme fused to the centromere protein B (CENP-B) (Figure 3.3A), and is not toxic to cells at the concentrations used in our experiments (Figure 3.3B). To show that CTH can recruit proteins from the cytosol to cellular structures, we targeted mCherry to mitochondria by uncaging CTH with 385 nm light. At
this wavelength, CTH requires less light than NTH to uncage (Figure 3.4A). The DEACM photocage can also be uncaged with 444 nm light, whereas NTH is only sensitive to shorter wavelengths (Figure 3.4B).

The photocleavage mechanism of DEACM

The general mechanism of DEACM is different from that of NVOC. Initially, light absorption of DEACM generates $\pi-\pi^*$ excited singlet state (Scheme 3.1)$^2$, which can undergo either unproductive, non-radiative decay and fluorescence relaxation, or the desired heterolytic C–X bond cleavage. There is evidence supporting $S_n1$ cleavage$^2$ at this carbon using $^{18}$O-labeled water. The solvent caged tight ion pair, (coumarin-4-yl)methyl cation and the leaving group conjugate base is the key intermediate. The cation either reacts directly with surrounding nucleophiles or solvent (usually water in biological systems) to form a new stable coumarylmethyl product. Although the heterolytic bond cleavage is very fast, the tight ion pair dominantly recombines to regenerate the ground-state DEACM caged compound in an alternative non-productive pathway (Scheme 3.1). Poor leaving groups such as alcohols, phenols, thiols, and amines can render (coumarin-4-yl)methyl derivatives resistant to photoheterolysis. Thus, caging these groups through a carbonate or carbamate linkage in our strategy is more efficiently released. Decarboxylation of the released carbamate anion, which is the rate-limiting step, is strongly dependent on the pH and on the nature of the released amine or amino acid, but this is usually rapid at the physiological pH. Concomitantly, our substrate, TMP,
exhibits fast release, likely because the aromatic amine leaving group is thermodynamically downhill.

Figure 3.1 | New dimerizers. (a) Chemical structures of CTH and TNH, (b) UV-Vis absorption spectra of CTH compared to NTH (previously developed).
Figure 3.2 | New dimerizers based on a modular design and the concepts. (A) Light-induced protein dimerization using CTH, a coumarin-based photocage that is more sensitive to light, and absorbs longer wavelengths, compared to previously developed NVOC photocage. (B) Light-induced reversal of protein dimerization using TNH. NVOC is inserted in the linker between TMP and Haloligand, providing a reversible spatiotemporal control of protein dimerization using light.
Scheme 3.1 | The photolysis mechanism of DEACM.
Figure 3.3 | Cell-permeability and cytotoxicity of CTH and TNH. (A) CTH enters live cells. HeLa cells stably expressing Haloenzyme-GFP fused to the centromere protein CENP-B were incubated with 10 μM CTH for various times, or with various concentrations of CTH for 60 min, then a dye that binds to Haloenzyme (JF585, HHMI Janelia Research Campus) was added to visualize any sites not occupied by CTH. (B) The Halo dye JF585 signal in A was plotted as a function of CTH concentration and incubation time. (C) CTH and TNH are not cytotoxic at the concentrations used in our experiments. Cells were incubated with the dimerizers for 24 h, which is much longer than our usual experimental time (2-3 h), and the AlamarBlue cell viability assay was used to measure toxicity. CTH and TNH are not cytotoxic up to 50 μM and 1 μM, respectively, which are at least 5-fold higher than the concentrations used in our experiments. Average values were obtained from three independent experiments.
Figure 3.4 | Light-induced protein dimerization at mitochondria as a proof-of-concept for CTH. (A) Cells expressing mCherry-eDHFR and Halo-GFP-ActA, which localizes to mitochondria, were incubated with 10 μM CTH or NTH. Cells were illuminated with a 100-ms pulse of wide-field UV (387 ± 5 nm), which was sufficient to uncage CTH but not NTH, followed by a 2-s pulse to uncage NTH. The localization mCherry at mitochondria can be easily observed when the dimerization occurs. (B) CTH is uncaged by 444 nm light whereas NTH is not. Cells expressing mCherry-eDHFR and Halo-GFP-ActA, as described above, were incubated with either 10 μM CTH or 10 μM NTH. mCherry-eDHFR was recruited to mitochondria by uncaging CTH with a 444-nm light pulse, but NTH was not affected. (C) Graph shows average intensity of mCherry at mitochondria, quantified as described in methods (Chapter 5).
Development of a new photocleavable dimerizer

Due to limited spatially controlled and slow reversibility of our first generation dimerizer by the addition of TMP competitor\(^1\), we developed a new dimerizer that can be cleaved with light to reverse dimerization. Taking advantage of the modular design of our system, we inserted a cleavable NVOC linker in between the Haloligand and TMP ligands to make TNH (6) (Figure 3.1A). The synthesis of this molecule is discussed in the next section of this chapter and more details and characterization can be found in Chapter 5. To show that TNH enters living cells and recruits proteins to cellular structures, we targeted mCherry-eDHFR to mitochondria. The kinetics of recruitment depend on TNH concentration (Figure 3.5), with the highest degree of dimerization (\(t_{1/2} \sim 2\) min) observed at 0.1 \(\mu\)M. A higher concentration of TNH (1 \(\mu\)M) is less effective, because independent occupancy of both protein binding sites with two different TNH molecules would lead to unproductive protein–ligand complexes (the hook effect as discussed in Chapter 1). A lower concentration (0.01 \(\mu\)M) required more time (\(t_{1/2} \sim 5\) min) to achieve maximum dimerization, which is \(\sim 80\%\) of the maximum dimerization obtained with 0.1 \(\mu\)M. TNH is not toxic to cells at these concentrations (Figure 3.3B). To test whether the recruitment of proteins can be reversed with light, we targeted multiple regions sequentially with a 405-nm laser (Figure 3.5B). mCherry was released rapidly in the illuminated regions, demonstrating reversal of dimerization with spatiotemporal control.
**Peroxisome transport assay as a functional test**

As a functional test of the two new dimerizers, we used them to control organelle transport. We previously showed that recruitment of kinesin or dynein motors to peroxisomes induces transport to the periphery or center of the cell, respectively. Recruiting the dynein adaptor Bicaudal-D (BICD) to peroxisomes by uncaging CTH led to peroxisome accumulation at the cell center, as expected (Figure 3.6). To show that transport can be halted by reversing dimerization, we used TNH to recruit an N-terminal fragment of kinesin light chain 1 (KLC1) to peroxisomes, and subsequently cleaved it with light on one side of the cell (Figure 3.7, yellow region). After incubation with TNH, but before photocleavage, peroxisomes were partially depleted from the interior of the cell, showing the recruitment of the organelle to BICD. After illumination, peroxisomes remained stationary on the cleaved side of the cell, but depletion continued on the uncleaved side. Furthermore, peroxisomes accumulated in peripheral regions, where MT plus ends are located, on the uncleaved side, as expected for kinesin-mediated transport (Figure 3.7A and B). These results demonstrate that organelle transport induced by dimerization can be arrested with spatial and temporal control by photocleaving the dimerizer.
Figure 3.5 | Light-induced reversal of protein dimerization at mitochondria using TNH.

(A) TNH recruitment kinetics depends on TNH concentration, and the recruitment can be reversed by cleavage of TNH with UV light. TNH was added at t=0 at the indicated concentrations. (B) Graph shows average intensity of mCherry at mitochondria, quantified as described in methods (chapter 5). Error bars represent standard deviation. Scale bars 10 μm. (C) Cells expressing mCherry-eDHFR and Halo-GFP-ActA were incubated with 100 nM TNH to recruit mCherry to mitochondria. Three regions (outlined in yellow, red or blue) were illuminated sequentially with a 405-nm laser to cleave TNH and release mCherry from mitochondria. Merged images show Halo-GFP-ActA in green and mCherry-eDHFR in magenta.
Figure 3.6 | Control of peroxisome transport with CTH. Cells expressing BICD-mCherry-eDHFR and PEX3-GFP-Halo, which localizes to peroxisomes, were treated with 10 μM CTH. CTH was uncaged with 444 nm light at t=0 to induce peroxisome transport towards the centrosome. Graph shows GFP intensity at the accumulation region (indicated by yellow arrows), which was quantified as described in methods (chapter 5). The kinetics of induced peroxisome transport are different depending on the time of light exposure. Error bars represent standard deviation (N=18 and 16 cells for 0.1 second and 3 seconds illumination time, respectively). Scale bars 10 μm.
Figure 3.7 | Control of peroxisome transport with TNH. (a) Cells expressing kinesin light chain 1 (KLC1)-mCherry-eDHFR and PE X3-GFP-Halo, which localizes to peroxisomes, were treated with 100 nM TNH at t = −4 min to induce peroxisome transport toward the cell periphery. Half the cell (yellow region) was illuminated with a 405-nm laser at t = 0 to cleave TNH. GFP intensity was quantified (f) in the interior region (1), where TNH was cleaved, and in interior and peripheral regions (2 and 3, respectively) that were not exposed to 405 nm light. Intensity in each region over time is shown as a fraction of the maximal intensity in that region. Scale bars, 10 μm.
Synthesis of CTH (Scheme 3.2)

The linker between TMP and Halo was previously synthesized using the same method to functionalize the chlorohexane with a carboxylic acid functional group, 21 (Chapter 2). 7-(Diethylamino)-4-(hydroxymethyl)coumarin (DEACM) was synthesized according to the modified procedures from literature. Oxidation of the allylic carbon of the coumarin, 28, using selenium dioxide (SeO₂) under reflux in p-xylene for three days gave the corresponding aldehyde, which was then reduced by sodium borohydride (NaBH₄) to obtain DEACM, 29, in 10% yield overall. Although we tried to improve the yield by adding SeO₂ multiple times and extending the time for refluxing, the yield is generally low according to many reported procedures, most likely due to the quality of SeO₂ and the complications during column chromatography. We found that different lots and brands of SeO₂ gave inconsistent results. However, by changing the conditions, we were able to improve the yield and reduce the cost of synthesizing TMP-NBoc, 25, which is the key intermediate for making both NTH, CTH, and TNH as well as other compounds in the future. We used 3-bromopropylamine, which is less expensive than the iodo-derivative, with a milder condition using 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in dimethyl sulfoxide (DMSO) at room temperature (rt) overnight. For CTH, TMP was DEACM-caged via carbamate linkage utilizing a different method used for NVOC cage. The chloroformate derivative of DEACM could also be employed to make DEACM-caged TMP, in a similar manner to the NTH, but it suffered from low yields due to sequential reactions and separation of isomers. Therefore, a better synthesis was devised for CTH. TMP-NBoc was added with 1.5 equivalent of 1,1'-carbonyl-diimidazole,
Hünig’s base, and a catalytic amount of 4-(dimethylamino)pyridine (DMAP), then refluxed in DCM for 2 h. The tandem liquid chromatography-mass spectrometry (LCMS) technique was used to monitor the reaction. The activated TMP intermediate with CDI can be trapped by methanol and detected with the mass to charge ratio (m/z) of 550.24, indicating the conversion to the starting material into the methanol adduct. DEACM was then added into the reaction flask, and stirred at rt overnight to gain the DEACM-TMP-NBoc, 30. Surprisingly, we found that using this method, exclusively, the para-DEACM caged isomer was observed as a major product. Unlike the DEACM chloroformate electrophile, this CDI-activated TMP intermediate offers a higher selectivity. More computational studies to determine the reactivity of this intermediate are needed to confirm this observation. The HATU reaction was performed using the same method previously described in chapter 1, to afford the final product, CTH, 5.
Scheme 3.2 | Synthesis of Coumarin-caged TMP-Halo ligand, CTH, (5).

Reagents and conditions: (a) i) SeO₂, p-xylene, reflux, 3 days; ii) NaBH₄, THF:EtOH (1:1), 25 °C, 4 h, 10% (two steps) (b) 48% HBrₐq, 100 °C, 30 min, 48% (c) tert-butyl-3-bromopropylcarbamate, DBU, DMSO, 25 °C, 16 h, 56% (d) i) CDI, DIEA, DMAP, DCM, reflux, 2 h; ii) 29, DIEA, DMAP, DCM, 25 °C, 16 h, 12% (two steps) (e) i) TFA:DCM (1:3), 25 °C, 2 h; ii) 21, HATU, DIEA, DMF, 25 °C, 16 h, 64% (two steps)
Synthesis of TNH (Scheme 3.3)

To synthesized TNH, the NVOC needed to be functionalized, so that it can be attached to both TMP and Haloligand. We envisioned that the methoxy group at the meta-position to the nitro group could be an attachment point and there was precedent that Haloligand could be linked to this functional group. Inspired by the literature precedent\textsuperscript{4}, we redesigned the linker for Haloligand while TMP linker remained the same. Haloligand, 33, was synthesized by mono-alkylation of the tetraethylene glycol (4EG, 31) with 1-bromo-6-chlorohexane, followed by the Appel reaction to convert the hydroxy group into bromo derivative, 32. For the NVOC moiety functionalization, Vanillin was benzyl protected, followed by the nitration reaction to selectively install a nitro group at the para-position to the methoxy group and ortho-position to the aldehyde group, 36. This was achieved by dropwise adding fuming nitric acid (>90% concentrated) in a flask with 35 dissolved in 1,2-dichloroethane at -30 \degree C for 2-3 h. Precaution and protective gears should be taken seriously as the fuming nitric acid is highly corrosive. Deprotection of the benzyl protecting group was done using 48\% aqueous solution of hydrobromic acid (HBr) at 90 \degree C for 3 h but it suffers from low yield (23\%). I later found a better method using pure trifluoroacetic acid (TFA) as both reagent and solvent, which results in a cleaner reaction and desired product without having to do a column chromatography (unpublished). The nitro vanillin, 36, was then reacted with the bromo derivative of Haloligand, 33, to obtain 37, which was then reduced by sodium borohydride (NaBH\textsubscript{4}) to achieved 38. A key step to make TNH is the formation of carbamate linkage, which is generally low-yielding. We found that activating the hydroxy group using 4-nitrophenyl...
chloroformate with DMAP in DCM as a solvent at room temperature overnight, then added to the deprotected TMP, gave the final product, TNH, 6 in a reasonable yield.

Scheme 3.3 | Synthesis of the photocleavable TMP-NVOC-Halo ligand, TNH, 3.

Reagents and conditions: (a) NaH, 1-bromo-6-chlorohexane, THF:DMF (2:1), 0 °C to 25 °C, 16 h, 40%; (b) PPh₃, CBr₄, THF, 25 °C, 16 h, 50%; (c) K₂CO₃, benzyl bromide, DMF, 80 °C, 16 h 99%; (d) i) Fuming HNO₃, 1,2-dichloroethane -30 °C to 25 °C, 3 h ii) 48% HBraq 90 °C, 3 h, 23% (two steps); (e) 33, K₂CO₃, DMF, 60 °C, 16 h, 30%; (f) NaBH₄, MeOH:Dioxane (1:1) 0 °C to 25 °C, 2 h, quantitative yield; (g) i) 4-Nitrophenyl chloroformate, DMAP, DCM, 25 °C, 16 h ii) Deprotected-25 (TMP-NH₂TFA salt) DIEA, DCM, 25 °C, 16 h 16% (two steps).
REFERENCES


The content in this chapter has been prepared for the following manuscript:


**Contributions**: C.A., H.Z., M.A.L. and D.M.C. designed the research. C.A. synthesized and characterized the dimerizer and performed biological experiments. C.A., H.Z., and D.Z.W. conducted the Mad1 experiment together. D.Z.W. contributed to the hydroxy-coumarin synthesis. C.A. drafted the manuscript. All authors edited the manuscript.
Summary

Many dynamic biological processes are regulated by protein-protein interactions, which are highly dynamic and often reversible. Chemically-induced dimerization (CID) has been used to study reversible cellular events, especially cell signaling networks. Previously, we reported two dimerizers CTH and TNH with which light can be used to turn on and off protein dimerization, respectively. However, to date no chemical dimerizer that can be both rapidly activated and deactivated with high spatiotemporal resolution has been reported. Herein, we present a novel chemical inducer of protein dimerization that can be rapidly turned on and off using single pulses of light at two orthogonal wavelengths. We demonstrate the utility of this molecule by controlling peroxisome transport and mitotic checkpoint signaling in living cells. Our system highlights and enhances the spatiotemporal control offered by the CID technique. This new tool will be useful for addressing and uncovering biological problems on a single cell or sub-cellular level.
Introduction to a light-controlled ON-OFF chemical switch for protein dimerization

To dissect complex biological interactions that are highly dynamic and often reversible, experimental tools that can reversibly manipulate protein interactions with comparable precision are necessary. Some major limitations of existing chemically-induced dimerization (CID) systems inspired us to develop a new optogenetic system by introducing light control to the CID technique (Chapter 2). We previously reported photocaged chemical dimerizers that allow protein recruitment with light to many subcellular locations, such as centromeres, kinetochores, mitochondria, peroxisomes, and centrosomes (Chapter 2 and 3), and a photocleavable dimerizer for which light can be used to reverse dimerization (Chapter 3). However, to study both the gain and loss of function of protein recruitment in the same cell with spatial and temporal control, a novel dimerizer that can be both activated and deactivated by light is required. Currently, no CID system with such ability has been reported. To accomplish this goal, we created a novel molecule based on the modular design of our previous dimerizers. Herein, we report a cell-permeable photocaged-photocleavable dimerizer that is capable of turning on and off protein dimerization using two orthogonal wavelengths of light (Figure 4.1). The dimerizer rapidly induces protein dimerization and dissociation within seconds upon illumination at various locations in the cell, such as mitochondria, centromeres, and plasma membrane. Furthermore, we demonstrate the utility of our new system by controlling peroxisome transport and mitotic checkpoint signaling with high spatiotemporal resolution. To the best of our knowledge, this is the first rapidly photo-
inducible and reversible chemical dimerizer that can be completely controlled by light on a timescale of seconds.

**Figure 4.1** | The light-controlled ON-OFF switch for protein dimerization. (A) Chemical structure of the optimized CTNH structure, 9. (B) the concept of using CTNH for light-induced dimerization and the reversal by light. (C) A superimposed model of the ternary complex.
Rational design of the light-controlled system

To create a light-induced reversible dimerizer, we take advantage of the modular design of our CIDs. Based on previously developed dimerizers (Chapter 3), we showed that [7-(diethylamino)coumarin-4-yl] methyl (DEACM) in CTH could be uncaged with 444 nm light. At this wavelength, NTH is unaffected due to the nitroveratryl (NV) having a more blue-shifted absorbance compared to DEACM (Chapter 3). Thus, we were able to correctly predict these two photocages could be used in conjunction with one another for sequential photoactivation. We employ the principle of photolabile functional groups and chromatic orthogonality, which allow for wavelength-selective photo-uncaging\(^1,2\). Although DEACM has a significantly higher photon absorption and high quantum efficiency at 440 nm, NV exhibits its highest absorption at 365 nm but almost no absorption at >400 nm. In addition, the absorption spectra of DEACM and NV are sufficiently blue-shifted to protect them from undesired cleavage at wavelengths commonly used for imaging (e.g., 488 nm) (Figure 4.2). Therefore, we selected this pair for use in our strategy. DEACM (blue moiety in the chemical structure) serves as a photocage and NV (purple moiety) is incorporated into the linker between the two ligands of the dimerizer (Figure 4.1A).

To function successfully in living cells, the dimerizer must enter the cell and form a stable complex with the first receptor (Haloenzyme) for the non-caged ligand. Illumination with the first wavelength (444 nm) selectively removes the DEACM photocaging group without cleaving the NV linker, which initiates binding to the receptor, eDHFR. Another important property is the stability of the chemically induced dimerized
protein complex, which should persist over the course of the experiment and undergo efficient photocleavage only after illumination with the second wavelength (387-405 nm), which photocleaves the dimerizer to reverse the protein dimerization (Figure 4.1B). After removing DEACM and activating TMP-NV-Halo induced protein dimerization, the photosensitive linker could be buried inside the protein complex, which would diminish the photocleavage efficiency. Thus, the photocleavable linker should be optimized to allow the NV group to be fully exposed outside the dimerized protein complex for efficient photolysis. We speculated that the addition of an ethylene glycol moiety between the caged TMP and the NV photocleavable linker attached to the Haloligand would be useful. We began with a crude model for the structure of the TMP-NV-Halo dimerizer complexed with Haloenzyme (PDB: 1BN7) and eDHFR (PDB: 1RG7) to visualize the space between the two ligands, TMP and Halo. The length of the designed chemical spacer in 9 was determined to be ~35 Å in the fully extended conformation. The NV group was also exposed outside the protein complex. Virtually no steric interactions were observed (Figure 4.1C).
Figure 4.2 | UV-Vis absorption spectra comparing between (A) TH with CTNH, and (B) TNH with CTNH.
Synthesis of the new dimerizers

Building upon our previously established system (Chapter 2 and 3), we designed and synthesized new dimerizers (Figure 4.3), composed of four modules: (7-Diethylaminocoumarin-4-yl)methyl (DEACM) photocage, a caged trimethoprim (TMP) ligand, a photocleavable nitroveratryl (NV) linker and a non-caged HaloTag (Halo) ligand. Employing the synthesis framework similar to that of CTH and TNH (Chapter 3), CTNH1, 7 (Figure 4.3) was first synthesized as the simplest version of CTNH with the shortest linker between TMP and Haloligand. CTNH, 7, was tested in live cell mitochondrial assays (results obtained by Dr. Ballister and later repeated by myself). We found that the dimerization was induced prior to light-activation, indicating two possibilities: either the molecule was contaminated with the non-caged TNH or 7 decomposed in the dark or quickly hydrolyzed in the cell media, giving rise to TNH. After re-synthesis and HPLC purification of several different batches of 7, the issue persisted. Therefore, we ruled out the possibility of contamination. To test whether 7 was decomposing to TNH in the dark, an HPLC assay was conducted to monitor the decomposition progression. It was observed that 7 underwent hydrolysis in the dark, and we suspected it was likely due to the slightly acidic TFA/water that may catalyze decomposition (data not shown). Changing the HPLC solvent to pure water still gave 7 that was contaminated with TNH. To confirm that 7 was being hydrolyzed in water, a live cell assay for protein dimerization at mitochondria was conducted. Cells expressing Haloenzyme-GFP-ActA (mitochondria targeting motif) and mCherry-eDHFR fusion proteins, were added with freshly HPLC purified 7, and then monitored live on the
microscope in GFP and mCherry channels, which do not uncage CTH normally. Without any photoactivation, protein dimerization occurred after ~15 min, indicating that 7 underwent hydrolysis or uncaging in the dark with the half-life of about 15 min. This led to the synthesis of 8 and 9 (Figure 4.3). It is possible that two different moieties in a small molecule interact, such as in Fluorescence Resonance Energy Transfer (FRET) or photo-induced electron transfer (PET). It is plausible that an intramolecular interaction between the two photo-sensitive groups, NV and DEACM contribute to the hydrolysis in the dark. Moreover, there is literature precedent that some photocaged molecules undergo dark-hydrolysis3-4. To improve the design of the dimerizer and prevent hydrolysis, NV and DEACM were separated farther away from each other. By adding a longer linker to 7; 8 and 9 was synthesized. The isomers were isolated to test if one would function better than the other. The separation of 8 from TNH was difficult as it runs closer to the uncaged contaminant than 7. Therefore, only 9 was successfully isolated. Using the same mitochondrial assay to test dimerizer 7, Compound 9 was no longer hydrolyzed in the dark, indicating that 9 is significantly more stable than Compound 7.

CTNH, 9 was synthesized by forming connections between three parts as follows (Scheme 4.1): DEACM-caged TMP, a functionalized ethylene glycol linker, and the 6-nitroveratryl alcohol attached to the Haloligand. The dianinopyrimidine ring of TMP was caged via a carbamate linkage, which rapidly decarboxylated upon photocleavage to release the native amine. For the full synthetic scheme, see Scheme 4.2. The regioisomer of the DEACM-caged TMP was confirmed by nuclear magnetic resonance (NMR). The correlation between the protons of -NH₂ and the two nitrogen atoms in the pyrimidine
ring of TMP was observed in the $^1$H-$^{15}$N HMBC (Figure 4.4). The correlation between -NH and the protons of the methylene bridging between the two aromatic rings of TMP was also observed in the Nuclear Overhauser Effect Spectroscopy (NOESY) results, confirming that the carbamate linkage on the nitrogen atom is ortho to the methylene bridge (Figure 4.4). More synthesis details can be found in Chapter 5.
Figure 4.3 | Chemical structures of CTNH derivatives synthesized in this study.
Scheme 4.1 | Retro-synthesis of CTNH (9).
Scheme 4.2 | Chemical synthesis of CTNH (9).

**Reagents and conditions:** (a) NaH, 1-bromo-6-chlorohexane, THF/DMF (2:1), 0 to 25 °C, 16 h, 40%  
(b) PPh₃, CBr₄, THF, 25 °C, 16 h, 50%  
(c) SeO₂, p-xylene, reflux 3 days; (d) NaBH₄, THF/EtOH (1:1), 25 °C, 4 h, 20% (2 steps)  
(e) 48% HBrₐq, 100 °C, 30 min, 48%  
(f) tert-butyl-3-bromopropylcarbamate, DBU, DMSO, 25 °C, 16 h, 42%  
(g) 15% Phosgene in Toluene, DIEA, THF, 0 °C, 2 h;  
(ii) 25, DIEA, DCM, rt, 16 h, 20%  
(h) K₂CO₃, benzyl bromide, DMF, 80 °C, 16 h, 99%  
(i) Fuming HNO₃, 1,2-dichloroethane -20 °C to 25 °C, 16 h, 50%  
(j) 48% HBrₐq, 90 °C, 3 h, 23%  
(k) 33, K₂CO₃, DMF, 60 °C, 16 h, 60%  
(l) NaBH₄, MeOH/Dioxane (1:1), 0 to 25 °C, 2 h, quantitative  
(m) 4-Nitrophenyl chloroformate, DMAP, DCM, 25 °C, 24 h  
(n) tert-butyl (3-(2-(2-(3-
aminopropoxy)ethoxy)ethoxy)propyl)carbamate, DIEA, DMAP, DCM, 25 °C, 16 h 16% (o) i) TFA/DCM (1:1), 0 to 25 °C, 2 h; ii) succinic anhydride, Et₃N, DCM, 25 °C, 16 h, 90% (2 steps) (p) 30b (DEACM-TMP-NH₂ TFA salt), DIEA, DMF, 25 °C, 16 h, 30% (2 steps)

Figure 4.4 | Characterization of DEACM caged TMP. (A) Key NOE and 15N-HMBC correlations observed in NMR experiment. (B) Optimized geometry structure of DEACM caged TMP in 3D, calculated by energy minimization using B3LYP/6-31G(d) using WebMO Gaussian.
CTNH cell-permeability and cytotoxicity

To determine the cell permeability and optimal concentration of CTNH, we performed the previously established Halo ligand dye displacement assay (chapter 2) with cells expressing centromere protein B (CENP-B) fused to GFP-Haloenzyme. Incubating cells with 20 µM CTNH for 30 min blocked more than 90% of the Halo ligand red dye (JF585) binding, indicating that CTNH enters the cells and occupies the haloenzyme active sites (Figure 4.5). Furthermore, CTNH is not toxic to cells at concentrations up to 20 µM for 24 or 48 h, indicating that it is suitable for long-term live cell experiments (Figure 4.6).
Figure 4.5 | Determination of concentration and incubation time for treatment of cells with CTNH. HeLa cells stably expressing CenpB-GFP-Halo were treated with CTNH (A) at 20 µM, then incubated for 15, 30, and 60 min. (B) at 5, 10, and 20 µM, then incubated for 1 hour. Untreated cells were used as a control and normalized to 100% uptake of CTNH. Subsequently, all cells were washed out and then, treated with Halo red dye (JF585) at 0.5 µM for 15 min, followed by 15 min wash-out prior to imaging. Confocal images reveal that treating cells with CTNH at 20 µM for 30 min is sufficient to block >90% of Halo red dye (JF585) binding, indicating that CTNH enters cells and covalently occupies Haloenzyme active site. Error bars represent s.d. (n >10 fields for each data point, multiple cells per field). a.u., arbitrary unit.
Figure 4.6 | AlamarBlue cell viability assay. HeLa cells were incubated with various concentrations of CTNH, DMSO and camptothecin as a positive and negative control, respectively, in 96-well plates. After 24 h and 48 h incubation, cells were washed with fresh media and stained with AlamarBlue reagent according to the manufacturer's protocol. A plate reader was used to detect the fluorescence signal from metabolically active cells. No significant cell death was observed with CTNH up to 20 µM and 48 h incubation, compared to the negative control. Each experiment was performed with four replicates. Averages were calculated and plotted from three independent experiments. Error bars represent s.d.
**Rapid light-controlled protein dimerization and reversal at multiple locations in living cells**

To validate our protein dimerization strategy, we initially targeted mitochondria, which are highly dynamic, membrane-bound organelles involved in oxidative metabolism, apoptotic signaling, and other critical cellular pathways\(^5,6\). To test the ability of CTNH to recruit and reverse the induced recruitment of a freely diffusible protein to mitochondria, we localized haloenzyme to the cytosolic face of the mitochondrial outer membrane using the C-terminal domain of the *Listeria monocytogenes* ActA protein. HeLa cells expressing eDHFR fused to mCherry (mCherry–eDHFR) and ActA fused to GFP and haloenzyme (Halo-GFP-Mito) were incubated with 20 µM CTNH for 30 min, followed by a 30-min wash-out. No protein dimerization was observed prior to illumination, indicating that caging the TMP ligand with DEACM is sufficient to completely block binding between the TMP ligand and eDHFR (Figure 4.7A). Upon whole-field illumination with 444-nm light, rapid localization of mCherry-eDHFR was observed at the mitochondria (Figure 4.7A). The interaction between TMP and eDHFR is not covalent but, in steady state, the localization is persistent (Figure 4.7A and B). Upon illumination of the whole field of view with ~387 nm light, mCherry-eDHFR was released from mitochondria, indicating that protein dimerization was effectively deactivated by photocleavage of the NV linker (Figure 4.7A and B). To verify that the loss of mCherry from mitochondria was due to photocleavage but not photobleaching, we conducted a control experiment compared to the non-cleavable dimerizer TH that was previously published. Using the same laser condition, CTNH was photocleaved with
lower laser power (~5%) while TH was not affected (Figure 4.7E). Furthermore, to demonstrate spatiotemporal control, we used a targeted 405-nm laser to achieve spatially defined reversal of the dimerization in subcellular regions of the cell. After global activation of mitochondrial recruitment of mCherry-eDHFR, three regions of interest were targeted for sequential photocleavage of CTNH (Figure 4.7C and D).

Next, we targeted centromeres, chromosomal loci that play a crucial role in cell division7,8, by fusing haloenzyme tagged with GFP to the centromere protein CENP-B. HeLa cells expressing CENPB-GFP-Halo together with mCherry-eDHFR were treated with 20 µM CTNH for 30 min, subsequently washed, and exposed to a short pulse of 444 nm light, leading to rapid recruitment of mCherry–eDHFR to centromeres within one second. Upon 387-nm illumination, this recruitment was rapidly reversed within one second (Figure 4.8A). Then, we used a targeted 405-nm laser to achieve rapid and spatially defined removal from an individual centromere (Figure 4.8A), demonstrating spatiotemporal control.

Finally, we targeted the plasma membrane, an important site for numerous cell signaling pathways9 where multiple biological phenomena have been studied using CID, including lipid signaling, Ras-regulated pathways, cell protrusion, and migration10-12. To demonstrate the recruitment of mCherry-eDHFR to the plasma membrane using CTNH, we used HeLa cells expressing cytosolic mCherry-eDHFR and haloenzyme-GFP fused to the CAAX motif, a common plasma membrane targeting sequence, (where C is cysteine, A is an aliphatic amino acid and X is any amino acid). The C-terminus of CAAX proteins undergoes a series of post-translational modifications, resulting in cell membrane
localization\textsuperscript{13}. As for mitochondria and centromeres, global recruitment of cytosolic mCherry-eDHFR was observed at the plasma membrane within one second upon 444-nm illumination and rapidly reversed upon 387-nm illumination (Figure 4.8B).

Together, these results indicate that a variety of subcellular locations, both cytosolic and nuclear, can be targeted to recruit and release a freely diffusive protein on the second timescale, which highlights the fast kinetics and spatial control of CTNH uncaging and cleavage using light. We foresee that CTNH will be widely applicable for rapid and reversible control of protein dimerization or localization at many different subcellular locations. Therefore, we applied this tool to control two highly dynamic biological processes: organelle transport and mitotic checkpoint signaling in live cells.
Mitochondrial fluorescence intensity after illumination (%)

405-nm laser intensity (%)
Figure 4.7 | Light-induced protein dimerization and reversal at mitochondria.

Cells expressing Halo–GFP-ActA (mitochondria targeting sequence) and mCherry–eDHFR were treated with 20 µM CTNH for 30 min, followed by a 30-min washout, then illuminated with 444-nm light to uncage CTNH and recruit mCherry-eDHFR to mitochondria. (A-B) Whole-field illumination with ~387 nm UV light released mCherry from mitochondria. For quantification (b), mitochondrial mCherry intensity were normalized such that the minimum value observed over the course of experiment for each field is 0 and the maximum value is 1. Each time point shows an average of 10 fields. Error bars represent s.d. (C-D) The specified regions (1, 2, and 3, highlighted by orange, red, and pink dashed lines, respectively) were sequentially targeted for 405-nm laser illumination to cleave CTNH. For quantification (d), mitochondrial mCherry intensity in each region was divided by the maximum value observed over the course of experiment in that region. Scale bars, 10 µm. (E) CTNH can be photocleaved whereas TH is not. Cells described above were treated with either 20 µM CTNH for 30 min, followed by a 30-min washout, or 0.1 µM TH (non-cleavable dimerizer) as indicated. Cells were illuminated with 444-nm light to uncage CTNH and recruit mCherry-eDHFR to mitochondria. After which the cells were illuminated again with 405-nm laser at different laser intensities. TH treated cells were illuminated with 405-nm laser under the same condition. The percentage of GFP and mCherry fluorescence intensities after 405-nm illumination were plotted as a function of 405-nm laser power. The loss of mCherry intensity as the 405-nm laser intensity increases to ~8% was due to photobleaching. However, the CTNH-treated cells showed the loss of mCherry intensity after 5%, indicating that CTNH underwent photocleavage.
Figure 4.8 | Light-induced protein dimerization and the reversal at centromeres and plasma membrane. HeLa cells were expressing CenpB–GFP-Halo (for centromere targeting), CAAX-GFP-Halo (for plasma membrane targeting), and cytosolic mCherry–eDHFR. (A) Cells were treated with 20 µM CTNH for 30 min, followed by 30-min washout prior to imaging. A 444-nm laser was used to activate the recruitment of mCherry-eDHFR to all centromeres within one second. Subsequently, a centromere (indicated by a yellow arrow) was targeted for spatiotemporal deactivation with a targetable 405-nm laser. The rapid photocleavage of CTNH resulted in reversal of the dimerization at the targeted centromere within one second. (B) Cells described above were treated under the same condition. After 444-nm laser photoactivation for mCherry recruitment to plasma membrane, 387-nm illumination was used to reverse the dimerization. Each field was imaged every second.
Reversible control of organelle transport within subcellular regions

Organelle transport is an essential biological process in eukaryotes and contributes to intracellular organization and cell polarity\textsuperscript{14}. Various tools have been developed in order to probe functions of organelle positioning\textsuperscript{15,16}. As a functional test for CTNH, we employed a previously reported peroxisome transport assay in living cells\textsuperscript{17,18}. Peroxisomes are membrane-bound organelles that play a crucial role in the catabolism of fatty acids, reduction of reactive oxygen species, and biosynthesis of plasmalogens and ether phospholipids, and are critical for the normal function of mammalian brains and lungs\textsuperscript{19}. Because peroxisomes are immobile under normal steady-state conditions, movement induced by light is easily observed, providing an ideal assay to test our new dimerizer.

To control peroxisome transport, we employed HeLa cells expressing PEX3-GFP-Halo, a peroxisome-targeting sequence fused to GFP-haloenzyme, and KLC1-mCherry-eDHFR, a fusion protein of mCherry-eDHFR linked to kinesin light chain 1 (KLC1) that binds and recruits kinesin heavy chain (Figure 4.9A). Prior to illumination, the peroxisomes spread throughout the cell, and KLC1-mCherry-eDHFR diffused in the cytosol (Figure 4.9B). After global activation of CTNH with a 444-nm laser, the peroxisomes were initially transported towards the plus ends of the microtubules at the cell periphery. Then, half of the cell was targeted for cleavage of CTNH with a 405-nm laser (Figure 4.9B, dashed yellow region). As a result, the induced peroxisome transport was disrupted in this region with no significant change in the peroxisome density after 405-nm illumination and no accumulation of peroxisomes at the cell periphery (Figure
In contrast, peroxisome transport continued in the other half of the cell, which was not illuminated with the 405-nm laser, leading to a decrease in GFP intensity over time in the interior of this region and an increase at the periphery (Figure 4.9B and C, regions 2 and 3). At the cell periphery, the ratio of the GFP intensity in the uncleaved region to the cleaved region increased significantly after 405-nm illumination, indicating that peroxisomes only accumulated in the uncleaved region (Figure 4.9D). Together, these results demonstrate that we can successfully employ light to induce and disrupt peroxisome transport with spatiotemporal control.
Figure 4.9 | Controlling peroxisome transport with light.

(A) Schematic and protein constructs. (B-D) HeLa cells expressing PEX3-GFP-Halo and KLC1-mCherry-eDHFR were incubated with 20 µM CTNH for 30 min followed by a 30-min washout prior to imaging. CTNH was uncaged with 444-nm light at t = 0 min to recruit KLC1-mCherry-eDHFR to peroxisomes. Half of the cell was illuminated with a 405-nm laser at t = 1 min to cleave CTNH and release KLC1 from peroxisomes. In a representative cell (b), dashed white and yellow lines indicate the cell outline and the photocleaved area, respectively, and brown square insets show KLC1 release from peroxisomes in the cleaved area. GFP intensity was quantified (c) in interior and peripheral regions in both the cleaved (regions 1 and 4) and uncleaved (regions 2 and 3) areas. Intensity in each region over time was shown as a fraction of the maximal intensity observed in that region. The average GFP intensity (as a proxy for peroxisome intensity) at the cell periphery was quantified (d) as a ratio of region 3 to region 4, both before uncaging with 444-nm light and after cleavage with 405-nm laser at the final timepoint (t = 35 min) over 9 cells. Each
dot represents an individual cell and the mean±SEM was shown (n = 9 cells pooled from three independent experiments). ****P < 0.005, paired Student’s t-test. Scale bars, 10 µm.

Activating and silencing mitotic checkpoint signaling

To achieve proper chromosome segregation in mitosis, the spindle checkpoint is initially activated until all chromosomes are properly attached to the spindle and then silenced at metaphase to allow the cell to progress to anaphase. For checkpoint proteins, localization to and release from kinetochores are essential steps in signal transduction that determine cell fate. With CTNH, we aimed to optically control both checkpoint activation and silencing by manipulating kinetochore localization of the checkpoint protein Mad1 (Figure 4.10A). HeLa cells expressing Halo-GFP-SPC25, which localizes to kinetochores, and eDHFR-mCherry-Mad1 were incubated with CTNH. To demonstrate that the checkpoint can be re-activated with light, we targeted metaphase cells and recruited Mad1 from the cytosol to kinetochores using 444-nm light. As a result, <20% of cells with Mad1 recruited to kinetochores proceeded to anaphase within 30 min while >80% of control cells that were not exposed to 444-nm light proceeded to anaphase normally (Figure 4.10B and D). To show that the recruitment can be reversed by light, we globally recruited Mad1 to kinetochores and used 405-nm laser to cleave CTNH in some of the 444-nm light activated cells, and all the cells were followed for >30 min (Figure 4.10C). Without cleavage, Mad1 remained at kinetochores, and <10% of the cells proceeded to anaphase within 30 min, indicating that the mitotic checkpoint was active. In comparison, >60% of the cells in which Mad1 was released from the
kinetochores entered anaphase (Figure 4.10C and D). This re-activation and silencing of the mitotic checkpoint is consistent with our previous observations using a dimerizer that is not caged but can be photocleaved. However, with CTNH, both checkpoint activation and silencing can be controlled with light. These results demonstrate that we can use CTNH to manipulate kinetochore function at the single-cell level.
Figure 4.10 | Activating and silencing the spindle assembly checkpoint.

(A) Experimental design using CTNH for light-induced recruitment of Mad1 to and release from kinetochores. (B-D) HeLa cells expressing Halo-GFP-SPC25 and mCherry-eDHFR-Mad1 were incubated with 10 µM CTNH for 1 h followed by a 30-min washout. (B) Upon illumination with 444-nm light in some cells, Mad1 was recruited to metaphase kinetochores and cells arrested in metaphase. (C) Without 444 nm light, cells proceeded to anaphase. (D) Quantification of cells arrested in metaphase, with error bars indicating standard deviation.
metaphase (left panel). Cells without Mad1 recruitment to metaphase kinetochores proceeded to anaphase normally (right panel). Brown square insets show the zoom-in regions in mCherry channel. (C) After global illumination with 444-nm light, Mad1 was released from kinetochores in some cells using a 405-nm laser to target the metaphase plate (c, right cell, yellow region), while other cells were not exposed to 405-nm light. Brown square insets show the zoom-in regions in mCherry channel of the left and right cell accordingly. (D) The percentage of cells entering anaphase within ~30 minutes after 444-nm uncaging in (b) and after 405-nm photocleavage in (c) was quantified for both cell populations. Error bars represent s.d. (number of cells in each condition, n = 15 for uncaged; n = 21 for caged; n = 20 for uncleaved; n = 25 for cleaved population, pooled from three independent experiments). Scale bars: 5 μm (b), and 10 μm (c).
Conclusions (Chapter 4)

In summary, we have developed a novel cell-permeable chemical inducer that can rapidly produce a discrete ON and then OFF state for spatiotemporal control of protein dimerization. Our results highlight the advantages of a hybrid chemical and genetic approach and of our modular design, which facilitates the development of new dimerizers with additional properties tailored for specific purposes. We demonstrate that many subcellular locations, such as plasma membrane, mitochondria, peroxisomes, centromeres and kinetochores, can be targeted with spatial precision for both recruitment and release of a cytosolic protein on a timescale of seconds. We demonstrated that peroxisome transport can be rapidly induced and disrupted on a subcellular length scale. Additionally, we applied this tool to manipulate kinetochore function by controlling mitotic checkpoint signaling. Our future goal is to use this dimerizer in conjunction with other orthogonal dimerizers to control multiple target proteins within a cell. This approach opens up a new avenue for a variety of biological studies, such as organelle transport, signal transduction, and cell division. We envision that our new chemical tool will be readily adopted to answer new biological questions that would be otherwise difficult to address with conventional methods.
Concluding remarks and future directions

One of the biggest advantages of our system is the modular design that facilitates the opportunity to develop new dimerizers with additional properties on the same protein platform using light as an external control. Both Haloligand and TMP are biorthogonal in mammalian cells, providing minimal interference with biological pathways. The dimerization can be rapidly activated and reversed using a single pulse of light; therefore, sustained dimerization and its reversal can be achieved at high spatial and temporal resolution on a single-cell or sub-cellular levels. The wavelengths used for photo-activation are non-invasive and suitable with common excitation wavelengths for imaging such as GFP and mCherry. To the best of our knowledge, CTNH is the most advanced dimerizer that has been reported. We demonstrate that our new optogenetic tools are applicable to reversibly controlled highly dynamic processes such as peroxisome transport and mitotic checkpoint signaling. We envision that our new tools will be readily adapted for studying many other biological events, such as apoptosis, autophagy, cell cycle, gene regulations, stem cell differentiation, enzyme activities, and numerous cell signaling pathways including in cancer and other diseases.

Our current efforts are to expand the CID toolkits based on our system. We are developing new molecules with functionally diverse properties, which highlight the power of our modular design (unpublished, Figure 4.11). HaloTag can be replaced by another widely utilized covalent protein labeling system called SNAP-Tag. We have synthesized non-caged TMP-SNAP, TS (10) and tested in a live cell assay (unpublished). 10 could globally induced eDHFR and SNAP fusion proteins as TH (1). We then
developed the photocaged analogs NVOC-TMP-SNAP, NTS (11) and NVOC-TMP-SNAPCl, NTSCI (12) but surprisingly, 11 failed to induce dimerization while 12 is not as cell-permeable as its counterpart NTH (2).

Furthermore, protein molecules can interact to form a physiologically active ternary complex. For example, the Silent Information Regulators (SIR) Sir2p, Sir3p, and Sir4p assemble to form a heterotrimeric protein complex that is required for gene silencing at telomeres. Conversely, Sir2p alone forms a homotrimeric histone deacetylase that is required for rDNA repression. The ability to simultaneously control the interaction of three different proteins at a specific subcellular location will allow us to further investigate and gain insight into more complex biological networks with the complementary protein dimerization techniques. To date, there have not been reports of a chemical tool that is able to accomplish this process; therefore, we envision that our CID system could be tethered with the third ligand to previous dimerizers, we can create a new platform for chemically-induced trimerization (CIT). We have synthesized the non-caged tri-TMP-Halo-SNAP, tTHS (13), tri-TMP-Halo-SNAPCl, tTHSCI (14), and the caged TMP tri-NVOC-TMP-Halo-SNAPCl, tNTHSCI (15) (Figure 4.11). Unfortunately, our preliminary results showed that these molecules were not cell-permeable enough to be useful, likely due to the high molecular weights. Further investigations and structural modifications are necessary to improve these molecules.

Finally, by attaching the SNAP ligand to the cage of NTH, SNAPCl-NVOC-TMP-Halo, SNTH (16) is synthesized which is able to swap dimerized protein for a different protein upon illumination. This novel swappable protein dimerizer is an exciting
tool by itself and can be combined with existing complementary tools for studying complex signaling networks in living cells.
Figure 4.11 | Chemical structures of dimerizers in progress.
REFERENCES


CHAPTER 5

MATERIALS, METHODS, SYNTHESIS, AND CHARACTERIZATION
General information for chemistry

All commercially available reagents and solvents were used as received. All chemicals were purchased from Sigma Aldrich and Fisher Scientific, unless otherwise specifically noted. Boc-anhydride, 7-(Diethylamino)-4-(methyl)coumarin, CDI, and NaBH₄ were purchased from Chem-Impex International. 2-(2-aminoethoxy)ethanol, NaH (60% in mineral oil), and 48% aq.HBr, tetraethylene glycol, fuming nitric acid, succinic anhydride, DMAP, and DBU were purchased from Acros Organics. 4-Nitrophenyl chloroformate was purchased from Alfa Aesar. 6-Chloro-1-iodohexane, 4,5-dimethoxy-2-nitrobenzyl chloroformate (NVOC) were purchased from both Sigma Aldrich and Fisher. 6-Bromo-1-chlorohexane and vanillin were purchased from TCI America. Tert-butyl-3-iodopropylcarbamate was purchased from Ace Synthesis LLC. Trimethoprim was purchased from Astatech, Inc, and Chem-Impex International. HATU was purchased from GenScript and Oakwood Chemical. Selenium dioxide and CBr₄ were purchased from Acros, Alfa Aesar, and Fisher.
Instrumentations

- Thin-layer chromatography was performed on Sorbent Technologies silica plates (250 µm thickness).
- Silica gel column chromatography was performed using Silicycle silica gel (55–65 Å pore diameter) by manual and an automated system. Automated flash column chromatography was performed using RediSep Rf silica gel on CombiFlash Rf+ system with internal UV detector and also evaporative light scattering detector (ELSD). The instrument was purchased from Teledyne Isco, Inc., NE., USA. Generous technical support is with the help from Constance D. Reber, the sale representative.
- High-performance liquid chromatography (HPLC) was performed for most of the final product purification using a JASCO PU-2080 Plus Intelligent HPLC Pump and Phenomenex Luna C18 columns (5 µm particle size, 100 Å pore size). HPLC-grade water with 0.1% TFA added and acetonitrile are commonly used for gradient elution but water with TFA is also used in some cases.
- Infrared (IR) spectra were obtained on Jasco FT-IR Spectrum BX system and reported as wavenumber of the absorption maxima between 4000 cm⁻¹ and 800 cm⁻¹ of only major peaks.
- Ultraviolet-visible absorption spectrophotometry was performed on a JASCO V-650 spectrophotometer with a PAC-743R multichannel Peltier using quartz cells with a 1 cm cell path length.
• Proton nuclear magnetic resonance spectroscopy (\(^1\)H NMR) and Carbon nuclear magnetic resonance spectroscopy (\(^{13}\)C NMR) spectra were recorded on a Bruker UNI400, AVII 500, and BioDRX 600 NMR (for most 2D NMR experiments) with the support from NMR facility directors; Dr. George Furst and Dr. Jun Gu.

• Low-resolution mass spectrometry (LCMS) or nominal mass accuracy. A walk-up LCMS machine was used to obtain data. The instrument model: Waters Acquity UPLC system equipped with a Waters TUV detector (254 nm) and a Waters SQD single quadrupole mass analyzer with electrospray ionization (ESI). LC gradient 500 uL/min: 30 second hold 95:5 (water:acetonitrile 0.1% v/v formic acid), 2 minute gradient to 5:95, and 30 second hold. Acquity UPLC BEH C18, 1.7um, 2.1x 50 mm column. Diode-array detector for dual wavelength of 254 nm and 330 nm detection was also used in the instrument.

• High-resolution mass spectrometry (HRMS) were obtained at the University of Pennsylvania’s Mass Spectrometry Service Center with the support from the former mass spectrometry facility director, Dr. Rakesh Kohli on a Micromass AutoSpec electrospray/chemical ionization spectrometer. An upgraded system supported by the new director, Dr. Charles W. Ross III., for accurate mass measurement analyses were conducted on either a Waters GCT Premier, time-of-flight (TOF), GCMS with electron ionization (EI), or an LCT Premier XE, TOF, LCMS with electrospray ionization (ESI). Samples were taken up in a suitable solvent for analysis. The signals were mass measured against an internal lock mass reference of perfluorotributylamine (PFTBA) for EI-GCMS, and leucine
enkephalin for ESI-LCMS. Waters software calibrates the instruments, and reports measurements, by use of neutral atomic masses. The mass of the electron is not included.

**Synthesis Notes**

- All syntheses were done in the dark room with some Red LED Strip light (650 nm) to prevent all light-sensitive compounds from accidental exposure to light and decomposition of the compounds.
- Reactions are typically done under inert atmosphere (argon or nitrogen)
- All solvent mixtures used in the synthesis and purification as well as TLC solvent are prepared as volume percent (v/v %).
- Room temperature (rt) is normally controlled to be at 25 °C
- Overnight reactions are stirred for ~16 hours
**Synthesis details for Scheme 2.1**

**Synthesis of tert-butyl (2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)carbamate, Halo-NBoc (20)**

19 was first synthesized following the precedent of Singh et. al\(^1\). To a solution of 2-(2-aminoethoxy)ethanol 18 in anhydrous EtOH was added Boc\(_2\)O at 0 °C. After stirring at room temperature (rt) for 2 h, the reaction mixture was evaporated. The product was then extracted with DCM and the combined organic layers were dried over Na\(_2\)SO\(_4\) and evaporated under vacuum to obtain the product 19 as colorless oil in quantitative yield, which was used for the next step without further purification. The spectral data were in agreement with the reported data\(^1\).

To a solution of 19 in a 2:1 mixture of THF and DMF at 0 °C was added NaH (60% in mineral oil). After stirring at 0 °C for 30 min, 6-chloro-1-iodohexane was added to the above solution. The reaction mixture was stirred overnight (16 h) and quenched with saturated NH\(_4\)Cl. The mixture was extracted with EtOAc, washed with H\(_2\)O and brine. The combined organic layers were dried over Na\(_2\)SO\(_4\) and concentrated. The crude product was purified by silica gel column chromatography using EtOAc:Hexanes (20% to 30%) to yield pure product 20 (49%) as colorless oil. The spectral data were in agreement with the reported data\(^1\).
Synthesis of 4-((2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)amino)-4-oxobutanoic acid, Halo-COOH (21)

Compound 21 was synthesized by modification of a procedure described by Passemard et al\textsuperscript{2}. To a solution of 20 (1.0 g, 3.21 mmol) in 10 mL anhydrous DCM at 0 °C was slowly added TFA (10 mL, 130 mmol). Thereafter the reaction mixture was warmed to rt and stirred for 2 h. After completion of the reaction as evident by TLC analysis, the solvent was removed under high vacuum to obtain the crude deprotected product as a TFA salt, which was used without further purification.

To a solution of the above deprotected product (0.72 g, 3.21 mmol) in DCM (5 mL) was slowly added triethylamine (19.26 mmol, 1.95 g) and subsequently succinic anhydride (9.63 mmol, 0.96 g) followed by stirring overnight (16 h) at 25°C. Next, the reaction mixture was washed with 1 M aq. HCl (3×5 mL) followed by brine. The organic layer was dried over MgSO\textsubscript{4} and concentrated under vacuum to afford 21 as a brown oil (90% in two steps).

\( R_f = 0.45 \) (5% MeOH:DCM).

\textbf{IR} (NaCl, thin film): \( \nu \) 3320, 2936, 2865, 1733, 1653, 1558, 1436, 1172, 1116 cm\textsuperscript{-1}.

\textbf{\textsuperscript{1}H NMR} (500 MHz, CDCl\textsubscript{3}) \( \delta \) ppm: 6.47 (s, 1H), 3.65 - 3.58 (m, 4H), 3.58 - 3.41 (m, 8H), 2.70-2.63 (m, 2H), 2.52 (dd, \( J = 7.5, 5.8 \) Hz, 2H), 1.77 (m, 6.8 Hz, 2H), 1.68 - 1.52 (m, 2H), 1.50 - 1.41 (m, 2H), 1.41 - 1.31 (m, 2H).
\(^{13}\)C NMR (126 MHz, CDCl\(_3\)) \(\delta\) ppm: 175.3, 172.5, 77.4, 71.5, 70.2, 70.2, 69.7, 45.2, 39.6, 32.6, 31.3, 30.3, 26.8, 25.4.

HRMS (ESI, m/z): Calcd. for C\(_{14}\)H\(_{26}\)ClNO\(_5\) [M+Na]\(^+\): 346.1397. Found: 346.1400.

The known compounds, \(22\) and \(23\), can be synthesized following the literature precedent\(^3\).

**Synthesis of tert-butyl (3-(4-((2,4-diaminopyrimidin-5-yl)methyl)-2,6-dimethoxyphenoxy)propyl)carbamate, TMP-NBoc (25)**

![Chemical Structure](image)

The known compound \(24\) was synthesized following Jing et al\(^4\).

Compound \(24\), tert-butyl N-(3-iodopropyl)carbamate, and cesium carbonate were dissolved in anhydrous DMF. The reaction mixture was heated at 70 °C for 7 h, followed by the removal of solvent under high vacuum. The crude mixture was purified by column chromatography with silica gel MeOH:DCM (10%, v/v) to yield product \(25\) as a brown amorphous solid (42%) The spectral data were in agreement with the reported data\(^4\).

\(R_f = 0.70\) (20% MeOH:DCM).

**UV-Vis** \((\lambda_{\text{max}}\) in DCM): 239, 282 nm.

**IR** (NaCl, thin film): \(v\) 3338, 1620, 1506, 1456, 1242, 1124 cm\(^{-1}\).
**$^1$H NMR** (500 MHz, MeOD) δ ppm: 7.51 (s, 1H), 6.51 (s, 2H), 3.94 (t, J = 5.9 Hz, 2H), 3.78 (s, 6H), 3.63 (s, 2H), 3.28 (d, J = 6.7 Hz, 2H), 1.88 - 1.76 (m, 2H), 1.43 (s, 9H).

**$^{13}$C NMR** (126 MHz, MeOD) δ ppm: 164.3, 163.1, 158.4, 155.9, 154.6, 136.4, 108.0, 106.6, 80.0, 72.3, 56.5, 39.1, 39.0, 34.4, 30.8, 28.8.

**HRMS** (ESI, m/z): Calcd. for C$_{21}$H$_{31}$N$_5$O$_5$ [M+H]$^+$: 434.2403. Found: 434.2413.

**Synthesis of tert-butyl (3-(4-((4-amino-2-(((4,5-dimethoxy-2-nitrobenzyl)oxy) carbonyl)amino)-pyrimidin-5-yl)methyl)-2,6-dimethoxyphenoxy)propyl)carbamate (26)**

Reaction conditions were adapted from Wysocki et al.$^5$. To a stirring solution of compound 25 (0.63 g, 1.45 mmol) and 1 equiv. of DIEA (0.188 g, 1.45 mmol) in DCM was added 4,5-dimethoxy-2-nitrobenzyl chloroformate (NVOC) (0.40 g, 1.45 mmol) in the dark. The mixture was stirred at room temperature overnight (16 h). The mixture was washed with H$_2$O and brine. The combined organic layers were dried over Na$_2$SO$_4$ and concentrated under vacuum. The crude product was purified by silica gel column chromatography using MeOH:DCM (1% to 5%) to yield pure product 26 (41%) as a yellowish semi-solid material.

$R_f$ = 0.50 (5% MeOH:DCM).
UV-Vis ($\lambda_{\text{max}}$ in DCM): 239, 282, 346 nm.

IR (NaCl, thin film): ν 3375, 2937, 1699, 1589, 1522, 1460, 1331, 1277, 1221, 1172, 1126, 1072, 914, 731 cm$^{-1}$.

$^1$H NMR (500 MHz, DMSO-d$_6$) δ ppm: 10.02 (s, 1H), 7.78 (s, 1H), 7.72 (s, 1H), 7.45 (s, 1H), 6.74 (d, J = 5.7 Hz, 1H), 6.66 (s, 2H), 6.60 (s, 2H), 5.43 (s, 2H), 3.91 (s, 3H), 3.87 (s, 3H), 3.80 (t, J = 6.3 Hz, 2H), 3.72 (s, 6H), 3.63 (s, 2H), 3.09 (q, J = 6.6 Hz, 2H), 1.70 (t, J = 6.7 Hz, 2H), 1.37 (s, 9H).

$^{13}$C NMR (126 MHz, DMSO-d$_6$) δ ppm: 162.4, 155.9, 155.6, 153.6, 152.9, 151.6, 138.9, 134.8, 134.6, 127.8, 111.7, 110.4, 108.1, 106.0, 77.5, 70.5, 62.7, 56.4, 55.9, 55.9, 39.1, 37.4, 32.9, 30.0, 28.3.

HRMS (ESI, m/z): Calcd. for C$_{31}$H$_{40}$N$_{6}$O$_{11}$ [M+H]$^+$: 673.2833. Found: 673.2835.

Synthesis of N1-(2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)-N4-(3-(4-((2,4-diaminopyridin-5-yl)methyl)-2,6-dimethoxyphenoxy)propyl)succinamide, TMP-Halo (1)

![Chemical Structure](image)

Compound 25 (60 mg, 0.138 mmol) was deprotected by directly dissolving in 4 mL 50% TFA:DCM and stirring at room temperature for 2 h. Next, excess TFA was removed under high vacuum to yield the TFA salt as a colorless solid. $R_f = 0.02$ in 10% MeOH:DCM. The crude product was used in the next step without further purification.
The crude mixture above was dissolved in 1 mL of DMF followed by addition of DIEA (17.9 mg, 25 µL, 0.138 mmol). After stirring for 10 minutes, a 1-mL solution of 21 (45 mg, 0.138 mmol) in DMF, DIEA (17.9 mg, 25 µL, 0.138 mmol), and HATU (52.6 mg, 0.138 mmol) was added to the crude mixture. The mixture was then allowed to stir for another 2 h before being concentrated under high vacuum. The crude mixture was purified by silica gel column chromatography using 3% MeOH:DCM to yield pure product 1 (69%) as a brown oil.

\[ \text{R}_f = 0.55 \text{ (5% MeOH:DCM)}. \]

**UV-Vis** (λ<sub>max</sub> in DCM): 210, 288 nm.

**IR** (NaCl, thin film): ν 3337, 2936, 1657, 1558, 1505, 1457, 1232, 1125 cm<sup>-1</sup>.

\[ ^1\text{H NMR} \ (500 \text{ MHz, CDCl}_3) \ δ \text{ ppm} : 7.57 \ (s, 1H), 7.06 \ (t, J = 5.8 \text{ Hz, 1H}), 6.62 \ (d, J = 6.3 \text{ Hz, 1H}), 6.38 \ (s, 2H), 5.58 \ (s, 2H), 5.40 \ (s, 2H), 4.01 \ (t, J = 5.5 \text{ Hz, 2H}), 3.79 \ (s, 6H), 3.67 - 3.25 \ (m, 22H), 2.50 \ (s, 4H), 1.87 \ (p, J = 5.7 \text{ Hz, 2H}), 1.81 - 1.69 \ (m, 2H), 1.59 \ (p, J = 6.9 \text{ Hz, 2H}), 1.49 - 1.39 \ (m, 2H), 1.39 - 1.29 \ (m, 2H). \]

\[ ^{13}\text{C NMR} \ (126 \text{ MHz, CDCl}_3) \ δ \text{ ppm}: 172.4, 172.0, 163.2, 153.6, 135.6, 133.6, 107.1, 105.3, 77.4, 72.4, 71.4, 70.4, 70.1, 69.9, 56.3, 45.2, 39.4, 38.0, 34.6, 32.6, 31.9, 31.8, 29.5, 29.3, 26.8, 25.5. \]

**HRMS** (ESI, m/z): Calcd. for C<sub>30</sub>H<sub>47</sub>ClN<sub>6</sub>O<sub>7</sub> [M+H]<sup>+</sup>: 639.3273. Found: 639.3267.
Synthesis of 4,5-dimethoxy-2-nitrobenzyl (4-amino-5-((21-chloro-5,8-dioxo-12,15-dioxato-4,9-diazahenicosyl)oxy)-3,5-dimethoxybenzyl)pyrimidin-2-yl)carbamate, para-NVOC-TMP-HTag (2)

Compound 26 was deprotected and reacted with compound 21 using the same procedure for 1 as mentioned above. The crude mixture of product 2 was purified by silica gel column chromatography using 1% MeOH:DCM to yield pure product 2 (70%) as brown oil.

\[ R_f = 0.45 \ (5\% \ MeOH:DCM). \]

**UV-Vis** (\(\lambda_{\text{max}}\) in DCM): 238, 280, 340 nm.

**IR** (NaCl, thin film): \(\nu\) 3417, 3359, 3149, 2939, 1726, 1647, 1624, 1585, 1520, 1466, 1325, 1271, 1250, 1223, 1124, 1070, 984, 912, 839, 729 cm\(^{-1}\).

**\(^1\)H NMR** (500 MHz, DMSO-d\(_6\)) \(\delta\) ppm: 9.97 (s, 1H), 7.83 (t, \(J = 5.7\) Hz, 1H), 7.78 (s, 1H), 7.73 (d, \(J = 10.2\) Hz, 2H), 7.45 (s, 1H), 6.61 (d, \(J = 15.9\) Hz, 4H), 5.43 (s, 2H), 3.92 (s, 2H), 3.87 (s, 2H), 3.82 (t, \(J = 6.3\) Hz, 2H), 3.72 (s, 6H), 3.61 (dd, \(J = 14.8, 8.2\) Hz, 4H), 3.53 - 3.42 (m, 6H), 3.37 (m, 6.2 Hz, 4H), 3.30 (s, 2H), 3.23 - 3.12 (m, 4H), 2.29 (m, 4H), 1.70 (m, 4H), 1.47 (m, 2H), 1.42 - 1.19 (m, 8H).

**\(^{13}\)C NMR** (126 MHz, DMSO-d\(_6\)) \(\delta\) ppm: 171.4, 171.1, 162.3, 155.8, 154.9, 153.6, 152.9, 151.5, 147.6, 138.9, 134.9, 134.5, 127.7, 111.7, 110.5, 108.1, 106.1, 70.4, 70.1,
69.5, 69.4, 69.1, 62.6, 56.4, 56.1, 55.9, 45.3, 38.5, 35.9, 32.9, 32.0, 30.8, 30.8, 29.8, 29.0, 29.0, 26.1, 24.9.

HRMS (ESI, m/z): Calcd. for C_{40}H_{56}ClN_{7}O_{13} \ [M+H]^+: 878.3703. Found: 878.3704.
Synthesis details for Scheme 3.2

Compounds 21 was synthesized as described previously.

Synthesis of 7-(diethylamino)-4-(hydroxymethyl)-2H-chromen-2-one, Coumarin-OH (29)

Compound 29 was synthesized following a modified procedure described by Schönleber et al.⁶.

Selenium dioxide (3.33 g, 30.0 mmol) was added to a solution of commercially available 4-methyl-7-diethylaminocoumarin (4.63 g, 20.0 mmol) in p-xylene (120 mL), and heated to reflux under inert atmosphere with vigorous stirring. After 24 h, the mixture was cooled down to room temperature (rt) and added additional selenium dioxide (1.50 g, 13.5 mmol). The crude mixture was continued refluxing for another 48 h. After refluxing for 3 days, the crude black mixture was filtered through celite and silica gel followed by concentration under reduced pressure at 50 °C. The dark brown residual oil was redissolved in EtOH:THF (1:1, 130 mL) in a 500 mL round-bottom flask, sodium borohydride (760 mg, 20.0 mmol) was added portionwise while stirring, and the solution was stirred for 4 h at rt. Thereafter, the crude mixture was carefully quenched and
hydrolyzed with 1 M HCl (20 mL) and diluted with H$_2$O (50 mL). The organic solvents (EtOH and THF) were partially removed under reduced pressure before the aqueous solution was extracted with DCM (3×100 mL). The organic phase was washed with H$_2$O and brine, dried over MgSO$_4$, and then concentrated in vacuo. The resulting crude mixture was purified by column chromatography (Acetone:DCM, 1:5) to obtain 494 mg (10% over two steps) of the alcohol product as a dark yellow semi-solid. The spectral data were in agreement with the reported data\textsuperscript{6}.

$R_f = 0.30$ (5% MeOH:DCM).

$^1$H NMR (500 MHz, Chloroform-$d$): 7.27 (d, $J = 9.0$ Hz, 1H), 6.51 (dd, $J = 9.0$, 2.6 Hz, 1H), 6.39 (d, $J = 2.6$ Hz, 1H), 6.25 (t, $J = 1.4$ Hz, 1H), 4.80 - 4.76 (m, 2H), 3.75 - 3.71 (m, 1H), 3.34 (q, $J = 7.1$ Hz, 4H), 1.15 (t, $J = 7.1$ Hz, 6H).

$^{13}$C NMR (126 MHz, Chloroform-$d$): 163.2, 156.0, 155.8, 150.6, 124.5, 108.8, 106.4, 105.1, 97.6, 60.7, 44.7, 12.5.

LCMS (ESI, m/z): Calcd. for C$_{14}$H$_{18}$NO$_3$ [M+H]$^+$: 248.13. Found: 248.27.

Synthesis of tert-butyl (3-(4-((2,4-diaminopyrimidin-5-yl)methyl)-2,6-dimethoxy-phenoxy)propyl)carbamate, TMP-NBoc (25)

24 was synthesized using modified procedure reported by Colloway et al\textsuperscript{7}.
Compound **24, TMP-OH** (4.14 g, 15.0 mmol) was dissolved in DMSO (45 mL) and DBU (2.51 g, 16.5 mmol) was added at room temperature (rt). After TMP-OH was completely dissolved and the solution turned deep red after about 10 minutes, tert-butyl-3-bromopropylcarbamate (3.93 g, 16.5 mmol) was added. The reaction mixture was stirred at rt overnight. Distilled water (150 mL) was added and the mixture was extracted with EtOAc (4×100 mL). The combined EtOAc solution was washed with water (100 mL) and brine, dried over MgSO₄, and evaporated. The crude mixture was purified by column chromatography with silica gel (10% to 20% MeOH:DCM) to yield 3.67 g (56%) of product **25** as light brown amorphous solid. The spectral data were in agreement with the reported data⁷.

**Rf** = 0.70 (20% MeOH:DCM).

**UV-Vis** (λmax in DCM): 239, 282 nm.

**¹H NMR** (500 MHz, MeOD): 7.51 (s, 1H), 6.51 (s, 2H), 3.94 (t, J = 5.9 Hz, 2H), 3.78 (s, 6H), 3.63 (s, 2H), 3.28 (d, J = 6.7 Hz, 2H), 1.88 - 1.76 (m, 2H), 1.43 (s, 9H).

**¹³C NMR** (126 MHz, MeOD): 164.3, 163.1, 158.4, 155.9, 154.6, 136.4, 108.0, 106.6, 80.0, 72.3, 56.5, 39.1, 39.0, 34.4, 30.8, 28.8.

Synthesis of tert-butyl (3-((4-amino-2-(((7-(diethylamino)-2-oxo-2H-chromen-4-yl)-methoxy)carbonyl)amino)pyrimidin-5-yl)methyl)-2,6-dimethoxyphenoxy)propyl)-carbamate, Coumarin-TMP-NBoc (30)

In a 50-mL round-bottom flask equipped with nitrogen, compound 25 (315 mg, 0.730 mmol) was dissolved in anhydrous and degassed DCM (5 mL), DIEA (150 µL, 0.730 mmol), DMAP (45 mg, 0.36 mmol). The resulting light brown suspension was stirred for 15 minutes at room temperature (RT). After the mixture became less cloudy, 1,1′-Carbonyldiimidazole (177 mg, 1.09 mmol, CDI, ≥98% purity, purchased from Chem Impex Int, Inc. and kept dried in a desiccator) was added to the mixture. Thereafter, the suspension was heated to reflux and the solution became transparent. Heating was continued to reflux, under nitrogen atmosphere, overnight. The reaction mixture was then cooled to rt. The reaction progression was monitored by LCMS as follows: a small amount of the solution was taken, dissolved in ~500 µL of MeOH, injected on LCMS (Waters, ESI+ mode). The formation of CDI-TMP-NBoc could be observed as a MeOH adduct of MeOH-trapped CDI-TMP-NBoc intermediate, [M+H⁺]: 550.24, the structure shown below.
In a separate small vial, Coumarin-OH, compound **29** (180 mg, 0.730 mmol), was dissolved in DCM (2 mL), DIEA (150 µL, 0.730 mmol), DMAP (45 mg, 0.36 mmol) and then added into the above solution. Subsequently, the mixture was heated to reflux overnight and reaction progression was monitored by TLC and LCMS. A small amount of the solution was taken, dissolved in ~500 µL of MeOH, injected on LCMS (Waters, ESI+ mode) for analysis. The addition of one Coumarin-OH was observed with two isomers. However, the major isomer was Coumarin-TMP-NBoc, **30**, LCMS (ESI+) [M+H\(^+\)]: 707.33. The reaction mixture was then cooled to rt, quenched with H\(_2\)O (20 mL), and diluted with DCM (30 mL) and separated. The organic layer was washed with saturated aqueous solution of NH\(_4\)Cl once and then brine, dried with anhydrous Na\(_2\)SO\(_4\), concentrated to get dark brown oil. The crude mixture was purified by column chromatography with silica gel (2% to 20% MeOH:DCM) to yield 60 mg (12% over two steps) of product **30** as dark brown amorphous solid.

\[ R_f = 0.40 \text{ (10\% MeOH:DCM).} \]

**UV-Vis** (\(\lambda_{\text{max}}\) in 20% DMSO/PBS buffer pH 7.4): 239, 282, 381 nm.

**IR** (NaCl, thin film): v 3309, 2934, 1717, 1605, 1527, 1506, 1424, 1356, 1206, 1126 cm\(^{-1}\).

\(^{1}\text{H NMR}\) (500 MHz, DMSO-\(d_6\)): 9.99 (s, 1H), 7.80 (s, 1H), 7.47 (d, \(J = 9.0\) Hz, 1H), 6.78 - 6.51 (m, 7H), 6.19 (d, \(J = 1.5\) Hz, 1H), 5.30 (d, \(J = 1.5\) Hz, 2H), 3.81 (t, \(J = 6.3\) Hz, 2H), 3.72 (s, 6H), 3.64 (s, 2H), 3.43 (q, \(J = 7.0\) Hz, 4H), 3.09 (q, \(J = 6.5\) Hz, 2H), 1.70 (p, \(J = 6.6\) Hz, 2H), 1.37 (s, 9H), 1.12 (t, \(J = 7.0\) Hz, 6H).
$^{13}$C NMR (126 MHz, DMSO-$d_6$): 155.7, 155.5, 155.0, 152.8, 151.4, 150.4, 134.5, 125.3, 111.8, 108.7, 106.0, 105.2, 104.6, 96.8, 77.4, 70.5, 61.1, 55.9, 44.0, 39.1, 37.32, 32.9, 30.0, 28.2, 12.3.

HRMS (ESI, m/z): Calcd. for C$_{36}$H$_{47}$N$_6$O$_9$ [M+H]$^+$: 707.3405 Found: 707.3405

(7-(diethylamino)-2-oxo-2H-chromen-4-yl)methyl (4-amino-5-(4-((21-chloro-5,8-dioxo-12,15-dioxo-4,9-diazahenicosyl)oxy)-3,5-dimethoxybenzyl)pyrimidin-2-yl)carbamate, Coumarin-TMP-Halo, CTH (5)

![Chemical structure](image)

Compound 30 (60 mg, 0.085 mmol) was deprotected by directly dissolving in 2 mL of TFA:DCM (1:3) and stirring at room temperature (rt) for 2 h. Next, excess TFA was removed and co-evaporated with DCM and DMF several times under high vacuum to yield the TFA salt as a light brown semi-solid. $R_f = 0.02$ in 10% MeOH:DCM. The primary amine salt can be stained on TLC using Ninhydrin solution, a pink spot close to baseline appeared upon heated. The crude product was used in the next step without further purification.

The crude mixture above was dissolved in 1 mL of DMF followed by addition of DIEA (~200 µL) and stirring at room temperature for 10 minutes. In a separate flask, a 1-
mL solution of 21 (28 mg, 0.085 mmol) in DMF, DIEA (~200 µL), and HATU (65 mg, 0.17 mmol) were stirred at rt for 10 minutes and then added to the above stirring mixture of the deprotected Coumarin-TMP-NH$_2$ TFA salt. The mixture was then allowed to stir at rt for another 16 h before being concentrated under high vacuum. The crude brown oil was dissolved in DCM (50 mL) and washed with saturated NH$_4$Cl solution and brine, dried with anhydrous Na$_2$SO$_4$, and concentrated under reduced pressure. The crude mixture was purified by silica gel column chromatography using 2% to 15% MeOH:DCM to yield pure product 5 (64% over two steps) as a brown oil.

R$_f$ = 0.40 (10% MeOH:DCM).

UV-Vis ($\lambda_{max}$ in 20% DMSO/PBS buffer pH 7.4): 238, 280, 381 nm.

$^1$H NMR (500 MHz, DMSO-$d_6$): 10.03 (s, 1H), 7.85 (t, $J = 5.6$ Hz, 1H), 7.81 - 7.72 (m, 2H), 7.44 (d, $J = 9.0$ Hz, 1H), 6.66 (m, 3H), 6.58 (s, 2H), 6.52 (d, $J = 2.5$ Hz, 1H), 6.17 (d, $J = 1.4$ Hz, 1H), 5.28 (d, $J = 1.5$ Hz, 2H), 3.80 (t, $J = 6.3$ Hz, 2H), 3.70 (s, 6H), 3.64 - 3.55 (m, 4H), 3.50 - 3.28 (m, 14H), 3.16 (m, 4H), 2.27 (m, 4H), 2.27 (s, 2H), 1.50 - 1.40 (m, 2H), 1.38 - 1.22 (m, 4H), 1.09 (t, $J = 7.0$ Hz, 6H).

$^{13}$C NMR (126 MHz, DMSO-$d_6$): 171.4, 171.2, 162.4, 160.8, 155.8, 155.7, 155.0, 152.91, 151.5, 150.4, 134.5, 125.3, 108.7, 106.0, 105.2, 104.6, 96.9, 70.5, 70.2, 69.6, 69.4, 69.1, 61.1, 55.9, 45.4, 44.0, 39.3, 38.6, 35.9, 32.9, 32.0, 30.9, 30.8, 29.8, 29.1, 26.1, 24.9, 12.3.

HRMS (ESI, m/z): Calcd. for C$_{45}$H$_{63}$ClN$_7$O$_{11}$ [M+H]$^+$: 912.4274 Found: 912.4276
Synthesis details for Scheme 3.3

Compound 31 – 38 (except compound 36) were synthesized using modified procedures reported by Erhart et al\textsuperscript{8} and Zimmermann et al\textsuperscript{9}.

Synthesis of 18-chloro-3,6,9,12-tetraoxaoctadecan-1-ol, 4EG-Halo (32)

\[
\text{HO-}\overset{\text{O}}{\text{O}}\overset{\text{O}}{\text{O}}\overset{\text{O}}{\text{O}}\overset{\text{Cl}}{\text{Cl}}
\]

Tetraethyleneglycol, 4EG, 31 (1.45 mL, 8.40 mmol) was dissolved in 18 mL of THF:DMF (2:1). Sodium hydride (60% in mineral oil, 368 mg, 9.20 mmol) was added portionwise at 0 °C. After stirring for 30 min at room temperature (rt), 1-bromo-6-chlorohexane (1.25 mL, 8.40 mmol) was added dropwise. The mixture was stirred at rt for 16 h. The excess of sodium hydride was carefully quenched with water and the crude mixture was poured into H₂O (50 mL) and extracted with EtOAc (3×50 mL). The combined organic layers were dried over Na₂SO₄ and concentrated under reduced pressure. The crude oil was purified by flash column chromatography using 2% to 5% MeOH:DCM to yield compound 32 as a light yellow oil (1.05 g, 40%). The spectral data were in agreement with the reported data\textsuperscript{8}.

\[ R_f = 0.55 \text{ (5\% MeOH:DCM).} \]
\(^{1}\)H NMR (500 MHz, Chloroform-\(d\)): 3.72 (t, \(J = 4.5 \) Hz, 2H), 3.69 - 3.55 (m, 14H), 3.53 (t, \(J = 6.7 \) Hz, 2H), 3.45 (t, \(J = 6.6 \) Hz, 2H), 2.67 (s, 1H), 1.77 (p, \(J = 6.9 \) Hz, 2H), 1.59 (p, \(J = 6.9 \) Hz, 2H), 1.49 - 1.31 (m, 4H).

\(^{13}\)C NMR (126 MHz, Chloroform-\(d\)): 72.7, 71.4, 70.8, 70.7, 70.7, 70.7, 70.5, 70.2, 61.9, 45.2, 32.7, 29.6, 26.8, 25.5.

HRMS (ESI, m/z): Calcd. for \(\text{C}_{14}\text{H}_{30}\text{ClO}_5 \) [M+H]\(^{+}\): 313.1782 Found: 313.1783

**Synthesis of 1-bromo-18-chloro-3,6,9,12-tetraoxaoctadecane, Br-4EG-Halo (33)**

\[
\begin{align*}
\text{Br} & \quad \text{O} & \quad \text{O} & \quad \text{O} & \quad \text{O} & \quad \text{Cl} \\
\end{align*}
\]

Compound \(32\) (2.10 g, 6.71 mmol) was dissolved in THF (20 mL). Triphenylphosphate (2.05 g, 7.82 mmol) and carbon tetrabromide (2.59 g, 7.82 mmol) were added portionwise at 0°C. The resulting mixture was stirred at room temperature (rt) for 16 h. The solvent was evaporated under reduced pressure and the crude oil was purified by flash column chromatography using 10% to 50% EtOAc:Hexanes to yield compound \(33\) as a light yellow oil (1.24 g, 50%). The spectral data were in agreement with the reported data\(^8\).

\(R_f = 0.60\) (40% EtOAc:Hexanes).
$^1$H NMR (500 MHz, Chloroform-$d$): 3.80 (t, $J = 6.3$ Hz, 2H), 3.68 - 3.61 (m, 10H), 3.59 - 3.55 (m, 2H), 3.52 (t, $J = 6.7$ Hz, 2H), 3.46 (q, $J = 6.5$ Hz, 4H), 1.77 (m, 2H), 1.59 (m, 2H), 1.49 - 1.30 (m, 4H).

$^{13}$C NMR (126 MHz, Chloroform-$d$): 71.4, 70.8, 70.8, 70.8, 70.7, 70.7, 70.3, 45.2, 32.7, 30.4, 29.6, 26.8, 25.6.

HRMS (ESI, m/z): Calcd. for C$_{14}$H$_{29}$BrClO$_4$ [M+H]$^+$: 375.0938 Found: 375.0939

**Synthesis of 4-(benzyloxy)-3-methoxybenzaldehyde (35)**

Compound 35 was synthesized using the same procedure described by Critchley et al.$^{10}$ In brief, commercially available vanillin 34 (10.0 g, 65.7 mmol) and benzylbromide (11.2 g, 65.7 mmol) were dissolved in 100 mL of DMF. K$_2$CO$_3$ (9.08 g, 65.7 mmol) was added and the mixture was stirred at 80°C for 16 h. The reaction was concentrated under high vacuum, quenched with saturated NH$_4$Cl (500 mL) and extracted with EtOAc three times. The combined organic layers were dried over Na$_2$SO$_4$ and concentrated under reduced pressure. The resulting product 35 (15.76 g, 99%) was used without further purification. The spectral data were in agreement with the reported data.$^{10}$

$R_f = 0.60$ (25% EtOAc:Hexanes).

$^1$H NMR (500 MHz, Chloroform-$d$): 9.84 (s, 1H), 7.44 (m, 3H), 7.41 - 7.36 (m, 3H), 7.36 - 7.29 (m, 1H), 6.99 (d, $J = 8.2$ Hz, 1H), 5.25 (s, 2H), 3.95 (s, 3H).
Synthesis of 4-(benzyloxy)-5-methoxy-2-nitrobenzaldehyde (35a)

Compound 35a was synthesized using the same procedure described by Critchley et al.\textsuperscript{10} In brief, the benzylated vanillin 35 (1.00 g, 4.13 mmol) and 1,2-dichloroethane (5 mL) were cooled to -30 °C under argon. Fuming nitric acid (highly corrosive) (2 mL) was added dropwise with caution, and the temperature was maintained at -15 °C for 3 h (with proper shielding equipment in a fume hood). Thereafter, the reaction mixture was poured into ice-water and extracted with EtOAc three times. Removal of the solvent gave a bright yellow amorphous solid, which was used in the next step without further purification.

\( R_f = 0.50 \) (25% EtOAc:Hexanes).

\(^1\text{H} \text{NMR} \) (500 MHz, Chloroform-\textit{d}): 10.44 (s, 1H), 7.67 (s, 1H), 7.48 -7.31 (m, 6H), 5.27 (s, 2H), 4.02 (s, 3H).

\(^{13}\text{C} \text{NMR} \) (126 MHz, Chloroform-\textit{d}): 187.9, 153.9, 151.6, 143.8, 135.0, 129.0, 128.8, 127.7, 125.9, 110.2, 109.1, 71.7, 56.9.
\textbf{HRMS (ESI, m/z): Calcd. for C}_{15}\text{H}_{14}\text{NO}_{5} [\text{M+H}]^{+}: 288.0872 \text{ Found: 288.0878}

\textit{Synthesis of 4-hydroxy-5-methoxy-2-nitrobenzaldehyde, Nitro-Vanillin (18)}

\begin{center}
\includegraphics[width=0.2\textwidth]{structure.png}
\end{center}

6-Nitro-o-benzylvanillin \textbf{35a} (4.12 g, 14.3 mmol) was dissolved in acetic acid (30 mL, 99\%), and heated to 85 °C. 48\% HBr\textsubscript{aq} (15 mL) was added to the mixture and stirred for 3 h. The mixture was poured into ice water and extracted with EtOAc three times. The combined organic layers were washed with saturated NaHCO\textsubscript{3}, brine and dried over Na\textsubscript{2}SO\textsubscript{4}. The solvent was evaporated under reduced pressure. The crude product was purified by flash column chromatography using 20\% to 60\% EtOAc:Hexanes to yield the desired product \textbf{36} as dark brown liquid (0.65 g, 23\% over two steps). The spectral data were in agreement with the reported data\textsuperscript{10}.

\begin{equation*}
R_{f} = 0.50 \text{ (40\% EtOAc:Hexanes).}
\end{equation*}

\textbf{HRMS (ESI, m/z): Calcd. for C}_{8}\text{H}_6\text{NO}_5 [\text{M-H}]^{-}: 196.0246 \text{ Found: 196.0245}
Synthesis of 4-((18-chloro-3,6,9,12-tetraoxaoctadecyl)oxy)-5-methoxy-2-nitrobenz-aldehyde (37)

\[ \text{H} \quad \text{NO}_2 \]
\[ \text{O} \quad \text{Me} \quad \text{O} \quad \text{O} \quad \text{O} \quad \text{O} \quad \text{Cl} \]

\( o \)-Hydroxy-6-nitrovanillin 36 (880 mg, 4.46 mmol) and \( \text{K}_2\text{CO}_3 \) (617 mg, 4.46 mmol) were dissolved in DMF (15 mL). Compound 33 (1.68 g, 4.46 mmol) was added to the above solution. The reaction mixture was stirred at 60°C overnight before being concentrated. Then, the crude mixture was poured into a solution of saturated \( \text{NH}_4\text{Cl} \) (300 mL) and extracted with \( \text{EtOAc} \) three times and dried over \( \text{Na}_2\text{SO}_4 \). The solvent was evaporated under reduced pressure and the crude product was purified by flash column chromatography using 40% to 80% \( \text{EtOAc:Hexanes} \) to yield the desired product 38 as brown oil (659 mg, 30%).

\( R_f = 0.50 \) (75% \( \text{EtOAc:Hexanes} \)).

\( ^1\text{H} \text{ NMR} \) (500 MHz, Chloroform-\( d \)): 10.43 (s, 1H), 7.70 (s, 1H), 7.39 (s, 1H), 4.35 - 4.29 (m, 2H), 3.99 (s, 3H), 3.95 - 3.90 (m, 2H), 3.75 - 3.70 (m, 2H), 3.69 - 3.59 (m, 8H), 3.59 - 3.54 (m, 2H), 3.52 (t, \( J = 6.7 \text{ Hz} \), 2H), 3.44 (t, \( J = 6.6 \text{ Hz} \), 2H), 1.76 (m, 2H), 1.62 - 1.53 (m, 2H), 1.48 - 1.30 (m, 4H).

\( ^{13}\text{C} \text{ NMR} \) (126 MHz, Chloroform-\( d \)): 187.9, 153.7, 152.0, 143.81, 125.7, 110.1, 108.9, 77.4, 71.4, 71.1, 70.8, 70.7, 70.2, 69.5, 69.5, 56.8, 45.2, 32.7, 29.6, 26.8, 25.6.

\( \text{HRMS} \) (ESI, m/z): Calcd. for \( \text{C}_{22}\text{H}_{35}\text{ClNO}_9 \)[M+H]\(^+\): 492.2000 Found: 492.1985
Synthesis of (4-((18-chloro-3,6,9,12-tetraoxaoctadecyl)oxy)-5-methoxy-2-nitrophe-nyl)methanol (38)

Compound 37 (610 mg, 1.24 mmol) was dissolved in 1:1 MeOH:Dioxane (16 mL). NaBH₄ (85.3 mg, 2.25 mmol) was added portionwise at 0°C. The mixture was stirred at room temperature (rt) for 2 h. Then, the mixture was poured into water, neutralized with a 1 M solution of HClₐq and extracted with DCM three times. The combined organic layers were dried over Na₂SO₄ and the solvent was evaporated under reduced pressure. The crude product was purified by flash column chromatography using 3% to 5% MeOH:DCM to yield compound 38 as brown oil in quantitative yield (613 mg).

Rᵣ = 0.55 (5% MeOH:DCM).

¹H NMR (500 MHz, Chloroform-𝑑): 7.76 (s, 1H), 7.16 (s, 1H), 4.94 (s, 2H), 4.24 (m, 2H), 3.96 (s, 3H), 3.93 - 3.88 (m, 2H), 3.75 - 3.59 (m, 10H), 3.56 (m, 2H), 3.52 (t, J = 6.7 Hz, 2H), 3.44 (t, J = 6.7 Hz, 2H), 2.81 (s, 1H), 1.80 - 1.70 (m, 2H), 1.58 (p, J = 6.8 Hz, 2H), 1.46 - 1.31 (m, 4H).

¹³C NMR (126 MHz, Chloroform-𝑑): 154.5, 147.3, 139.6, 132.8, 111.2, 110.2, 71.4, 71.0, 70.8, 70.7, 70.2, 69.6, 69.2, 62.8, 56.5, 45.2, 32.7, 29.6, 26.8, 25.5.
**HRMS** (ESI, m/z): Calcd. for C\textsubscript{22}H\textsubscript{36}ClINaO\textsubscript{9} [M+Na]\textsuperscript{+}: 516.1976 Found: 516.1971

Synthesis of 4-((18-chloro-3,6,9,12-tetraoxaoctadecyl)oxy)-5-methoxy-2-nitrobenzyl (3-(4-((2,4-diaminopyrimidin-5-yl)methyl)-2,6-dimethoxyphenoxy)propyl)-carbamate, TMP-NVOC-Halo, TNH (6)

![Chemical Structure](image)

Compound 38 (430 mg, 0.870 mmol), 4-nitrophenyl chloroformate (263 mg, 1.31 mmol), and DMAP (210 mg, 1.74 mmol) were dissolved in DCM (5 mL). The solution was stirred at room temperature (rt) under nitrogen atmosphere overnight.

Compound 25, TMP-NHBoc (430 mg, 0.992 mmol) was deprotected by directly dissolving in 4 mL of TFA:DCM (1:3) and stirring at rt for 2 h. Then, excess TFA was removed and co-evaporated with DCM several times under high vacuum to yield the TMP-NH\textsubscript{2} TFA salt as a light brown solid. R\textsubscript{f} = 0.02 in 10% MeOH:DCM. The primary amine salt can be stained on TLC using Ninhydrin solution, a pink spot close to baseline appeared upon heating. The crude TMP-NH\textsubscript{2} TFA salt was dissolved in a mixture of DIEA (150 µL) and DCM (2 mL) and subsequently added to the above solution containing compound 38, which had been stirred overnight. The reaction mixture was
stirred at rt under nitrogen atmosphere overnight. The reaction was quenched with water (10 mL), concentrated under high vacuum before being extracted with DCM three times. The combined organic layers were dried over Na$_2$SO$_4$ and the solvent was evaporated under reduced pressure. The crude product was purified by flash column chromatography using 2% to 20% MeOH:DCM to yield compound 6 as brown viscous oil (120 mg, 16% over two steps).

$R_f = 0.70$ (15% MeOH:DCM).

$^1$H NMR (500 MHz, Chloroform-$d$): 7.71 (s, 1H), 7.50 (s, 1H), 6.99 (s, 1H), 6.38 (m, 3H), 5.79 (s, broad, 1H), 5.48 (s, 2H), 4.20 (t, $J = 4.8$ Hz, 2H), 4.02 (t, $J = 5.5$ Hz, 2H), 3.88 (m, 3H), 3.84 (s, 3H), 3.77 (s, 6H), 3.69 (m, 2H), 3.66 - 3.57 (m, 10H), 3.54 (m, 2H), 3.48 (m, 4H), 3.42 (t, $J = 6.6$ Hz, 2H), 1.90 (p, $J = 5.7$ Hz, 2H), 1.78 - 1.68 (m, 2H), 1.55 (p, $J = 6.8$ Hz, 2H), 1.46 - 1.27 (m, 4H).

$^{13}$C NMR (126 MHz, Chloroform-$d$): 163.2, 156.2, 154.1, 153.6, 147.3, 139.6, 135.6, 133.3, 129.1, 110.5, 110.1, 105.2, 71.9, 71.3, 70.9, 70.6, 70.6, 70.1, 69.45, 69.0, 63.3, 56.3, 56.1, 45.1, 39.4, 34.4, 32.6, 29.5, 26.7, 25.5.

HRMS (ESI, m/z): Calcd. for C$_{39}$H$_{58}$ClN$_6$O$_{13}$ [M+H]$^+$: 853.3750 Found: 853.3748
Synthesis details for Scheme 4.2

Compound 25, 29, 33, and 38 were previously synthesized.

Synthesis of tert-butyl (3-(4-((2-amino-4-(((7-(diethylamino)-2-oxo-2H-chromen-4-yl)methoxy)carbonyl)amino)pyrimidin-5-yl)methyl)-2,6-dimethoxyphenoxy)propyl)-carbamate (ortho-DEACM-TMP-NBoc, 30b)

\[
\begin{align*}
\text{Et}_2\text{N} & \quad \text{O} & \quad \text{NH} & \quad \text{OMe} \\
\text{O} & \quad \text{N} & \quad \text{N} & \quad \text{OMe} \\
\text{H}_2\text{N} & \quad \text{N} & \quad \text{OMe} \quad \text{NHBOc}
\end{align*}
\]

30b was synthesized using a modified procedure reported by Nadler et al\textsuperscript{12}.

Briefly, previously synthesized 7-Diethylamino-(4-hydroxymethyl)coumarin, DEACM-OH, 29 (150 mg, 0.607 mmol) was dissolved in dry THF, added DIEA (0.2 mL, 1.21 mmol), and then stirred for 10 min at 0 °C. After dropwise addition of cold phosgene (1 mL, 1.50 mmol, 15 wt % solution in toluene), stirring was continued under the exclusion of light at 0 °C for additional 2 h. The reaction mixture was transferred onto a mixture of EtOAc and ice-cold H\textsubscript{2}O (1:1 v/v, 200 mL), the liquid was separated and the organic layer was dried over Na\textsubscript{2}SO\textsubscript{4}. The solvent was removed under reduced pressure and the crude [7-(diethylamino)-coumarin-4-yl]-methylchloroformate was further dried under high vacuum conditions.
In a separate flask, a solution of the TMP-NBoc (25, 132 mg, 0.303 mmol) in DCM (2 mL) was treated with DIEA (0.2 mL, 1.21 mmol) and DMAP (0.303 mmol, 37.0 mg), and then stirred for 10 min at 0 °C. The previously prepared [7-(diethylamino)-coumarin-4-yl]-methyl chloroformate was dissolved in DCM (1 ml) and slowly added to the reaction mixture of TMP-NBoc, 25, DIEA, and DMAP. The reaction mixture was stirred, protected from light and allowed to reach rt overnight. The reaction mixture was then transferred onto a mixture of DCM and water (1:1 v/v, 300 mL), the layers were separated, the organic layer washed with H₂O (1 x 100 mL), saturated aqueous NH₄Cl solution (2 x 100 mL), and brine (1 x 100 mL) and dried over MgSO₄. The solvent was removed under reduced pressure and the crude mixture was purified by column chromatography with silica gel (2% to 20% MeOH:DCM) to yield 30 mg (15 %) of product 30b as dark brown solid.

R<sub>t</sub> = 0.30 (10% MeOH:DCM).

UV-Vis (λ<sub>max</sub> in 20% DMSO/PBS buffer pH 7.4): 239, 282, 381 nm.

<sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>): 9.83 (s, 1H), 8.05 (s, 1H), 7.41 (dd, J = 9.1, 5.2 Hz, 1H), 6.67 (t, J = 5.4 Hz, 1H), 6.62 (dd, J = 9.2, 2.6 Hz, 1H), 6.49-6.46 (m, 5H), 6.18 (s, 1H), 5.32 (s, 2H), 3.86-3.73 (m, 4H), 3.69 (s, 6H), 3.36 (q, J = 6.9 Hz, 4H), 3.11 (q, J = 6.6 Hz, 2H), 1.71 (p, J = 6.8 Hz, 2H), 1.35 (d, J = 3.9 Hz, 9H), 1.07 (t, J = 7.0 Hz, 6H).

<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>): 162.5, 160.9, 160.2, 156.6, 155.9, 155.7, 153.0, 152.5, 151.1, 150.5, 135.7, 134.8, 125.3, 114.6, 108.8, 105.7, 105.3, 104.9, 96.9, 77.5, 70.7, 61.8, 55.8, 44.1, 37.5, 33.4, 30.1, 28.3, 12.3
HRMS (ESI, m/z): Calcd. for C_{36}H_{47}N_{6}O_{9} [M+H]^+ : 707.3405 Found: 707.3405

Synthesis of tert-butyl (3-(2-(2-(3-aminopropoxy)ethoxy)ethoxy)propyl)carbamate (39)

The known 39 was synthesized according to the reported procedure\textsuperscript{11}.

Briefly, bis(3-aminopropyl)diethylene glycol (26.85 g, 26.7 mL, 122 mmol) was dissolved in 1,4-dioxane (50 mL) and a solution of Boc-anhydride (4.37 g, 23.5 mmol) in dioxane (30 mL) was added dropwise at room temperature within 5 h. The mixture was stirred for additional 5 h and then the solvent was evaporated. The resulting yellowish oil was dissolved in water (50 mL) and extracted with DCM (4 × 50 mL). The organic phases were combined and washed with brine (4 × 30 mL). The extraction procedure and the subsequent washing were repeated. The resulting organic solution was dried (MgSO\textsubscript{4}), which subsequently was filtered off to give colorless oil after evaporation and drying under high vacuum (4.86 g, 65% rel. to Boc\textsubscript{2}O). TLC indicated only traces of doubly protected diamine ($R_f = 0.95$, 10% MeOH:DCM) and no starting material. The spectral data were in agreement with the reported data\textsuperscript{11}.

$R_f = 0.70$ (10% MeOH:DCM).

$^1$H NMR (500 MHz, Chloroform-d): 5.11 (s, 1H), 3.67 - 3.46 (m, 13H), 3.31 - 3.07 (m, 5H), 2.86 (t, J = 6.6 Hz, 2H), 1.76 (m, 4H), 1.41 (s, 9H).
$^{13}$C NMR (126 MHz, Chloroform-$d$): 156.3, 79.0, 70.7, 70.7, 70.6, 70.3, 70.2, 69.7, 69.6, 39.8, 38.6, 32.1, 29.8, 28.6.

HRMS (ESI, m/z): Calcd. for C$_{15}$H$_{32}$N$_2$O$_5$ [M+Na]$^+$: 343.2209 Found: 343.2209

**Synthesis of tert-butyl (1-(4-((18-chloro-3,6,9,12-tetraoxaoctadecyl)oxy)-5-methoxy-2-nitrophenyl)-3-oxo-2,8,11,14-tetraoxa-4-azaheptadecan-17-yl)carbamate (40)**

\[
\text{\begin{array}{c}
\begin{array}{c}
\text{BocH} \\
\text{O} \\
\text{O} \\
\text{\textbf{NO}}_2 \\
\text{OMe}
\end{array}
\end{array}}
\]

Compound 38 (1.79 g, 3.624 mmol), 4-nitrophenyl chloroformate (1.10 g, 5.435 mmol), DIEA (1.3 mL, 7.247 mmol), and DMAP (0.44 g, 3.624 mmol) were dissolved in DCM (18 mL). The mixture was stirred at room temperature (rt) under nitrogen atmosphere overnight.

Next, compound 39 (1.16 g, 3.624 mmol) and DIEA (0.6 mL, 3.624 mmol) were dissolved in DCM (5 mL), and then added to the above solution containing compound 38, which had been stirred overnight. The reaction mixture was continued stirring at rt for another 16 h. The reaction was quenched with H$_2$O (20 mL) and partitioned into two layers. The organic phase was extracted with saturated aqueous NH$_4$Cl solution (2 x 20 mL), brine (1 x 20 mL), and dried over MgSO$_4$. The combined organic layers were evaporated under reduced pressure. The crude mixture was purified by flash column chromatography with silica gel using 2% to 20% MeOH:DCM to yield the product 40 as brown oil (120 mg, 16%).
R_f = 0.60 (5% MeOH:DCM).

^1^H NMR (500 MHz, Chloroform-d): 7.76 (s, 1H), 6.99 (s, 1H), 5.58 (s, 1H), 5.47 (s, 2H), 4.96 (s, 1H), 4.23 (t, J = 4.9 Hz, 2H), 3.93 (s, 3H), 3.92 - 3.84 (m, 2H), 3.71 (dd, J = 6.1, 3.6 Hz, 2H), 3.69 - 3.40 (m, 34H), 3.33 (q, J = 6.1 Hz, 2H), 3.20 (q, J = 6.3 Hz, 2H), 1.83 - 1.68 (m, 7H), 1.58 (p, J = 6.9 Hz, 2H), 1.42 (s, 13H), 1.39 - 1.29 (m, 3H).

^1^3^C NMR (126 MHz, Chloroform-d): 156.2, 154.1, 147.5, 139.8, 128.8, 110.6, 110.2, 79.0, 77.4, 71.4, 71.0, 70.8, 70.8, 70.7, 70.7, 70.3, 70.3, 70.2, 69.8, 69.7, 69.6, 69.2, 63.4, 56.4, 45.2, 39.6, 38.6, 32.7, 3, 29.6, 29.5, 28.6, 26.8, 25.5.

HRMS (ESI, m/z): Calcd. for C_{38}H_{66}ClN_{3}O_{15} [M+Na]^+: 862.4080 Found: 862.4078

**Synthesis of 1-((4-(18-chloro-3,6,9,12-tetraoxaoctadecyl)oxy)-5-methoxy-2-nitrophe-nyl)-3,19-dioxo-2,8,11,14-tetraoxa-4,18-diazadocosan-22-oic acid (41)**

![Chemical structure of compound 41](image_url)

Compound 41 was synthesized by modification of a procedure described by Passemard et al.\(^2\).

To a solution of 40 (1.15 g, 1.368 mmol) in 5 mL anhydrous DCM at 0 °C was slowly added TFA (5 mL, 130 mmol). Thereafter the reaction mixture was warmed to room temperature (rt) and stirred for 2h. After completion of the reaction as evident by
TLC analysis, the solvent was removed under high vacuum to obtain the crude product as a TFA salt, which was used in the next step without further purification.

To a solution of the above deprotected product in DCM (5 mL) was slowly added DIEA to adjust the pH until >7 and added a catalytic amount of DMAP (33 mg, 0.2737 mmol). Subsequently, the above mixture was added succinic anhydride (166 mg, 1.66 mmol) followed by stirring overnight at rt. Next, the reaction mixture was diluted with DCM (5 mL), washed with 1 M aqueous solution of HCl (3x5 mL) followed by brine. The organic layer was dried over MgSO$_4$ and concentrated under vacuum to afford 41 (1.04 g, 90%, two steps) as brown oil, $R_f = 0.40$ (10% MeOH:DCM), which was used in the next step without further purification.

**Synthesis of (7-(diethylamino)-2-oxo-2H-chromen-4-yl)methyl (2-amino-5-(4-((1-(4-((18-chloro-3,6,9,12-tetraoxaoctadecyl)oxy)-5-methoxy-2-nitrophenyl)-3,19,22-trioxo-2,8,11,14-tetraoxa-4,18,23-triazahexacosan-26-yloxy)-3,5-dimethoxybenzyl)-pyrimidin-4-yl)carbamate, CTNH (9)**

![Chemical structure of CTNH (9)](image)

Compound 30b (30 mg, 0.04244 mmol) was deprotected by directly dissolving in 2 mL of TFA/DCM (1:2) and stirring at room temperature (rt) for 2 h. Next, excess TFA was removed and co-evaporated with DCM and DMF several times under high vacuum to yield
the TFA salt as a light brown solid. \( R_f = 0.02 \) in 10% MeOH/DCM. The primary amine salt can be stained well on TLC by Ninhydrin solution, the pink spot close to baseline appeared upon heated. The crude product was used in the next step without further purification.

The crude mixture above was dissolved in 1 mL of DMF followed by addition of DIEA (100 µL) and stirring at room temperature (rt) for ~15 minutes. In a separate flask, a 1 mL solution of 41 (54 mg, 0.06367 mmol) in DMF, DIEA (100 µL), and HATU (28 mg, 0.07428 mmol) were stirred at rt for ~10 minutes and then mixed with the above stirring mixture of deprotected DEACM-TMP-NH₂. The mixture was then stirred at rt for another 16 h before being concentrated under high vacuum. The crude brown oil was dissolved in DCM (50 mL) and washed with saturated NH₄Cl solution, brine, and dried with anhydrous Na₂SO₄. The crude mixture was then concentrated under reduced pressure and purified by silica gel column chromatography using 2% to 15% MeOH:DCM to yield pure product 9 (30%) as a brown oil. The product was further purified by preparative HPLC using H₂O/acetonitrile gradient.

\( R_f = 0.40 \) (10% MeOH:DCM).

UV-Vis \( (\lambda_{\text{max}} \text{ in 20% DMSO/PBS buffer pH 7.4}) \) : 250, 311, 397 nm.

\(^1\)H NMR \( (500 \text{ MHz, DMSO-}d_6) \) : 9.96 (s, 1H), 8.01 (s, 1H), 7.77 (s, 1H), 7.76 - 7.68 (m, 4H), 7.44 (m, 2H), 7.38 (s, 1H), 7.16 (s, 1H), 6.67 (m, 2H), 6.60 (s, 1H), 6.57 (d, \( J = 4.1 \text{ Hz, 2H} \)), 6.52 (t, \( J = 3.0 \text{ Hz, 2H} \)), 6.43 (d, \( J = 10.7 \text{ Hz, 2H} \)), 6.17 (s, 1H), 6.13 (s, 1H), 5.29 (d, \( J = 9.5 \text{ Hz, 5H} \)), 4.18 (t, \( J = 4.5 \text{ Hz, 2H} \)), 3.88 (s, 3H), 3.77 (m, 6H), 3.73 - 3.64 (m, 11H), 3.64 - 3.54 (m, 7H), 3.54 - 3.23 (m, 61H), 3.16 (m, 5H), 3.04 (m, 4H),
2.26 (s, 4H), 1.76 (d, J = 2.7 Hz, 3H), 1.72 - 1.52 (m, 9H), 1.52 - 1.39 (m, 2H), 1.39 - 1.15 (m, 7H), 1.10 (t, J = 7.0 Hz, 11H).

$^{13}$C NMR (126 MHz, DMSO-$d_6$) : 171.2, 171.2, 168.9, 162.3, 155.8, 155.7, 155.6, 155.0, 153.5, 152.9, 152.9, 152.8, 151.4, 150.4, 146.9, 139.3, 134.9, 134.5, 128.0, 125.4, 125.3, 111.8, 110.8, 109.5, 108.7, 108.7, 106.0, 105.7, 105.2, 104.8, 104.6, 96.8, 70.5, 70.4, 70.1, 69.9, 69.8, 69.8, 69.7, 69.5, 69.5, 69.4, 68.7, 68.5, 68.0, 67.8, 62.2, 61.7, 61.1, 56.2, 55.9, 55.8, 45.3, 44.0, 37.6, 35.9, 35.8, 33.2, 32.9, 32.0, 30.9, 29.8, 29.6, 29.3, 29.0, 26.1, 24.9, 22.6, 12.3.

HRMS (ESI, m/z): Calcd. for C$_{68}$H$_{99}$ClN$_9$O$_{22}$ [M+H]$^+$: 1428.6593 Found: 1428.6576
**General methods for biology**

**Dimerizer treatment and storage.**

Dimerizers were typically dissolved in DMSO at 10 mM and stored in amber plastic microcentrifuge tubes at −80 °C for long-term storage. For experiments, an aliquot was diluted in cell culture medium to final working concentration (10 µM for NTH and CTH, 100 nM for TNH, and 20 µM for CTNH, unless otherwise stated) and kept at −20 °C, then warmed to 37 °C when ready to use. Care was taken to minimize exposure of dimerizers to light and heat before experiments. Working quickly in low brightness of white light, LED red light installed, or low levels of red filtered light used for differential interference contrast (DIC) microscopy did not cause detectable premature cleavage of our dimerizers. However, CTNH is more prone to photouncaging.

**Plasmids.**

All plasmids in this study are derived from pEM705, which contains a CAG promoter for constitutive expression, obtained from E.V. Makeyev (Nanyang Technological University, Singapore)\(^\text{13}\). More information about plasmids used in this thesis can be found in our published data\(^\text{14-16}\). For plasmids in prepared manuscript (Chapter 4) are summarized in the table below.
<table>
<thead>
<tr>
<th>No.</th>
<th>Plasmid name</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Halo-GFP-ActA</td>
<td>C-terminal 47 amino acids of the <em>Listeria monocytogenes</em> ActA gene, conferring mitochondrial outer membrane targeting.</td>
</tr>
<tr>
<td>2</td>
<td>mCherry-eDHFR</td>
<td>mCherry fused to N-terminus of <em>E. Coli</em> dihydrofolatereductase (eDHFR)</td>
</tr>
<tr>
<td>3</td>
<td>PEX3-GFP-Halo</td>
<td>N-terminal 42 amino acids of the human Pex3 gene, which confer peroxisome targeting</td>
</tr>
<tr>
<td>4</td>
<td>KLC1-mCherry-eDHFR</td>
<td>Residues 1-175 of rat Kinesin-1 light chain, augmented with motor and motor effector domains</td>
</tr>
<tr>
<td>5</td>
<td>Halo-GFP-SPC25</td>
<td>SPC25 sequence fused to C-terminus of Halo-GFP</td>
</tr>
<tr>
<td>6</td>
<td>mCherry-eDHFR-Mad1</td>
<td>Mad1 sequence fused to C-terminus of mCherry-eDHFR</td>
</tr>
<tr>
<td>7</td>
<td>GFP-Halo-CAAX-IRES-mCherry-eDHFR</td>
<td>CAAX motif fused to GFP-Halo C-terminus</td>
</tr>
</tbody>
</table>

**Cell culture and cell lines.**

All experiments were performed with HeLa acceptor cells as described, originally obtained from E.V. Makeyev, Nanyang Technological University, Singapore. Cells were cultured in growth medium (Dulbecco’s modified Eagle’s medium with 10% FBS and 1% penicillin–streptomycin, all purchased from ThermoScientific, Life Technologies) at 37 °C in a humidified atmosphere with 5% CO₂. All cell experiments were performed with either stably or transiently transfected with a plasmid for bicistronic
expression of mCherry–eDHFR and indicated Haloenzyme-GFP constructs as the following protocols.

Stable cell lines were created by recombinase-mediated cassette exchange using the HILO recombinase-mediated cassette exchange (RMCE) system\textsuperscript{14}. Briefly, a monoclonal HeLa acceptor cell line with LoxP and Lox2272 recombination sites at a single chromosomal locus was co-transfected with a donor plasmid containing a transgenic cassette flanked by LoxP and Lox2272 sites and a Cre recombinase expressing plasmid, followed by antibiotic selection for a marker in the donor cassette. Cells were cultured as described above until reaching \textasciitilde60\% confluency. In a single well of a 6-well plate cells were transfected with 1\,\mu g of donor plasmid and 10\,ng of Cre plasmid, pEM784 (the Lampson lab) using either 3\,\mu l of Fugene 6 (Promega) or Lipofectamine 2000 (ThermoFisher). Two days after transfection, 1-2\,\mu g/ml puromycin was added to the growth medium for the selection of stable cell lines. The transfected cells were grown for \textasciitilde7-14 days to select for a single colony. Transient transfections were performed as described above using 1\,\mu g of plasmid DNA and a transfection reagent of choice, 24-48\,h prior to imaging experiments.

A cell line stably expressing Haloenzyme-GFP fused to CENP-B was used in Figure\,2.8, 3.3, and 4.5. A stable cell line stably integrated with CENP-B-GFP-Halo-IRES-mCherry-eDHFR was used in Figure\,2.9, 2.10, and 4.8. All mitochondrial recruitment assays were performed using a stable cell line constitutively expressing Halo-GFP-Mito-IRES-mCherry-eDHFR. The plasma membrane recruitment (Figure\,4.8B) was performed using transient transfection of GFP-Halo-CAXX-IRES-mCherry-eDHFR
plasmid. Peroxisome recruitment experiments (Figure 3.6, 3.7, and 4.9) were performed by transiently transfecting either BICD-mCherry-eDHFR or KLC1-mCherry-eDHFR plasmid into cells stably expressing PEX3-GFP-Halo.

The kinetochore experiments (Figure 4.10) were performed with Haloenzyme-GFP-tagged kinetochore proteins (SPC25) stably expressed in cell lines. The top 5% GFP-positive cells were sorted with flow cytometry. The mCherry-eDHFR-Mad1 plasmid was transiently transfected into the stable kinetochore GFP-positive cells using Lipofectamine 2000, 24 h before imaging following the manufacturer’s protocol. Hoechst staining was used to verify absence of bacterial contamination.

**Dimerizer cytotoxicity assay.**

AlamarBlue (Molecular Probes, catalog no. DAL1100) cell viability assays were performed following the protocol provided by the manufacturer (Thermo Fisher Scientific). The assay is based on conversion of a water-soluble dye, resazurin, into a fluorescent and colorimetric indicator by metabolically active cells. Damaged and nonviable cells generate lower signal due to reduced metabolic activity. HeLa cells were cultured in a 96-well assay plate (Corning Incorporated Costar, 3603), ~1 × 10⁴ cells per well. Cells were treated with various concentrations of a dimerizer. In the same plate for each experimental replicate, DMSO (vehicle) was used as a positive control and either Blasticidin S HCl (10 µg/mL; Thermo Fisher Scientific, catalog no. A1113903) or Camptothecin (Sigma Aldrich) was used as a negative control. The final DMSO concentration was kept at 0.5% for CTH and TNH, and 1% for CTNH in 100 µL media.
across all wells. After incubating with the dimerizers for 24 h or 48 h, cells were washed once with 200 µL of fresh DMEM, then incubated with the AlamarBlue reagent (10× dilution in DMEM medium) for 2 h to allow conversion of resazurin to resorufin. The fluorescence signal was measured at 37 °C with 550 nm excitation and 590 nm emission wavelengths, using a Tecan plate reader (model, Infinite M1000 PRO) operated by Tecan i-control software. The fluorescence signal was background subtracted based on wells without cells and calculated as a percentage of the signal from DMSO control cells. Each dimerizer concentration was tested in quadruplicate, and data were averaged over three independent experiments.

**Image acquisition.**

For live imaging, cells were seeded on 22 × 22 mm glass coverslips (no. 1.5; Fisher Scientific) coated with poly-D-lysine (Sigma-Aldrich) in single wells of a six-well plate. When ready for imaging, coverslips were mounted in magnetic chambers (Chamlide CM-S22-1, LCI) with cells maintained in L-15 medium without phenol red (Invitrogen) supplemented with 10% FBS and 1% penicillin–streptomycin at 37 °C on a heated stage in an environmental chamber (Incubator BL; PeCon GmbH). Images were acquired with a spinning disk confocal microscope (DM4000; Leica) with a ×100 1.4 NA objective, an XY Piezo-Z stage (Applied Scientific Instrumentation), a spinning disk (Yokogawa), an electron multiplier charge-coupled device camera (ImageEM; Hamamatsu Photonics), and a laser merge module equipped with 488- and 593-nm lasers.
(LMM5; Spectral Applied Research) controlled by MetaMorph software (Molecular Devices).

Z-stacks for both GFP and mCherry channel were taken with 0.5 µm spacing for a total of 3 µm in plasma membrane, centromeres, mitochondria and kinetochore experiments, and a total of 5 µm thickness was taken for peroxisome motor experiments. The time interval and duration of acquisition were specified in the figures, except for Mad1 experiments, time series were taken with time interval and duration manually decided based on cell cycle progression. All mages were analyzed using ImagJ or Fiji.\(^{17}\)

**Dimerization and photo-activation.**

For NTH, CTH, and CTNH experiments, cells were incubated with dimerizers at the concentration and duration as indicated in the figures, which was normally followed by a 30-min washout to remove unbound dimerizer. For TNH experiments, 100 nM TNH was added directly to cells on the microscope stage.

For whole-cell photouncaging or photocleavage, UV light is from a mercury arc lamp (Osram HXP R 120W/45C Vis) was filtered through a 387/11 nm band pass filter (Semrock, part # FF01-387/11 as a component in a DAPI filter cube) and focused through the objective.

For targeted uncaging or cleavage, photo-control was performed with an iLas2 illuminator system (Roper Scientific), equipped with a 405-nm laser (CrystaLaser LC, model # DL405-050-O; output of 27 mW after fiber coupling) controlled using the iLas2 software module within MetaMorph. For targeted irradiation with the 405-nm laser, a
region of interest was defined using MetaMorph. ~7% laser power and 20 repetitions were used for CTH and CTNH uncaging, and 8% laser power and 40 repetitions were used for TNH and CTNH cleavage, unless otherwise stated.

**Protein-ligand docking and alignment**

Using the Autodock tool integrated in Docking Server (http://www.dockingserver.com) for protein-ligand docking, we prepared individual ligands, Trimethoprim (TMP) and Haloligand (HTag) with a small part of the linker that we use to make the CTNH dimerizer. Energy minimization and total charges of the ligand structures were performed using the PM6 calculation. The CTNH dimerizer structure was energy minimized using an MM2 calculation in Chem3D 15.0 to obtain the fully extended structure of the linker. Docking was performed according to the default parameter (Autogrid) set up and automatic determination of the known binding site of each protein-ligand complex based on individual PDB crystal structures (1RG7 for eDHFR and 1BN7 for HaloTag protein).

Next, the Coot software was used to manually align the coordinates of the CTNH dimerizer to the individual TMP and Haloligand to obtain the CTNH-bound ternary protein complex. The distance of the linker between TMP and Haloligand was measured using the Chimera software. The distance was determined to be ~35 Å. The superimposed ternary complex structure does not exhibit steric hindrance between the eDHFR and HaloTag proteins. The photocleavable portion of the linker was observed exposed to water near the HaloTag protein.
REFERENCES


APPENDIX
NMR SPECTRA FOR CHAPTER 2
$^1$H NMR spectrum of 20 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 20 in CDCl$_3$ (126 MHz).
$^1$H NMR spectrum of 21 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 21 in CDCl$_3$ (126 MHz).
$^1$H NMR spectrum of 25 in MeOD (500 MHz).
$^{13}$C NMR spectrum of 25 in MeOD (126 MHz).
$^1$H NMR spectrum of 26 in DMSO- $d_6$ (500 MHz).
$^{13}$C NMR spectrum of 26 in DMSO-$d_6$ (126 MHz).
$^1$H-$^{15}$N HMBC NMR spectrum of 26 in DMSO-$d_6$ (600 MHz);
whole spectrum (top) and zoom-in (bottom)
$^{1}H$ NMR spectrum of 1 in CDCl$_3$, (600 MHz).
$^1$H NMR spectrum of 1 in CDCl$_3$ (126 MHz).
$^{1}$H NMR spectrum of 2 in DMSO-$d_6$ (500 MHz).
$^{13}$C NMR spectrum of 2 in DMSO-$d_6$ (126 MHz).

\[ \text{Chemical Structure} \]
$^1$H-$^{15}$N HMBC NMR spectrum of 2 in DMSO-d$_6$ (600 MHz)
NMR SPECTRA FOR CHAPTER 3
$^1$H NMR spectrum of 29 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 29 in CDCl$_3$ (126 MHz).
$^{1}$H NMR spectrum of 30a in DMSO-$d_6$ (500 MHz).
$^{13}$C NMR spectrum of 30a in DMSO-$d_6$ (126 MHz).
$^1$H-$^{15}$N HMBC NMR spectrum of 30a in DMSO-$d_6$ (126 MHz).
The H NMR spectrum of 5 in DMSO-d$_6$ (500 MHz).
$^{13}$C NMR spectrum of 5 in DMSO-$d_6$ (126 MHz).
$^1$H- $^{15}$N HMBC NMR spectrum of 5 in DMSO-$d_6$ (126 MHz).
$^{1}H$ NMR spectrum of 32 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 32 in CDCl$_3$ (126 MHz).
\(^1\)H NMR spectrum of 33 in CDCl\(_3\) (500 MHz).
$^{13}$C NMR spectrum of 33 in CDCl$_3$ (126 MHz).
$^1$H NMR spectrum of 35 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 35 in CDCl$_3$ (126 MHz).
$^{1}$H NMR spectrum of 35a in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 35a in CDCl$_3$ (126 MHz).
$^1$H NMR spectrum of 37 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 37 in CDCl$_3$ (126 MHz).
$^1$H NMR spectrum of 38 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 38 in CDCl$_3$ (126 MHz).
$^1$H NMR spectrum of 6 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 6 in CDCl$_3$ (126 MHz).
NMR SPECTRA FOR CHAPTER 4
\[ ^1H \text{NMR spectrum of 39 in CDCl}_3 (500 \text{ MHz}). \]
$^{13}$C NMR spectrum of 39 in CDCl$_3$ (126 MHz).
$^1\text{H NMR}$ spectrum of 40 in CDCl$_3$ (500 MHz).
$^{13}$C NMR spectrum of 40 in CDCl$_3$ (126 MHz).
$^1$H NMR spectrum of 25 in DMSO-$d_6$ (500 MHz).
$^{13}$C NMR spectrum of 25 in DMSO-$d_6$ (126 MHz).
DEPT-135 NMR spectrum of 25 in DMSO-$d_6$
COSY NMR spectrum of 25 in DMSO-$d_6$
NOESY NMR spectrum of 25 in DMSO-$d_6$
$^{1}\text{H}^{13}\text{C}$ HSQC NMR spectrum of 25 in DMSO-$d_6$
$^1$H-$^{13}$C HMBC NMR spectrum of 25 in DMSO-$d_6$
$^1$H-$^15$N HMBC NMR spectrum of 25 in DMSO-$d_6$
$^1$H NMR spectrum of 30b in DMSO-$d_6$ (500 MHz).
$^{13}$C spectrum of 30b in DMSO-d$_6$ (126 MHz).
DEP-135 NMR spectrum of 30b in DMSO-$d_6$
COSY NMR spectrum of 30b in DMSO-$d_6$
NOESY NMR spectrum of 30b in DMSO-$d_6$ (whole spectrum)
NOESY NMR spectrum of 30b in DMSO-$d_6$ (zoom-in)
$^{1}H-^{13}C$ HSQC NMR spectrum of 30b in DMSO-$d_6$
$^1$H-$^{13}$C HMBC NMR spectrum of 30b in DMSO-$d_6$
1H-15N HMBC NMR spectrum of 30b in DMSO-d6
$^1$H NMR spectrum of 9 in DMSO-$d_6$ (500 MHz)
$^3$C NMR spectrum of 9 in DMSO-$d_6$ (126 MHz)