2008

Semi-Synthesis of the Transcription Factor SMAD2 Containing Caging Groups and Phosphoaminoacid Analogues

Michael Eric Hahn

SEMI-SYNTHESIS OF THE TRANSCRIPTION FACTOR SMAD2 CONTAINING CAGING GROUPS AND PHOSPHOAMINOACID ANALOGUES

A Thesis Presented to the Faculty of

The Rockefeller University

In Partial Fulfillment of the Requirements for

The degree of Doctor of Philosophy

by

Michael Eric Hahn

June 2008

SEMI-SYNTHESIS OF THE TRANSCRIPTION FACTOR SMAD2 CONTAINING CAGING GROUPS AND PHOSPHOAMINOACID ANALOGUES

Michael Eric Hahn

The Rockefeller University 2008

Post-translational modification (PTM) of a protein refers to any chemical change that occurs to the protein after its ribosomal synthesis. The seemingly endless number of PTMs can endow proteins with new functionalities that are not present in the unmodified proteins. In order to study the functions of PTMs on a given protein, it is often necessary to have access to pure preparations of the modified protein and its analogues. Traditional biological methods frequently do not allow for the isolation of significant amounts of pure modified proteins, therefore chemical methods are often employed in this regard. Protein semi-synthesis is a chemical method that entails the melding of at least two protein fragments in which at least one fragment is isolated from a biotic source while another fragment is synthesized by chemical methods. This framework enables the installation of PTMs into the complete protein through chemical control over the synthesized fragment without the need to synthesize the entire protein molecule, which is often a practically impossible task.

This thesis describes efforts employing the protein semi-synthesis technique known as expressed protein ligation (EPL) to the study of the cellular signaling protein Smad2. Smad2 is activated by phosphorylation, which is the best characterized reversible PTM. Once phosphorylated, Smad2 accumulates in the nucleus of cells where it helps to direct transcriptional changes that affect cell behavior. Techniques were developed to

cage Smad2 by directly blocking the activating phosphates on Smad2 with bulky photoremovable groups. This affords the investigator control over the timing and localization of Smad2 activity by judicious application of light of the appropriate wavelength to remove the photocaging group. This approach can be employed to generate caged analogues of any phosphorylated protein. A parallel caging approach was developed that relies upon indirect blockade of Smad2 phospho-dependent activity through the installation of a caging group on the C-terminus of the protein. This approach was compatible with a fluorescence reporter that is fluorescent only when the photocaging group is removed, thus allowing for selective monitoring of the activated form of the protein. This protein was introduced into live cells and upon activation allowed for real time visualization of the active protein through fluorescence microscopy.

The activity of Smad2 is dependent upon differential protein-protein interactions that the phosphorylated protein is able to participate in while the non-phosphorylated protein is not. In an effort to identify new binding partners that are sensitive to the phosphorylated state of Smad2, methods were developed to install stable phospho-analogues (phosphonates) into Smad2. These analogues were used to identify a candidate Smad2-binding protein, PRMT5, that preferentially binds non-phosphorylated Smad2. Studies are ongoing to determine if this interaction has physiological relevance.

Acknowledgments

I would like to thank Professor Tom W. Muir for his infectious enthusiasm and optimism for science. His creativity and rigorous approach to scientific problems is truly inspiring.

I greatly appreciate the efforts and advice offered by my FAC committee members Drs. Seth Darst, Sandy Simon, and Joan Massagué. I would like to recognize Dr. David Lawrence of the University of North Carolina for serving as my external examiner.

I am deeply indebted to all past and present members of the Muir Laboratory who have shared their knowledge and expertise freely. I am especially pleased to acknowledge Drs. Jean-Philippe Pellois and Miquel Vila-Perelló for their direct collaboration in much of the work discussed in this thesis. I would also like to recognize several members of the laboratory by name who have made insightful suggestions to this work: Mr. Edmund Schwartz, Dr. Matthew Pratt, Dr. Kyle Chiang, Dr. Jennifer Ottesen, Dr. Mande Holford, Dr. Gholson Lyon, Dr. Roseanne Hofmann, Mr. Robert Flavell, Mr. Robert McGinty, and Dr. Vasant Muralidharan.

Several other Tri-Institutional scientists have helped me greatly and I would like to list their names here to acknowledge them. They are Dr. Joshua Weinger, Dr. Aaron Steiner, Ms. Ariel Levine, Mr. Matthew Sekedat, Dr. Gopal Sapkota, Mr. Claudio Alarcón, Professor Tarun Kapoor, and Professor Hiro Funabiki.

I wish to thank the NIH, the Keck Foundation, and the William Randolph Hearst Foundation for funding. I would also like to thank Alison North and the members of the Rockefeller Bioimaging Facility as well as Ewa Folta-Stogniew of the Keck Facility at Yale for technical assistance.

I am grateful to the staff of the Tri-Institutional MD-PhD Office and the Rockefeller Dean's office and would like to especially thank Dr. Olaf Andersen and Mrs. Ruth Gotian.

I could not have done any of this without the constant support and encouragement of my parents, Mrs. Ellen Hahn and Mr. Steven Hahn, and my brother, Mr. David Hahn. Above all, I would like to express my sincere thanks to my wife and partner, Dr. Allison Alcivar Hahn, without whom this thesis would not be possible. She has stood by me through good times and hard times. Words can not express my appreciation and love for her.

TABLE OF CONTENTS

Section **Page**

Section **Page**

LIST OF FIGURES

Section **Page**

CHAPTER 1

CHAPTER 2

Section **Page**

CHAPTER 3

Section **Page**

CHAPTER 4

CHAPTER 5

Section **Page**

CHAPTER 6

Chapter 1: Introduction

All living organisms must respond to changes in their environment to maintain homeostasis. The dynamic processes that enable these changes operate on many levels, including the molecular level, which is ultimately the locus of control. In the process known as signal transduction, information in the form of a molecular signal arrives at the outside of a cell and is sensed by a receptor at the cell membrane. In turn, the receptor acknowledges receipt of this information by communicating with and influencing other molecules inside the cell including proteins, lipids, ions, and small molecules. No matter the size or physical properties of each type of signaling molecule, they may all be regarded as information packets that relay information for the ultimate goal of communicating the original signal so that the cell can respond properly.

Each step in signal transduction requires interaction between signaling molecules. These interactions relay the signal from one molecule to the next, resulting in what is often referred to as a signaling cascade. As these interactions ultimately dictate cellular behavior, they are subject to regulation by a variety of mechanisms. One major mechanism for the regulation of protein-protein interactions is post-translational modification (Hunter, 2000; Walsh, 2005). These modifications serve as signaling switches that effectively turn protein activity on and off and can operate intermolecularly or intramolecularly (Seet et al., 2006). An intermolecular mechanism can allow or forbid one protein from interacting with another, whereas an intramolecular mechanism can allosterically modulate the activity of a protein. One illustrative example of regulation by the post-translational modification of an intramolecular protein-protein interaction is provided by the protein kinase Src (Martin, 2001; Sicheri et al., 1997; Xu et al., 1999; Xu

et al., 1997). When Tyr527 of Src is phosphorylated, an SH2 domain of the same polypeptide binds the pTyr in the context of its adjacent peptide sequence. This leads to inactivation of Src, thus shutting down the processes downstream of it. However, the phosphate on Tyr527 can be removed by phosphatases, leading to the activation of the kinase function of Src, which goes on to transduce this signal by phosphorylating downstream effector proteins, initiating a signaling cascade. Mutation of Tyr527, causes the kinase domain to become constitutively active, resulting in high level phosphorylation of downstream molecules. This in turn can lead to malignant transformation as in the famous case of v-Src. This example and others like it underlie the importance of studying signal transduction.

1.1. Brief historical framework of protein phosphorylation

The best understood type of reversible protein post-translational modification is phosphorylation (Cohen, 2002; Pawson and Scott, 2005). As alluded to in the example of Src above, protein phosphorylation refers to the installation of a phosphoryl group onto an amino acid side-chain of a protein (Figure 1.1). While working at the Rockefeller Institute in 1932, Phoebus Levene and Fritz Lipmann reported the discovery that the protein vitellin contains serine-*O*-phosphoric acid (Lipmann and Levene, 1932). Prompted by the observation that the pool of phosphoproteins undergoes a high rate of turnover in highly metabolic tissues such as tumors, Burnett and Kennedy undertook an investigation to determine if the installation of phosphoryl groups into proteins is a direct enzyme-mediated process (Burnett and Kennedy, 1954). Using the known phosphoprotein casein as substrate, they discovered an enzyme that they termed “protein

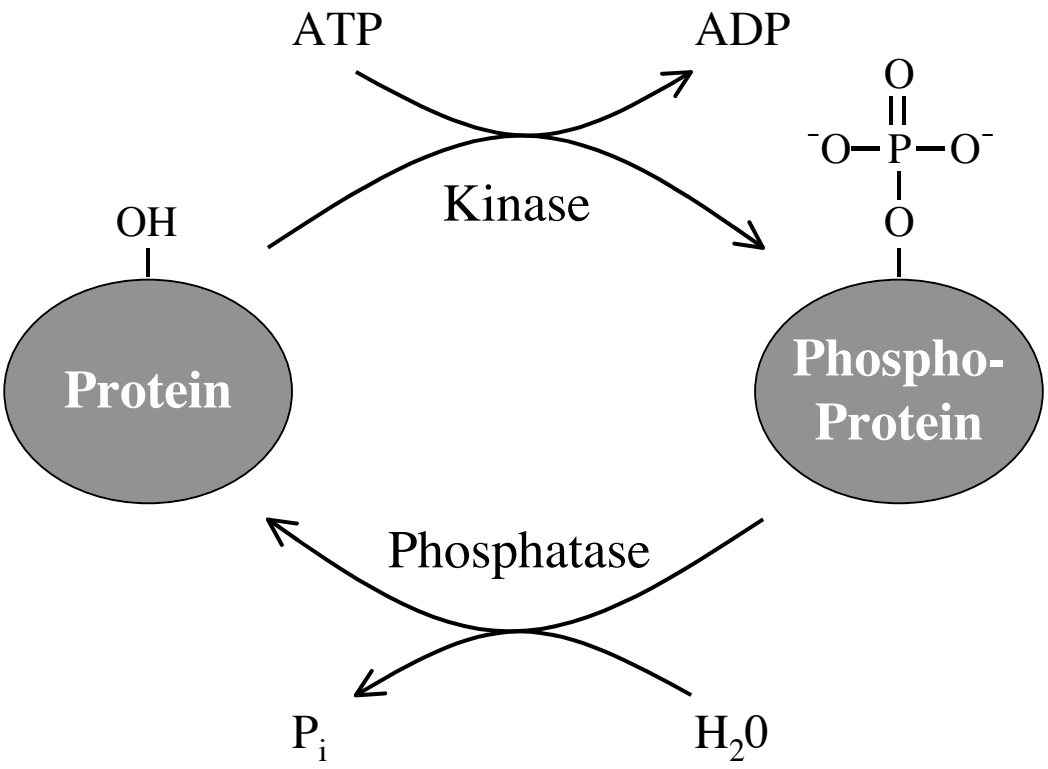


Figure 1.1. Reversible protein phosphorylation. A protein that is a substrate of a kinase is phosphorylated in a reaction involving transfer of the terminal phosphoryl group of the co-substrate ATP to the protein to create a phosphoprotein. Phosphatases catalyze the reverse reaction, hydrolysis of the phosphate from the phosphoprotein.

phosphokinase" that catalyzes the formation of phosphoserine in the context of the intact casein molecule. In addition to reporting the first description of a kinase, these insightful authors hinted at the root of substrate specificity of kinases: "One may surmise that the (phosphorylated) serine residues of casein occur in an amino acid sequence not present in other proteins tested…"

It would take an additional 20 years for Krebs, Kemp, Daile, Carnegie, Engstrom, and others to show that the primary determinant of kinase specificity is the sequence immediately surrounding the phosphorylation site (Daile et al., 1975; Humble et al., 1975; Kemp et al., 1976; Kemp et al., 1975; Zetterqvist et al., 1976). One of the major reasons for this 20-year delay is that the proper tools were not yet available to dissect this problem. Around the same time, the peptide chemist Bruce Merrifield of the Rockefeller Institute and his contemporaries were developing methods for the solid-phase synthesis of peptides (Merrifield, 1996; Merrifield, 1963). These would prove to be vital reagents in the quest to understand the substrate specificity of kinases, since their synthesis allowed investigators to systematically vary the amino acid sequence surrounding the phosphorylated residue of model peptide substrates. This highlights an important theme that still exists today in modern biomedical science: the development of many biological concepts must await development of the proper tools to study them. These tools are often developed by chemists who apply their knowledge and abilities to the synthesis of defined biomolecules and analogues.

By the 1940s, Cori, Cori, and Green had determined that the enzyme glycogen phosphorylase exists in two interconvertible forms, termed *a* and *b* (Cori and Cori, 1945; Cori and Green, 1943). This set the stage for the groups of Sutherland, Krebs, and Fischer

to investigate the mechanism of conversion between forms *a* and *b* (Cowgill and Cori, 1955; Fischer and Krebs, 1955; Krebs and Fischer, 1956; Rall et al., 1956a; Rall et al., 1956b; Sutherland and Wosilait, 1955; Wosilait and Sutherland, 1956). These investigators determined that glycogen phosphorylase *a* is phosphorylated while glycogen phosphorylase *b* is generated by dephosphorylation of *a*. They purified the kinase and phosphatase that carries out the cyclical phosphorylation/dephosphorylation of glycogen phosphorylase. This was the first time that scientists had in hand a complete interconvertible enzyme system, containing the regulated enzyme as well as the two enzymes that were responsible for the conversion.

From these examples grew concepts upon which generations of scientists would explore other phosphoproteins, kinases, phosphatases, and their biological role. Protein phosphorylation has been found to regulate all aspects of cellular physiology. This introduction will focus on chemical approaches to the study of protein phosphorylation.

1.2. Chemistry and protein phosphorylation

The progress of understanding signal transduction has been intimately linked with developments in chemistry. In fact the contributions of chemists have been central to the field and perhaps it is artificial to draw a line between chemists and biologists working in this arena. Nevertheless, it is useful to examine the contributions of biological chemists to the study of protein phosphorylation. In the past five to ten years, these investigators have developed increasingly sophisticated tools enabling the discovery of new insights into the mechanism of protein phosphorylation and its consequences. These include tools that have clarified the enzymatic mechanism of phosphorylation by kinases, tools to control

the function of kinases, phosphatases, and phosphoproteins, as well as tools to monitor the function of these proteins. Herein, these advances will be discussed both in terms of the chemistry behind them and the biological insight that their use has attained or is expected to attain in the future.

1.2.1. Chemical dissection of protein kinase mechanism

The groups of Kaiser and Knowles working at the Rockefeller University and Harvard University, respectively, performed an elegant experiment to determine that kinase-mediated phosphorylation occurred by direct transfer of a phosphoryl group from ATP to substrate and did not involve a covalent enzyme-phosphate intermediate (Ho et al., 1988). They used chiral ATP containing ^{16}O, ^{17}O, and ^{18}O in the γ-phosphate and a synthetic peptide as substrates for the kinase PKA. Once acted upon by the kinase, the investigators analyzed the phosphorylated product by dephosphorylation with alkaline phosphatase in the presence of the chiral acceptor molecule (*S*)-butane-1,3-diol. Following derivatization and analysis by NMR, the configuration of the phosphate could be determined unambiguously and compared to the known configuration of the starting material, chiral ATP. The kinase-mediated reaction proceeded with inversion of configuration, therefore direct transfer of the phosphoryl group from ATP to the peptide substrate was inferred.

There are two limiting mechanisms for direct transfer of a phosphoryl group from ATP to an acceptor: that involving an associative transition state (S_N2-like) and the other involving a dissociative transition state (S_N1-like; Figure 1.2) (Shen et al., 2005). There had been much controversy between different groups working in this field as to whether

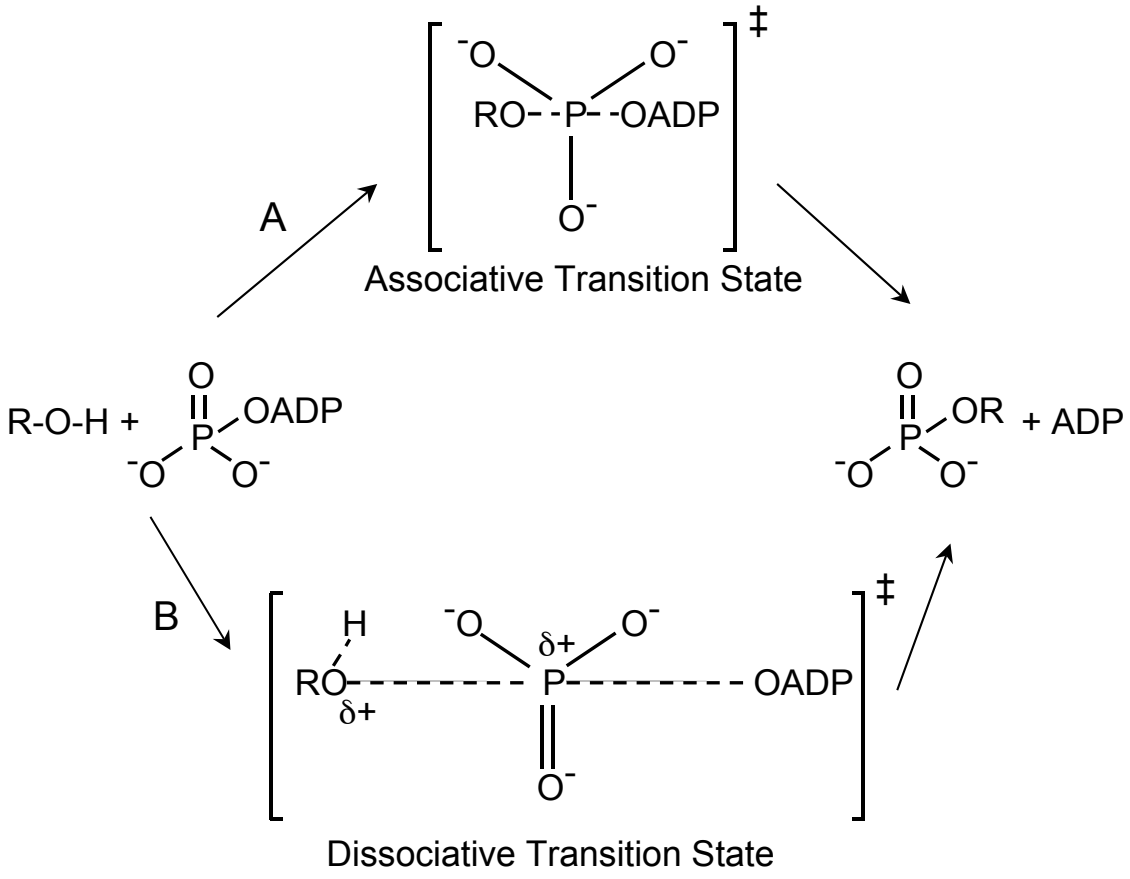


Figure 1.2. Transition states for phosphotransfer reactions. The phosphoryl group of the γ-phosphate of ATP can be transferred to an acceptor by two alternative limiting mechanisms. One involves an associative transition state in which the bond between the incoming acceptor nucleophile and phosphorous is well-formed. The other involves a dissociative transition state in which the bond between the incoming nucleophile and the phosphorous is not well-formed and the bond between the phosphorous and the leaving group is almost completely absent.

the transition state of kinase-mediated phosphorylation reactions is associative or dissociative (Kim and Cole, 1997). Cole and his colleagues addressed this problem by applying the classical physical organic chemistry technique known as Brønsted analysis to the phosphorylation of a peptide substrate catalyzed by the tyrosine kinase Csk (Kim and Cole, 1997, 1998). The pK_a of the incoming nucleophile (the tyrosine phenol of the peptide substrate) of a series of Csk substrates was altered by various fluorine substitutions of the ring and the k_{cat} values of Csk-catalyzed phosphorylation of the resulting substrates were determined. By plotting the log of the k_{cat} values as a function of substrate pK_a, the quantity β_{nuc} can be determined. This quantity is a measure of the dependence of the transition state on the nature of the incoming nucleophile. A β_{nuc} close to unity indicates a large dependence on the nucleophile, thus suggesting the existence of an associative transition state. However, Cole and colleagues found that the β_{nuc} for Csk-catalyzed phosphorylation is close to zero (Kim and Cole, 1997, 1998), indicating that the formation of the transition state has very little dependence on the nature of the nucleophile, consistent with a dissociative transition state.

1.2.2. Inhibitors and activators of protein kinases and phosphatases

Since protein phosphorylation is such a central component of cellular regulation, the aberrant activity of protein kinases and phosphatases can lead to a variety of diseases. For example, the constitutively active kinase activity of the BCR-ABL oncoprotein drives chronic myeloid leukemia (Deininger et al., 2005). The ABL tyrosine kinase inhibitor imatinib has proven exceedingly useful for the treatment of CML (Deininger et al., 2005). For this reason, there has been a great deal of effort in both academic and industrial

settings to develop drugs that modulate the activity of other protein kinases. It is beyond the scope of this introduction to exhaustively discuss all of the recent developments in this area. Therefore, a small, representative selection of particularly exciting and promising developments is presented (for a more comprehensive discussion, see Knight and Shokat, 2005).

It is quite difficult to design inhibitors specific for one kinase since there are over 500 kinases in human that each share the same catalytic mechanism. The consequences of using non-selective kinase inhibitors may be severe both in scientific investigation and as therapeutics. In the case of using kinase inhibitors as research tools, if the inhibitor is not specific it is impossible to discern which kinase is responsible for an observed phenotype. If a non-specific kinase inhibitor is to be used as a drug, it is likely that it would cause toxicity.[1] This has inspired researchers to develop methods that would produce more specific kinase inhibitors.

One promising method is to design bisubstrate inhibitors of kinases. A bisubstrate kinase inhibitor consists of a molecule bearing two portions, one that binds the ATP-binding site and another that binds the protein substrate site. Theoretically, this type of

[1] Recent data has shown that some moderately selective kinase inhibitors (termed multikinase inhibitors) are useful drugs, calling into question the paradigm that the greater the selectivity the better the drug (for example, see Wilhelm et al., 2006). The efficacy of multikinase inhibitors is correlated with which kinases they inhibit. For example, sorafenib, a kinase inhibitor designed to inhibit Raf, also inhibits VEGFR 1/2/3, PDGFRβ, FGFR1, c-Kit, Flt-3, and RET, all of which are implicated in angiogenesis or tumorigenesis. Therefore, this drug is useful in treating certain cancers.

inhibitor can achieve greater selectivity and potency as compared to a monosubstrate inhibitor since it capitalizes on the free energy of both substrate-binding sites. A key design feature of a bisubstrate inhibitor is to join the two fragments with a linker of the appropriate length to mimic the transition state of the reaction. The work of Cole and colleagues discussed above showed that kinase-mediated phosphorylation takes place with a dissociative transition state (Kim and Cole, 1997, 1998; Shen et al., 2005). The knowledge of the dissociative nature of the transition state has an important implication for effective inhibitor design; it suggests that the distance between the incoming oxygen of the nucleophile and the phosphorous is approximately 5 Å (Mildvan, 1997). Therefore, an effective bisubstrate inhibitor would have to position the phosphorous of the nucleotide portion of the inhibitor and the peptide substrate about 5 Å away from each other. Exploiting this subtle yet crucially important point, Cole, Hubbard, and colleagues designed a bisubstrate inhibitor of the tyrosine kinase IRK with a suitably long spacer between the two components of the inhibitor (Parang et al., 2001). This molecule (Figure 1.3A, compound **1**) inhibited IRK with a K_i of 370 nM and showed 100-fold selectivity for IRK over Csk. Furthermore, they were able to co-crystallize the bisubstrate inhibitor in complex with IRK. This demonstrated that the design principles did indeed aid in the function of the inhibitor as the linker distance between the two subunits of the inhibitor was 5 Å. Merging the concepts learned in these studies with traditional medicinal chemistry approaches may yield highly potent, selective, bioavailable kinase inhibitors.

Shokat and colleagues have taken an integrated approach using chemistry, protein engineering, and genetics to achieve specific inhibitors of kinases for research purposes (Shogren-Knaak et al., 2001). Their strategy makes use of the “bump-hole” approach

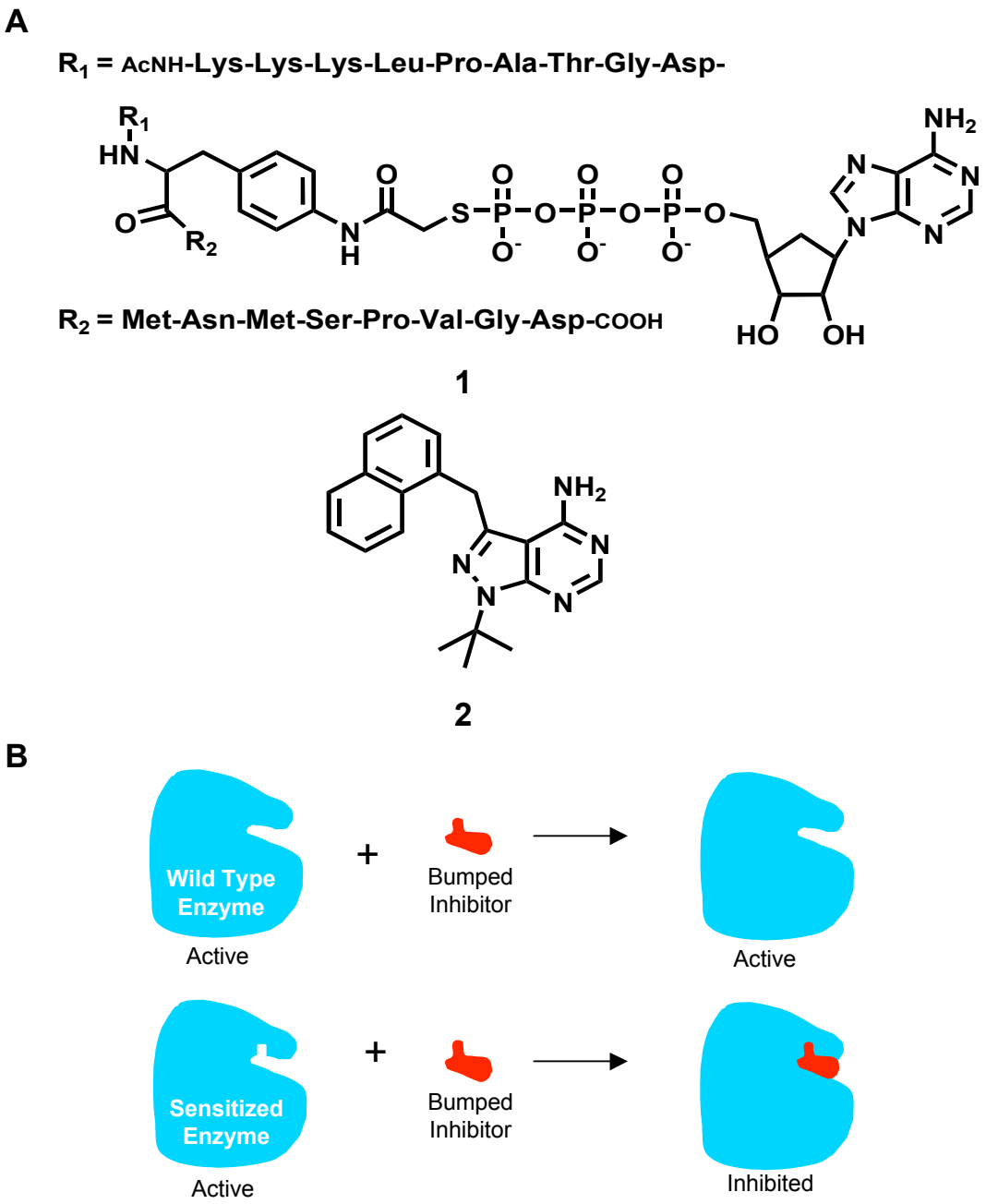


Figure 1.3. Kinase inhibitors and the Bump-Hole Approach. A. 1 = bisubstrate inhibitor of IRK. **2** = NM-PP1. **B.** In the bump-hole approach, a sensitized enzyme is made by mutagenesis to form a cavity at or near the active site. An inhibitor is derivatized with a bulky substituent that complements the cavity in the sensitized enzyme, but does not bind to the wild type enzyme, resulting in selective inhibition of the sensitized enzyme.

(Figure 1.3B). This involves mutation of the enzyme of interest to create a hole in the active site that will accommodate a bumped inhibitor that does not inhibit wild type enzymes of the same class due to steric clash between the bump and the wild type active site (Belshaw et al., 1995; Hwang and Miller, 1987; Kapoor and Mitchison, 1999; Powers and Walter, 1995). Shokat and colleagues discovered a conserved "gatekeeper" residue in the active site of many kinases, which controls access of compounds to the ATP-binding site (Bishop et al., 1998). This residue is frequently leucine, methionine, phenylalaninine, or threonine, all of which contain a bulky side chain (Zhang et al., 2005). By mutating this residue to one with a smaller side chain, such as alanine or glycine, extra space is built into the structure of the kinase, sensitizing it to inhibition by a bumped inhibitor, where the bump is complementary in size to the hole in the kinase (Bishop et al., 1998; Bishop et al., 2000). For example, NM-PP1 (Figure 1.3A, compound **2**) is a derivative of the broad spectrum kinase inhibitor PP1 in which a napthyl group has been added to the exocyclic carbon. This bulky derivative cannot fit inside the active site of wild type kinases due to steric clash of the napthyl group with the gatekeeper residue. Mutant kinases carrying a gatekeeper mutation rendering them sensitive to inhibition by bumped inhibitors are termed analog-sensitive (AS) kinases.

The "hole" in many AS kinases (~70% of those tested) does not significantly impact the catalytic activity and specificity of the enzyme (Zhang et al., 2005). However, up to 30% of kinases do not retain activity when their gatekeeper residue has been trimmed from a bulky amino acid to alanine or glycine (Zhang et al., 2005). To extend the bump-hole approach to these kinases, Shokat and colleagues performed a genetic screen by exploiting the bacterial aminoglycoside kinase APH(3')-IIIa, which is

structurally homologous to protein kinases. They discovered second site suppressor mutations that yield active kinase when both the gatekeeper and second site are mutated (Zhang et al., 2005). Among the suppressors found, one (N87T) was chosen for further study due to its close proximity to the gatekeeper residue in the β-sheet of the N-terminal lobe of the kinase. Introduction of β-branched amino acids, such as the threonine of site N87T, into β-sheets can stabilize the sheet (Minor and Kim, 1994; Otzen and Fersht, 1995). The authors reasoned that this stabilization offset the destabilization to the same β-sheet that occurs upon mutation of the gatekeeper residue and restored activity to the double-mutant kinase. This approach has led to the acquisition of active versions of the protein kinases Cdc5, MEKK1, GRK2, Pto, and Plk1, each containing a gatekeeper residue mutation. Thus far, MEKK1 and Plk1 have been shown to be analog sensitive. This approach and similar rationally-designed approaches (Kenski et al., 2005) promise to greatly increase the number of kinases accessible to inhibition by the bump-hole strategy.

Through extensive collaborations, this technology has been widely used and has helped to uncover a great deal of kinase biology. A particularly illustrative example is the use of AS-kinase inhibition in the study of the temporal requirements for cJun N-terminal kinase (Jnk) signaling in mediating survival and apoptosis (Ventura et al., 2006). In response to tumor necrosis factor (TNF), Jnk kinase displays a biphasic activation pattern with respect to time. The early phase is characterized by robust kinase activity peaking at approximately ten minutes. This is followed by a period of Jnk activity that is lower in amplitude, but lasts longer; on the order of 12 hours. Paradoxically, under certain physiological conditions, Jnk kinase activity signals apoptosis while under other

physiological conditions, it acts as a survival factor. Ventura et al. wondered if the two different phases of Jnk activity were responsible for this paradox: the same protein is involved in two opposite processes, survival and apoptosis. These investigators used mutant AS-Jnk that is active and sensitive to the inhibitor 1NM-PP1 to resolve this issue. They found that Jnk signaling for longer than 1hour but less than 6 hours is required for a full proaptotic response after UV-induced stress. In contrast, they found that only the first 30 minutes of Jnk activity is required for Jnk-mediated survival signaling in cells treated with TNF. This study highlights the power of small molecule approaches to study biological problems. The timecourse of a biological response must be greater than the time it takes the investigator to perturb the process in order to probe it. While cells displaying diminished Jnk activity are readily available by genetic knockout or siRNA-mediated knockdown, these perturbations take longer than the response of Jnk to TNF in survival signaling. Therefore, the role of this rapidly induced Jnk activity would have gone unnoticed in experiments using genetic depletion of Jnk activity. Targeting of the protein activity directly by small molecule inhibitors affords rapid inhibition of the target. In this case, the rapidity was sufficient to distinguish the effects of the same protein acting just hours apart.

This technology also enabled the discovery of unexpected biochemical properties of important kinases that may not have been possible to find with other techniques. The unfolded protein response is a transcriptional program that increases the cell's ability to rescue unfolded proteins once they are sensed in the ER lumen (Bernales et al., 2006). The transmembrane kinase Ire1 is the central protein of the unfolded protein response. When the amount of unfolded proteins in the lumen of the endoplasmic reticulum reaches

a threshold level, Ire1 oligomerizes and the subunits trans-autophosphorylate, leading to activation of Ire1's cytoplasmic RNase domain (Bernales et al., 2006). This RNase catalyzes the cleavage of a unique mRNA, termed *HAC1*^{u}, which is the first step in its maturation toward becoming *HAC1*^{i}, a competent template for ribosomal translation. The product of *HAC1*^{i} translation is the Hac1 transcription factor, which induces the expression of genes of the unfolded protein response. The kinase activity of Ire1 is required for this signaling module to function properly, but the process by which autophosphorylation results in activation of the RNase of Ire1 was mysterious. Using AS mutants of Ire1 that bind the bumped inhibitor 1NM-PP1, the authors found that the kinase activity was required to stabilize the kinase active site in an open conformation (Papa et al., 2003). This open conformation allows the kinase to bind ADP, which allosterically stimulates the RNase domain's activity. ADP levels are high when cellular energy levels are low. Therefore, Ire1 may function as a node that integrates information on the folded state of proteins in the ER and the levels of cellular energy stores. It is possible that both conditions need to be satisfied for a complete unfolded protein response.

Bishop and colleagues recently reported efforts aimed at extending this "bump-hole" approach to protein phosphatases (Bishop and Blair, 2006; Hoffman et al., 2005). These investigators mutated isoleucine 219 of protein tyrosine phosphatase 1B (PTP1B) to alanine to create a "hole" in the active site of the enzyme. Addition of a methyl group to the six-membered ring of a known generic PTP inhibitor (Andersen et al., 2000) yielded the bumped inhibitor **3** (Figure 1.4) that was greater than thirty-fold selective, as measured by K_i, for the mutant enzyme versus the wild type enzyme. Surprisingly,

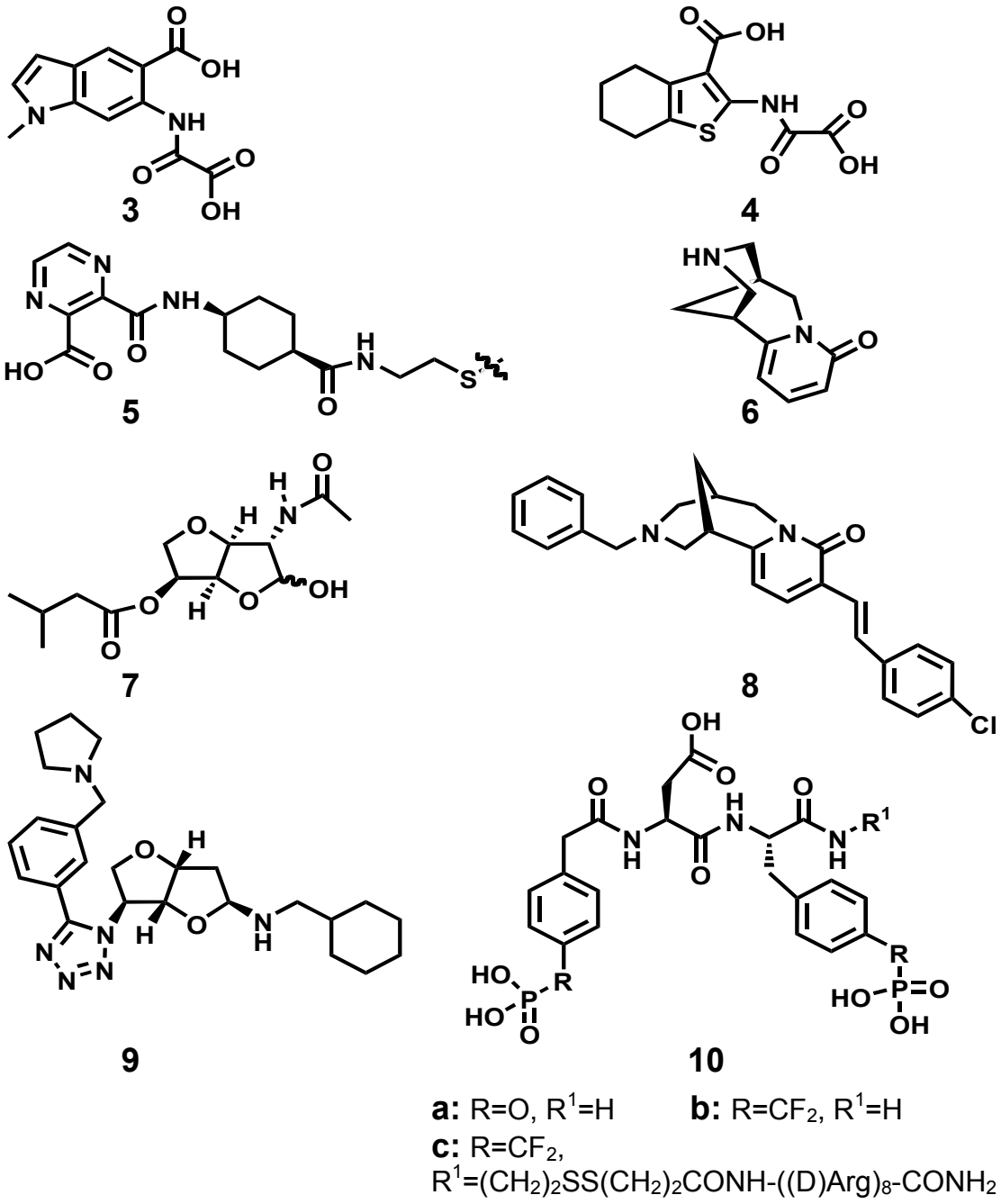


Figure 1.4. Phosphatase Inhibitors, their natural product scaffolds, and phosphate mimics. 3 = Allele-specific bumped inhibitor. **4** = PTP1B inhibitor. **5** = phosphate mimic discovered with tethering technology. **6** = Cytisine. **7** = Furanodictin A. **8** = VE-PTP inhibitor based on **6**. **9** = Shp-2 inhibitor based on **7**. **10** = PTP1B binder **(a),** inhibitor **(b),** and cell permeable inhibitor **(c).**

compound **4** (Figure 1.4) (Iversen et al., 2000), which was designed to and indeed does target wild type PTP1B (K_i=6.0 μM) was more potent for PTP1B with the I219A mutation (K_i=230 nM) (Bishop and Blair, 2006). This result is noteworthy because it implies that I219 acts as a true gatekeeper for PTPs (Bishop and Blair, 2006). Exploitation of this important information will hopefully bring phosphatase chemical genetics up to speed with the kinases.

In recent years phosphatases have emerged as genuine drug targets, motivating interest in the development of inhibitors for wild type phosphatases (Bialy and Waldmann, 2005). For example, PTP1B is being targeted for inhibition by many groups in academia and industry due to its role in diabetes and obesity (Dube and Tremblay, 2005). Working at Sunesis, Erlanson and Wells pioneered an approach termed "tethering" for the discovery of lead compounds (Erlanson et al., 2003). Tethering involves mutating an amino acid in or near to the active site of the enzyme under study to cysteine. Then, a relatively small library of disulfide-containing molecules is added to the enzyme in the presence of catalytic amounts of reducing agent. A molecule that binds to the active site of the protein will be in the vicinity of the engineered cysteine and thus is susceptible to disulfide exchange resulting in a covalent bond between the protein and the small molecule. In the presence of a catalytic quantity of reducing agent this process is dynamic and will ultimately result in labeling of the protein with the molecule from the library that has the most free energy of binding. Mass spectrometry of the intact protein-small molecule conjugate can identify the small molecule, which can serve as a lead compound for further optimization. Erlanson et al. used this technique with PTP1B as a

template to identify a novel phosphotyrosine mimetic, compound **5** (Figure 1.4), that may in the future be a useful constituent of phosphatase inhibitors (Erlanson et al., 2003).

Waldmann and colleagues have taken an innovative approach toward the synthesis of small molecule libraries and applied these libraries to the identification of protein phosphatase inhibitors (Noren-Muller et al., 2006). Their approach, termed "biology-oriented synthesis," takes inspiration from the structural core of well-known natural products, which have been refined by evolution to be exceptionally good drugs. The skeleton of the natural products Cytisine (Figure 1.4, compound **6**) and Furanodictin A (Figure 1.4, compound **7**) were derivatized with various moieties including alkyl, alkenyl, acyl, and aryl chains, and tetrazole and triazole rings. Screening the resulting libraries against various protein phosphatases yielded inhibitors **8** and **9** (Figure 1.4) of VE-PTP and Shp-2, respectively, with K_is in the low micromolar range. Screening libraries derived from biology-oriented synthesis is not limited to phosphatases and should prove applicable to a wide range of enzymes.

Zhang, Lawrence, and colleagues capitalized on the unexpected finding that PTP1B contains a site proximal to the active site that also binds phosphotyrosine (Puius et al., 1997). Using similar logic to Cole for the development of bisubstrate inhibitors, these investigators reasoned that a small molecule inhibitor that occupies both sites would be exceptionally potent and selective due to additivity of binding energy (Shen et al., 2001). A modestly sized library of phosphotyrosine-containing molecules was synthesized and tested for binding to a catalytically inactive mutant of PTP1B that maintains the ability to bind substrate (Shen et al., 2001). Compound **10a** (Figure 1.4) was found to bind with a K_d of ~30 nM to the mutant enzyme. To convert this into a

useful inhibitor for wild type PTP1B, the investigators exchanged the labile phosphates with non-hydrolyzable difluorinated phosphonate analogues. They generated compound **10b** (Figure 1.4), which inhibits PTP1B with a K_i of 2.4 nM and exhibits remarkable selectivity.

Highly charged molecules such as **10b** typically show poor cellular permeability. One method to deliver impermeable molecules to cells is by derivatization with a cell penetrating peptide (CPP). Lee et al. attached the cell penetrating peptide $((D)\text{-Arg})_8$ to the inhibitor **10b** through a disulfide linkage (Lee et al., 2005). The resulting conjugate, **10c** (Figure 1.4), displayed a K_i three orders of magnitude lower than the parent inhibitor **10b**, indicating that it no longer effectively inhibits PTP1B. This is due to a "self-silencing" effect whereby the negatively charged phosphonates interact intramolecularly with the positively charged arginine side chains. Once delivered to cells, the disulfide bond linking the parent inhibitor to the CPP is reduced by intracellular thiols. As it is no longer intramolecular, the CPP no longer quenches the inhibitor. Thus, the inhibitor functions only when delivered to cells. Structural studies of **10b** in complex with PTP1B showed that the design principles were indeed responsible for the observed potency and selectivity; both the active site and proximal site are engaged by **10b** (Sun et al., 2003).

1.2.3. Rescue of mutant kinase function by small molecule complementation

Recently, investigators have succeeded in using small molecules that have the opposite effect on protein function as the inhibitors described above; these molecules function as direct activators of protein kinases (Qiao et al., 2006; Williams et al., 2000). The first step in this approach is to identify and mutate an active site residue that is

required for catalysis, thus inactivating the enzyme. A small molecule containing a similar functional group as the mutated residue is then added to the enzyme to restore its catalytic abilities. Early work showed that this strategy could rescue the activity of mutant serine proteases (Carter and Wells, 1987; Craik et al., 1987) and aminotransferases (Toney and Kirsch, 1989). Cole and colleagues extended this approach to achieve small molecule activation of the tyrosine kinases Csk and Src in living cells (Figure 1.5) (Qiao et al., 2006; Williams et al., 2000). They identified a conserved arginine in these kinases that forms hydrogen bonds to both the phenolic oxygen of the tyrosine to be phosphorylated and the catalytic aspartic acid residue of the enzyme. Since this arginine plays a critical role in catalysis, mutating it to alanine resulted in substantial loss of activity of both enzymes (~3,000 fold reduction in k_{cat} for Csk and ~200 fold reduction in k_{cat} for Src, relative to each wild type kinase). With these inactive enzymes in hand, Cole and colleagues screened various small molecules to determine if any of them could complement the activity of the mutant kinases. The molecules that were found to confer activity all contained two nitrogen atoms capable of donating hydrogen bonds and had the potential to have one positive charge, much like the guanidinium side chain of arginine. Of these, imidazole led to the largest increase in mutant kinase activation (~60 fold increase in velocity for Csk in the presence of 50 mM imidazole and ~100 fold increase for Src in the presence of 25 mM imidazole, relative to the mutant kinase in the absence of imidazole). Importantly, in the presence of 25 mM imidazole, mutant Src displayed ~50% of the activity of wild type Src, indicating that robust rescue of mutant kinase function can be achieved (Figure 1.5) (Qiao et al., 2006).

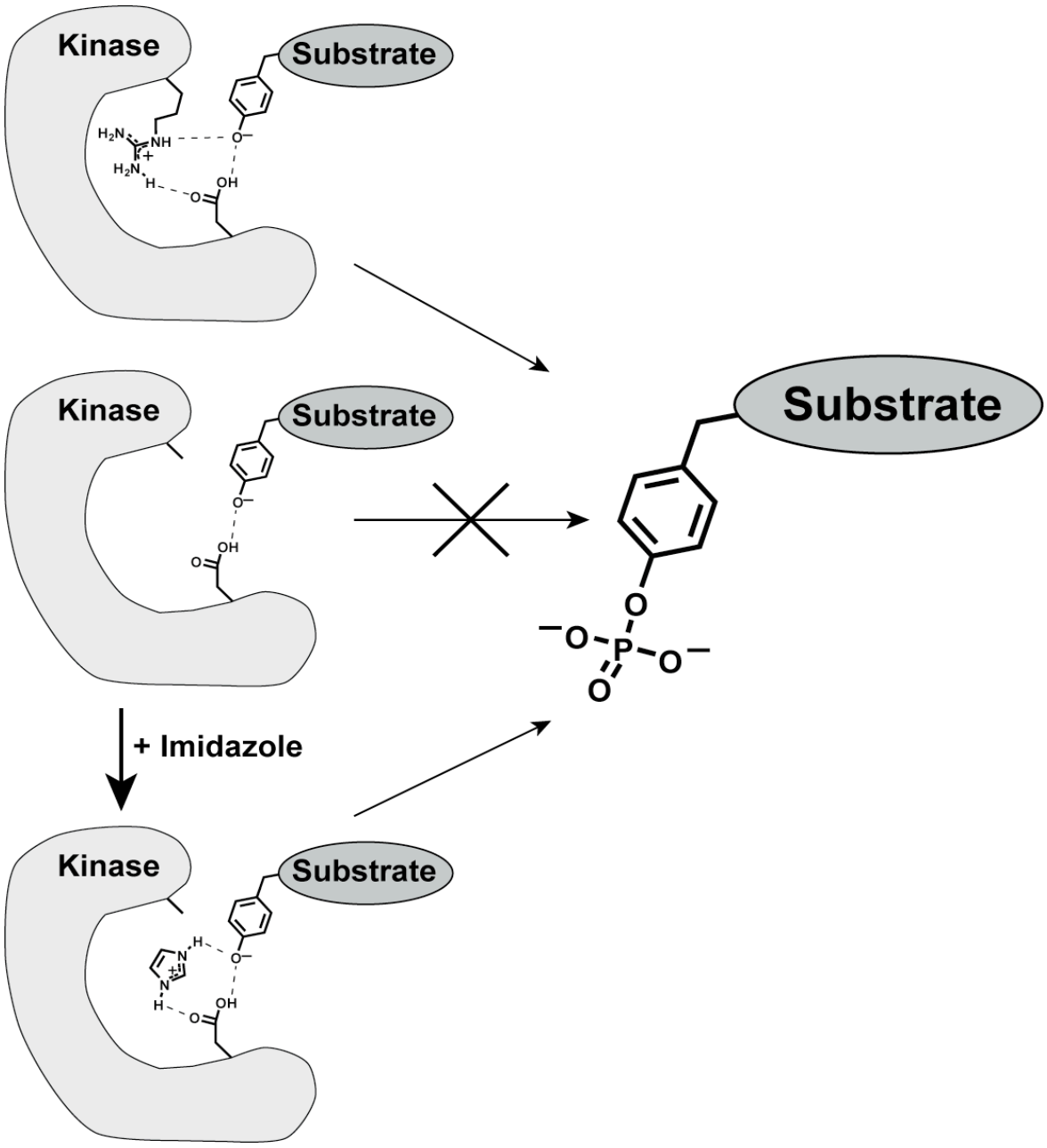


Figure 1.5. Rescue of mutant kinase function by small molecule complementation. **A.** Hydrogen bonds of an arginine of a wild type kinase help to orient the catalytic aspartic acid of the kinase and the tyrosine phenol of the substrate, leading to substrate phosphorylation. **B.** A kinase with a mutation of the arginine to alanine can not catalyze substrate phosphorylation. **C.** Addition of exogenous imidazole to the mutant kinase rescues its function, leading to small molecule control over protein phosphorylation.

The successful rescue of Src function *in vitro* prompted Qiao et al. to explore whether mutant Src can be rescued in live cells with imidazole, which is cell permeable (Qiao et al., 2006). Addition of imidazole to cells expressing mutant Src resulted in an increase in global protein tyrosine phosphorylation as well as an increase in Src activation-loop autophosphorylation. Interestingly, just four minutes after washout of imidazole, activation-loop phosphorylation was reduced to background, indicating that rescue is reversible and providing insight into the kinetics of phosphatase-mediated dephosphorylation in cells. Since imidazole treatment of these cells resulted in rapid Src activation, the investigators performed a comparative analysis of phosphotyrosine-containing proteins in the absence of Src activity and after 5 minutes of Src stimulation with imidazole. This resulted in identification of 18 previously unknown Src targets, of which 8 were confirmed with immunoprecipitation studies. Likewise, gene expression changes following rapid Src activation were determined. The expression pattern varied significantly from previous results using cells expressing constitutively active Src. These results highlight the benefit of using rapidly-acting small molecules to control protein function.

1.2.4. Identification of kinase substrates

Owing to their importance in cellular regulation, hundreds of researchers have focused on identifying substrates of protein kinases. Methods for this purpose, among others, include direct enrichment of phosphoproteins with immunochemical or metal reagents, mass spectrometry, 2-dimensional gel electrophoresis, and protein chip

technology. Here, only those methods that are rooted in organic chemistry will be discussed.

Several investigators have used chemical reactions to convert phosphorylated residues to a new epitope that can be selectively enriched or monitored by downstream applications, typically mass spectrometry (Oda et al., 2001; Zhou et al., 2001). In one incarnation of this approach, the well-known lability of phosphoserine and phosphothreonine to highly basic conditions was used to convert these residues to a biotinylated site, enabling enrichment of the phosphorylated proteins or proteolytic peptides thereof (Figure 1.6A-C) (Oda et al., 2001). Basic conditions lead to β-elimination of the phosphate in these residues resulting in a dehydroalanine residue, which is a Michael acceptor capable of reacting with nucleophiles, such as thiols. Following protection of native cysteine residues by oxidation, Chait and colleagues performed β-elimination and then reacted the dehydroalanine residues with ethanedithiol to create unique sulfhydryl groups at each site of serine and threonine phosphorylation. Then, a compound consisting of biotin linked to a thiol-reactive maleimide group was added, resulting in selective biotinylation at sites of former serine and threonine phosphorylation. Tryptic digestion followed by avidin affinity chromatography yields peptides that can be analyzed by mass spectrometry to determine the protein to which they belong and in ideal cases, the site of phosphorylation can be determined.

Hathaway and colleagues and Shokat and colleagues took advantage of the β-elimination pathway to convert sites of serine and threonine phosphorylation to lysine analogues, enabling lysine-specific proteolytic digestion at these sites (Figure 1.6A,D-F) (Knight et al., 2003; Rusnak et al., 2002). Following base-induced β-elimination, the

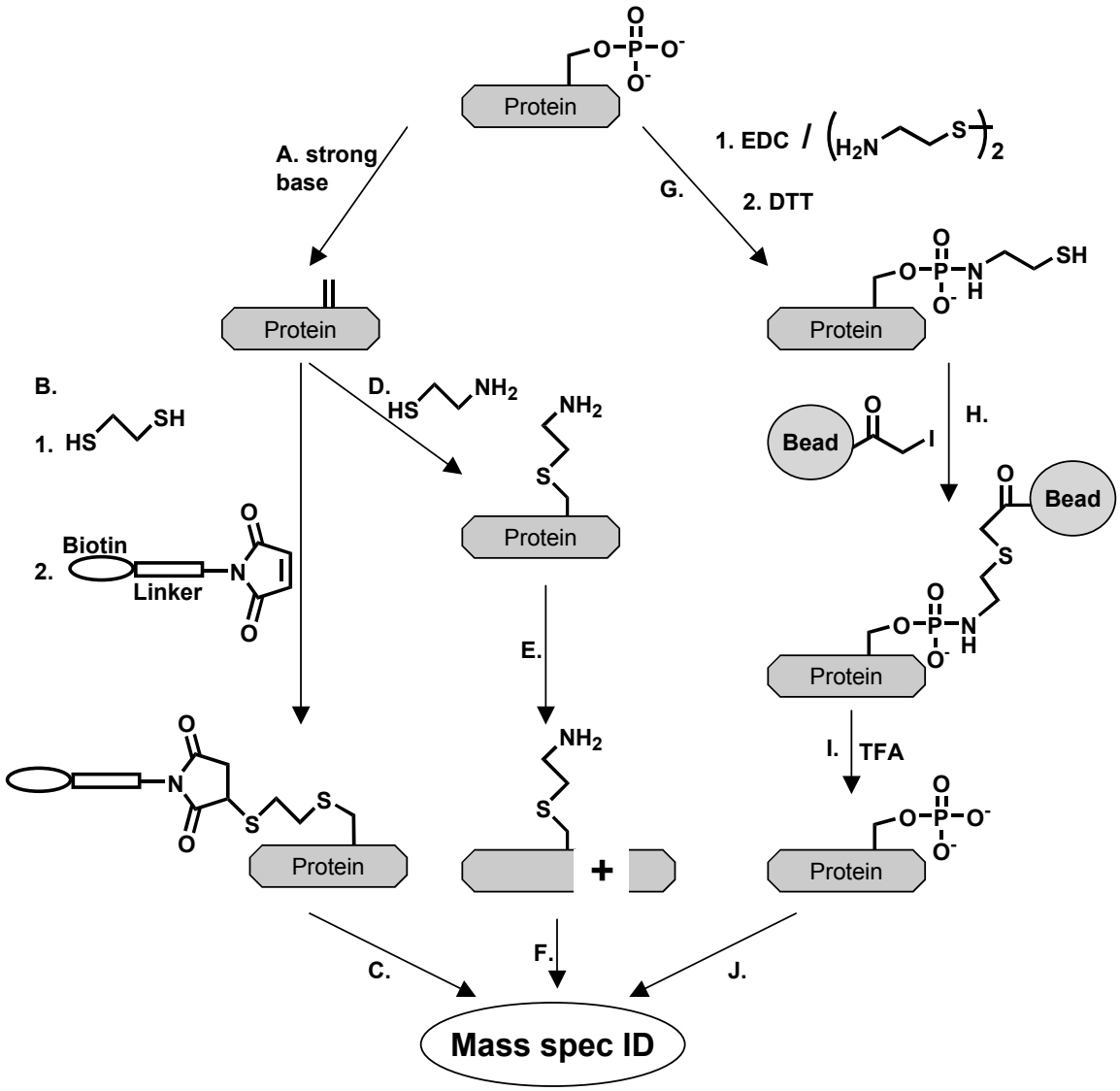


Figure 1.6. Chemical Approaches for the identification of phosphoproteins. A. Strong base induces β-elimination of phosphate at pSer/pThr residues, resulting in a dehydroalanine (DHA). **B.** Installation of biotin at pSer/pThr sites. **C.** Enrichment by avidin chromatography followed by mass spec analysis. **D.** Conversion of DHA to a lysine isostere. **E.** Proteolysis results in new peptide fragments. **F.** Identification by mass spec. **G.** Conversion of phosphates to phospohramidates. **H.** Enrichment by covalent chromatography. **I.** Cleavage and purification of phosphorylated protein. **J.** Identification by mass spec. Note: only pSer is shown. Protecting group steps not shown.

resulting dehydroalanine is reacted with cysteamine (2-aminoethanethiol) to form the lysine isosteres aminoethylcysteine and β-methylaminoethylcysteine at sites of serine and threonine phosphorylation, respectively. Cleavage of the resulting proteins with lysine-specific proteases yields peptides with novel termini, which can be analyzed by mass spectrometry. Alternatively, by derivatizing polymer supports with cysteamine through an acid-labile carbamate linkage, this strategy can be used to enrich phosphorylated proteins on the solid phase (Knight et al., 2003). Following enrichment, strong acid liberates formerly phosphorylated proteins or peptides from the solid support, which can then be subjected to lysine-specific proteolysis and mass spectrometry. This solid-phase strategy is particularly promising as it allows for enrichment and selective proteolysis, thus significantly reducing the complexity of the sample.

Aebersold and co-workers reported another chemical means to selectively enrich phosphorylated proteins (Figure 1.6G-J) (Zhou et al., 2001). After blocking of amines and thiols, phosphorylated residues are converted to phosphoramidates by carbodiimide-mediated condensation of phosphate with an exogenous amine. In this step, carboxylic acids of the peptides or proteins under study are converted to acid-stable amides. Treatment with strong acid regenerates the phosphates, which are then subjected to another round of phosphoramidate formation using cysteamine as the exogenous amine. At this stage, the result is conceptually the same as that described above for the sequence utilized by Chait, et al. (Oda et al., 2001); the peptides display thiol residues only at sites of former phosphorylation. Instead of reacting these thiols with a maleimide linked to biotin, these investigators chose to directly conjugate the peptides to a solid support by reaction of the thiols with glass beads containing thiol-reactive iodoacetyl groups.

Washing the beads and subsequent treatment with strong acid cleaves the phosphoramidate, releasing the phosphorylated peptide or protein, which can then be analyzed by mass spectrometry. The major advantage of this approach over those utilizing β-elimination is that phosphotyrosine is subject to carbodiimide-phosphoramidation, but not to base induced β-elimination. Therefore, peptides or proteins containing phosphotyrosine will be enriched and can be detected. Disadvantages include the greater number of required steps, thus reducing the yield of phosphorylated peptides or proteins, and the resulting analyte still contains phosphorylated aminoacids, which are known to ionize poorly compared to their non-phosphorylated counterparts in mass spectrometry. In the β-elimination approaches, the phosphates are permanently converted to other moieties, thus increasing their likelihood of identification.

The approaches described above offer powerful solutions to phosphoprotein and modification site identification, but they cannot be used to determine which kinase is responsible for each phosphorylation event. Some chemical strategies that link a given phosphorylation event to its kinase are discussed below.

The bump-hole approach to protein kinase inhibition (see section **2.2**) can operate in an alternative mode to enable discovery of substrates of a kinase of interest (Shah et al., 1997; Shogren-Knaak et al., 2001). Instead of creating a bulky inhibitor, a bulky ATP analogue that is still capable of transferring the phosphoryl group of its γ-phosphate to substrates is used together with the AS-kinase of interest carrying a hole in its active site. A single phenyalanine to glycine mutation in the ATP-binding pocket of Cdk1 allowed the kinase to utilize N^6-(benzyl)-ATP, an analogue not accepted by wild type kinases (Ubersax et al., 2003). Incubating yeast extract with Cdk1/Clb2 and N^6-(benzyl)-ATP

radiolabeled at the γ-phosphate resulted in the specific phosphorylation of many proteins. 169 new substrates were identified, 11 of which were confirmed with further experiments. Among the targets, several regulatory molecules and protein machines involved in cell-cycle events were identified, providing a basis for further detailed physiological analysis.

O'Shea, Shokat, and colleagues recently described a proteomic screen using the bump-hole approach for the high-throughput identification of yeast protein kinase substrates (Dephoure et al., 2005). This strategy used a previously described yeast library in which each of 4250 strains expressed an epitope-tagged version of one protein (Ghaemmaghami et al., 2003). Upon addition of N^6-(benzyl)-[γ-^{32}P]ATP and an analog sensitive version of the Pho85-Pcl1 yeast cyclin-dependent kinase complex (containing the gatekeeper F82G mutation in the Pho85 subunit) to lysates of each strain, the targets of Pho85-Pcl1 were labeled with radioactive phosphate. The tagged protein in each lysate was then immunopurified and screened by SDS-PAGE followed by autoradiography and immuoblotting to determine the protein targets of Pho85-Pcl1. Subsequent analysis confirmed that 24 of the 34 hits in the primary screen were direct substrates. The investigators then compared the phosphorylation specificity of the Pho85-Pcl1 complex versus Pho85 in complex with a different cyclin, Pho80, using substrates discovered in the screen and other previously known substrates. Of these, 13 substrates were specifically phosphorylated by one but not the other complex, while 14 substrates were phosphorylated with approximately equal efficiency by both complexes.

Interfacing analog-sensitive kinases with a cellular library containing all known proteins fused to an affinity handle to determine kinase substrates is not possible for most

model organisms. Shokat and colleagues have described a new method to determine kinase substrates in cells from mice expressing endogenous levels of analog sensitive kinases via knock-in methodology (Allen et al., 2005; Allen et al., 2007). In place of radiolabeled bumped ATP, bumped ATP-γ-S is used as substrate for the kinase. ATP-γ-S was added to digitonin-permeabilized mouse embryonic fibroblasts (MEFs) from mice expressing AS-Erk2 resulting in thiophosphorylation of Erk2 targets at native sites of phosphorylation (Allen et al., 2007). Then, *p*-nitrobenzylmesylate is added to alkylate the thiophosphates to convert them into unique thiophosphodiesters that can be recognized by specific antibodies. Importantly, these antibodies do not cross-react with the thioether resulting from alkylation of cysteine residues by *p*-nitrobenzylmesylate. Proteins are then immunoprecipitated by the thiophosphodiester specific antibody to purify the proteins that had been thiophosphorylated by Erk2. These proteins are then subjected to SDS-PAGE and mass spectrometry for phosphoprotein identification. This strategy has sufficient sensitivity to detect the Erk2 substrate Tpr. It is known that ATP-γ-S is generally not as efficient a substrate for kinases as ATP. Therefore, the investigators performed preliminary *in vitro* experiments using 15 different analog sensitive kinases for phosphorylation of known protein substrates with bumped ATP-γ-S. Of these, 13 were able to catalyze sufficient levels of thiophosphorylation to enable detection. However, not all of the substrates of Erk2 visualized by immunoblotting could be identified by subsequent mass spectrometry, questioning the sensitivity of the approach. Perhaps the sensitivity could be improved by interfacing this strategy with either of the β-elimination-mediated phosphoprotein identification techniques discussed above. For this to be successful, a means to perform orthogonal β-elimination at sites of

thiophosphorylation while leaving native phosphates intact must be devised. It is very likely that a thiophosphodiester will undergo β-elimination under more mild basic conditions than that required for native phosphomonoesters due to the better leaving group ability of the benzylated thiophosphate compared to phosphate.

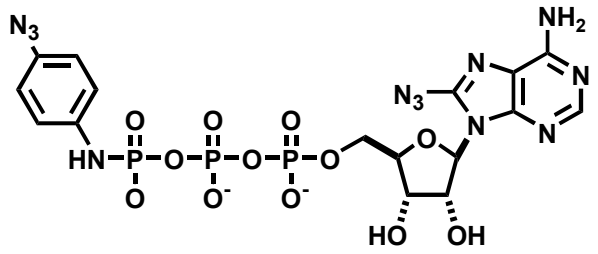
1.2.5. Identification of a kinase for a known phosphoprotein

The techniques discussed above all focus on discovering substrates of protein kinases. It is possible that future efforts using these and other phosphoproteomic strategies will provide an exhaustive inventory of *all* kinase substrates matched to their kinases. In the meantime, it is of critical importance to be able to perform the reverse experiment: the identification of kinases for *known* protein phosphorylation events. Recently, chemistry-inspired strategies have been reported toward reaching this goal.

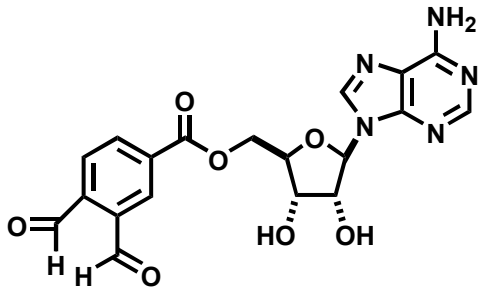
Cole and co-workers took advantage of their ability to construct bisubstrate inhibitors of protein kinases (discussed in section **2.2**) to prepare an affinity bait for the identification of the protein kinase responsible for a specific phosphorylation event (Shen and Cole, 2003). To increase the chance of success, these authors used the entire protein substrate to prepare the bait instead of just a small peptide. Src kinase is itself a substrate for phosphorylation by Csk. A bisubstrate inhibitor consisting of epitope-tagged Src fused to ATP was prepared by expressed protein ligation (see section **2.8.1.2**). Incubation of the bisubstrate inhibitor with Csk in the presence of cell extract, followed by pull-down of inhibitor through its epitope-tag revealed that this method could facilitate the detection of protein-protein interaction between Csk and Src.

Another strategy from the Cole group is based on the use of a generic kinase-targeted bivalent photo-crosslinking reagent (Figure 1.7, compound **11**) (Parang et al., 2002). The reagent is an ATP analogue that when bound to a kinase:substrate complex positions one of its photoreactive azide groups near the kinase and its other photoreactive azide group toward the protein substrate. Upon UV irradiation, crosslinking between the azides and proteins converts the ternary complex to one species, capable of being pulled-down. The authors showed that this molecule can be used to crosslink Csk and Src. For future experiments using **11** to discover new kinase-substrate pairs, the fact that the crosslink can be reversed enzymatically with phosphodiesterase is quite attractive. In addition, the crosslink should be susceptible to simple treatment with acid to cleave the phosphoramidate linker.

Shokat and co-workers designed and synthesized a mechanism-based crosslinker (**12**) to isolate kinase:substrate complexes for the identification of the kinase responsible for a known phosphorylation event (Figure 1.7) (Maly et al., 2004). The generic protein kinase inhibitor 5'-fluorosulfonylbenzoyl adenosine (FSBA) occupies the ATP-binding site of kinases where it covalently labels the kinase's catalytic lysine. To convert this general covalent inhibitor into a specific crosslinker, Maly et al. replaced the mono-reactive aryl sulfonyl fluoride in FSBA with a bi-reactive *o*-phthaldialdehyde (OPA) group. The OPA group is capable of forming intermolecular crosslinks by first reacting with an amine on one molecule to form an imine. This imine then reacts with a thiol from another molecule, resulting in a stable isoindole linkage between the two molecules. Kinase substrates are phosphorylated on serine, threonine, and tyrosine, but not cysteine, which is the only amino acid that contains the necessary thiol group for crosslinking by



11



12

Figure 1.7. Bifunctional crosslinkers for isolating kinase-substrate complexes.

11 = UV-activated crosslinking reagent. **12** = Mechanism-based crosslinking reagent.

this reagent. This missing piece is provided by mutation of the serine, threonine, or tyrosine of interest to cysteine. Initial studies using the serine/threonine kinases Akt1, PKA, p38 MAP kinase, and caseine kinase II together with their corresponding peptide pseudosubstrates containing the required cysteine mutation were incubated with crosslinker **12**. In all cases, crosslinking occurred. The three-piece crosslinking reaction carried out with pure components *in vitro* occurred with a yield of ~25%. The full scope of this promising approach for the identification of new interactions awaits further evaluation.

1.2.6. Fluorogenic probes of protein kinase function for live cell imaging

Traditional enzymology has provided deep insight into biomolecular function (Kornberg, 2003). However, it has been difficult to transport enzymological rigor to the study of cellular proteins in their native environment. This is because cellular proteins exist in a complex milieu whereas traditional enzymology studies are typically conducted with well-defined homogenous samples. The successful extension of the practice of enzymology to *in vivo* settings requires tools that are exquisitely specific with regard to what they are measuring. Only these tools are able to surmount the obstacle that exists in the heterogeneous world that is the living cell. The last decade has witnessed remarkable advances that live up to this challenge. For example, green fluorescent protein and its derivatives have been used to produce FRET-based sensors that reliably report protein activation status in live cells. Using this approach, several investigators have designed and utilized fluorescent-protein based FRET reporters of kinase activity (Sato et al., 2002; Ting et al., 2001). These probes typically consist of a single, genetically encoded

polypeptide containing two fluorescent proteins that undergo FRET: a phosphorylation substrate sequence for a given kinase of interest and a modular binding domain that specifically binds the phosphorylation sequence when it is phosphorylated (for example, an SH2 domain for a phosphotyrosine peptide sequence). When the endogenous kinase phosphorylates the probe, the distance between the two fluorescent proteins shortens, giving rise to an increase in FRET which can be monitored in live cells by fluorescence microscopy.

Fluorescent-protein based FRET sensors of kinase activity have been extensively used owing to their straightforward means of construction (standard molecular cloning), ease of delivery to cells (standard plasmid transfection), and ease of portability between laboratories. However, these tools do have some limitations, such as the change in FRET signal upon phosphorylation is typically only twenty to thirty percent (Sato et al., 2002; Ting et al., 2001). While this has been sufficient to monitor the activity of several kinases, the small change in signal upon phosphorylation could prove to be problematic for measuring particularly low abundance kinases or those that have slow turnover rates. Additionally, since each kinase activity probe requires two spectrally distinct fluorescent proteins, and hence occupies two optical channels, opportunities for imaging more than one event of interest is severely limited.

To overcome these limitations, chemists have synthesized peptidic probes of kinase activity (Figure 1.8) (Lawrence and Wang, 2007; Rothman et al., 2005b). To eliminate the requirement of two fluorophores per molecule of interest, investigators have concentrated on probe designs that require only one fluorophore. How then can a synthetic peptide probe be constructed with one fluorophore to sense and report kinase

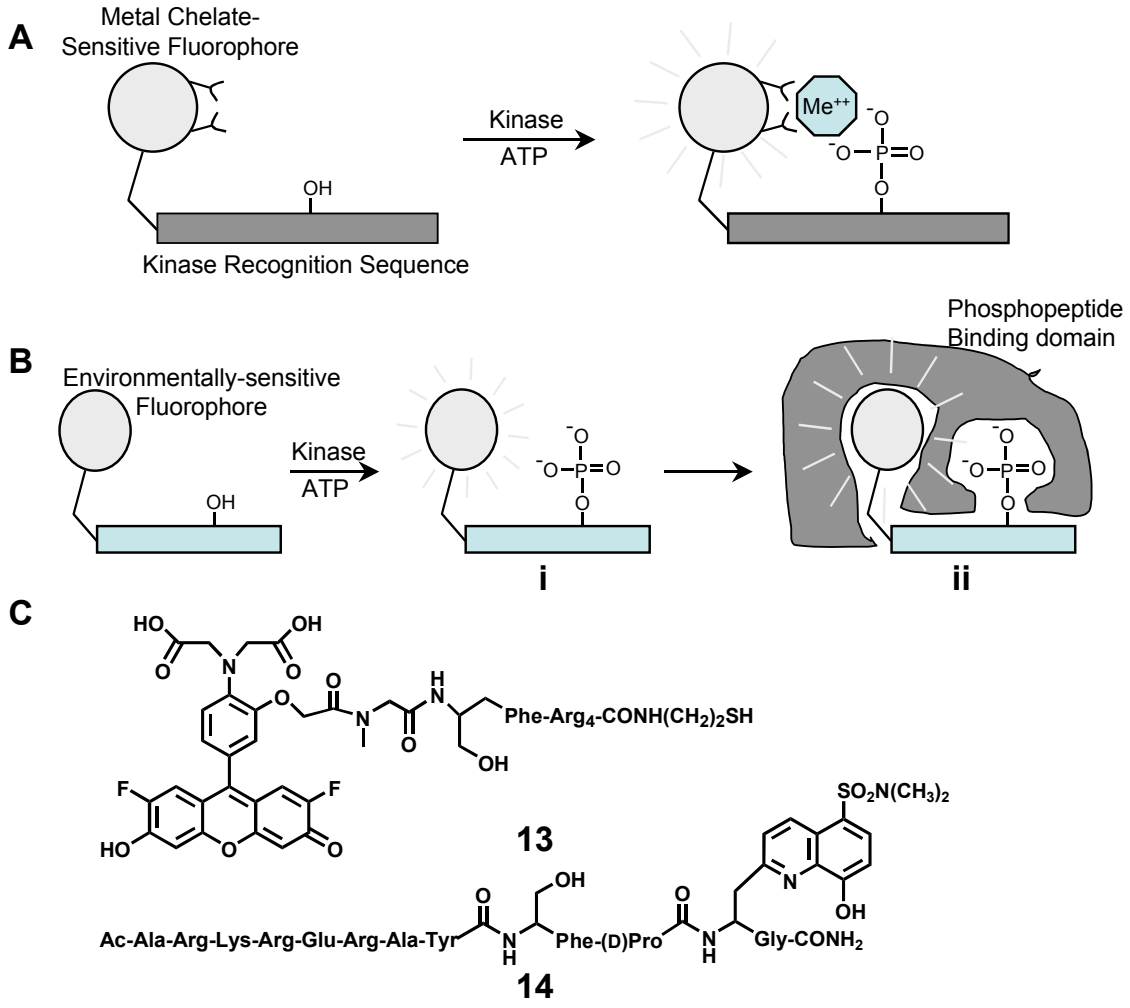


Figure 1.8. Peptide-based fluorescent kinase probes. A. Upon phosphorylation, a peptide containing the recognition sequence of the kinase of interest and a metal-chelate-sensitive (MCS) fluorophore chelates a divalent metal ion, leading to enhancement of fluorescence. **B.i.** A probe containing the recognition sequence of the kinase of interest and an environmentally-sensitive fluorophore (ESF). The polarity of the local environment is altered upon phosphorylation, leading to an increase in fluorescence. **ii.** An intensified increase in fluorescence is obtained upon phosphorylated-probe binding to a modular phosphopeptide binding domain. **C.** Structures of probes discussed in the text. **13** = MCS probe of PKC. **14** = MCS probe of Akt.

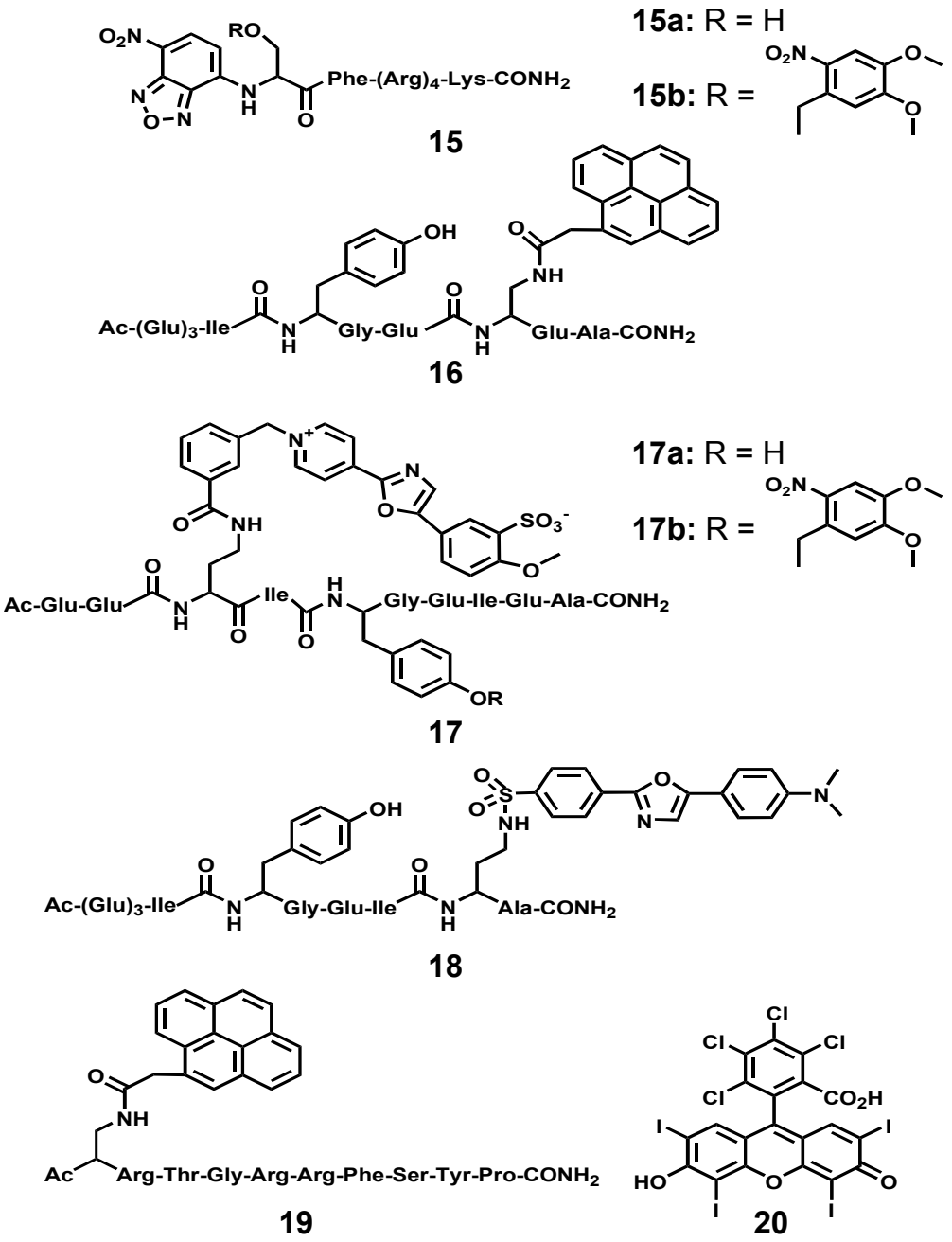


Figure 1.8, continued. 15a = ESF probe of PKC. **15b** = caged ESF probe of PKC. **16** = ESF probe of Src family kinases. **17a** = ESF probe of Src family kinases. **17b** = caged ESF probe of Src family kinases. **18** = ESF probe of Src family kinases that intensifies upon binding a SH2 domain. **19** = ESF probe of PKA that intensifies upon binding 14-3-3. **20** = Rose Bengal, the quencher used to further increase the dynamic range of probe **19**.

activity with a more robust change in signal compared to fluorescent protein FRET probes? Phosphorylation turns an uncharged hydroxyl of serine or threonine, or phenol in the case of tyrosine, into a negatively-charged phosphate-bearing residue. A properly positioned phosphate residue can readily participate in coordination of a divalent metal cation. If the phosphorylated peptide contains a properly positioned latent metal-chelate-sensitive fluorophore, then the phosphate and fluorophore may simultaneously coordinate an ion, enhancing the fluorescent properties of the fluorophore (Figure 1.8A). Lawrence and colleagues designed and synthesized the peptide-based PKC probe **13** containing a metal-chelate sensitive fluorophore (Figure 1.8C) (Chen et al., 2002). Upon phosphorylation of the peptide by PKC, a Mg^{2+} ion binds to the probe, bridging the fluorophore and phosphate. This leads to a fluorescence increase of ~3-fold. Utilizing the same principle, Shults and Imperiali have devised probes for the kinases Abl, PKC, PKA, MAPK2, Akt, Pim2 and Cdk2 containing the metal-chelate sensitive fluorophore Sox (Figure 1.8C, probe **14**) (Shults and Imperiali, 2003; Shults et al., 2005;) Shults et al., 2006). These sensors have demonstrated an impressive fold increase in fluorescence upon phosphorylation as high as ~9-fold in the presence of 10 mM Mg^{2+}. In the presence of lower concentrations of Mg^{2+} that better mimic endogenous concentration, the fold increase in fluorescence is more modest. Metal chelator based kinase probes have not yet been used inside live cells, but have proved useful in characterizing kinase activity in real time in cell lysates (Shults and Imperiali, 2003; Shults et al., 2005). For this class of sensors to be useful for live cell analysis, the requirement for supra-physiologial levels of cations must be addressed and in the case of the Sox fluorophore ($\lambda_{max,excitation}$ = 360 nm),

it would be beneficial if the molecule could be engineered so that its absorbs at a more biologically-compatible wavelength.

Lawrence and co-workers have presented an alternative class of peptide-based kinase activity probes containing environmentally-sensitive fluorophores. An example of this class is sensor **15a** (Figure 1.8C), which contains the environmentally-sensitive NBD fluorophore in close proximity to the residue targeted for phosphorylation by the kinase of interest (Yeh et al., 2002). Upon phosphorylation, the local environment changes drastically owing to the installation of the negatively charged phosphate (Figure 1.8B). This leads to a 2.0-fold increase in the fluorescence of the probe when excited at its $\lambda_{max,excitation}$ of 460 nm and a 2.5-fold increase when excited at 520 nm. This represents a substantial improvement over fluorescent-protein based FRET probes. This was the first fluorescence-based kinase sensor with a fold increase greater than 100% to be validated inside live cells. Upon activation of the PKC pathway in live cells by administration of the phorbol ester TPA, sensor **15a** reported real-time PKC activity. PKC activity is detectable in as little as four minutes and plateaus around fifteen minutes. Importantly, specificity of the probe for reporting PKC activity in the background of the greater than 500 other cellular kinases was demonstrated by the lack of fluorescence increase upon co-administration of TPA and a PKC-specific inhibitor.

For certain applications it is desirable to control when the probe is able to sense activity. For example, to monitor kinase activity during different phases of the cell cycle without having to continually re-administer the probe to the cells is advantageous. To this end, Lawrence and colleagues introduced a caging group on the phosphorylatable hydroxyl of sensor **15b** (Figure 1.8C) (Veldhuyzen et al., 2003). When the caging group

is present, phosphorylation cannot occur and hence the sensor is silent. Irradiation of the cells with UV light deprotects the hydroxyl group of the phosphorylatable residue, enabling it to be phosphorylated if PKC is active. This caging strategy endows probe **15** with the ability to monitor PKC more than once per application.

For live cell-based sensing of protein kinase activity it is clearly advantageous to use a probe that senses in an autonomous fashion, that is, one that does not require binding to additional molecules or ions. This is perhaps why Lawrence's environmentally-sensitive fluorescence strategy is the only peptide-based strategy of those discussed above that proved to be applicable to imaging kinase activity in live cells. Wang et al. recently extended the applicability of this type of sensor to the protein tyrosine kinases (Wang et al., 2006a). Probe **16** displayed a 4.7-fold increase in pyrene fluorescence upon phosphorylation (Figure 1.8C). This impressive increase without requiring any intermolecular interactions post-phosphorylation is explained by the close proximity of the phosphorylated tyrosine to the fluorophore. Before phosphorylation, non-phosphorylated tyrosine and pyrene interact via π-π stacking, leading to suppression of pyrene fluorescence. Upon phosphorylation, the electron density of the aromatic ring of the tyrosine side chain is reduced, consequently reducing the interaction between the ring and pyrene. This unleashes full fluorophore activity, resulting in the large fluorescence increase. In follow-up studies, Wang et al. modified this class of probe to produce additional probes, for example probe **17a**, containing fluorophores with optical properties more suitable for live-cell analysis (Figure 1.8C) (Wang et al., 2006b). In addition, this probe was caged on tyrosine as described above for probe **17b** (Figure

1.8C) (Wang et al., 2006b). This afforded a probe that could sense multiple rounds of Src kinase activity in live cells.

An additional peptide-based strategy has recently emerged from the Lawrence laboratory. This strategy uses an environmentally-sensitive fluorophore conjugated to a kinase substrate that binds to a third-party protein upon phosphorylation (Figure 1.8B) (Wang and Lawrence, 2005). For example, Wang et al. have described sensor **18** that is phosphorylated by Src and binds to the SH2 domain of Lck (Figure 1.8C). When the phosphorylated sensor binds the SH2 domain, the fluorophore is now positioned in a less polar environment leading to increases in fluorescence of 7.2-fold. Recently, Sharma et al. described another sensor (**19**) using this design strategy that senses PKA activity by binding to a 14-3-3 phosphoserine binding domain (Figure 1.8C) (Sharma et al., 2007). Additionally, several dye molecules were screened for their ability to quench the fluorescence of pyrene in sensor **19** via non-covalent interactions, resulting in the identification of Rose Bengal (**20**) which significantly quenches the fluorescence of **19** (Figure 1.8C) (Sharma et al., 2007). Upon binding of the phosphorylated sensor to the 14-3-3 domain, the non-covalent interaction between pyrene and Rose Bengal is disrupted, leading to a striking 64-fold increase in fluorescence. While the fold increase of fluorescence of these probes upon phosphorylation is impressive, the requirement for a third party, and in some cases fourth party, molecule to elicit a robust increase in fluorescence is a limitation for cell-based studies. Therefore, it will be highly desirable to develop probes that do not require additional molecules yet retain the robust increases described above. As this field is moving at a rapid pace, improvements will undoubtedly be made in the near future toward this goal.

The probes described above set a solid foundation for future studies aimed at monitoring kinase activation in live cells and tissues. Looking forward, it is reasonable to expect that these types of sensors will prove to be extremely valuable for clinical medicine. Inhibition of kinase activity is proving to be a useful clinical paradigm. Lawrence has pointed out that as drugs of this class increase in number and sophistication, a parallel increase in methods to determine drug-sensitivity of pathological samples will be required. Peptide-based kinase probes will be useful for screening new kinase inhibitors and may play a role in molecular pathology and non-invasive radiology for defining the pathophysiology of individual clinical samples.

1.2.7. Covalent inhibitors as probes of protein kinase and phosphatase activity

Activity-based protein profiling (ABPP) combines chemical synthesis with proteomics to determine the active enzyme complement of a given biological sample (Evans and Cravatt, 2006). Chemical probes designed to covalently label the active sites of a class of enzymes are synthesized with an appended reporter tag, such as a fluorophore or biotin. Alternatively, the probe may contain a small bio-orthogonal group such as an azide or alkyne enabling detection in a second step by reacting the bio-orthogonal group with an appropriately derivatized group (Ovaa et al., 2003; Speers et al., 2003). Incubation of an ABPP-probe with a biological sample results in the covalent labeling of the active enzymes of the probe-directed class. The proteins are then separated and detected *via* the reporter tag. Probes of many enzyme classes, including hydrolases and oxidoreductases have been synthesized and applied to the characterization of various

biological samples. Recently, several groups have synthesized probes for protein kinases and phosphatases.

Some of the existing strategies for activity based profiling of tyrosine phosphatases takes advantage of the ability of these enzymes to dephosphorylate substrates resembling phosphotyrosine. One strategy employs fluorinated phenylphosphate derivatives **21** and **22** (Figure 1.9) attached to a fluorescent probe (Lo et al., 2002; Zhu et al., 2003). The phosphate group of these probes undergoes hydrolysis in the active site of protein tyrosine phosphatases. The resulting phenoxide rearranges to form a highly electrophilic quinone methide, which can react with a nearby nucleophilic side-chain of the phosphatase, thus covalently labeling the phosphatase with a fluorescent probe. In another strategy, Zhang and co-workers synthesized α-bromobenzylphosphonate probe **23** (Figure 1.9) containing either biotin or a rhodamine derivative (Kumar et al., 2004; Kumar et al., 2006). This probe covalently binds protein tyrosine phosphatases, most likely by labeling the active site cysteine. Using the α-bromobenzylphosphonate probes, Zhang and colleagues compared global protein tyrosine phosphatase activity between a non-transformed cell line and a transformed cell line (Kumar et al., 2004). Additionally, they compared global protein tyrosine phosphatase activity across six different transformed cell lines (Kumar et al., 2006). The results showed that the pattern of protein tyrosine phosphatase activity is different in each cell line tested. This information may serve both as a starting point for the determination of how the various active phosphatase complements of each cell line contribute to the specific cellular phenotype and as a clinically useful marker to categorize disease and inform treatment.

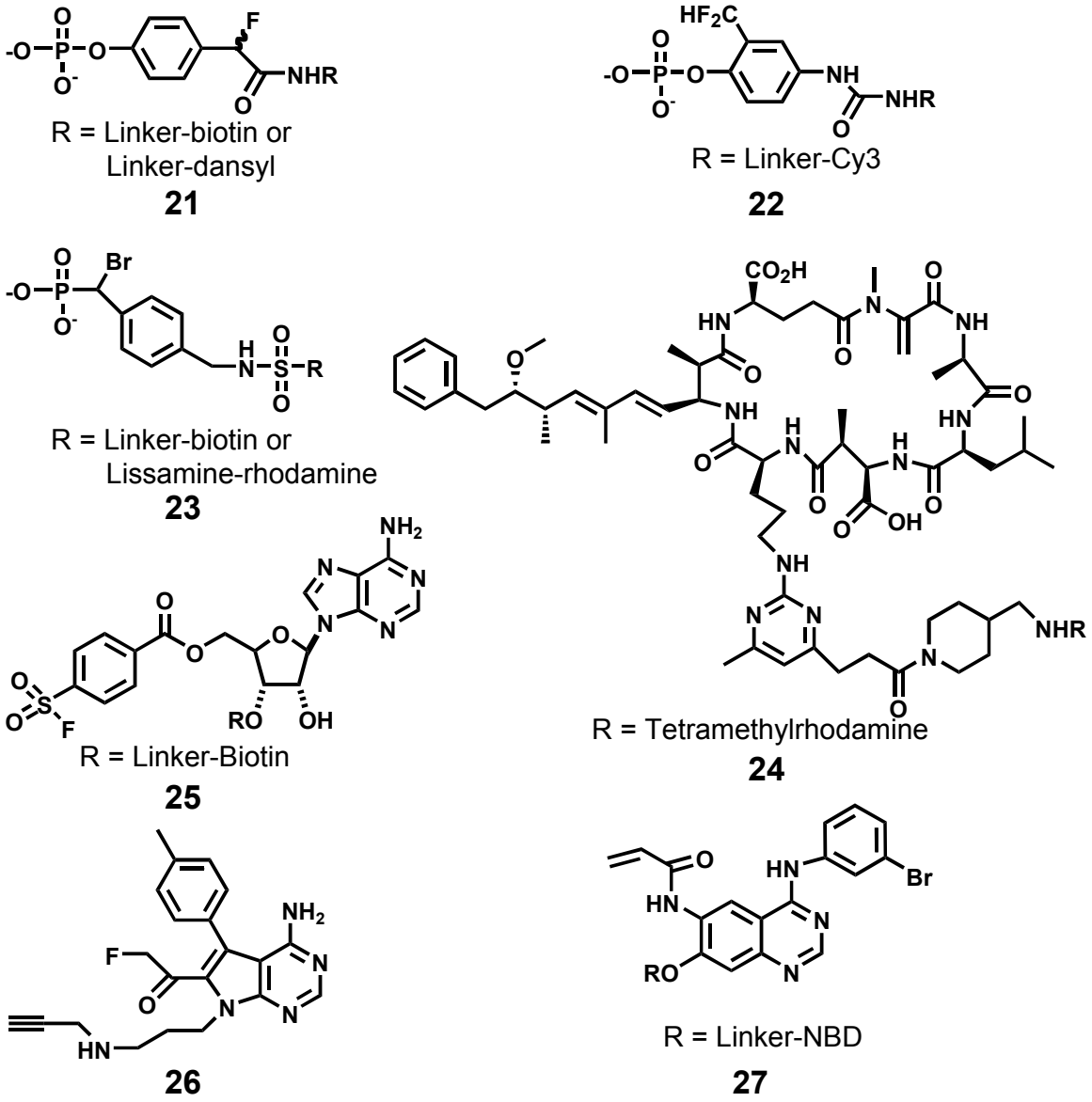


Figure 1.9. Activity-based probes of protein phosphatases and kinases. 21 through **23** = Protein tyrosine phosphatase probes. **24** = PP1 and PP2-A probes based on the natural product microcystin. **25** = General protein kinase probe. **26** = Clickable RSK probe. **27** = Allele selective probe used to study EGFR. These probes are potentially useful for any AS-kinase with either a natural or mutant cysteine near to the active site.

Shreder et al. derivatized the known natural product microcystin, an inhibitor of class PP-1 and PP-2A protein serine/threonine phosphatases, with a rhodamine-based fluorophore (**24**) (Shreder et al., 2004). These investigators showed that probe **24** specifically reports on the activity of PP-1 and PP-2A class phosphatases in a complex proteome.

Very recently, activity based probe methodology has been applied to monitoring kinase activity. Ratcliffe et al. derivatized the generic covalent protein kinase inhibitor FSBA with biotin to enable detection of active kinases by western blotting with avidin (Figure 1.9, compound **25**). Thus far, the authors have validated the approach *in vitro* with the purified protein kinases Alk5 and Cdk2 (Ratcliffe et al., 2007).

The other examples of ABPP probes of protein kinases published thus far have focused on the construction of probes specific for one kinase instead of the traditional use of activity based protein profiling of an entire family of enzymes simultaneously. In one example, Taunton and colleagues generated a probe specific for the C-terminal kinase domain (CTD) of wild type p90 RSK (Figure 1.9, compound **26**) (Cohen et al., 2007). The probe design is based on a previously described irreversible covalent inhibitor of the enzyme that targets a cysteine residue in the kinase ATP-binding site. Derivatization of the inhibitor with a short alkynyl linker furnished probe **26**. The CTD of RSK is known to indirectly facilitate the activation of the N-terminal kinase domain (NTD) of this tandem kinase domain-containing protein. However, it was unclear whether this CTD-mediated activation of the NTD was the sole mechanism for NTD activation. To tease this apart, Cohen et al. determined the effect of CTD inhibition (probe **26** was used to verify that inhibition was complete) on NTD activation by two different stimuli known to

activate the NTD, the phorbol ester PMA and the endotoxin LPS. PMA-induced NTD activation decreased as a result of CTD inhibition, verifying that the CTD activates NTD in response to this stimulus. In contrast, inhibition of the CTD had no effect on NTD activation in response to LPS, suggesting that the transduction mechanisms for NTD activation downstream of LPS and PMA are distinct.

Extending the "bump-hole" approach of kinase inhibition discussed above, Shokat and colleagues designed allele specific probes of tyrosine kinases (Blair et al., 2007). Inspiration for this came from the development of irreversible covalent inhibitors of a small fraction of kinases such as RSK (described above) and EGFR that naturally contain a cysteine in their ATP-binding site. These investigators synthesized probe **27** (Figure 1.9), which contains an acrylamide moiety designed to covalently label this cysteine, a bulky group that selectively binds a mutant kinase engineered to contain a hole by subtractive mutagenesis of its gatekeeper residue, and a NBD fluorophore. For kinases that do not naturally have a cysteine in the required position, it is also necessary to install a cysteine by mutation of the spatially-equivalent residue. Using this strategy, Blair et al. monitored the activity of the EGF receptor and correlated it with known downstream effectors. They yielded direct evidence that downstream pathway activation is proportional to the fraction of EGF receptors kinases with available active sites capable of phosphorylating targets.

One potential drawback of these profiling approaches is that correlation of kinase or phosphatase activity (as measured by the covalent probe) to downstream cellular phenotypes may be hampered since the probe inactivates the enzyme. Therefore, when used at saturating conditions, these probes do not provide any information on the effects

of kinase activity after introduction of the probe. For experiments aimed at correlating protein kinase activity with downstream cellular changes in real time after probe administration, it may be more appropriate to utilize the fluorescence-based peptide sensors described in the previous section that do not destroy the activity of the kinase under investigation.

1.2.8. Semi-synthesis of defined phosphorylated proteins and their analogues

Site-directed mutagenesis is invaluable to the study of proteins, both *in vitro* and *in vivo*. However, such studies are limited by functional groups present within the 20 genetically-encodable amino acids. Chemical manipulation extends this approach by incorporating new functional groups (including unnatural amino acids, non-genetically encodable optical and biophysical probes, and post-translational modifications) into proteins. Chemical techniques for protein modification range from direct labeling with reactive probes to sophisticated strategies involving chemical synthesis merged with native or engineered biosynthetic pathways within living cells. These techniques are discussed here followed by a review of their application to the study of protein phosphorylation.

1.2.8.1. Semi-synthetic methods

1.2.8.1.1. Targeted modification of intact proteins with reactive probes

This method makes use of reactive amino acid side chains that naturally exist in proteins, most commonly the ε-amino group of lysine and the β-thiol group of cysteine (Figure 1.10A). These side chains are modified by addition of a probe derivatized to

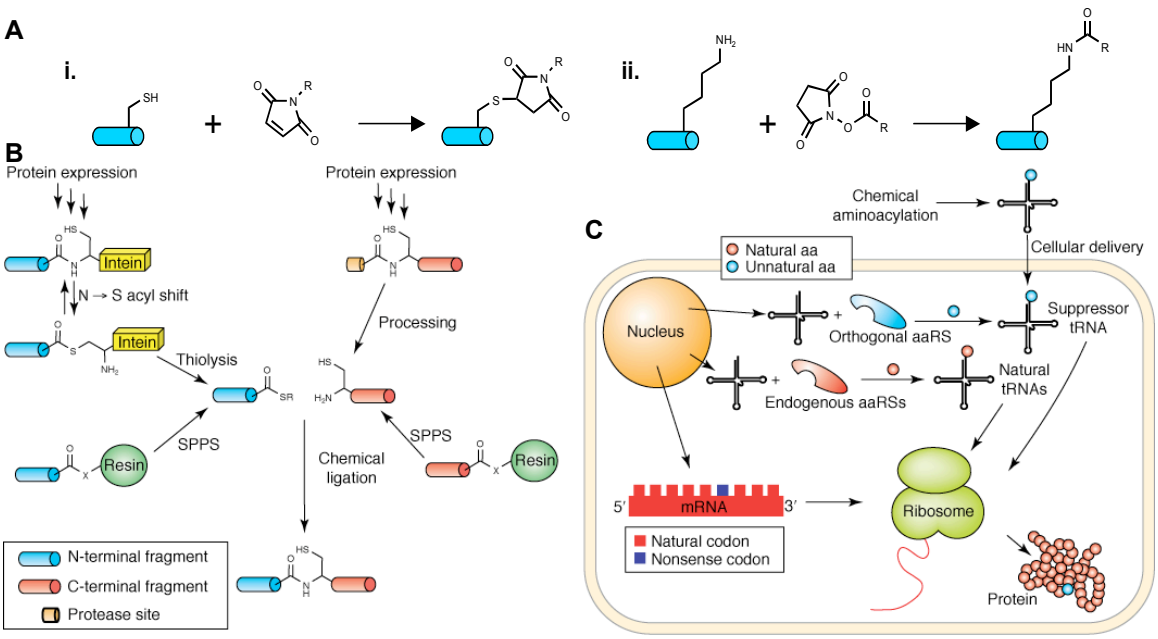


Figure 1.10. Protein semi-synthesis techniques. A. Targeted modification takes advantage of the reactivity of certain amino acid residues in proteins. Most commonly, cysteine and lysine are targeted. **i.** Cysteines react with maleimides to yield thioethers. **ii.** Lysines react with succinimidyl esters to yield amides. R = probe. **B.** In expressed protein ligation (EPL), two protein fragments of synthetic or recombinant origin are joined by a native peptide bond. One fragment contains an N-terminal cysteine which reacts with a C-terminal thioester of another fragment via native chemical ligation. Recombinant α-thioester proteins are produced by expressing the protein fused to a mutant intein, resulting in an intermediate thioester that can be intercepted by thiols. Recombinant N-terminal cysteine proteins are usually produced by proteolytic processing of a fusion protein. The corresponding synthetic fragments are readily prepared by solid phase peptide synthesis (SPPS) and other synthetic methods. **C.** In the nonsense suppression technique, an unnatural amino acid is inserted site-specifically into a protein of interest by infiltration of the protein biosynthetic pathway. A nonsense codon in the mRNA encoding the protein of interest is recognized by a suppressor tRNA charged with an unnatural amino acid. The suppressor tRNA can be charged in one of two ways. In the chemical approach, the tRNA is aminoacylated *in vitro* and subsequently delivered to cells. In the biosynthetic approach, an orthogonal aminoacyl-tRNA synthetase (aaRS) is expressed in cells that charges only a mutant suppressor tRNA.

contain an appropriately reactive group, such as an NHS-ester (to label lysine) or maleimide (to label cysteine).

Targeted modification has a rich history owing to its technical simplicity and the wide range of appropriately derivatized commercially-available probes. One drawback of targeted modification is that all residues of a protein that react with a particular probe are subject to being labeled. For example, if a protein contains 5 cysteines and a maleimide-containing probe is added to it, then all 5 residues are likely to be labeled. In some cases greater specificity has been achieved by limiting the concentration of probe relative to protein. Specificity can also be achieved by interfacing genetics with targeted modification. For example, if all the cysteines in a protein are eliminated except for the cysteine at the desired labeling site and the protein remains functional, then that cysteine can be selectively labeled (Zou et al., 2002). Finally, specificity can also be achieved in some cases by an elegant strategy using a bifunctional molecule consisting of a reactive probe fused to a homing moiety. For example, an active site cysteine of an enzyme can be selectively labeled by using a thiol-reactive moiety fused to an active site-directed substrate (Curley and Lawrence, 1998). It is noted that this mode of achieving specificity could also be used as an additional strategy to generate activity based probes of protein kinases.

1.2.8.1.2. Expressed protein ligation

Expressed protein ligation (EPL) elaborates upon the immensely successful native chemical ligation method whereby two synthetic peptides are ligated together by a peptide bond (Dawson et al., 1994). By taking advantage of the biosynthesis of α-

thioester and N-terminal Cys protein fragments, EPL readily allows for the addition of unnatural functionality to a recombinant protein framework, extending the ligation concept to proteins of all sizes (Figure 1.10B) (Muir, 2003). Methods to produce recombinant proteins bearing these functional groups have been developed. Erlanson, Verdine, and co-workers first described a method to generate recombinant proteins with N-terminal cysteines by proteolytic removal of a biosynthetic N-terminal protecting group (Erlanson et al., 1996). Since then, many alternative strategies for the generation of N-terminal cysteine proteins have been developed, involving proteolysis of fusion proteins by inteins (Hackenberger et al., 2006; Mathys et al., 1999; Southworth et al., 1999), methionine aminopeptidase (Gentle et al., 2004; Iwai and Pluckthun, 1999), and leader peptidase (Hauser and Ryan, 2007; Muir, 2003). Muir, Cole, and co-workers first described a biosynthetic strategy to produce recombinant proteins bearing C-terminal α-thioesters and used these to prepare semi-synthetic proteins *via* EPL (Muir et al., 1998). The production of recombinant proteins with C-terminal thioesters is made possible by intercepting a thioester intermediate formed in the process of protein splicing (Noren et al., 2000) by an exogenous thiol. With the ready availability of recombinant N-terminal cysteine and C-terminal α-thioester proteins, it is now possible to generate full-length, folded proteins harboring multiple site-specific chemical modifications *via* EPL (Hahn and Muir, 2005; Muir, 2003).

1.2.8.1.3. Suppressor mutagenesis

Another means of introducing unnatural functionality into proteins is the nonsense suppression mutagenesis method pioneered by the Schultz laboratory (Figure 1.10C)

(Ellman et al., 1991). This technique has been used extensively *in vitro* and recent improvements to the strategy have enabled its use *in vivo* (Xie and Schultz, 2005). Nonsense suppression allows an unnatural amino acid to be incorporated site-specifically into a protein, provided that the amino acid can be accepted by the ribosome. The new functionality can be a property of the new amino acid itself or the amino acid can be further elaborated by bio-orthogonal derivatization with other molecules (Chin et al., 2002; Deiters et al., 2003). The latter should dramatically expand the scope of the approach by allowing bulky probes to be attached post-translationally.

1.2.8.2. Applications to phosphoproteins

1.2.8.2.1. Site-specific introduction of phosphorylated residues

Investigators requiring phosphorylated proteins have traditionally generated them by carrying out enzymatic phosphorylation reactions with the isolated protein of interest and the appropriate kinase. This method is not ideal due to the difficulty involved in generating active kinases, inability to control the stoichiometry of phosphorylation, and the inability to control the site of phosphorylation in proteins that are phosphorylated on multiple residues. EPL provides an attractive alternative since it enables site-specific installation of phosphates to generate homogenous, well-defined phosphoproteins. Indeed, the first published use of EPL was to generate phosphorylated Csk (Muir et al., 1998). Since then, several phosphorylated proteins have been generated by EPL, some of which will be discussed here. The focus of this discussion will be on proteins involved in TGFβ signaling due to their relevance to this thesis.

Huse et al. prepared a tetra-phosphorylated semi-synthetic type I TGFβ receptor (TβRI) using EPL (Huse et al., 2000; Huse et al., 2001). Prior to this study, it was well known that hyperphosphorylation of the GS (glycine/serine rich) region of the receptor leads to its activation (Wieser et al., 1995; Wrana et al., 1994). However, the physical basis for activation by phosphorylation in this system was not known. In principle, phosphorylation of TβRI could cause a structural change that results in repositioning of active site residues into a catalytically-competent conformation. Alternatively, the hyperphosphorylated region could act as a binding surface for the protein substrate Smad2/3. These alternatives could not be reconciled with standard mutagenesis techniques and required semi-synthesis of TβRI with defined hyperphosphorylation of its GS region. Huse et al. succeeded in the synthesis of this extremely challenging protein and went on to show that hyperphosphorylation of the GS region enables TβRI to bind Smad2 *en route* to phosphorylation (Huse et al., 2000; Huse et al., 2001). Furthermore, when phosphorylated, TβRI no longer binds FKBP12, an endogenous inhibitor of the kinase. Thus, hyperphosphorylation of the GS region of TβR1 converts it from a binding site for an inhibitor (FKBP12) to a docking site for its substrate (Smad2/3).

The protein Smad2 transduces signals received at the cell membrane to the nucleus upon TGFβ binding to its receptor (Massague et al., 2005). In the nucleus, Smad2 acts as a transcription factor capable of modulating the expression of hundreds of genes. Smad2 is activated for this role by phosphorylation of two serine residues near its C-terminus. Wu et al. used EPL to prepare the MH2 domain of Smad2 with phosposerine at positions 465 and 467 (Wu et al., 2001). The crystal structure of a homotrimer of this protein revealed the mechanism of intermolecular phosphoserine binding, thereby

formally defining the MH2 domain as a modular phosphoserine binding domain, analogous to the SH2 domain for phosphotyrosine. Later work by Chacko et al. applied a similar strategy to the Smad2/Smad4 heterotrimeric complex, allowing them to determine the crystal structure of a heterotrimer composed of two doubly-phosphorylated Smad2 molecules and one molecule of Smad4 (Chacko et al., 2004). This study revealed the molecular basis of non-phosphorylated Smad4 participation in this trimer. More recently, Hill and colleagues used phosphorylated Smad2 prepared by EPL to show that Smad2-dependent transcription takes place exclusively on chromatin templates and cannot take place on naked DNA (Ross et al., 2006). This is in contrast to most other transcription factors that can drive transcription from both naked DNA and DNA packaged in chromatin. EPL can also be used to generate phosphorylated protein standards used in the determination of phosphorylation stoichiometry. This has been accomplished for Smad2 phosphorylation and will be described in section **4.2** of this thesis.

1.2.8.2.2. Introduction of stable phosphoaminoacid analogues

In certain circumstances, phosphorylated molecules are not suitable experimental reagents to determine the function of a specific phosphorylation event. For example, a phosphoprotein will likely not remain phosphorylated throughout the entire time course of an experiment *in vivo* due to the action of endogenous phosphatases. To circumvent this problem, biological investigations of serine and threonine phosphorylation have frequently used site-directed mutagenesis to create aspartic or glutamic acid at the phosphorylation sites. These mutated amino acids mimic phosphoserine and phosphothreonine by virtue of their carboxylate-containing side chains. In addition these

residues are not subject to the action of phosphatases. However, these are imperfect mimics and furthermore, there is no genetically-encodable mimic of tyrosine phosphorylation.

This has inspired investigators to use EPL to install amino acids at phosphorylation sites that closely resemble phosphoserine, phosphothreonine, and phosphotyrosine. The class of phosphorylation isosteres that are used most frequently for this purpose are phosphonates (Schwarzer and Cole, 2005). In phosphonates, the bridging oxygen connecting the amino acid side chain to the phosphoryl group is replaced by a methylene or fluorinated-methylene group. These phosphonates closely mimic the natural phosphates, both in size, charge, and polarity (Figure 1.11). Phosphonates have also been used extensively in medicinal chemistry to generate non-hydrolyzable phosphatase inhibitors and phosphoamino acid binding domain ligands (Burke, 2006).

The first application of a phosphonylated protein generated by EPL comes from the work of Cole and colleagues in their investigation of the cellular function of phosphorylation of the protein tyrosine phosphatase SHP-2 (Lu et al., 2001). Semi-synthetic SHP-2 was constructed with the non-hydrolyzable phospho-Tyr (pTyr) mimetic phosphonomethylenephenylalanine (Pmp, **29**) in place of a naturally occurring pTyr residue (Figure 1.11). Microinjection of the protein into cells activated the MAP kinase pathway, defining a role for this phosphorylation event in cellular regulation. Moreover, this is the first investigation to use a protein generated by EPL inside living cells, a theme that has been adopted by other groups since publication of this pioneering study. These investigators went on to install difluoromethylenephenylalanine (F_2Pmp, **30**) in Shp-2 (Figure 1.11) (Lu et al., 2003). This protein, which contains a difluorophosphonate, was

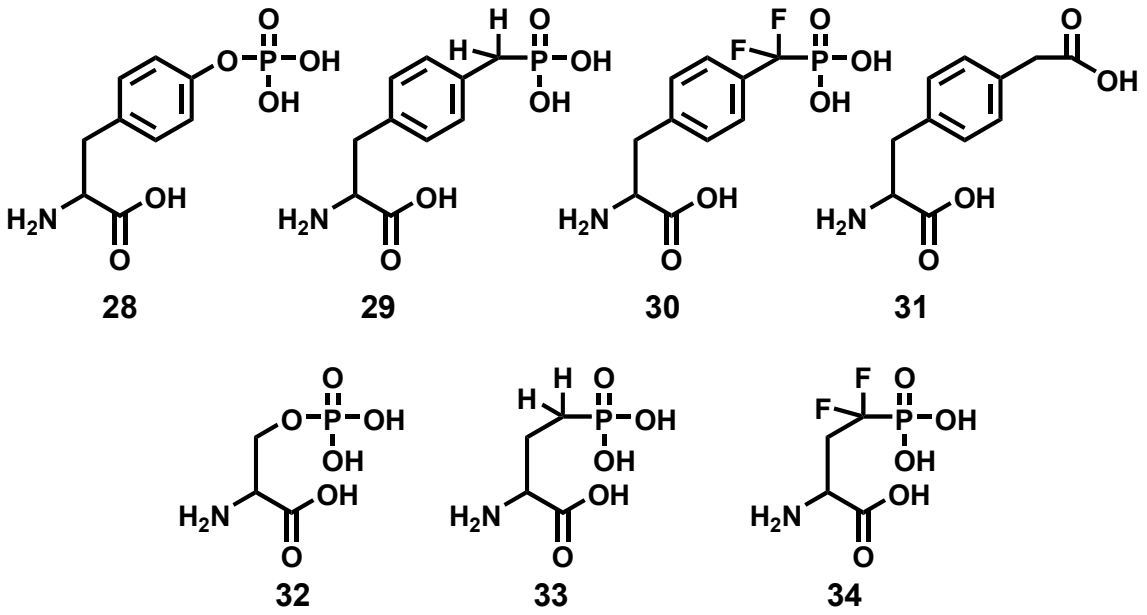


Figure 1.11. Phosphoserine, phosphotyrosine, and their unnatural non-hydrolyzable analogues. The number after each abbreviation is the second pK_a of the phosphate or phosphonate. **28** = phosphotyrosine (pTyr, 5.7[1]), **29** = phosphonomethylenephenylalanine (Pmp, 7.1[1]), **30** = difluorophosphonomethylenephenylalanine (F_2Pmp, 5.1[1,2]), **31** = *p*-carboxymethyl-phenylalanine (*p*CMF), **32** = phosphoserine (pSer, 5.7[3]), **33** = phosphonomethylenealanine (Pma, 7.1[3]), **34** = difluorophosphonomethylenealanine (Pfa, 5.1[3]).

---

[1] Reference: Domchek, et al., 1992

[2] Reference: Smyth, et al., 1992

[3] Reference: Zheng, et al., 2005

~50% more active than the corresponding phosphonate-containing protein. This indicates that F_2Pmp is a better mimic of pTyr as compared to Pmp. This is expected since the difluorophosphonate (F_2Pmp) and phosphate (pTyr) are both completely ionized at physiological pH (carrying a net charge of -2), whereas the phosphonate (Pmp) is only ~50% ionized and therefore carries a net charge of -1.5 (Figure 1.11) (Domchek et al., 1992; Smyth, 1992).

In another study, Zheng et al. replaced the naturally occurring phosphothreonine at residue 31 in the melatonin rhythm enzyme AANAT with phosphonomethylenealanine (Pma, **33**, Figure 1.11), a non-hydrolyzable pSer/pThr mimetic (Zheng et al., 2003). Microinjection of this protein provided direct evidence that phosphorylation results in increased stability of AANAT, most likely *via* a direct interaction with the protein 14-3-3. Similar results were found when another phosphorylation site of AANAT (serine-205) was converted to difluorophosphonomethylenealanine (Pfa, **34**, Figure 1.11) (Zheng et al., 2005). This suggests that bivalent binding of 14-3-3 to AANAT at sites nucleated by phosphorylation of serine-205 and threonine-31 may lead to synergistic stabilization of the enzyme.

Recently, Schultz and colleagues used nonsense suppression to install the phosphotyrosine analogue *p*-carboxymethyl-L-phenylalanine (*p*CMF, **31**) into Stat1 at position 701, which is a natural site of tyrosine phosphorylation (Figure 1.11) (Xie et al., 2007). Stat1 with *p*CMF at position 701 bound to a DNA oligonucleotide containing a Stat1-responsive element, providing experimental verification for the ability of *p*CMF to mimic phosphotyrosine in intact proteins. Sites of tyrosine phosphorylation that are not within close proximity of the N- or C-termini can readily be accessed and no additional

synthesis is required. However, *p*CMF is likely to be an inferior analogue of phosphotyrosine compared to Pmp and F_2Pmp, both of which have been successfully installed into proteins by EPL. In future applications, when choosing which analogue to utilize, one will have to consider this balance of site accessibility and degree of phospho-mimicry. This example highlights the complimentarity of these two approaches to unnatural amino acid mutagenesis of proteins.

1.2.8.2.3. Introduction of caged phosphoaminoacids and analogues into peptides and proteins

A 'caged' molecule is one that has been modified, usually covalently, in such a way that it is rendered inactive, but is readily re-activated upon irradiation with light of the appropriate wavelength and intensity (Kaplan et al., 1978; Mayer and Heckel, 2006). The use of caged neurotransmitters and metal ions such as Ca^{2+} has led to the mapping of complex inter-neuronal connectivity, the elucidation of the roles of individual neurotransmitters at synapses, the determination of specific enzyme activity *in vivo,* and to the better understanding of many other processes (Nerbonne, 1996).

Employing this strategy of molecular photo-control in the study of peptides and proteins has shed new light on the role of enzymes and signaling proteins in their native contexts since their activities are under spatial and temporal control (Lawrence, 2005). Furthermore, the dose of active peptide or protein is under exquisite control as either the intensity or duration of the applied light pulse is easily varied. Light in the far UV range has the additional advantages of being non-destructive and non-invasive.

It is generally simpler to synthesize a peptide than to carry out protein semi-synthesis, motivating several investigators to prepare caged, bioactive peptides

(Lawrence, 2005; Shigeri et al., 2001). Whereas uncaging of a protein is designed to activate it, uncaging of a peptide in many cases is used to inhibit a cellular protein. Robust methods for the synthesis of caged phosphopeptides have led investigators to use them in live cell experiments (Rothman et al., 2002, 2003). The key step in the synthesis of a caged phosphopeptide is the installation of the caged phosphate onto the amino acid of choice using a phosphoramidite that has a caging group on one of its oxygens. Nguyen et al. synthesized a caged phosphopeptide-inhibitor (**35**) of the cellular 14-3-3 family of proteins, which contain a modular phosphoserine/phosphothreonine binding domain (Figure 1.12) (Nguyen et al., 2004). Several distinct genes give rise to 14-3-3 proteins that have overlapping function. Therefore, ablation of 14-3-3 activity by knockout technology is intractable owing to the requirement for deletion of several loci. Since 14-3-3 proteins carry out their function by binding other proteins, Nguyen et al. designed their caged phosphopeptide inhibitor to disrupt these interactions by competitive binding upon uncaging. For experiments with live cells, the caged peptide inhibitor was delivered by a disulfide-linked peptide from the Antennaepedia protein, which functions as a cell penetrating peptide. These experiments showed that 14-3-3 proteins regulate many phases and checkpoints of the cell cycle. Caged phosphotyrosine peptides have also been validated in live-cell experiments addressing the role of tyrosine phosphorylation of the protein FAK in cell migration (Figure 1.12, compound **36**) (Humphrey et al., 2005).

Bastiaens and co-workers took advantage of a caged PTP1B substrate to probe the spatial regulation of PTP1B activity in live cells (Figure 1.12, compound **37**) (Yudushkin et al., 2007). By labeling PTP1B with a fluorophore capable of undergoing FRET with a fluorophore on the peptide substrate, the fraction of PTP1B bound to substrate could be

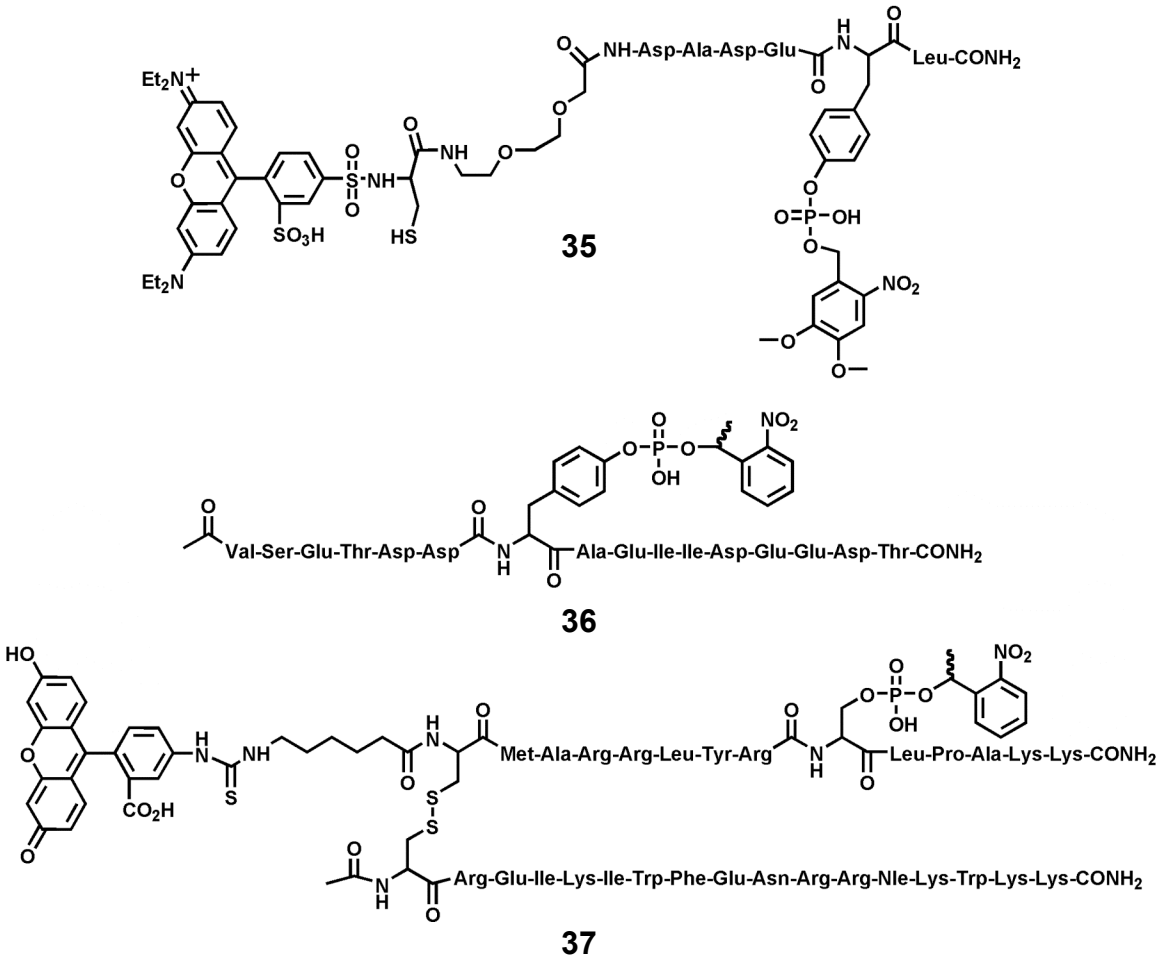


Figure 1.12. Caged phosphopeptides. 35 = caged phosphopeptide-inhibitor of 14-3-3 proteins, **36** = caged phosphopeptide-inhibitor of FAK cellular activity, **37** = caged substrate of PTP1B.

determined. Using the caged substrate, the investigators discovered that the k_{cat} of PTP1B in cells is lower than *in vitro*. Further experiments indicated that PTP1B activity was lower at the cell periphery than the peri-nuclear region of the cell. The elucidation of the biochemical mechanism of spatial regulation will have to await further studies.

Targeted modification has been used to create caged proteins (Curley and Lawrence, 1999; Marriott et al., 1998). However, this method generally leads to the installation of multiple caging groups at all sites of the protein capable of reacting with the caging group. This can be problematic because solubility or folding of the protein of interest could be lost when decorated with multiple caging groups. Also, a protein with many caging groups may require application of significantly more UV light over more time for complete uncaging, thereby limiting the temporal resolution of the approach.

In the past decade, researchers have attempted to solve these problems by devising strategies that furnish proteins with caging groups at specified locations, which will be referred to as site-specific caging (Lawrence, 2005). In developing methods to cage proteins site-specifically, it is desirable to create a strategy that will be generally applicable for use with many different protein families. As nature frequently uses post-translational modifications to augment protein activity, a caging strategy targeting one of these events would be useful in caging any protein regulated by the given modification. Phosphorylation is an extremely well-defined and ubiquitous post-translational modification that frequently acts as a molecular switch to regulate protein activity (Hunter, 2000). Therefore, a robust method of caging proteins on phosphate will be applicable to the caging and study of many proteins.

Several groups have reported methods to cage proteins regulated by phosphorylation. Bayley and co-workers prepared protein kinase A (PKA) caged on an activating phospho-threonine residue by targeted chemical modification of an enzymatically-generated thiophospho-Thr (Zou et al., 2002). Importantly, the native cysteines of PKA were mutated to alanine so that the only site of modification on the protein would be the thiophosphate. Complementary to this approach, Lawrence and colleagues designed a strategy for caging proteins that are inactivated by Ser/Thr phosphorylation events (Ghosh et al., 2002). This involves site-directed mutagenesis of the phosphorylatable Ser/Thr to Cys, which is then reacted with a caging group containing a carboxylate to mimic phosphate. Using this method, Ghosh et al. prepared a caged version of cofilin, a protein involved in cell motility. Importantly, the native cysteines of cofilin did not react with the caging group, presumably because they were protected in disulfides. While the enzymatic activities of cofilin were already characterized, it was unclear how activation of the protein affects motility *in vivo*. Spatially-localized activation of cofilin resulted in directional migration, thus implicating it as a component of the cellular "steering wheel" (Ghosh et al., 2004).

We have described two strategies for the semi-synthesis by EPL of caged Smad2, an intracellular mediator of TGFβ signaling (Hahn and Muir, 2004; Pellois et al., 2004). Caged and non-fluorescent Smad2 was prepared in such a way that UV irradiation simultaneously triggers protein activation and fluorescence (Pellois et al., 2004). In another approach, Smad2 was caged directly on two activating pSer residues (Hahn and Muir, 2004). This strategy is well suited to the caging of any protein activated by phosphorylation, thus providing an entry point to the study of many signaling pathways.

This work forms the majority of this thesis and will be described in detail in later chapters. Vogel and Imperiali have also succeeded in preparing a caged phosphoprotein by EPL (Vogel and Imperiali, 2007). They chose the phosphotyrosine-containing protein paxillin as their target. Semi-synthetic paxillin variants were prepared with tyrosine, phosphotyrosine, or caged phosphotyrosine at position 31. Synthetic C-terminal α–thioester peptides corresponding to the N-terminus of paxillin (residues 2-36) were ligated to a recombinant fragment of paxillin (residues 38-557) containing a N-terminal cysteine. The caged phosphotyrosine containing protein could be uncaged with UV light at 365 nm as judged by western blotting with a phospho-specific paxillin antibody. This study verifies that EPL is generally applicable for caging phosphoproteins since phosphotyrosine was successfully caged in an intact protein.

Rothman et al. used nonsense suppression to generate caged phosphoproteins by *in vitro* translation in rabbit reticulocyte lysates (Rothman et al., 2005a). They found that the suppression efficiencies of caged serine, tyrosine, and threonine were lower than the parent unphosphorylated and uncaged amino acids. Nevertheless, the authors succeeded in the production of VASP containing caged phosphoserine at position 153. Following UV irradiation, the protein was uncaged and, therefore, competent for phosphorylation by PKA.

1.3. Summary

This introduction provides a framework for the use of chemistry to understand various aspects of protein phosphorylation. Many key concepts in the history of protein phosphorylation required tools that originated in chemistry. The examples described

above of chemistry-inspired technologies and their use for the dissection of mechanism and consequences of protein phosphorylation are of undoubted utility for future investigations. Furthermore, it is expected that several new innovative techniques for studying protein phosphorylation will emerge from the chemical biology community.

The remaining chapters of this thesis will describe several uses of protein semi-synthesis to the study of phosphorylation of the cellular signaling protein Smad2.

Chapter 2: A general strategy for the preparation of caged phosphoproteins and its application to Smad2[1]

2.1. Background

The ability to activate proteins with spatial and temporal control inside live cells allows for quantitative kinetic measurements of protein function to be made in a biologically relevant context. Proteins that contain photo-labile protecting groups appended to functionalities required for biological activity are light-activatable and provide a means to enable such analyses (Figure 2.1A). An increasing number of reports describing construction of these reagents, known as caged proteins, have appeared in the recent literature (Curley and Lawrence, 1999; Lawrence, 2005). Herein, a semi-synthetic route to the preparation of caged phosphoproteins is described. This strategy has been applied to the cellular signaling protein Smad2.

Various caging groups have been used for the construction of caged compounds. The first caged compound described, caged ATP, contains the photolabile *o*-nitrobenzyl (*o*-NB, **31**, Figure 2.1B) group attached to the γ-phosphate of ATP. Upon UV irradiation, the nitro group rearranges to form an *aci*-nitro intermediate that cyclizes and then decomposes to release ATP and a nitrosobenzaldehyde by-product (Figure 2.1C). Most other caged compounds described to date also contain the *o*-NB group or one of its

[1] The work described in this chapter resulted in the following publication: Hahn, M.E., and Muir, T.W., Photocontrol of Smad2, a multiphosphorylated cell-signaling protein, through caging of activating phosphoserines. *Angewandte Chemie International Edition*, **43**, 5800-5803 (2004).

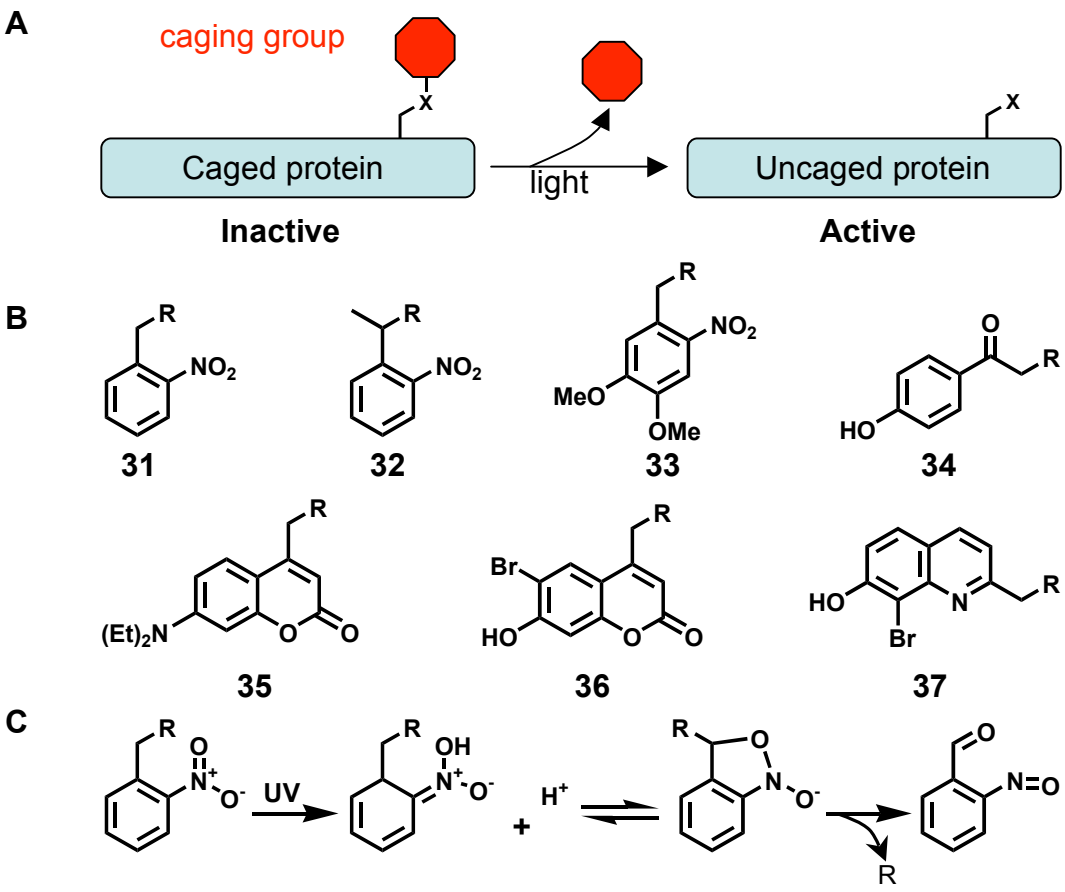


Figure 2.1. Caged proteins and caging groups. A. A caged protein contains a caging group (red) that blocks a residue (X) of the protein that is required for activity, leading to protein inactivation. Upon irradiation of light of the appropriate wavelength, the caging group is removed and protein activity is restored. **B.** Examples of caging groups in current use. R indicates the compound that is being caged. **31** = *o*-nitrobenzyl group, **32** = *o*-nitrophenylethyl group, **33** = 3,4-dimethoxy-6-nitrobenzyl group, **34** = 4-hydroxyphenacyl group, **35** = 7-N,N-diethyl aminocoumarin, **36** = 6-bromo-7-hydroxycoumarin-4-ylmethyl group, **37** = 8-Bromo-7-hydroxyquinoline group. **C.** Mechanism of *o*-nitrobenzyl group photolysis. Absorption of UV light causes conversion of the *o*-nitrobenzyl group into an *aci*-nitro intermediate, which cyclizes and decomposes to release R and a nitrosobenzaldehyde by-product.

improved derivatives. Alkylation of the exocyclic methylene of the *o*-NB group increases the rate of photolysis. For example, methylation at this position generates the *o*-nitrophenylethyl (*o*-NPE, **32**, Figure 2.1B) caging group that can be photolyzed faster than the parent *o*-NB group (Holmes, 1997). Furthermore, the ketone-containing photolysis by-product of the *o*-NPE group is less reactive and therefore less prone to causing damage in cells as compared to the aldeyde-containing photolysis product of the *o*-NB group. Installation of methoxy substituents on the ring of either of these caging groups (**33,** Figure 2.1B) results in red-shift of the absorbance spectrum enabling efficient photocleavage at less damaging wavelengths.

Many other caging groups have been developed that display different properties useful for various applications. The 4-hydroxyphenacyl group (**34**, Figure 2.1B) displays superior photolysis kinetics compared to *o*-NB based caging groups and has been used to cage a thiophospho-protein and a protein phosphatase on sulfur (Arabaci et al., 1999; Park and Givens, 1997; Zou et al., 2002). The coumarin-based DECM caging group **35** (Figure 2.1B) can be photolyzed with non-toxic light in the visible range (Shembekar et al., 2005). This is a useful feature for studying processes in exceptionally UV sensitive cells. Other coumarin- and quinoline-based caging groups (**36** and **37**, Figure 2.1B) can be photolyzed with two-photon techniques enabling uncaging studies in deep tissue (Furuta et al., 2004; Furuta et al., 1999; Goard et al., 2005; Zhu et al., 2006).

2.2. Smad2: a target for phosphoprotein caging

Smad2 is a key element of the intracellular response to cytokines of the TGFβ superfamily, which are involved in a myriad of normal and disease processes including

development, tissue homeostasis, and cancer (Massague et al., 2005; Siegel and Massague, 2003). Binding of TGFβ to its cognate receptor complex results in phosphorylation of the last two serines of the Smad2 C-terminal sequence CSSMS (residues 463-467; Figure 2.2) (Macias-Silva et al., 1996; Wu et al., 2001). These phosphorylation events activate Smad2, rendering the protein competent to both homo-trimerize and interact with Smad4, a binding partner required for many downstream functions (Lagna et al., 1996). Activated Smad2 accumulates in the nucleus, where it regulates transcriptional programs by interacting with a host of other proteins and target promoters.

These differential protein-protein interactions and the localization of Smad2 provide a basis for understanding how this molecule functions in a cell. However, these descriptions are static and do not adequately describe the dynamics underlying these signaling events. Little is known about where many of these protein-protein interactions are initiated and for how long they exist. The kinetics of nuclear import and export as well as the importance of signal strength, duration, and localization are all poorly understood. Therefore Smad2 was targeted for caging because the ability to activate this protein with temporal and spatial control allows one to directly address some of these fundamental issues. Furthermore, the caging strategy was designed to be generally applicable to other proteins modulated by phosphorylation.

2.3. General strategy for photocaging Smad2 on phosphoserine residues

The caging strategy discussed here takes advantage of protein phosphorylation, the post-translational modification most often used to regulate protein activity (Hunter,

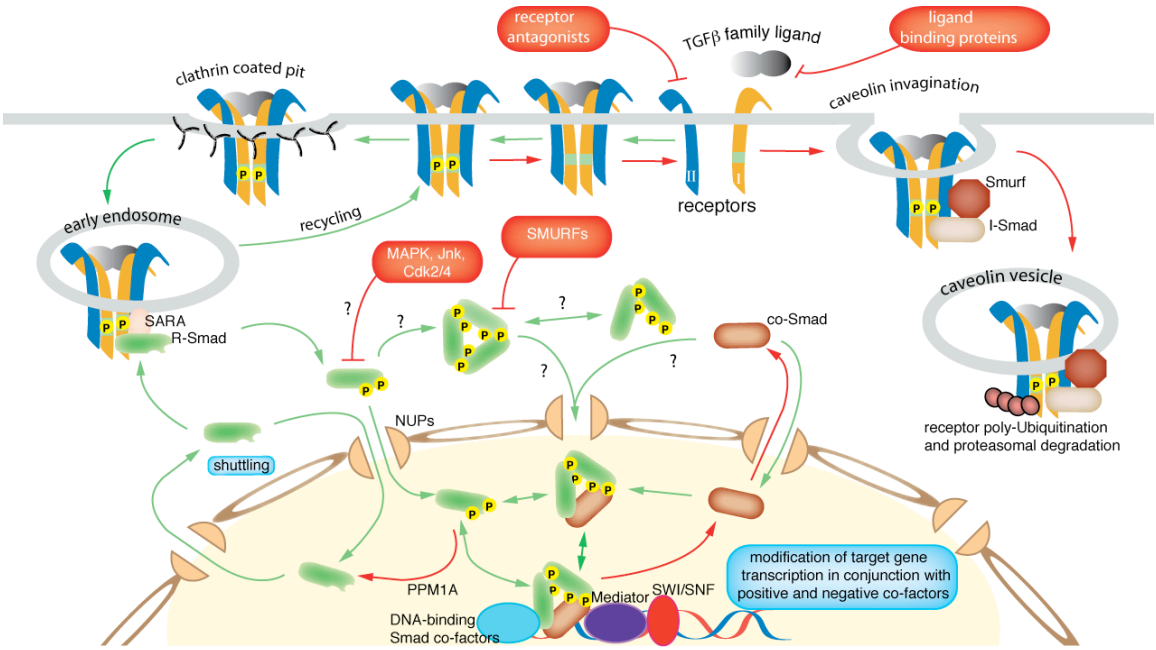


Figure 2.2. The intracellular response of Smads to TGFβ Signals. A TGFβ family ligand engages the TGFβ receptor to induce an active signaling complex of type I and type II receptors. This complex can be degraded by the proteasome in a ubiquitin-dependent manner after endocytosis through caveolin vesicles. The active receptor complex can also be internalized via clathrin-dependent endocytosis to early endosomes, where it is active. R-Smads can be recruited to the active receptor by SARA where they are activated by phosphorylation. Active R-Smads can interact with other active R-Smads and the co-Smad, Smad4. These R-Smad containing complexes accumulate in the nucleus where they associate with other transcriptional co-factors to modulate the expression of TGFβ-responsive genes. General components of the transcription apparatus (e.g. the Mediator complex) and chromatin remodeling factors (e.g. the SWI/SNF complex) help mediate these transcriptional changes. Activated Smad2 can be dephosphorylated in the nucleus by PPM1A to terminate the signal. Smads can be degraded following poly-ubiquitination by ubiqutin ligases such as SMURFs. Green and red arrows indicate positive and negative regulation of the pathway, respectively. Question marks indicate areas under active investigation in the last few years.

2000). Much recent effort has been directed at the preparation of caged analogues of phosphopeptides and phosphoproteins (Lawrence, 2005). Expressed protein ligation (EPL) was used as the center point of a semi-synthetic scheme for the preparation of Smad2 caged on activating phosphorylated residues (**40**; Figure 2.3) (Muir, 2003). This approach offers several advantages including the ability to produce caged proteins of any size in quantities sufficient for various biological applications without the need for mutagenesis (Muir, 2003). Additionally, EPL readily allows for the installation of multiple caged phosphates in a homogenous manner (Huse et al., 2000). This characteristic is of significant importance since many proteins, including Smad2, are controlled by multi-site phosphorylation (Cohen, 2000).

Here, we chose to work with the MH2 domain of Smad2 (residues 241-467; MW ~25 kDa), since its molecular weight enables precise characterization by chromatography and electrospray mass spectrometry (ESI-MS). The MH2 domain mediates many of the functions of Smad2, including receptor recognition, homo- and hetero-oligomerization, and nuclear import (Shi and Massague, 2003).

2.4. Semi-synthesis of Smad2-MH2 caged on phosphoserine residues

Preparation of the caged phosphoprotein commenced with the synthesis of the corresponding doubly-caged phosphopeptide **39** using the 9-fluorenylmethoxycarbonyl (Fmoc) strategy (Figure 2.4). Key to the synthesis of **39** was: (1) orthogonal trityl (Trt) protection of the side-chains of the two serines to be phosphorylated and (2) incorporation of the N-terminal cysteine (Cys) required for EPL as *tert*-butoxycarbonyl-1,3-thiazolidine-4-carboxylic acid (Boc-Thz). The latter allowed for the thiol of Cys to be

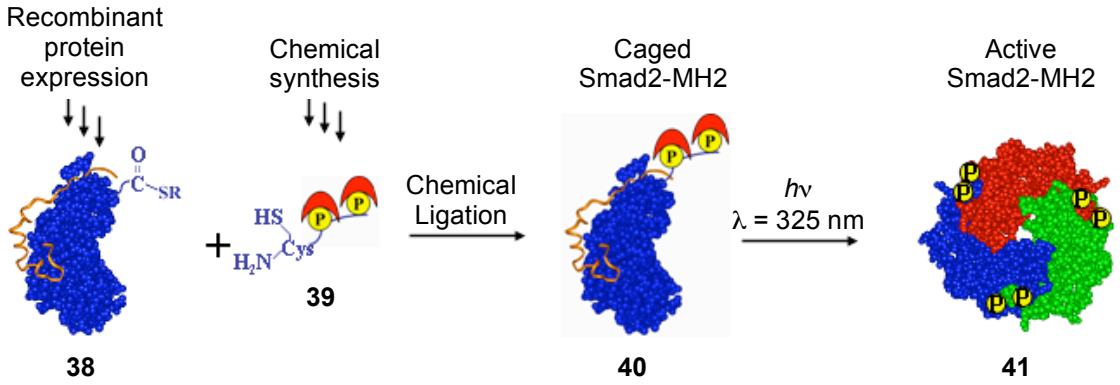


Figure 2.3. Semi-synthesis of caged Smad2-MH2. Expressed protein ligation was used to ligate a recombinant Smad2-MH2-α-thioester/SARA-SBD protein complex (**38**) to the doubly-caged phospho-peptide **39**, giving the caged Smad2-MH2/SARA-SBD hetero-dimer (**40**). Caged Smad2-MH2 is activated by exposure to UV light and subsequently releases SARA-SBD and forms a homo-trimer (**41**). Smad2-MH2 is shown in globular form, SARA-SBD is shown in orange, phosphorylated residues are symbolized by yellow circles, and caging groups are symbolized by red crescents.

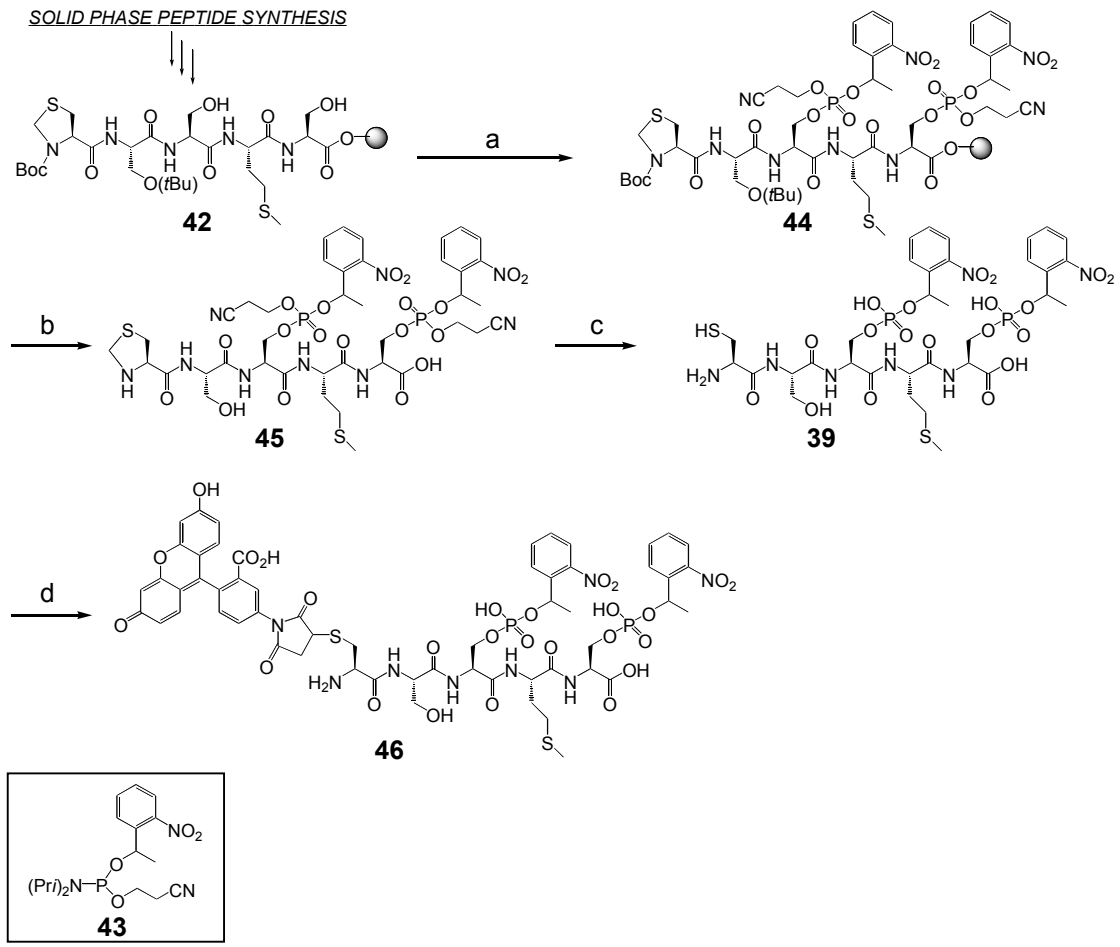


Figure 2.4. Synthesis of doubly-caged phospho-peptide 39. a) (i). *O*-1-(2-nitrophenyl) ethyl-*O*'-β-cyanoethyl-*N*,*N*-diisopropylphosphoramidite,4,5-dicyano-imidazole (**43**, shown in caption below scheme), DMF (anhydrous); (ii). 1M *t*BuOOH, DCM (anhydrous); b) 92.5% TFA, 2.5% EDT, 2.5% TIS, 2.5% H_2O; c) 1% DBU, DMF then 0.5 M $MeONH_2 \cdot HCl$, H_2O; d) fluorescein-5-maleimide.

protected during the critical phosphorylation step and provided a convenient method for Cys deprotection following cleavage of the peptide from the solid support (Villain et al., 2001). Following chain assembly and selective unmasking of the two serines, the resulting peptidyl-resin **42** was dried extensively and reacted with *O*-1-(2-nitrophenyl)ethyl-*O′*-β-cyanoethyl-*N,N*-diisopropylphosphoramidite (**43**) (Rothman et al., 2002), which contains the *o*-NPE group. This caging group was chosen since Rothman et al. had previously demonstrated its utility for the synthesis of caged phosphopeptides. *Tert*-butylhydroperoxide was then used to oxidize the intermediate phosphites to the desired phosphates, yielding **44**. Notably, the undesired oxidation of the thioether of methionine to the sulfoxide was largely avoided (<5%) by limiting the time of oxidation to 20 minutes. Attempted on-resin removal of the β-cyanoethyl protecting groups in the setting of two juxtaposed phosphates resulted in significant amounts of β-elimination of the protected phosphate moiety (Kuder et al., 2000). Interestingly, this side reaction was not found to occur when the deprotection step was carried out on **45** in solution following cleavage from the resin. This suggests that the elimination was facilitated by the C-terminal ester linkage between the peptide and the solid support. Smooth removal of the β-cyanoethyl groups was therefore carried out under optimized conditions in solution using the hindered amidine 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU). Methoxylamine was then added in situ to convert Thz to Cys. The crude product contained one major peak (~75% by reversed-phase high-performance liquid chromatography) (Figure 2.5A), which was subsequently purified to homogeneity to give the desired peptide caged on two phosphorylated serine residues (**39**) in ~10% yield (Figure 2.5B). The identity of the product was verified by tandem MS (Figure 2.6).

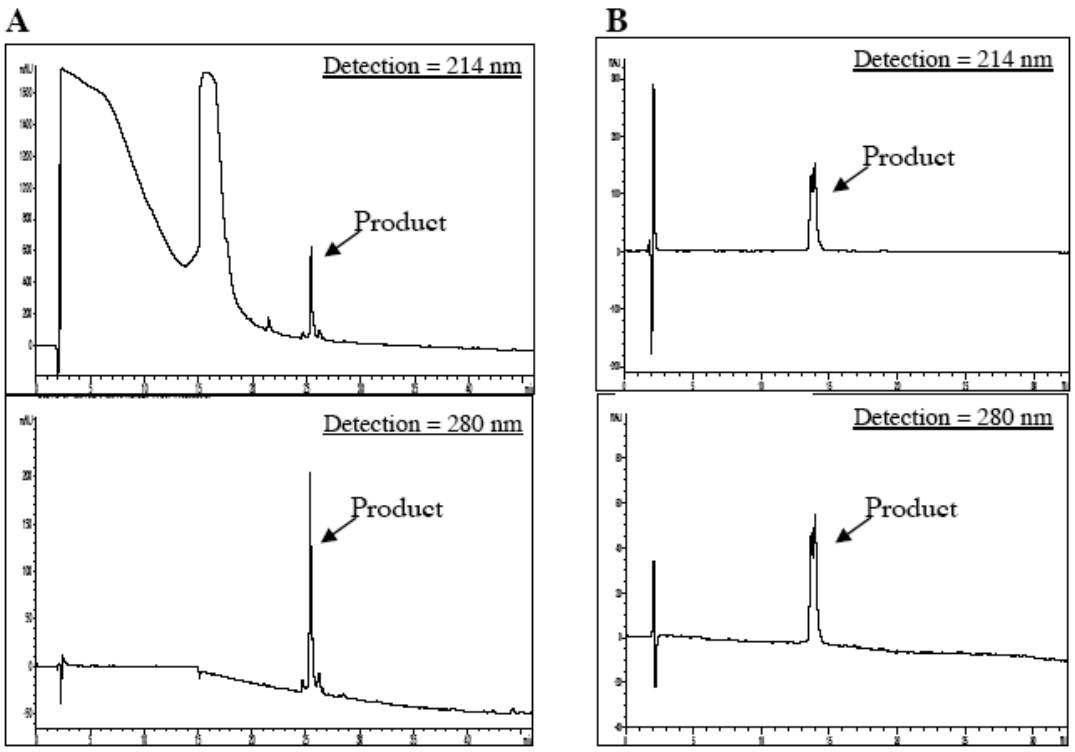
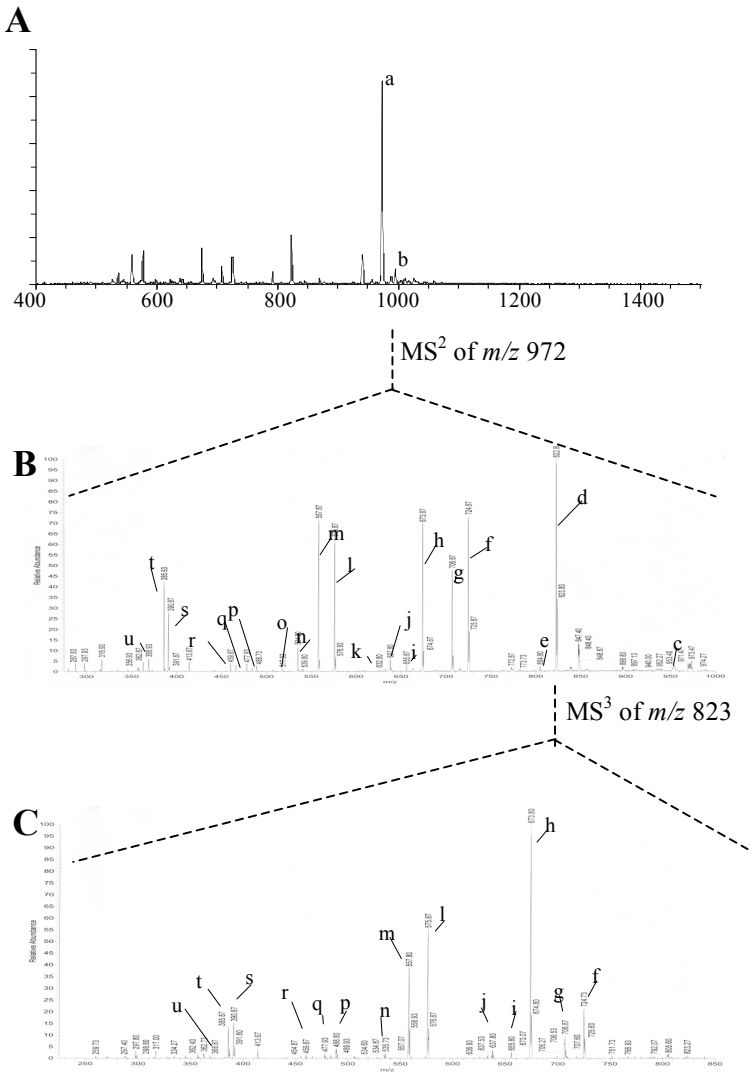


Figure 2.5. RP-HPLC characterization of 39. **A**. RP-HPLC analysis of crude synthetic material. A gradient of 0 – 73%B over 30 minutes was used with detection at 214 nm (top panel) and 280 nm (bottom panel). In the top panel, the large absorbance before 20 minutes is due to DMF from the reaction mixture, which does not absorb at 280 nm (bottom panel). **B**. RP-HPLC of purified synthetic material. A gradient of 20 – 30%B over 30 minutes was used with detection at 214 nm (top panel) and 280 nm (bottom panel). The pure product is separated on this shallow gradient due to the racemic nature of the benzylic carbon of the 2-nitrophenylethyl protecting groups.



code	*m/z*	description
a	972.0	**39** + H^+
b	994.0	**39** + Na^+
c	953.40	**39** – H_2O
d	822.80	**39** – NPE
e	804.80	**d** – H_2O
f	724.97	**d** – P
g	706.87	**f** – H_2O
h	673.87	**d** – NPE
i	655.87	**h** – H_2O
j	637.80	b_4
k	632.80	y_3 – NPE
l	575.87	**h** – P
m	557.97	**l** – H_2O
n	534.90	**k** – P
o	517.93	**n** – NH_2
p	488.73	b_4 – NPE
q	477.93	**l** – P
r	459.87	**q** – H_2O
s	390.87	**q** – P
t	385.93	**n** – NPE
u	368.93	**20** – NH_2

Figure 2.6. Representative mass spectra of 39. **A**. ESI-MS of **39** with a single quadruple mass detector. **B.** MS^2 of the parent ion of **39** at 972 Da was performed on an ion trap instrument. **C.** MS^3 of the fragment at *m/z* 823 was performed on an ion trap instrument. MS^3 of all major peaks was performed and spectra are consistent with the structure of **39**. Major fragments identified in the spectra above are listed in the table along with a description.

The caged peptide **39** was labeled with fluorescein-5-maleimide (thus generating peptide **46**) and subjected to low intensity UV irradiation (312 nm, 2 mW/cm^2) followed by RP-HPLC to determine the kinetics and quantum yield of uncaging (Figure 2.7A,B). Interestingly, both possible singly-caged peptides were observed in approximately equal amounts after irradiation for an intermediate length of time, indicating that the efficiency of photolysis was equivalent for both caging groups (Figure 2.7B,ii). Photolysis followed first-order kinetics with a rate constant of 4.9 x 10^{-3} s^{-1}, corresponding to a quantum yield of uncaging of 0.16 per caging group (Figure 2.8). Brief (<5 seconds) exposure of the peptide to the output of a He-Cd laser (325 nm, 4.74 W/cm^2) resulted in near quantitative conversion (>97%) to the uncaged peptide (Figure 2.7C).

A recombinantly expressed Smad2-MH2 domain (residues 241-462) bearing a C-terminal thioester (**38**) was prepared as previously described (Wu et al., 2001). This protein was expressed as a fusion with a modified GyrA intein and chitin binding domain. Following affinity purification of the protein over chitin beads, Smad2-MH2-α-thioester was released by incubation with the thiol 2-mercaptosulfonic acid (MESNa). A complex of Smad2-MH2-α-thioester with the minimal Smad binding domain of SARA (SARA-SBD, residues 665-721) was formed by incubation with excess SARA-SBD and purified by cation exchange chromatography. SARA-SBD binds the MH2 domain in a region known as the hydrophobic corridor and was used here to prevent precipitation of the MH2 domain during subsequent steps.

The resulting pure protein complex was concentrated to 0.25 mM and a 4-fold molar excess of the caged peptide **39** was added to initiate the ligation reaction (Figure 2.3). The reaction was monitored by RP-HPLC, ESI-MS, and SDS-PAGE (sodium

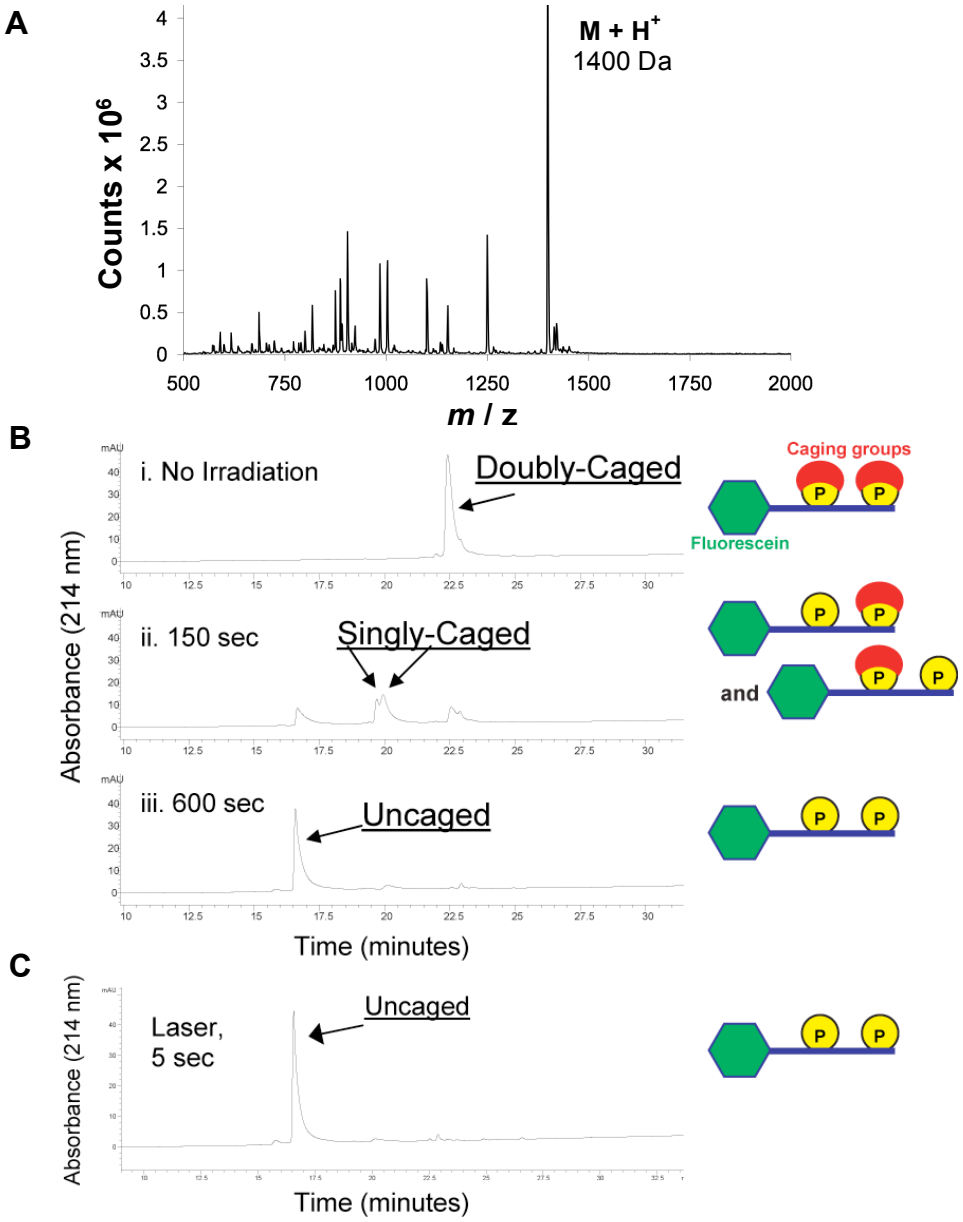


Figure 2.7. Representative RP-HPLC of 46 before and after photolysis. A. Mass spectrum of **46**. Expected M_r is 1400.2 Da. **B. 46** was irradiated for the times indicated with UV light (312 nm, 2 mW/cm^2) and analyzed by RP-HPLC (0 – 73%B). The eluting species were detected at 443 nm (the absorbance maximum of the fluorescein label in RP-HPLC solvents). In the 150 sec trace, the doublet at approximately 20 minutes is presumed to be due to the two singly-caged species, as both peaks have the same molecular weight. **C. 46** was irradiated with the 325 nm laser for 5 seconds and analyzed by RP-HPLC. This treatment led to the near quantitative (>97%) conversion of the caged peptide to the uncaged peptide.

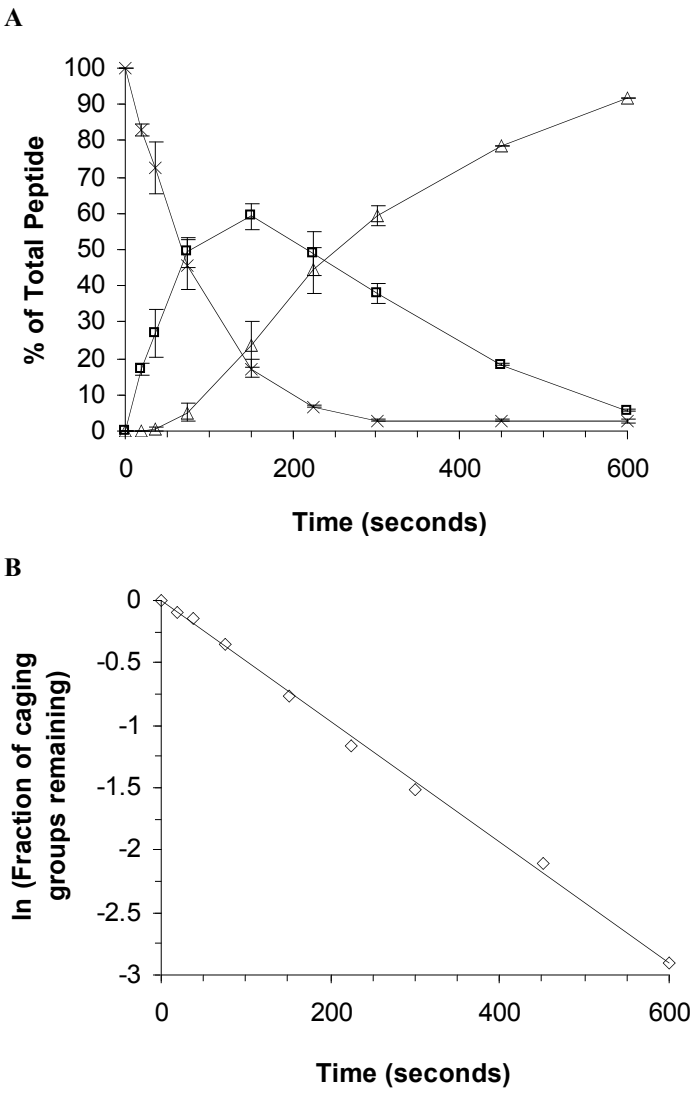


Figure 2.8. Photolysis kinetics of doubly-caged phospho-peptide labeled with flourescein-5-maleimide (46). A. A solution of **46** at 10 mM was irradiated with low intensity UV light (312 nm, 2 mW/cm^2) for the times indicated and subjected to RP-HPLC for quantitation of the doubly-caged (×), singly-caged (□), and uncaged forms (Δ). The percentage of each is plotted versus time of irradiation. **B.** The fraction of caging groups remaining was calculated and the natural logarithm at each time point plotted versus time of irradiation, along with a line of best fit, which yielded a first-order rate constant for photolysis of 4.9×10^{-3} and an R^2 of 0.99. For both **A** and **B**, the mean of two experiments is plotted and for **A** the standard deviation is represented by error bars.

dodecyl sulfate-polyacrylamide gel electrophoresis) and was complete after 12 hours (Figure 2.9A). The caged protein **40** was purified by preparative size exclusion chromatography (SEC) and its identity was confirmed by HPLC, SEC, and ESI-MS (Figure 2.9B-D).

2.5. Characterization of photocaged Smad2 and validation of the caging strategy

To be deemed effective, the caged protein should behave as if it was non-phosphorylated in the absence of UV light and should display all the properties of the active, doubly-phosphorylated Smad2-MH2 when uncaged by UV light. We therefore proceeded with studies designed to determine the oligomerization state of the caged protein before and after UV irradiation. Integration of the RP-HPLC elution peaks corresponding to Smad2-MH2 and SARA-SBD allows quantitation of the stoichiometry of Smad2-MH2 binding to SARA-SBD. A 1:1 hetero-complex between Smad2-MH2 and SARA-SBD would display a ~4:1 peak area ratio (detection at 214 nm) since Smad2-MH2 is approximately four times larger than SARA-SBD. The caged protein mimicked non-phosphorylated Smad2-MH2, as it bound SARA-SBD in a 1:1 molar ratio (Figure 2.9B). This hetero-dimeric arrangement of Smad2-MH2 and SARA-SBD was verified by SEC coupled with multi-angle laser light scattering (MALLS) detection at a loading concentration of 5 μM (Figure 2.10B) (Folta-Stogniew and Williams, 1999). At higher loading concentrations (25-50 μM), MALLS analysis indicated that the caged protein had a slight residual tendency to form homo-trimers (Figure 2.10A,B). Importantly, the tendency of the doubly caged protein to homo-oligomerize is concentration-dependent, such that at physiologically relevant concentrations (<5 μM), this behavior is no longer

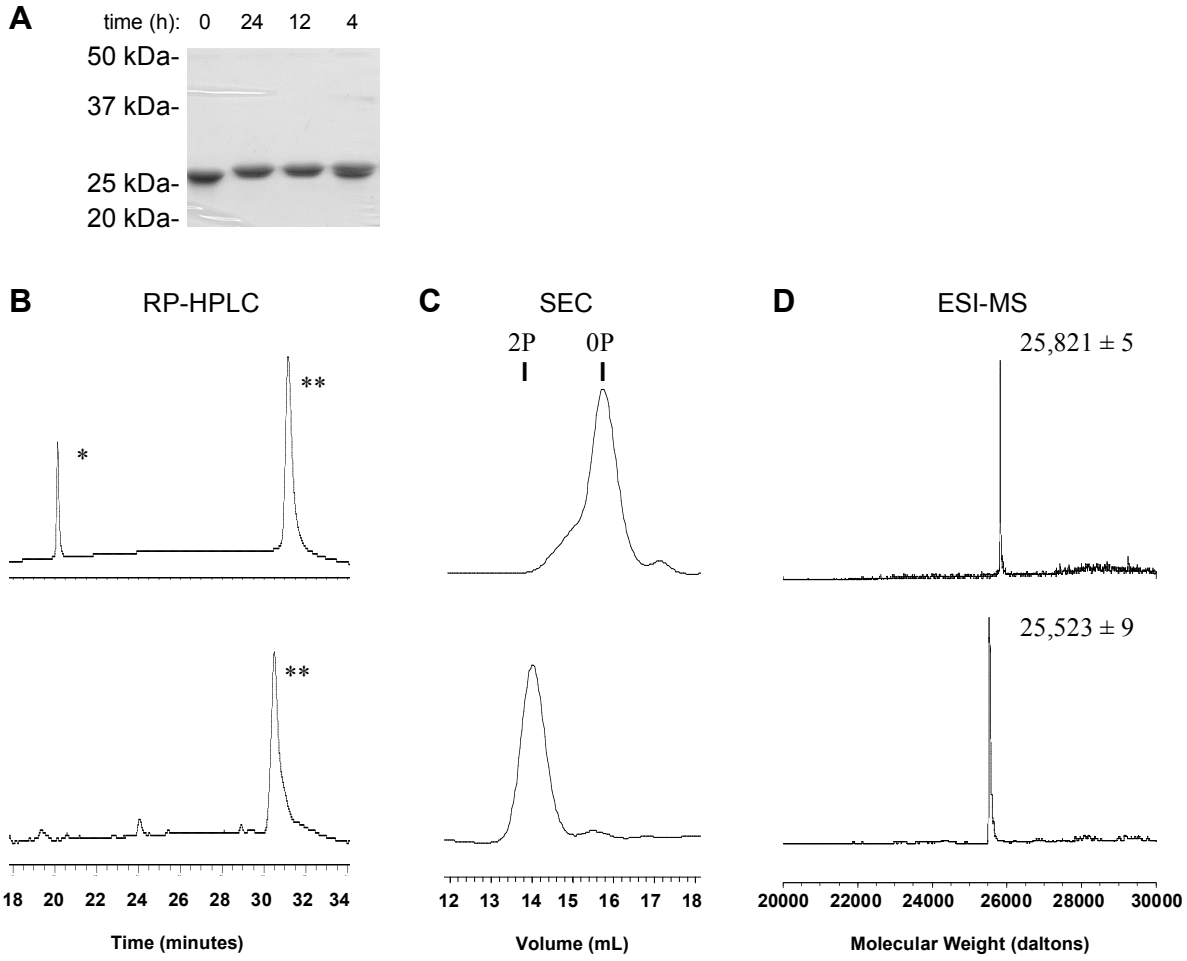


Figure 2.9. Characterization of caged Smad2-MH2 (40). A. SDS-PAGE of ligation timecourse. **B-D.** Caged Smad2-MH2 (top panels) was converted to uncaged Smad2-MH2 (**41**; bottom panels) by irradiation for 5 seconds with a He-Cd laser (325 nm, 4.74 W/cm^2). **B.** RP-HPLC of the caged Smad2-MH2(**)/SARA-SBD(*) complex. **C.** The homo-oligomeric status of caged and uncaged Smad2-MH2 (at 5 μM) was assessed by SEC with detection at 280 nm. The elution positions of non-phosphorylated Smad2-MH2 (0P) and doubly-phosphorylated Smad2-MH2 (2P) controls are indicated. **D.** Reconstructed molecular weight from ESI-MS indicates that the caged protein (expected MW = 25,818 Da) was assembled successfully. ESI-MS of the uncaged protein (expected MW = 25,519 Da) indicates quantitative removal of the caging groups after laser irradiation.

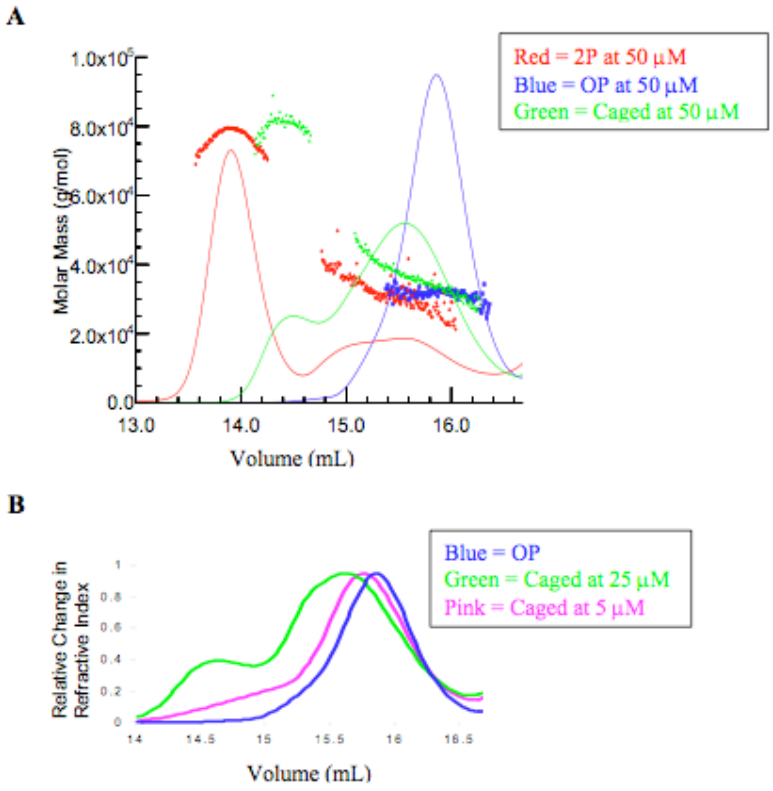


Figure 2.10. SEC/MALLS analysis of caged Smad2-MH2. A. Caged Smad2-MH2 (caged), doubly-phosphorylated Smad2-MH2 (2P) and non-phosphorylated Smad2-MH2 (0P) were applied to a S200 SEC column at a loading concentration of 50 µM. The eluant was analyzed by UV, refractive index (RI), and multi-angle light scattering (MALLS) detectors. The solid curves correspond to the RI signal and the points represent the molecular weight of the eluting species at each second derived from RI and MALLS measurements. At a loading concentration of 50 µM, the caged protein elutes in two peaks, the majority of which is part of a hetero-dimer with SARA-SBD. **B**. The normalized RI traces for the caged protein at 25 µM and 5 µM and for the non-phosphorylated protein are shown. The residual tendency of the caged protein to homo-trimerize at high concentrations is greatly reduced at 5 µM.

observed (Figure 2.10B). Brief irradiation (<5 seconds) of the caged protein with the output of the He-Cd laser followed by SEC, RP-HPLC, and ESI-MS demonstrated that the caging groups were quantitatively removed from the protein and that SARA-SBD was released from Smad2-MH2 in favor of homo-trimerization, thus generating protein **41** (Figure 2.9B-D).

As a step toward the ultimate goal of using caged phosphoproteins in live cells to study the kinetics of biological signaling and transport processes, we set out to determine the behavior of caged Smad2-MH2 in a nuclear import assay. For this, we labeled cysteines of the Smad2-MH2 variants with Texas Red C2-maleimide (Figure 2.11A). When incubated with digitonin-permeabilized HeLa cells in the presence of SARA-SBD, non-phosphorylated Smad2-MH2 (OP) is excluded from the nucleus, whereas phosphorylated Smad2-MH2 (2P) accumulates in the nucleus (Figure 2.11B) (Xu et al., 2000). UV irradiation had no effect on the localization pattern of phosphorylated and non-phosphorylated Smad2-MH2 control proteins (Figure 2.11B). In the same assay, we found that caged Smad2-MH2 was excluded from the nucleus, whereas uncaging of the protein with UV light led to clear nuclear accumulation (Figure 2.11B). This demonstrates that the caged and uncaged proteins behave as desired in a biological context.

2.6. Summary

Smad2-MH2 caged on two activating phosphate residues has been prepared by a semi-synthetic route. The molecule described represents the first report of a protein caged on phosphate. In principle, this approach can be applied to the construction of a caged

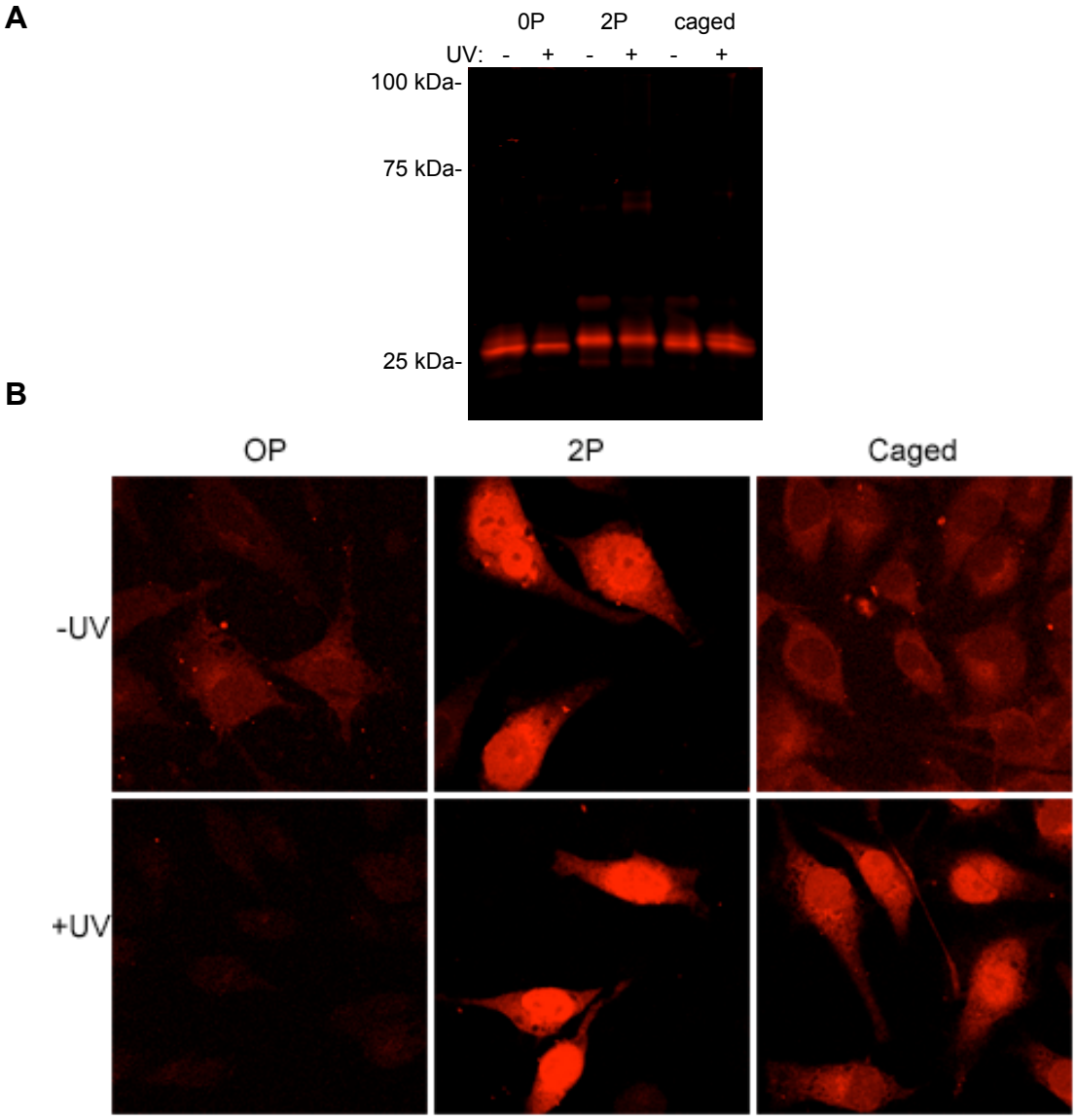


Figure 2.11. Nuclear import assay of Smad2-MH2 variants labeled with Texas Red C_2-maleimide (TRM). A. Smad2-MH2 variants (6 μM, 0P = non-phosphorylated, 2P = doubly-phosphorylated) were labeled at 4 °C overnight with TRM, purified by SEC and analyzed by fluorescence scanning of an SDS-PAGE gel before and after treatment with UV light (as in Figure 2.8). **B.** The proteins were incubated separately at 1.5 μM with digitonin-permeabilized HeLa cells for 20-minutes at room temperature in the presence of 4.5 μM GST-SARA-SBD, an ATP-regenerating system, and 1 mg/mL BSA. After the import reaction, cells were washed, fixed, and analyzed by confocal microscopy for the localization of each Smad2-MH2 variant.

version of any protein activated by phosphorylation. Further studies using this class of reagent can be expected to yield quantitative insight into the kinetics of Smad2 nuclear import and export.

Chapter 3: A caging strategy for simultaneous triggering of Smad2 activity and fluorescence[1]

3.1. Background

In the previous chapter, the successful semi-synthesis of the MH2 domain of Smad2 caged on its two C-terminal phosphoserines was described. While this is a useful reagent for controlling the time and site of protein activation, it is not possible to conditionally visualize the protein only when it is active. The ability to selectively monitor only the uncaged active form of the protein would provide a great advantage for live-cell imaging of Smad2. This would, in principle, allow for direct measurement of the extent of uncaging and activity of Smad2 following a dose of UV light. The fluorescent signal from the protein could be followed in time and space to determine the dynamic localization of the protein. Furthermore, this measurable fluorescence activity may be correlated to downstream cellular events.

[1] The work described in this chapter was completed in collaboration with Dr. Jean-Philippe Pellois while he was a postdoctoral associate in the Muir laboratory and resulted in the following publication: Pellois, J-P., Hahn, M.E., and Muir T.W., Simultaneous triggering of protein activity and fluorescence. *Journal of the American Chemical Society*, **126**, 7170 (2004).

3.2. Design of a caging strategy for Smad2 enabling the simultaneous activation of the protein and its fluorescence

It is not straightforward to design a protein so that a change in its fluorescence occurs upon activation. For inspiration, we turned to previous examples of induced fluorescence for measuring biomolecular activity (Figure 3.1). Fluorescence quenching methods have been used extensively to create probes that report the activity or presence of biomolecules such as proteases and DNA (Tyagi and Kramer, 1996). For example, a peptide containing a fluorophore and quencher in close proximity exhibits very little, if any, fluorescence due to intramolecular quenching (Figure 3.1A) (Weissleder et al., 1999). Further, if the peptide contains a protease recognition site between the fluorophore and quencher, then proteolysis will result in loss of proximity of the fluorophore and quencher. This leads to a large increase in fluorescence, which effectively reports the presence of active protease. In another example, probes known as molecular beacons are used to detect DNA or other related molecules (Figure 3.1B) (Tyagi and Kramer, 1996). A molecular beacon consists of a single-stranded deoxyoligonucleotide containing a fluorophore and a quencher forced into close proximity by formation of an engineered DNA hairpin structure (Figure 3.1B) (Tyagi and Kramer, 1996). In the presence of its complement, the hairpin structure is lost and the beacon participates in a double-helix in which the fluorophore and quencher are far apart, leading to a large increase in fluorescence (Figure 3.1B). Thus, a molecular beacon can be used to fluorescently monitor the presence of specific nucleic acid sequences.

We envisioned integrating this concept of induced increase in distance between a fluorophore and quencher with protein caging to construct a photoactivatable form of

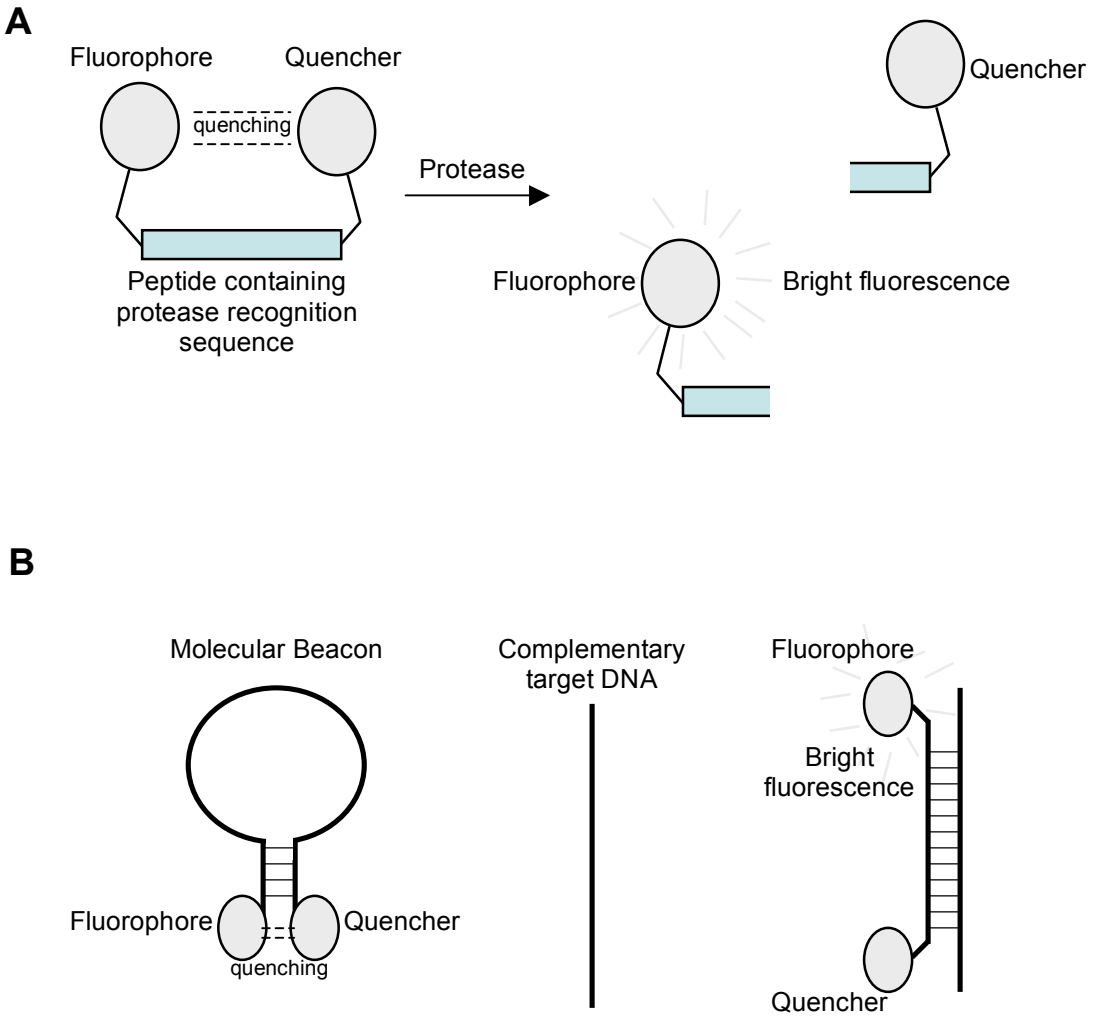


Figure 3.1. Induced fluorescence probes of biomolecules. A. A peptide consisting of a protease recognition sequence labeled on either end with a fluorophore and quencher is processed by a protease to yield two peptides that can diffuse away from each other. This leads to a robust increase in fluorescence of the portion of the peptide labeled with the fluorophore. **B.** A molecular beacon contains a flourophore and a quencher on either end of a short DNA stem loop. In the absence of a complementary target, the stem loop is stable and the fluorophore is quenched owing to the close proximity of the fluorophore and quencher. Upon hybridizing to a complementary target sequence, the molecular beacon takes part in a double helix. In this conformation, the fluorophore and quencher are no longer close enough in space for efficient quenching, thus fluorescence is restored.

Smad2 that becomes fluorescent upon activation. The requirements for this were: (1) labeling of Smad2 with a fluorophore and quencher that are close to each other before uncaging, (2) installation of a caging group that suppresses protein activity, and (3) linking uncaging to an increase in distance between the fluorophore and quencher to generate an increase in fluorescence. Inspection of the structure of the trimeric phosphorylated Smad2-MH2 domain provided a basis for rational design to satisfy these three requirements (Figure 3.2) (Wu et al., 2001). Methionine 466 is solvent exposed and does not participate in any protein-protein interactions in the Smad2 trimer (Figure 3.2B). Therefore, position 466 was selected as the site of fluorophore installation. The C-terminal carboxylate of Smad2 at position 467 participates in a network of hydrogen bonds that stabilizes the trimer (Figure 3.2C). We reasoned that installation of a bulky caging group at this position would prevent the trimerization-dependent activity of Smad2. Finally, a quencher could be linked to the caging group such that it would be in close proximity to position 466 of the protein while the caging group is intact, but would diffuse away with the cleaved caging group upon UV irradiation. In principle, the successful synthesis of caged and quenched Smad2-MH2 (**47**) following this design would provide a means to simultaneously activate and visualize the protein (**48**) (Figure 3.3A).

3.3. Validation of fluorescent labeling and caging strategies

Before preparing a caged and quenched version of Smad2-MH2 containing all of the modifications discussed above, we set out to systematically determine if each modification performed as expected in isolation. For this, we prepared control proteins

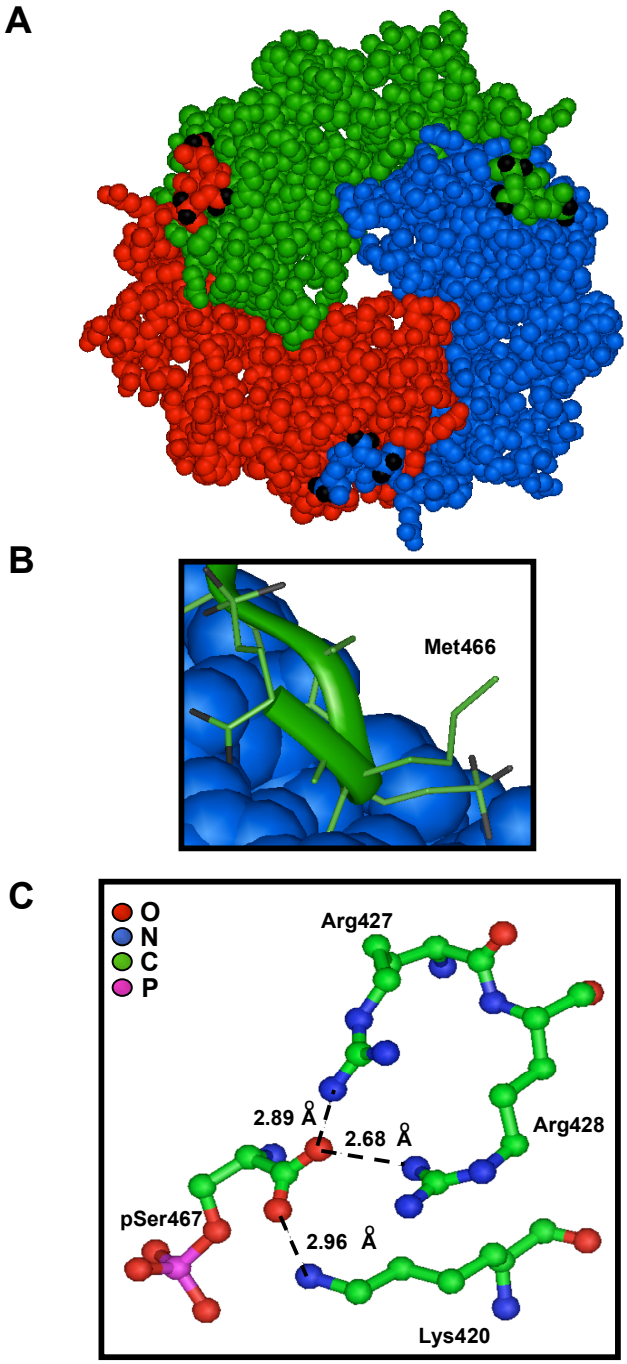


Figure 3.2. Structure of the Smad2-MH2 trimer. A. Space-filling model of the Smad2-MH2 homo-trimer (drawn from pdb code 1khx). Phosphoserine residues 465 and 467 are shown in black. **B.** Close up of methionine 466 shows that it is oriented toward solvent and does not make any protein contacts. **C.** The C-terminal carboxylate participates in hydrogen bonds that stabilize the homo-trimer.

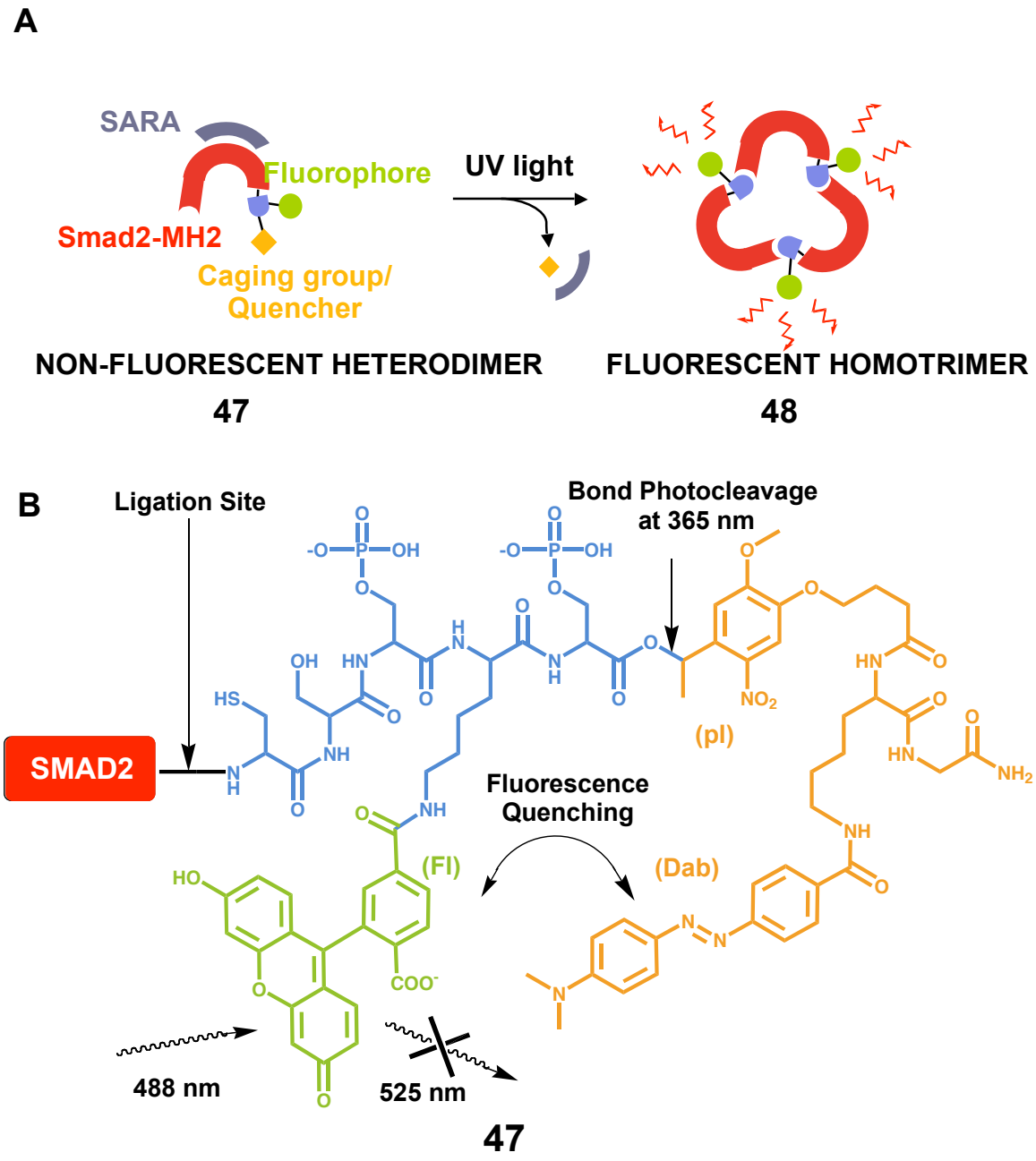


Figure 3.3. Caged and quenched Smad2-MH2 (47). A. Schematic depicting caged and quenched Smad2-MH2. Upon UV irradiation, the caging group and quencher are removed, leading to homo-trimerization and fluorescence of the protein. **B.** Structure of peptide **52** ligated to Smad2. Shown in blue is the phosphopeptide skeleton. Shown in green is carboxyflourescein. Shown in orange is the caging group-quencher unit.

49, **50**, and **51** by employing a similar semi-synthetic strategy (using Smad2 residues 241-462 in complex with SARA-SBD) as was used for the phosphocaged MH2 domain discussed in chapter 2 (Figure 3.4A). To determine if Smad2-MH2 is stable to UV light, protein **49** was analyzed before and after irradiation at 365 nm. Size exclusion chromatography and mass spectrometry demonstrated that **49** was unchanged after exposure to UV light (Figure 3.4B). Next, methionine 466 of Smad2 was mutated to lysine in protein **50**. This lysine was labeled at its ε-amine with the fluorophore carboxyfluorescein[2] and the protein was analyzed by size exclusion chromatography and mass spectrometry (Figure 3.4C). Provided that the modification at position 466 does not interfere with the protein structure, then **50** should homotrimerize due to the presence of phosphoserine at positions 465 and 467. Indeed, size exclusion chromatography analysis of **50** demonstrated that it is homotrimeric (Figure 3.4C,i).

Having validated position 466 as a suitable location for fluorescent labeling and verifying that Smad2-MH2 is stable to UV light, we went on to test if the caging strategy outlined in section 3.2 would be suitable for Smad2-MH2. In addition to containing phosphoserine at positions 465 and 467, protein **51** contains a caging group at its C-

[2] Carboxyfluorescein is not the best fluorophore for imaging in living systems. Fluorescein derivatives photobleach more readily than other fluorophores. Its use in these proteins was motivated primarily by the fact that it is a well-known fluorophore that interfaces well with solid phase peptide synthesis. Moreover, several variables and unknowns were being evaluated simultaneously during the development of the caging and quenching strategy so we wanted to minimize additional unknowns resulting from the use of less well characterized fluorophores.

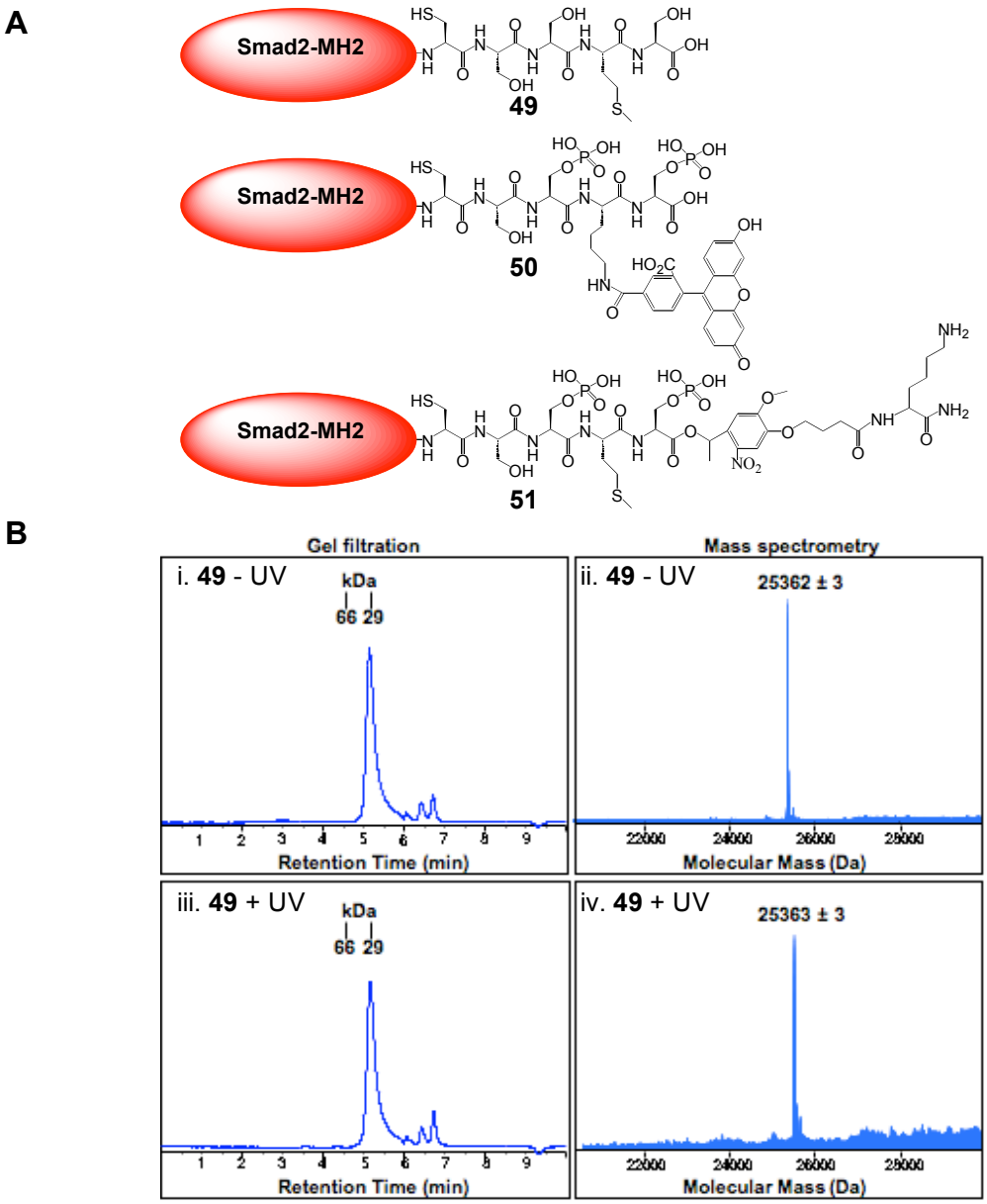


Figure 3.4. Verification of fluorescent labeling and caging strategies. A. Control proteins used. **B.** Irradiation of **49** with UV light does not lead to any apparent change in oligomeric structure (compare i and iii) or molecular mass (compare ii and iv) as demonstrated by ESI-MS and SEC (expected M_r is 25,358.8 Da).

C

D

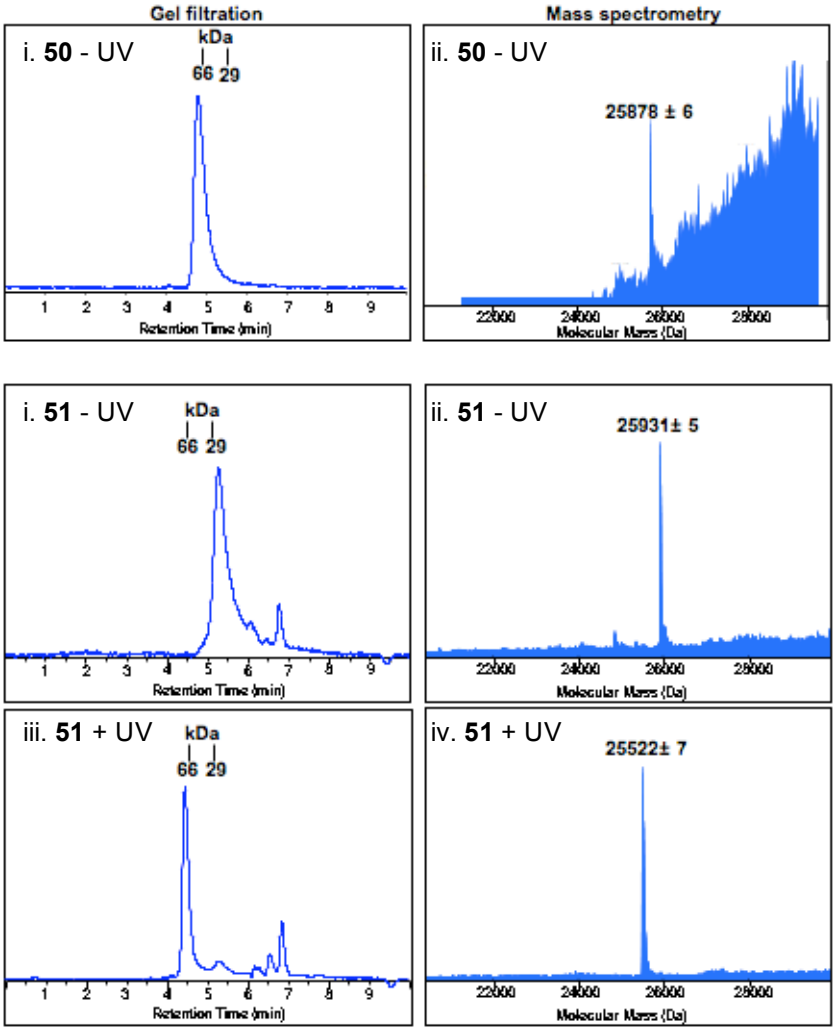


Figure 3.4, continued. C. 50 is homotrimeric (i) and was assembled correctly (ii) as judged by ESI-MS (expected M_r is 25,874.1 Da). **D.** Before irradiation, **51** elutes at a retention time consistent with it taking part in a heterodimer with SARA-SBD (i). After irradiation, **51** is homotrimeric (iii). Irradiation leads to loss of the caging group as demonstrated by mass spectrometry (compare ii and iv; expected M_r of **51** – UV is 25,928.2 Da and **51** + UV is 25,518.7 Da). UV irradiation in all applicable panels was performed for 2 minutes at 365 nm (~ 10 mW/cm^2). In all panels showing gel filtration traces, homotrimeric Smad2-MH2 elutes at approximately 4.5 minutes whereas the heterodimeric complex of Smad2-MH2 and SARA-SBD elutes at approximately 5.5 minutes. Peaks after ~6 minutes represent buffer components.

terminus. To judge the caging efficiency, **51** was analyzed by size exclusion chromatography and mass spectrometry before and after irradiation with UV light (Figure 3.4D). Before irradiation, **51** participated in a heterodimer with SARA-SBD (Figure 3.4D, i and iii), as did the inactive, non-phosphorylated protein **49** (Figure 3.4B, i). Upon irradiation with UV light, the caging group was removed (Figure 3.4D, iv) and **51** formed homotrimers (Figure 3.4D, iii), indicating that uncaging activated the protein. These analyses demonstrated that each modification required for the production of caged and quenched Smad2-MH2 was independently functional.

3.4. Semi-synthesis of caged and quenched Smad2

In section 3.3, each modification required for caged and quenched Smad2-MH2 was tested individually. Since they all performed as desired, we went on to unite them in one protein to prepare caged and quenched Smad2-MH2. Smad2-MH2-α-thioester (residues 241-462) in complex with SARA-SBD was prepared and ligated to peptide **52** to generate caged and quenched Smad2-MH2 (**47**, Figure 3.3). Peptide **52** is highly modification-dense and contains all the elements required for expressed protein ligation (N-terminal cysteine), fluorescence (carboxyfluorescein), quenching (dabcyl), caging (4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy]butanoic acid), and activity after uncaging (phosphoserine residues). Key to the synthesis of peptide **52** was a series of orthogonal protection/deprotection steps enabling installation of the fluorophore and quencher (Figure 3.5A). Also, limiting the time of Fmoc deprotection after the installation of the phosphoserine residues to 3 minutes minimized piperidine-induced β-elimination of the phosphate residue. This peptide was extremely challenging to

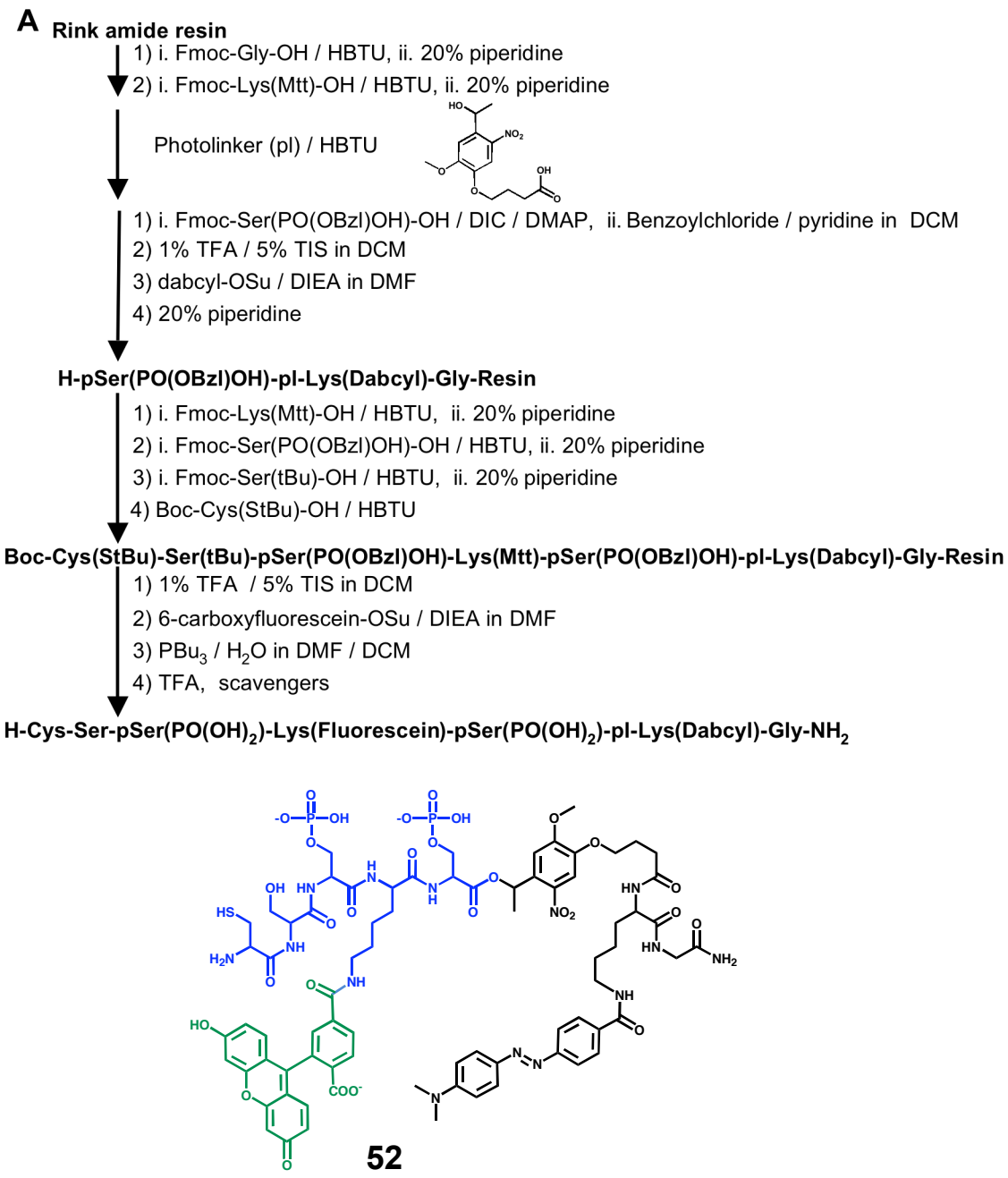
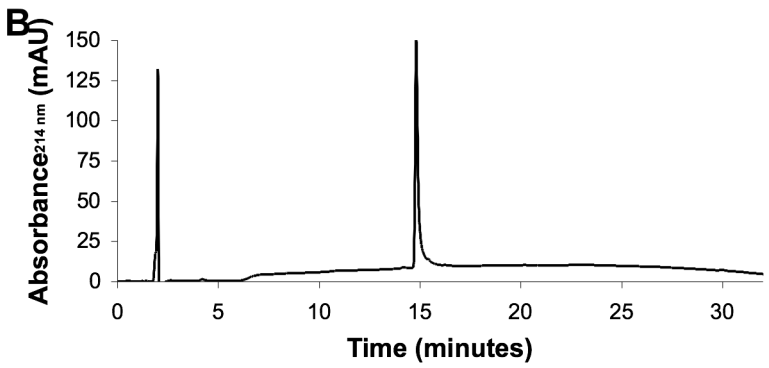
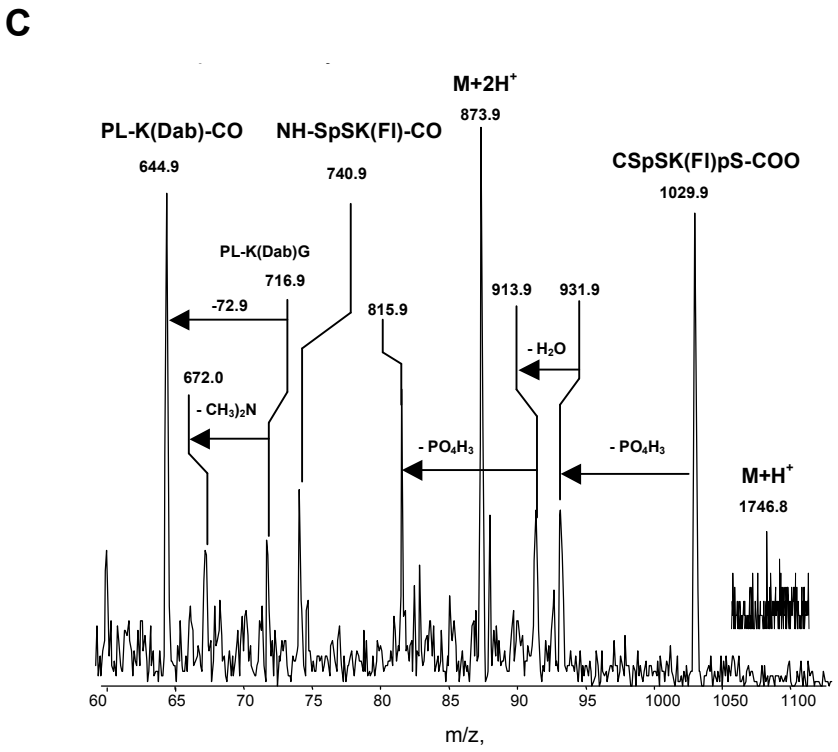


Figure 3.5. Synthesis of peptide 52 used to prepare caged and quenched Smad2. A. Synthetic scheme and structure of final product. **B.** RP-HPLC (20-73%B) of **52. C.** Mass spectrum of **1**. Expected M_r is 1744.6 Da. **D.** UV-VIS spectrum of peptide **52**.





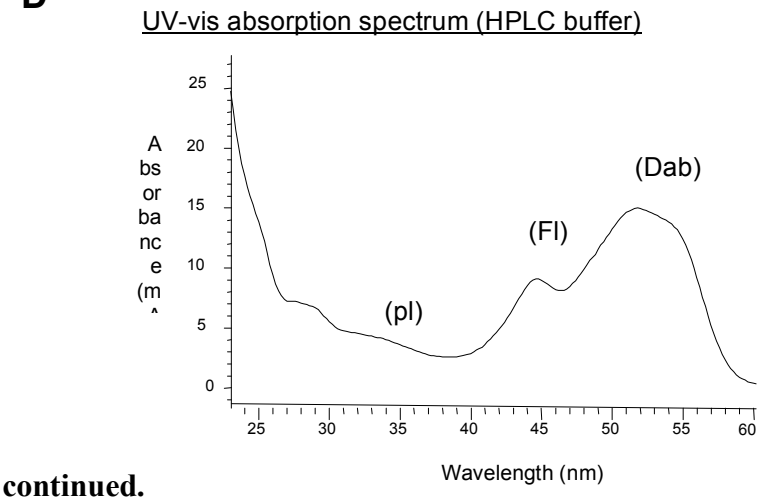


Figure 3.5, continued.

synthesize owing to the presence of many modifications on a small peptide skeleton. Fortunately, synthesis was successful in that the quantity of peptide generated was sufficient for expressed protein ligation reactions. Reversed-phase HPLC, mass spectrometry, and spectrophotometry confirmed the identity of the peptide (Figure 3.5B-D).

Expressed protein ligation was carried out with the Smad2-MH2-α-thioester/SARA-SBD complex and peptide **52** to create the caged and quenched protein **47** (Figure 3.3). The ligation reaction was initiated by adding peptide **52** to a final concentration of 1 mM to a solution of Smad2-MH2-α-thioester (**38**) at 0.25 mM. The reaction proceeded for 24 hours at 4 °C (Figure 3.6A) at which point the ligated protein was purified by preparative size-exclusion chromatography (SEC).

3.5. Characterization of caged and quenched Smad2-MH2

Mass spectrometry confirmed the identity of the caged and quenched Smad2-MH2 protein (Figure 3.6D), which is in complex with SARA-SBD in 1:1 stoichiometry, as demonstrated by integration of the Smad2-MH2 and SARA-SBD peaks of the RP-HPLC traces shown in Figure 3.6B. Analytical size exclusion chromatography demonstrated that before irradiation with UV light, the caged and quenched protein exists as a heterodimer with SARA-SBD (Figure 3.6C). UV irradiation led to cleavage of the caging group and quencher from the protein, as confirmed by mass spectrometry (Figure 3.6D). The uncaged protein no longer binds to SARA-SBD, but instead forms homotrimers, as expected (Figure 3.6B,C). These data indicate that the protein design

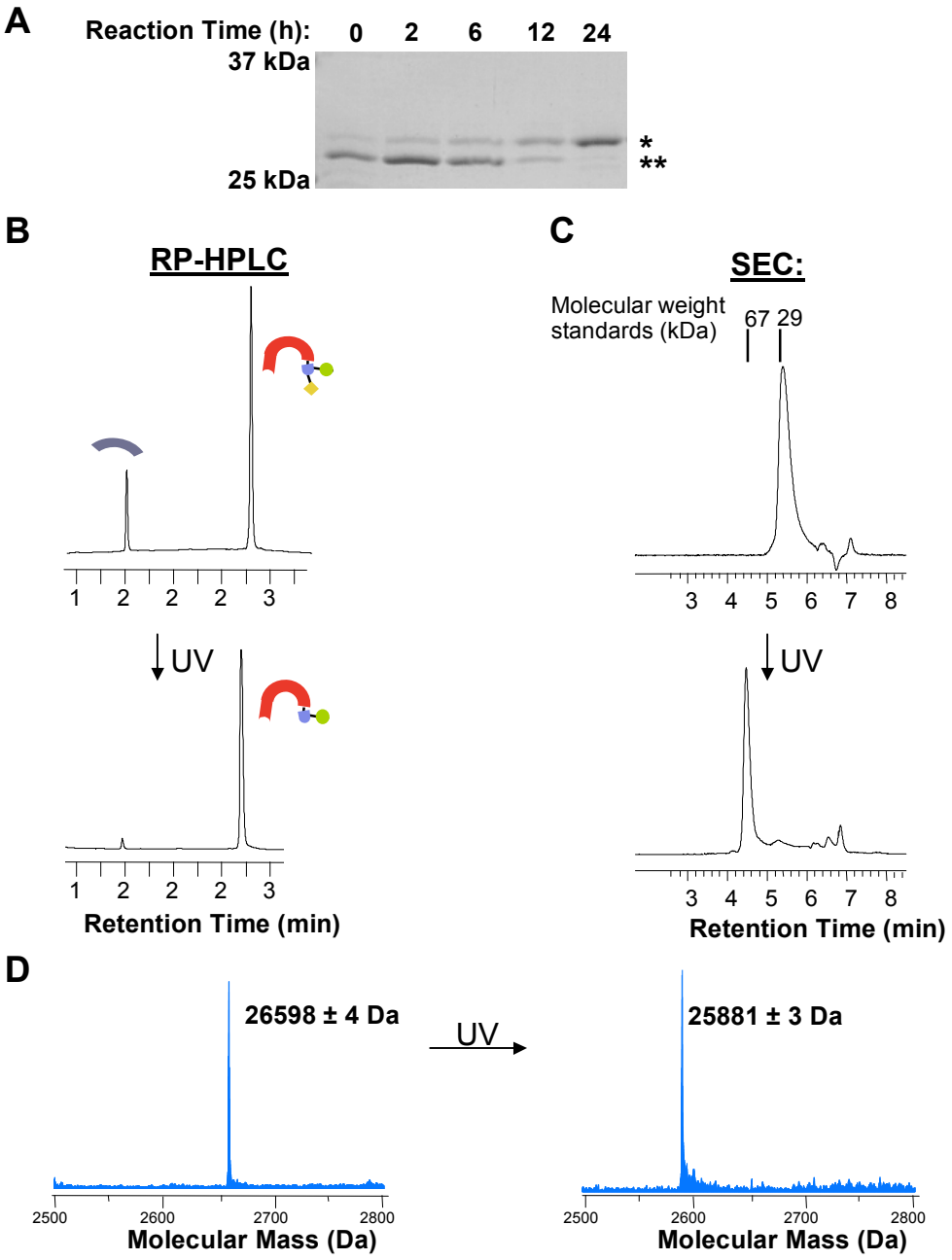


Figure 3.6. Characterization of caged and quenched Smad2-MH2 (47). A. SDS-PAGE analysis of ligation reaction at the indicated timepoints. Smad2-MH2-α-COSR is indicated by ** and the ligation product is indicated by *. **B.** RP-HPLC of caged protein before and after UV irradiation, demonstrating release of SARA-SBD upon photocleavage. **C.** SEC demonstrates that caged Smad2-MH2 elutes at an apparent molecular weight consistent with it being in a hetero-dimeric complex with SARA-SBD. Uncaged Smad2-MH2 elutes at an apparent molecular weight consistent with it being in a homo-trimeric complex. **D.** Reconstructed mass spectrum of caged and uncaged protein. Expected M_r of caged protein is 26,590.8 Da and uncaged protein is 25,874.1 Da.

principles discussed above were successful for the production of a caged version of Smad2-MH2.

As the goal of this work was to link protein uncaging to an increase in fluorescence, we next monitored the fluorescence of the protein after UV irradiation. Upon irradiation with low-intensity UV light at 365 nm, the fluorescence of the protein conjugate increased with time (Figure 3.7A). The photolysis was first order with a rate constant of 4.0×10^{-3} s^{-1} and the quantum yield of uncaging was 0.03 (Figure 3.7B). This contrasts with the observed quantum yield of 0.16 of photocleavage of the caging groups of the phosphocaged protein discussed in chapter 2. Since the two proteins share the same core caging group, it is inferred that the lower quantum yield of the caged and quenched protein is due to optical and/or electronic interference of the fluorophore or quencher with the caging group. However, irradiation at 325 nm for 3 seconds with the output of a high-intensity UV laser resulted in complete photolysis and conversion to the uncaged protein (Figure 3.7A). The maximal fluorescence increase upon uncaging was 26-fold. This is a robust increase that can be used to selectively monitor the uncaged form of the protein. Furthermore, since the quantum yield of photolysis was determined, it is possible to uncage either a fraction or all of the caged protein at one time, thus uncaging is titratable.

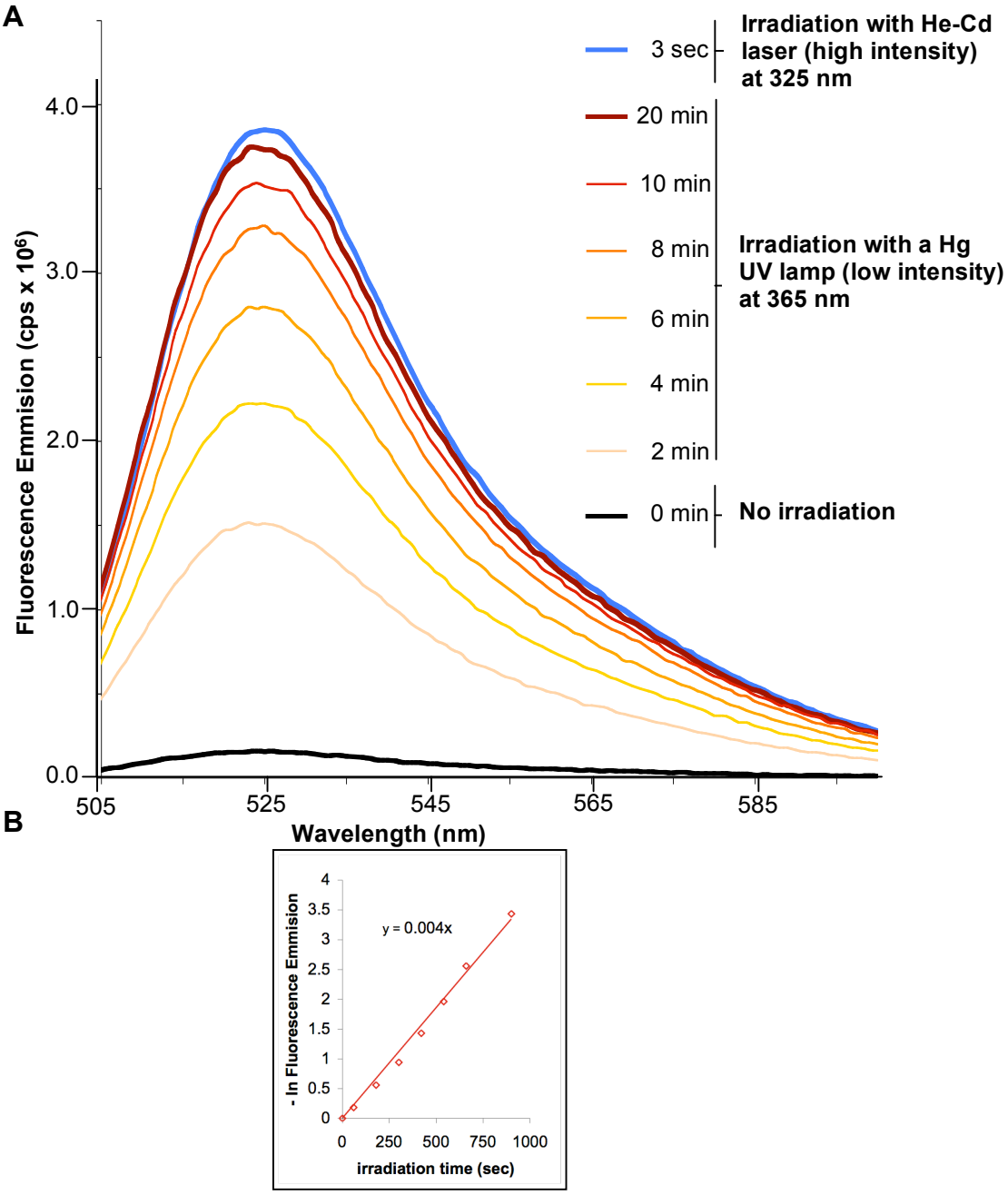


Figure 3.7. Photoactivation and fluorescence of caged and quenched Smad2-MH2 (47). A. In the absence of UV irradiation, the caged and quenched protein is essentially non-fluorescent due to intramolecular quenching of fluorescein by dabcyl. Low intensity UV irradiation leads to an increasing fraction of uncaged protein over time. Irradiation with high intensity UV light from a laser completely uncages the protein, resulting in a 26-fold increase in fluorescence. **B.** The log plot used to find the rate constant of uncaging is shown.

3.6. Photoactivation of fluorescence in live cells and *Xenopus* embryos[3]

Having successfully demonstrated that the caging strategy enables photoactivation of fluorescence *in vitro*, we undertook an investigation to determine conditions for photoactivation of fluorescence inside live systems. For these studies, a model protein containing glutamic acids in place of the phosphoserines (Smad2-MH2-2E, **53**) was used for synthetic ease. Peptide **54** (Figure 3.8A) was synthesized in the same manner as peptide **52**, with the exception of substitution of glutamic acid for phosphoserine at positions 465 and 467 of Smad2. The caged and quenched glutamic acid-containing protein was prepared in the same manner as described in section 3.4 by chemical ligation of peptide **54** to Smad2-MH2-α-thioester.

This protein was microinjected into the cytoplasm of live HeLa cells that were then irradiated with low intensity UV light (Figure 3.8B). Total cell fluorescence intensity increased for the first ten minutes of irradiation and reached a plateau of fourteen-fold by twenty minutes of irradiation (Figure 3.8C). The fraction of **53** increased in the nucleus following photoactivation. Interestingly, the observed nuclear fluorescence after uncaging was not uniform in distribution (Figure 3.8D). Instead, punctate fluorescence patterns were observed in the nucleus. While it is not clear why this occurs, it is intriguing to speculate that Smad2 clusters at molecular addresses inside the nucleus, possibly corresponding to promoter binding sites. However, it is possible that these results are artifacts due to lack of part of the linker of Smad2 and its MH1 domain in **53**.

[3] The work described in this section was carried out in collaboration with Ariel Levine of the Brivanlou laboratory.

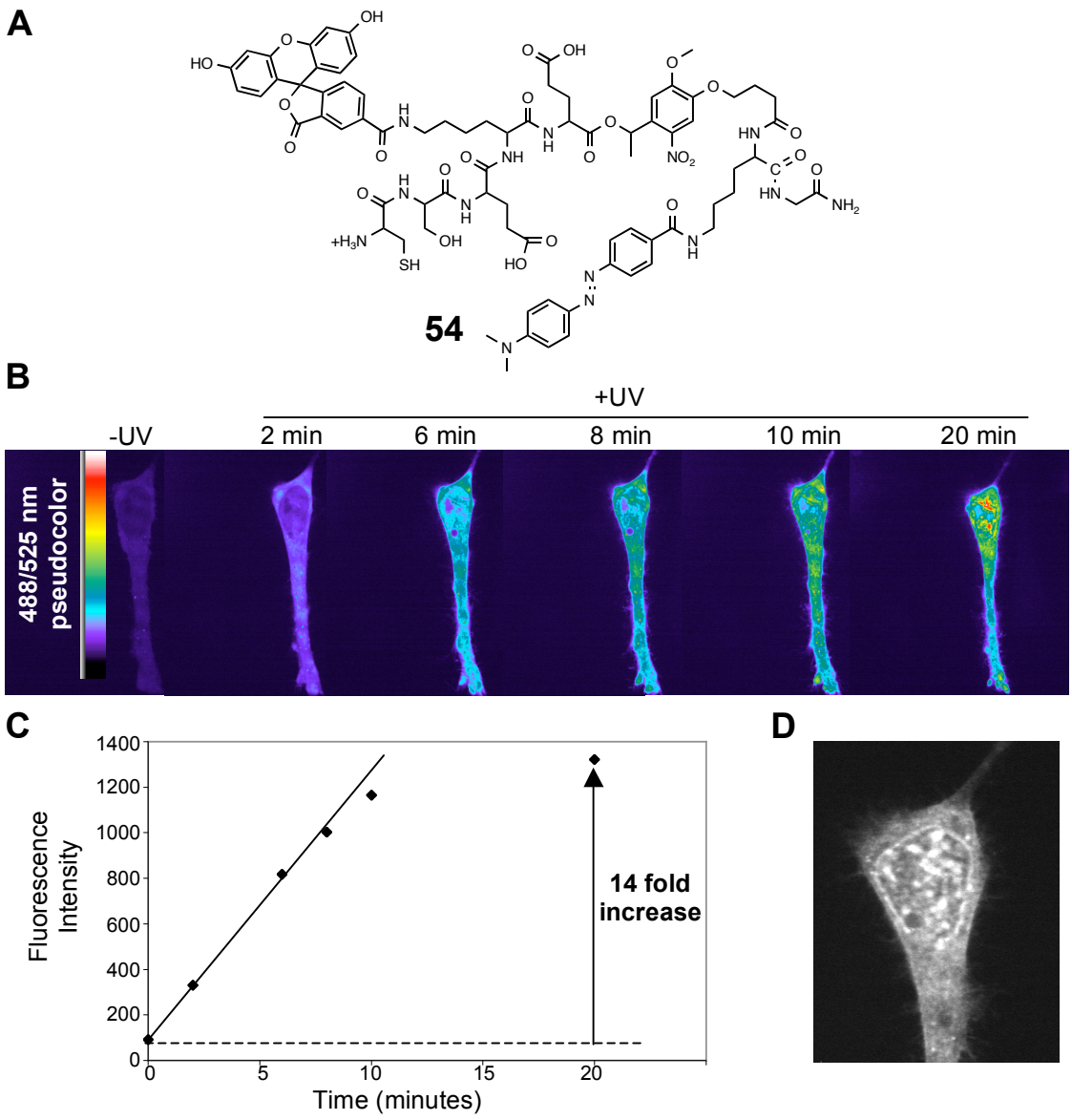


Figure 3.8. Photoactivation of caged and quenched Smad2-MH2-2E (53) in live cells. **A.** Structure of peptide **54** used to prepare Smad2-MH2(2E). **B.** Caged and quenched Smad2-MH2-2E was injected into the cytoplasm of live HeLa cells. The cells were continuously irradiated with UV light through the microscope's standard DAPI excitation filter. Confocal images were taken at the times indicated. **C.** The total fluorescence intensity of the cell in **B** was measured at each timepoint and plotted. After 20 minutes of continuous irradiation, fluorescence increased 14-fold. **D.** Grayscale image of the 10 minute irradiation timepoint shows non-homogenous, punctate Smad2 fluorescence in the nucleus and fluorescence at the nuclear envelope.

It is also possible that the concentration of **53** once microinjected into cells is higher than the endogenous Smad2 concentration.

The caged and quenched protein **53** was also microinjected into the animal pole of live two-cell stage *Xenopus laevis* embryos. The embryos were allowed to develop overnight. Next, the animal regions were explanted, placed under a coverslip and imaged by confocal fluorescence microscopy before and after photoactivation with a 325 nm laser. Rhodamine-dextran served as a cytosolic marker. Representative images of caged and uncaged Smad2-MH2 in animal region explants clearly indicate that photoactivation of fluorescence was achieved in these samples (Figure 3.9).

Next, we evaluated the biological activity of semi-synthetic Smad2-MH2 variants in live *Xenopus laevis* embryos by means of an Activin responsive luciferase assay. Like TGFβ, stimulation of cells with Activin leads to the phosphorylation of Smad2 and Smad3 (Massague et al., 2005). Control studies using non-phoshorylated and doubly-phosphorylated Smad2-MH2 were undertaken. Each protein was microinjected along with an Activin-responsive luciferase construct into embryos, which were allowed to develop until the late blastula stage. At this point, their animal poles were explanted and lysed and luciferase assays were conducted. Both proteins induced luciferase expression above background (Figure 3.10, compare lanes 3 and 4 to lane 1). Injection of Activin mRNA served as a positive control. This produced strong luciferase induction as compared to both proteins (Figure 3.10, compare lane 2 to lanes 3 and 4). Luciferase induction due to doubly phosphorylated Smad2-MH2-2P and non-phosphorylated Smad2-MH2-0P injection was not significantly different from one another (Figure 3.10, compare lanes 3 and 4). It is possible that the reason for this lack of significant luciferase

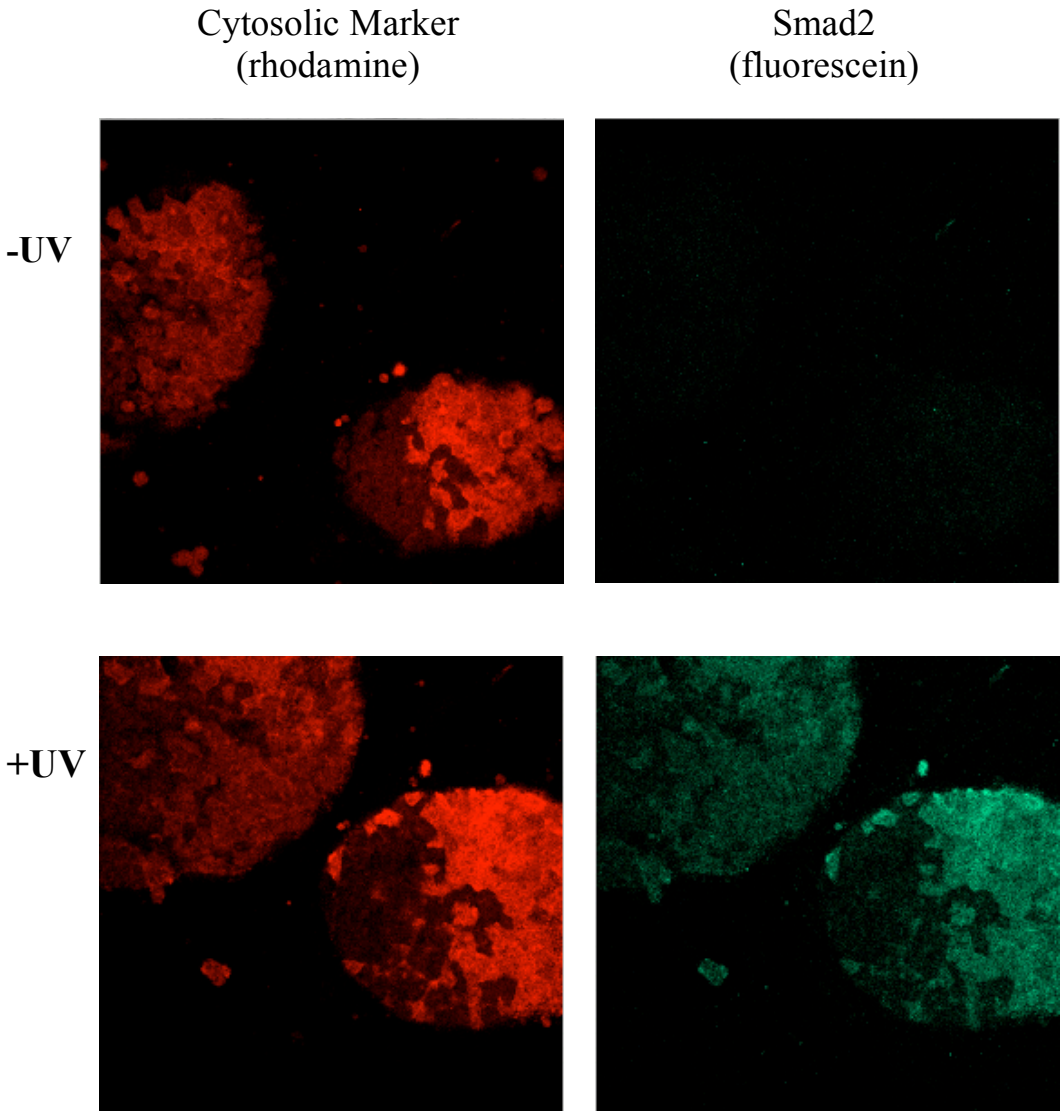


Figure 3.9. Photoactivation of caged and quenched Smad2-MH2-2E (53) in live *Xenopus* embryos. Frog embryos at the 2-cell stage were injected in the animal pole with a solution containing caged Smad2-MH2 and were allowed to develop to the late blastula stage. Animal regions were explanted, placed under a coverslip and imaged by confocal fluorescence microscopy before and after photoactivation with a 325 nm laser. Rhodamine-dextran serves as a cytosolic marker. Each frame is 1.3 mm x 1.3 mm.

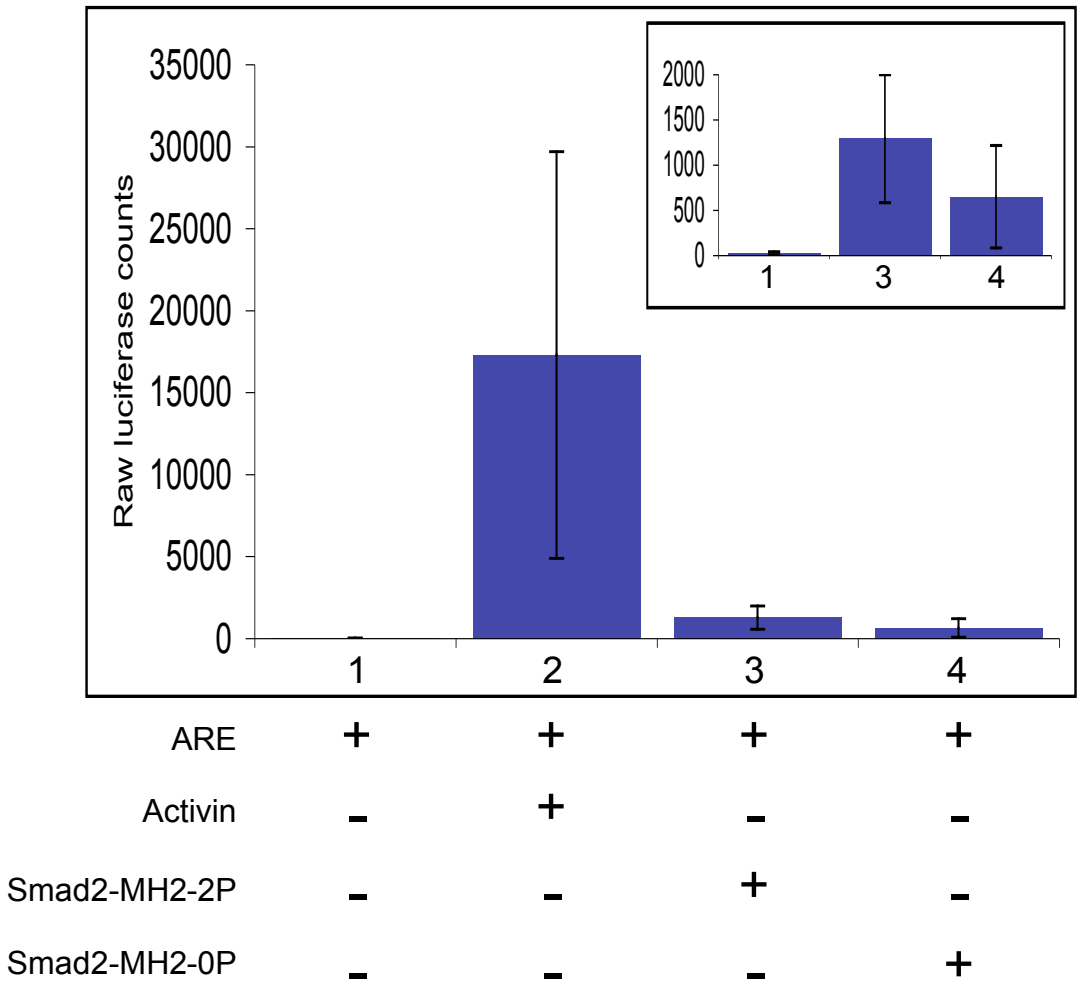


Figure 3.10. ARE luciferase assays of Smad2-MH2 variants in *Xenopus* embryos. 2-cell stage embryos (n=3) were injected with an ARE-responsive luciferase reporter (ARE). Embryos were co-injected as indicated above with either Activin mRNA (positive control), 5 nM doubly-phosphorylated Smad2-MH2-2P, or 5 nM non-phosphorylated Smad2-MH2-0P. Activin mRNA injection leads to a robust increase in luciferase activity over background (compare lanes 1 and 2). Injection of Smad proteins also leads to an increase in luciferase activity over background (compare lane 1 with lanes 3 and 4) that is more modest than that induced by Activin mRNA (compare lane 1 with lanes 3 and 4, see inset). The activity of the phosphorylated MH2 is not statistically different (S.D.) than the non-phosphorylated protein (compare lanes 3 an 4, see inset).

induction of the phosphorylated protein compared to the non-phosphorylated protein is related to the fact that only a fragment of the protein was used that does not contain the MH1 domain or the full linker region. This led us to pursue methods that would allow us to interface our caging strategy with full-length Smad2, the subject of the next chapter.

3.7. Summary

Rational protein design led to the semi-synthesis of a protein that upon uncaging and protein activation exhibits a robust increase in fluorescence. Therefore, the active protein can be monitored selectively by its increased fluorescent signal. Proof of principle experiments demonstrated that these proteins can be uncaged and fluorescence can be activated in live systems such as HeLa cells and *Xenopus* embryos. While the results with the caged Smad2-MH2 proteins described in this chapter were encouraging, they do not necessarily reflect physiological reality since they lacked the MH1 and part of the linker of Smad2. In order to achieve more realistic results, we required full-length caged Smad2. Therefore, we proceeded to develop a protocol for the production of full-length Smad2-α-thioester on the preparative scale. Efforts toward this are detailed in the next chapter.

Chapter 4: Tunable Photoactivation of a Posttranslationally Modified Signaling Protein and its Unmodified Counterpart in Live Cells[1]

4.1. Background

Chapter 3 described the development of a caging strategy enabling simultaneous activation and fluorescence monitoring of a fragment of Smad2. Biochemical and *in vivo* data demonstrated that implementation of the caging strategy led to the production of a Smad2-MH2 analogue that could be activated by UV light. Furthermore, the robust increase in fluorescence upon uncaging makes this caging strategy suitable for live-cell imaging studies. However, imaging of the caged and quenched Smad2-MH2 protein **47** after uncaging in the context of intact cells may not provide physiologically relevant results since the protein is just a fragment of Smad2. Full-length Smad2 contains an MH1 domain and an intact linker region that is not present in the caged and quenched Smad2-MH2 protein **47** (Massague et al., 2005). The MH1 domain of R-Smads is the portion of the protein capable of binding DNA at Smad-responsive promoters (Kim et al., 1997; Massague et al., 2005). The MH1 domain binds co-factors required for specific, high-affinity promoter binding (Massague et al., 2005). Smad2 contains a 30-amino acid insert

---

1 The work described in this chapter was completed in collaboration with Dr. Jean-Philippe Pellois and Dr. Miquel Vila-Perelló while they were postdoctoral associates in the Muir laboratory and resulted in the following publication: Hahn, M.E., Pellois, J-P., Vila-Perelló, M., and Muir T.W., Tunable Photoactivation of a Posttranslationally Modified Signaling Protein and its Unmodified Counterpart in Live Cells. *ChemBioChem*, in press.

of uncertain physiological significance in its MH1 domain (Massague et al., 2005; Yagi et al., 1999). Both the MH1 domain and linker region are subject to regulation by phosphorylation and other posttranslational modifications (Massague et al., 2005). Clearly, it is vital to use full-length Smad2 in any attempt to extract physiologically-relevant data from imaging studies. Therefore, we sought to apply the semi-synthetic caging strategy described in chapter 3 to full-length Smad2.

4.2. Determination of the fractional phosphorylation stoichiometry of Smad2 following treatment of cells with TGFβ

Hill and co-workers have studied kinetic aspects of Smad2 nucleocytoplasmic shuttling upon phosphorylation (Nicolas et al., 2004; Schmierer and Hill, 2005). The general method used in these studies relies upon expression of Smad2 fused to a fluorescent protein (FP) in live cells. Smad2 localization is then followed by confocal fluorescence microscopy after treatment with TGFβ. This method can be used to unambiguously monitor activated Smad2 only if the entire cellular pool of the protein is phosphorylated in response to stimulation of the cells by TGFβ since fluorescence of the FP is not coupled to Smad2 phosphorylation. This is problematic because most posttranslational modifications (PTMs) are thought to be installed substoichiometrically (Mann and Jensen, 2003). A potential advantage of our method is that Smad2 fluorescence is linked to activation via uncaging.

We sought to measure the extent of Smad2 phosphorylation upon treatment of cells with TGFβ. For this purpose, we designed an immunoblotting assay to determine the fraction of total Smad2 molecules that are phosphorylated at any given time following

TGFβ treatment (Figure 4.1). Extracts of HeLa and HaCaT cells that had been treated with TGFβ for 45 minutes were found to contain the maximal amount of C-terminally phosphorylated Smad2 relative to uninduced cells (Fig. 4.1A), which is consistent with previous studies (Inman et al., 2002). Standards generated using defined mixtures of purified non-phosphorylated and phosphorylated Flag-Smad2[2] (Figure 4.1B) were analyzed by Western blotting for both total Smad2 and phosphorylated Smad2 levels (Fig. 4.1C). Quantitative analysis of these blots allowed the generation of a standard curve (Figure 4.1D) from which we determined the fractional stoichiometry of Smad2 phosphorylation in various cell lines following treatment with TGFβ for 45 minutes (Figure 4.1C). As shown in Figure 4.1E, the fraction of phosphorylated Smad2 varied significantly between these cell lines, underlining the key point that PTMs can be present at substoichiometric levels depending on the experimental context. These data demonstrate that our strategy of selectively monitoring active Smad2 has advantages over other methods. We note that this immunoblotting method is a simple alternative to existing mass spectrometry approaches for determining phospho-protein levels (Gerber et al., 2003) and can indeed be applied to the quantification of other PTMs in cases where the modified protein standards and PTM-specific antibodies are available.

[2] The phosphorylated Flag-Smad2 standard was prepared by EPL. Section 4.3 details the preparation of full-length Flag-Smad2-α-thioester. Semi-synthesis was carried out in the same manner as described in section 4.4.

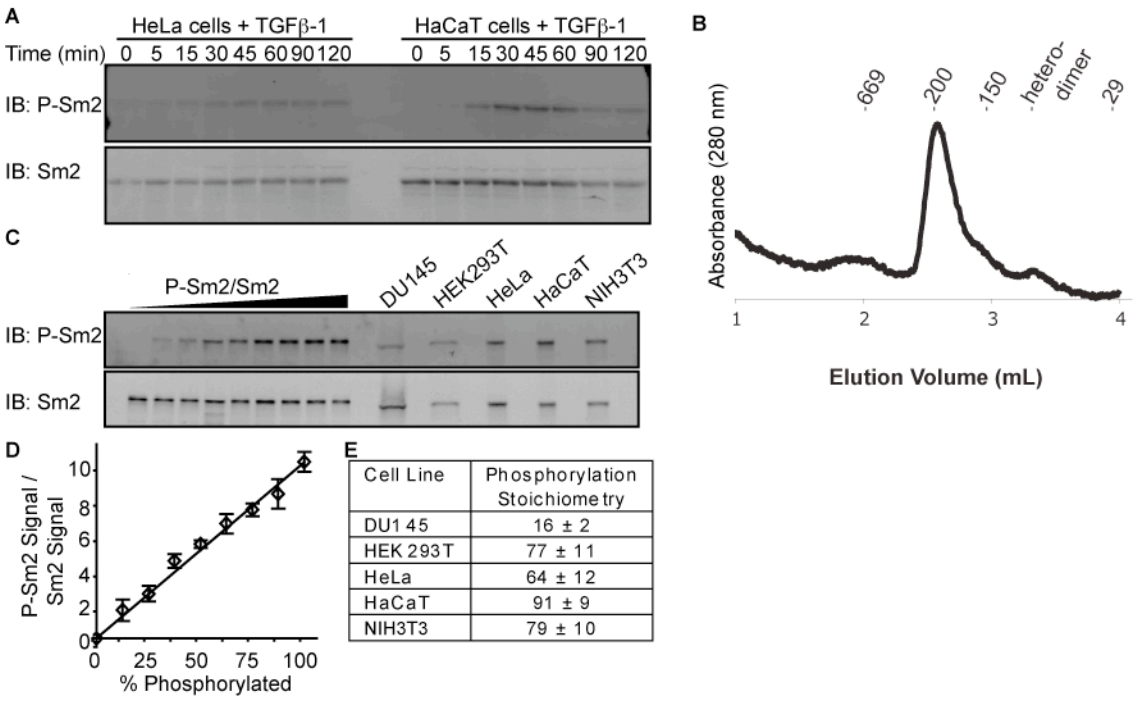


Figure 4.1. Determination of the fractional stoichiometry of Smad2 C-terminal phosphorylation induced by TGFβ. **A.** Extracts of the indicated cell lines treated with TGFβ for the indicated times were analyzed by immunoblotting with antibodies directed to phosphorylated Smad2 (P-Sm2) and total Smad2 (Sm2). **B.** Doubly-phosphorylated semi-synthetic Flag-Smad2 standard used in **C** was analyzed by size exclusion chromatography to confirm that it is homotrimeric. Elution positions of molecular weight standards (kDa) and a Flag-Smad2/MBP-SARA-SBD heterodimer are indicated. Expected M_r of homotrimeric Flag-Smad2 is 165.1 kDa. **C.** Recombinant non-phosphorylated Flag-Smad2 was mixed with an increasing fraction of phosphorylated Flag-Smad2 and analyzed by immunoblotting alongside extracts of the indicated cell lines that had been treated with TGFβ for 45 minutes. **D.** Comparison of the integrated intensities of the bands in panel **C** resulting from immunoblotting with antibodies for total Smad2 and phosphorylated Smad2 yields a standard curve from which Smad2 phosphorylation stoichiometry can be determined. Error is S.D. (n = 3). **E.** The fractional stoichiometry of Smad2 C-terminal phosphorylation in various cell lines is shown. Error is S.D. (n = 3).

4.3. Expression and purification of full-length Smad2-α-thioester

We designed caged Smad2 (**55**) so that photolysis of a single bond would result in both protein and fluorescence activation (Pellois et al., 2004), ensuring that any fluorescent signal observed is due to active Smad2 (Figure 4.2). Preparation of a complex molecule of this type is non-trivial, but should be possible using the protein semi-synthesis approach expressed protein ligation (EPL) (Muir, 2003), which is especially useful for building complex functionality into polypeptides. For this we required recombinant protein-α-thioester building block **57** (Figure 4.3), which corresponds to Smad2 lacking the last 5 residues (residues 1-461). Fusing a protein of interest to a modified intein and subsequently treating the fusion with exogenous thiol releases the protein as an α-thioester.

Preliminary experiments indicated that simply appending full-length Smad2 to an intein did not result in sufficient yield of pure protein for our needs. Therefore, we developed a modified expression and purification strategy. Soluble expression of Smad2-intein in *E. coli* was greatly enhanced by fusing the protein to a N-terminal SUMO-Flag sequence. However, the protein still lacked suitable stability and precipitated during purification. We reasoned that performing lysis in the presence of a Smad2 binding partner might aid in the stabilization of the protein during subsequent purification steps. Indeed, bacterial lysis in the presence of the Smad binding domain (SBD) of SARA (Tsukazaki et al., 1998; Wu et al., 2000) fused to Maltose-Binding-Protein (MBP) improved the stability of the protein (Figure 4.4A). Thiolysis of the intein and subsequent removal of the SUMO domain by selective proteolysis (Mossessova and Lima, 2000)

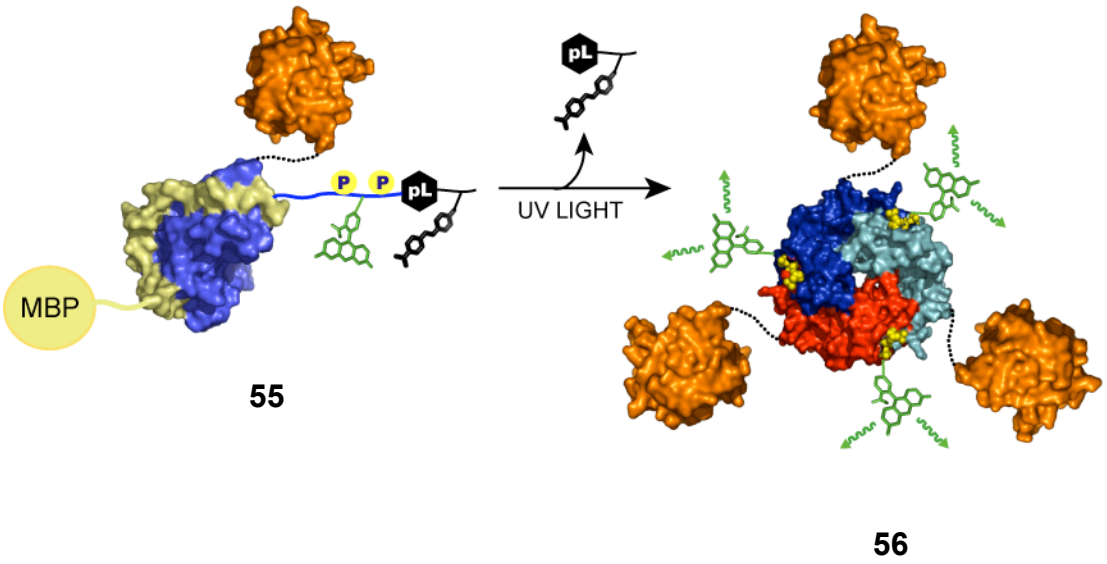


Figure 4.2. Design of caged full-length Smad2 (55). The MH1 domain (pdb code 1MHD) is modeled in orange and the MH2 domain is in blue in complex with SARA-SBD in gold (pdb code 1DEV). The linker between the two domains is depicted by a dotted line. Before irradiation the protein is non-fluorescent and unable to homotrimerize due to the presence of the bulky photolinker-quencher unit (black). Upon irradiation with UV light, the photolinker-quencher unit is released, allowing Smad2 to dissociate from SARA, form homotrimers and become fluorescent. The phosphoserine residues that mediate oligomerization are shown in yellow and carboxyfluorescein is shown in green. The MH1 domain is modeled from the crystal structure of the Smad3 MH1 domain. MBP = maltose binding protein. The individual MH2 domains of the Smad2 homotrimer are in blue, cyan, and red.

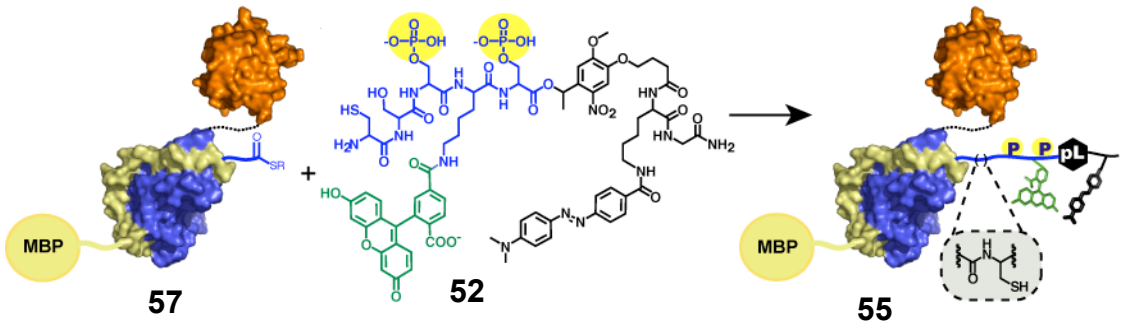


Figure 4.3. Semi-synthesis of caged full-length Smad2 (55) by expressed protein ligation. Full-length Smad2-α-thioester in complex with MBP-SARA-SBD (**57**) was ligated to synthetic peptide **52** to yield caged Smad2 (**55**). Peptide **52** contains a N-terminal cysteine for EPL, two phosphoserine residues (highlighted in yellow), carboxyfluorescein (green), and a photolinker-quencher unit (black). The linker between the two domains is depicted by a dotted line.

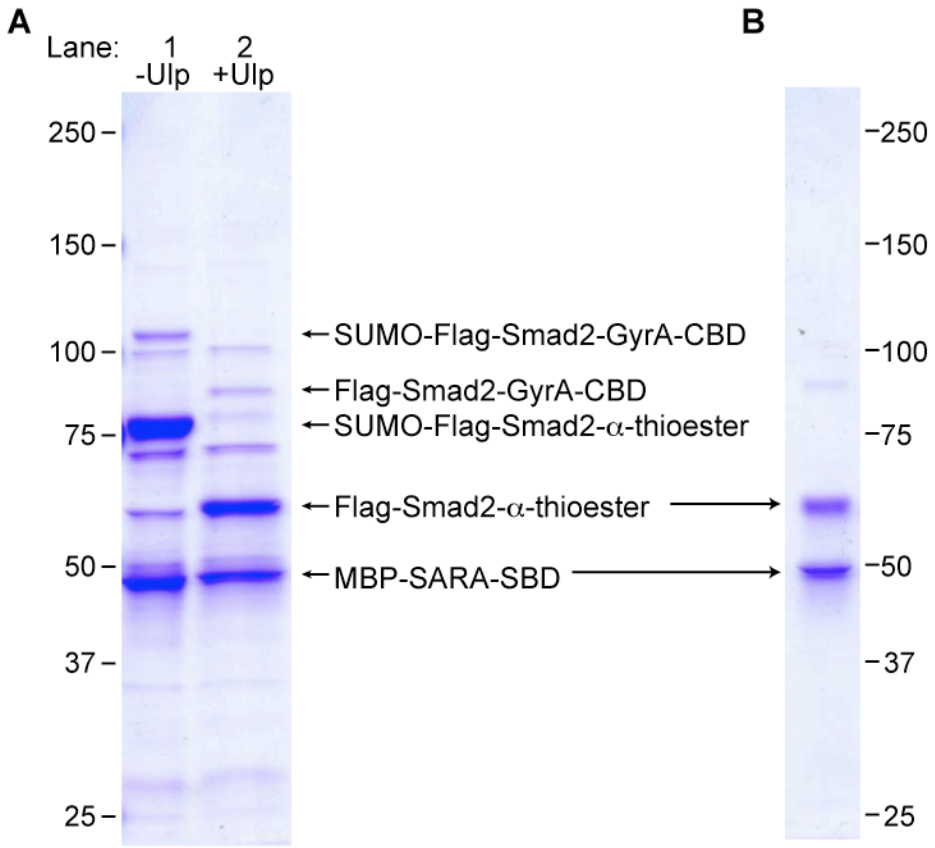


Figure 4.4. Cleavage of SUMO from SUMO-Flag-Smad2-α-thioester. A. Chitin column fractions containing SUMO-Flag-Smad2-α-thioester were treated (+Ulp) or not (-Ulp) with SUMO protease (Ulp) overnight at 4°C and analyzed by SDS-PAGE. Identities of the bands corresponding to various forms of Smad2 and MBP-SARA-SBD are indicated. **B.** The protein represented in panel **A**, lane 2, was further purified by anion exchange chromatography and analyzed by SDS-PAGE. The expected M_r of Flag-Smad2-α-thioester (**57**) and MBP-SARA-SBD is 54.2 kDa is 48.9 kDa, respectively. Gels were stained with coomasie.

gave full-length Flag-Smad2-α-thioester/MBP-SARA-SBD (**57**) in milligram amounts following purification (Figure 4.4B).

4.4. Semi-synthesis and characterization of caged full-length phosphorylated Smad2

Following production of sufficient quantities of purified protein **57**, we went on to prepare caged full-length Smad2 (**55**) by EPL (Figure 4.3). For this, we used peptide fragment **52**, which contains the last five amino acids of Smad2 as well as all of the functionalities required for ligation, caging, fluorescence, and quenching (Figures 3.5 and 4.3). The ligation reaction was initiated by combining purified synthetic peptide **52** and recombinant protein complex **57** under physiological conditions at 4 ^{o}C (Figure 4.3). The reaction was monitored by SDS-PAGE with fluorescence imaging, which indicated that the reaction was complete after two days. The ligation product **55** was then purified by size-exclusion chromatography and analyzed by SDS-PAGE (Figure 4.5A).

The purified ligation product **55** exists as a heterodimer with SARA-SBD and was able to form homotrimers upon uncaging with UV light (Figure 4.5A,B). Surprisingly, some of the Smad2 protein eluted on an SEC column at a volume consistent with it remaining in complex with SARA-SBD even though the photolysis reaction was essentially complete under the irradiation conditions employed (325 nm at 4.74 W/cm^2 for 3 seconds). We wondered whether the SARA-SBD in the mixture could compete with trimerization of the uncaged protein. To test this idea, we treated purified doubly-phosphorylated full-length Smad2[3] with increasing amounts of SARA-SBD and indeed

[3] This protein was prepared by EPL in the same manner as the caged and quenched protein described in section 4.4.

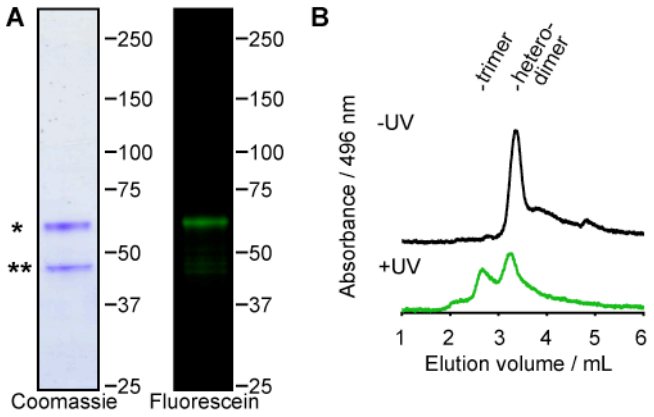


Figure 4.5. Characterization of caged full-length Smad2 (55). A. SDS-PAGE of caged protein **55**, which is a complex between semi-synthetic fluorescent Smad2 (*) and MBP-SARA-SBD (**). Both proteins are detected by coomassie staining of the SDS-PAGE gel, whereas fluorescence imaging of the gel detects only the semi-synthetic Smad2 (through its residual carboxyfluorescein fluorescence). Expected M_r of caged Flag-Smad2 and MBP-SARA-SBD is 55.1 kDa and 48.9 kDa, respectively. **B.** Caged full-length Flag-Smad2 (**55**) was either irradiated with the output of a laser at 325 nm for 3 seconds (+UV) or left in the dark (-UV). The protein was then analyzed by size exclusion chromatography. Elution positions of authentic trimeric and heterodimeric (with MBP-SARA-SBD) standards are shown. **C.** Phosphorylated full-length Smad2 exists in an equilibrium of homotrimer and heterodimer in the presence of SARA-SBD. Each trace corresponds to the SEC profile of phosphorylated full-length Smad2 in the absence (a) or presence (b-e) of MBP-SARA-SBD. The ratio of Smad2 to Sara is approximately 2:1 in b, 1:1 in c, 1:2 in d, and 1:4 in e. **D.** Fluorescence spectra of caged full-length Flag-Smad2 (**55**) in vitro before and after irradiation with UV light at 365 nm. The inset indicates the energy of light applied and fold increase of Smad2 fluorescence (520 nm) after irradiation.

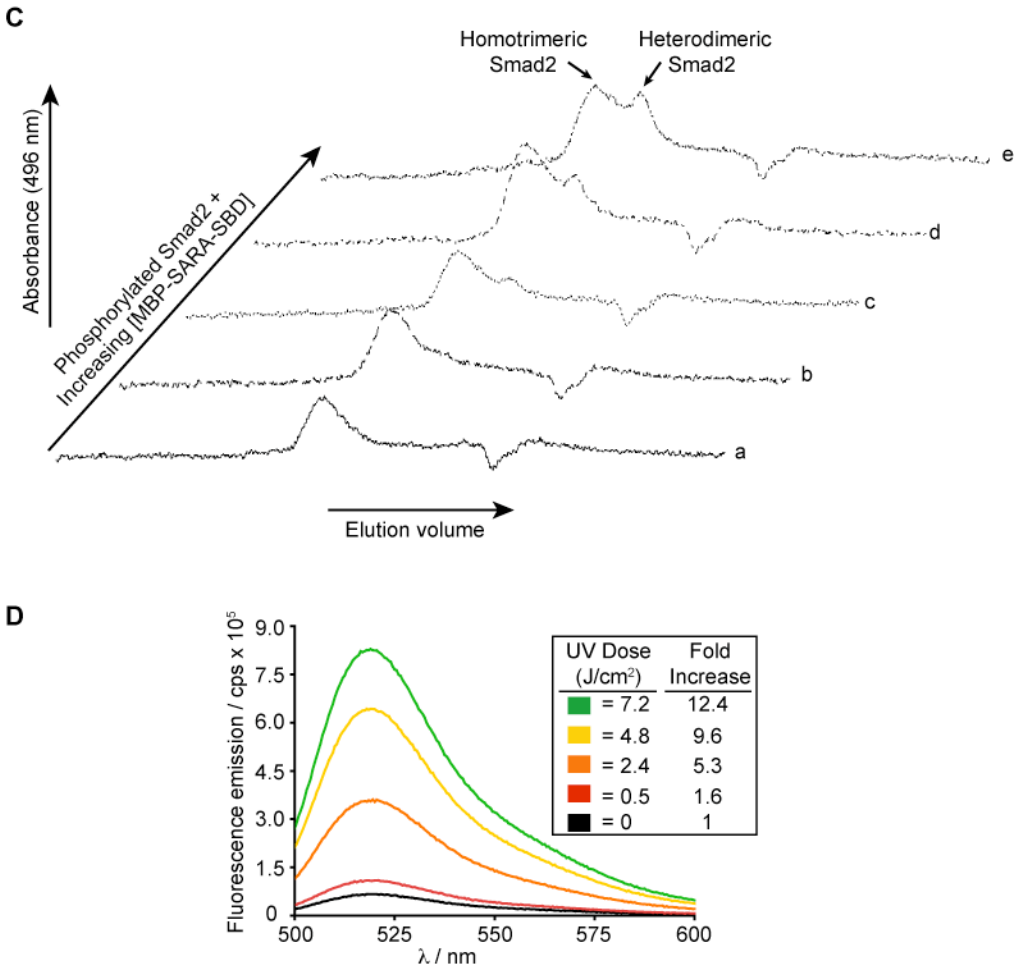


Figure 4.5, continued.

observed dose-dependent disruption of the Smad2 trimer, as measured by size exclusion chromatography (SEC; Figure 4.5C). This is in stark contrast with the doubly-phosphorylated MH2 domain alone which forms stable homotrimers even in the presence of a large excess of SARA-SBD (Ottesen et al., 2004).

A ~12-fold increase in carboxyfluorescein fluorescence was attained upon UV photocleavage of the purified caged protein, indicating release of the photolabile unit containing the dabcyl quencher (Figure 4.5D). This was lower than the corresponding value of 26-fold for the caged Smad2-MH2 domain (**47**) described in chapter 3 (Figure 3.7A). Since the same fluorophore/quencher system was utilized in both proteins, it is unclear why the fold increase of fluorescence upon photoactivation is different between them. It is known that the MH1 and MH2 domains can interact with each other (Hata et al., 1997). This may lead to close proximity of the MH1 domain to the C-terminus of the same polypeptide where the quencher and fluorophore reside. This in turn may change the polarity of the environment immediately surrounding the fluorophore and quencher to perturb the quenching efficiency or fluorescence of carboxyfluorescein, potentially resulting in the observed difference in fold activation of fluorescence.

Next we studied the utility of our approach in live cells. Photocaged Smad2 (**55**) was microinjected into the cytosol of live HaCaT cells and monitored by confocal microscopy. The fluorescence signal from carboxyfluorescein before photoactivation was barely detectable over background, indicating that dabcyl effectively quenches carboxyfluorescein inside cells (Figure 4.6A). Application of a continuous low dose of UV light allowed us to examine whether photoactivation could be titrated (Figure 4.6B). Increasing the dose of UV light led to an increase in fluorescence emission, which

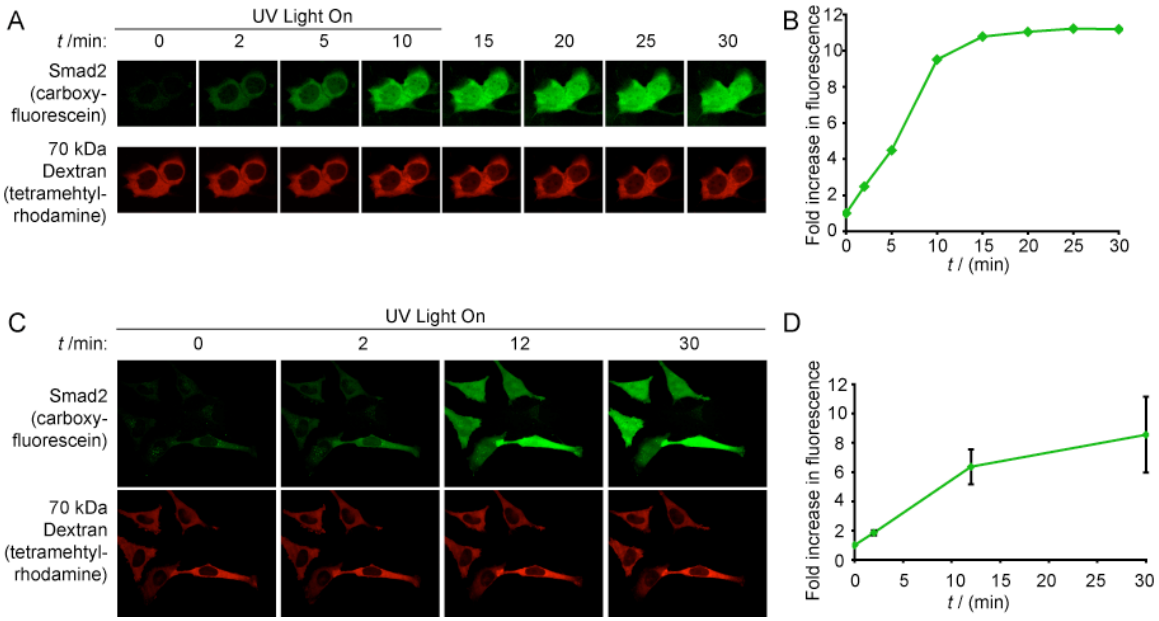


Figure 4.6. Titration of phosphorylated Smad2 activity in live cells. **A.** Caged full-length Flag-Smad2 (**55**) and 70 kDa tetrametylrhodamine-dextran (cytosolic marker) were microinjected into live HaCaT cells, which were then irradiated with low intensity UV light at 360 nm for 10 minutes. Confocal images were obtained at the times indicated. **B.** Quantification of Smad2 fluorescence in the cells from panel A. **C.** Caged full-length Flag-Smad2 (**55**) and 70 kDa tetrametylrhodamine-dextran (cytosolic marker) were microinjected into live HeLa cells, which were then irradiated with low intensity UV light at 360 nm for 10 minutes. Confocal images were obtained at the times indicated. **D.** Quantification of Smad2 fluorescence in the cells from panel **B**. Error bars are ± S.D.

reached a maximum of 12-fold over the non-irradiated sample (Fig. 4.6B). Similar results were obtained in HeLa cells (Figure 4.6C,D). This robust change in signal allows discrimination of the caged and uncaged versions of the phospho-protein in live cells, thereby validating a key part of our design, namely direct visualization of titratable protein activity.

4.5. Kinetic analysis of Smad2 nuclear import in live HeLa cells

Uncaging provides temporal control over protein function allowing kinetic analysis of cellular processes (Lawrence, 2005). As demonstrated above, our system has the added benefit that it is possible to selectively monitor the movement of the uncaged phospho-protein. To illustrate this point, we monitored the nuclear import of **55** after a short (60 s) UV exposure (Figure 4.7A). The mean nuclear fluorescence in each cell was normalized to total cell fluorescence at each time point and plotted against time (Figure 4.7B). The import kinetic curves are sigmoidal in nature with an apparent lag phase suggestive of one or more intermediate events. The data were analyzed by non-linear least squares curve-fitting to a stretched exponential function (see Methods section 7.31). While this analysis does not yield a true rate constant, it does provide a useful empirical description of complex kinetic processes (Hamada and Dobson, 2002; Morozova-Roche et al., 1999). In particular, the parameter *n* is a measure of cooperativity in the process and was found to be greater than 1 in all cases (Figure 4.7B) consistent with the presence of a lag phase in the nuclear import of our fluorescent construct under this experimental regime. This lag phase could reflect many intermediate events including dissociation of uncaged **56** from SARA-SBD, oligomerization of uncaged **56** and/or interactions of

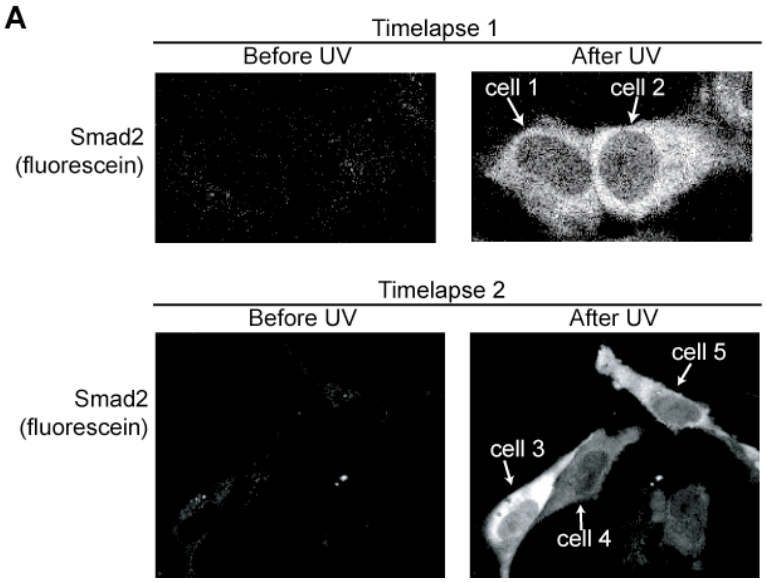


Figure 4.7. Nuclear import kinetics of caged Flag-Smad2 (55) following uncaging. **A.** Caged Flag-Smad2 (**55**) was microinjected into live HeLa cells, which were then irradiated with UV light for 60 seconds. Images were taken every 10 seconds. Images before and after UV irradiation (365 nm) are shown. Each cell is identified by a label and arrow. **B.** Mean nuclear fluorescence in each cell was normalized to total cell fluorescence at each time point and plotted against time (black dots), where t = 0 refers to the time at which the irradiation was switched off. The data were fit to the stretched exponential function (Equation 1). Values of k_s (s^{-1}) and n with errors (S.E.) for each cell are given in the inset of each plot. R^2 was greater than 0.95 for all fitted curves.

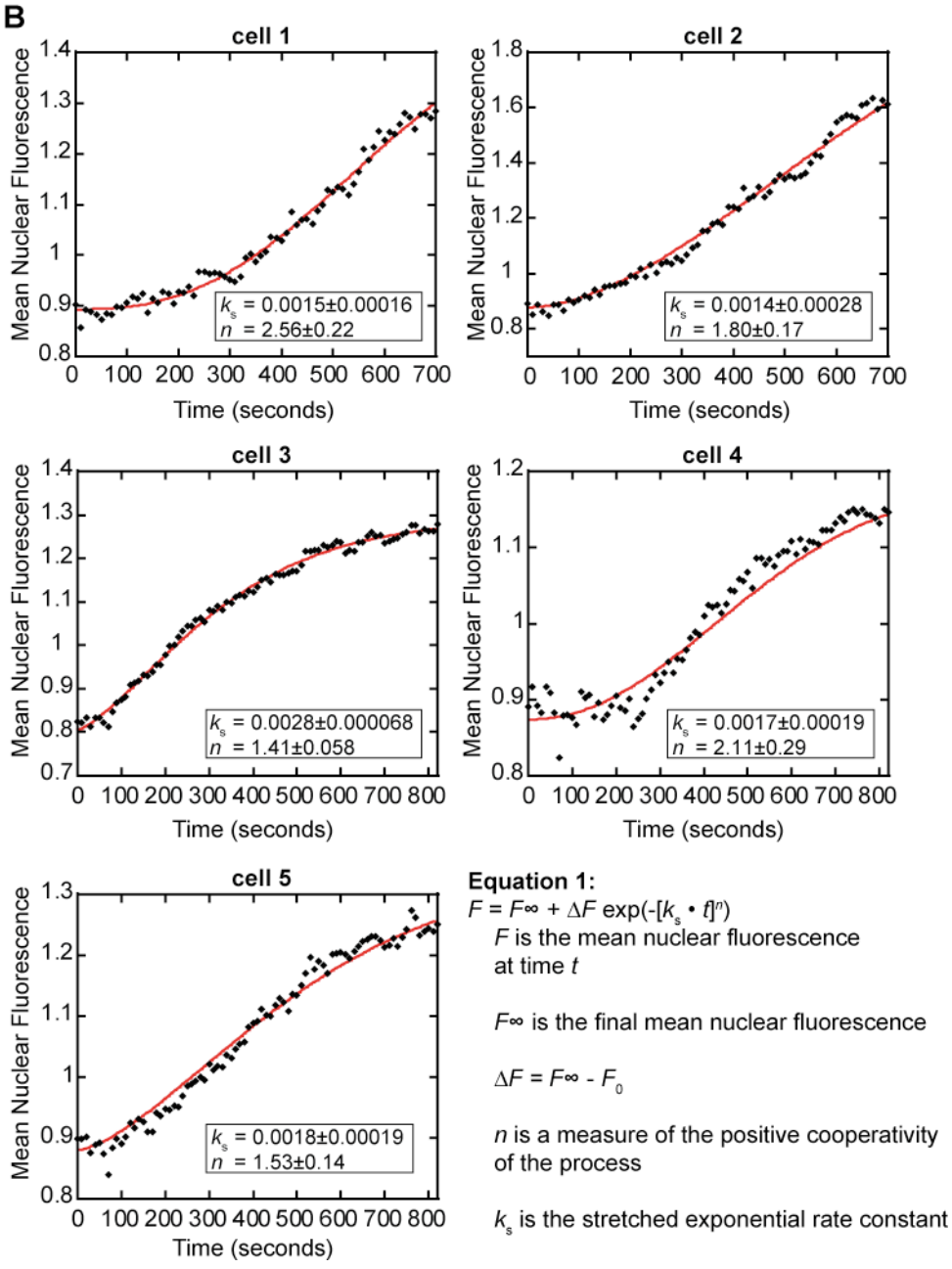


Figure 4.7, continued.

uncaged **56** with other cellular components such as microtubules (Dong et al., 2000) or the nuclear pore complex (Xu et al., 2002). Regardless of the origins of this lag phase, the kinetic data demonstrate that it is possible to resolve discrete kinetic events in complex cellular translocation processes using this strategy.

4.6. Simultaneous activation of two different forms of full-length Smad2 in live cells

In order to better understand the effects of phosphorylation on Smad2 dynamic localization, it would be useful to compare the phosphorylated form of the protein to its non-phosphorylated counterpart. Other investigators have accomplished this by comparing the dynamic properties of GFP-labeled Smad2 in cells treated with TGFβ to GFP-labeled Smad2 in untreated cells (Nicolas et al., 2004; Schmierer and Hill, 2005). While cells do communicate and interact with one another, each cell is its own unique system.

Cells have variable geometry, may have different protein concentrations, and may be in different phases of the cell cycle. These and other phenotypic differences may impact the behavior of the same protein in different cells. Therefore, quantitative comparisons between proteins residing in different cells may be prone to artifacts.

To address this issue, we envisioned monitoring the phosphorylated and non-phosphorylated proteins in the same cell at the same time. We set out to accomplish this by preparing a non-phosphorylated caged full-length Smad2 analogue labeled with an orthogonal fluorophore (tetramethylrhodamine) followed by simultaneous imaging of the phosphorylated and non-phosphorylated forms of the protein in the same cell. To demonstrate the potential of this approach, we prepared protein **58** by ligating protein-α-

thioester **57** to peptide **59** (Figures 4.8 and 4.9). Protein **58** is a caged and quenched version of non-phosphorylated Smad2 containing the red fluorophore tetramethylrhodamine. The two proteins **55** and **58** differ in two important ways: i. **55** is phosphorylated on serine residues 465 and 467 while **58** is not (the serines are replaced by alanines in **58**), and ii. they contain orthogonal fluorophores. Proteins **55** and **58** share the same core caging group and were found to exhibit similar uncaging kinetics (Figure 4.10A). The maximal increase in fluorescence of the tetramethylrhodamine labeled protein **58** after UV photoactivation was 6-fold as compared to 12-fold for the carboxyfluorescein labeled protein **55** (Figure 4.5D and Figure 4.9B). The observation that dabcyl quenched carboxyfluorescein more efficiently than tetramethylrhodamine presumably reflects the greater spectral overlap between carboxyfluorescein emission and dabcyl absorbance as compared to that of tetramethylrhodamine emission and dabcyl absorbance.

Encouraged by the *in vitro* behavior of protein **58**, we moved on to ask whether it was possible to simultaneously activate proteins **55** and **58** in the same cell. Accordingly, proteins **55** and **58** were co-injected into HeLa cells that were then treated with a short (60 seconds) UV pulse and then imaged by confocal fluorescence microscopy after a further 5 minutes (Figure 4.10B,C). Both proteins were successfully uncaged as judged by the increase in fluorescence in both the green (protein **55**) and red (protein **58)** channels. In keeping with the *in vitro* results, the quenching of protein **58** was not as efficient as for protein **55** in cells, although protein fluorescence was clearly activated in both channels (Figure 4.10B). In addition, protein **55** could be detected both in the nucleus and cytoplasm after uncaging, whereas protein **58** was cytoplasmic. This

A

Rink amide resin

1) i. Fmoc-Gly-OH / HBTU, ii. 20% piperidine
2) i. Fmoc-Lys(Mtt)-OH / HBTU, ii. 20% piperidine

i. Fmoc -aminoethyl photolinker (pl) / HBTU, ii. 20% piperidine

1) i. Fmoc-Ala-OH / HBTU
2) Benzoylchloride / pyridine in DCM
3) 1% TFA / 5% TIS in DCM
4) dabcyl-OSu/DIEA in DMF
5) 20% piperidine

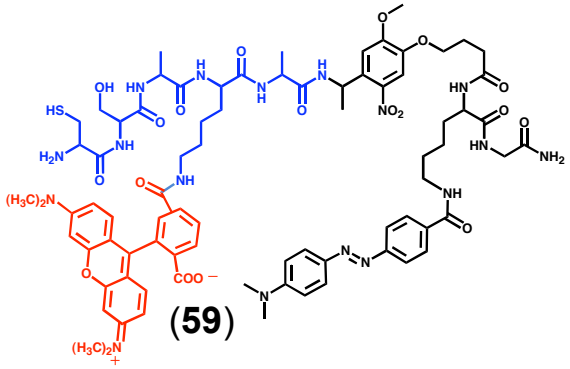
H-Ala-pl-Lys(Dabcyl)-Gly-Resin

1) i. Fmoc-Lys(Mtt)-OH / HBTU, ii. 20% piperidine
2) i. Fmoc-Ala-OH / HBTU, ii. 20% piperidine
3) i. Fmoc-Ser(tBu)-OH / HBTU, ii. 20% piperidine
4) Boc-Cys(StBu)-OH / HBTU

Boc-Cys(StBu)-Ser(tBu)-Ala-Lys(Mtt)-Ala-pl-Lys(Dabcyl)-Gly-Resin

1) 1% TFA / 5% TIS in DCM
2) 5(6)-TAMRA / HBTU in DMF
3) PBu_3 / H_2O in DMF / DCM
4) TFA, scavengers

H-Cys-Ser-ALa-Lys(5(6)-TAMRA)-Ala-pl-Lys(Dabcyl)-Gly-NH_2



(**59**)

Figure 4.8. Synthesis of peptide 59. A. Synthetic scheme and structure of final product. **B.** RP-HPLC (30-50%B) of peptide **59**. The peptide is a mixture of two regioisomers since 5(6)-TAMRA was used in the synthesis. **C.** ESI-MS of peptide **59**. Expected M_r is 1606.7 Da.

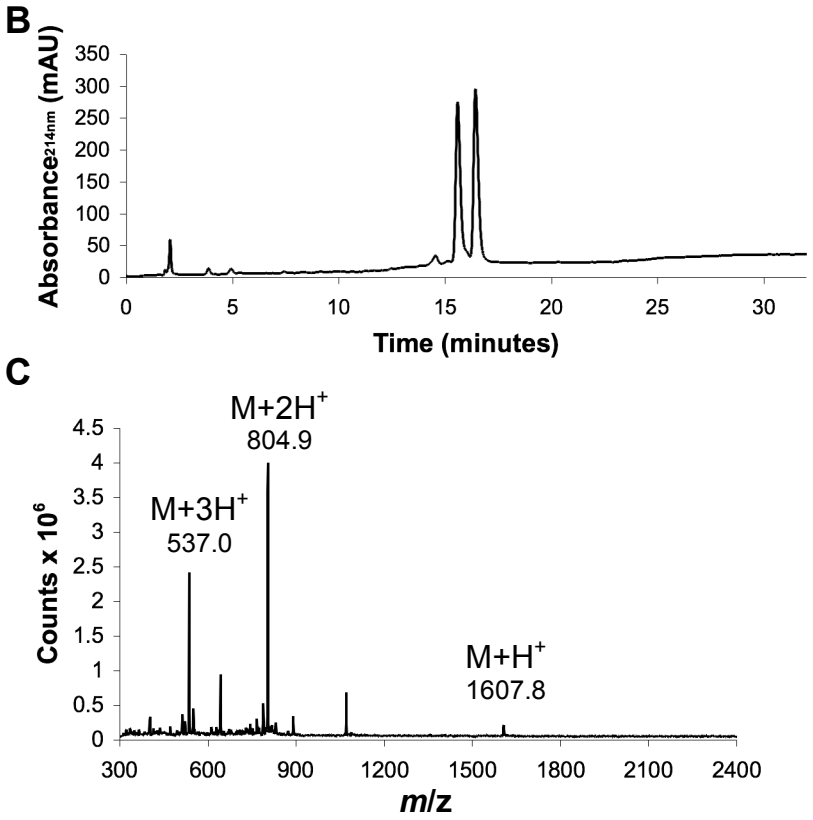


Figure 4.8, continued.

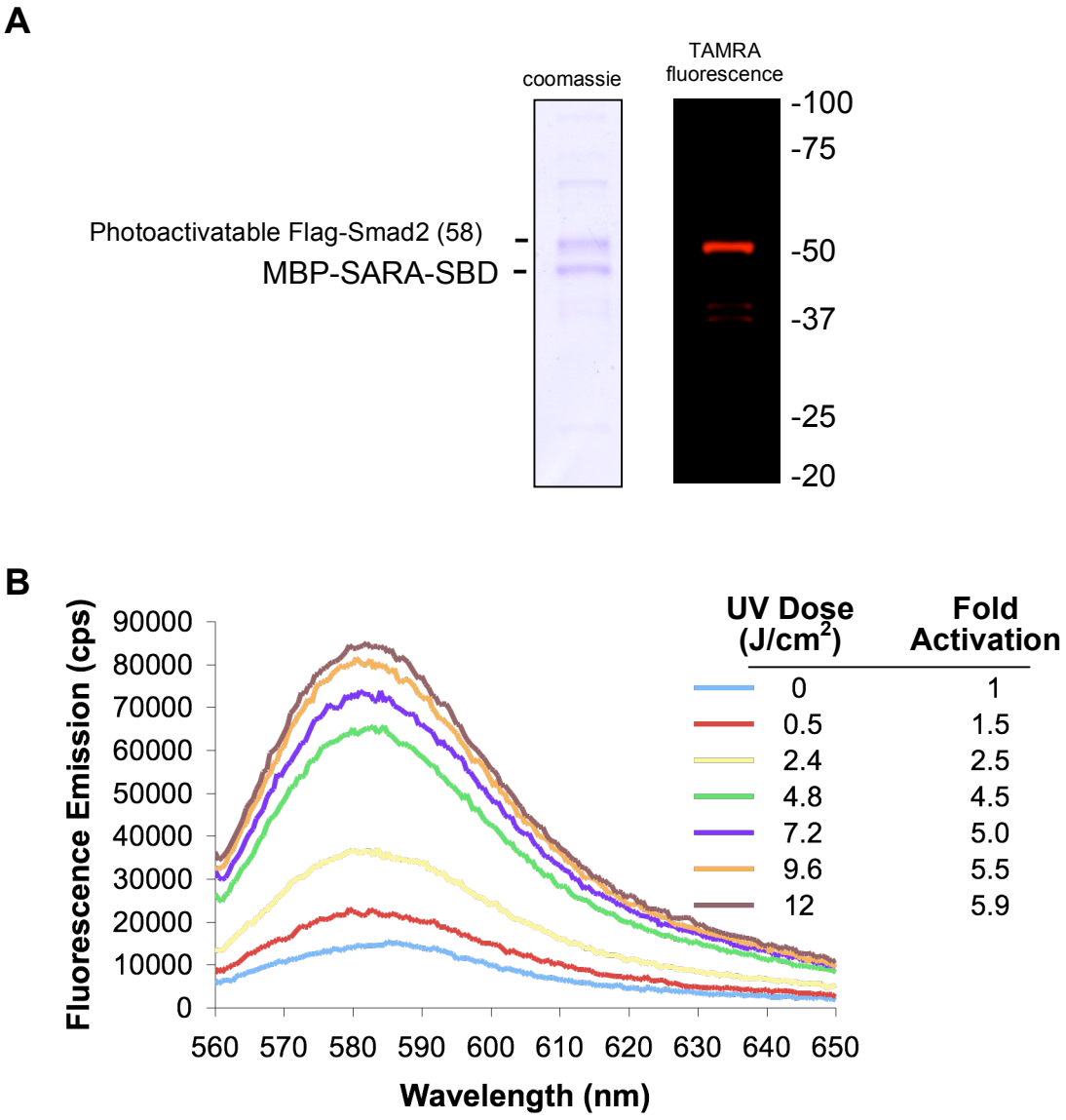


Figure 4.9. Semi-synthesis and photoactivation of non-phosphorylated photoactivatable Flag-Smad2 (58). A. SDS-PAGE of photoactivatable Flag-Smad2 prepared by EPL in complex with MBP-SARA-SBD. **B.** Fluorescence spectra of photoactivatable Flag-Smad2 *in vitro* before and after irradiation with UV light at 365 nm. The energy of light applied and fold increase of Smad2 fluorescence (582 nm) after irradiation are indicated.

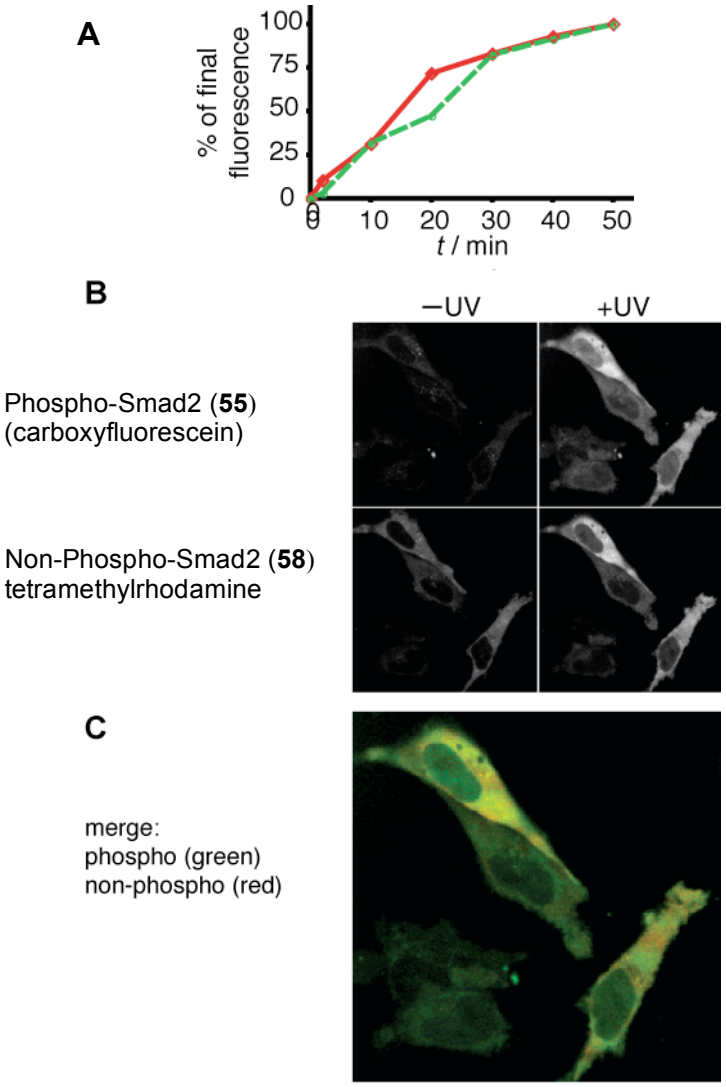


Figure 4.10. Simultaneous activation of two different forms of caged full-caged Smad2 in the same live cells. A. *In vitro* uncaging kinetics of caged phosphorylated Smad2 labeled with carboxyfluorescein (**55**) and its non-phosphorylated counterpart labeled with tetramethylrhodamine (**58**). Proteins **55** (dashed green line) and **58** (solid red line) were irradiated with light at 365 nm (Irradiance = 4 mW/cm^2) and fluorescence spectra were recorded at each time point. The percent of total uncaging was determined by dividing the fluorescence emission (at $\lambda =$ 520 nm for carboxyfluorescein; $\lambda =$ 582 nm for tetramethylrhodamine) at each time point by that of the final time point. **B and C.** Phosphorylated (**55**) and non-phosphorylated (**58**) caged forms of Smad2 were co-injected into the cytoplasm of live HeLa cells. Confocal images before (-UV) and 5 minutes after (+UV) irradiation (60 second pulse) with UV light at 360 nm are shown.

observation is consistent with the known localization behavior of phosphorylated and non-phosphorylated Smad2 (Macias-Silva et al., 1996; Massague et al., 2005; Schmierer and Hill, 2005).

4.7. Summary

In the experiments presented here, semi-synthetic protein chemistry methodology was used to construct the signaling protein Smad2 bearing two activating PTMs, a fluorescent donor-quencher pair and a photo-cleavable trigger. The successful semi-synthesis of this ornately modified ~50 kDa protein illustrates the broad scope of protein ligation technology. In principle, this approach can be applied to other proteins that are modulated by phosphorylation and their non-phosphorylated counterparts, especially if structural information of the protein is known. The requirements are: i. chemical access to the phosphorylation sites, ii. installation of a fluorophore at a position where it does not significantly affect protein structure and function (typically a location on the surface of the protein), iii. installation of a quencher distal to a photolinker near to the site of the fluorophore. These moieties can be installed on amino acid side chains or at the termini of the protein or even on functional groups of other classes of biomolecules (Tang and Dmochowski, 2005).

The direct comparison of imaging data on different proteins acquired by observing them in separate cells is prone to artifacts due to cell-to-cell heterogeneity. To eliminate this problem, it is desirable to image different proteins simultaneously in the same cell (Schultz et al., 2005). Such multiparameter experiments are often carried out by co-expressing two or more proteins fused to different fluorescent proteins (FPs).

However, the activation state of a protein is uncoupled from the fluorescence of the FP to which it is fused. Therefore, dynamic FRET-based methods that report the activation of two or more proteins in the same cell have been devised (Brumbaugh et al., 2006; Peyker et al., 2005; Schultz et al., 2005). Our approach uniquely complements these FRET-based approaches in that it allows for simultaneous activation and subsequent selective monitoring of more than one protein in the same cell. In this report we have demonstrated the simultaneous monitoring of the activated form of a posttranslationally modified protein and its non-modified, inactive form. In principle, this approach can be extended to simultaneous activation and monitoring of more than one differentially posttranslationally modified form of the same protein and also to posttranslationally modified forms of two or more different proteins. Future studies will be aimed at identifying optimal fluorophore/quencher pairs and uncovering new biological insights.

Chapter 5: A search for proteins whose interaction with Smad2 is regulated by phosphorylation identifies PRMT5 as a potential Smad2 binding partner

5.1. Background

Smad2 function is modulated by interaction with other proteins (Massague et al., 2005). Perhaps the most interesting of these interactions are those that are modulated by Smad2 phosphorylation. Since phosphorylation is a dynamic post-translational modification, it follows that interactions that are regulated by phosphorylation are also dynamic. Dynamic interactions allow cells to respond rapidly to changes in their environments and are therefore vital to the maintenance of cellular homeostasis (Pawson and Nash, 2003).

A variety of protein-protein interaction screens have been applied in the identification of novel Smad-interacting proteins. For example, He et al. prepared recombinant baits for *in vitro* interaction assays containing the linker region and MH2 domain of Smad2 fused to GST (He et al., 2006). The critical serine residues of the Smad2 C-terminus that are phosphorylated in response to TGFβ were both mutated to aspartic acid (Asp). This mutation was employed to mimic the naturally occurring phosphoserines (pSer) since the carboxylate of the Asp side chain resembles the charge, polarity, and size of a phosphate. Incubation of this bait with HeLa lysates and its subsequent co-precipitation with interacting proteins led to the isolation of TIF1γ. Further analysis of the TIF1γ-Smad2/3 interaction revealed a role for TIF1γ as an alternative to Smad4 in TGFβ-regulated signaling and showed that this interaction was vital for differentiation of hematopoeitic precursors into erythrocytes.

5.2. Design of a novel probe for the discovery of phosphorylation-regulated Smad2 binding partners

Taking the general framework of He et al. for the identification of Smad2 interaction proteins one step further requires incorporation of a more authentic pSer mimetic at positions 465 and 467 of Smad2. To accomplish this, EPL was utilized to install native pSer (**32**) in these positions (Figure 5.1A). pSer was chosen since it is the natural modification of TGFβ-receptor phosphorylated Smad2. However, pSer suffers from the drawback that its phosphate group is labile to phosphatases that are likely to be active in the lysates used in interaction screens. This can be minimized to some extent by the inclusion of phosphatase inhibitors, but this is not ideal since their inclusion does not guarantee that potential Smad2 phosphatases would be completely inhibited. Furthermore, at the time this work began, the identity of Smad-phosphatases were unknown and their identification was the central motivation in the undertaking of this project. In principle, the interaction between Smad2 and a potential phosphatase could itself be phospho-specific. It follows that if one or both of the phosphates recognized by a potential Smad2 phosphatase were removed, then the binding affinity of the Smad2/phosphatase complex would weaken, likely to a degree that would not support identification in an interaction screen (Figure 5.1B).

To surmount this obstacle, the non-hydrolyzable pSer analogue phosphonomethylenealanine (Pma; **33**) was employed for its superior mimicry of pSer in comparison to Asp or Glu and its stability to removal by phosphatases (Figure 5.1A). The phosphonate group has been used by others to mimic phosphate on a variety of molecules

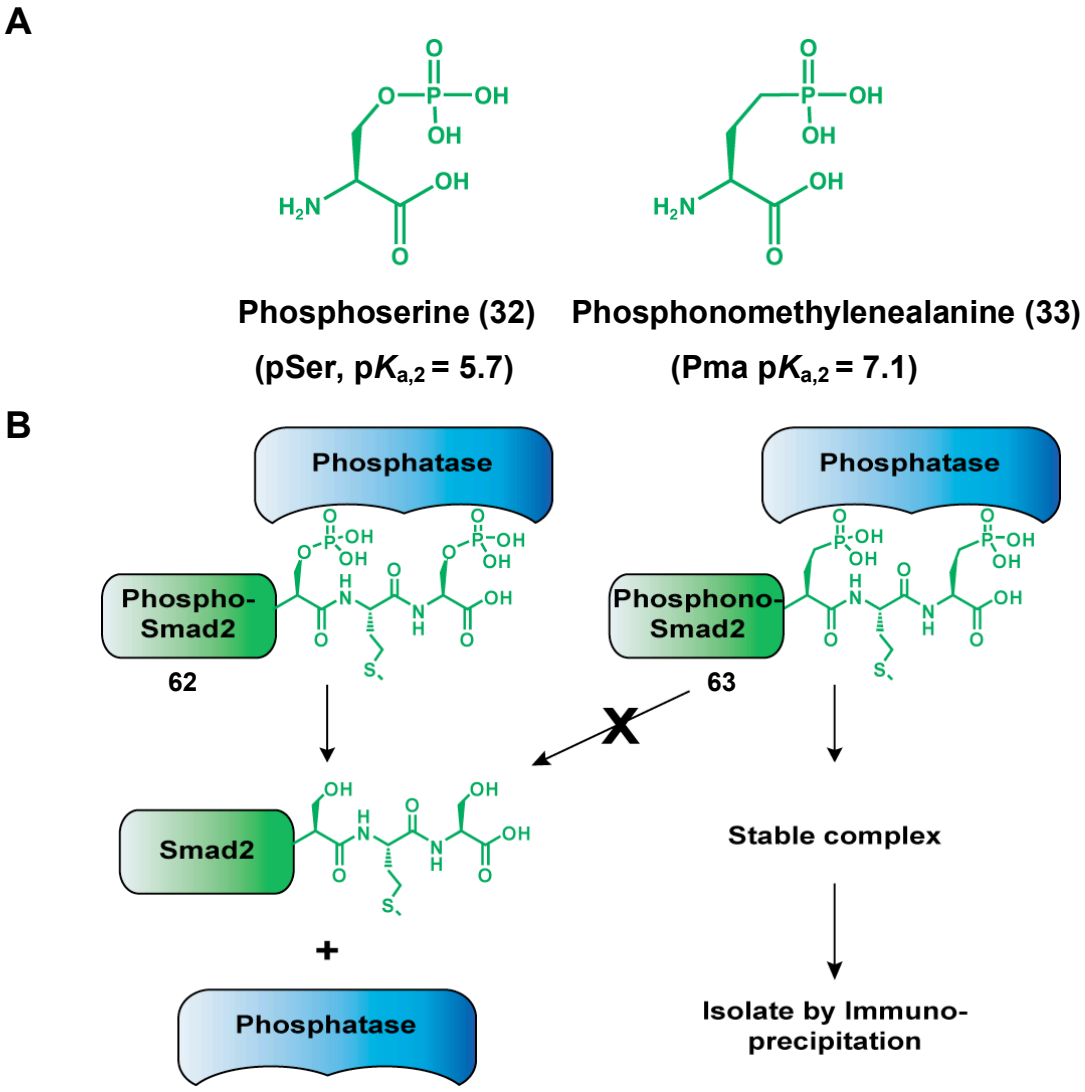


Figure 5.1. Principle of Phosphonylated Smad2 Protein-Protein Interaction Screen. **A.** pSer and Pma differ only in the replacement of the γ-oxygen of pSer with a methylene group. **B.** Phosphono-Smad2-MH2 (**63**) may interact more strongly with a putative Smad phosphatase (as compared to phospho-Smad2-MH2 (**62**)) since its phosphono groups can not be removed by the phosphatase.

including, but not limited to, metabolites, enzyme inhibitors, and proteins (Berkowitz et al., 2000; Lee et al., 2005; Zheng et al., 2003). The key to its utility is that the bridging oxygen linking the biological molecule, in this case serine of Smad2, to phosphorous is replaced with a methylene (-CH_2-) group, yielding a phosphonate. This substitution does not dramatically alter the steric and electronic properties of the resulting phosphonate as compared to the phosphate and ensures that the phosphonate is not labile to phosphatases (Figure 5.1A,B).

5.3. Synthesis of Pma-containing peptides

Cole and co-workers have reported the incorporation of Pma into synthetic peptides prepared by Fmoc-based SPPS without protection of the phosphonate moeity (Zheng et al., 2003). Commercially available D,L-Pma was obtained and protected with Fmoc using Fmoc-succinimidyl carbonate. The resulting Fmoc-D,L-Pma (hereafter Fmoc-Pma) was used in an attempted synthesis of NH_2-Cys-Ser-Pma-Met-Pma-OH using the benzylic hydroxyl of Wang resin as the anchor to the solid support. No attempt was made to separate the enantiomers of Fmoc-Pma since the final peptide product would contain 25% of the configurationally correct NH_2-Cys-Ser-L-Pma-Met-L-Pma-OH. Furthermore, an additional 50% of the peptide product would contain L-Pma in one of the two positions. This was judged to be sufficient for initial studies. Unexpectedly, Wang resin could only be loaded with Pma to 5% of the expected loading stoichiometry, making the synthesis by this strategy impractical. This was likely due to intramolecular nucleophilic attack of the unprotected phosphonate on the ester linking the carboxylic

acid of the loaded Fmoc-Pma to the benzylic hydroxyl of Wang resin, which releases Fmoc-PMA from the solid support.

As an alternative, the peptide was re-designed to contain a C-terminal amide in place of the wild-type carboxylic acid. This substitution is not ideal since the C-terminal carboxylic acid is known to make contacts in some, but not all, Smad2 protein-protein interactions (Chacko et al., 2004; Qin et al., 2001; Wu et al., 2001). Nevertheless, this route was pursued since its successful completion would result in a viable reagent for the interaction screen. Using Rink amide resin, the loading of Fmoc-Pma was improved to 75%, a typical value for loading of peptidyl resins. This was possible in comparison to Wang resin since the linkage between Fmoc-Pma and Rink amide resin is an amide, which is stable to nucleophilic attack by the unprotected phosphonate.

5.4. Semi-synthesis and *in vitro* characterization of Smad2-MH2 interaction baits containing pSer and Pma

Flag-Smad2-MH2 (241-462; hereafter Flag-Smad2-MH2) was expressed as a fusion to GyrA-CBD (**60** = Flag-Smad2-MH2-GyrA-CBD), purified, and processed as described for Smad2-MH2-GyrA-CBD (see section 2.4 and methods) to generate Flag-Smad2-MH2-COSR (**61**). The synthetic peptides NH_2-Cys-Ser-pSer-Met-pSer-OH (**68**) and NH_2-Cys-Ser-Pma-Met-Pma-$CONH_2$ (**71**) were each ligated to Flag-Smad2-MH2-COSR (**61**) to form the corresponding semi-synthetic proteins, hereafter referred to as Flag-Smad2-MH2-2P (**62**) for the doubly phosphorylated protein and Flag-Smad2-MH2-2Pma (**63**) for the doubly phosphonylated protein (Figure 5.2A).

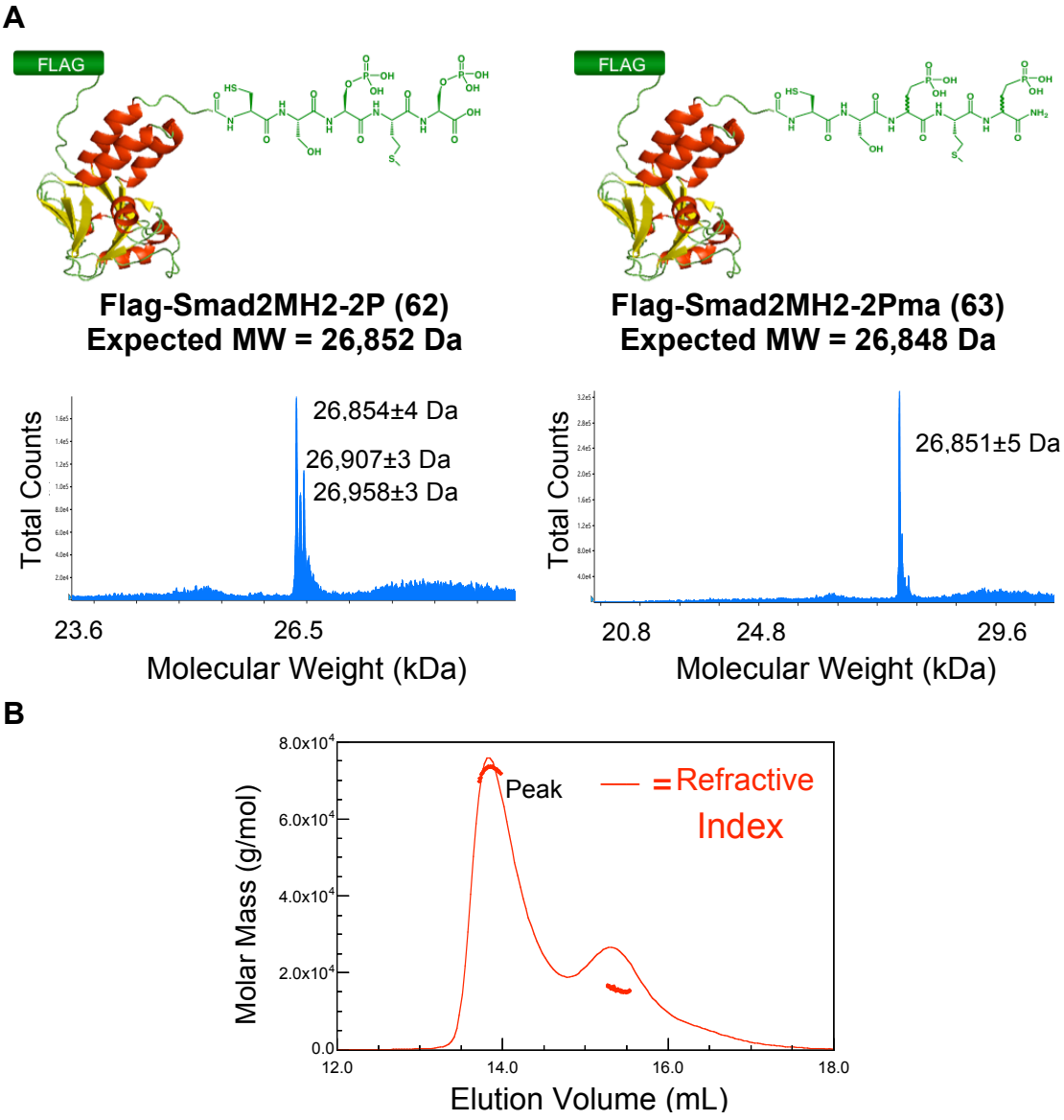


Figure 5.2. *in vitro* characterization of Flag-Smad2MH2 baits. A. Flag-Smad2MH2-2P (**62**) and Flag-Smad2-2Pma (**63**) were prepared by EPL and verified by ESI-MS. The peaks at 26,907 Da and 26,958 Da in the Flag-Smad2-MH2-2P (**62**) spectrum represent adducts with one and two Fe^{3+} ions, respectively. **B.** Flag-Smad2-2Pma (**63**) was analyzed by SEC-MALLS. Peak 1 contained Flag-Smad2-2Pma (**63**). The expected molar mass of a Flag-Smad2-2Pma (**63**) homo-trimer is 80,544 Da. A molar mass of 72,610 ± 87 Da was observed. The smaller peak at ~15.3 mL is not proteinaceous.

These two proteins were analyzed by ESI-MS, which verified that they were assembled as expected (Figure 5.2A). SEC analysis of Flag-Smad2-MH2-2P (**62**) confirmed that it formed homo-trimers as previously described for Smad2-MH2-2P (**41**). Surprisingly, SEC analysis of Flag-Smad2-MH2-2Pma (**63**) suggested that this protein formed homo-trimers as well, despite the presence of racemic Pma and a C-terminal amide.

Due to the unexpected nature of this result, Flag-Smad2-MH2-2Pma (**63**) was analyzed by SEC-MALLS to determine the solution molecular weight of the protein complex. This analysis confirmed that Flag-Smad2-MH2-2Pma (**63**) formed higher order homo-oligomers. However, the average molar mass of these oligomers was 72,610 ± 87 Da, while the expected molar mass of a Flag-Smad2-MH2-2Pma (**63**) homo-trimer is 80,544 Da (Figure 5.2B). Note, previous SEC-MALLS studies indicated that Smad2-MH2-2P (**41**) containing exclusively L-pSer at positions 465 and 467 and a C-terminal acid exhibits the expected solution molar mass for a homo-trimer (see section 2.5). The solution molar mass determined for Flag-Smad2-MH2-2PMA (**63**) may indicate that the protein exists in a dynamic equilibrium that exchanges quickly between monomeric, homo-dimeric, and homo-trimeric forms as compared to the timescale of SEC analysis, which is approximately 30 minutes. This feature may be advantageous for discovering phospho-specific Smad2 interacting proteins since many of these proteins may bind to a transiently formed Smad2-2P monomer or homo-dimer.

5.5. Interaction screen using Flag-Smad2-MH2 baits containing pSer and PMA identifies PRMT5 as a potential Smad2 interacting protein[1]

HeLa nuclear lysates were mixed with Flag-Smad2-MH2-2P (**62**) or Flag-Smad2-MH2-2Pma (**63**) in the presence of an excess of Smad2-MH2-0P (**64**). The non-phosphorylated non-epitope-tagged Smad2-MH2-0P (**64**) was added as a competitor in an attempt to enrich proteins that selectively bind the phosphorylated protein. Proteins bound to Flag-Smad2-MH2-2P (**62**) or Flag-Smad2-MH2-2Pma (**63**) were then isolated by co-immunoprecipitation and were separated by SDS-PAGE (Figure 5.3A). Several protein bands were differentially present in alternate lanes. Each lane of the gel was cut into pieces and the proteins in each piece were extracted, trypsinized and analyzed by MALDI-Qq-TOF mass spectrometry (Figure 5.3A). This analysis revealed a protein that appeared in the immunoprecipitates only when Flag-Smad2-MH2-2Pma (**63**) was used as bait. This protein, Protein Arginine Methyltransferase 5 (PRMT5) is a member of the PRMT family which contains at least 7 other mammalian members (Figure 5.3B) (Bedford and Richard, 2005). These proteins transfer methyl groups from *S*-adenosylmethionine to specific arginine residues in proteins (Figure 5.3C).

PRMT5 is a type II transferase, which indicates that it is able to generate symmetrical ω–N^G,N'^G-dimethyl arginine (sDMA) in proteins (Figure 5.3C) (Branscombe et al., 2001; Rho et al., 2001). As an obligate intermediate along the pathway from arginine to sDMA is monomethyl arginine, PRMT5 is also capable of forming this modified amino acid in proteins. Numerous substrates of PRMT5 have been

[1] The mass spectrometry analysis was performed in collaboration with Matthew Sekedat in the laboratory of Professor Brian Chait.

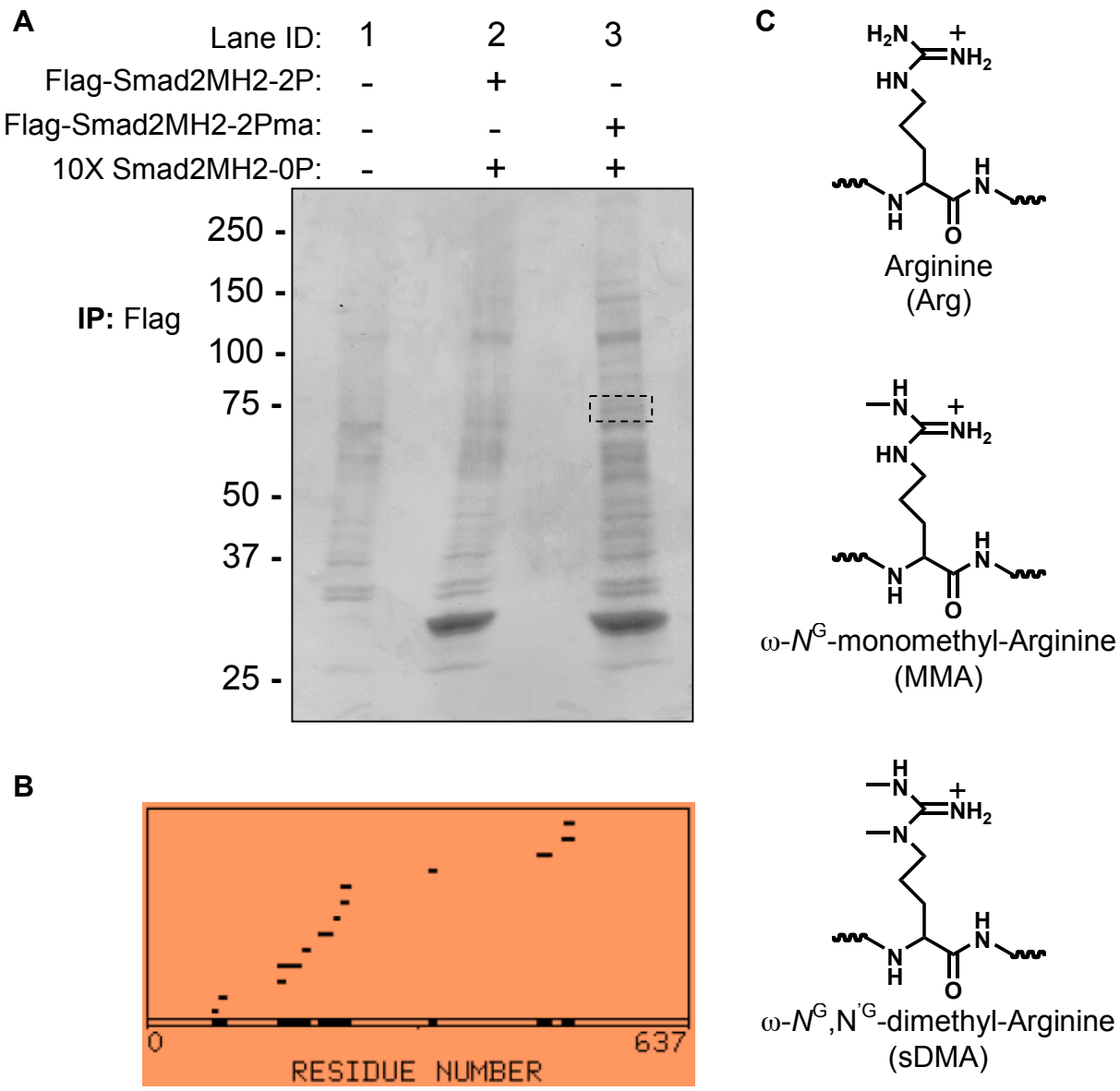


Figure 5.3. Identification of PRMT5 as a potential Smad2 interacting protein. A. Flag-Smad2MH2-2P (**62**) and Flag-Smad2-2Pma (**63**) were incubated with HeLa nuclear extract in the presence of competitor untagged non-phosphorylated Smad2MH2-0P. Immunoprecipitation and subsequent MS-based identification revealed PRMT5 in lane 3. The region of the gel from which PRMT5 was isolated is highlighted by the dashed rectangle. **B.** Diagram of PRMT5 tryptic peptides (represented by black bars at their position in the sequence) identified in the mass spectrometry analysis. The overall sequence coverage of this 637 amino acid protein was 23%. **C.** PRMT5 catalyzes the formation of MMA and sDMA at Arg residues in proteins.

discovered and, in some cases, their methylation has been linked to a physiological function (Bedford and Richard, 2005). PRMT5-mediated methylation of arginines in the Sm ribonucleoproteins is required for their proper integration into snRNPs, components of the pre-mRNA splicing machinery (Friesen et al., 2001; Meister et al., 2001). Methylation of specific arginines in the N-terminal tails of histones H3 and H4 by PRMT5 has been shown to regulate gene expression (Ancelin et al., 2006; Pal et al., 2004; Pal et al., 2003). PRMT5 also methylates FCPI, a phosphatase that dephosphorylates the CTD of RNA polymerase II and is required for proper transcription elongation and RNA polymerase II recycling (Amente et al., 2005).

From these and other studies, it is clear that PRMT5 has a multitude of cellular functions. This makes elucidation of the significance of the Smad2/PRMT5 interaction of particular interest as it may represent a point of integration for Smad2 with any of these other processes.

5.6. Validation of the Smad2/PRMT5 interaction

Proteins identified in a MS-based protein-protein interaction screen may be false positives. To validate PRMT5 as an authentic Smad2 interacting protein, we turned to co-immunoprecipitation followed by western blotting. Since PRMT5 was initially identified in the MS experiments when FLAG-Smad2-MH2-2Pma (**63**) was used as bait, it was expected that PRMT5 would preferentially bind phosphorylated Smad2 variants over non-phosphorylated variants. To test this, co-immunoprecipitation and immunoblotting experiments were performed using Flag-tagged Smad2 variants and HeLa nuclear lysates (Figure 5.4A). This analysis showed that Flag-Smad2 variants co-precipitated with

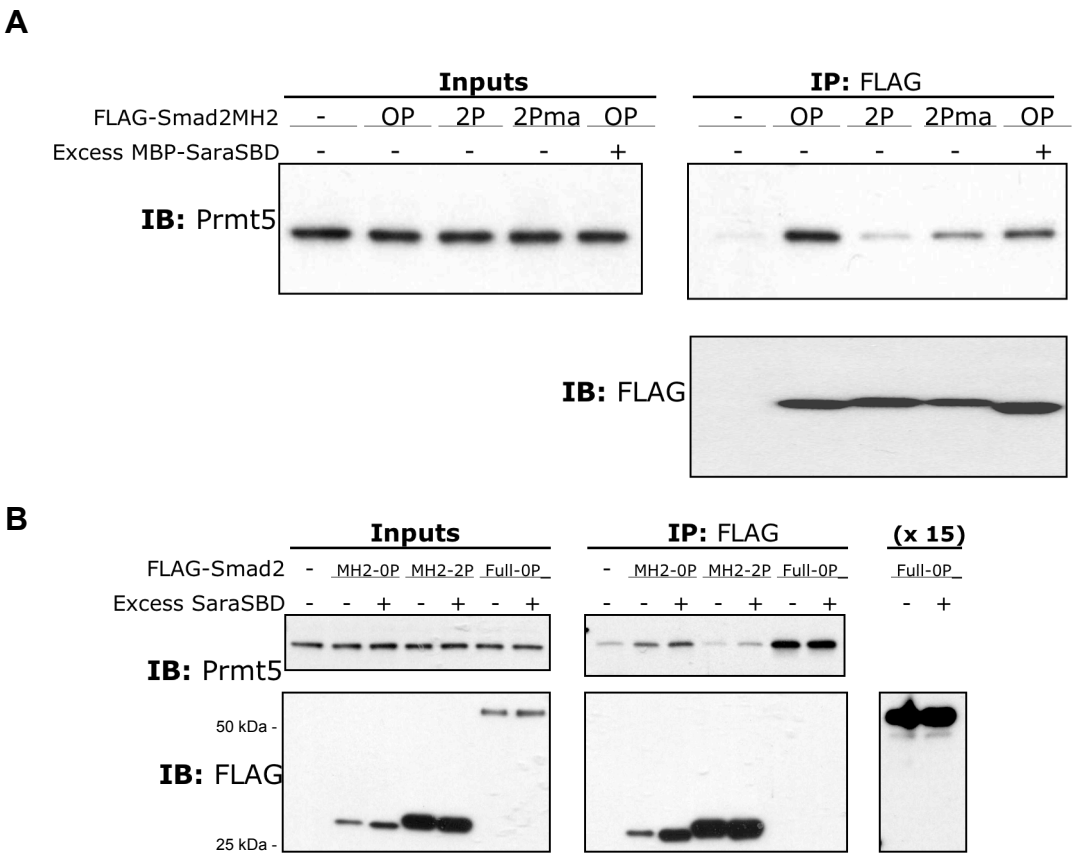


Figure 5.4. PRMT5 interacts preferentially with non-phosphorylated, full-length Smad2. Flag-Smad2MH2 variants (panels **A** and **B**) and non-phosphorylated full-length Flag-Smad2 **(B)** were incubated with HeLa nuclear extract and the resulting complexes were immunoprecipitated with anti-FlagM2 agarose followed by immunoblotting as indicated. The full-length Flag-Smad2 immunoprecipitates were analyzed again at higher loading to show that full-length Flag-Smad2-0P was present. No significant modulation of the interaction was detected when excess SaraSBD was included. Protein IDs: Flag-Smad2-MH2-0P = **60**, Flag-Smad2-MH2-2P = **62**, Flag-Smad2-MH2-2Pma = **63**, Full-length Flag-Smad2-0P = **57**.

endogenous PRMT5, but that the interaction was strongest with Flag-Smad2-MH2-0P (**61**), weakest with Flag-Smad2-MH2-2P (**62**), and of intermediate strength with Flag-Smad2-MH2-2Pma (**63**). In the mass spectrometry screen discussed in section 5.5, PRMT5 was found to co-precipitate with Flag-Smad2-MH2-2Pma (**63**) in the presence of a ten-fold excess of non-tagged Smad2-MH2-0P (**64**) competitor protein. Perhaps the levels of endogenous PRMT5 in the HeLa nuclear lysates used in the screen were high enough to saturate the excess binding sites of Smad2-MH2-0P (**64**), leaving the remainder available for binding to Flag-Smad2MH2-2Pma (**63**).

Analogous experiments demonstrated that full length Flag-Smad2 (**57**) interacted considerably stronger with PRMT5 than did Flag-Smad2-MH2 (**62**; Figure 5.4B). This interesting observation has many potential explanations. It is possible that the binding site for PRMT5 on Smad2 is bidentate and that one part is found in the MH2 domain while the other is found in the linker region or the MH1 domain. Alternatively, the structure of the MH2 domain may be modulated in the presence of the remainder of the protein such that the binding site for PRMT5 in the MH2 domain is more accessible. This is quite possible as it is known that the MH1 and MH2 domains of Smad2 can interact with each other directly (Hata et al., 1997).

To determine if the MH1 domain of Smad2 is able to interact with PRMT5, His_6-SUMO-Smad2-MH1 (**65**; containing residues 1–185 of Smad2) was mixed with HeLa cell nuclear extracts containing endogenous PRMT5. The four extracts used in this study differ from one another. The parent nuclear extract (NE) was prepared in the same manner as the nuclear extract used in the interaction screen (section **5.5** and Methods section 7.33). The lysate referred to as NE_C was further clarified by high-speed

centrifugation at 100,000 x *g* before being added to His_6-SUMO-Smad2-MH1 (**65**). NE_{TX} and NE_D were treated with 1% Triton-X-100 or DNAse I, respectively, before being added to His_6-SUMO-Smad2-MH1 (**65**). Following co-precipitation of His_6-SUMO-Smad2-MH1 (**65**) and its interacting proteins with Ni^{2+}-NTA beads, western blotting was carried out to determine if PRMT5 interacted with the MH1 domain of Smad2 (Figure 5.5A). PRMT5 from NE and NE_C co-precipitated with the MH1 domain with equal efficiency. However, co-precipitation of the MH1 domain with endogenous PRMT5 was diminished when NE_{TX} or NE_D were used as the source of PRMT5. This decreased efficiency of interaction in the presence of Triton-X-100 indicates that the interaction is sensitive to non-ionic detergents, which is a common finding for protein-protein interactions. More interesting is the slightly decreased efficiency of interaction with PRMT5 from NE_D. It is possible that an interaction between Smad2 and PRMT5 occurs only in the presence of DNA or that DNA serves as a scaffold to increase the efficiency of interaction either directly or indirectly. The potential significance of this finding will be discussed in chapter 6.

All of the above interaction studies used recombinant or semi-synthetic Smad2 variants as bait to test for Smad2 interaction with endogenous PRMT5. To perform the reverse co-precipitation, Flag-PRMT5 was expressed in HEK293T cells. Immuno-precipitation of total cell lysates of these cells with anti-FlagM2 agarose revealed that endogenous Smad2 co-precipitated with Flag-PRMT5 (Figure 5.5B). As a negative control, HEK293T cells were transfected with an irrelevant plasmid. Smad2 was not detected when lysates from these cells were subjected to anti-FlagM2 immunoprecipitation. Several attempts were made to co-immunoprecipitate endogenous

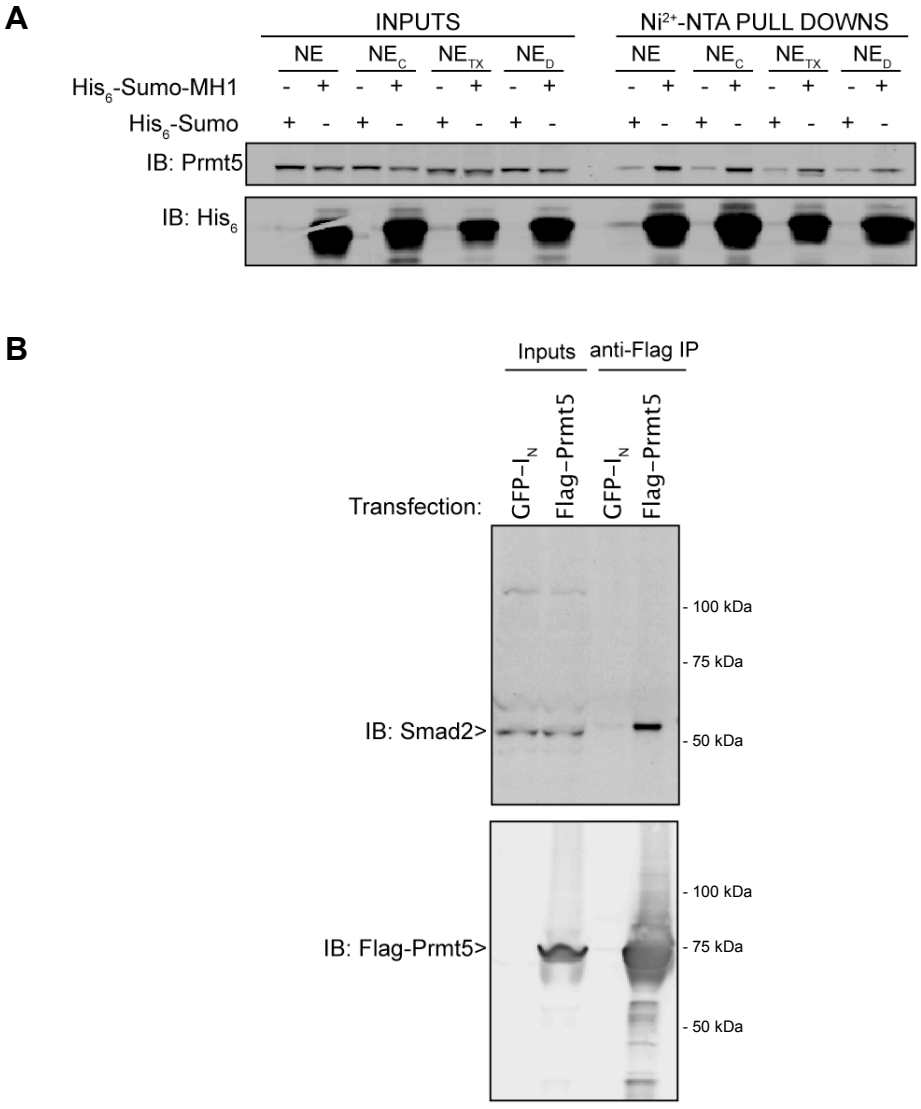


Figure 5.5. PRMT5 interacts with the MH1 domain of Smad2 and endogenous Smad2 co-immunoprecipitates with Flag-PRMT5 expressed in HEK293T cells. A. His_6-Sumo-Smad2-MH1 (**65**) or His_6-Sumo (**66**) control proteins and their interacting proteins from HeLa cell nuclear extracts were precipitated with Ni^{2+}-NTA beads and analyzed by SDS-PAGE and immunoblotting for PRMT5. The parent nuclear extract (NE) was alternately centrifuged at 100,000 x *g* (NE_c), treated with 1% Triton-X-100 (NE_{TX}), or treated with DNAse I (NE_D). **B.** HEK293T cells were transfected with a plasmid encoding Flag-PRMT5 or an irrelevant plasmid (GFP-I_N) as a control. Anti-FlagM2 immunoprecipitates were analyzed by SDS-PAGE and immunoblotting for Smad2.

Smad2 and endogenous PRMT5 using antibodies against both proteins. Endogenous interaction could not be detected in these experiments. In the future, experiments using other cell lines or tissues, alternative antibodies for immunoprecipitation, and alternative lysis conditions could be performed to verify this interaction for endogenous proteins.

5.7. Attempts to determine if Smad2 is a direct substrate of PRMT5

The most straightforward hypothesis regarding the function of this interaction is that Smad2 may be a substrate for PRMT5. To test this, the Flag-Smad2MH2 bands were excised from the gels originally used for the MS-identification of PRMT5 and submitted for further trypsinization and MS-analysis to determine if any of the resulting peptides corresponding to Smad2 had the +14 Da signature of methylation. No methylated peptides were found that did not exist in controls. Next, Smad variants were tested in an *in vitro* methylation assay using *S*-Adenosyl-[*methyl*-^{14}C] methionine as the methyl donor in the presence of immuno-purified Flag-PRMT5. Methylation of Smad2MH2 variants was not observed.

5.8. Summary

A screen for proteins that preferentially interact with phosphorylated Smad2 was carried out using Smad2 variants prepared by expressed protein ligation containing phosphoserine as well as the stable phosphate analogue phosphonomethylenealanine. This resulted in the identification of PRMT5 as a potential interaction partner for Smad2. The observation that PRMT5 appears to bind non-phosphorylated Smad2 more efficiently than phosphorylated Smad2 is unexpected since the screen was designed for the

discovery of phospho-specific binding proteins. It is noted that co-immunoprecipitation studies are subject to false positive results. Therefore, more work is required to further validate that the interaction between Smad2 and PRMT5 is authentic and physiologically significant. The fact that the interaction as studied so far has been shown to be negatively modulated by Smad2 phosphorylation is intriguing and suggests that the interaction may play a role in TGFβ signaling. The possible implications of these results and experiments to be done in the future will be discussed in chapter 6.

Chapter 6: Discussion

6.1. Development of protein caging strategies

To measure molecular processes that occur rapidly, investigators require tools that allow equally rapid perturbation of the molecules under study. Since the introduction of caged ATP in 1978 (Kaplan et al., 1978), numerous studies have used caging strategies to quickly activate a molecule of interest, both *in vitro* and in live systems (Mayer and Heckel, 2006). Due to difficulty in their preparation, reports on caged proteins have lagged behind that of caged small molecules. In the last decade, protein engineering techniques have become increasingly sophisticated and have enabled the preparation of increasingly complex proteins bearing several unnatural functionalities. Among these are caged proteins, which have yielded insight into several biological systems (Lawrence, 2005). In this thesis, two protein caging strategies have been described and both were applied to the transcription factor and signal transduction protein Smad2.

In the first strategy, we took advantage of the fact that Smad2 is activated by phosphorylation of serines 465 and 467 (Macias-Silva et al., 1996; Massague et al., 2005). The phosphoserine residues of one activated Smad2 molecule make important contacts to residues in an adjacent Smad subunit in the context of homo- and hetero-oligomeric Smad complexes (Chacko et al., 2004; Wu et al., 2001). Therefore, if these phosphate residues are blocked by covalently attached bulky groups, we expected that the protein would behave as if it was not phosphorylated and hence, it would be inactive. To examine this, we installed *o*-nitrobenzyl caging groups onto phosphoserines 465 and 467 of Smad2-MH2 using the protein semi-synthesis technique expressed protein ligation (EPL) (Muir, 2003). In the absence of UV light, the resulting caged phosphoprotein (**40**)

behaved as if it was not phosphorylated (Figures 2.9-2.11). Upon irradiation with UV light, the caging groups were removed and the protein had the properties expected of the active, doubly-phosphorylated Smad2-MH2 domain (Figures 2.9-2.11). This study was the first published report of a protein caged on a phosphoaminoacid (Hahn and Muir, 2004).

In the second strategy we attached a bulky caging group to the C-terminus of Smad2-MH2. While the C-terminus of Smad2 is not subject to reversible posttranslational modification like serines 465 and 467, it is nevertheless important for Smad2 activation as it also makes contacts with residues in adjacent subunits in the context of Smad homo- and hetero-oligomers (Figure 3.2) (Chacko et al., 2004; Wu et al., 2001). Therefore, the presence of the caging group on the C-terminus led to inactivation of Smad2 until it was removed by UV light (Figure 3.6). At this point, the uncaged protein is identical to and exhibited the properties of the active, doubly-phosphorylated protein (Figure 3.6). After demonstrating that this caging method was viable, we interfaced it with a fluorescence activation strategy to enable the simultaneous activation and fluorescence monitoring of the protein. This involved installation of the fluorophore carboxyfluorescein at position 466 of Smad2-MH2 and installation of the quencher dabcyl distal to the caging group to quench the fluorescence of carboxyfluorescein while the caging group was intact (**47**, Figure 3.3). Upon irradiation with UV light, the caging group and quencher were removed from the protein, leading to both protein activation and restoration of robust carboxyfluorescein fluorescence (Figures 3.6 and 3.7).

Both of these caging strategies were validated *in vitro* and in cellular contexts. For imaging of Smad2 in live cells, we elected to pursue the C-terminal caging strategy to

selectively monitor only the active Smad2 variants. This is the chief advantage of this strategy over that of caged phosphates. However, for certain applications it may be beneficial to use the caged phosphate strategy. For example, in experiments requiring prolonged incubation of a caged protein in a living system, such as the experiments involving *Xenopus* embryos (Figure 3.9), caged phosphates may perform better than the C-terminal caging strategy owing to their greater stability to cellular phosphatases. In the C-terminal strategy, the activating phosphates are unprotected and, therefore, are subject to dephosphorylation by phosphatases. In comparison, caged phosphates are unnatural phosphodiesters and, as such, are not substrates for endogenous protein phosphatases that catalyze hydrolysis of phosphate from phosphomonoesters in proteins.

6.2. Comparison of caging strategies to protein visualization using GFP and its variants

Fluorescence imaging is the method of choice when observing the subcellular localization of a protein in live cells. For this purpose, the investigator has many options when choosing a fluorophore. The most commonly used fluorophores are GFP and its variants (Shaner et al., 2005). The primary advantage of these fluorophores is that they are genetically encodable and, therefore, do not require any chemical synthesis or *in vitro* protein labeling before use. However, these fluorophores are large proteins that may impact the behavior of the protein to which they are fused. An important recent advance in the fluorescent protein field is the development of photoactivatable GFP (paGFP) and other photoactivatable fluorescent proteins (Ando et al., 2004; Patterson and Lippincott-Schwartz, 2002). These fluorophores allow the investigator to rapidly turn on

fluorescence of the probe by irradiation with light of the appropriate wavelength, providing a defined starting timepoint for kinetic investigations. However, the activity of the protein to which paGFP is fused is uncoupled from the fluorescence activation event.

In order to enable coincidence of fluorescence activation and protein activation, we designed the C-terminal caging and quenching strategy described in chapters 3 and 4 (Figures 3.3 and 4.2). Upon photolysis of one bond, caged and quenched Smad2 was activated and became fluorescent (Figures 3.6, 3.7, and 4.5). This allowed us to monitor the kinetics of nuclear accumulation of phosphorylated Smad2 in live cells (Figure 4.7). Quantitation of the imaging data revealed a sigmoidal nuclear import kinetic curve (Figure 4.7B). This indicated that nuclear accumulation of the uncaged Smad2 molecule occurred with at least two kinetically significant steps that are coupled to each other. While the exact molecular events that give rise to this kinetic behavior are not clear, the observation of discrete kinetic steps in live cells is a noteworthy achievement. In the future, it might be possible to define the molecular processes that give rise to these individual kinetic components by observing Smad2 nuclear import in cells that have been genetically or pharmacologically depleted of known proteins that act in the Smad pathway. Another key advantage to our protein visualization system is that it allowed for the monitoring of two versions of Smad2 in the same cell at the same time (Figure 4.10). This multiparameter imaging allows for the comparison of the behavior of one protein with another in live cells because cell-to-cell variability is eliminated (Schultz et al., 2005).

Our semi-synthetic construction of caged proteins and their use in live systems does have drawbacks. They require significant efforts in chemical synthesis and,

therefore, are not as accessible as genetically-encodable fluorophores. However, expressed protein ligation methodology has been adopted by many laboratories and has proven to be a robust method for the construction of semi-synthetic proteins (Muir, 2003). Therefore, we feel that laboratories with experience in protein chemistry would benefit from the application of the caging strategies discussed in this thesis to their proteins of interest.

Another drawback to our caging strategies is that the caged protein must be prepared *in vitro* followed by delivery to the system under study. In the experiments described herein, microinjection was used to deliver caged proteins to live cells and *Xenopus* embryos (Figures 3.8, 3.9, 4.6, 4.7, and 4.10). Microinjection is not an ideal delivery system for all applications, but is useful in many instances such as those described here. In the future, it may be possible to carry out caged protein semi-synthesis inside live cells using previously described approaches (Giriat and Muir, 2003; Lue et al., 2004; Yeo et al., 2003). This would improve the applicability of caged proteins in living systems by obviating the need to deliver the protein after *in vitro* semi-synthesis since the reaction would be carried out directly in the cells under investigation.

6.3. Future improvements to the caging strategies

Both caging strategies could undoubtedly benefit from future improvements. For example, other caging groups could be used in place of *o*-nitrobenzyl derivatives that are better suited for use in live systems. The coumarin-based caging group DECM (**35**, Figure 2.1) is photolyzed with visible light, which is less toxic to cells than the UV light required for *o*-nitrobenzyl photolysis (Shembekar et al., 2005). Also, many confocal

microscopes are now equipped with laser lines in the vicinity of 405 nm; a wavelength that will uncage DECM groups. Therefore, the use of DECM would broaden the applicability of caged proteins since most investigators have access to microscopes with the appropriate lasers for uncaging. Use of the caging groups Bhc and Bhq (**36** and **37**, Figure 2.1) that can be photolyzed with two-photon techniques would allow caged proteins to be activated in relatively thick tissue (Furuta et al., 2004; Furuta et al., 1999; Goard et al., 2005; Zhu et al., 2006). Also, the use of these caging groups would allow for better control over spatial activation of a caged protein since two-photon activation can be confined to a specific z-plane.

Carboxyfluorescein and tetramethylrhodamine are not the best fluorophores for imaging in living systems. Fluorescein derivatives photobleach more readily than other flourophores. Tetramethylrhodamine tends to destablize the proteins to which it is fused. Their use herein was motivated primarily by the fact that these are well-known fluorophores that interface well with solid phase peptide synthesis. Moreover, several variables and unknowns were being evaluated simultaneously during the development of the caging and quenching strategy (section **3.3**) so we wanted to minimize additional unknowns resulting from the use of less well characterized fluorophores. Now that the caging and quenching strategy has been validated, it could be further optimized. Substitution of these fluorophores with brighter, more water-soluble alternatives would improve the applicability of the caged proteins. Coupled to the use of alternative fluorophores is the use of alternative quenchers. Tetramethylrhodamine was not quenched as efficiently by dabcyl as was carboxyfluorescein in proteins **58** and **55**, respectively (Figures 4.5D and 4.9B). A systematic screen using various fluorophores and

quenchers could be employed to identify optimal fluorophore/quencher pairs. Examples of alternative commercially available quenchers that could be tested include Atto quenchers (Atto-tec), Iowa Black (IDT), QXL (Anaspec), DDQ (Eurogentec), QSY (Invitrogen), and BHQ (Biosearch) (Marras, 2006). Perhaps the identification of optimal fluorophores/quencher pairs would result in more stable, more efficiently quenched caged analogues of full-length phosphorylated and non-phosphorylated Smad2. These could be used to more accurately compare the kinetics of nuclear import of these two forms of the protein in live cells following the general framework described in sections **4.5 and 4.6**.

6.4. Construction of Smad2-MH2 containing phosphonates led to the identification of PRMT5 as a potential interacting partner of Smad2

We used EPL to prepare semi-synthetic Flag-Smad2-MH2 containing phosphonomethylenealanine (Pma) at positions 465 and 467 (Figure 5.1). This protein was used as a bait in a protein-protein interaction screen to identify binding partners of Smad2. The protein PRMT5 was discovered in this screen and may be a relevant interaction partner of Smad2. Preliminary experiments indicated that PRMT5 binds preferentially to non-phosphorylated Smad2 (Figures 5.4). In addition, there seems to be two binding sites for PRMT5 in Smad2: one in the MH1 domain and the other in the MH2 domain (Figures 5.4 and 5.5). However, attempts to detect interaction of endogenous PRMT5 and Smad2 have not been successful to date. Therefore, future studies should examine the endogenous PRMT5/Smad2 interaction in cells by further co-immunoprecipitation experiments. We note that it is possible that a true, biologically-relevant protein-protein interaction may be of low affinity or may be transient and may

therefore not be possible to detect with co-immunoprecipitation techniques. Such interactions require alternative methods for their detection. In this regard, our laboratory has recently succeeded in installing cross-linking groups into Smad2 that could potentially stabilize transient interactions of Smad2 with other proteins (Vila-Perello et al., 2007). This and other cross-linking reagents may be useful in validating the Samd2/PRMT5 interaction with endogenous proteins. If this interaction is indeed validated for endogenous proteins, it will be of interest to determine whether these proteins interact directly or indirectly. *In vitro* interaction assays using purified preparations of each protein would help to determine if these proteins interact directly or indirectly through the intermediacy of one or more other molecules, perhaps DNA (Figure 5.5A).

Protein arginine methylation has recently gained recognition as a regulatory posttranslational modification in signal transduction processes (Bedford and Richard, 2005). It remains possible that Smad2 is a substrate of PRMT5. Now that a robust bacterial expression system has been developed for full length Smad2 (see section 4.3 and 4.4), *in vitro* methylation assays using radioactive methyl donors should be performed using full length Smad2 with and without C-terminal phosphorylation. In addition, optimized mass spectrometry-based methods have been developed for identification of methylated arginine in proteins (Ong et al., 2004; Rappsilber et al., 2003). These methods could be adapted to Smad2 and other Smad proteins to search for arginine methylation in these proteins.

6.5. Potential experiments to define the physiological relevance of the Smad2/Prmt5 interaction

In addition to the biochemical assays discussed in the previous section, a focused attempt should be made to determine if Smad2 and PRMT5 influence each other's function *in vivo*. As a starting point, PRMT5 could be overexpressed or knocked-down followed by luciferase assays for TGFβ responsiveness or real-time PCR assays of endogenous Smad-responsive genes to determine if their expression is altered. PRMT5 may be directed by Smads to alter histone modifications at the promoters of Smad-responsive genes, as has been shown for other promoters, including *cyclin E1*, *ST7*, *NM23*, and *Myogenin* (Dacwag et al., 2007; Fabbrizio et al., 2002; Pal et al., 2004). Chromatin immunoprecipitation with PRMT5 antibodies or antibodies directed to methylated histone H3R8 or H4R3 epitopes (known methylation targets of PRMT5) followed by quantitative PCR of selected Smad-responsive genes should be considered in experimental evaluation of this possibility (Ancelin et al., 2006; Pal et al., 2004).

Detailed studies of Smad2 localization have demonstrated that a small population of Smad2 exists in the nucleus even in the absence of phosphorylation (Inman et al., 2002; Massague et al., 2005; Nicolas et al., 2004; Reguly and Wrana, 2003; Schmierer and Hill, 2005; Xu et al., 2002; Xu and Massague, 2004). It appears that C-terminal Smad2 phosphorylation does not trigger Smad2 nuclear import *per se*, rather it influences the proportion of Smad2 in the nucleus by altering the affinity of Smad2 for nuclear and cytoplasmic retention factors (Massague et al., 2005; Xu and Massague, 2004). What, if anything, might be the role of non-phosphorylated Smad2 in the nucleus? One function of nuclear non-phosphorylated Smad2 might be to interact with and direct PRMT5 to certain

Smad-responsive promoters. Once there, PRMT5 may methylate histones H3 and H4 on arginines 8 and 3, respectively. These modifications are typically linked to transcriptional repression (Fabbrizio et al., 2002; Pal et al., 2004), but recent studies have shown that they may also lead to promoter activation (Dacwag et al., 2007). Is it possible that non-phosphorylated Smad2 binds to and leads to repression of some of the same genes that phosphorylated Smad2 activates? In this regard, it may be insightful to analyze global gene expression in wild type cells and in cells lacking Smad2 to differentially identify candidate genes whose expression are derepressed upon loss of Smad2. Alternatively, the set of genes, if any, that are regulated by non-phosphorylated Smad2/PRMT5 may be different than those regulated by phosphorylated Smad2. Regardless of which genes may be regulated by Smad2/PRMT5, it is noteworthy that most gene expression regulation by PRMT5 acts through histone remodeling mechanisms involving the SWI/SNF remodeling complex (Dacwag et al., 2007; Pal et al., 2004; Pal et al., 2003). Furthermore, Brg1, an ATPase subunit of the SWI/SNF complex, has been identified as a Smad binding partner and plays a crucial role in the expression of certain known Smad2 genes (He et al., 2006; Ross et al., 2006). The histone acetyltransferase p300 is recruited by active Smad2 to histone H3 at some Smad target genes where is acetylates lysine 9 of H3, which correlates with active transcription at these genes (Ross et al., 2006). Since PRMT5 is known to methylate the adjacent arginine 8, one can speculate that lysine 9 acetylation and arginine 8 methylation of histone H3 may be synergistic or antagonistic modifications in the activation of Smad-responsive promoters. Further experimentation will determine if this is the case or if it is simply a molecular coincidence.

Very recent studies have implicated various phosphatases in the dephosphorylation of R-Smads (Chen et al., 2006; Duan et al., 2006; Knockaert et al., 2006; Lin et al., 2006). One family of these phosphatases, SCP1-3, which dephosphorylates the linker region of Smad1, 2, and 3 and the C-terminal phosphates of Smad1 (Knockaert et al., 2006; Sicheri et al.), is related to FCP1, which was previously identified as a substrate for PRMT5 (Amente et al., 2005). Given this, it is possible that FCP1, SCP, or other family members may function in a SCP/PRMT5/R-Smad signaling axis.

6.6. A putative side reaction discovered in the synthesis of phosphonate-containing peptides suggests a new design for peptide-based fluorescent kinase probes

During the development of our protein-protein interaction screen using phosphonylated Flag-Smad2-MH2 as bait, we attempted to prepare a peptide acid with Pma at the C-terminus by loading Wang resin with Fmoc-Pma that was not protected on the phosphonate (see section **5.3**). The yield of this loading reaction was extremely low (<5%). One possible explanation is that the unprotected phosphonate attacked the ester bond linking Pma to the Wang linker, leading to cleavage of the ester bond and removal of Fmoc-Pma from the resin as an asymmetric phosphoanhydride (Figure 6.1A). To test this explicitly, a peptide could be synthesized on Wang resin containing protected phosphoserine as the C-terminal amino acid such that it would be attached to the Wang linker by an ester bond. If the proposed mechanism is correct, and assuming that the phosphate of phosphoserine has similar nucleophilicity as the phosphonate of Pma, then orthogonal deprotection of the phosphate should lead to auto-cleavage of the peptide

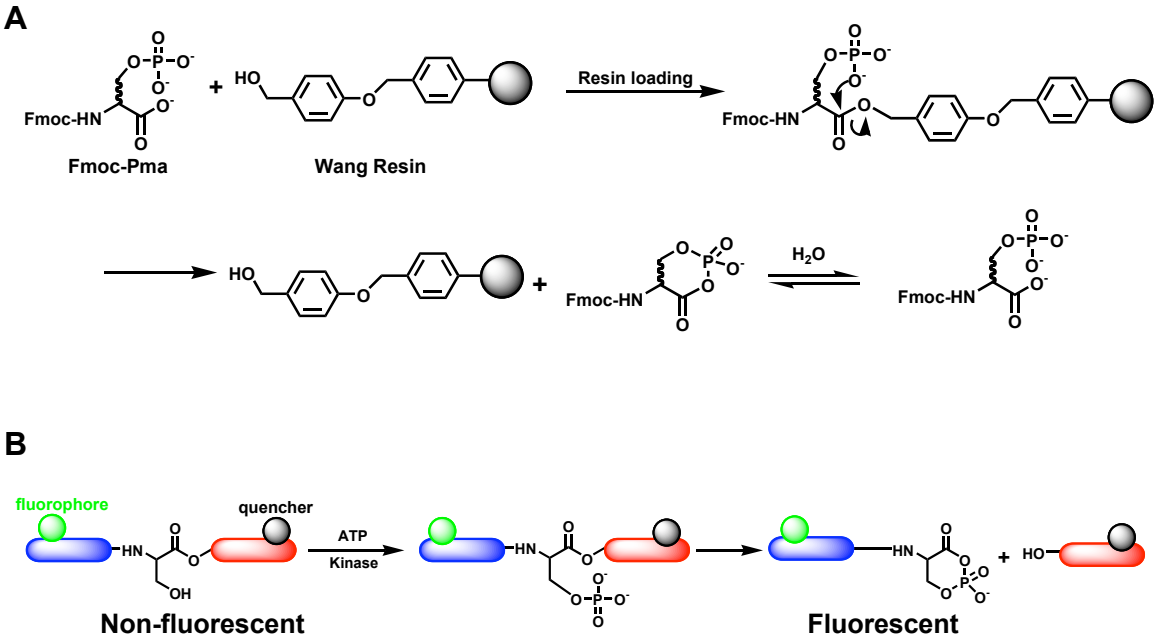


Figure 6.1. A putative side reaction suggests a new design for a fluorogenic serine/threonine kinase probe. A. Following loading of Fmoc-Pma onto Wang resin, the unprotected phosphonate attacks the ester linking Fmoc-Pma to the resin, resulting in the release of Fmoc-Pma from the resin as an asymmetric phosphoanhydride, which is subject to hydrolysis to regenerate Fmoc-Pma. **B.** Proposed design of a peptide-based fluorogenic kinase probe. The probe contains a fluorophore and a quencher such that it is non-fluorescent. It also contains a serine (or threonine) in the context of the recognition sequence of a kinase of interest. The serine is linked to the amino acid immediately C-terminal to it by an ester bond. Upon phosphorylation, the ester bond is cleaved by intramolecular attack by the incipient phosphate group, by analogy to the reaction described in panel **A**. This results in peptide cleavage and release of quenching, restoring the fluorescence of the fluorophore, which can then be monitored.

from the resin. This could be monitored readily by installing a fluorophore N-terminal to the phosphoserine and monitoring the increase of fluorescence in the supernatant of a resin suspension following deprotection of the phosphate.

If this reaction does indeed occur, it suggests a design for a fluorogenic serine/threonine kinase probe. In section **1.2.6** of this thesis, fluorescent kinase probes were discussed (Figure 1.8). Throughout the development of these probes, investigators have attempted to achieve two goals in probe design. First, a large increase in fluorescence upon probe phosphorylation by the kinase being monitored is essential. Second, a probe that does not require any third-party molecules such as protein binding partners or metal ions is favored over one that does require third-party participation for *in vivo* applications. However, the most robust increases in fluorescence have come from probes that do require third-party participation (see discussion in section **1.2.6**). To address this issue, a probe of the general design as shown in Figure 6.1B could be synthesized. Like the previously described probes, this hypothetical probe contains a kinase recognition sequence and is phosphorylated by a serine/threonine kinase of interest. Upon phosphorylation, the phosphate would attack the ester bond present between the phosphorylated amino acid and the amino acid immediately C-terminal to it. Before phosphorylation, the fluorescence of the fluorophore is quenched by the presence of an intramolecular quencher in much the same way as in a molecular beacon and in the caged proteins described in this thesis (Figures 3.3 and 4.2). Upon phosphorylation and cleavage of the ester bond, the fluorophore and quencher would no longer reside in the same molecule so a robust increase in fluorescence would occur. If successful, this class of fluorogenic kinase probes would be immensely useful because a large increase in

fluorescence could be achieved upon phosphorylation by the kinase being monitored without requiring any additional molecular interactions.

6.7. Conclusions

The protein semi-synthesis approach expressed protein ligation has been used to prepare several analogues of the transcription factor and signaling molecule Smad2. These include caged versions of the protein that have been validated both *in vitro* and *in vivo*. One caging strategy involves direct attachment of caging groups to the phosphoserine residues at positions 465 and 467 of Smad2. This technique is generalizable to the preparation of other caged phosphoproteins (Vogel and Imperiali, 2007). The other caging approach was linked to a conditional fluorescence activation strategy such that only the active, uncaged protein was fluorescent. This strategy is particularly useful for monitoring the localization of proteins upon uncaging.

An analogue of Smad2 was created that contains the stable phosphoserine mimic phosphonomethylenealanine at positions 465 and 467 of Smad2. Use of this protein as bait in a protein-protein interaction screen led to the identification of PRMT5 as a potential Smad2 interaction partner. Preliminary experiments indicated that PRMT5 preferentially interacts with non-phosphorylated Smad2 over phosphorylated Smad2 and that there may be two binding sites for PRMT5 on Smad2. These preliminary findings require further experimentation for validation. Future experiments have been suggested for this purpose as well as to uncover the possible physiological role of the Smad2/PRMT5 interaction.

Protein semi-synthesis is a powerful method for the construction of proteins containing multiple unnatural groups, as demonstrated herein. In the past decade, many investigators have begun to succeed in applying semi-synthetic proteins to biological problems. Future efforts in this field will undoubtedly yield new insight into biological systems.

Chapter 7: Materials and Methods

7.1. Chemicals

All Fmoc amino acid derivatives, resins, 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy)-butanoic acid, and dabcyl-OSu were from Novabiochem unless stated otherwise (San Diego, CA). *tert*-butoxylcarbonyl-1,3-thiazolidine-4-carboxylic acid (Boc-Thz) was from Bachem (King of Prussia, PA). Boc-Cys(StBu)-OH was purchased from Fluka (Milwaukee, WI). O-(1H-benzotriazole-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU) was from Qbiogene (Carlsbad, CA). O-(7-Azabenzotriazole-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU) was from Fluka. Piperidine and DIEA were from Applied Biosystems (Foster City, CA). *O*-1-(2-nitrophenyl)ethyl-*O′*-β-cyanoethyl-*N,N*-diisopropylphosphoramidite was prepared as described (Rothman et al., 2002). 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy)-butanoic acid was from Novabiochem. Complete protease inhibitor tablets were from Roche and were used at the manufacturer's recommended concentration. Tris(2-carboxyethyl)phosphine hydrochloride was from Strem (Newburyport, MA). All other reagents were from Sigma-Aldrich unless stated otherwise (St. Louis, MO).

7.2. Solvents

Solvents for HPLC and for general synthetic purposes were from Fisher (Fairlawn, NJ). Anhydrous solvents were from Aldrich (Milwaukee, WI) or Acros (Morris Plains, NJ).

7.3. Fluorophores

Fluorescein-5-maleimide, Texas Red-C_2-maleimide, and fluorescent dextrans were from Molecular Probes (Eugene, OR). 6-carboxyfluorescein-OSu and 5(6)-tetramethylrhodamine were from Anaspec (San Jose, CA).

7.4. Chromatography

Analytical gradient reversed-phase HPLC was performed on Hewlett-Packard 1100 series instruments and a Vydac C18 column (5 micron, 4.6 x 150 mm). Flow rate was 1 mL/min and routine UV detection was at 214 and 280 nm. Semi-preparative (Vydac 15-20 micron, 10 x 250 mm), preparative (Vydac 10 micron, 22 x 250 mm), and process (15-20 micron, 50 x 250 mm) RP-HPLC were performed on a Waters DeltaPrep 4000 system fitted with a Waters 486 tunable absorbance detector. Flow rate was 5 mL/min (semi-preparative), 16 mL/min (preparative), or 30 mL/min (process). All runs used linear gradients of 0.1% aqueous TFA (solvent A) and 90% aqueous acetonitrile plus 0.1% TFA (solvent B). Native protein purification was performed on classic and ÄKTA FPLC systems (Amersham Biosciences, Piscataway, NJ). Detailed procedures, columns, and gradients used for protein purification are included in the sections below describing each protein. Analytical SEC of proteins described in chapter 2 was performed on a Superdex 200 column (Amersham Biosciences) in aqueous running buffer containing 20 mM HEPES, 150 mM NaCl, 0.5 mM EDTA, 2 mM DTT at pH 8.0 at a flow rate of 0.5 mL/min. Analytical SEC of proteins described in chapter 3 was performed on a Zorbax GF-250, column (Agilent, Palo Alto, CA). The flow rate was 0.5 mL/min, and the buffer contained 1 mM DTT, 150 mM NaCl, 25 mM Tris (pH 7.2). SEC of proteins described in

chapter 4 was performed on a TOSOH SuperSW3000 HPLC column at 25 °C at a flow rate of 0.35 mL/min. The running buffer was 20 mM HEPES, 150 mM NaCl, 0.05% NaN_3, pH 7.3. Detection was at 280 nm or 496 nm, which monitors fluorescein absorbance. SEC standards (Sigma) included proteins cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), BSA (66 kDa), alcohol dehydrogenase (150 kDa), and blue dextran (2,000 kDa).

7.5. Mass spectrometry

Mass spectrometric analysis was routinely performed on all synthetic peptides and expressed proteins on a Sciex API-100 single quadrupole spectrometer or a Micromass ZQ single quadrupole spectrometer. For MS/MS experiments, a Finnigan LCQ ion trap instrument was used. Calculated protein masses were obtained using the PeptideMass tool on the Expasy server (http://us.expasy.org).

7.6. General peptide synthesis procedures

All peptides were synthesized manually using standard Fmoc SPPS on Wang resin for peptide acids and Rink amide MBHA resin for peptide carboxamides. Standard chain assembly was carried out for ~2 hours with HBTU activation using a 4.4-fold excess of amino acid over the resin in DMF with DIEA as base. A stream of dry N_2 was used to agitate the reaction mixture. Fmoc removal was carried out with 20% piperidine in DMF (1 x 3 minutes, followed by 1 x 10 minutes, unless stated otherwise). Washing of peptidyl-resins between coupling cycles with DMF was carried out with alternating batch and flow washes over a 3 minute period. Cleavage of peptides from the resin was

achieved with 92.5% TFA, 2.5% triisopropylsilane (TIS), 2.5% ethanedithiol, and 2.5% H_2O (cleavage cocktail), unless stated otherwise. Crude peptide products were precipitated and washed with cold Et_2O and dissolved in solvent A. A minimal amount of solvent B was added to aid dissolution of any non-dissolved material and peptides were lyophilized. Thz in applicable peptides was deprotected by adding 0.5 M methoxylamine (aq) to the peptide and agitating the resulting solution for 6 hours at 37 °C. For peptides that did not dissolve in 0.5 M methoxylamine (aq), acetonitrile was added to aid dissolution. Acetonitrile did not adversely affect the efficiency of the Thz deprotection step. All peptides were purified by RP-HPLC. In addition to the usual washing steps between coupling cycles, an additional wash with 10% DIEA in DMF was used for peptides containing Fmoc-Ser(PO(OBzl)OH)-OH to ensure complete removal of piperidine from piperidine/phosphate counter-ion adducts.

7.7. Data for peptides prepared using standard methods

For H-CSSMS-OH (**67**), the protected amino acid building blocks used were: Boc-Thz-OH for Cys, Fmoc-Ser(*t*Bu)-OH for Ser, and Fmoc-Met-OH for Met. H-CSSMS-OH was purified by process RP-HPLC. Identity of the peptide was confirmed by ESI-MS (found *m/z* 514.0, expected 513.6 $M+H^+$).

For H-CSpSMpS-OH (**68**), the protected amino acid building blocks used were: Boc-Thz-OH for Cys, Fmoc-Ser(*t*Bu)-OH for Ser, Fmoc-Ser(PO(OBzl)OH)-OH for pSer, and Fmoc-Met-OH for Met. H-CSpSMpS-OH was purified by semi-preparative RP-HPLC. Identity of the peptide was confirmed by ESI-MS (H-CSpSMpS-OH found *m/z* 674.0, expected 673.5 $M+H^+$).

For H-CSpSK(Fl)pS-OH (**69**), the protected amino acid building blocks used were: Boc-Cys(Trt)-OH for Cys, Fmoc-Ser(tBu)-OH for Ser, Fmoc-Ser(PO(OBzl)OH)-OH for pSer, and Fmoc-Lys(MTT)-OH for Lys. Following chain assembly, the Mtt group of Lys was removed with 1% TFA/5% TIS in DCM for 30 minutes. 6-carboxyfluorescein-OSu, dissolved in DMF containing DIEA, was then coupled to the Lys. Identity of the peptide was confirmed by ESI-MS (H-CSpSK(Fl)pS-OH found *m*/z 1029.9, expected 1029.8 $M+H^+$).

7.8. Synthesis of H-CSpS(NPE)MpS(NPE)-OH (39)

A synthetic scheme for this peptide can be found in Figure 2.4. Resin loading was carried out using the method of Sieber (Sieber, 1987). Unreacted hydroxyl functionalities on the resin were capped with benzoyl chloride. Double coupling was routinely used to ensure quantitative acylations. The two residues to be phosphorylated and caged were incorporated as Fmoc-Ser(Trt)-OH. Met was incorporated as Fmoc-Met-OH. The non-phosphorylated Ser was incorporated as Fmoc-Ser(*t*Bu)-OH. Cys was incorporated as Boc-Thz-OH. Following chain assembly, the resin was treated with 2% TFA, 5% TIS in DCM for 30 minutes. After extensive washing with DMF and DCM, the resin (~0.04 mmoles) was dried overnight under high vacuum. All subsequent manipulations were carried out in the dark. Phosphitylation was carried out at room temperature by adding *O*-1-(2-nitrophenyl)ethyl-*O′*-β-cyanoethyl-*N,N*-diisopropylphosphoramidite (**43**, 0.6 mmoles, 7.5 equivalents per each free hydroxyl) and 4,5-dicyanoimidazole (1.2 mmoles) in 2 mL of anhydrous DMF to the peptidyl-resin. The reaction mixture was stirred gently with a magnetic stir-bar. Care was taken to maintain the reaction mixture in an

anhydrous, anaerobic environment by carrying out the reaction under argon. After 20 hours, the resin was washed with anhydrous DMF followed by anhydrous DCM and suspended in 5 mL of anhydrous DCM. 1 mL of a 5.5 M solution of *t*BuOOH in decane was added dropwise to the stirred resin suspension. After 20 minutes, the resin was drained and washed well with DCM, DMF, Et_2O, DMF, and DCM and placed on high vacuum overnight. The dry resin was suspended in 2 mL of cleavage cocktail and stirred vigorously for 3.5 hours. The cleavage mixture was added dropwise to 35 mL of cold Et_2O and the resulting suspension was placed at -20 °C. The peptide was collected by centrifugation and the Et_2O was decanted. Fresh Et_2O was added and the centrifugation and decanting steps were repeated. The precipitated peptide was dissolved in 20% solvent B and lyophilized. To remove the remaining protecting groups, the peptide was then dissolved in DMF containing 1% DBU and stirred at room temperature. After 3 minutes, methoxylamine in water was added to a final concentration of 0.5 M and the solution was agitated for 6 hours at 37 °C. The crude mixture contained **39** as the major product (~75%) as determined by RP-HPLC (Figure 2.5). The peptide was purified by preparative RP-HPLC and its identity was confirmed by ESI-MS (found *m/z* 972.0, expected 972.2 $M+H^+$) and by MS/MS experiments (Figure 2.6). The isolated yield after synthesis and purification was ~10%.

7.9. Synthesis of H-CSpSKpS-pl-K-NH_2 (70)

The peptide was prepared with Fmoc-Lys(Boc)-OH, 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy)-butanoic acid, Fmoc-Ser(PO(OBzl)OH)-OH, Fmoc-Met-OH, Fmoc-Ser(tBu)-OH, and Boc-Cys(Trt)-OH. The synthesis steps were as follows: 1) coupling of

Fmoc-Lys(Boc)-OH and deprotection; 2) coupling of 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy)-butanoic acid with HBTU; 3) coupling of Fmoc-Ser(PO(OBzl)OH)-OH with DIC (290 μL, 1.85 mmol), DMAP (9 mg, 74 μmol), and DIEA (644 μL, 3.7 mmol) in DMF/DCM (1:1, v/v) for 4h at 4 °C; 4) capping of unreacted alcohol groups with benzoylchloride (155 μL, 1.34 mmol), pyridine (156 μL, 1.93 mmol) in DCM (3.8 mL) for 5h; 5) Fmoc deprotection; 6) coupling of Fmoc-Met-OH and deprotection; 7) coupling of Fmoc-Ser(PO(OBzl)OH)-OH and deprotection; 8) coupling of Fmoc-Ser(tBu)-OH and deprotection; 8) coupling of Boc-Cys(Trt)-OH; 9) TFA deprotection and cleavage. The peptide was purified by semi-preparative RP-HPLC. The identity of the peptide was verified by ESI-MS (found *m/z* 1082.9, expected 1084.0 $M+H^+$).

7.10. Synthesis of H-CSpSK(Fl)pS-pl-K(Dab)G-NH_2 (52)

A synthetic scheme for this peptide can be found in Figure 3.5. The peptide was prepared with Fmoc-Gly-OH. Fmoc-Lys(Mtt)-OH, 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy)-butanoic acid, Fmoc-Ser(PO(OBzl)OH)-OH, Fmoc-Ser(tBu)-OH, Boc-Cys(StBu)-OH, and the succinimidyl esters of dabcyl and 6-carboxyfluorescein. The synthesis steps were as follows: 1) coupling of Fmoc-Gly-OH and deprotection; 2) coupling of Fmoc-Lys(Mtt)-OH and deprotection; 3) coupling of 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy)-butanoic acid; 4) coupling of Fmoc-Ser(PO(OBzl)OH)-OH with DIC (290 μL, 1.85 mmol), DMAP (9 mg, 74 μmol), and DIEA (644 μL, 3.7 mmol) in DMF/DCM (1:1, v/v) for 4h at 4 °C; 5) capping of unreacted alcohol groups with benzoylchloride (155 μL, 1.34 mmol), pyridine (156 μL, 1.93 mmol) in DCM (3.8 mL) for 5h; 6) deprotection of the Mtt group by treatment of the resin with 1%TFA/5% TIS in

DCM for 30 min; 7) coupling of dabcyl-OSu (266 mg, 815 μmol) in DIEA (644 μL, 3.7 mmol) and DMF for 8 hrs; 8) Fmoc deprotection; 9) coupling of Fmoc-Lys(Mtt)-OH and deprotection; 9) coupling of Fmoc-Ser(PO(OBzl)OH)-OH and deprotection; 10) coupling of Fmoc-Ser(tBu)-OH and deprotection; 11) coupling of Boc-Cys(StBu)-OH; 12) removal of Mtt group; 13) coupling of 6-carboxyfluorescein-OSu (525 mg, 1.1 mmol) in DIEA (644 μL, 3.7 mmol) and DMF for 12 hrs; 14) deprotection of the StBu group by treatment with PBu_3 (1.8 mL, 7.2 mmol), H_2O (518 μL, 28.8 mmol), in 11 mL of DMF/DCM (1:1, v/v) for 4 hrs; 15) TFA deprotection and cleavage. The peptide was purified by semi-preparative RP-HPLC. The peptide was characterized by analytical RP-HPLC, UV-Vis absorption, and ESI-MS (found *m/z* 1746.8, expected 1746.6 $M+H^+$). Analytical data is presented in Figure 3.5.

7.11. Synthesis of H-CSEK(Fl)E-pl-K(Dab)G-NH_2 (54)

Synthesis of this peptide was carried out in the same manner as described in section **7.10**, except Fmoc-Glu(O*t*Bu)-OH was used in place of Fmoc-Ser(PO(OBzl)OH)-OH. The identity of the peptide was verified by ESI-MS (found *m/z* 835.7, expected 835.8 $M+2H^+$).

7.12. Synthesis of H-CSAK(Rho)A-pl-K(Dab)G-NH_2 (59)

Synthesis of this peptide was carried out in the same manner as described in section **7.10**, except Fmoc-Ala-OH was used in place of Fmoc-Ser(PO(OBzl)OH)-OH and 5(6)-tetramethylrhodamine was used in place of 6-carboxyfluorescein. 5(6)-

tetramethylrhodamine was activated with HBTU and DIEA. The identity of the peptide was verified by ESI-MS (found *m/z* 1607.8, expected 1607.7 $M+H^+$)

7.13. Synthesis of H-CSPmaMPma-OH (71)

Fmoc-D,L-Pma (Pma = phosphonomethylenealanine) was synthesized from D,L-Pma (obtained from Tocris) and Fmoc-OSu (obtained from Fluka) as described by Zheng et al. (Zheng et al., 2003). Peptide synthesis was initiated by loading of Rink Amide MBHA resin (0.167 mmole) using a threefold excess of pre-activated Fmoc-D,L-Pma (0.5 mmole). Activation of Fmoc-D,L-Pma was accomplished upon the addition of 0.475 mmole HATU and 1.5 mmole of DIEA to Fmoc-D,L-Pma in DMF. Subsequent chain elongation was performed as described in section **7.6** using HATU as the activation agent throughout. The peptide was purified by semi-preparative RP-HPLC. The peptide was characterized by ESI-MS (found *m/z* = 681.2, expected 681.2 Da $M + H^+$).

7.14. Preparation of fluorescently labeled peptide 46

Purified **39** (0.7 mg, 0.72 μmoles) was dissolved to a final concentration of 0.1 mM in 50 mM Tris-HCl containing 1 mM TCEP at pH 7.5. The solution was allowed to stand at room temperature for 1 hour. Fluorescein-5-maleimide was then added to a final concentration of 1 mM from a fresh stock solution in DMF. The reaction mixture was agitated for 3 hours at room temperature and the fluorescently labeled peptide (**46**) was purified by semi-preparative RP-HPLC. The identity of the peptide was confirmed by ESI-MS (found *m/z* 1399.0, expected 1399.3 Da $M+H^+$).

7.15. *In vitro* photolysis

Light illumination was performed with a collimated light source from Oriel (Stratford, CT) equipped with a 200 W Hg lamp. The irradiance was measured using a monochromic photometer (model 840-c, Newport, Irvine, CA). Selective irradiation at 312 nm or 365 nm was performed using an analytical line filter (11 nm and 9.4 nm bandwidths, respectively) obtained from Oriel. Alternatively, photolysis was performed by placing the protein solution in the beam of a He-Cd laser (325 nm, 4.74 W/cm^2) for 3 seconds.

7.16. Determination of photolysis kinetics and quantum yield of peptide 46 and protein 47

For peptide **46** (data for the quantum yield determination can be found in Figures 2.7 and 2.8), light at 312 nm was focused on the sample. An irradiance of 2 mW/cm^2 was used, as measured with a digital photometer from Newport (Irvine, CA). The concentration of **46** was 10 μM in an aqueous buffer containing 20 mM Tris-HCl, 100 mM NaCl, 5% glycerol, 2 mM DTT at pH 7. Irradiation times were: 0 sec, 20 sec, 37 sec, 75 sec, 150 sec, 225 sec, 300 sec, 450 sec, and 600 sec. The irradiated solution was injected onto an analytical RP-HPLC column and eluted with a linear gradient of solvent B while monitoring at 443 nm (the absorbance maximum of fluorescein under acidic conditions). The identities of the eluting species were confirmed by ESI-MS (**46** found *m/z* 1399.0, expected 1399.3 Da M+H$^+$; **46** – 1 caging group found *m/z* 1250.0, expected 1250.2 Da M+H$^+$, **46** – 2 caging groups found *m/z* 1101.0, expected 1101.2 M+H$^+$). The two singly-caged photolysis intermediates were considered as one species for the purposes of

photolysis kinetics and quantum yield calculations. The relative amount of each form was determined by calculating the ratio of each peak area to the sum of the peak areas for all forms. Experiments were conducted in duplicate and the average values were used for all subsequent calculations. The half-life for photolysis of each caging group was determined by plotting the natural logarithm of the fraction of caging groups remaining on the peptide versus time of irradiation. Photolysis quantum yield (Φ) was determined from the following equation (Hasan et al., 1997):

$$\Phi = \frac{0.693\ (6.02 \times 10^{20})}{2.303\ t_{1/2}\ \varepsilon_{\lambda}\ I}$$

where $t_{1/2}$ is the calculated half-life of the photolysis reaction (sec), ε_{λ} is the molar extinction coefficient per caging group at the photolysis wavelength λ, and I is the light intensity (photons cm^{-2} sec^{-1}). ε_{λ} at λ = 312 nm was determined to be 2,613 M^{-1} cm^{-1} by subtracting the portion of the total extinction coefficient of **46** due to the fluorescein label from the total extinction coefficient and dividing the resulting quantity by 2. For protein **47** (the data for the quantum yield determination can be found in Figure 3.7), the quantum yield of photolysis was calculated with the same equation. The half-life ($t_{1/2}$) was calculated from the log plot of protein fluorescence emission at time t divided by protein fluorescence emission at time zero versus the irradiation time. The experiments were repeated two to three times for each illumination conditions and the mean half-life was calculated. The molar extinction coefficients of 4-[4-(1-hydroxyethyl) -2-methoxy-5-nitrophenoxy)-butanoic acid were determined in 1 mM DTT, 150 mM NaCl, 25 mM Tris (pH 7.2) (ε_{365}= 4500 M^{-1} cm^{-1}).

7.17. Generation of Smad2-MH2 α-thioester (38)

Smad2-MH2 (residues 241-462) was expressed in *E. coli* BL21(DE3) cells as a fusion protein with a modified GyrA intein and a C-terminal chitin binding domain in the pTXB1 vector from New England BioLabs (Beverly, MA) as previously described (Wu et al., 2001). The expression vector for Smad2-MH2-GyrA-CBD will be referred to as pMH2. Transformed cells were grown at 37 °C to an OD_{600} of 0.65 in LB supplemented with ampicillin. Protein expression was then induced by the addition of 0.5 mM IPTG to the growth medium. The cells were allowed to grow for an additional 3.5 hours at 37 °C at which time they were collected by centrifugation. After this point, all manipulations were performed either on ice or at 4 °C. The cell pellet was resuspended in lysis buffer (20 mM Tris-HCl, 200 mM NaCl, 200 mM 1,6-hexanediol at pH 7.5) supplemented with Complete protease inhibitor tablets from Roche Diagnostics (Mannheim, Germany). Following lysis by passage through a French press, the insoluble fraction was removed by centrifugation and the soluble fraction was then filtered through a 5 μm filter. The clarified soluble fraction was incubated with chitin resin pre-equilibrated in lysis buffer for 20 hours. The resin was then poured into a column and washed with 7.5 column volumes (CVs) of lysis buffer, followed by 1.5 CVs of cleavage buffer (100 mM HEPES, 200 mM NaCl, 200 mM 1,6-hexanediol at pH 8.0), followed by 1 CV of cleavage buffer supplemented with 50 mM 2-mercaptoethanesulfonic acid (MESNa). The resin was then incubated for 20 hours in 2.5 CVs of cleavage buffer supplemented with 50 mM MESNa. The protein was eluted from the column with cleavage buffer supplemented with 50 mM MESNa. The fractions containing Smad2-MH2 α-thioester (as assessed by SDS-PAGE) were combined and 2 molar equivalents of the Smad binding domain of SARA (SARA-

SBD, residues 665-721) were then added to form the Smad2-MH2/SARA-SBD hetero-complex. SARA-SBD was prepared by Tev protease catalyzed proteolysis of a bacterially expressed GST fusion protein that was purified over GSH resin. The heterocomplex was then purified from excess SARA-SBD by cation exchange chromatography at pH 6.0 using an SP FF column (Amersham Biosciences). A linear gradient from 0 to 1 M NaCl was used. The flow rate was 2 mL/min. The heterocomplex was simultaneously buffer-exchanged into ligation buffer (100 mM HEPES, 200 mM NaCl, 50 mM MESNa) and concentrated to 6.25 mg/mL using a 10 kDa molecular weight cut-off ultrafiltration membrane from Millipore (Billerica, MA). Proteins were either used immediately in expressed protein ligation reactions or flash frozen with liquid N_2 and stored at -80 °C.

7.18. Expressed Protein Ligation of Smad2-MH2 domain semi-synthetic proteins

Expressed protein ligation (EPL) was performed in the dark at 5 °C by adding the appropriate N-terminal cysteine-containing synthetic peptide (1 mM, unless otherwise specified) to the Smad2-MH2 α-thioester/SARA-SBD heterocomplex (0.25 mM) in ligation buffer. The reaction was followed by SDS-PAGE, RP-HPLC, and ESI-MS. Prior to RP-HPLC analysis, the protein was reduced with TCEP at pH 7.5. Reactions were complete within 24 hours. The semi-synthetic proteins were then separated from unreacted peptide by SEC (Superdex 75, Amersham Biosciences) in 20 mM Tris-HCl, 100 mM NaCl, 2 mM DTT, and 5% glycerol at pH 7.0. Fractions were analyzed by SDS-PAGE, RP-HPLC, and ES-MS.

7.19. Multi-angle Laser Light Scattering (MALLS)

SEC/MALLS was performed as described at the W.M. Keck Foundation Biotechnology Resource Laboratory at Yale University (Folta-Stogniew and Williams, 1999). A Superdex-200 column (Amersham Biosciences) was used with a flow rate of 0.5 mL/min. MALLS measurements were taken every second. The loading concentration for the doubly-phosphorylated Smad2-MH2 standards was 50 μM. Loading concentrations for caged Smad2-MH2 (**40**) were 50 μM, 25 μM, 12.5 μM, and 5 μM.

7.20. Fluorescence Spectroscopy

Experiments were conducted at 20 °C in a stirred 1 cm-pathlength cuvette using a SPEX FL-311C fluorimeter. Excitation was at 488 nm with a 2 nm slit and the fluorescence emission was monitored from 498 to 700 nm through a 3 nm slit. Protein solutions were 1 μM in 20 mM Tris-Cl, 150 mM NaCl, pH 7.5.

7.21. Protein Labeling with Texas Red C_2-Maleimide (TRM) and Nuclear Import Assay

Caged (**40**) and non-phosphorylated control proteins (6 μM in 20 mM Tris-HCl, 150 mM NaCl, 30 mM TCEP at pH 7.5) were labeled in the dark at 4 °C for 13.5 hours with 50 μM TRM added from a 2.5 mM stock solution in DMSO. Doubly-phosphorylated control protein was labeled in the same manner, except 75 μM TRM was used to achieve a level of labeling equivalent to that of the OP and caged proteins. Excess TRM was removed by SEC (Quick Spin Sephadex G-50 columns columns, Roche, Basel, Switzerland) and the proteins were simultaneously exchanged into import buffer (20 mM HEPES, 110 mM

potassium acetate, 5 mM, sodium acetate, and 2 mM magnesium acetate at pH 7.3). For the nuclear import assay, proteins were diluted to 1.5 μM in import mixture (import buffer + 2 mM DTT, 0.5 mM EGTA, 1 mM MgATP, 5 mM creatine phosphate, 20 units/mL creatine phosphokinase, 1 mg/mL BSA, 4.5 μM GST-SARA-SBD, 1 μg/mL aprotinin, 1 μg/mL leupeptin, and 1 μg/mL pepstatin). Prior to incubation with cells, proteins were either left in the dark (- UV) or irradiated with the He-Cd laser for 5 seconds (+ UV) as described above. Nuclear import assays with digitonin-permeablized HeLa cells were carried using a standard protocol, as previously described (Adam et al., 1992; Xu et al., 2000). The import reaction was carried out for 20 minutes at room temperature. Following the import reaction, samples were washed well, fixed with 4% paraformaldehyde, mounted, and analyzed by laser-scanning confocal microscopy (LSM 510, Carl Zeiss, Thornwood, NY) at the Rockefeller University Bio-Imaging Resource Center.

7.22. Semi-synthesis of Smad2-MH2-CSEK(fluorescein)E-pL-K(dabcyl)-G (53)

This protein was constructed by EPL of Smad2-MH2-α-thioester (**38**) and H-CSEK(Fl)E-pl-K(Dab)G-NH_2 (**54**, see sections 7.17 and 7.18). The identity of the protein was confirmed by manual reconstruction of ESI-MS data (found 26,521 ± 5 Da, expected 26,517).

7.23. Uncaging of Smad2-MH2-CSEK(fluorescein)E-pL-K(dabcyl)-G (53) in live *Xenopus laevis* embryos

2-cell stage embryos were microinjected in the animal pole and allowed to develop until the late blastula stage. A total of 20 nL/embryo was injected with a solution containing the tracer Rhodamine-dextran (70 kDa) with or without protein **53** (3.5 mM). For uncaging and imaging studies (Figure 3.9), animal regions were explanted, placed under a coverslip and imaged with a Zeiss LSM 5 Pascal confocal microscope before and after photoactivation with the 325 nm laser for 3 seconds as described in section 7.15.

7.24. ARE luciferase assays in *Xenopus laevis* embryos

A total of 20 nL/embryo was injected containing various combinations (as indicated in Figure 3.10) of 20 pg of ARE-luciferase reporter, 40 pg Activin mRNA, non-phosphorylated Smad2-MH2-0P (5 nM), or doubly-phosphorylated Smad2-MH2-2P (5 nM). Embryos were allowed to develop until the late blastula stage. At this point, the animal regions were explanted and luciferase activity was assayed using the manufacturer's recommended procedure (Promega Luciferase Assay System). Data is reported in Figure 3.10 as mean of 3 independent experiments, error bars = ± S.D.

7.25. General cell culture methods, TGFβ stimulation of cells, and western blotting to determine phosphorylation stoichiometry of Smad2

HeLa, HaCaT, HEK293T, and DU145 cells were maintained in DMEM supplemented with 10% FBS. NIH3T3 cells were maintained in DMEM supplemented with 10% CS. For the data in Figure 4.1, cells were serum starved for 6 hours, then the media was

replaced by DMEM containing 100 pM TGFβ-1 (a kind gift from C. Alarcón and J. Massagué, MSKCC). Samples of cells were stimulated with TGFβ-1 for 0, 5, 15, 30, 45, 60, 90, and 120 minutes before harvesting. Cells were washed with DPBS and then lysed in 2X SDS-containing loading buffer. Lysates and recombinant phosphorylated (generated by EPL, see section 7.29) or non-phosphorylated Flag-Smad2 protein standards (standards were added at defined ratios) were separated by SDS-PAGE (4-15%, Criterion, BioRad) and transferred to PVDF (0.2 mm, BioRad). Membranes were blocked with 5% milk in DPBS, probed for phosphorylated Smad2 (rabbit anti-phospho-Smad2 Ser465/467, Cell Signaling Technology), washed, and incubated with a fluorescent anti-rabbit antibody (IRDye 680 goat anti-rabbit, Li-Cor Biosciences). The membrane was washed then imaged on an Odyssey scanner (Li-Cor) in the 700 nm mode. The membrane was stripped (Western Re-Probe Reagent, Calbiochem), re-blocked, and re-probed for total Smad2 (mouse anti-Smad2/3, BD Transduction Labs), washed, incubated with a fluorescent anti-mouse secondary antibody (IRDye 800CW goat anti-mouse, Li-Cor), washed, and imaged on an Odyssey scanner in the 800 nm mode. Band intensities were quantified with the software included with the Odyssey scanner (Li-Cor). For each band corresponding to Smad2, the ratio of phospho-Smad2 intensity to total Smad2 intensity was determined. A standard curve was generated using this ratio for each recombinant sample of known phospho-Smad2 fractional stoichiometry. The endogenous phospho-Smad2 fractional stoichiometry was determined by comparing the ratio of each unknown sample to the standard curve. Averages and standard deviations were determined from three experiments.

7.26. Cloning of His_6-SUMO-Flag-Smad2 into pTXB1

The internal *Sap*I restriction site in human *Smad2* was removed by site-directed mutagenesis (QuikChange II XL, Stratagene) using the 5' primer (5'- C AAC CAG GAA TTT GCT GCG CTT CTG GCT CAG TCT GTT AAT C) and its reverse complement. Full length *Smad2* cDNA lacking the last five codons was amplified by PCR using the 5' primer (5'-GGT GGT CAT ATG TCG TCC ATC TTG CCA TTC ACG C) and the 3' primer (5'-GA TGA TGC TCT TCC GCA ACG CAC TGA AGG GGA TCC). The PCR product was digested with *Nde*I and *Sap*I, purified, and cloned into pTXB1 (New England Biolabs, Ipswich, MA) using the *Nde*I and *Sap*I sites to create pJM1. The DNA encoding yeast SUMO was amplified by PCR from a plasmid containing the *Smt3* gene from *S. cerevisiae* using the 5' primer (5'-GAC GCT CAT ATG GGT CAT CAC CAT CAT CAT CAC GGG TCG GAC TCA GAA GTC AAT CAA GAA GCT) and 3' primer (5'-GAC AAC CAT ATG ACC ACC AAT CTG TTC TCT GTG AGC CTC AAT AAT ATC). The 5' primer was designed to introduce an N-terminal His_6 tag and both primers contained the *Nde*I restriction site. The PCR product was digested, purified, and cloned into pJM1 using the *Nde*I site to create pJM11. The *Nde*I site 5' of *Smt3* was destroyed by site-directed mutagenesis using the 5' primer (5'-GAA GGA GAT ATA CGT ATG GGT CAT CAC CAT CAT C) and its reverse complement. Following this, a Flag tag was inserted at the remaining *Nde*I site between Sumo and Smad2 using synthetic phosphorylated oligonucleotides (5'-T ATG GGA AGC GAC TAC AAA GAC GAT GAC GAC AAG GGA CA) and (5'-TAT GTC CCT TGT CGT CAT CGT CTT TGT AGT CGC TTC CCA) to create pMVP7, encoding His_6-SUMO-Flag-Smad2(1-462)-

GyrA-CBD. All clones were verified by DNA sequencing.

7.27. Cloning of MBP-SARA-SBD

DNA encoding SARA-SBD was amplified from an expression plasmid encoding GST-SARA-SBD (corresponding to amino acids 665-721 of SARA) by PCR using the 5' primer (5'- CGC GGA TCC ATG AGT GCC TCA AGC CAG AGC) and the 3' primer (5'- AT GCG GTC GAC TTA TCT GGG CTG AGC CAC TTC TGC TCC). The DNA was digested with *Bam*HI and *Sal*I and inserted into pMAL-c2X (New England Biolabs) digested with the same enzymes to yield pJM6, which encodes MBP-SARA-SBD.

7.28. Expression and purification of full-length Flag-Smad2 α-thioester/MBP-SARA-SBD complex (57)

E. coli BL21(DE3) cells (Novagen) transformed with either pMVP7 (encoding His_6-SUMO-Flag-Smad2(1-462)-GyrA-CBD) or pJM6 (encoding MBP-SARA-SBD) were grown to an OD_{600} of 0.6 in LB medium supplemented with ampicillin and induced with 0.5 mM IPTG at 37 °C for 4 h. Cells were harvested by centrifugation and combined in a ratio of 3:2::pMVP7:pJM6 by weight, then resuspended in lysis buffer containing 20 mM Tris-Cl, 200 mM NaCl, 10 mM $MgCl_2$, 200 mM 1,6-hexanediol (which stabilizes the protein), pH 7.5, supplemented with protease inhibitors (Complete, Roche) and DNAse I (Sigma). Following cell lysis using a French press, the soluble fraction was collected by centrifugation, filtered through a 5 μm filter, and applied to chitin resin (New England Biolabs) pre-equilibrated with lysis buffer and incubated overnight at 4 °C. The beads were washed with 10 column volumes of lysis buffer and 10 column volumes of 100 mM

Hepes, 200 mM NaCl, 200 mM 1,6-hexanediol, pH 7.3 (buffer A). The SUMO-Flag-Smad2 moiety was cleaved from the resin by incubation of the beads with 3 column volumes of buffer A supplemented with 100 mM MESNa (2-mercaptoethanesulfonic acid, Sigma) for 60 hr at 4 °C. This cleavage step yields His_6-SUMO-Flag-Smad2 with a C-terminal thioester in complex with MBP-SARA-SBD. His_6-SUMO protease (Invitrogen) was added to a final concentration of 0.006 U/mL and the mixture was incubated overnight at 4 °C then passed through a bed of Ni^{2+}-NTA beads (Novagen) to remove His_6-SUMO protease and His_6-SUMO. The Flag-Smad2 α-thioester/MBP-SARA-SBD complex (**57**) was further purified by anion exchange chromatography (Q Sepharose XL, GE Healthcare) at a flow rate of 2 mL/min. Elution was accomplished with a linear gradient of buffer A (20 mM Tris-Cl, 50 mM NaCl, 100 mM 1,6-hexanediol, 75 mM MESNa, pH 7.5) to buffer B (= Buffer A with 1 M NaCl) over 10 column volumes. Fractions containing the desired protein were concentrated (10 kDa cutoff, PES membrane, Vivaspin, Vivascience) to ~15 μM with respect to Flag-Smad2, which was then used in subsequent EPL reactions.

7.29. Expressed protein ligation reactions of full-length Smad2 variants

Expressed protein ligation was performed at 4 °C by adding the appropriate chemically synthesized peptide (**52**, **59**, or **68**) at a final concentration 100-300 μM to a ~15 μM solution of protein thioester **57**. Ligation reactions proceeded for 48–72 hours, then each ligated protein was treated with tris(2-carboxyethyl) phosphine hydrochloride (100 mM) for 20 min and purified by size exclusion chromatography (Superdex 200, GE Healthcare). The flow rate was 0.5 mL/min, and the buffer contained 20 mM HEPES,

150 mM NaCl, 1 mM TCEP, pH 7.3. Peak fractions were snap frozen in liquid nitrogen and stored at -80 °C. The semi-synthetic protein prepared using peptide **52** is **55** (see Figure 4.2). The semi-synthetic protein prepared using peptide **59** is **58** (see Figure 4.9).

7.30. Live-cell Imaging Experiments

HeLa and HaCaT cells were maintained in DMEM supplemented with 10% FBS. For live cell imaging, cells were cultured on separate 35 mm plates (P35G-1.5-7-C-grid, MatTek Corp., Ashland, MA). Frozen aliquots of the caged proteins **56** and **58** were thawed on ice, concentrated to ~100 μM, and were microinjected in the cytoplasm of cells using a commercial microinjection system (Femtojet, Eppendorf). During imaging, cells were incubated in Leibovitz's L-15 Medium and placed under an inverted microscope (Axiovert 200M, Carl Zeiss) equipped with a spinning disk confocal apparatus and a chamber maintained at 37 °C. Images were collected with a Hamamatsu ORCA-ER digital CCD camera mounted on the microscope with a 40× / 1.4 NA oil objective, and filter sets with excitation wavelengths of 488 and 568 nm and emission wavelengths of 525 and 620 nm, respectively. Photolysis of the microinjected protein was performed using irradiation from the microscope's Xenon lamp mounted with a narrow-band filter at 360 nm (corresponding to the standard DAPI excitation light) or at 365 nm with a 200 W Hg lamp fitted with a fiber optic cable (for co-injection experiments involving proteins **56** and **58**). Integrated whole-cell fluorescence intensities were measured after background subtracting the image in a region containing no cells with Metamorph software.

7.31. Nuclear import of caged full-length Smad2 after UV uncaging in live cells

Cells were microinjected and imaged as described in section 7.30. Photolysis of the microinjected protein was performed using irradiation at 365 nm from a 200 W Hg lamp fitted with a fiber optic cable. Integrated whole-cell fluorescence intensities and nuclear fluorescence intensities were measured after background subtracting the image in a region containing no cells with Metamorph software. The nuclear import curves were modeled by the stretched exponential function (Hamada and Dobson, 2002; Morozova-Roche et al., 1999):

$$F = F\infty + \Delta F \exp(-[k_s \bullet t]^n)$$

Where F and $F\infty$ are the fluorescence intensity at time t and the final fluorescence intensity, respectively, and ΔF is the amplitude of the fluorescence change. k_s is the stretched exponential rate constant and is a useful parameter for characterizing complex sigmoidal rate curves. n is a measure of positive cooperativity of the process and was found to be greater than 1 in all cases, suggesting import of Smad2 is a cooperative process with at least two steps that contribute to k_s. The data were fit to the equation using Kaleidagraph software.

7.32. Preparation of Flag-Smad2-MH2-α-thioester and expressed protein ligation with H-CSpSMpS-OH and H-CSPmaMPma-OH to generate proteins 62 and 63

Nucleotides encoding a Flag tag were inserted into the vector pMH2 (pMH2 = pTXB1 containing nucleotides encoding Smad2-MH2(241-462) 5' to the Smad2-MH2 initiator ATG by the same procedure used to insert the Flag tag into full-length Smad2 encoded in

plasmid pMVP7 (see section 7.26). Protein expression and purification was carried out in an identical manner as for the non-tagged Smad2-MH2-α-thioester (see section 7.17), except that SARA-SBD was not added. Expressed protein ligation reactions were carried out as described (section 7.18) with the peptides H-CSpSMpS-OH and H-CSPmaMPma-OH, where pS = phosphoserine and Pma = phosphonomethylenealanine. ESI-MS for both of these proteins is shown in Figure 5.2.

The identities of these proteins were confirmed by manual reconstruction of ESI-MS data (pSer-containing protein **62** found 26,854 ± 4 Da, expected 26,852 Da; Pma-containing protein **63** found 26,851 ± 5 Da, expected).

7.33. Preparation of HeLa cell nuclear extracts.

All steps were performed at 4 °C, unless otherwise specified. Pelleted HeLa-S3 cells (7 x 10^8 cells) resulting from 5 L of suspension culture (obtained on ice from the National Cell Culture Center) were divided into 4 equal portions. Each portion was washed twice with 10 mL of PBS and resuspended in 10 mL of TM2 buffer (10 mM Tris, 2 mM $MgCl_2$, pH 7.4) containing protease inhibitors (Roche Complete Protease Inhibitor Cocktail tablets). The cells were incubated at room temperature for 1 minute and then placed on wet ice for 5 minutes. Digitonin (Sigma) was added to a final concentration of 0.1% and the cells were mixed gently for 5 minutes. Cells were lysed with 6 strokes of the tight-fitting type "B" pestle in a Dounce homogenizer. Lysis was complete as determined by microscopic observation of a sample of the homogenate. The homogenate was centrifuged at 800 RPM for 10 minutes. The pellet containing the nuclei was washed once with 5 mL of TM2 buffer. 5 mL of 60% iodixanol (OptiPrep, Axis-Shield) and 5

mL of buffer D (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM tricine, pH 7.8) were added, followed by centrifugation at 10,000 x *g* for 20 minutes. The pellet containing the purified nuclei was washed with 5 mL of TM2 buffer and resuspended in 1 mL of TM2 buffer. To the resuspended nuclei, 9 mL of NE buffer (10 mM Tris, 100 mM KCl, 0.5 mM EDTA, 5% glycerol, 5 mM $MgCl_2$, pH 7.9) was added. The total resuspended nuclei were divided into 3 portions. One portion (hereafter referred to as NE) was sonicated 6 times for 20 seconds at 35% power, aliquoted, flash frozen with liquid N_2, and stored at -80 °C. Another portion (NE_{TX}) was treated with 1% Triton-X-100 before sonication, aliquoting, and flash freezing. The last portion (NE_D) was treated with 10 μg of DNAse I (Sigma) for 5 minutes before sonication, aliquoting, and flash freezing. Final protein concentration of extracts = 2.5 mg/mL as determined by Bradford assay. For use, all extracts were removed from the freezer and placed at 4 °C to thaw.

7.34. Preparative immunoprecipitation of Flag-Smad2-MH2 interacting proteins from HeLa cell nuclear extracts

All manipulations were carried out at 4 °C, unless otherwise specified. HeLa nuclear extract (NE) was centrifuged at 13,000 RPM for 10 minutes. The supernatant was removed and incubated with Flag-Smad2-MH2 variants. 12 μg of Flag-Smad2-MH2-2P (**62**) or 12 μg of Flag-Smad2-MH2-2Pma (**63**) were each incubated with 0.4 mL of NE (2.5 mg/mL total protein) in the presence of a 10-fold excess of untagged Smad2-MH2-0P (**38**) as competitor. The samples and an additional sample (mock) containing only NE (not containing any recombinant Smad protein), were then incubated with mild agitation on a rotating platform overnight. 5% of each sample by volume was removed for later

analysis (inputs). The remainder of each sample was incubated with 100 μL of anti-FlagM2 agarose beads (Sigma) and agitated by mild rotation for 6 hours. The supernatant was removed by centrifugation and the beads were then washed twice with 1 mL of NE buffer. The beads were drained and bound proteins were eluted by addition of 0.2 mL of a solution containing 3X Flag peptide (Sigma) at 0.2 mg/mL. Following incubation for 1 hour, the supernatant containing eluted proteins was removed by centrifugation. The eluted proteins were precipitated from the supernatant with TCA, washed with ice cold ethanol, dried, and re-dissolved in 16 μL of SDS-PAGE loading buffer. Samples were then separated by SDS-PAGE and stained with coomassie (GelCode Blue, Pierce). Extreme care was taken to minimize contamination of these samples with environmental and human proteins, such as keratin, that could interfere with downstream MS identification experiments. For example, ultraclean SDS-PAGE running running buffer (Biorad) was used. The work bench and all compatible equipment were cleaned with Windex before each manipulation. This has been found to be the most effective agent for cleaning potentially contaminating proteins from surfaces (personal communication with Matthew Sekedat, laboratory of Brian Chait, The Rockefeller University)

7.35. Extraction of proteins from the gel and digestion with trypsin

Each coomassie-stained gel lane was excised and chopped into 2 mm pieces with a gel slicer. Each piece was destained in 55% ammonium bicarbonate (100 mM)/45% acetonitrile with serial washes (4 °C). Gel pieces were dehydrated in acetonitrile, then re-hydrated in 50 mM ammonium bicarbonate, and again dehydrated in acetonitrile. Proteins were digested in-gel with 75 ng of trypsin (Roche Diagnostics, Basel,

Switzerland) per gel band in 50 mM ammonium bicarbonate for 6 h at 37 °C. Tryptic peptides were extracted from the gel pieces with an 8-µl slurry of 1 volume of POROS R2 20 reverse-phase resin (Applied Biosystems, Foster City, CA) to 10 volumes of 5% formic acid/0.2% trifluoroacetic acid (TFA) at 4°C for 8 h. C18 Ziptips (Millipore, Billerica, MA) were conditioned with sequential washes of 0.1% TFA, 50% methanol/20% acetonitrile/0.1% TFA, and 0.1% TFA. POROS R2 20 resin was transferred to the conditioned Ziptips. The POROS R2 20 resin was washed in the Ziptip twice with 0.1% TFA. Tryptic peptides were eluted onto a matrix-assisted laser desorption ionization (MALDI) target compact disk (Krutchinsky et al., 2001) with one-third saturated 2,5-dihydroxybenzoic acid (Lancaster Synthesis, Windham, NH) in 50% methanol: 20% acetonitrile: 0.1% TFA.

7.36. Mass spectrometric identification of proteins

Tryptic peptides were identified by matrix assisted laser desorption ionization (MALDI) mass spectrometric analysis for protein identification (Beavis and Chait, 1996). Mass analysis of the tryptic peptides was performed with a MALDI-quadrupole-quadrupole-time-of-flight (mQqTOF) mass spectrometer (Centaur Technology, Austin, TX; MDS Sciex, Concord, ON, Canada) modified with a compact disk sample stage as described previously (Krutchinsky et al., 2001). The highly accurate MALDI-mass spectrometry (MS) data was used to search the National Center for Biotechnology Center non-redundant protein database with the program ProFound to identify proteins from the tryptic peptide masses (Zhang and Chait, 2000). In addition to matching peptide masses to proteins, ProFound provides lists of peptide masses that can be used to confirm protein

identifications by MS/MS analysis and lists of peptide masses that were not assigned to a protein in the MS analysis. To search for more proteins not identified in the MALDI-MS analysis and to confirm proteins identified in the single-stage MS analysis, amino acid sequence information for the tryptic peptides was obtained by MALDI-MS/MS fragmentation. Using the program M-IT sequencer, confirmation and unassigned masses obtained in the mQqTOF analysis were used to prepare instrument files for MALDI-MS/MS analysis (Krutchinsky et al., 2001). MALDI-MS/MS analysis of the same samples on the same MALDI target compact disk was performed using a modified LCQ Deca XP ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) as described previously (Krutchinsky et al., 2001). MALDI-MS/MS data was used to search the National Center for Biotechnology Information nonredundant protein database with the program Sonar (Genomic Solutions, Ann Arbor, MI) to identify proteins from the tryptic peptide fragmentation masses (Field et al., 2002). Identified proteins were assigned to their corresponding band in the Coomassie-stained gel. In addition to PRMT5, other proteins found in this screen include various heterogeneous ribonucleoproteins that were identified to using both protein **62** and protein **63** as bait. Carbamoyl phosphate synthetase I was identified when protein **63** was used as bait, but not when protein **62** was used as bait.

7.37. Interaction of endogenous PRMT5 with semi-synthetic and recombinant Smad2 variants

Co-immunoprecipitation was performed as in section 7.34, except that competitor untagged protein was not added. In addition to the baits Flag-Smad2-2P (**62**) and Flag-

Smad2-2Pma (**63**), additional samples were prepared using the following proteins as baits: Flag-Smad2-MH2-0P (**61**) and full-length Flag-Smad2-0P (**57**). The non-phosphorylated proteins **61** and **57** lack the last five amino acids of Smad2 and were used after isolation as C-terminal α-thioesters. Samples were run in separate lanes on SDS-PAGE gels and analyzed by western blotting for the presence of Flag-Smad2 variants using anti-FlagM2 mouse monoclonal antibody (Sigma) and for PRMT5 using a monoclonal mouse anti-PRMT5 antibody (BD Trandsuction Labs). Visualization was with an anti-mouse secondary antibody conjugated to HRP (Amersham) using the ECL reagent (Amersham).

7.38. Preparation of recombinant His_6-Sumo-Smad2-MH1 (65)

The MH1 domain (residues 1-185) of Smad2 was cloned into the pET-Sumo vector (Invitrogen) using the TA cloning strategy. Briefly, the MH1 domain was amplified by PCR from pMVP7 with *Taq* polymerase and the primers 5'-TCGTCCATCTTGCCAT TCACGCCG-3' and 5'-TCAATCTCGGTGTGTCGGGGCAC-3' and then ligated to the linearized vector by means of the 3'-A overhangs added by *Taq* polymerase to obtain the plasmid named pMVP14. The His_6-Sumo control construct was made by inserting a stop codon immediately after the Sumo domain in pMVP14 using the QuikChange kit (Stratagene). *E. coli* BL21(DE3) transformed with the obtained plasmids were grown in LB containing 50 μg/mL of kanamycin at 37 °C for 3 h (until A_{600} = 0.6) and then expression was induced for 3 h with 1 mM IPTG. Cells were resuspended in 50 mM phosphate buffer, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol, pH 8 and lysed by means of a French press. Soluble fractions were purified over a Ni^{2+}-NTA

column and purified proteins were eluted with 250 mM imidazole containing buffer after extensive washing with lysis buffer and wash buffer (50 mM phosphate buffer, 300 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, pH 8). Purified proteins thus obtained were dialyzed against 20 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, pH 8, aliquoted and flash frozen. His_6-Sumo-Smad2-MH1 is protein **65** and His_6-Sumo is protein **66**. These two proteins were kindly provided by Miquel Vila-Perelló.

7.39. Interaction of recombinant His_6-Sumo-Smad2-MH1 (65) with endogenous PRMT5

Aliquots of HeLa nuclear extracts NE, NE_{TX}, and NE_D (section 7.33) were thawed at 4 °C. Nuclear extract NE_C was prepared by high speed centrifugation of NE at 100,000 x *g* for 30 minutes at 4 °C. 45 μL of His_6-Sumo-Smad2-MH1 (**65**) or His_6-Sumo (**66**), each at 2 mg/mL, were incubated for 16 hours at 4 °C with each of the four nuclear extracts (220 μL of each extract at 2.5 mg/mL total protein). Following incubation, each sample was supplemented with 10 mM imidazole to reduce non-specific binding in the next step. Each sample was then mixed with 10 μL of Ni^{2+}-NTA beads (Novagen) for 4 hours at 4 °C with gentle agitation on a nutator. The beads were collected by centrifugation and the supernatant was drained. The beads were then washed with 5 x 150 μL of Ni^{2+}-NTA wash buffer (Novagen) containing 10 mM imidazole. Bound proteins were eluted into 60 μL of Ni^{2+}-NTA wash buffer containing 0.5 M imidazole by incubation of the solution with the beads for 30 minutes. The eluate was collected by centrifugation and 20 μL of 4X SDS-PAGE was added followed by boiling for 5 minutes. The samples were resolved by SDS-PAGE and analyzed by western blotting for His_6-Sumo-Smad2-MH1 with an

anti-penta His monoclonal mouse antibody (Qiagen) and for PRMT5 with an anti-PRMT5 rabbit polyclonal antibody (Upstate). Visualization was performed on an infrared fluorescence scanner (Odyssey, Licor) with anti-mouse and anti-rabbit secondary antibodies conjugated to infrared fluorophores (Licor).

7.40. Interaction of endogenous Smad2 with Flag-PRMT5 expressed in HEK293T cells

HEK293T cells at ~70% confluency in 10-cm dishes were transfected using Fugene 6 (Roche) with either 5 μg of Flag-PRMT5 expression vector (a kind gift of Stephen Nimer, MSKCC) or 5 μg of mock plasmid encoding EGFP-I_N using the manufacturer's recommended transfection procedure. Only one plasmid was used per dish i.e., plasmids were not co-transfected. Following transfection, cells were incubated at 37 °C for 46 hours at which point they were washed with 7 mL of cold PBS and drained. 0.45 mL of cold lysis buffer (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 5 mM NaF, 10 mM β-glycerophosphate, 5 mM sodium pyrophosphate, 10% glycerol, 2.5 mM DTT, 1% NP-40, pH 7.5, containing Roche complete protease inhibitors) was added to each plate and incubated for 5 minutes on ice. Cells were scraped with a rubber policeman and transferred into a 1.5 mL microfuge tube. Cells were homogenized with 10 passes through a 2 mL Dounce homogenizer using the tight fitting type "B" pestle. Homogenates were sonicated 4 x 3 seconds at 35% power with 5 seconds in between each pulse and centrifuged for 15 minutes at 13,000 RPM at 4 °C. The supernatant was used in subsequent steps. 5% of each sample was saved for later analysis (inputs). The remainder of each sample was incubated with 25 μL of anti-FlagM2 agarose beads

(Sigma) in a small column format for 2 hours at 4 °C with gentle agitation on a nutator. The beads were drained and washed 7 x 1 mL with lysis buffer. Bound proteins were eluted by incubation for 1 hour at 4 °C followed by 10 minutes at room temperature with 60 μL of a solution containing 3X Flag peptide (Sigma) at 0.4 mg/mL. 20 μL of 4X SDS-PAGE loading buffer was added to the eluate, which was then boiled for 5 minutes. Samples were run in separate lanes on SDS-PAGE gels and analyzed by western blotting for the presence of Smad2 using an anti-Smad2 rabbit polyclonal antibody (Upstate) and for Flag-PRMT5 using a polyclonal rabbit anti-Flag antibody (Sigma). Visualization was performed on an infrared fluorescence scanner (Odyssey, Licor) with an anti-rabbit antibody conjugated to an infrared fluorophore (Licor).

References

Adam, S. A., Sterne-Marr, R., and Gerace, L. (1992). Nuclear protein import using digitonin-permeabilized cells. Methods Enzymol *219*, 97-110.

Allen, J. J., Lazerwith, S. E., and Shokat, K. M. (2005). Bio-orthogonal affinity purification of direct kinase substrates. J Am Chem Soc *127*, 5288-5289.

Allen, J. J., Li, M., Brinkworth, C. S., Paulson, J. L., Wang, D., Hubner, A., Chou, W. H., Davis, R. J., Burlingame, A. L., Messing, R. O.*, et al.* (2007). A semisynthetic epitope for kinase substrates. Nat Methods *4*, 511-516.

Amente, S., Napolitano, G., Licciardo, P., Monti, M., Pucci, P., Lania, L., and Majello, B. (2005). Identification of proteins interacting with the RNAPII FCP1 phosphatase: FCP1 forms a complex with arginine methyltransferase PRMT5 and it is a substrate for PRMT5-mediated methylation. FEBS Lett *579*, 683-689.

Ancelin, K., Lange, U. C., Hajkova, P., Schneider, R., Bannister, A. J., Kouzarides, T., and Surani, M. A. (2006). Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. Nat Cell Biol *8*, 623-630.

Andersen, H. S., Iversen, L. F., Jeppesen, C. B., Branner, S., Norris, K., Rasmussen, H. B., Moller, K. B., and Moller, N. P. (2000). 2-(oxalylamino)-benzoic acid is a general, competitive inhibitor of protein-tyrosine phosphatases. J Biol Chem *275*, 7101-7108.

Ando, R., Mizuno, H., and Miyawaki, A. (2004). Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. Science *306*, 1370-1373.

Arabaci, G., Guo, X. C., Beebe, K. D., Coggeshall, K. M., and Pei, D. (1999). alpha-Haloacetophenone Derivatives As Photoreversible Covalent Inhibitors of Protein Tyrosine Phosphatases. J. Am. Chem. Soc. *121*, 5085-5086.

Beavis, R. C., and Chait, B. T. (1996). Matrix-assisted laser desorption ionization mass-spectrometry of proteins. Methods Enzymol *270*, 519-551.

Bedford, M. T., and Richard, S. (2005). Arginine methylation an emerging regulator of protein function. Mol Cell *18*, 263-272.

Belshaw, P. J., Schoepfer, J. G., Liu, K. Q., Morrison, K. L., and Schreiber, S. L. (1995). Rational Design of Orthogonal Receptor-Ligand Combinations. Angewandte Chemie-International Edition in English *34*, 2129-2132.

Berkowitz, D. B., Bose, M., Pfannenstiel, T. J., and Doukov, T. (2000). alpha-fluorinated phosphonates as substrate mimics for glucose 6-phosphate dehydrogenase: the CHF stereochemistry matters. J Org Chem *65*, 4498-4508.

Bernales, S., Papa, F. R., and Walter, P. (2006). Intracellular signaling by the unfolded protein response. Annu Rev Cell Dev Biol *22*, 487-508.
Bialy, L., and Waldmann, H. (2005). Inhibitors of protein tyrosine phosphatases: next-generation drugs? Angew Chem Int Ed Engl *44*, 3814-3839.

Bishop, A. C., and Blair, E. R. (2006). A gatekeeper residue for inhibitor sensitization of protein tyrosine phosphatases. Bioorg Med Chem Lett *16*, 4002-4006.

Bishop, A. C., Shah, K., Liu, Y., Witucki, L., Kung, C., and Shokat, K. M. (1998). Design of allele-specific inhibitors to probe protein kinase signaling. Curr Biol *8*, 257-266.

Bishop, A. C., Ubersax, J. A., Petsch, D. T., Matheos, D. P., Gray, N. S., Blethrow, J., Shimizu, E., Tsien, J. Z., Schultz, P. G., Rose, M. D.*, et al.* (2000). A chemical switch for inhibitor-sensitive alleles of any protein kinase. Nature *407*, 395-401.

Blair, J. A., Rauh, D., Kung, C., Yun, C. H., Fan, Q. W., Rode, H., Zhang, C., Eck, M. J., Weiss, W. A., and Shokat, K. M. (2007). Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. Nat Chem Biol *3*, 229-238.

Branscombe, T. L., Frankel, A., Lee, J. H., Cook, J. R., Yang, Z., Pestka, S., and Clarke, S. (2001). PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. J Biol Chem *276*, 32971-32976.

Brumbaugh, J., Schleifenbaum, A., Gasch, A., Sattler, M., and Schultz, C. (2006). A dual parameter FRET probe for measuring PKC and PKA activity in living cells. J Am Chem Soc *128*, 24-25.

Burke, T. R., Jr. (2006). Design and synthesis of phosphonodifluoromethyl phenylalanine (F2Pmp): a useful phosphotyrosyl mimetic. Curr Top Med Chem *6*, 1465-1471.

Burnett, G., and Kennedy, E. P. (1954). The enzymatic phosphorylation of proteins. J Biol Chem *211*, 969-980.

Carter, P., and Wells, J. A. (1987). Engineering enzyme specificity by "substrate-assisted catalysis". Science *237*, 394-399.

Chacko, B. M., Qin, B. Y., Tiwari, A., Shi, G., Lam, S., Hayward, L. J., De Caestecker, M., and Lin, K. (2004). Structural basis of heteromeric smad protein assembly in TGF-beta signaling. Mol Cell *15*, 813-823.

Chang, C. y., Fernandez, T., Panchal, R., and Bayley, H. (1998). Caged Catalytic Subunit of cAMP-Dependent Protein Kinase. J. Am. Chem. Soc. *120*, 7661-7662.

Chen, C. A., Yeh, R. H., and Lawrence, D. S. (2002). Design and Synthesis of a Fluorescent Reporter of Protein Kinase Activity. J. Am. Chem. Soc. *124*, 3840-3841.

Chen, H. B., Shen, J., Ip, Y. T., and Xu, L. (2006). Identification of phosphatases for Smad in the BMP/DPP pathway. Genes Dev *20*, 648-653.

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002). Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J Am Chem Soc *124*, 9026-9027.

Cohen, M. S., Hadjivassiliou, H., and Taunton, J. (2007). A clickable inhibitor reveals context-dependent autoactivation of p90 RSK. Nat Chem Biol *3*, 156-160.

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. Trends Biochem Sci *25*, 596-601.

Cohen, P. (2002). The origins of protein phosphorylation. Nat Cell Biol *4*, E127-130.

Cori, G. T., and Cori, C. F. (1945). The enzymatic conversion of phosphorylase-a to phosphorylase-b. Journal of Biological Chemistry *158*, 321-332.

Cori, G. T., and Green, A. A. (1943). Crystalline muscle phosphorylase II. Prosthetic group. Journal of Biological Chemistry *151*, 31-38.

Cowgill, R. W., and Cori, C. F. (1955). The conversion of inactive phosphorylase to phosphorylase b and phosphorylase a in lobster muscle extract. J Biol Chem *216*, 133-140.

Craik, C. S., Roczniak, S., Largman, C., and Rutter, W. J. (1987). The catalytic role of the active site aspartic acid in serine proteases. Science *237*, 909-913.

Curley, K., and Lawrence, D. S. (1998). Photoactivation of a Signal Transduction Pathway in Living Cells. J. Am. Chem. Soc. *120*, 8573-8574.

Curley, K., and Lawrence, D. S. (1999). Light-activated proteins. Curr Opin Chem Biol *3*, 84-88.

Dacwag, C. S., Ohkawa, Y., Pal, S., Sif, S., and Imbalzano, A. N. (2007). The protein arginine methyltransferase Prmt5 is required for myogenesis because it facilitates ATP-dependent chromatin remodeling. Mol Cell Biol *27*, 384-394.

Daile, P., Carnegie, P. R., and Young, J. D. (1975). Synthetic substrate for cyclic AMP-dependent protein kinase. Nature *257*, 416-418.

Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994). Synthesis of proteins by native chemical ligation. Science *266*, 776-779.

Deininger, M., Buchdunger, E., and Druker, B. J. (2005). The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood *105*, 2640-2653.

Deiters, A., Cropp, T. A., Mukherji, M., Chin, J. W., Anderson, J. C., and Schultz, P. G. (2003). Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. J Am Chem Soc *125*, 11782-11783.

Dephoure, N., Howson, R. W., Blethrow, J. D., Shokat, K. M., and O'Shea, E. K. (2005). Combining chemical genetics and proteomics to identify protein kinase substrates. Proc Natl Acad Sci U S A *102*, 17940-17945.

Domchek, S. M., Auger, K. R., Chatterjee, S., Burke, T. R., Jr., and Shoelson, S. E. (1992). Inhibition of SH2 domain/phosphoprotein association by a nonhydrolyzable phosphonopeptide. Biochemistry *31*, 9865-9870.

Dong, C., Li, Z., Alvarez, R., Jr., Feng, X. H., and Goldschmidt-Clermont, P. J. (2000). Microtubule binding to Smads may regulate TGF beta activity. Mol Cell *5*, 27-34.

Duan, X., Liang, Y. Y., Feng, X. H., and Lin, X. (2006). Dephosphorylation of Smad1 in the BMP signaling pathway by PPM1A. J Biol Chem.

Dube, N., and Tremblay, M. L. (2005). Involvement of the small protein tyrosine phosphatases TC-PTP and PTP1B in signal transduction and diseases: from diabetes, obesity to cell cycle, and cancer. Biochim Biophys Acta *1754*, 108-117.

Dunn, N. R., Koonce, C. H., Anderson, D. C., Islam, A., Bikoff, E. K., and Robertson, E. J. (2005). Mice exclusively expressing the short isoform of Smad2 develop normally and are viable and fertile. Genes Dev *19*, 152-163.

Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J., and Schultz, P. G. (1991). Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. Methods Enzymol *202*, 301-336.

Erlanson, D. A., Chytil, M., and Verdine, G. L. (1996). The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. Chem Biol *3*, 981-991.

Erlanson, D. A., McDowell, R. S., He, M. M., Randal, M., Simmons, R. L., Kung, J., Waight, A., and Hansen, S. K. (2003). Discovery of a new phosphotyrosine mimetic for PTP1B using breakaway tethering. J Am Chem Soc *125*, 5602-5603.

Evans, M. J., and Cravatt, B. F. (2006). Mechanism-Based Profiling of Enzyme Families. Chem. Rev. *106*, 3279-3301.

Fabbrizio, E., El Messaoudi, S., Polanowska, J., Paul, C., Cook, J. R., Lee, J. H., Negre, V., Rousset, M., Pestka, S., Le Cam, A., and Sardet, C. (2002). Negative regulation of transcription by the type II arginine methyltransferase PRMT5. EMBO Rep *3*, 641-645.

Feng, X. H., and Derynck, R. (2005). Specificity and versatility in tgf-beta signaling through Smads. Annu Rev Cell Dev Biol *21*, 659-693.

Field, H. I., Fenyo, D., and Beavis, R. C. (2002). RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. Proteomics *2*, 36-47.

Fischer, E. H., and Krebs, E. G. (1955). Conversion of phosphorylase b to phosphorylase a in muscle extracts. J Biol Chem *216*, 121-132.

Folta-Stogniew, E., and Williams, K. M. (1999). Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. J. Biomol. Tech. *10*, 51-63.

Friesen, W. J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G. S., Van Duyne, G., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001). The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol Cell Biol *21*, 8289-8300.

Furuta, T., Takeuchi, H., Isozaki, M., Takahashi, Y., Kanehara, M., Sugimoto, M., Watanabe, T., Noguchi, K., Dore, T. M., Kurahashi, T.*, et al.* (2004). Bhc-cNMPs as either water-soluble or membrane-permeant photoreleasable cyclic nucleotides for both one- and two-photon excitation. Chembiochem *5*, 1119-1128.

Furuta, T., Wang, S. S., Dantzker, J. L., Dore, T. M., Bybee, W. J., Callaway, E. M., Denk, W., and Tsien, R. Y. (1999). Brominated 7-hydroxycoumarin-4-ylmethyls: photolabile protecting groups with biologically useful cross-sections for two photon photolysis. Proc Natl Acad Sci U S A *96*, 1193-1200.

Gentle, I. E., De Souza, D. P., and Baca, M. (2004). Direct production of proteins with N-terminal cysteine for site-specific conjugation. Bioconjug Chem *15*, 658-663.

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A *100*, 6940-6945.

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Ghosh, M., Ichetovkin, I., Song, X., Condeelis, J. S., and Lawrence, D. S. (2002). A new strategy for caging proteins regulated by kinases. J Am Chem Soc *124*, 2440-2441.

Ghosh, M., Song, X., Mouneimne, G., Sidani, M., Lawrence, D. S., and Condeelis, J. S. (2004). Cofilin promotes actin polymerization and defines the direction of cell motility. Science *304*, 743-746.

Giriat, I., and Muir, T. W. (2003). Protein semi-synthesis in living cells. J Am Chem Soc *125*, 7180-7181.

Goard, M., Aakalu, G., Fedoryak, O. D., Quinonez, C., St Julien, J., Poteet, S. J., Schuman, E. M., and Dore, T. M. (2005). Light-mediated inhibition of protein synthesis. Chem Biol *12*, 685-693.

Hackenberger, C. P. R., Chen, M. M., and Imperiali, B. (2006). Expression of N-terminal Cys-protein fragments using an intein refolding strategy. Bioorganic & Medicinal Chemistry *14*, 5043-5048.

Hahn, M. E., and Muir, T. W. (2004). Photocontrol of Smad2, a multiphosphorylated cell-signaling protein, through caging of activating phosphoserines. Angew Chem Int Ed Engl *43*, 5800-5803.

Hahn, M. E., and Muir, T. W. (2005). Manipulating proteins with chemistry: a cross-section of chemical biology. Trends Biochem Sci *30*, 26-34.

Hamada, D., and Dobson, C. M. (2002). A kinetic study of beta-lactoglobulin amyloid fibril formation promoted by urea. Protein Sci *11*, 2417-2426.

Hasan, A., Stengele, K. P., Geigrich, H., Cornwell, P., Isham, K. R., Sachleben, R. A., Pfleiderer, W., and Foote, R. S. (1997). Photolabile protecting groups for nucleosides: Synthesis and photodeprotection rates. Tetrahedron *53*, 4247-4264.

Hata, A., Lo, R. S., Wotton, D., Lagna, G., and Massague, J. (1997). Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4. Nature *388*, 82-87.

Hauser, P. S., and Ryan, R. O. (2007). Expressed protein ligation using an N-terminal cysteine containing fragment generated in vivo from a pelB fusion protein. Protein Expr Purif *54*, 227-233.

He, W., Dorn, D. C., Erdjument-Bromage, H., Tempst, P., Moore, M. A., and Massague, J. (2006). Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. Cell *125*, 929-941.

Hegeman, A. D., Harms, A. C., Sussman, M. R., Bunner, A. E., and Harper, J. F. (2004). An isotope labeling strategy for quantifying the degree of phosphorylation at multiple sites in proteins. J Am Soc Mass Spectrom *15*, 647-653.

Ho, M., Bramson, H. N., Hansen, D. E., Knowles, J. R., and Kaiser, E. T. (1988). Stereochemical course of the phospho group transfer catalyzed by cAMP-dependent protein kinase. J. Am. Chem. Soc. *110*, 2680-2681.

Hoffman, H. E., Blair, E. R., Johndrow, J. E., and Bishop, A. C. (2005). Allele-specific inhibitors of protein tyrosine phosphatases. J Am Chem Soc *127*, 2824-2825.

Holmes, C. P. (1997). Model Studies for New o-Nitrobenzyl Photolabile Linkers: Substituent Effects on the Rates of Photochemical Cleavage. J. Org. Chem. *62*, 2370-2380.

Humble, E., Berglund, L., Titanji, V., Ljungstrom, O., Edlund, B., Zetterqvist, O., and Engstrom, L. (1975). Non-dependence on native structure of pig liver pyruvate kinase when used as a substrate for cyclic 3',5'-AMP-stimulated protein kinase. Biochem Biophys Res Commun *66*, 614-621.

Humphrey, D., Rajfur, Z., Vazquez, M. E., Scheswohl, D., Schaller, M. D., Jacobson, K., and Imperiali, B. (2005). In situ photoactivation of a caged phosphotyrosine peptide derived from focal adhesion kinase temporarily halts lamellar extension of single migrating tumor cells. J Biol Chem *280*, 22091-22101.

Hunter, T. (2000). Signaling--2000 and beyond. Cell *100*, 113-127.

Huse, M., Holford, M. N., Kuriyan, J., and Muir, T. W. (2000). Semisynthesis of Hyperphosphorylated Type I TGF beta Receptor: Addressing the Mechanism of Kinase Activation. J. Am. Chem. Soc. *122*, 8337-8338.

Huse, M., Muir, T. W., Xu, L., Chen, Y. G., Kuriyan, J., and Massague, J. (2001). The TGF beta receptor activation process: an inhibitor- to substrate-binding switch. Mol Cell *8*, 671-682.

Hwang, Y. W., and Miller, D. L. (1987). A mutation that alters the nucleotide specificity of elongation factor Tu, a GTP regulatory protein. J Biol Chem *262*, 13081-13085.

Inman, G. J., Nicolas, F. J., and Hill, C. S. (2002). Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. Mol Cell *10*, 283-294.

Iversen, L. F., Andersen, H. S., Branner, S., Mortensen, S. B., Peters, G. H., Norris, K., Olsen, O. H., Jeppesen, C. B., Lundt, B. F., Ripka, W.*, et al.* (2000). Structure-based design of a low molecular weight, nonphosphorus, nonpeptide, and highly selective inhibitor of protein-tyrosine phosphatase 1B. J Biol Chem *275*, 10300-10307.

Kaplan, J. H., Forbush, B., 3rd, and Hoffman, J. F. (1978). Rapid photolytic release of adenosine 5'-triphosphate from a protected analogue: utilization by the Na:K pump of human red blood cell ghosts. Biochemistry *17*, 1929-1935.

Kapoor, T. M., and Mitchison, T. J. (1999). Allele-specific activators and inhibitors for kinesin. Proc Natl Acad Sci U S A *96*, 9106-9111.

Kemp, B. E., Benjamini, E., and Krebs, E. G. (1976). Synthetic hexapeptide substrates and inhibitors of 3':5'-cyclic AMP-dependent protein kinase. Proc Natl Acad Sci U S A *73*, 1038-1042.

Kemp, B. E., Bylund, D. B., Huang, T. S., and Krebs, E. G. (1975). Substrate specificity of the cyclic AMP-dependent protein kinase. Proc Natl Acad Sci U S A *72*, 3448-3452.

Kenski, D. M., Zhang, C., von Zastrow, M., and Shokat, K. M. (2005). Chemical genetic engineering of G protein-coupled receptor kinase 2. J Biol Chem *280*, 35051-35061.

Kim, J., Johnson, K., Chen, H. J., Carroll, S., and Laughon, A. (1997). Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. Nature *388*, 304-308.

Kim, K., and Cole, P. A. (1997). Measurement of a Bronsted Nucleophile Coefficient and Insights into the Transition State for a Protein Tyrosine Kinase. J. Am. Chem. Soc. *119*, 11096-11097.

Kim, K., and Cole, P. A. (1998). Kinetic Analysis of a Protein Tyrosine Kinase Reaction Transition State in the Forward and Reverse Directions. J. Am. Chem. Soc. *120*, 6851-6858.

Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., Gibson, B. W., and Shokat, K. M. (2003). Phosphospecific proteolysis for mapping sites of protein phosphorylation. Nat Biotechnol *21*, 1047-1054.

Knight, Z. A., and Shokat, K. M. (2005). Features of selective kinase inhibitors. Chem Biol *12*, 621-637.

Knockaert, M., Sapkota, G., Alarcon, C., Massague, J., and Brivanlou, A. H. (2006). Unique players in the BMP pathway: small C-terminal domain phosphatases dephosphorylate Smad1 to attenuate BMP signaling. Proc Natl Acad Sci U S A *103*, 11940-11945.

Kornberg, A. (2003). Ten commandments of enzymology, amended. Trends Biochem Sci *28*, 515-517.

Krebs, E. G., and Fischer, E. H. (1956). The phosphorylase b to a converting enzyme of rabbit skeletal muscle. Biochim Biophys Acta *20*, 150-157.

Kretzschmar, M., Doody, J., Timokhina, I., and Massague, J. (1999). A mechanism of repression of TGFbeta/ Smad signaling by oncogenic Ras. Genes Dev *13*, 804-816.

Krutchinsky, A. N., Kalkum, M., and Chait, B. T. (2001). Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. Anal Chem *73*, 5066-5077.

Kuder, N., Zelinski, T., Pathak, T., Seitz, O., and Waldmann, H. (2000). Synthesis of a triply phosphorylated pentapeptide from human tau-protein. Bioorg Med Chem *8*, 2433-2439.

Kumar, S., Zhou, B., Liang, F., Wang, W. Q., Huang, Z., and Zhang, Z. Y. (2004). Activity-based probes for protein tyrosine phosphatases. Proc Natl Acad Sci U S A *101*, 7943-7948.

Kumar, S., Zhou, B., Liang, F., Yang, H., Wang, W. Q., and Zhang, Z. Y. (2006). Global analysis of protein tyrosine phosphatase activity with ultra-sensitive fluorescent probes. J Proteome Res *5*, 1898-1905.

Lagna, G., Hata, A., Hemmati-Brivanlou, A., and Massague, J. (1996). Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. Nature *383*, 832-836.

Lawrence, D. S. (2005). The preparation and in vivo applications of caged peptides and proteins. Curr Opin Chem Biol *9*, 570-575.

Lawrence, D. S., and Wang, Q. (2007). Seeing is believing: peptide-based fluorescent sensors of protein tyrosine kinase activity. Chembiochem *8*, 373-378.

Lee, S. Y., Liang, F., Guo, X. L., Xie, L., Cahill, S. M., Blumenstein, M., Yang, H., Lawrence, D. S., and Zhang, Z. Y. (2005). Design, construction, and intracellular activation of an intramolecularly self-silenced signal transduction inhibitor. Angew Chem Int Ed Engl *44*, 4242-4244.

Lin, X., Duan, X., Liang, Y. Y., Su, Y., Wrighton, K. H., Long, J., Hu, M., Davis, C. M., Wang, J., Brunicardi, F. C.*, et al.* (2006). PPM1A functions as a Smad phosphatase to terminate TGFbeta signaling. Cell *125*, 915-928.

Lipmann, F. A., and Levene, P. A. (1932). Serinephosphoric acid obtained on hydrolysis of vitellinic acid. J Biol Chem *98*, 109-114.

Lo, L. C., Pang, T. L., Kuo, C. H., Chiang, Y. L., Wang, H. Y., and Lin, J. J. (2002). Design and synthesis of class-selective activity probes for protein tyrosine phosphatases. J Proteome Res *1*, 35-40.

Loudwig, S., Nicolet, Y., Masson, P., Fontecilla-Camps, J. C., Bon, S., Nachon, F., and Goeldner, M. (2003). Photoreversible inhibition of cholinesterases: catalytic serine-labeled caged butyrylcholinesterase. Chembiochem *4*, 762-767.

Lu, W., Gong, D., Bar-Sagi, D., and Cole, P. A. (2001). Site-specific incorporation of a phosphotyrosine mimetic reveals a role for tyrosine phosphorylation of SHP-2 in cell signaling. Mol Cell *8*, 759-769.

Lu, W., Shen, K., and Cole, P. A. (2003). Chemical dissection of the effects of tyrosine phosphorylation of SHP-2. Biochemistry *42*, 5461-5468.

Lue, R. Y., Chen, G. Y., Hu, Y., Zhu, Q., and Yao, S. Q. (2004). Versatile protein biotinylation strategies for potential high-throughput proteomics. J Am Chem Soc *126*, 1055-1062.

Macias-Silva, M., Abdollah, S., Hoodless, P. A., Pirone, R., Attisano, L., and Wrana, J. L. (1996). MADR2 is a substrate of the TGFbeta receptor and its phosphorylation is required for nuclear accumulation and signaling. Cell *87*, 1215-1224.

Maly, D. J., Allen, J. A., and Shokat, K. M. (2004). A mechanism-based cross-linker for the identification of kinase-substrate pairs. J Am Chem Soc *126*, 9160-9161.

Mann, M., and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. Nat Biotechnol *21*, 255-261.

Marras, S. A. (2006). Selection of fluorophore and quencher pairs for fluorescent nucleic acid hybridization probes. Methods Mol Biol *335*, 3-16.

Marriott, G., Ottl, J., Heidecker, M., and Gabriel, D. (1998). Light-directed activation of protein activity from caged protein conjugates. Methods Enzymol *291*, 95-116.

Martin, G. S. (2001). The hunting of the Src. Nat Rev Mol Cell Biol *2*, 467-475.

Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev *19*, 2783-2810.

Mathys, S., Evans, T. C., Chute, I. C., Wu, H., Chong, S., Benner, J., Liu, X. Q., and Xu, M. Q. (1999). Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. Gene *231*, 1-13.

Mayer, G., and Heckel, A. (2006). Biologically active molecules with a "light switch". Angew Chem Int Ed Engl *45*, 4900-4921.

Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C., and Fischer, U. (2001). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. Curr Biol *11*, 1990-1994.

Meister, G., and Fischer, U. (2002). Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. Embo J *21*, 5853-5863.

Mendel, D., Ellman, J. A., and Schultz, P. G. (1991). Construction of a light-activated protein by unnatural amino acid mutagenesis. J. Am. Chem. Soc. *113*, 2758-2760.

Merrifield, B. (1996). The chemical synthesis of proteins. Protein Sci *5*, 1947-1951.

Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. J. Am. Chem. Soc. *85*, 2149-2154.

Mildvan, A. S. (1997). Mechanisms of signaling and related enzymes. Proteins *29*, 401-416.

Minor, D. L., Jr., and Kim, P. S. (1994). Measurement of the beta-sheet-forming propensities of amino acids. Nature *367*, 660-663.

Morozova-Roche, L. A., Jones, J. A., Noppe, W., and Dobson, C. M. (1999). Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme. J Mol Biol *289*, 1055-1073.

Mossessova, E., and Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol Cell *5*, 865-876.

Muir, T. W. (2003). Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem *72*, 249-289.

Muir, T. W., Sondhi, D., and Cole, P. A. (1998). Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci U S A *95*, 6705-6710.

Nerbonne, J. M. (1996). Caged compounds: tools for illuminating neuronal responses and connections. Curr Opin Neurobiol *6*, 379-386.

Nguyen, A., Rothman, D. M., Stehn, J., Imperiali, B., and Yaffe, M. B. (2004). Caged phosphopeptides reveal a temporal role for 14-3-3 in G1 arrest and S-phase checkpoint function. Nat Biotechnol *22*, 993-1000.

Nicolas, F. J., De Bosscher, K., Schmierer, B., and Hill, C. S. (2004). Analysis of Smad nucleocytoplasmic shuttling in living cells. J Cell Sci *117*, 4113-4125.

Noren, C. J., Wang, J., and Perler, F. B. (2000). Dissecting the Chemistry of Protein Splicing and Its Applications. Angew Chem Int Ed Engl *39*, 450-466.

Noren-Muller, A., Reis-Correa, I., Jr., Prinz, H., Rosenbaum, C., Saxena, K., Schwalbe, H. J., Vestweber, D., Cagna, G., Schunk, S., Schwarz, O.*, et al.* (2006). Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. Proc Natl Acad Sci U S A *103*, 10606-10611.

Oda, Y., Nagasu, T., and Chait, B. T. (2001). Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat Biotechnol *19*, 379-382.

Ong, S. E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods *1*, 119-126.

Ottesen, J. J., Huse, M., Sekedat, M. D., and Muir, T. W. (2004). Semisynthesis of phosphovariants of Smad2 reveals a substrate preference of the activated T beta RI kinase. Biochemistry *43*, 5698-5706.

Otzen, D. E., and Fersht, A. R. (1995). Side-chain determinants of beta-sheet stability. Biochemistry *34*, 5718-5724.

Ovaa, H., van Swieten, P. F., Kessler, B. M., Leeuwenburgh, M. A., Fiebiger, E., van den Nieuwendijk, A. M., Galardy, P. J., van der Marel, G. A., Ploegh, H. L., and Overkleeft, H. S. (2003). Chemistry in living cells: detection of active proteasomes by a two-step labeling strategy. Angew Chem Int Ed Engl *42*, 3626-3629.

Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P., and Sif, S. (2004). Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. Mol Cell Biol *24*, 9630-9645.

Pal, S., Yun, R., Datta, A., Lacomis, L., Erdjument-Bromage, H., Kumar, J., Tempst, P., and Sif, S. (2003). mSin3A/histone deacetylase 2- and PRMT5-containing Brg1 complex is involved in transcriptional repression of the Myc target gene cad. Mol Cell Biol *23*, 7475-7487.

Papa, F. R., Zhang, C., Shokat, K., and Walter, P. (2003). Bypassing a kinase activity with an ATP-competitive drug. Science *302*, 1533-1537.

Parang, K., Kohn, J. A., Saldanha, S. A., and Cole, P. A. (2002). Development of photo-crosslinking reagents for protein kinase-substrate interactions. FEBS Lett *520*, 156-160.

Parang, K., Till, J. H., Ablooglu, A. J., Kohanski, R. A., Hubbard, S. R., and Cole, P. A. (2001). Mechanism-based design of a protein kinase inhibitor. Nat Struct Biol *8*, 37-41.

Park, C. H., and Givens, R. S. (1997). New Photoactivated Protecting Groups. 6. p-Hydroxyphenacyl: A Phototrigger for Chemical and Biochemical Probes. J. Am. Chem. Soc. *119*, 2453-2463.

Patterson, G. H., and Lippincott-Schwartz, J. (2002). A photoactivatable GFP for selective photolabeling of proteins and cells. Science *297*, 1873-1877.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445-452.

Pawson, T., and Scott, J. D. (2005). Protein phosphorylation in signaling--50 years and counting. Trends Biochem Sci *30*, 286-290.

Pellois, J. P., Hahn, M. E., and Muir, T. W. (2004). Simultaneous triggering of protein activity and fluorescence. J Am Chem Soc *126*, 7170-7171.

Petersson, E. J., Brandt, G. S., Zacharias, N. M., Dougherty, D. A., and Lester, H. A. (2003). Caging proteins through unnatural amino acid mutagenesis. Methods Enzymol *360*, 258-273.

Peyker, A., Rocks, O., and Bastiaens, P. I. (2005). Imaging activation of two Ras isoforms simultaneously in a single cell. Chembiochem *6*, 78-85.

Powers, T., and Walter, P. (1995). Reciprocal stimulation of GTP hydrolysis by two directly interacting GTPases. Science *269*, 1422-1424.

Puius, Y. A., Zhao, Y., Sullivan, M., Lawrence, D. S., Almo, S. C., and Zhang, Z. Y. (1997). Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: a paradigm for inhibitor design. Proc Natl Acad Sci U S A *94*, 13420-13425.

Qiao, Y., Molina, H., Pandey, A., Zhang, J., and Cole, P. A. (2006). Chemical rescue of a mutant enzyme in living cells. Science *311*, 1293-1297.

Qin, B. Y., Chacko, B. M., Lam, S. S., de Caestecker, M. P., Correia, J. J., and Lin, K. (2001). Structural basis of Smad1 activation by receptor kinase phosphorylation. Mol Cell *8*, 1303-1312.

Rall, T. W., Sutherland, E. W., and Wosilait, W. D. (1956). The relationship of epinephrine and glucagon to liver phosphorylase. III. Reactivation of liver phosphorylase in slices and in extracts. J Biol Chem *218*, 483-495.

Rall, T. W., Wosilait, W. D., and Sutherland, E. W. (1956). The interconversion of phosphorylase a and phosphorylase b from dog heart muscle. Biochim Biophys Acta *20*, 69-76.

Rappsilber, J., Friesen, W. J., Paushkin, S., Dreyfuss, G., and Mann, M. (2003). Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. Anal Chem *75*, 3107-3114.

Ratcliffe, S. J., Yi, T., and Khandekar, S. S. (2007). Synthesis and characterization of 5'-p-fluorosulfonylbenzoyl-2' (or 3')-(biotinyl)adenosine as an activity-based probe for protein kinases. J Biomol Screen *12*, 126-132.

Reguly, T., and Wrana, J. L. (2003). In or out? The dynamics of Smad nucleocytoplasmic shuttling. Trends Cell Biol *13*, 216-220.

Rho, J., Choi, S., Seong, Y. R., Cho, W. K., Kim, S. H., and Im, D. S. (2001). Prmt5, which forms distinct homo-oligomers, is a member of the protein-arginine methyltransferase family. J Biol Chem *276*, 11393-11401.

Ross, S., Cheung, E., Petrakis, T. G., Howell, M., Kraus, W. L., and Hill, C. S. (2006). Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. Embo J *25*, 4490-4502.

Rothman, D. M., Petersson, E. J., Vazquez, M. E., Brandt, G. S., Dougherty, D. A., and Imperiali, B. (2005). Caged phosphoproteins. J Am Chem Soc *127*, 846-847.

Rothman, D. M., Shults, M. D., and Imperiali, B. (2005). Chemical approaches for investigating phosphorylation in signal transduction networks. Trends Cell Biol *15*, 502-510.

Rothman, D. M., Vazquez, M. E., Vogel, E. M., and Imperiali, B. (2002). General method for the synthesis of caged phosphopeptides: tools for the exploration of signal transduction pathways. Org Lett *4*, 2865-2868.

Rothman, D. M., Vazquez, M. E., Vogel, E. M., and Imperiali, B. (2003). Caged phospho-amino acid building blocks for solid-phase peptide synthesis. J Org Chem *68*, 6795-6798.

Rusnak, F., Zhou, J., and Hathaway, G. (2002). Identification of phosphorylated and glycosylated sites in peptides by chemically targeted proteolysis. Journal of Biomolecular Techniques *13*, 228-237.

Sapkota, G., Knockaert, M., Alarcon, C., Montalvo, E., Brivanlou, A. H., and Massague, J. (2006). Dephosphorylation of the linker regions of Smad1 and Smad2/3 by small C-terminal domain phosphatases has distinct outcomes for bone morphogenetic protein and transforming growth factor-beta pathways. J Biol Chem *281*, 40412-40419.

Sato, M., Ozawa, T., Inukai, K., Asano, T., and Umezawa, Y. (2002). Fluorescent indicators for imaging protein phosphorylation in single living cells. Nat Biotechnol *20*, 287-294.

Schmierer, B., and Hill, C. S. (2005). Kinetic analysis of Smad nucleocytoplasmic shuttling reveals a mechanism for transforming growth factor beta-dependent nuclear accumulation of Smads. Mol Cell Biol *25*, 9845-9858.

Schultz, C., Schleifenbaum, A., Goedhart, J., and Gadella, T. W., Jr. (2005). Multiparameter imaging for the analysis of intracellular signaling. Chembiochem *6*, 1323-1330.

Schwarzer, D., and Cole, P. A. (2005). Protein semisynthesis and expressed protein ligation: chasing a protein's tail. Curr Opin Chem Biol *9*, 561-569.

Seet, B. T., Dikic, I., Zhou, M. M., and Pawson, T. (2006). Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol *7*, 473-483.

Shah, K., Liu, Y., Deirmengian, C., and Shokat, K. M. (1997). Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. Proc Natl Acad Sci U S A *94*, 3565-3570.

Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. Nat Methods *2*, 905-909.

Sharma, V., Agnes, R. S., and Lawrence, D. S. (2007). Deep quench: an expanded dynamic range for protein kinase sensors. J Am Chem Soc *129*, 2742-2743.

Shembekar, V. R., Chen, Y., Carpenter, B. K., and Hess, G. P. (2005). A protecting group for carboxylic acids that can be photolyzed by visible light. Biochemistry *44*, 7107-7114.

Shen, K., and Cole, P. A. (2003). Conversion of a tyrosine kinase protein substrate to a high affinity ligand by ATP linkage. J Am Chem Soc *125*, 16172-16173.

Shen, K., Hines, A. C., Schwarzer, D., Pickin, K. A., and Cole, P. A. (2005). Protein kinase structure and function analysis with chemical tools. Biochim Biophys Acta *1754*, 65-78.

Shen, K., Keng, Y. F., Wu, L., Guo, X. L., Lawrence, D. S., and Zhang, Z. Y. (2001). Acquisition of a specific and potent PTP1B inhibitor from a novel combinatorial library and screening procedure. J Biol Chem *276*, 47311-47319.

Shi, Y., and Massague, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685-700.

Shigeri, Y., Tatsu, Y., and Yumoto, N. (2001). Synthesis and application of caged peptides and proteins. Pharmacology and Therapeutics *91*, 85-92.

Shogren-Knaak, M. A., Alaimo, P. J., and Shokat, K. M. (2001). Recent advances in chemical approaches to the study of biological systems. Annu Rev Cell Dev Biol *17*, 405-433.

Shreder, K. R., Liu, Y., Nomanhboy, T., Fuller, S. R., Wong, M. S., Gai, W. Z., Wu, J., Leventhal, P. S., Lill, J. R., and Corral, S. (2004). Design and Synthesis of AX7574: A Microcystin-Derived, Fluorescent Probe for Serine/Threonine Phosphatases. Bioconjug Chem *15*, 790-798.

Shults, M. D., Carrico-Moniz, D., and Imperiali, B. (2006). Optimal Sox-based fluorescent chemosensor design for serine/threonine protein kinases. Anal Biochem *352*, 198-207.

Shults, M. D., and Imperiali, B. (2003). Versatile fluorescence probes of protein kinase activity. J Am Chem Soc *125*, 14248-14249.

Shults, M. D., Janes, K. A., Lauffenburger, D. A., and Imperiali, B. (2005). A multiplexed homogeneous fluorescence-based assay for protein kinase activity in cell lysates. Nat Methods *2*, 277-283.

Sicheri, F., Moarefi, I., and Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. Nature *385*, 602-609.

Sieber, P. (1987). An improved method for anchoring of 9-fluorenylmethoxycarbonyl-amino acids to 4-alkoxybenzyl alcohol resins. Tetrahedron Letters *28*, 6147-6150.

Siegel, P. M., and Massague, J. (2003). Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. Nat Rev Cancer *3*, 807-821.

Smyth, M. S., Ford, Jr., H. Burke, Jr., T. (1992). A general method for the preparation of benzylic alpha,alpha difluorophosphonic acids; non-hydrolyzable mimetics of phosphotyrosine. Tetrahedron Letters *33*, 4137-4140.

Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q., and Perler, F. B. (1999). Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. Biotechniques *27*, 110-114, 116, 118-120.

Speers, A. E., Adam, G. C., and Cravatt, B. F. (2003). Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. J Am Chem Soc *125*, 4686-4687.

Sun, J. P., Fedorov, A. A., Lee, S. Y., Guo, X. L., Shen, K., Lawrence, D. S., Almo, S. C., and Zhang, Z. Y. (2003). Crystal structure of PTP1B complexed with a potent and selective bidentate inhibitor. J Biol Chem *278*, 12406-12414.

Sutherland, E. W., Jr., and Wosilait, W. D. (1955). Inactivation and activation of liver phosphorylase. Nature *175*, 169-170.

Tang, X., and Dmochowski, I. J. (2005). Phototriggering of caged fluorescent oligodeoxynucleotides. Org Lett *7*, 279-282.

ten Dijke, P., and Hill, C. S. (2004). New insights into TGF-beta-Smad signalling. Trends Biochem Sci *29*, 265-273.

Ting, A. Y., Kain, K. H., Klemke, R. L., and Tsien, R. Y. (2001). Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. Proc Natl Acad Sci U S A *98*, 15003-15008.

Toney, M. D., and Kirsch, J. F. (1989). Direct Bronsted analysis of the restoration of activity to a mutant enzyme by exogenous amines. Science *243*, 1485-1488.

Tsukazaki, T., Chiang, T. A., Davison, A. F., Attisano, L., and Wrana, J. L. (1998). SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. Cell *95*, 779-791.

Tyagi, S., and Kramer, F. R. (1996). Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol *14*, 303-308.

Ubersax, J. A., Woodbury, E. L., Quang, P. N., Paraz, M., Blethrow, J. D., Shah, K., Shokat, K. M., and Morgan, D. O. (2003). Targets of the cyclin-dependent kinase Cdk1. Nature *425*, 859-864.

Veldhuyzen, W. F., Nguyen, Q., McMaster, G., and Lawrence, D. S. (2003). A light-activated probe of intracellular protein kinase activity. J Am Chem Soc *125*, 13358-13359.

Ventura, J. J., Hubner, A., Zhang, C., Flavell, R. A., Shokat, K. M., and Davis, R. J. (2006). Chemical genetic analysis of the time course of signal transduction by JNK. Mol Cell *21*, 701-710.

Vila-Perello, M., Pratt, M. R., Tulin, F., and Muir, T. W. (2007). Covalent capture of phospho-dependent protein oligomerization by site-specific incorporation of a diazirine photo-cross-linker. J Am Chem Soc *129*, 8068-8069.

Villain, M., Vizzavona, J., and Gaertner, H. (2001). Proc. Second Int. Seventeeth Am. Peptide Symp., 107-108.

Vogel, E. M., and Imperiali, B. (2007). Semisynthesis of unnatural amino acid mutants of paxillin: protein probes for cell migration studies. Protein Sci *16*, 550-556.

Walsh, C. T. (2005). Posttranslational Modification of Proteins: Expanding Nature's Inventory (Portland, OR: Roberts & Company Publishers).

Wang, Q., Cahill, S. M., Blumenstein, M., and Lawrence, D. S. (2006). Self-reporting fluorescent substrates of protein tyrosine kinases. J Am Chem Soc *128*, 1808-1809.

Wang, Q., Dai, Z., Cahill, S. M., Blumenstein, M., and Lawrence, D. S. (2006). Light-regulated sampling of protein tyrosine kinase activity. J Am Chem Soc *128*, 14016-14017.

Wang, Q., and Lawrence, D. S. (2005). Phosphorylation-driven protein-protein interactions: a protein kinase sensing system. J Am Chem Soc *127*, 7684-7685.

Weissleder, R., Tung, C. H., Mahmood, U., and Bogdanov, A., Jr. (1999). In vivo imaging of tumors with protease-activated near-infrared fluorescent probes. Nat Biotechnol *17*, 375-378.

Wieser, R., Wrana, J. L., and Massague, J. (1995). GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. Embo J *14*, 2199-2208.

Wilhelm, S., Carter, C., Lynch, M., Lowinger, T., Dumas, J., Smith, R. A., Schwartz, B., Simantov, R., and Kelley, S. (2006). Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. Nat Rev Drug Discov *5*, 835-844.

Williams, D. M., Wang, D., and Cole, P. A. (2000). Chemical rescue of a mutant protein-tyrosine kinase. J Biol Chem *275*, 38127-38130.

Wosilait, W. D., and Sutherland, E. W. (1956). The relationship of epinephrine and glucagon to liver phosphorylase. II. Enzymatic inactivation of liver phosphorylase. J Biol Chem *218*, 469-481.

Wrana, J. L., Attisano, L., Wieser, R., Ventura, F., and Massague, J. (1994). Mechanism of activation of the TGF-beta receptor. Nature *370*, 341-347.

Wu, G., Chen, Y. G., Ozdamar, B., Gyuricza, C. A., Chong, P. A., Wrana, J. L., Massague, J., and Shi, Y. (2000). Structural basis of Smad2 recognition by the Smad anchor for receptor activation. Science *287*, 92-97.

Wu, J. W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D. J., Kyin, S., Muir, T. W., Fairman, R.*, et al.* (2001). Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. Mol Cell *8*, 1277-1289.

Wu, N., Deiters, A., Cropp, T. A., King, D., and Schultz, P. G. (2004). A genetically encoded photocaged amino acid. J Am Chem Soc *126*, 14306-14307.

Xie, J., and Schultz, P. G. (2005). Adding amino acids to the genetic repertoire. Curr Opin Chem Biol *9*, 548-554.

Xie, J., Supekova, L., and Schultz, P. G. (2007). A Genetically Encoded Metabolically Stable Analogue of Phosphotyrosine in Escherichia coli. ACS Chem Biol.

Xu, L., Chen, Y. G., and Massague, J. (2000). The nuclear import function of Smad2 is masked by SARA and unmasked by TGFbeta-dependent phosphorylation. Nat Cell Biol *2*, 559-562.

Xu, L., Kang, Y., Col, S., and Massague, J. (2002). Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGFbeta signaling complexes in the cytoplasm and nucleus. Mol Cell *10*, 271-282.

Xu, L., and Massague, J. (2004). Nucleocytoplasmic shuttling of signal transducers. Nat Rev Mol Cell Biol *5*, 209-219.

Xu, W., Doshi, A., Lei, M., Eck, M. J., and Harrison, S. C. (1999). Crystal structures of c-Src reveal features of its autoinhibitory mechanism. Mol Cell *3*, 629-638.

Xu, W., Harrison, S. C., and Eck, M. J. (1997). Three-dimensional structure of the tyrosine kinase c-Src. Nature *385*, 595-602.

Yagi, K., Goto, D., Hamamoto, T., Takenoshita, S., Kato, M., and Miyazono, K. (1999). Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. J Biol Chem *274*, 703-709.

Yeh, R. H., Yan, X., Cammer, M., Bresnick, A. R., and Lawrence, D. S. (2002). Real time visualization of protein kinase activity in living cells. J Biol Chem *277*, 11527-11532.

Yeo, D. S., Srinivasan, R., Uttamchandani, M., Chen, G. Y., Zhu, Q., and Yao, S. Q. (2003). Cell-permeable small molecule probes for site-specific labeling of proteins. Chem Commun (Camb), 2870-2871.

Yudushkin, I. A., Schleifenbaum, A., Kinkhabwala, A., Neel, B. G., Schultz, C., and Bastiaens, P. I. (2007). Live-cell imaging of enzyme-substrate interaction reveals spatial regulation of PTP1B. Science *315*, 115-119.

Zetterqvist, O., Ragnarsson, U., Humble, E., Berglund, L., and Engstrom, L. (1976). The minimum substrate of cyclic AMP-stimulated protein kinase, as studied by synthetic peptides representing the phosphorylatable site of pyruvate kinase (type L) of rat liver. Biochem Biophys Res Commun *70*, 696-703.

Zhang, C., Kenski, D. M., Paulson, J. L., Bonshtien, A., Sessa, G., Cross, J. V., Templeton, D. J., and Shokat, K. M. (2005). A second-site suppressor strategy for chemical genetic analysis of diverse protein kinases. Nat Methods *2*, 435-441.

Zhang, W., and Chait, B. T. (2000). ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. Anal Chem *72*, 2482-2489.

Zhang, X., Jin, Q. K., Carr, S. A., and Annan, R. S. (2002). N-Terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. Rapid Commun Mass Spectrom *16*, 2325-2332.

Zheng, W., Schwarzer, D., Lebeau, A., Weller, J. L., Klein, D. C., and Cole, P. A. (2005). Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. J Biol Chem *280*, 10462-10467.

Zheng, W., Zhang, Z., Ganguly, S., Weller, J. L., Klein, D. C., and Cole, P. A. (2003). Cellular stabilization of the melatonin rhythm enzyme induced by nonhydrolyzable phosphonate incorporation. Nat Struct Biol *10*, 1054-1057.

Zhou, H., Watts, J. D., and Aebersold, R. (2001). A systematic approach to the analysis of protein phosphorylation. Nat Biotechnol *19*, 375-378.

Zhu, Q., Huang, X., Chen, G. Y., and Yao, S. Q. (2003). Activity-based fluorescent probes that target phosphatases. Tetrahedron Letters *44*, 2669-2672.

Zhu, Y., Pavlos, C. M., Toscano, J. P., and Dore, T. M. (2006). 8-Bromo-7-hydroxyquinoline as a photoremovable protecting group for physiological use: mechanism and scope. J Am Chem Soc *128*, 4267-4276.

Zou, K., Cheley, S., Givens, R. S., and Bayley, H. (2002). Catalytic subunit of protein kinase A caged at the activating phosphothreonine. J Am Chem Soc *124*, 8220-8229.