ENZYME-CATALYZED EXPRESSED PROTEIN LIGATION

by

Samuel Henager

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

August, 2017

Abstract

Expressed protein ligation involves the chemoselective reaction of recombinant protein thioesters produced via inteins with N-Cys containing synthetic peptides and has proved to be a valuable method for protein semisynthesis. Expressed protein ligation requires a cysteine residue at the ligation junction which can limit its use. Here we employ subtiligase, a re-engineered form of the protease subtilisin, to ligate a range of synthetic peptides, without the requirement of an N-terminal cysteine, to a variety of recombinant protein thioesters in rapid fashion. We have further broadened the scope of subtiligase-mediated protein ligations by employing a second-generation form (E156Q/G166K subtiligase) and a newly developed form (Y217K subtiligase) for ligation junctions with acidic residues.

We have applied subtiligase-mediated expressed protein ligation to the generation of tetraphosphorylated, monophosphorylated, and non-phosphorylated forms of the tumor suppressor lipid phosphatase PTEN. In this way, we have demonstrated that the natural sequence around the ligation junction produced by subtiligase rather than cysteine-mediated ligation is necessary to confer the dramatic impact of tail phosphorylation on driving PTEN's closed conformation and reduced activity. We thus propose that subtiligase-mediated expressed protein ligation is an attractive traceless technology for precision analysis of protein post-translational modifications.

Thesis Advisor: Dr. Philip Cole

Second Reader: Dr. Jungsan Sohn

To my family and friends, without whom none of this would have been possible.

Acknowledgements

I am especially grateful to Dr. Nam Chu, who carried out the experiments with ubiquitin and subtiligase Y216K described in chapter 2 and who has been very helpful throughout my thesis research, and Stephanie Henriquez, who assisted with expressing and purifying semisynthetic PTEN.

I would also like to thank Amy Weeks, from the Wells lab at UCSF who provided the subtiligase M222A described in chapter 3.

I am also grateful to the other members of the Cole lab for all their help and friendship throughout my time at Hopkins.

Contents

Abstract ii

Acknowledgements iv

List of Tables vii

List of Figures viii

Chapter 1: Introduction 1

Phosphorylation and Protein Semisynthesis 1

Serine Proteases and Subtiligase 7

Phosphorylation and Signal Transduction 13

Phosphatases 14

The PTEN/PI3K/AKT Signaling Pathway 15

PTEN and Cancer 17

Structure of PTEN 20

Regulation of PTEN 23

PTEN Phosphorylation 23

Summary 28

Chapter 2: Protein Semisynthesis with Subtiligase 30

Introduction 30

Methods 33

Results 46

Discussion 61

Chapter 3: Creating Wild Type Semisynthetic Phospho-PTEN 63

Introduction....63
Methods....64
Results....72
Discussion....95
Bibliography....90
Curriculum Vitae....108

Tables

Table 1. 35

Primers used for GST cloning and mutagenesis.

Table 2. 36

Primers used for ubiquitin mutagenesis.

Table 3. 37

Primers used for subtiligase mutagenesis.

Table 4. 44

Peptides used for GST and ubiquitin ligations.

Table 5. 53

Ubiquitin-thioester ligations catalyzed by subtiligase.

Table 6. 60

GST-thioester ligations catalyzed by subtiligase.

Table 7. 66

Primers used for PTEN and subtiligase mutagenesis

Table 8. 68

Peptides used for PTEN ligations

Table 9. 94

Catalytic activity and tail phosphorylation sensitivity to alkaline phosphatase for semisynthetic PTEN constructs.

Figures

Figure 1. 5

Mechanism of Expressed Protein Ligation

Figure 2. 8

Serine protease mechanism

Figure 3. 11

Subtiligase ligation mechanism

Figure 4. 12

Crystal structure of subtiligase with a protein inhibitor

Figure 5. 16

The PTEN/PI3K/AKT pathway

Figure 6. 21

Partial crystal structure of PTEN.

Figure 7. 25

PTEN phosphorylation sites.

Figure 8. 32

Schematic of Native Chemical Ligation versus subtiligase-catalyzed ligation.

Figure 9 41

GST-thioester constructs.

Figure 10 45

MALDI mass spectra of peptides used for GST and ubiquitin ligations.

Figure 11.....47

General scheme for ubiquitin ligations.

Figure 12.....48

Time-course analysis of a subtiligase-catalyzed ubiquitin ligation.

Figure 13.....50

Mass spectrometry analysis of the ligated ubiquitin product.

Figure 14.....51

Subtiligase-mediated hydrolysis of the ubiquitin-thioester.

Figure 15.....52

Results of subtiligase-catalyzed ubiquitin ligations

Figure 16.....55

Analysis of subtiligase-catalyzed and native chemical ligations.

Figure 17.....56

Introducing the Y217K mutation in the active site of subtiligase.

Figure 18.....58

General scheme for GST ligations.

Figure 19.....59

Western blot analysis of subtiligase-catalyzed GST ligations

Figure 20.....65

General scheme for PTEN ligations.

Figure 21.....69

MALDI mass spectra of peptides used for PTEN ligations

Figure 22.....74

Y379-4p-PTEN ligation and purification.

Figure 23.....75

Y379-n-PTEN ligation and purification.

Figure 24.....76

Y379-1p380-PTEN ligation and purification.

Figure 25.....77

Y379-1p382-PTEN ligation and purification.

Figure 26.....78

Y379-1p383-PTEN ligation and purification.

Figure 27.....79

Y379-1p385-PTEN ligation and purification.

Figure 28.....81

Subtiligase M222A ligations with non-phosphorylated PTEN tail peptides.

Figure 29.....82

Subtiligase M222A ligations with phosphorylated PTEN tail peptides.

Figure 30.....85

Western blot analyses of phospho-PTEN.

Figure 31.....88

Enzymatic activity of PTEN.

Figure 32.....89

Enzymatic activity of semisynthetic PTEN.

Figure 33.....92

Sensitivity of 4p-PTEN to treatment with alkaline phosphatase

Figure 34.....93

Sensitivity of 1p-PTEN is treatment with alkaline phosphatase

Chapter 1: Introduction

Phosphorylation and Protein Semisynthesis

Characterizing the effects of phosphorylation on a protein can be challenging. Pure, homogeneously modified protein is necessary for most *in vitro* biochemical assays, but because of the dynamic nature of phosphorylation *in vivo*, protein purified from cell lysate is typically not homogeneously modified.[1-3] Purifying recombinant protein and allowing a kinase to phosphorylate it *in vitro* is also not guaranteed to result in a homogenous population of phospho-protein.[1-3] Thus the effect of many phosphorylation events is poorly understood.[2]

Protein semisynthesis is one method to generate pure, homogeneously modified protein. In general terms, a semisynthetic protein is any purified recombinant protein that is chemically modified.[4, 5] This allows for the precise, stoichiometric addition of post-translational modifications at virtually any residue within a protein. Current methods for introducing phospho-residues or their mimics include: conventional site-directed mutagenesis with Asp/Glu, total synthesis, unnatural amino acid mutagenesis, cysteine modification, Staudinger ligation, and expressed protein ligation.[2, 4, 5]

The dynamic nature of phosphorylation *in vivo* complicates functional analysis. To remedy this, long-lived, non-enzymatically labile, phosphonate mimics can be used. Analogs for phospho-serine, -threonine, and –tyrosine where the oxygen of the phosphoester linkage is replaced with a methylene (CH_2) group have

proven useful.[6-8] Difluoromethylene (CF_2) phosphonates, which more closely mimic the pKa and charge state of physiological phosphorylations have also been used.[7]

Solid phase peptide synthesis allows for peptide chains of up to about 50-60 residues to be created with acceptable yields and purity. Beyond that, peptide chains must be synthesized separately and then ligated together, usually via the chemoselective native chemical ligation reaction between the C-terminal thioester of one peptide and the N-terminal cysteine of a second.[9] In native chemical ligation, the thiol of the cysteine acts as a nucleophile to attack the C-terminal thioester in a transthioesterification reaction. The resulting thioester between the two peptides rapidly rearranges to form a native amide bond. The enzyme subtiligase has also been used for total synthesis with peptide esters in this context.[10] Total synthesis allows complete control over what residues and modifications are included in the protein, but the reaction conditions and piece-wise assembly process can reduce yields, and make refolding challenging for larger proteins.

Unnatural amino acid mutagenesis is another method for introducing post-translational modifications into recombinant expressed proteins. This process involves expressing orthogonal aminoacyl-tRNA synthetase/tRNA pairs in the host system and supplying the desired unnatural amino acid in the expression media. The novel synthetase will create a tRNA corresponding to a stop codon (usually amber, UAG) carrying the desired amino acid. Stop codons in the gene of interest will then code for the unnatural residue.[11] This method has been used to incorporate posttranslational modifications such as phosphoserine[12] and acetylated lysine[13], as well as completely unnatural moieties such as cross-linking groups.[14, 15]

The presence of endogenous phosphatases can make using this technique to study protein phosphorylation difficult, however. Recently, an advance to introduce phosphono-Tyr using nonsense suppression appears promising.[16]

Cysteine alkylation is another useful method for introducing a wide variety of modifications into a protein. This method chemoselectively converts cysteine to dehydroalanine.[17] Dehydroalanine can then be used to affix many different mimics of post-translational modifications, including phosphoserine mimics.[18] This method is limited by the reaction conditions, and cannot be used if there are additional cysteines in the protein. Additionally, the use of mimics instead of the native modifications can introduce its own complications.

The Staudinger reaction can be used in two different ways to create a semisynthetic protein. One method is to react the appropriate azido amino acid with a phosphite reagent to yield an N-linked phosphoramide mimic.[19] An azidohomoalanine (for phosphoserine or phosphothreonine) or 4-azidophenylalanine must be first be incorporated into the protein at the desired site for this method. The second method is a more general technique for semisynthesis, in which a peptide with a C-terminal thioester is ligated to another with an N-terminal azide via a reaction with a phosphinothiol.[20]

Expressed protein ligation uses the native chemical ligation reaction described above to link an intein-derived protein thioester to the N-terminal cysteine of a synthetic peptide.[21] The resulting protein thus has a recombinant N-terminus and synthetic C-terminus. Expressed protein ligation can also be used with a synthetic peptide thioester and the proteolytically revealed N-terminal cysteine of

a recombinant protein,[22] which allows the incorporation of N-terminal post-translational modifications. Expressed protein ligation has been used with a variety of post-translational modifications, including acetylation[23] and phosphorylation,[24, 25] and with cross-linking reagents.[25] It is particularly useful when a cluster of PTMs are present in a confined protein segment, since unnatural amino acid mutagenesis is less efficient with multiple substitutions are required. Inteins have also been used to create cyclized proteins,[26] and to add fluorophores to protein C-termini.[27] Split inteins, in which the N- and C-termini of the intein are expressed separately as fusions with a protein of interest can also be used to access proteins that might be cytotoxic if expressed otherwise.[28] Figure 1 shows the general scheme for expressed protein ligation.

To create the protein-thioester necessary for expressed protein ligation, the protein of interest is expressed as a fusion with a modified intein. Inteins are protein domains that splice themselves out from between two flanking peptides, called "exteins", and then ligate those two peptides together. The process is analogous to self-splicing RNA introns, hence the name "intein."[29] The N-terminal residue of an intein is either cysteine or serine, the C-terminal residue is a conserved asparagine, and the first residue of the C-terminal extein is a cysteine, serine, or threonine. In the canonical intein reaction, the N-terminal residue of the intein attacks the amide bond between the intein and the N-terminal extein forming an ester/thioester linkage. This is followed by a transacylation reaction in which the N-terminal residue of the C-terminal extein attacks the N-terminal ester/thioester.

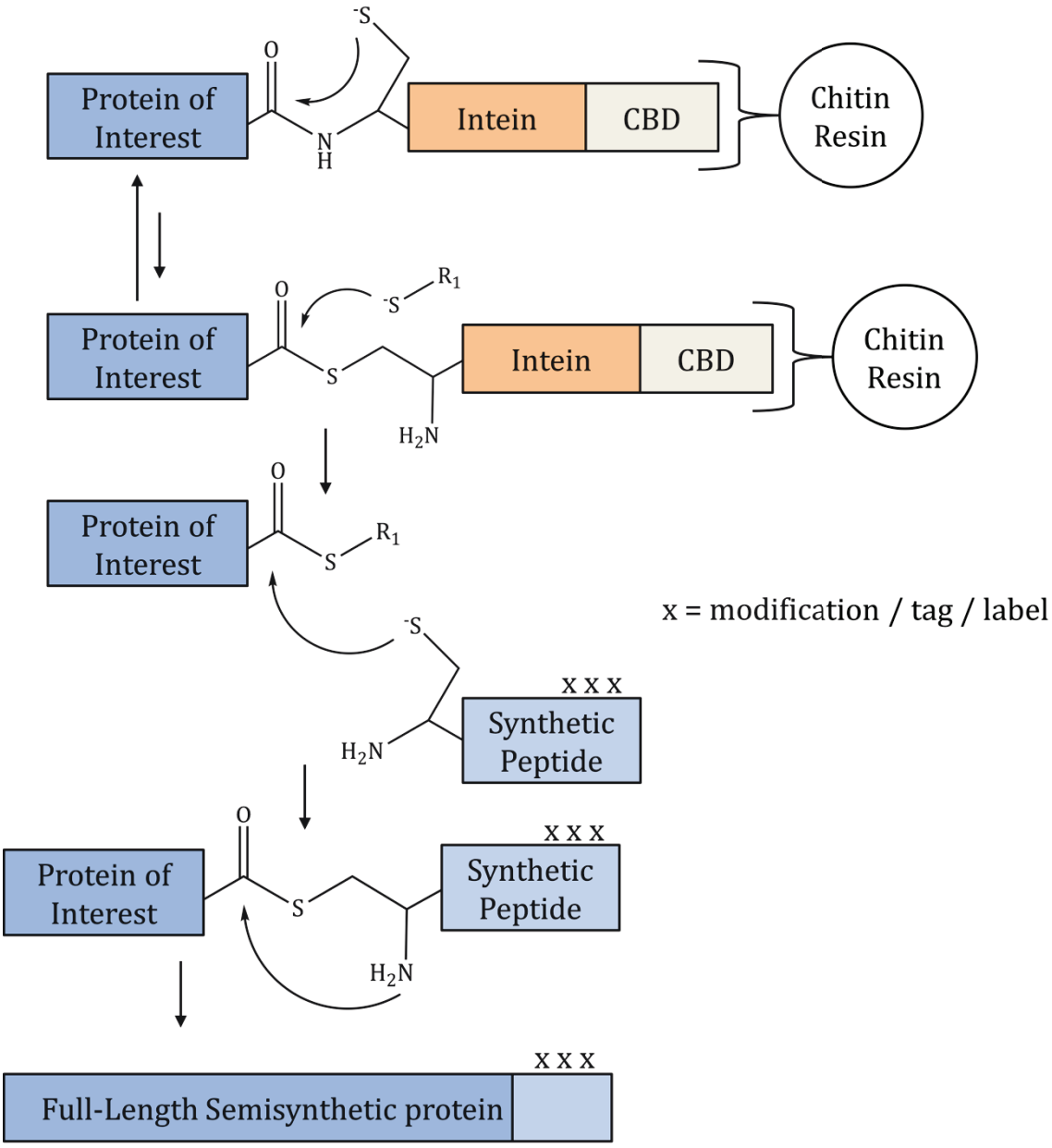


Figure 1. Mechanism of Expressed Protein Ligation. The protein of interest is expressed as an intein-CBD-fusion protein that is purified on chitin resin. The protein of interest is cleaved from the intein using a small molecule thiol, such as mercaptoethanesulfonate (MESNA). A peptide containing a cysteine at its N-terminus and the post-translational modifications to be introduced to the protein of interest is then ligated to the C-terminus of the protein of interest.

To finish the splicing, the N-terminal asparagine residue breaks the amide bond between the intein and the C-terminal extein by cyclizing to form a succinimide, which releases the intein. The remaining ester/thioester linkage between the N- and C-terminal exteins rapidly rearranges into an amide bond.[29]

For the purposes of expressed protein ligation, an intein with an N-terminal cysteine has had the C-terminal asparagine residue of the intein mutated to alanine. The intein is also fused with a C-terminal chitin binding domain that allows for easy purification on chitin resin. Unable to form the succinimide that would finish the splicing process, the intein stalls after forming the initial thioester.[30] This intermediate can be intercepted by the addition of a thiol, such as mercaptoethanesulfonate (MESNA). A synthetic peptide containing an N-terminal cysteine can, via the native chemical ligation reaction, displace the MESNA and form a native amide bond.[9, 21]

As virtually any synthetic peptide can be used, this is a powerful technique for creating semisynthetic proteins to study post-translational modifications, and has been used to great effect with proteins such as PTEN,[24, 25] SAHH,[23] and CK2.[31] However, there are several drawbacks to this technique. One is that the synthetic peptide used must contain an N-terminal cysteine. If no cysteine is readily available in the protein of interest, then one must be introduced, possibly altering the structure and/or function of the protein. In addition, for some protein-peptide pairs, the expressed protein ligation reaction can be very slow, requiring several days to go to completion. To remedy some of these shortcomings, this thesis

describes the creation of semisynthetic proteins from intein-derived protein thioesters and synthetic peptides using the engineered protein ligase, subtiligase.

Serine Proteases and Subtiligase

Proteases, also known as peptidases, are enzymes that catalyze the cleavage of a backbone amide bond in a protein substrate.[32] There are six recognized types of proteases: serine, threonine, cysteine, aspartic, glutamic, and metallo-.[32-34] Serine, threonine, and cysteine proteases use a nucleophilic alcohol (serine and threonine) or thiol (cysteine) to cleave amide bonds, while aspartic, glutamic, and metallo-proteases use a nucleophilic water molecule to do the same.[32] Proteins known as asparagine peptide lyases also have proteolytic activity, but are exclusively autoproteolytic.[35] This family of proteins includes inteins.

The conserved active site of serine proteases consists of a catalytic triad of serine, histidine, and aspartate.[36] During the proteolytic reaction (Figure 2), the aspartate and histidine activate the hydroxyl of the serine which nucleophilically attacks a carbonyl carbon in the backbone of the protein substrate. The unstable tetrahedral intermediate resolves by breaking the carbon-nitrogen bond in the substrate backbone, forming an ester intermediate between the protease and the N-terminal fragment of the substrate protein, and releasing the C-terminal fragment. A nucleophilic water molecule then attacks this intermediate, releasing the N terminal fragment and restoring the enzyme.

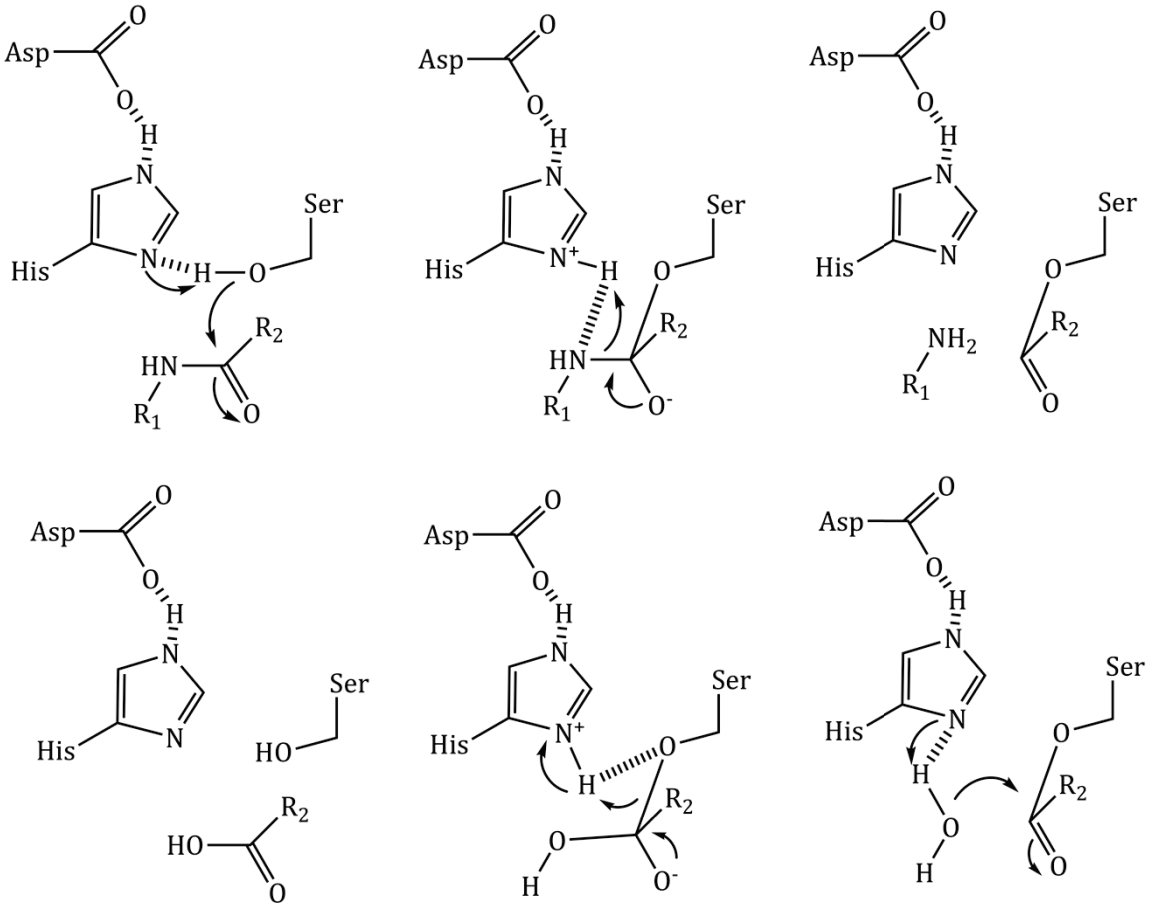


Figure 2. Serine protease mechanism. After binding to a substrate, the activated serine alcohol of the protease nucleophilically attacks a carbonyl carbon in the backbone of the protein substrate. The tetrahedral intermediate resolves by breaking the carbon-nitrogen bond in the substrate backbone, forming an ester intermediate between the protease and the N-terminal fragment of the substrate protein, and releasing the C-terminal fragment. A nucleophilic water molecule then attacks this intermediate, releasing the N terminal fragment and restoring the enzyme.

Serine proteases have diverse substrates and varying specificities. Some, such as trypsin, which cleaves after any lysine or arginine residue[37] , are highly promiscuous, while other, such as TEV, will only bind and cleave a specific sequence.[38] Proteases are crucial to many processes, including blood clotting (thrombin),[39] digestion (pepsin),[40] apoptosis (caspases),[41] and protein degradation (proteasome).[42]

Subtilisin is a bacterial serine protease that is widely expressed. First isolated from *Bacillus subtilis*, it can also be found in many other *Bacillus* species.[43] A relatively promiscuous enzyme, subtilisin prefers to bind and cleave around hydrophobic residues.[43-45] However, mutations to residues involved in substrate binding have shown that the specificity of subtilisin can be changed while not appreciably impacting its catalytic efficiency.[45-47]

In an early example of semisynthesis, the catalytic serine (Ser221) of subtilisin was chemically converted to cysteine to see if a serine protease could be converted to a cysteine protease.[48] This "thiol-subtilisin" still had some proteolytic activity, though its activity was lower than that of wild type subtilisin[49] Later it was shown that while thiolsubtilisin likely proceeded via the same proteolytic mechanism[48, 49] as subtilisin, the thioester intermediate that formed had a >1000-fold increase in aminolysis (that is, attack by a nucleophilic amine) relative to hydrolysis over the native enzyme.[50] This is likely due to the greater lability of thioesters toward amines than water.[51] Selenosubtilisin, where the catalytic serine was replaced by selenocysteine, showed an even greater preference for aminolysis over hydrolysis.[52] This preference for aminolysis was exploited to turn

thiolsubtilisin into a peptide ligase and it was shown that it could effectively ligate a peptide with a C-terminal alkyl-ester to another peptide.[50] In this reaction (Figure 3), the cysteine of thiolsubtilisin attacks the ester carbonyl forming a thioester intermediate, which is in turn attacked by the nitrogen of the N-terminus of the second peptide, forming an amide bond with the first peptide and releasing thiolsubtilisin. However, this modified enzyme still had much lower activity than the native enzyme. The decreased activities of thiolsubtilisin and selenosubtilisin were thought to be caused by steric crowding in the active site from the larger sulfur atom and longer carbon-sulfur bond.[45] To relieve this crowding, a second mutation, P225A was introduced, which shifted C221 about 0.3 Å in the active site.[45] This new subtilisin, named "subtiligase" still had ~500-fold increase in aminolysis relative to hydrolysis over the wild type enzyme, and very little hydrolytic activity.[45] An additional five mutations (M50F, N76D, N109S, K213R, N218S) were introduced to subtiligase and were shown to increase its thermal stability and ability to tolerate detergents and chaotropes.[53]

Subtilisin interacts with 7 residues of its substrate (Figure 4).[43-45] These are numbered P_1 – P_4 and P_1' – P_3', with P_1 and P_1' being N-terminal to and C-terminal to the cleaved bond, respectively. The most important residues for substrate recognition are P_1 and P_4. Both are strongly preferred to be hydrophobic.[45, 46, 53, 54] Subtiligase has similar substrate preferences,[46, 53, 54] and was expected then to interact primarily with the four C-terminal residues (P_1 – P4) of the peptide ester substrate and the three N-terminal residues (P_1' – P_3') of the amine nucleophile substrate.[53]

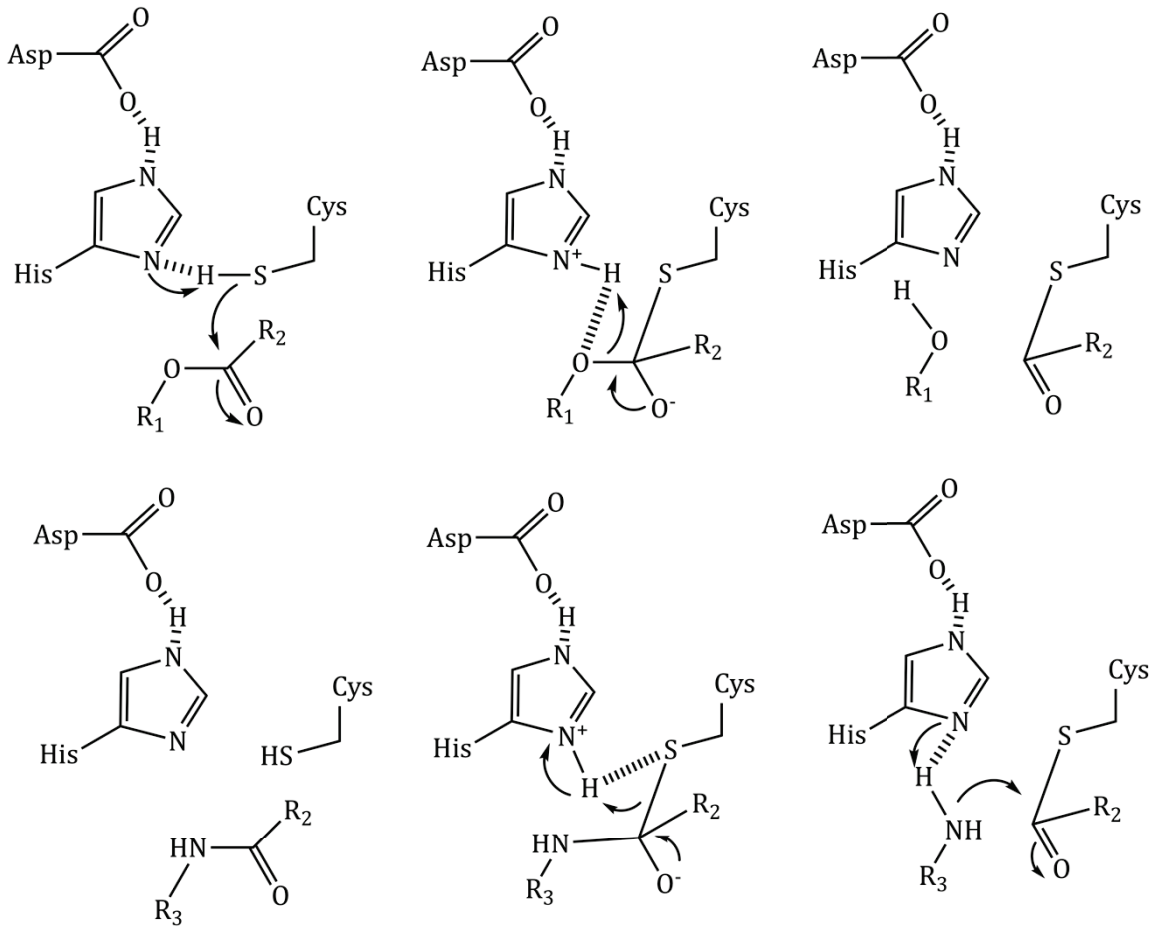


Figure 3. Subtiligase ligation mechanism. After binding to the N-terminal substrate fragment, the activated cysteine thiol of subtiligase nucleophilically attacks the substrate ester. The tetrahedral intermediate resolves by breaking the carbon-oxygen bond in the substrate, forming a thioester intermediate between subtiligase and the N-terminal fragment of the substrate, and releasing an alcohol. The nucleophilic amine of the C-terminal substrate fragment then attacks this intermediate, releasing the fully ligated protein and restoring the enzyme.

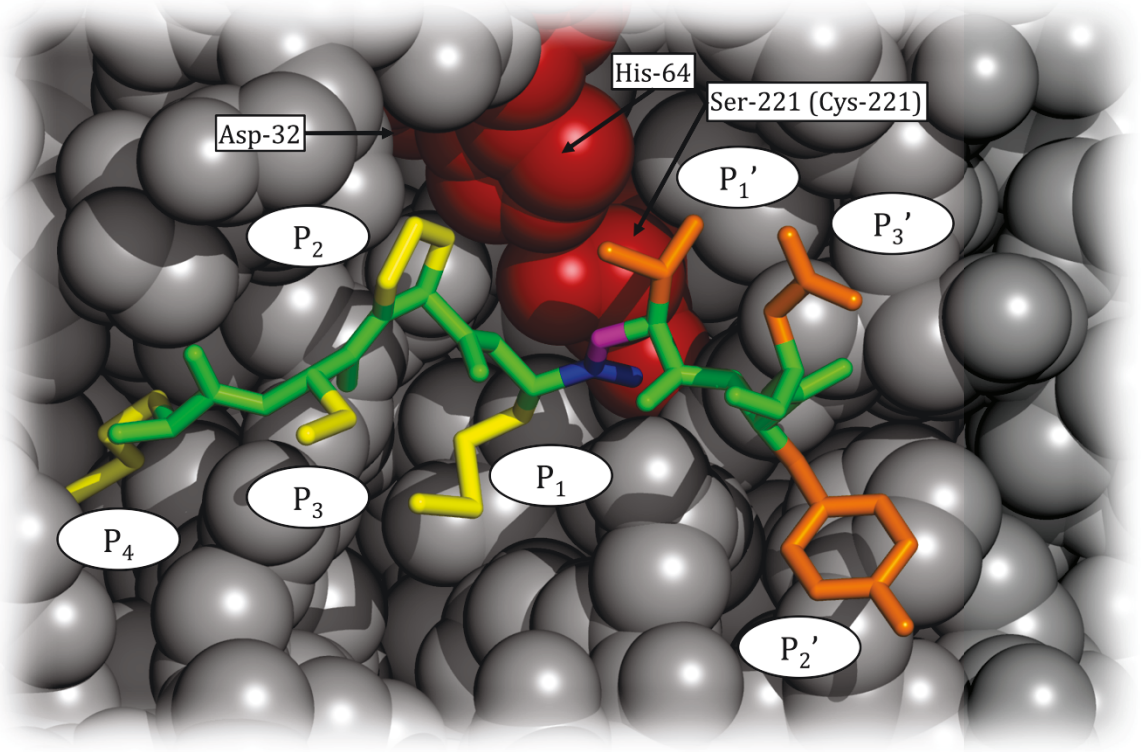


Figure 4. Crystal structure of subtilisin (grey) with a protein inhibitor (green). Only the seven residues of the inhibitor (P_1 – P_4, yellow; and P_1' – P_3', orange) that interact with the active site of subtilisin are shown. The catalytic triad, (Asp-32, His-64, Ser-221), is shown in red. In subtiligase Ser-221 is mutated to Cys. The substrate bond cleaved by subtilisin is shown in pink and blue. The side chains of residues P_1 and P_4 fit into relatively deep pockets in the subtilisin active site, while those of P_2 and P_3 are oriented away from the enzyme. The side chains of residues P_1' and P_2' and the backbone of P_3' also interact with subtilisin. (PDB ID: 2SIC[55])

Like subtilisin, subtiligase preferred hydrophobic residues at P_1 and P_4.[45] At P_1' and P2' subtiligase preferred hydrophobic or polar residues, but cysteine, residues with beta-branching (i.e. threonine or valine) or acidic residues were less preferred.[53] Subtiligase (and subtilisin) is very tolerant of mutations to its active site, allowing its substrate specificity to be modulated to allow for a wide variety of residues around the ligation site.[45, 46] In addition to creating semisynthetic proteins,[10] subtiligase has also been used to tag protein N-termini for proteomic analyses.[56, 57]

Phosphorylation and Signal Transduction

Signal transduction is the process by which cells take in, process, and react to information from their environment. The many complex signally networks within a cell allow for the simultaneous processing of many diverse inputs and the necessary responses to those stimuli. Among these responses are: growth, senescence, death, migration, gene expression, or the release of signaling molecules. Signal processing can be extremely complex, and involve the interactions and activities of hundreds of proteins, or very simple, involving only a handful of interactions.[37]

Within the cell, signals may be related directly via protein-protein interaction, second messengers (e.g. ROS, Ca^{2+}, PIP3, cAMP, etc.), or protein post-translational modifications (e.g. phosphorylation, acylation, ubiquitylation, etc.).[37] One of the most common modifications is phosphorylation, which in eukaryotes involves a phosphate group (PO_3^{2-}) being attached to the hydroxyl group of a serine, threonine, or tyrosine amino acid. *In vivo* phosphorylation states are dynamic and

reversible.[3, 58] Phosphorylation modifications affect proteins by altering conformation, enzymatic activity, localization within the cell, or affinity for binding partners.[37, 58] Small molecules in the cell, such as lipids (e.g. inositol) and sugars (e.g. glucose)[37] can also be phosphorylated. Enzymes known as kinases are responsible for attaching phosphate groups while phosphatases remove them.[58] The human genome contains over 500 kinase genes and nearly 200 phosphatase genes.[59, 60] Many kinases can phosphorylate multiple substrates and different kinases often have overlapping substrates. The same is true for phosphatases, giving cells a fine degree of control over the amount and localization of phospho-proteins.[3, 58]

Phosphatases

Phosphatases can be grouped according to protein fold, and catalytic mechanism. Of the 189 known human phosphatases, over half belong to the CC1 fold, in which a catalytic cysteine is responsible for dephosphorylation.[59] This includes the protein tyrosine phosphatase family, dual-specificity phosphatase family, and the PTEN lipid phosphatase family. These phosphatases contain a conserved CX_5R motif located in the P-loop of the active site. The arginine and N-H backbone groups of the other P-loop residues bind and stabilize the substrate phosphate. These enzymes proceed through a phospho-enzyme intermediate. The nucleophilic cysteine attacks the phosphorus of the substrate phosphate, displacing O-R', which is protonated by an aspartate residue. The aspartate then deprotonates a water molecule which hydrolyzes the phospho-enzyme intermediate and regenerates the enzyme.[61]

The PTEN/PI3K/AKT Signaling Pathway

The PTEN/PI3K/AKT signaling pathway (Figure 5) is a central regulatory pathway that impacts cell growth and metabolism.[62] Many of the proteins involved in this signaling pathway are mutated in cancer, leading to constitutive activation and uncontrolled growth.[62-65] Initiation of the signaling cascade begins when a growth factor, such as EGF, binds to a receptor tyrosine kinase, in the case of EGF, EGFR. This causes the receptor to dimerize and autophosphorylate tyrosine residues on the intracellular domain of the receptor. These phosphorylated tyrosines recruit SH2-domain-containing proteins, such as phosphoinositide-3 kinase (PI3K), which has an SH2 domain in its p85 regulatory subunit.[66] This binding allosterically activates the P110 catalytic subunit of PI3K,[66] which then phosphorylates phosphatidylinositol-4,5-bisphosphate (PIP_2), converting it to phosphatidylinositol-3,4,5-trisphosphate (PIP_3).[67] PIP_2 and PIP_3 are membrane associated lipid second messengers that recruit proteins to the plasma membrane via association with phosphatidylinositide phosphate sensing domains, such as PH, ENTH, FYVE, and PHOX domains.[68] The kinase AKT is recruited to the plasma membrane via its PH domain by PIP_3, and is activated by other kinases such as PDK1 and the mTORC2 complex.[62] Once activated, AKT then phosphorylates a myriad of other downstream substrates, including GSK3,[69] FOXO,[70] and TSC2 which in turn allows activation of mTORC1.[71] Ultimately, AKT activation leads to increased cell growth, proliferation, invasiveness, and resistance to apoptosis.[62, 69-73]

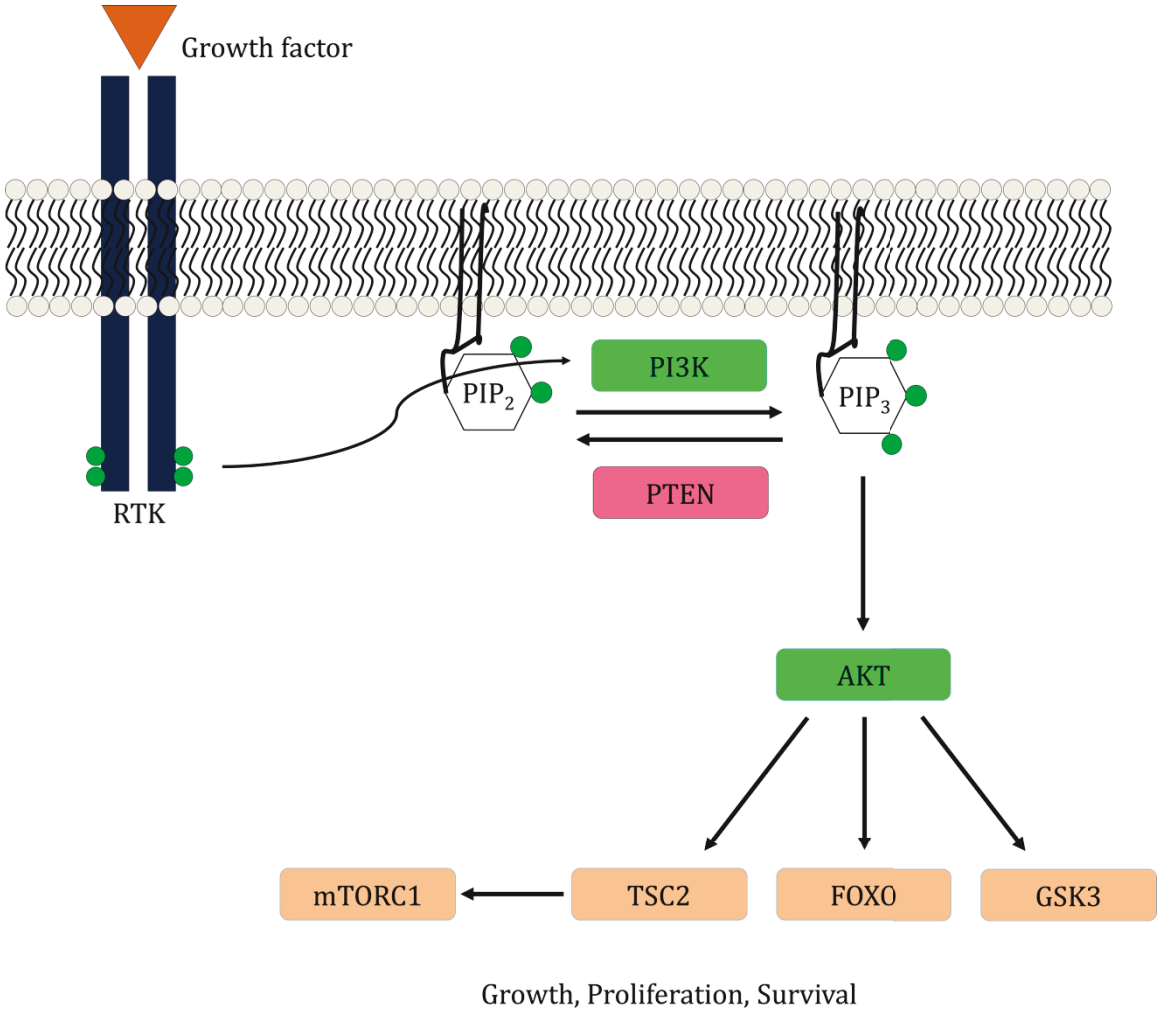


Figure 5. The PTEN/PI3K/AKT pathway. In response to growth factor binding, receptor tyrosine kinases (RTKs) dimerize and autophosphorylate on intracellular tyrosine residues. This leads to the activation of PI3K via its SH2 domain-containing p85 regulatory subunit. PI3K phosphorylates PIP_2, generating PIP_3 which recruits AKT to the membrane, where AKT becomes activated. AKT then phosphorylates multiple downstream substrates, which leads to the activation of pro-growth, pro-survival pathways. PTEN negatively regulates this pathway by dephosphorylating PIP_3, thus preventing AKT activation.

Due to the pro-growth and proliferative effects of the PI3K/AKT pathway, it should come as no surprise that many of its component proteins are mutated in a wide variety of cancers.[62, 64, 72, 73] The accumulation of gain-of-function mutations in positive regulators and loss-of-function mutations in negative regulators leads to constitutive activation in which the pathway remains stuck "on" and cannot be suppressed by normal cellular mechanisms. This allows cancer cells to grow, proliferate, and metastasize. The most commonly mutated protein in this pathway is PTEN.[74-76] Several kinases in the pathway are targeted by gain-of-function mutations, including receptor tyrosine kinases, PI3K, and AKT, which leads to constitutive activation.[62, 63] Gene amplification has also been shown to increase the protein levels of many of these oncogenic kinases.[62, 65]

PTEN and Cancer

PTEN (**P**hosphatase and **TEN**sin homolog deleted on chromosome **TEN**) is a lipid phosphatase that dephosphorylates PIP_3 to PIP_2, thus serving as a key negative regulator of AKT activation and its pro-growth effects.[77, 78] PTEN was discovered in 1997 when deletion mapping of chromosome 10 revealed a potential tumor suppressor at the 10q23 locus.[76, 79, 80] The protein product was found to have sequence similarity to protein tyrosine phosphatases and tensin, an actin filament binding protein and was thus given the name PTEN.[79] MMAC1 (Mutated in Multiple Advanced Cancers) was also an early name for PTEN.[76] Inactivating mutations in PTEN were also found in a number of brain, breast, and prostate cancers.[79] Germline mutations in PTEN were associated with cancer-predisposition syndromes such as

Cowden's disease.[81] PTEN knockout mice were non-viable and heterozygous deletion of PTEN in mice caused an increase in tumor development. In contrast to the classic "two-hit" model of cancer development in which mutations must accumulate in two or more oncogenic or tumor-suppressor genes before tumors begin to form, the ability of PTEN to prevent cancer appears to be dose-dependent.[75, 82] In this continuum model, the expression and/or activity of PTEN is proportional to its ability to prevent neoplastic transformation independent of mutations in other regulatory genes.

The presence of the canonical CX_5R motif in PTEN indicated that it could act as a phosphatase, but no physiologically relevant substrate could be identified.[79] While it could dephosphorylate some model substrates, this activity was very low compared to other phosphatases known to dephosphorylate tyrosine residues. It was however, shown to be much more active toward a peptide consisting entirely of glutamates and phospho-tyrosine.[83] This inactivity toward model substrates and seeming preference for acidic substrates led to the discovery that PTEN readily dephosphorylates PIP_3, converting it to PIP_2.[77] Further, it was shown that PTEN could dephosphorylate many inositol phosphates, but that its activity was specific for phosphorylation at the 3' position. PTEN mutations observed in Cowden's disease were shown to ablate PTEN activity toward PIP substrates and introducing active PTEN into cancer cell lines led to decreased levels of PIP_3 and deactivation of downstream effectors.[78] In PTEN-deficient LnCaP prostate cancer cells, reintroduction of PTEN led to increased apoptosis, showing its potency as a tumor suppressor. However, PTEN-deficient glioblastoma cells did not show this

phenotype upon treatment with PTEN, even though downstream signaling was affected, indicating that reconstituting PTEN activity is not a silver bullet in all cases.

Though the lipid phosphatase activity of PTEN is thought to be the main mode for its tumor suppression, it also possesses a tumor suppressive role in the nucleus.[84-88] Some of this activity may be independent of its phosphatase activity, as even a catalytically dead PTEN mutant (C124S) showed similar tumor-suppressive activity to wild type PTEN, though other studies have shown that phosphatase activity is necessary.[89, 90] Though the nature of PTEN's nuclear function remains unclear, it appears that PTEN regulates chromosomal stability through interactions with multiple proteins, including PLK1,[90] EG5,[89] p53,[91] the APC-CDH1 complex,[92] and members of the mitotic checkpoint complex.[93] Breast cancers with nuclear PTEN have better prognoses that those without, and in general, tumors lacking nuclear PTEN are more aggressive than tumors with nuclear PTEN.[75]

In cancer, PTEN is targeted for inactivation by many methods. These include: gene deletion, gene silencing, mRNA-mediated transcript reduction, and loss-of-function mutations.[75, 94-98] PTEN gene sequences from patient tumor samples reveal abundant point mutations at nearly every residue, apart from residues in its C-terminal tail. Arginine-130, which is necessary for catalysis, has the highest mutational frequency. Residues involved in plasma membrane binding and residues at the interface of the phosphatase and C2 domains are also hotspots for mutations. In general, the absence of PTEN activity in cancer and tumor cells is associated with poor prognosis.

Structure of PTEN

PTEN is a 403-residue protein comprised of an N-terminal phosphatase domain, a C2 domain, and a C-terminal 52-residue tail. A crystal structure of the core of PTEN (the phosphatase and C2 domains) was solved in 1999 (Figure 6).[99] This structure is missing three portions that were deleted, presumably because their flexibility was interfering with crystal formation. The three deleted portions were: The N-terminus (residues 1-7), an internal loop in the C2 domain (residues 286-309), and the entire C-terminal tail (residues 352-403). In addition to those deletions, residues 8 – 14, and 282 – 312 are absent from the crystal structure.

The phosphatase domain of PTEN consists of six helices surrounding a central beta sheet.[99] This domain is structurally akin to dual-specificity phosphatases such as VHR. The active site contains the signature P-loop and conserved CX_5R sequence. The active site is relatively deep, to accommodate the large inositol-phosphate group of PIP_3. PTEN's mechanism, as described above is identical to that of other cysteine phosphatases. Cysteine-124 is the nucleophilic cysteine required for catalysis, and arginine-130 serves to stabilize the bound substrate.[83] Aspartate-92 on an adjacent loop serves as a general acid to protonate the leaving phosphate group during catalysis.[83] The active sites of PTEN, and its homologs, are unique in possessing several lysine residues that are suggested to stabilize interactions with the highly negatively-charged PIP_3 substrate.[99] The N-terminus of this domain (and thus of PTEN) has been proposed to interact with PIP_2 and appears to be important for membrane binding.[100, 101]

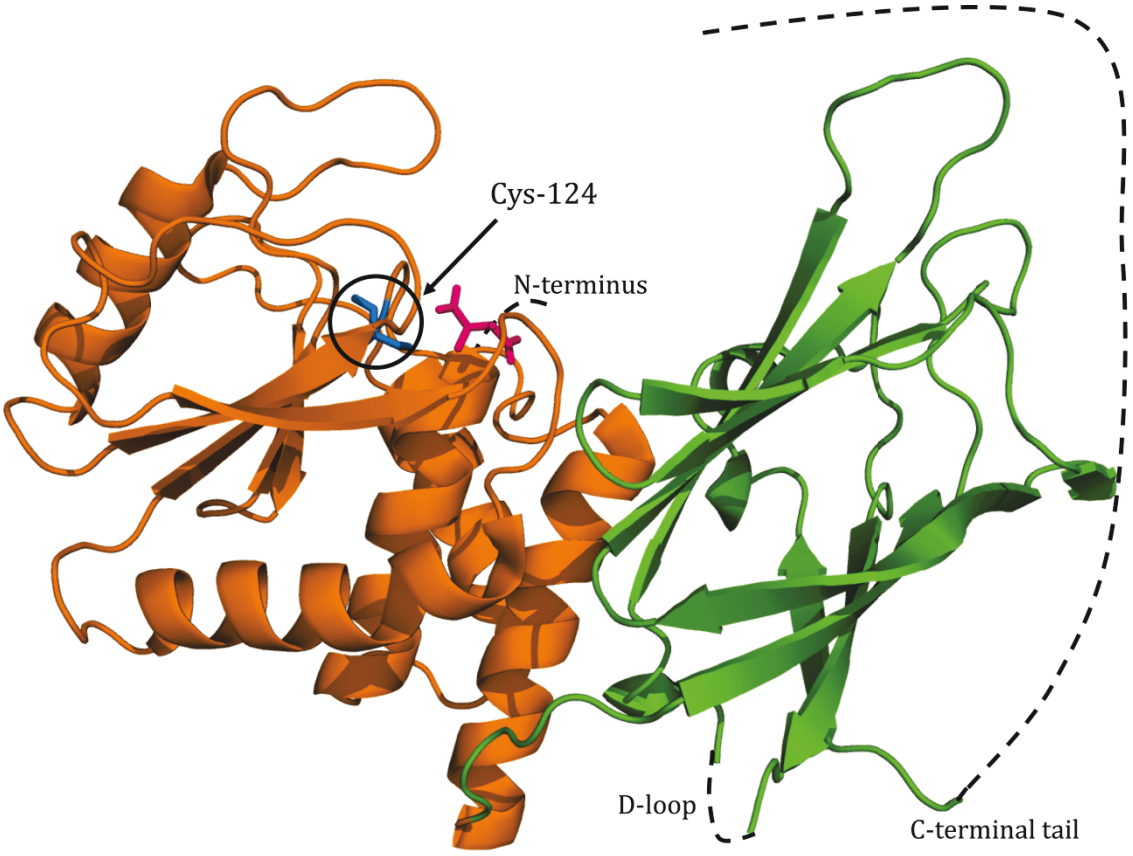


Figure 6. Partial crystal structure of PTEN. The crystal structure of PTEN contains globular phosphatase (orange) and C2 (green) domains. Omitted from the crystal structure were the N-terminus, internal D-loop, and C-terminal tail (dashed lines). The catalytic cysteine-124 is highlighted in blue. PTEN was co-crystalized with tartrate, shown in pink. (PDB ID: 1D5R[99])

The C2 domain of PTEN consists of two antiparallel beta sheets, and is structurally similar to C2 domains found in PI3K, PLCδ1, or PKC. Many C2 domains in other proteins have been shown to mediate Ca^{2+}-dependent membrane recruitment, but in contrast, the PTEN C2 domain does not appear to bind calcium. Two solvent exposed loops, Cα2 and CBRIII, contain multiple lysine and arginine residues and are important for membrane binding and PTEN's tumor suppressor activity.[99] The interface between the phosphatase and C2 domains is relatively large and involves hydrophobic and hydrogen bond interactions.[99] Some of the most common PTEN mutations found in cancer are located at this interface, highlighting its importance for PTEN's function as a tumor suppressor, and indicating that perturbations in one domain may be structurally transmitted to the other.

The C-terminal tail of PTEN consists of the final 52 residues. It contains two PEST (proline, glutamate, serine, and threonine) sequences and a PDZ interaction motif at residues 399-403. The tail is highly acidic, with 15 glutamate or aspartate residues clustered in the final 30 residues. The tail is known to be phosphorylated at several positions by multiple kinases including casein kinase 2 (CK2) and glycogen synthase kinase 3-beta (GSK3β).[102, 103] When PTEN was crystallized, the C-terminal tail was omitted to enable crystallization, so its structure is unknown. However, analysis of PTEN mutants, as well as small-angle X-ray scattering (SAXS) and cross-linking data indicate that when unphosphorylated, the tail is unstructured, but when phosphorylated it interacts with the C2 and phosphatase domains of PTEN.[24, 25]

Regulation of PTEN

PTEN activity is tightly regulated by a variety of mechanisms in the cell. Expression of PTEN is controlled by DNA methylation,[98, 104-106] and by a number of transcription factors, including CBF-1,[94] c-JUN,[107, 108] p53,[95] and Snail1.[96] PTEN mRNA levels are modulated by miRNA.[97] The PTEN protein itself is regulated by posttranslational modifications, including phosphorylation,[24, 102, 103, 109, 110] acetylation,[111, 112] sumoylation,[113] and ubiquitylation.[100, 114-117] Like many other cysteine phosphatases, PTEN can also be inactivated by reactive oxygen species that promote the disulfide bridge formation between the catalytic cysteine and another cysteine residue in the phosphatase domain.[118] PTEN is also regulated by the lipid composition of the plasma membrane, specifically the presence of PIP_2 and phosphatidylserine.[24, 100, 119] PTEN activity can also be modulated by interaction with other proteins, including the p85 subunit of PI3K,[120] MAGI-2,[121] and neutral endopeptidase.[122] As mentioned earlier, PTEN is also regulated by its subcellular localization. PTEN is found at the plasma membrane, in the cytoplasm, or in the nucleus. Localization of PTEN is regulated by phosphorylation and ubiquitylation,[24, 25, 100, 117, 123] with phosphorylation sequestering PTEN from the plasma membrane and preventing ubiquitylation, monoubiquitylation promoting nuclear localization, and polyubiquitylation leading to proteasomal degradation.

PTEN Phosphorylation

PTEN phosphorylation is one of the most studied forms of PTEN regulation. PTEN can be phosphorylated at multiple sites in the C2 domain and the

C-terminal tail (Figure 7). In the C2 domain, serine-229, threonine-232, threonine-319, and threonine-321 are phosphorylated by the RHOA associated protein kinase (ROCK1)[124, 125] and tyrosine-336 is phosphorylated by the protein kinase Rak.[126] In the C-terminal tail, serine-363, threonine-366, serine-370, serine-380, threonine-382, threonine-383, and serine-385 are phosphorylated by CK2, and GSK3β.[102, 103, 109, 123]

ROCK1 was shown to be able to phosphorylate PTEN at Ser229, Thr232, Thr319, and Thr321 *in vitro*, and activation of ROCK1 *in vivo* was shown to regulate a PTEN-dependent decrease of AKT signaling.[124, 125] How these phosphorylation events regulate PTEN activity is unknown. All four sites are on non-membrane binding surfaces of the C2 domain, so it is possible that they alter the interaction of PTEN with another regulatory protein, either via recruitment or inhibition, or they could allosterically modulate the catalytic activity of PTEN.

The protein tyrosine kinase Rak is also capable of phosphorylating PTEN at Tyr336.[126] This phosphorylation may increase the stability of PTEN, as shown by the more rapid degradation of an unphosphorylatable Y336F PTEN mutant compared to wild type. In addition, Rak levels were positively correlated with PTEN expression and its inhibitory effects on AKT signaling. Tyr336 is in the Cα2 loop of PTEN, which is important for membrane binding, and the presence of a charged phosphate group at this residue may interfere with this process.

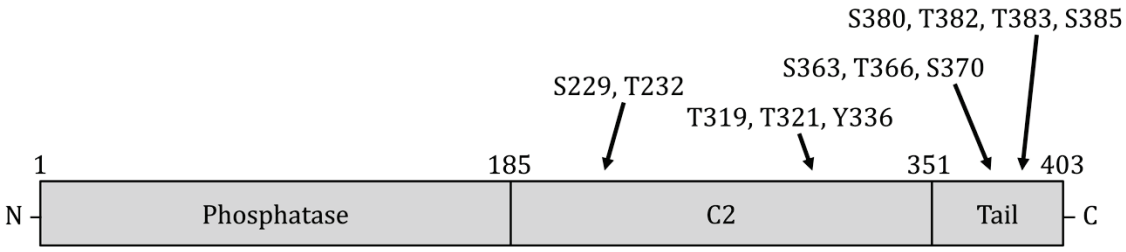


Figure 7. PTEN Phosphorylation sites. PTEN can be phosphorylated at several sites in its C2 domain and C-terminal tail. The C2 domain is phosphorylated at Ser229, Thr232, Thr319, and Thr321 by RHOA associated protein kinase (ROCK) and at Tyr336 by the protein kinase Rak. The C-terminal tail is phosphorylated at Ser363, Thr366, Ser370, Ser380, Thr382, Thr383, and Ser385 by the kinases CK2 and GSK3β.

One E3 ubiquitin ligase for PTEN, NEDD4-1,[114, 116, 127] is thought to be active at the plasma membrane, so phosphorylation at Tyr336 could be preventing PTEN from co-localizing with NEDD4-1, and thus preventing PTEN from becoming polyubiquitylated and degraded.

The C-terminal tail is the most abundant site of PTEN phosphorylation.[103, 109] Multiple studies have identified phosphorylation at Ser363, Thr366, Ser370, Ser380, Thr382, Thr383, and Ser385 by the kinases CK2 and GSK3β.[103, 109, 123, 128] How phosphorylation at Ser363, Thr366, and Ser370 affects PTEN is not well understood. Although an unphosphorylatable T366A PTEN mutant was shown to be more stable *in vivo* than wild type PTEN, indicating that phosphorylation at Thr366 may regulate PTEN stability.[129]

Phosphorylation at Ser380, Thr382, Thr383, and Ser385 is primarily by CK2, though GSK3β may also contribute.[103, 109, 123, 128, 130] Introducing an S385A mutation to PTEN was shown to drastically reduce the level of tail phosphorylation, indicating that phosphorylation at Ser385 primes for phosphorylation at Ser380, Thr382, and Thr383.[131] NMR experiments with PTEN tail peptides showed that CK2 first phosphorylates Ser385, then Ser380, then Thr383, and finally Thr382.[109] This was done *in vitro* and in the absence of the full PTEN protein, so the actual phosphorylation process may be different.

It is thought that the ratio of phosphorylated to unphosphorylated PTEN likely varies based on CK2 activity. In pancreatic beta cells, treatment with leptin was shown to rapidly increase the level PTEN phosphorylation,[132, 133] and in TM cells, treatment with TGF-β was also associated with an increase in PTEN

phosphorylation.[134] No specific conditions have been reported that lead to a decrease in cellular PTEN phosphorylation. In the experiments in pancreatic beta cells, PTEN phosphorylation decreased to basal levels within an hour after treatment, indicating that there is some phosphatase that acts on PTEN, but none has yet been identified. PTEN has shown possible auto-phosphatase activity *in vitro* but this is controversial.[135]

Phosphorylation at the Ser380/Thr382/Thr383/Ser385 cluster has been shown to cause a conformational change in PTEN.[24, 25] Semisynthetic PTEN, in which residues 1 – 378 were expressed in insect cells as a fusion with a modified intein and chitin binding domain (CBD), was ligated via native chemical ligation to a synthetic peptide corresponding to residues 379 – 403 of the C-terminal tail. Phospho-Ser380, -Thr382, -Thr383, and -Ser385 were incorporated into the tail peptide. Tyr379 was mutated to Cys to accommodate the native chemical ligation reaction. Activity assays with soluble di-C6 PIP_3 substrates and PIP_3 incorporated into lipid vesicles showed that tetra-phosphorylated PTEN was less active than non-phosphorylated PTEN and had a lower affinity for lipid membranes.[24] Tetra-phosphorylated PTEN was also resistant to dephosphorylation by alkaline phosphatase, and despite the addition of four highly charged phosphate groups, it eluted earlier in anion exchange chromatography. These data indicated that the phosphorylated groups were interacting with the core of PTEN. A phosphorylated tail peptide was also able to inhibit PTEN activity in trans. Further experiments with mono-, di-, and tri-phosphorylated semisynthetic PTEN showed that each phosphorylation event contributed to PTEN inhibition.[25] Cross-linking and analysis

of PTEN mutants indicate that the phosphorylated tail interacts with the N-terminus, the Cα2 loop, and the CBRIII loop, which are all also implicated in membrane binding.[99, 136, 137] Thus, the phosphorylated tail may act as a mimic of a phospholipid membrane.

One limitation of these studies with semisynthetic PTEN is the presence of the Y379C mutation necessary to facilitate the native chemical ligation reaction. This mutation is immediately adjacent to the phosphorylated cluster of residues and may interfere with the interaction between the tail and core of PTEN. Tyr379 mutations have been shown to result in increased PTEN levels in the nucleus and plasma membrane, implying a more open conformation.[117] To determine how the Y379C mutation affects the activity of tetra-phosphorylated PTEN, it is necessary to create wild type, semisynthetic, tetra-phosphorylated PTEN.

Summary

The ability to install post-translational modifications, biophysical probes, unnatural amino acids, isotopic labels, and drug-like small molecules in a site-specific manner into proteins of any size offers enormous potential for both fundamental and applied biomedical research. Semisynthesis using expressed protein ligation has been exploited frequently in the construction of proteins containing diverse chemical modifications. While it is a powerful method, the scope of EPL is narrowed by the requirement of a Cys at the ligation junction. Cys is one of the least frequently encoded residues in proteins,[138-140] and this sharply confines the

flexibility in the ligation position or requires the introduction of Cys at unnatural locations, sometimes perturbing protein function.

Much research has been done on how phosphorylation affects the activity and localization of PTEN. Experiments done with semisynthetic phosphorylated PTEN have shown that phosphorylation on the C-terminal tail of PTEN increases protein stability, lowers catalytic activity, and decreases membrane affinity. These studies were all done with PTEN containing a tyrosine to cysteine mutation immediately adjacent to the cluster of phosphorylated residues, which has been shown to cause anomalous membrane association *in vivo*. Thus, it is desirable to create a wild type, semisynthetic phosphorylated PTEN. This thesis will demonstrate how the engineered protein ligase subtiligase can be used to create semisynthetic proteins from intein-generated protein thioesters and synthetic peptides, and will demonstrate the use of subtiligase to create wild type, semisynthetic, phosphorylated PTEN.

Chapter 2: Protein Semisynthesis with Subtiligase

Introduction

As mentioned in the previous chapter, precision-guided control of the chemical structure of proteins for biological study has been an elusive challenge. The ability to site-specifically install post-translational modifications, biophysical probes, unnatural amino acids, isotopic labels, and drug-like small molecules into proteins of any size offers enormous potential for both fundamental and applied biomedical research.[2, 141, 142] Major strides have been made in total synthesis,[143, 144] semisynthesis,[145, 146] and nonsense codon suppression strategies[147, 148] for protein chemical manipulation but each has significant technical limitations. Semisynthesis using expressed protein ligation, a method developed in 1998,[21, 149] has been exploited frequently in the construction of proteins containing diverse chemical modifications.

In standard expressed protein ligation, a recombinant protein fragment is fused to an intein, which is then reacted with a thiol to generate the isolated recombinant protein fragment possessing a C-terminal thioester.[21, 149] An N-Cys containing synthetic peptide is then added to the protein thioester which undergoes transthioesterifcation followed by rearrangement to a conventional amide bond (Figure 1). While powerful, the scope of expressed protein ligation is narrowed by the requirement of a cysteine at the ligation junction. Cysteine is one of the least frequently encoded residues in proteins,[139, 140] and this sharply constrains the

ligation position or requires the introduction of cysteine at unnatural locations. Such cysteine replacements are not always well-tolerated and can confer aberrant behaviors relative to the wild type protein sequence. Methods to desulfurize cysteine[150] or circumvent the Cys requirement[151] for a ligation are under development but have not yet proved robust. In addition, expressed protein ligation can sometimes require reaction times lasting 3 days at room temperature which can lead to protein loss or denaturation.

Over two decades ago, subtiligase was introduced as an engineered mutant (S221C, P225A) of the bacterial protease subtilisin that is devoid of amidase activity but can still catalyze the ligation of peptide fragments to each other if one contains a C-terminal ester.[45, 53] Subtiligase has been used in a variety of protein chemistry settings,[10, 54, 152] but not yet applied to ligations with recombinant protein thioester fragments. In this chapter, we investigate the use of subtiligase as an alternative to conventional non-enzymatic expressed protein ligation (Figure 8).

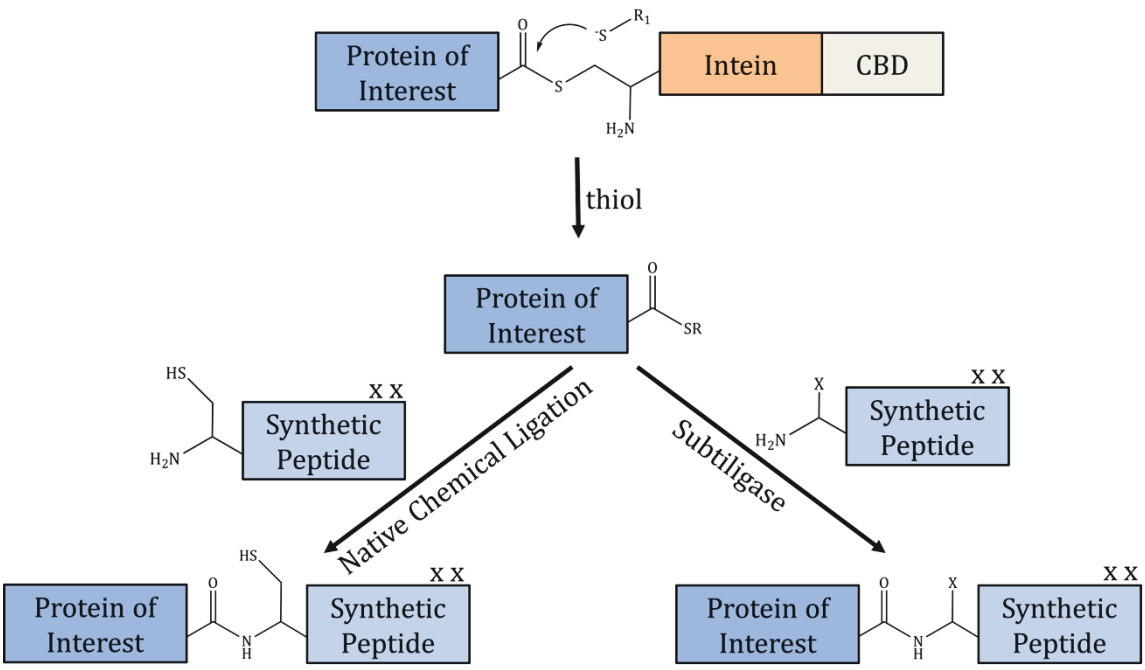


Figure 8. Schematic of Native Chemical Ligation versus subtiligase-catalyzed ligation. For expressed protein ligation, after the protein of interest is cleaved from the intein with a thiol, it is incubated with a synthetic peptide containing an N-terminal cysteine. Via the native chemical ligation reaction, the cysteine reacts with the thioester and forms a native amide bond. For subtiligase, the protein thioester is isolated, then incubated with subtiligase and a synthetic peptide. This synthetic peptide does not need an N-terminal cysteine. Subtiligase enzymatically ligates the two together, producing the full-length product.

Methods

Site-directed mutagenesis of GST, ubiquitin, and subtiligase

Glutathione S-transferase (GST) was cloned out of the pGEX-6P-1 vector. 1 µl of template was incubated with 200 nM forward and reverse primers and 2 mM dNTP mix (0.5 mM each dNTP) in 50 µl reaction buffer (20 mM Tris–HCl, 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 0.1 mg/ml BSA, pH 8.8) along with 1 µl Pfu Ultra II polymerase (Agilent). PCR for mutagenesis was carried out as follows: 95 °C for 30 s; then 18 cycles of 95 °C for 30 s, 55 °C for 1 min, and 68 °C for 30 s; followed by a final 2 min at 68 °C; then 4 °C for 5 min. Amplification of the GST gene was confirmed by agarose gel electrophoresis. The amplified GST gene was isolated from the agarose gel using a gel extraction kit (Qiagen), and the ends of the gene were trimmed using the NdeI and SmaI restriction enzymes (NEB). The gene was incubated with 1 unit/µl SmaI in NEBuffer 4 (50 mM potassium acetate, 20 mM Tris–acetate, 10 mM magnesium acetate, 1mM DTT, pH 7.9) for 2 h at 25 °C, then incubated with 1 unit/µl NdeI for 2 h at 37 °C. The pTYB2 vector was also digested in this way. To insert the gene into the vector, the digested gene and vector were incubated with 40 units/µl T4 ligase in T4 buffer (50 mM Tris–HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5) for 18 h at 16 °C. The completed ligation was transformed into chemically competent *E. coli* DH5α. The sequence was confirmed by Sanger sequencing. An overlooked SmaI site near the C terminus of GST resulted in the truncation of the final eight residues of GST and a single nucleotide frameshift of the intein–CBD relative to the GST. Quikchange mutagenesis was used to add Phe-Ala-Ala-Tyr to the C terminus of GST and remove the frameshift. 1 µl of template

was incubated with 200 nM forward and reverse primers and 2 mM dNTP mix in 50 µl reaction buffer along with 1 µl Pfu Ultra II polymerase. PCR for mutagenesis was carried out as follows: 95 °C for 30 s; then 18 cycles of 95 °C for 30 s, 55 °C for 1 min, and 68 °C for 7.5 min; followed by a final 5 min at 68 °C; then 4 °C for 5 min. After PCR, the reaction was incubated with 0.4 units/µl DpnI at 37 °C for 1 h to digest methylated DNA. The mutated plasmid was transformed into chemically competent *E. coli* DH5α. Quikchange mutagenesis was also used to further mutate the C termini of GST and ubiquitin, as well as to mutate subtiligase. Tables 1 – 3 show the primers used for GST, ubiquitin, and subtiligase cloning and mutagenesis, respectively.

B. Subtilis transformation with subtiligase

This was carried out analogously to previously described methods.[10, 45, 46, 53] *E. coli–B. subtilis* shuttle plasmid pPW 04 containing the pre-prosubtiligase sequence was purified from E. coli K12 ER1821 (NEB) and transformed into B. subtilis BG2864 (ΔaprE ΔnprE ΔflaA::kan, ATCC). B. subtilis was grown on LB agar + 25 µg/ml kanamycin. 2xYT media (5 ml) was inoculated with a single colony and grown overnight at 37 °C in a spinning incubator. Cells were pelleted and resuspended in 5 ml bacillus medium A (80 mM K_2HPO_4, 45 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 4 mM $C_6H_5O_7Na_3$, 5 mM $MgCl_2$, 50 µg/ml tryptophan, 0.5% glucose, 0.02% amicase). 1 ml of this suspension was added to 50 ml medium A and grown at 37 °C until 90 min after the culture had exited log-phase growth.

Primer	Sequence (5' - 3')
GST Cloning Forward	AAAAAACATATGTCCCCTATA--CTAGGTTATTGGAAAATTAAG
GST Cloning Reverse	TTTTTTCCCGGGGTCACGAT--GCGGCCGCTCGAGTCGACCCG
GST FAAY insertion, F	GGAATTCCCGGGT**TTCGCCGCC**--**TACTGCTTTGCCAAGGGTACC**
GST FAAY insertion, R	GGTACCCTTGGCAAAGCA**GTAG**--**GCGGCGAA**ACCCGGGAATTCC
GST Y216G, F	GGTTTCGCCGCC**GGC**TGCTTTGCCAAGG
GST Y216G, R	CCTTGGCAAAGCA**GCC**GGCGGCGAAACC
GST Y216R, F	GGTTTCGCCGCC**CGC**TGCTTTGCCAAGG
GST Y216R, R	CCTTGGCAAAGCA**GCG**GGCGGCGAAACC
GST Y216T, F	GGTTTCGCCGCC**ACC**TGCTTTGCCAAGG
GST Y216T, R	CCTTGGCAAAGCA**GGT**GGCGGCGAAACC
GST Y216E, F	GGTTTCGCCGCC**GAG**TGCTTTGCCAAGG
GST Y216E, R	CCTTGGCAAAGCA**CTC**GGCGGCGAAACC
GST F213D, F	CCGGAATTCCCGGGT**GAT**GCCGCCTACTGCTTTGC
GST F213D, R	GCAAAGCAGTAGGCGGC**ATC**ACCCGGGAATTCCGG
GST F213S, F	CCGGAATTCCCGGGT**AGC**GCCGCCTACTGCTTTGC
GST F213S, R	GCAAAGCAGTAGGCGGC**GCT**ACCCGGGAATTCCGG
GST F213S-G216, F	CCGGAATTCCCGGGT**AGC**GCCGCCGGCTGCTTTGC
GST F213S-G216, R	GCAAAGCAGCCGGCG**GCG**CTACCCGGGAATTCCGG

Table 1. Primers used for GST cloning and mutagenesis. **Bold** indicates mutated residues.

Primer	Sequence (5' – 3')
Ubiquitin G76Y, F	GGTCCTGCGTCTGAGAGGT**TAT**TGCATCACGGGAGATGCAC
Ubiquitin G76Y, R	GTGCATCTCCCGTGATGCA**ATA**ACCTCTCAGACGCAGGACC
Ubiquitin G76E, F	GTGCTAAGGCTAAGAGGT**GAA**TGCATCACGGGAGATGCAC
Ubiquitin G76E, R	GTGCATCTCCCGTGATGCA**TTC**ACCTCTTAGCCTTAGCAC
Ubiquitin G76A, F	GTGCTAAGGCTAAGAGGT**GCG**TGCATCACGGGAGATGCAC
Ubiquitin G76A, R	GTGCATCTCCCGTGATGCA**CGC**ACCTCTTAGCCTTAGCAC
Ubiquitin G76H, F	GTGCTAAGGCTAAGAGGT**CAT**TGCATCACGGGAGATGCAC
Ubiquitin G76H, R	GTGCATCTCCCGTGATGCA**ATG**ACCTCTTAGCCTTAGCAC
Ubiquitin G76F, F	GTGCTAAGGCTAAGAGGT**TTT**TGCATCACGGGAGATGCAC
Ubiquitin G76F, R	GTGCATCTCCCGTGATGCA**AAA**ACCTCTTAGCCTTAGCAC
Ubiquitin G76W, F	GTGCTAAGGCTAAGAGGT**TGG**TGCATCACGGGAGATGCAC
Ubiquitin G76W, R	GTGCATCTCCCGTGATGCA**CCA**ACCTCTTAGCCTTAGCAC
Ubiquitin G76V, F	GTGCTAAGGCTAAGAGGT**GTT**TGCATCACGGGAGATGCAC
Ubiquitin G76V, R	GTGCATCTCCCGTGATGCA**AAC**ACCTCTTAGCCTTAGCAC
Ubiquitin G76I, F	GTGCTAAGGCTAAGAGGT**ATT**TGCATCACGGGAGATGCAC
Ubiquitin G76I, R	GTGCATCTCCCGTGATGCA**AAT**ACCTCTTAGCCTTAGCAC
Ubiquitin G76L, F	GTGCTAAGGCTAAGAGGT**CTG**TGCATCACGGGAGATGCAC
Ubiquitin G76L, R	GTGCATCTCCCGTGATGCA**CAG**ACCTCTTAGCCTTAGCAC
Ubiquitin G76R, F	GTGCTAAGGCTAAGAGGT**CGT**TGCATCACGGGAGATGCAC
Ubiquitin G76R, R	GTGCATCTCCCGTGATGCA**ACG**ACCTCTTAGCCTTAGCAC
Ubiquitin G76D, F	GTGCTAAGGCTAAGAGGT**GAT**TGCATCACGGGAGATGCAC
Ubiquitin G76D, R	GTGCATCTCCCGTGATGCA**ATC**ACCTCTTAGCCTTAGCAC
Ubiquitin L73F-G76Y, F	CCTTACATCTTGTGCTAAGG**T**- -**TT**AGAGGT**TAT**TGCATCACG
Ubiquitin L73F-G76Y, R	CGTGATGCA**ATA**ACCTCT**AA**- -**A**CCTTAGCACAAGATGTAAGG
Ubiquitin L73A-G76Y, F	CCTTACATCTTGTGCTAAGG**G**- -**CG**AGAGGT**TAT**TGCATCACG
Ubiquitin L73A-G76Y, R	CGTGATGCA**ATA**ACCTCT**CGC**- -CCTTAGCACAAGATGTAAGG
Ubiquitin L73E-G76Y, F	CCTTACATCTTGTGCTAAGG**G**- -**AA**AGAGGT**TAT**TGCATCACG
Ubiquitin L73E-G76Y, R	CGTGATGCA**ATA**ACCTCT**TTC**- -CCTTAGCACAAGATGTAAGG

Table 2. Primers used for ubiquitin mutagenesis. **Bold** indicates mutated residues.

Subtiligase E156Q-G166K, F	CGTTGCGGCAGCCGGTAAC**AAT**GGCACTTCCGGCAG- -CTCAAGCACAGTG**AAA**TACCCTGGTAAATACCCTTC
Subtiligase E156Q-G166K, R	GAAGGGTATTTACCAGGGTA**TTT**CACTGTGCTTGAG- -CTGCCGGAAGTGCC**ATT**GTTACCGGCTGCCGCAACG
Subtiligase Y217K, F	GGAAACCGGTACGGGGCG**AAA**AG- -CGGTACGTGCATGGCATCTGCGC
Subtiligase Y217K, R	GCGCAGATGCCATGCACGTACCG- -CT**TTT**CGCCCCGTACCGGTTTCC

Table 3. Primers used for subtiligase mutagenesis. **Bold** indicates mutated residues.

Then, 0.5 ml of this culture was added to 5 ml medium B (80 mM K_2HPO_4, 45 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 4 mM $C_6H_5O_7Na_3$, 5 mM $MgCl_2$, 2.5 μg/ml tryptophan, 0.5% glucose, 0.005% amicase). Next, 300 μl of this culture was added to tubes containing 2 μg of plasmid pPW 04 and grown at 37 °C. After 2 h, 1 ml 2× YT medium containing 0.5 μg/ml chloramphenicol was added, and the culture was grown for an additional 1 h. Culture was spun down (4,000 x g, 2 min), resuspended in 150 μl 2xYT, and plated on LB agar containing 5 μg/ml chloramphenicol.

Subtiligase expression and purification

Parent and mutant subtiligase were expressed and purified using the following methods adapted from previous procedures.[10, 45, 46, 53] Transformed *B. subtilis* was streaked out on LB agar + 10 μg/ml chloramphenicol. 2xYT media (5 ml) containing 10 μg/ml chloramphenicol was inoculated with one colony and grown overnight. 1 ml of overnight culture was added to 1 l 2× YT media containing 10 μg/ml chloramphenicol and grown at 37 °C for 24 h. Cells were pelleted and discarded, and then the supernatant was slowly treated with 600 g $(NH_4)_2SO_4$ while stirring on ice at 4 °C. The solution was stirred for an additional 1 h, during which a brown protein precipitation was observed. The mixture was centrifuged (10,000 × g, 30 min, 4 °C), and the saved pellet was re-dissolved in 75 ml sodium acetate buffer (25 mM CH_3CO_2Na, 5 mM DTT, pH 4.5). To this solution was added 300 ml ethanol, resulting in a pale brown precipitate, and the mixture was stirred for 30 min at 4 °C. The mixture was then centrifuged (5,000 × g, 15 min, 4 °C), and the pellet was saved and re-dissolved in 50 ml sodium acetate buffer, then dialyzed

thoroughly against the same buffer. After dialysis, the mixture was centrifuged (17,500 × g, 15 min, 4 °C) and the pellet discarded. The supernatant was loaded onto a Mono S cation-exchange column (GE Healthcare Life Sciences). Protein was eluted in a 0–400 mM NaCl gradient in sodium acetate buffer, pH 4.5. Fractions containing subtiligase were dialyzed overnight against 100 mM BICINE, 5mM DTT, pH 8.0, then concentrated to ~100 μM in an Amicon Ultra 10 kDa MWCO filter unit, flash frozen, and stored at −80 °C. Protein was estimated to be greater than 95% by SDS–PAGE, and the stock concentration was determined by comparison to bovine serum albumin standards stained with Coomassie blue.

GST expression and purification.

GST was subcloned into a pTYB2 vector, which contains the VMA intein from *S. cerevisiae*, and subjected to insertional mutagenesis as described above, resulting in the appending of four residues (X-Ala-Ala-X) to the natural GST C terminus followed by the intein. The GST plasmids were expressed in *E. coli* BL21. LB media (5 ml) containing 100 μg/ml carbenicillin was inoculated with a single colony and grown overnight at 37 °C in a spinning incubator. 3.5 ml from overnight cultures was used to inoculate 1 l LB media containing 100 μg/ml carbenicillin, and the cultures were grown in shaker flasks at 37 °C until OD_{600} = 0.7, then 1 ml 1M IPTG was added to induce expression and the cultures were incubated overnight while shaking at 16 °C. Cells were pelleted and then resuspended in 40 ml lysis buffer (250 mM NaCl, 50 mM HEPES, 1 mM EDTA, 10% glycerol, pH = 7.5). Unused cell pellets were flash frozen and stored at -80 °C until needed. *E. coli* cells were

lysed by French press, the lysate was pelleted (17,500 × g, 15 min, 4 °C), and the supernatant was loaded onto 5 ml of chitin resin (NEB). Resin was washed with 150 ml wash buffer (250 mM NaCl, 25 mM HEPES, 0.1% Triton X-100, pH 7.5) then incubated overnight in cleavage buffer (250 mM NaCl, 50 mM HEPES, 300 mM MESNA, pH 7.5). To generate GST–biotin, 10 mM peptide containing an N-terminal cysteine and C-terminal Lys-biotin was also added to the cleavage buffer. The cleavage buffer was eluted from the resin, and the buffer was exchanged with 150 mM NaCl, 50 mM MES, pH 6.0 using an Amicon 10 kDa MWCO filter unit. The GST thioester proteins were concentrated to >1 mg/ml, flash frozen, and stored at −80 °C; and they appeared to be >95% pure by Coomassie-stained SDS–PAGE (Figure 9) Protein concentration was determined by SDS–PAGE referenced to bovine serum albumin standards using Coomassie staining.

Ubiquitin expression and purification.

Human ubiquitin was C-terminally fused with Mxe intein into a pTXB1 vector and subjected to mutagenesis as described above to generate L73X and G76X residues. The ubiquitin plasmids were expressed in *E. coli* Rosetta(DE3). LB media (15 ml) containing 100 μg/ml ampicillin and 10 μg/ml chloramphenicol was inoculated with a single colony and grown overnight at 37 °C in a shaking incubator. 10 ml from overnight cultures was used to inoculate 1 l LB media containing 100 μg/ml ampicillin and 10 μg/ml chloramphenicol, and the cultures were grown in shaker flasks at 37 °C until OD_{600} = 0.7, then 1 ml 1M IPTG was added to induce expression, and the cultures were further incubated for 3 h at 37 °C.

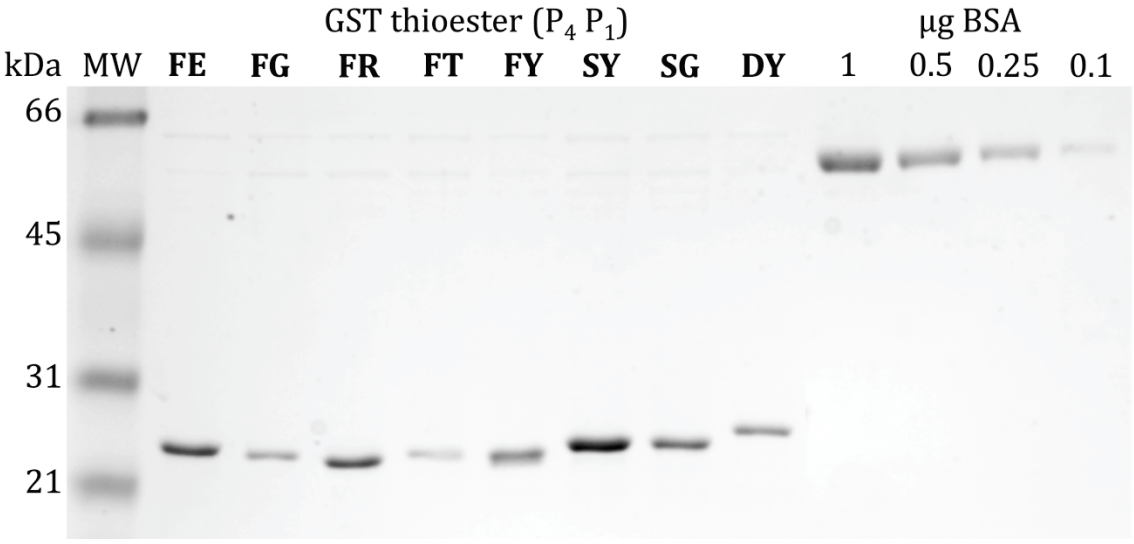


Figure 9. GST-thioester constructs. SDS-PAGE analysis of all GST-thioesters used for ligations. MW, molecular weight markers.

Cells were pelleted and then resuspended in 20 ml lysis buffer (250 mM NaCl, 50 mM HEPES, 1 mM EDTA, 10% glycerol, pH 7.5).

E. coli cells were lysed by French press, the lysate was pelleted (17,500 × g, 30 min, 4 °C), and the supernatant was loaded onto 5 ml of chitin resin (NEB). Resin was washed with 150 ml wash buffer (250 mM NaCl, 25 mM HEPES, 0.1% Triton X-100, pH 7.5) then incubated overnight in cleavage buffer (250 mM NaCl, 50 mM HEPES, 300 mM MESNA, pH 7.5). The cleavage buffer was eluted from the resin, and the buffer was exchanged with 150 mM NaCl, 50 mM MES, pH 6.0 using an Amicon 3.0 kDa MWCO filter unit. The ubiquitin thioester proteins were concentrated to 1 mM, flash frozen and stored at −80 °C; and they appeared to be >95% pure by Coomassie-stained SDS–PAGE. Protein concentration was determined by SDS–PAGE referenced to bovine serum albumin standards using Coomassie staining.

Peptide synthesis.

Peptides were synthesized either on a Prelude or PS3 peptide synthesizer, from Protein Technologies, using standard Fmoc-based solid-phase peptide synthesis. Peptides were synthesized by double coupling every residue. Fmoc groups were deprotected for five times, 10 min each with 20% piperidine in DMF. Coupling times were 1.5 h. Biotinylated peptides were synthesized using N-ε-biotin-lysine Wang resin (Iris Biotech or P3 Biosystems). All peptides were deprotected and cleaved from resin using reagent K (82.5:2.5:5:5:5–trifluoroacetic acid:ethane dithiol:water:thioanisole:phenol) then purified using reverse-phase C18

HPLC and lyophilized. Table 4 lists the peptide sequences for these studies. Peptide structures were confirmed using MALDI mass spectrometry (Figure 10), and peptide concentrations were determined by amino acid analysis.

Ubiquitin and GST ligations.

The standard ligation reaction conditions employed 40 μM protein thioester, 3 mM biotinylated peptide, 0.5 μM subtiligase in a buffer containing 100 mM BICINE, 5 mM $CaCl_2$, pH 8.0 for 90 min at 25 °C before quenching with SDS loading dye. A no-subtiligase control was also included, as well as a zero time point where the reaction was quenched before adding subtiligase. Aliquots (3 μl) of quenched reaction mixtures were run on 12% SDS–PAGE and detected by Coomassie blue staining or western blot for ubiquitin or GST proteins, respectively. For western blot, after transfers to membranes using an iBlot system, the membranes were blocked overnight in 50 mg/ml BSA in TBS-T. Membranes were incubated with anti-biotin HRP-linked antibody (Sigma no. A4541) at a 1:10,000 dilution for 1 h, washed several times by TBS-T, and developed with Amersham ECL western blotting detection reagents (GE Healthcare) and imaged by a GeneSys imaging system for 4 min. Band intensities were quantified using ImageJ software (https://imagej.nih.gov/). For native chemical ligation, 100 μM of ubiquitin-thioester was incubated with 3 mM of peptide for up to 20 minutes. The reaction was quenched by 50 mM cysteine on ice for 20 minutes and the excess of cysteine was removed by trichloroacetic acid precipitation.

ID	Sequence
1	GLSGRGKGGK-Biotin
2	MTSGRGKGGK-Biotin
3	SISGRGKGGK-Biotin
4	ALSGRGKGGK-Biotin
5	ELSGRGKGGK-Biotin
6	RLSGRGKGGK-Biotin
7	PLSGRGKGGK-Biotin
8	YLSGRGKGGK-Biotin
9	GGSGRGKGGK-Biotin
10	FLSGRGKGGK-Biotin
11	CRGKGGKGLGKGGAK-Biotin

Table 4. Peptides used for GST and Ubiquitin Ligations.

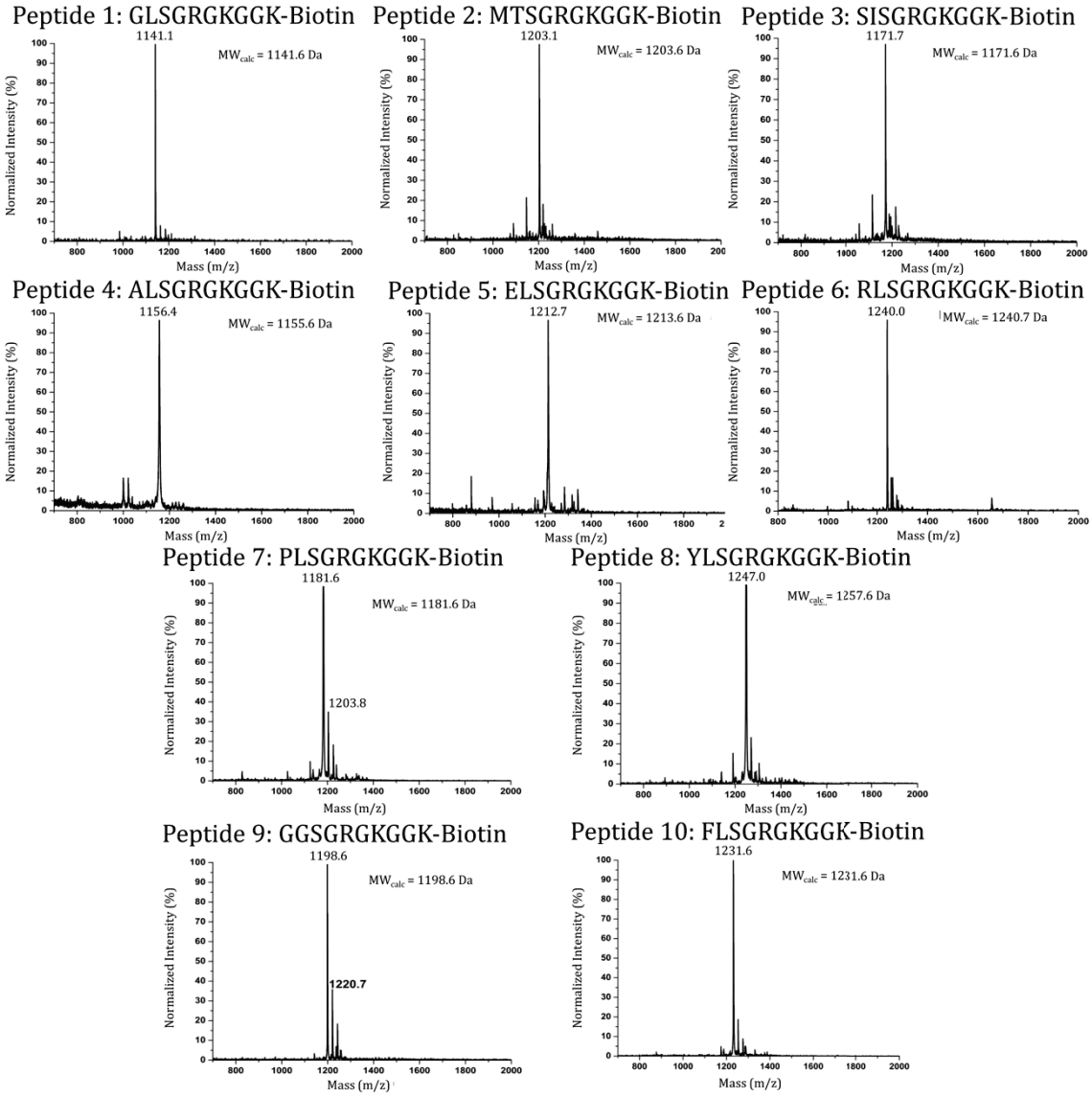


Figure 10. MALDI mass spectra of peptides used for GST and ubiquitin Ligations

Purification of ubiquitin ligations and tryptic digestion

The completed ubiquitin ligation was loaded onto 500 µl pre-blocked (10 mM biotin followed by 0.1 M Glycine, pH 2.8, followed by 500 mM NaCl, 50 mM Na_2HPO_4, 5 mM DTT, pH 7.0) mono-avidin agarose resin (Thermo) on ice. The agarose was washed with PBS buffer and the biotinylated ligation product was eluted with 2 mM biotin in PBS buffer. The eluate with 2.5 µg of ligation product was subjected to trypsin digestion with 50 ng trypsin in 50 mM Tris-HCl, pH 8.0 at 37°C for 3 hrs and stopped by 0.1% TFA, and analyzed by MALDI-MS. MS data was analyzed using the Protein Analysis Worksheet (Genomic Solutions).

Results

Subtiligase-catalyzed ligations with ubiquitin

We initially examined subtiligase-catalyzed ligation with ubiquitin (Ub) C-terminal thioester, produced from an intein fusion protein after reaction with MESNA (Figure 11). The extreme C-terminal residues of Ub are LRGG and it is noteworthy that the known preferences of subtiligase include a hydrophobic residue at the P_4 position[45, 53, 54] (see Figure 4), which is Leu in Ub. A synthetic decapeptide (GLSGRGKGGK(biotin), 1 mM) was reacted with 100 µM Ub thioester, 7.5 µM subtiligase, 100 mM bicine at pH 8, and 5 mM Ca^{+2} at room temperature and the reaction was monitored by Coomassie-stained SDS-PAGE. As shown in Figure 12, the ligation appeared to plateau by 60 min at 50% ligated product and had already achieved half-maximal conversion by 5 min.

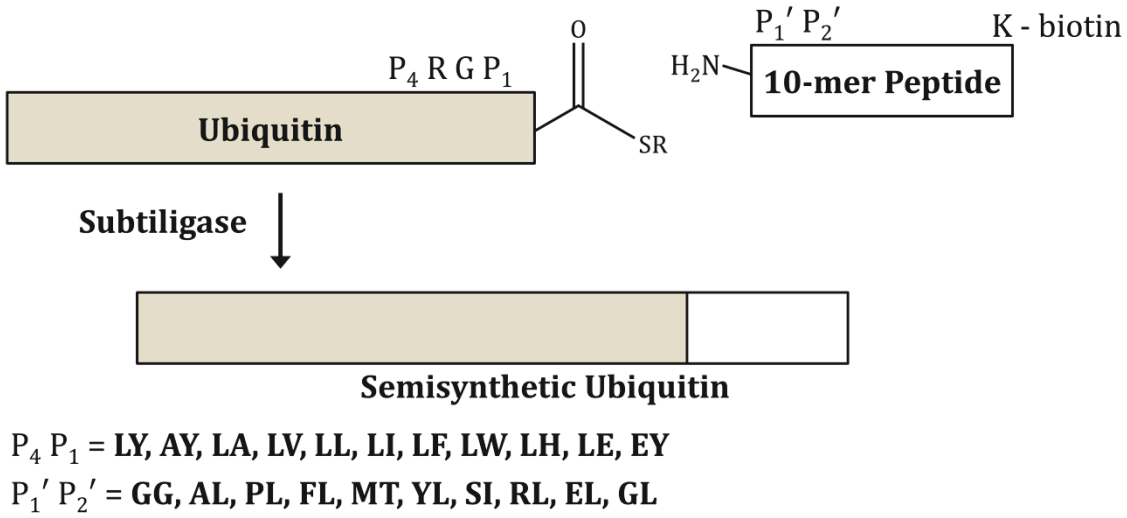


Figure 11. General scheme for ligations between ubiquitin and 10-mer biotinylated peptides.

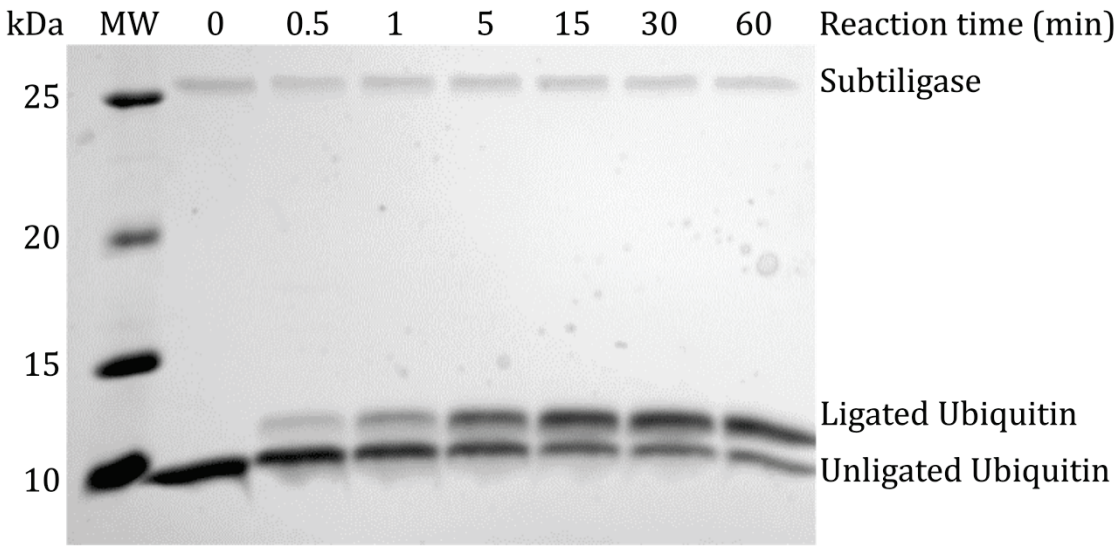


Figure 12. Time-course analysis of a subtiligase-catalyzed ligation between a ubiquitin thioester (P_4 P_1 = LY) and peptide 1 (P_1' P_2' = GL). MW, molecular weight markers.

Mass spectrometry confirmed the predicted mass and peptide linkage for ligation (Figure 13). Interestingly, the recovered ubiquitin showed evidence that hydrolysis of the thioester occurred in a fashion accelerated by subtiligase (Figure 14), which could explain why the reaction stalled at 50% conversion.

Prior structural and reactivity data obtained with subtiligase and synthetic peptide oxyester ligations suggested that the P_4, P_1, P_1', and P_2' positions were the most influential for ligation efficiency.[45, 46, 53, 54] Thus, to further optimize and understand the generality of the reaction, we next explored Ub variants that contained a range of residues at the P_4 (Leu, Ala, Glu) and P_1 (Tyr, Leu, Ile, Val, Phe, Ala, Trp, His, Gly) positions and a set of synthetic acceptor peptides with diversity introduced at the P_1' (Gly, Ala, Arg, Phe, Met, Ser, Tyr, Pro, Glu) and P_2' (Leu, Thr, Ile, Gly) positions (Table 4 and Figure 10). As shown in Figure 15 and Table 5, the majority of the conversion efficiencies were in the range of 50-60%. Notable exceptions to these relatively good subtiligase-catalyzed Ub ligation efficiencies were when P_4, P_1, or P_1' was glutamate or aspartate or when P_1' was proline, which were previously known to be poor substrates for oxyester peptide ligations.[10, 45, 46, 53, 54]

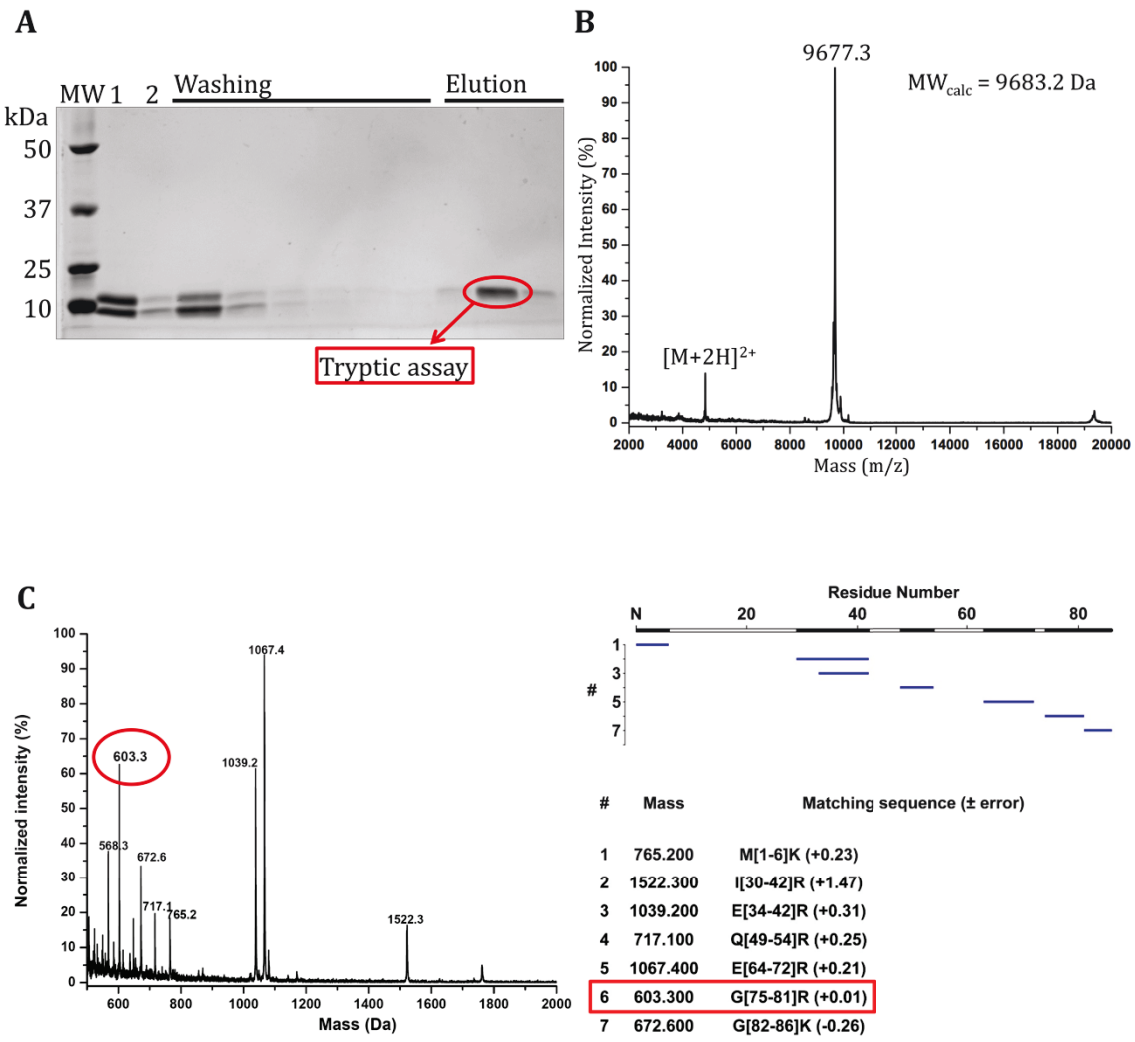


Figure 13. Mass spectrometry analysis of the ligated ubiquitin product. (**A**) Monoavidin purification of ligated ubiquitin product. MW, molecular weight markers; Lane 1, ubiquitin ligation loaded on monoavidin resin; Lane 2, Flow through from monoavidin resin. (**B**) MALDI-MS analysis of purified ubiquitin ligation product showing the expected mass from the ligation. (**C**) MALDI-MS analysis of tryptic digest of purified ubiquitin ligation product. The highlighted mass corresponds to the peptide containing the ligation junction.

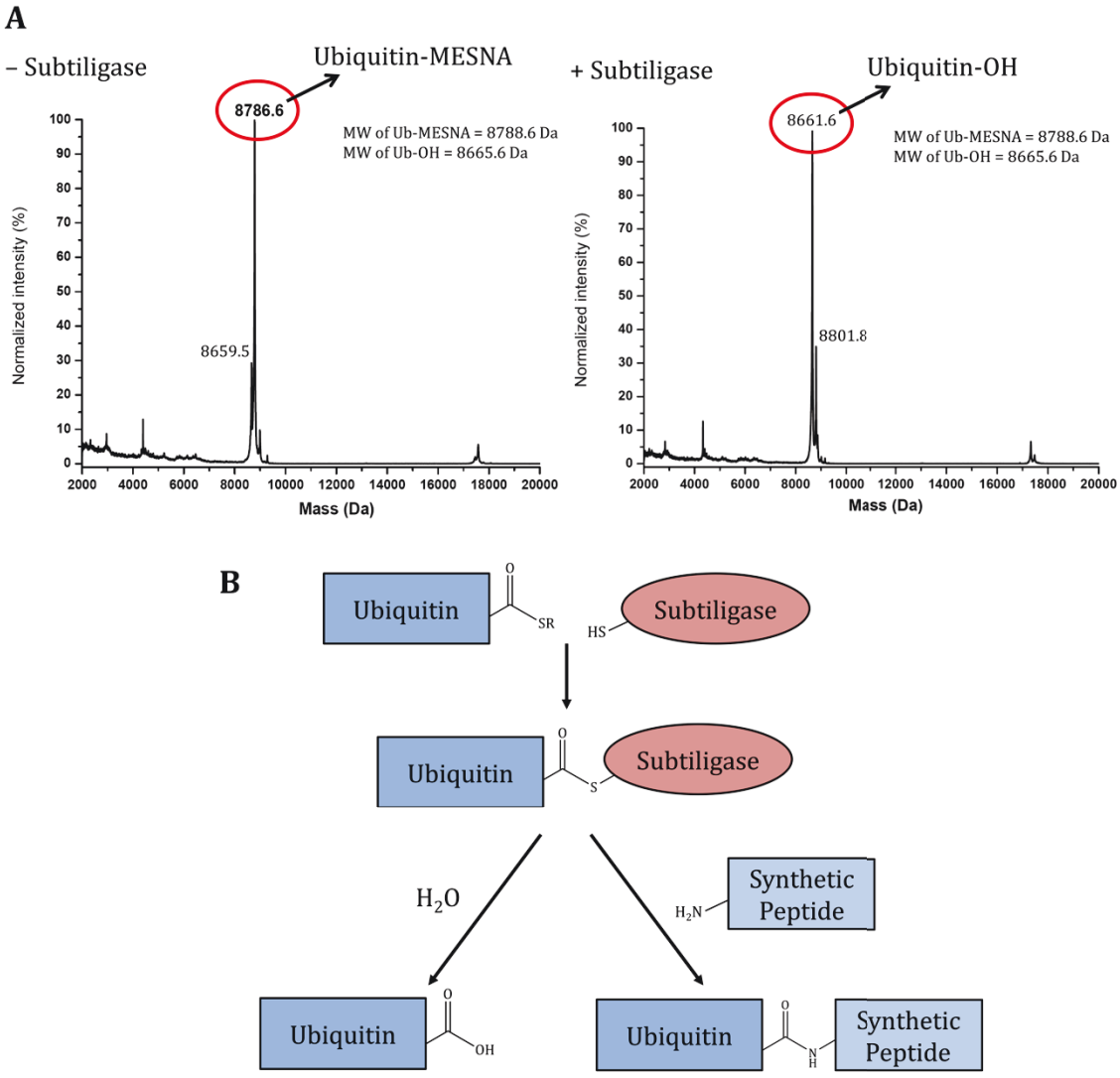


Figure 14. Subtiligase-mediated hydrolysis of ubiquitin-thioester. (**A**) MALDI-MS analysis of ubiquitin thioester after incubation with or without subtiligase. (**B**) Proposed partitioning of the acylated enzyme intermediate between hydrolysis and aminolysis.

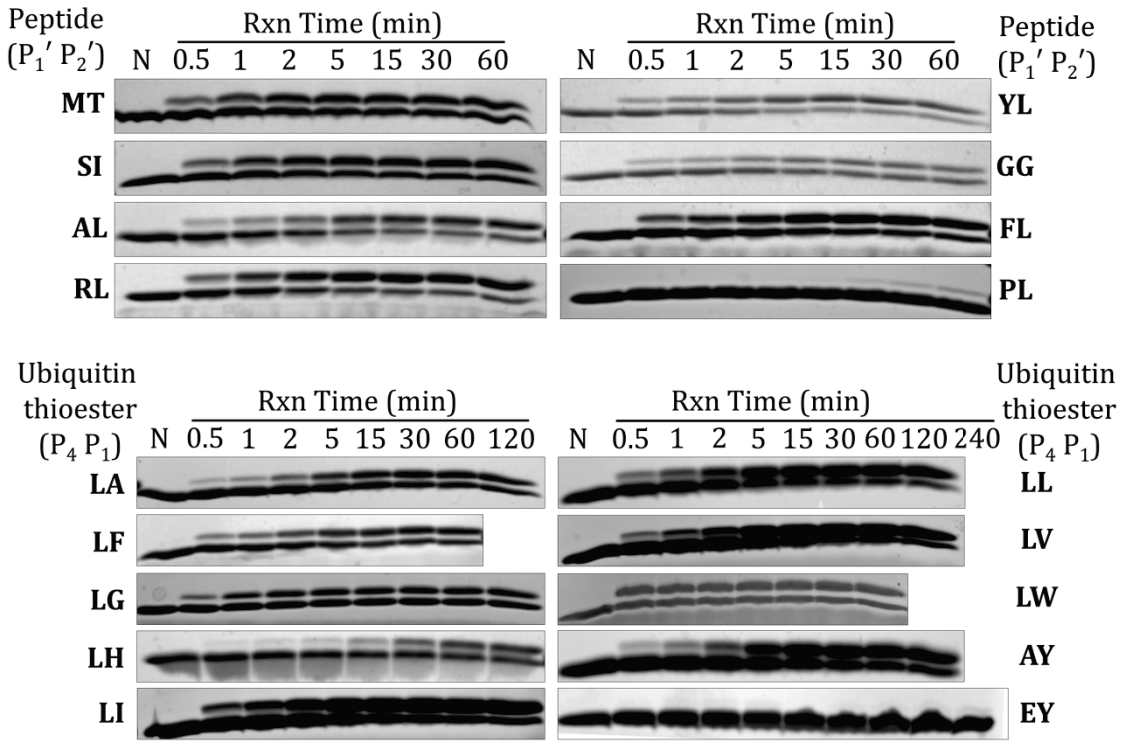


Figure 15. Results of Subtiligase-catalyzed ubiquitin ligations Top: SDS-PAGE analysis of subtiligase-catalyzed ligations between ubiquitin-MESNA (P_4 P_1 = LY) and the indicated peptides (peptides 2-4, and 6-10, Table 4). Representative of two replicates as reported in Table 5. Bottom: SDS-PAGE analysis of subtiligase-catalyzed ligations between the indicated ubiquitin-MESNA thioester and peptide 1 (P_1' P_2' = GL). Representative of two replicates as reported in Table 5.

Ubiquitin thioester (P_4 – P_1)	Peptide (P_1' and P_2')	Percent Conversion
Leu-Arg-Gly-Tyr	Gly-Gly	47 ± 1
	Ala-Leu	55 ± 2
	Pro-Leu	9 ± 1
	Phe-Leu	61 ± 1
	Met-Thr	50 ± 1
	Tyr-Leu	60 ± 2
	Ser-Ile	58 ± 1
	Arg-Leu	63 ± 2
	Glu-Leu	20 ± 2*
	Gly-Leu	69 ± 2
Ala-Arg-Gly-Tyr	Gly-Leu	56 ± 3
Leu-Arg-Gly-Ala		47 ± 6
Leu-Arg-Gly-Val		63 ± 6
Leu-Arg-Gly-Leu		61 ± 3
Leu-Arg-Gly-Ile		62 ± 2
Leu-Arg-Gly-Phe		55 ± 3
Leu-Arg-Gly-Tyr		69 ± 2
Leu-Arg-Gly-Trp		63 ± 0.1
Leu-Arg-Gly-His		56 ± 3
Leu-Arg-Gly-Glu		48 ± 1
Glu-Arg-Gly-Tyr		5

Table 5. Ubiquitin thioester ligations catalyzed by subtiligase. Percent conversions are averages of two replicates based on quantifying Coomassie-stained SDS-PAGE and errors represent the span of duplicates performed on separate occasions.
*, Performed with Y217K subtiligase.

To compare the efficiency of subtiligase-mediated expressed protein ligation to that of the standard chemoselective Cys-mediated ligation, we performed the reactions with an N-Cys-containing peptide substrate head-to-head with a subtiligase-catalyzed reaction that replaced the peptide N-terminal cysteine with a glycine (Figure 16). This experiment revealed that the initial subtiligase reaction rate was approximately 3-fold faster than that of the uncatalyzed Cys-mediated reaction. These results underscore the rapid kinetics of the enzyme-catalyzed ligation versus the Cys-mediated reaction.

Creating Y216K subtiligase

In cases involving acidic residues like glutamate at the C-terminal ligation junction (P_1), the introduction of two additional mutations (E156Q, G166K) to standard subtiligase, referred to here as QK subtiligase has been shown to increase the ligation efficiency in prior oxyester peptide ligations.[45, 54] We also designed and prepared Y217K subtiligase based on structural studies on subtilisin[153] (Figure 17a,b) and its potential to complement acidic residues at the P_1' position. Assessment of QK subtiligase and Y217K subtiligase was performed with some of these challenging Ub ligations. In the case of glutamate at P_1, the conversion efficiency improved from 33% with standard subtiligase to ~50% with either QK or Y217K subtiligase (Figure 17c,d). In cases of glutamate at P_1', the conversion efficiencies were significantly improved, about 2-fold, with Y217K subtiligase relative to the standard or QK forms (Figure 17c,d). These modified subtiligase forms thus broaden the scope of enzyme-catalyzed expressed protein ligation.

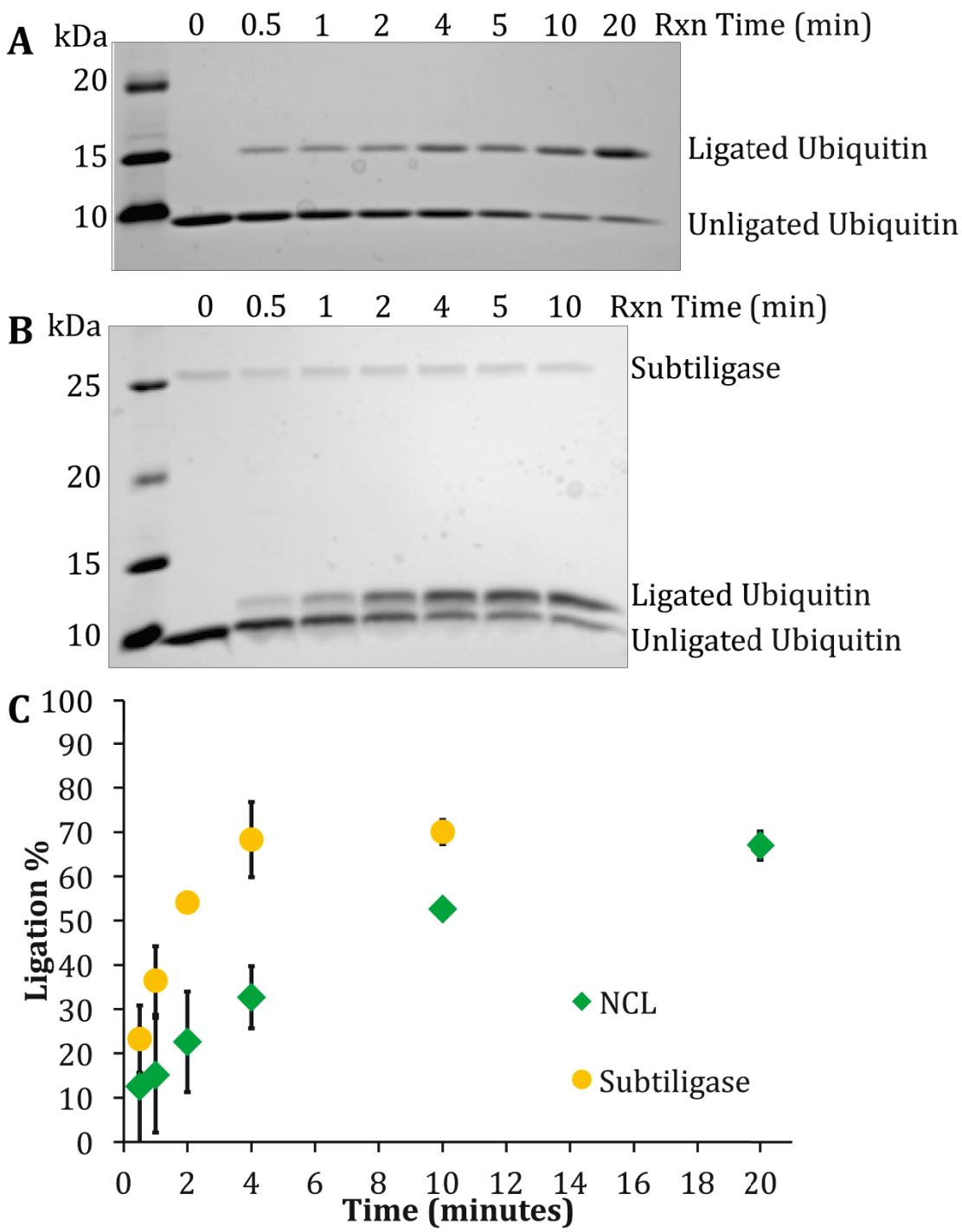


Figure 16. Analysis of subtiligase-catalyzed and native chemical ligations. (**A**) SDS-PAGE analysis of native chemical ligation between ubiquitin-thioester (P_4 P_1 = LY) and peptide 11 (P_1' P_2' = CR). (**B**) SDS-PAGE analysis of subtiligase-catalyzed ligation between ubiquitin-thioester (P_4 P_1 = LY) and peptide 1 (P_1' P_2' = GL) Each ligation was run twice, and the span of values is shown in (**C**).

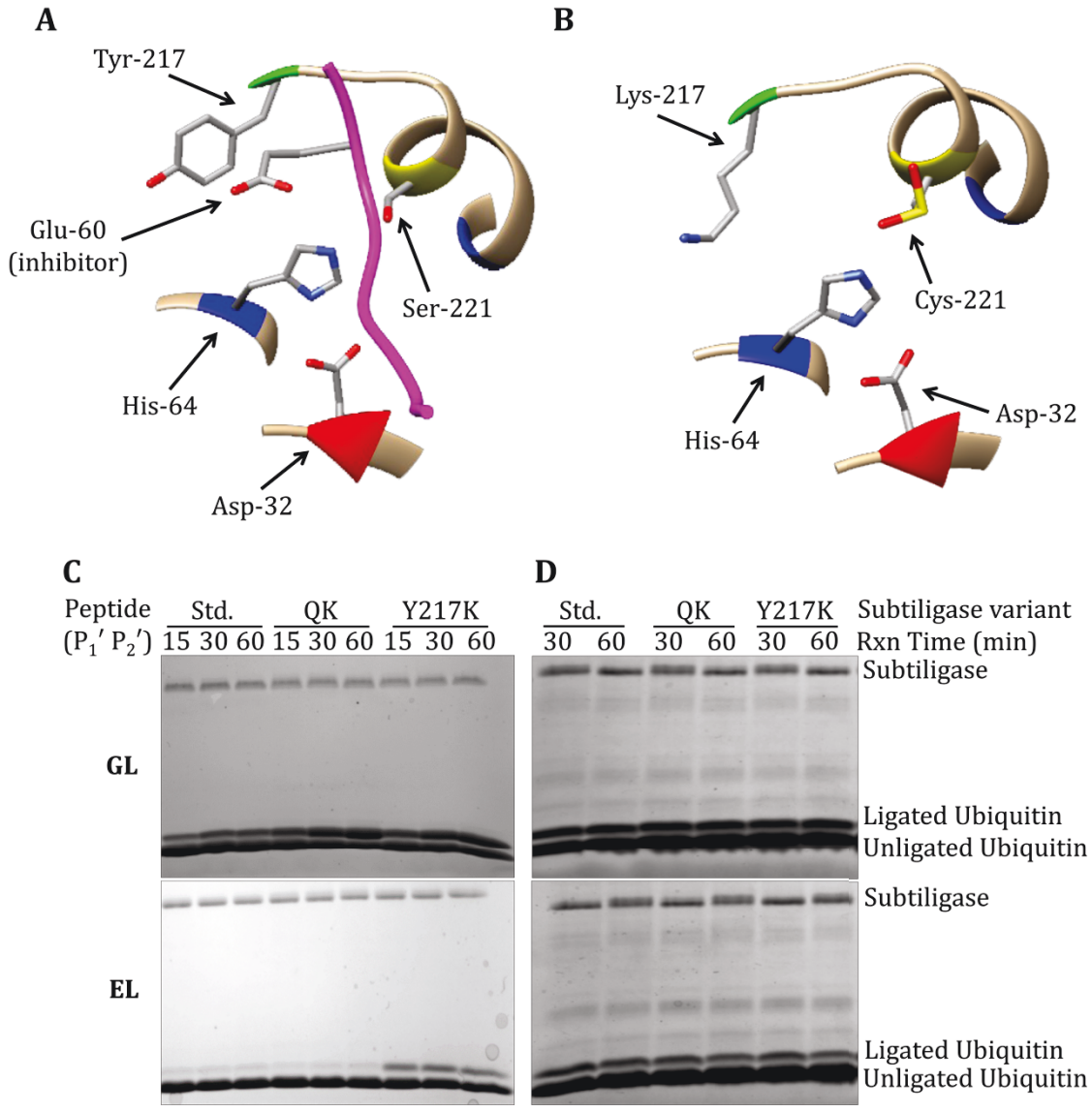


Figure 17. Introducing the Y217K mutation in the active site of subtiligase. (**A**) Crystal structure of subtilisin complexed with a chymotrypsin inhibitor (pink) containing a glutamate residue in the P_1' site (PDB ID: 2SNI[154]). (**B**) Crystal structure of S221C, Y217K subtilisin with the sidechain of Lys-217 oriented toward His-64 and potentially complementary to acidic residues in the P_1' site (adapted from PDB ID: 1SUD[155]). The catalytic triad (Asp-32, His-64, Ser/Cys-221) is highlighted. (**C**), SDS-PAGE analysis of subtiligase ligations between ubiquitin-MESNA (P_4 P_1 = LE) and peptides 1 and 5. (**D**), SDS-PAGE analysis of subtiligase ligations between ubiquitin-MESNA (P_4 P_1 = LD) and peptides 1 and 5.

Subtiligase-catalyzed ligations with GST

To further assess the generality of subtiligase-mediated expressed protein ligation, we explored the method with a series of GST-thioesters containing a range of C-terminal P_4 (Phe, Ser, Asp) and P_1 (Tyr, Thr, Arg, Gly, Glu) residues. These reactions were followed by quantitative western blotting based on the presence of a biotin attached to a set of 10-mer peptides containing varying P_1'-P_2' residues (Met-Thr, Gly-Leu, and Ser-Leu) (Figure 18). Of these 24 different GST thioester reactions, 18 showed conversion percentages in the range of 30-60% whereas 6 showed minimal or no product formation (5% conversion) (Figure 19 and Table 6). As with the Ub thioester ligations, an acidic residue (aspartate) at P_4 was not tolerated. Interestingly, with serine at P_4, reaction was favorable when P_1 was tyrosine (conversions close to 60%) but not when P_1 was glycine (5%). However, glycine at P_1 was acceptable when P_4 was Phe with conversions of ~30%. Given the Ub thioester ligations described above, perhaps the most surprising result was that glutamate at P_1 was well-tolerated when P_4 was phenylalanine with conversions of ~60%. These results highlight how there can be an interplay among the C-terminal amino acids influencing ligation efficiency.

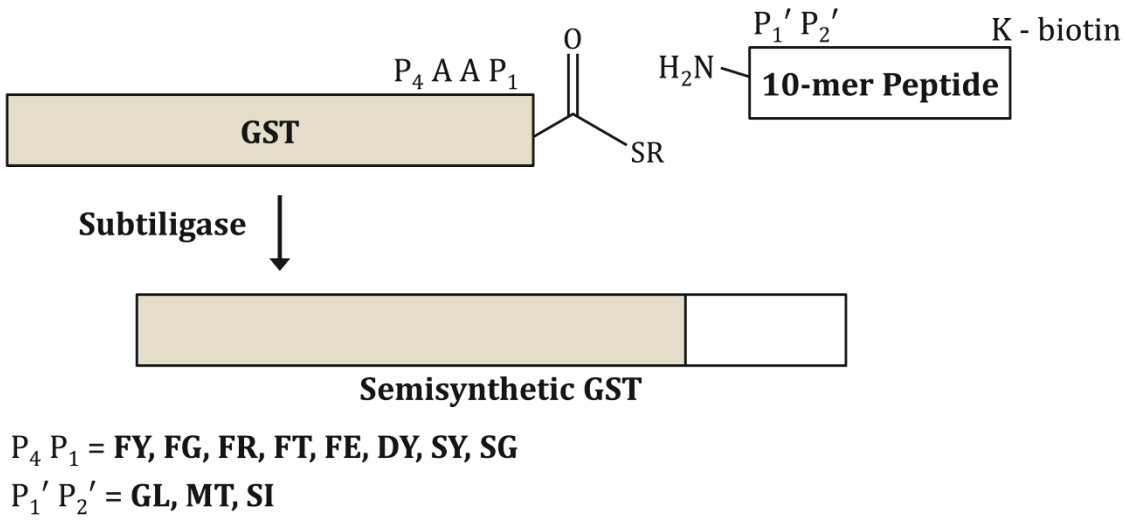


Figure 18. General scheme for ligations between GST and 10-mer biotinylated peptides.

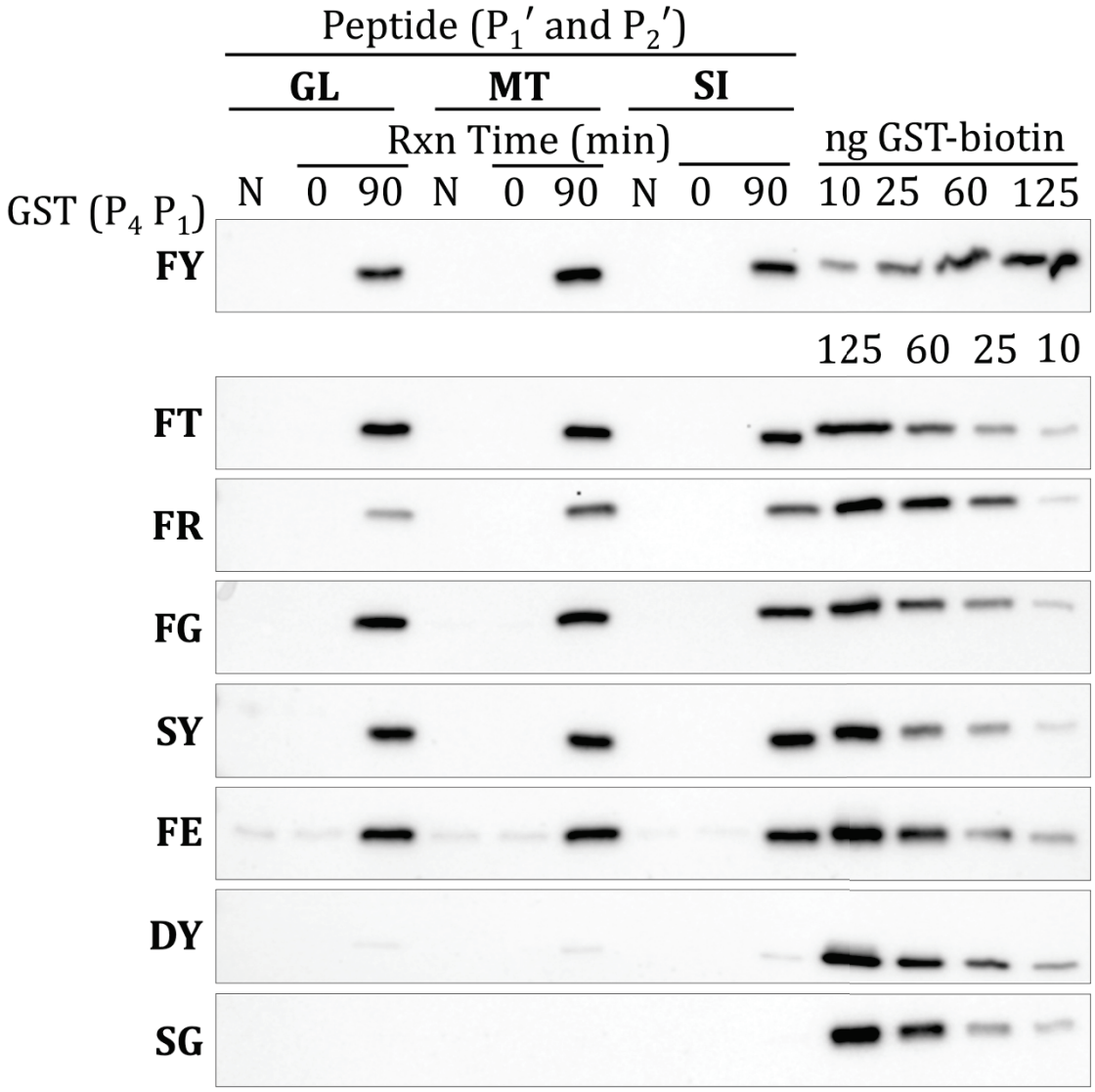


Figure 19. Western blot analysis of subtiligase-catalyzed ligations between the indicated GST-MESNA thioester and peptides 1-3 (Table 4). Representative of two replicates as reported in Table 6.

GST thioester (P_4 - P_1)	Peptide (P_1' and P_2')		
	Gly-Leu	Met-Thr	Ser-Ile
	Percent Conversion		
Phe-Ala-Ala-Tyr	53 ± 3	40 ± 3	42 ± 0.4
Phe-Ala-Ala-Thr	53 ± 3	49 ± 6	49 ± 4
Phe-Ala-Ala-Arg	52 ± 4	35 ± 3	41 ± 5
Phe-Ala-Ala-Gly	35 ± 0.3	32 ± 5	36 ± 6
Ser-Ala-Ala-Tyr	61 ± 3	56 ± 4	64 ± 10
Ser-Ala-Ala-Gly	5	5	5
Asp-Ala-Ala-Tyr	5	5	5
Phe-Ala-Ala-Glu	63 ± 1	58 ± 2	59 ± 4

Table 6. GST thioester ligations catalyzed by subtiligase. Percent conversions are averages of two replicates based on quantifying western blots by using a biotin tag in the peptides and errors represent the span of duplicates performed on separate occasions.

Discussion

Here we have established that the engineered protein ligase, subtiligase, offers benefits as a facile catalyst for enzyme-catalyzed expressed protein ligation. The two principal advantages over conventional expressed protein ligation are the ability to avoid a cysteine at the ligation junction and the speed of ligation. We have found that for a relatively wide variety of residues proximal to the recombinant protein thioester C-terminus (P_4 to P_1), as well as a range of residues within the synthetic peptide (P_1' and P_2') are well-tolerated by the first generation subtiligase (S221C, P225A). The preference established for subtiligase through peptide oxyester ligations that P_4 should be a hydrophobic residue was observed with the recombinant protein thioester ligations here. It is noteworthy that residues at P_1 can influence the acceptability at P_4, as peptides containing serine at P_4 were efficiently ligated when P_1 was a tyrosine but not a glycine. However, a P_1 of glycine was tolerated by subtiligase when P_4 was phenylalanine. Another example of the interdependence was the case of glutamate at P_1, which was poorly ligated by the first generation subtiligase when P_4 was leucine, but was better accepted when P_4 was phenylalanine.

In the cases examined, first generation subtiligase-mediated ligations were inefficient when P_4, P_1 or P_1' were glutamate, with the exception noted above. The previously described QK subtiligase variant complements a glutamate at P_1 and the newly designed Y217K subtiligase aids in ligations with glutamate at P_1 or P_1'. As hypothesized, Y217K subtiligase appears to complement acidic residues at P_1' and is thus likely to be a generally useful tool for enzyme-catalyzed expressed

protein ligation. These findings suggest that it may be feasible to design or screen for other mutants of subtiligase to overcome challenging ligations.

The maximal conversions of subtiligase-mediated expressed protein ligations appear to be around 70%. Since these ligations are performed with a large excess of synthetic peptide, we believe this plateauing stems from competing hydrolysis vs. peptide aminolysis of the covalent subtiligase cysteine-protein intermediate. It would thus be advantageous for next generation subtiligases be engineered to inhibit such thioester hydrolysis, promoting completion of the ligation reactions.

Chapter 3: Creating Wild Type Semisynthetic Phospho-PTEN

Introduction

We next turned our attention to an application of subtiligase-mediated expressed protein ligation in the context of the key tumor suppressor lipid phosphatase PTEN. PTEN is responsible for the conversion of phosphatidylinositol trisphosphate (PIP_3) to phosphatidylinositol bisphosphate (PIP_2) by dephosphorylation at the 3' position.[77] PTEN is known to be phosphorylated on a cluster of four C-terminal serine and threonine residues (S380, T382, T383, S385) and these phosphorylations can inhibit the enzyme and prevent membrane association by inducing an intramolecular conformational change in PTEN.[24, 25, 110, 156] Conventional cysteine-mediated expressed protein ligation has been used previously to generate 380,382,383,385-tetraphosphorylated PTEN,[24] as well as mono-, di-, and tri-phosphorylated PTENs.[25] In these case, a Y379C mutation was introduced into PTEN since there is no nearby cysteine available for the ligation reaction. It has subsequently been discovered[117] that Y379C PTEN behaves anomalously in cells, mimicking a non-phosphorylatable form with enhanced membrane localization relative to wild type. Thus, generating a tetraphosphorylated PTEN form without including the cysteine mutation is a desirable goal to dissect the contributions of Tyr379 to PTEN conformation and function. The natural PTEN sequence immediately preceding the C-terminal phosphorylation cluster, 374-

PDHYRYS-380, suggested, based on the ligation results above with ubiquitin and GST, that positioning Tyr-377 as the P_1 residue and Pro-374 as the P_4 residue could be favorable for subtiligase-mediated expressed protein ligation (Figure 20).

Methods

r-PTEN expression and purification

Quikchange mutagenesis was used to covert t-PTEN[24, 25] to r-PTEN. Table 7 lists the primers used for mutagenesis. A pFastBac1 baculovirus vector containing the PTEN-intein-CBD fusion (PTEN residues 1 – 377) was used to make bacmid and then baculovirus in SF-21 insect cells using standard methods,[24] and the baculovirus was then used to infect High Five insect cells with M.O.I. 1.0. After growing infected High Five cells in Express Five SFM Media (Gibco) for 48 h at 27 °C, they were pelleted (700 × g, 10 min, 4 °C) and then resuspended in 1/20 of the medium used for culture, pelleted again (discarding the supernatant), and then flash frozen and stored at −80 °C. Resuspended cells from 200 ml culture were lysed in a 40 ml homogenizer in 30 ml lysis buffer (150 mM NaCl, 50 mM HEPES, 1 mM EDTA, 10% Glycerol, 0.1% Triton X-100) containing three dissolved protease tablets (Roche). The lysate was then centrifuged (17,500 x g, 40 min, 4 °C), and the supernatant was added to a 10-ml bed of powdered cellulose (Sigma). After 1 h of incubation at 4 °C, the lysate was drained from the cellulose and then bound to 5 ml chitin resin (NEB). The resin was then washed with 150 ml washing buffer (250 mM NaCl, 25 mM HEPES, 0.1% Triton X-100, pH 7.5), and incubated overnight in cleavage buffer (250 mM NaCl, 50 mM HEPES, 300 mM MESNA, pH 7.5).

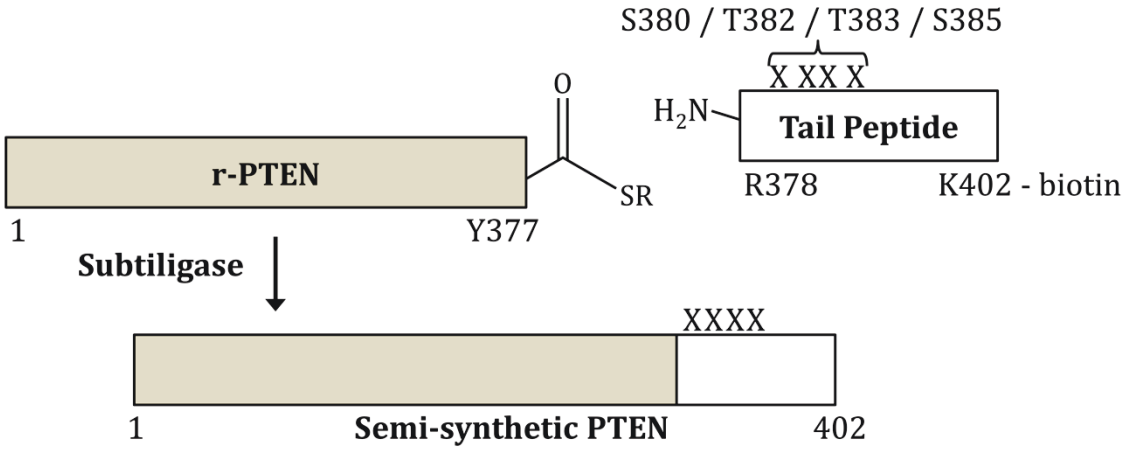


Figure 20. General scheme for ligations between r-PTEN and C-terminal tail peptides.

PTEN R378 deletion, F	CAATGAACCTGATCATTAT*TGCATCACGGGAGATGC
PTEN R378 deletion, R	GCATCTCCCGTGATGCA*ATAATGATCAGGTTCATTG
Subtiligase M222A, F	CGTACAGCGGTACGTGC**GCA**GCATCTGCGCACGTTGC
Subtiligase M222A, R	GCAACGTGCGCAGATGC**TGC** GCAC GTACCGCTGTACG

Table 7. Primers used for PTEN and subtiligase mutagenesis. Asterisks indicate deleted residues.

The cleavage buffer was eluted from the resin, and the buffer was exchanged with 150 mM NaCl, 50 mM MES, pH 6.5 using an Amicon 10 kDa MWCO filter unit. The PTEN thioester protein produced was shown to be >80% pure by Coomassie-stained SDS–PAGE and concentrated to greater than 1 mg/ml. If the r-PTEN protein thioester was not used within a day, it was flash frozen and stored at −80 °C. For stable t-PTEN generation, 50 mM Cys was added to the cleavage buffer, and it was otherwise handled as described above.

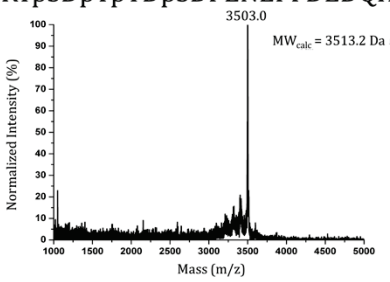
Peptide synthesis

Peptides corresponding to residues 378–403 of PTEN were synthesized by double coupling residues 386–402 for 1.5 h each, triple coupling residues Asp-384 and Asp-381 for 1.5 h each, and double coupling phosphorylated residues for 3 h each. Phosphate groups were monoprotected by O-benzyl groups during the synthesis. All peptides were deprotected and cleaved from resin using reagent K (82.5:2.5:5:5:5–trifluoroacetic acid:ethane dithiol:water:thioanisole:phenol) then purified using reverse-phase C18 HPLC and lyophilized. Table 8 lists the peptides used for PTEN ligations. Peptide structures were confirmed using MALDI mass spectrometry (Figure 21), and peptide concentrations were determined by amino acid analysis.

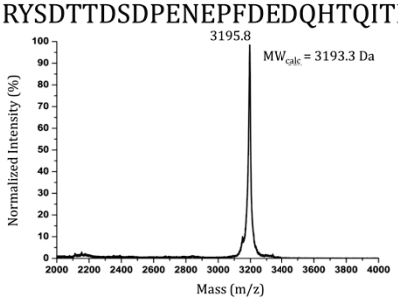
ID	Sequence
4p tail	RYpSDpTpTDpSDPENEPFDEDQHTQITK-Biotin
np tail	RYSDTTDSDPENEPFDEDQHTQITK-Biotin
1p-380 tail	RYpSDTTDSDPENEPFDEDQHTQITK-Biotin
1p-382 tail	RYSDpTTDSDPENEPFDEDQHTQITK-Biotin
1p-383 tail	RYSDTpTDSDPENEPFDEDQHTQITK-Biotin
1p-385 tail	RYSDTTDpSDPENEPFDEDQHTQITK-Biotin

Table 8. Peptides used for PTEN ligations.

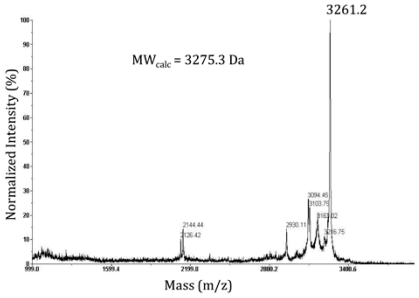
4p-tail:
RYpSDpTpTDpSDPENEPFDEDQHTQITK-Biotin



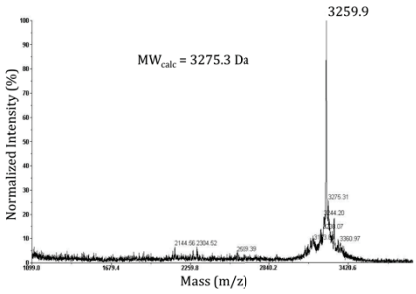
np-tail:
RYSDTTDSDPENEPFDEDQHTQITK-Biotin



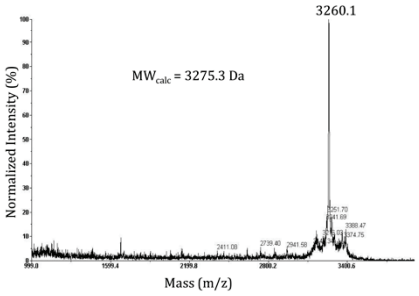
1p-380-tail:
RYpSDTTDSDPENEPFDEDQHTQITK-Biotin



1p-382-tail:
RYSDpTTDSDPENEPFDEDQHTQITK-Biotin



1p-383-tail:
RYSDTpTDSDPENEPFDEDQHTQITK-Biotin



1p-385-tail:
RYSDTTDpSDPENEPFDEDQHTQITK-Biotin

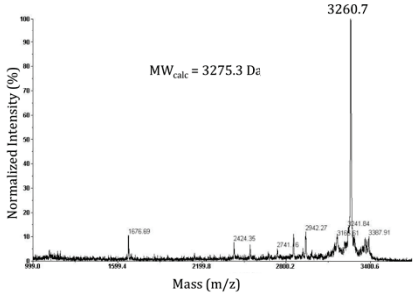


Figure 21. MALDI mass spectra of peptides used for PTEN ligations.

PTEN ligation and purification

Under standard conditions (1 ml), a reaction mixture containing r-PTEN thioester (~1 mg/ml) C-terminal peptide (~10 mg/ml) in buffer (100 mM BICINE, 100 mM $CaCl_2$, pH 8.0) was treated with subtiligase (~25 μM final). Note that higher subtiligase and increased Ca^{2+} concentrations relative to the ubiquitin and GST cases appeared to improve the ligation yield, presumably because of the slower rate of this ligation. A 3 μl aliquot was saved as a negative control before adding subtiligase. The reaction was split into 150 μl aliquots and incubated at 25 °C. After 4 h, 3 μl of the mixture was taken for SDS–PAGE analysis. The remainder of the mixture was injected onto a Superdex 75 size-exclusion column (GE Healthcare Life Sciences). Size-exclusion chromatography was performed with a flowrate of 0.6 ml/min in a buffer containing 500 mM NaCl, 50 mM Na_2HPO_4, 5 mM DTT, pH 7.0; and 0.3 ml fractions were collected, and 5 μl of each fraction was analyzed by Coomassie-stained SDS–PAGE. Fractions containing ligated PTEN were combined and stored at 4 °C overnight and then loaded onto 500 μl pre-blocked (10 mM biotin followed by 0.1 M Glycine, pH 2.8, followed by 500 mM NaCl, 50 mM Na_2HPO_4, 5 mM DTT, pH 7.0) mono-avidin resin (Thermo) on ice. Resin was washed sequentially with 5 ml of 500 mM NaCl, 50 mM Tris, 10 mM DTT, pH 8.0 for 1 h followed by 5 ml of 1 M NaCl, 50 mM Tris, 10 mM DTT, pH 8.0 for 1 h. Next, the resin was equilibrated with 5 ml 150 mM NaCl, 50 mM Tris, 10 mM DTT, pH 8.0 and then eluted with 10 mM biotin in the same buffer. The resin was incubated with 100 μl of elution buffer for 10 – 15 min at room temperature, drained, then incubated with an additional 100 μl of elution buffer. This process was repeated for up to a total of 500 μl. The purified

semisynthetic protein was shown to be >90% pure by Coomassie-stained SDS–PAGE. Mono-phosphorylated and non-phosphorylated PTENs were prepared and purified in the same manner, but Na_2HPO_4 was excluded from all buffers.

MEFs and western blot analysis.

Murine embryo fibroblasts (MEF) obtained from the Stivers lab at Johns Hopkins University or purchased from the ATCC, shown to be mycoplasma negative, were grown in DMEM, high glucose (Thermo Fisher) with 10% fetal bovine serum (FBS) at 37 °C, 5% CO_2. Once the cells had reached ~70% confluency, the media was exchanged with DMEM with 2% FBS. Cells were then either treated with 50 μg/ml 4,5,6,7-tetrabromobenzotriazole (TBB) or DMSO. After 12–15 h incubation at 37 °C, cells were washed with PBS then lysed with RIPA buffer plus 1 mM PMSF. The lysate was pelleted at 15,000 × g (10 min, 4 °C). The protein concentration in the supernatant was determined by BCA assay (Thermo Fisher). Supernatant with 50 μg of protein was run on a 10% SDS–PAGE gel along with Y379-4p-PTEN standards. Western blot membranes were incubated with either anti-phospho-PTEN antibody 44A7 (Cell Signaling no. 9549) or anti-PTEN antibody N-19 (Santa Cruz Biotech no. sc-6818) at 1:1,000 dilutions. Western blots were developed and analyzed as described above.

Alkaline phosphatase sensitivity assay

This was adapted from previous methods.[24, 25] PTEN (1 μg) was incubated with 5 μM calf intestine phosphatase (NEB) in 20 μl of reaction buffer (50

ng/μl PTEN) for a total of 4 h. Reaction buffer consisted of 50 mM Tris, 10 mM BME, pH 8.0. Samples were taken from the reaction at various time points, diluted ten-fold in SDS loading dye, and run on SDS–PAGE for western blot analysis. 10 μl were loaded for the Cys-PTEN and 2 μl were loaded for the Tyr-PTEN. Blots were analyzed using anti-phospho-PTEN antibody (Cell Signaling no. 9554) at 1:1,000 dilution. Images were analyzed using ImageJ software, and Prism 6 software (Graphpad) was used to determine phosphorylation half-life fit to a standard exponential decay. Replicates with different preps showed similar half-lives (within 20%).

Phosphatase activity assay

PTEN activity toward a water-soluble substrate (diC6-PIP_3, from Avanti Polar Lipids) was determined by the evolution of inorganic phosphate as measured with a malachite green[157] detection kit (R and D Biosystems). Assays for were conducted as described previously.[24] Briefly, 0.5 – 1.5 μg PTEN was incubated with 160 μM diC6 PIP_3 for 10 minutes in 25 μl reaction buffer (50 mM Tris, 10 mM BME, pH 8.0) at 30˚C. Samples were quenched with malachite green reagent (R and D Biosystems) and absorbance was measured at 620 nm.

Results

Creating semisynthetic phosphorylated PTEN

A PTEN construct corresponding to residues 1-377 of PTEN was expressed in insect cells as a fusion protein with a modified intein and CBD. After

purification on chitin resin, MESNA was used to generate the corresponding PTEN thioester (r-PTEN). This was ligated to a synthetic 380,382,383,385-tetraphosphorylated peptide corresponding to residues 378-402 of PTEN, and modified with a biotin at Lys-402. This ligation proceeded to ~30% completion (Figure 22a), and the desired semisynthetic tetraphosphorylated product (Y379-4p-PTEN) was purified by size exclusion and mono-avidin affinity chromatography exploiting the C-terminal biotin (Figure 22b,c). Ligations were also carried out with the corresponding unphosphorylated synthetic peptide and monophosphorylated synthetic peptides, creating Y379-n-PTEN (~50% completion), Y379-1p380-PTEN (~25%), Y379-1p382-PTEN (~30%), Y379-1p383-PTEN (~40%), and Y379-1p385-PTEN (~40%) (Figures 23 – 27). The increasing efficiency of the monophosphorylated peptides as the phosphorylation site shifted away from the ligation junction, and the greatly increased efficiency of the non-phosphorylated peptide compared with the tetraphosphorylated-peptide suggests that phospho-amino acids near the ligation junction may negatively influence interactions with subtiligase.

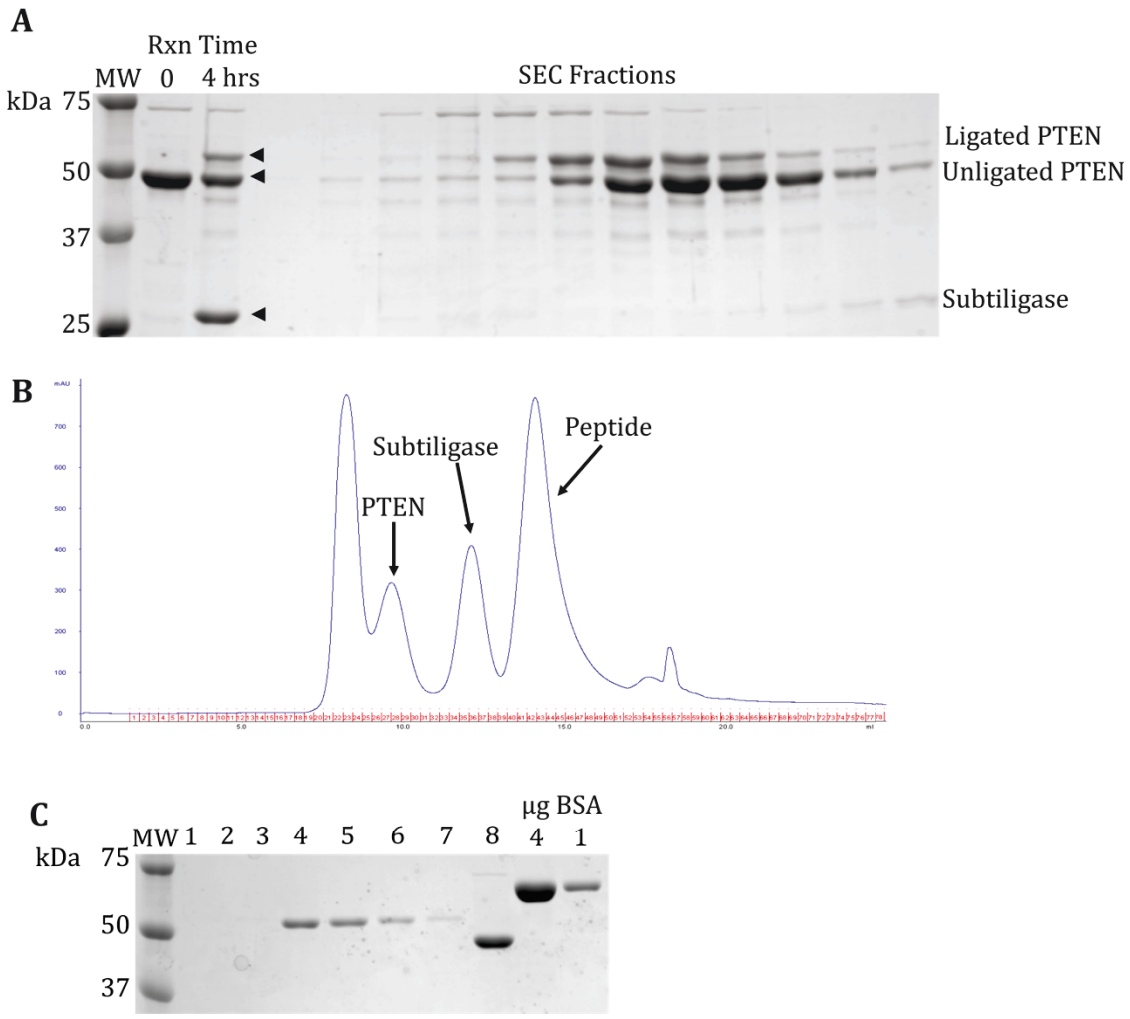


Figure 22. Y379-4p-PTEN ligation and purification. (**A**) SDS-PAGE analysis of the ligation between r-PTEN-MESNA and the 4p-tail peptide. (**B**) Size-exclusion chromatogram of the ligation. (**C**) Representative SDS-PAGE analysis of monoavidin purification of semisynthetic PTEN. MW, molecular weight markers; Lane 1, 0.5 M NaCl wash; Lane 2, 1 M NaCl wash; Lane 3, 150 mM NaCl rinse; Lanes 4-7, 10 mM biotin elution; Lane 8, r-PTEN-MESNA.

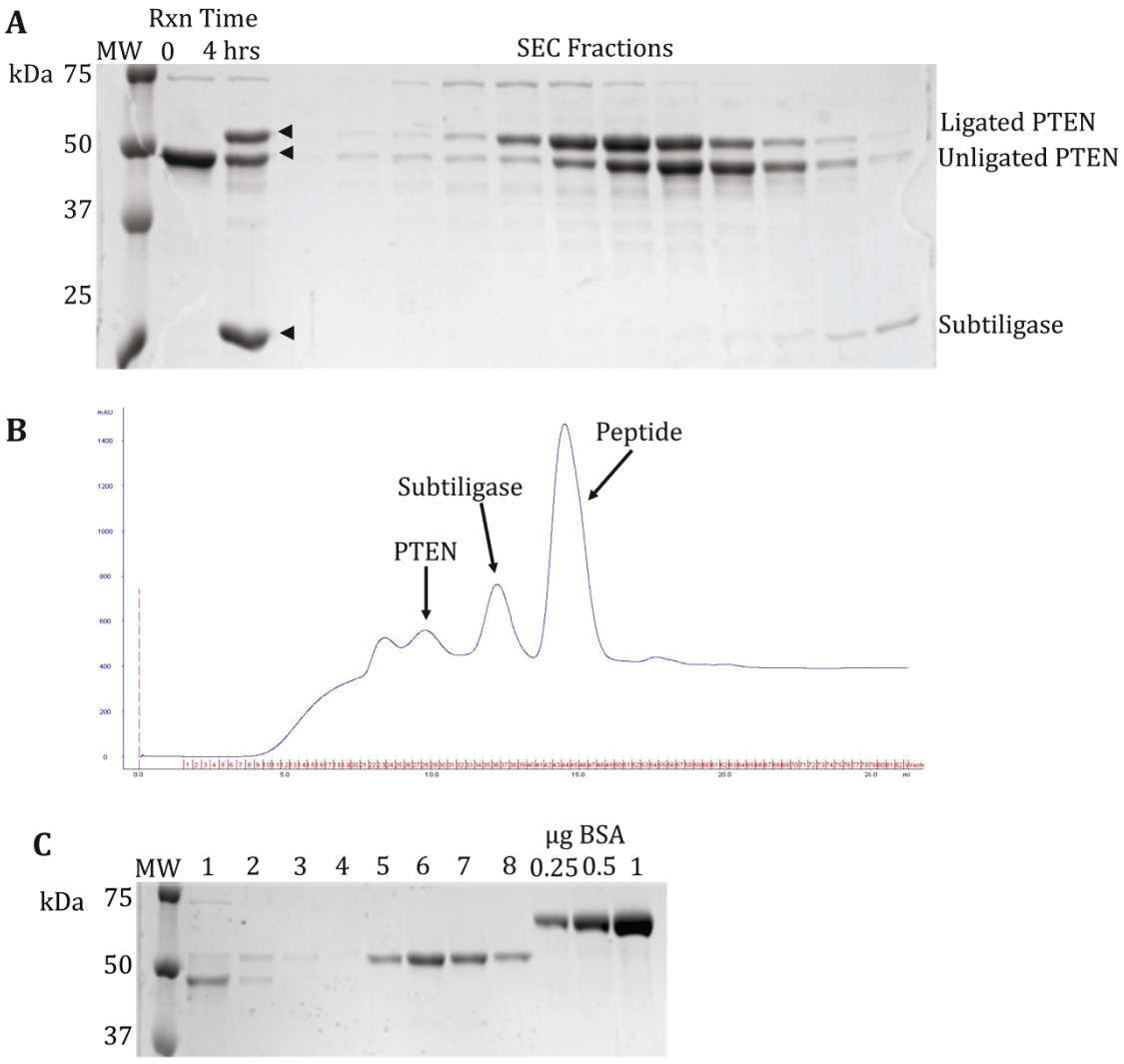


Figure 23. Y379-n-PTEN ligation and purification. (**A**) SDS-PAGE analysis of the ligation between r-PTEN-MESNA and the non-phosphorylated tail peptide. (**B**) Size-exclusion chromatogram of the ligation. (**C**) Representative SDS-PAGE analysis of monoavidin purification of semisynthetic PTEN. MW, molecular weight markers; Lane 1, 0.5 M NaCl wash; Lane 2, 1 M NaCl wash; Lane 3, 150 mM NaCl rinse; Lanes 4-8, 10 mM biotin elution.

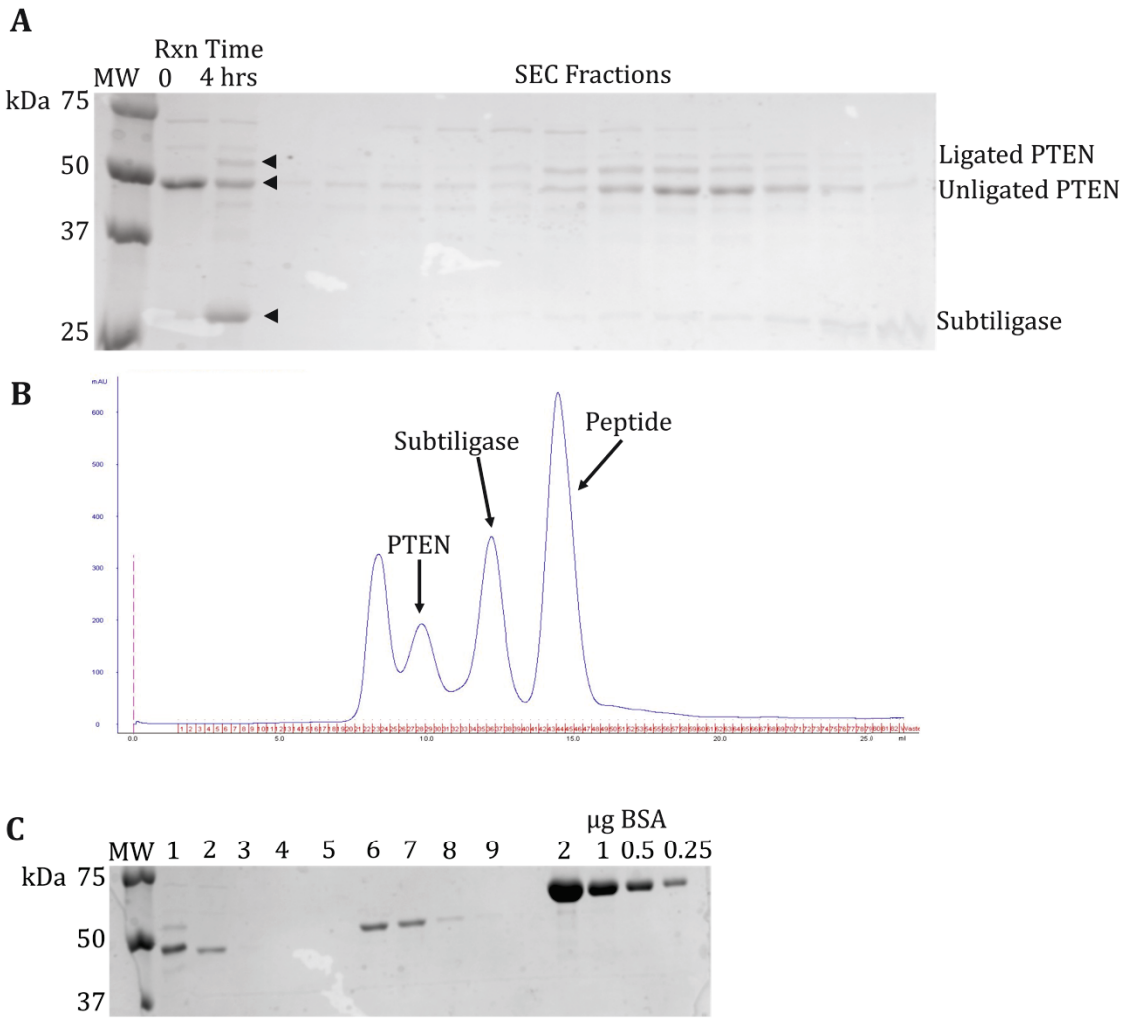


Figure 24. Y379-1p380-PTEN ligation and purification. (**A**) SDS-PAGE analysis of the ligation between r-PTEN-MESNA and the 1p380 tail peptide. (**B**) Size-exclusion chromatogram of the ligation. (**C**) Representative SDS-PAGE analysis of monoavidin purification of semisynthetic PTEN. MW, molecular weight markers; Lane 1, Input from SEC fractions; Lane 2, 0.5 M NaCl wash; Lane 3, 1 M NaCl wash; Lane 4, 150 mM NaCl rinse; Lanes 5-9, 10 mM biotin elution.

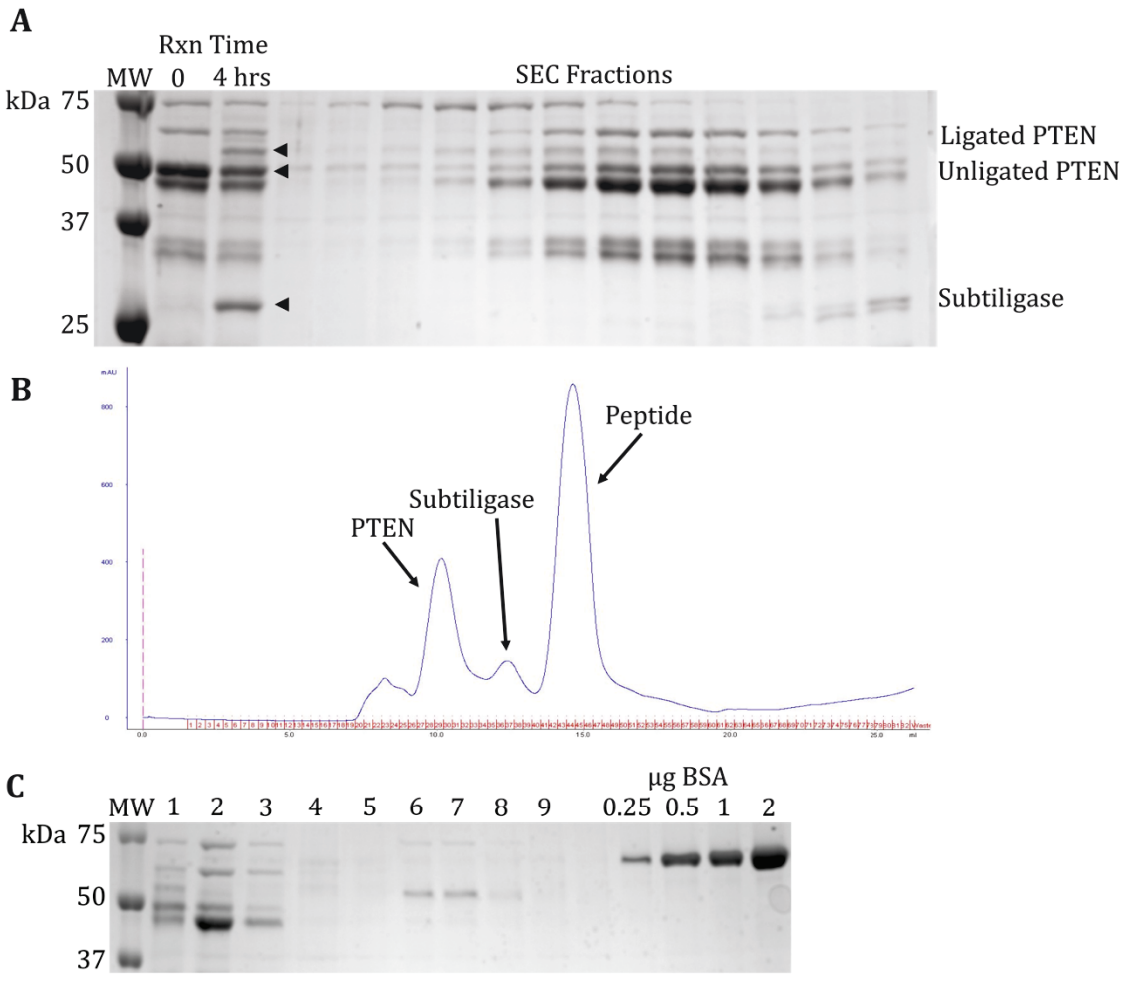


Figure 25. Y379-1p382-PTEN ligation and purification. (**A**) SDS-PAGE analysis of the ligation between r-PTEN-MESNA and the 1p382 tail peptide. (**B**) Size-exclusion chromatogram of the ligation. (**C**) Representative SDS-PAGE analysis of monoavidin purification of semisynthetic PTEN. MW, molecular weight markers; Lane 1, Input from SEC fractions; Lane 2, 0.5 M NaCl wash; Lane 3, 1 M NaCl wash; Lane 4, 150 mM NaCl rinse; Lanes 5-9, 10 mM biotin elution.

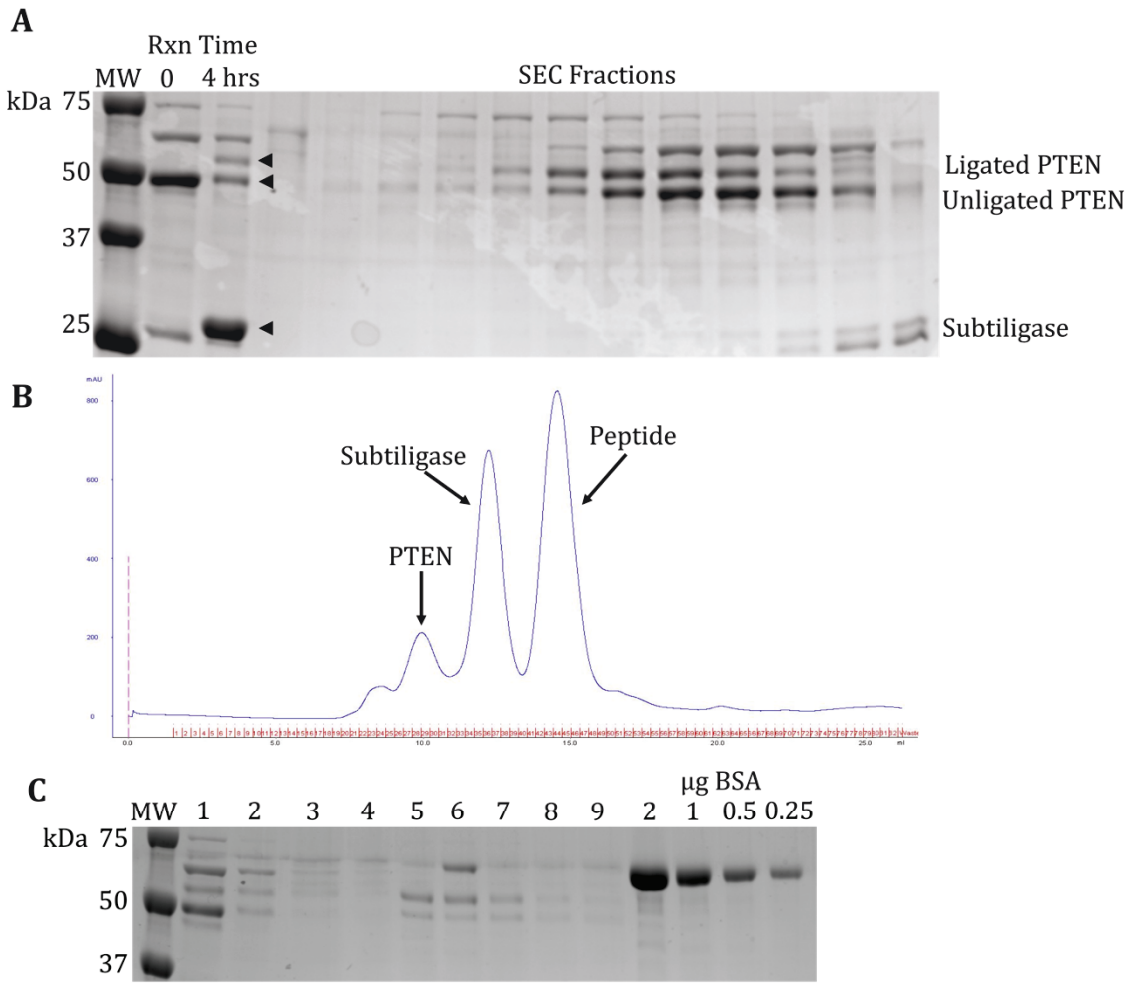


Figure 26. Y379-1p383-PTEN ligation and purification. (**A**) SDS-PAGE analysis of the ligation between r-PTEN-MESNA and the 1p383 tail peptide. (**B**) Size-exclusion chromatogram of the ligation. (**C**) Representative SDS-PAGE analysis of monoavidin purification of semisynthetic PTEN. MW, molecular weight markers; Lane 1, Input from SEC fractions; Lane 2, 0.5 M NaCl wash; Lane 3, 1 M NaCl wash; Lane 4, 150 mM NaCl rinse; Lanes 5-9, 10 mM biotin elution.

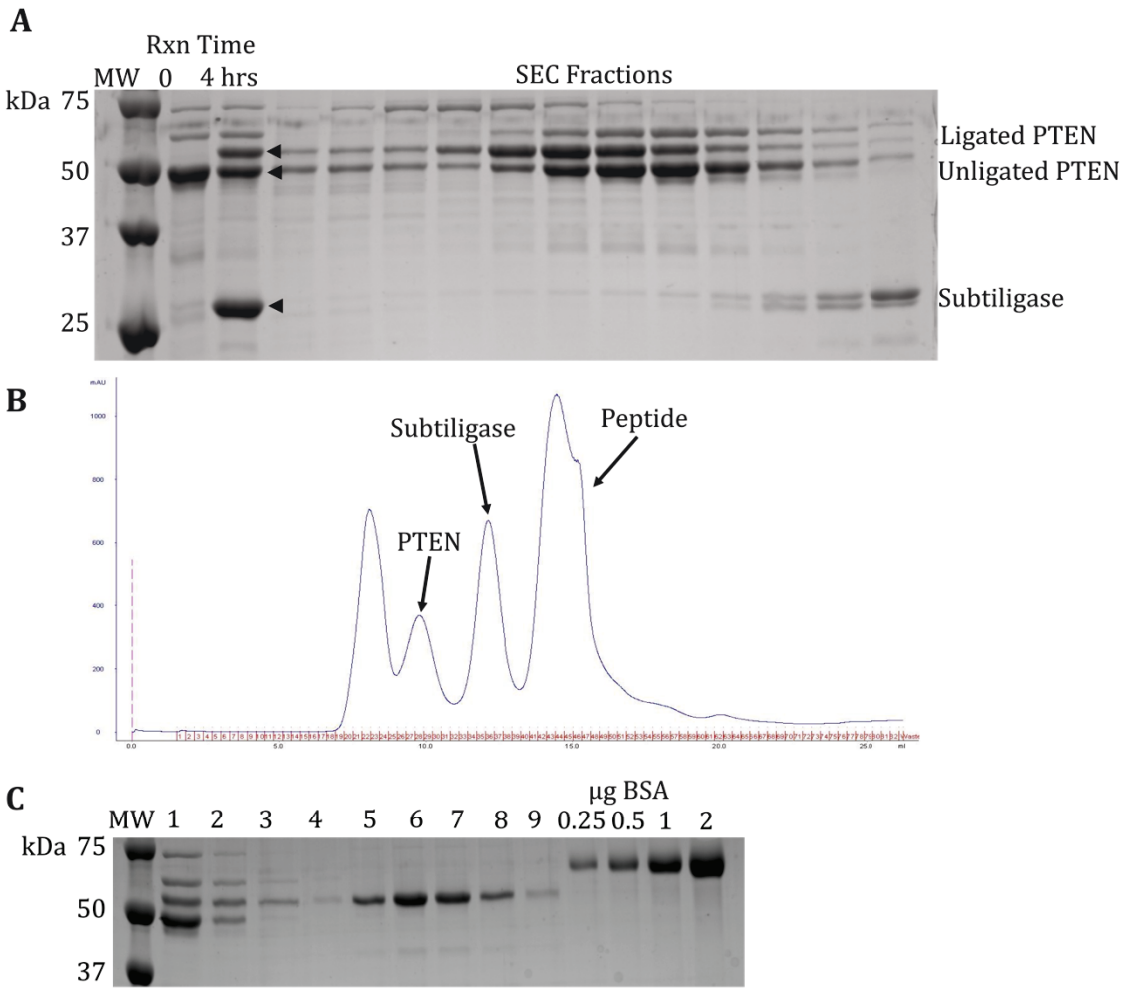


Figure 27. Y379-1p385-PTEN ligation and purification. (**A**) SDS-PAGE analysis of the ligation between r-PTEN-MESNA and the 1p385 tail peptide. (**B**) Size-exclusion chromatogram of the ligation. (**C**) Representative SDS-PAGE analysis of monoavidin purification of semisynthetic PTEN. MW, molecular weight markers; Lane 1, Input from SEC fractions; Lane 2, 0.5 M NaCl wash; Lane 3, 1 M NaCl wash; Lane 4, 150 mM NaCl rinse; Lanes 5-9, 10 mM biotin elution.

Subtiligase M222A

A new subtiligase mutant (M222A) had been developed to increase the yield of ligations. We examined the ability of this subtiligase to ligate GST-MESNA thioesters (P_4 P_1 = FY) to PTEN tail peptides (P_1' P_2' = RY). This mutant performed approximately 30% better than standard subtiligase, giving 55% conversion, compared to 44% for standard subtiligase (Figure 28) Ligations between GST-thioesters (P_4 P_1 = FY) and the four mono-phosphorylated tail peptides with subtiligase M222A all ran to > 50% conversion, indicating that this new subtiligase can accommodate phosphorylated peptides (Figure 29a,b). Unfortunately, when subtiligase M222A was used for a ligation between r-PTEN-MESNA and the 1p-Thr-383 tail peptide, 10% ligation was observed after four hours (Figure 29c), indicating that the while the M222A mutation may have increased the yield of ligations, it also altered the substrate preference of subtiligase. Prolines near the ligation junction are not generally favored by subtiligase,[45, 46, 53, 54] so the lack of ligation may be due to the proline in the P_4 position of r-PTEN preventing subtiligase M222A from productively interacting with the PTEN-thioester.

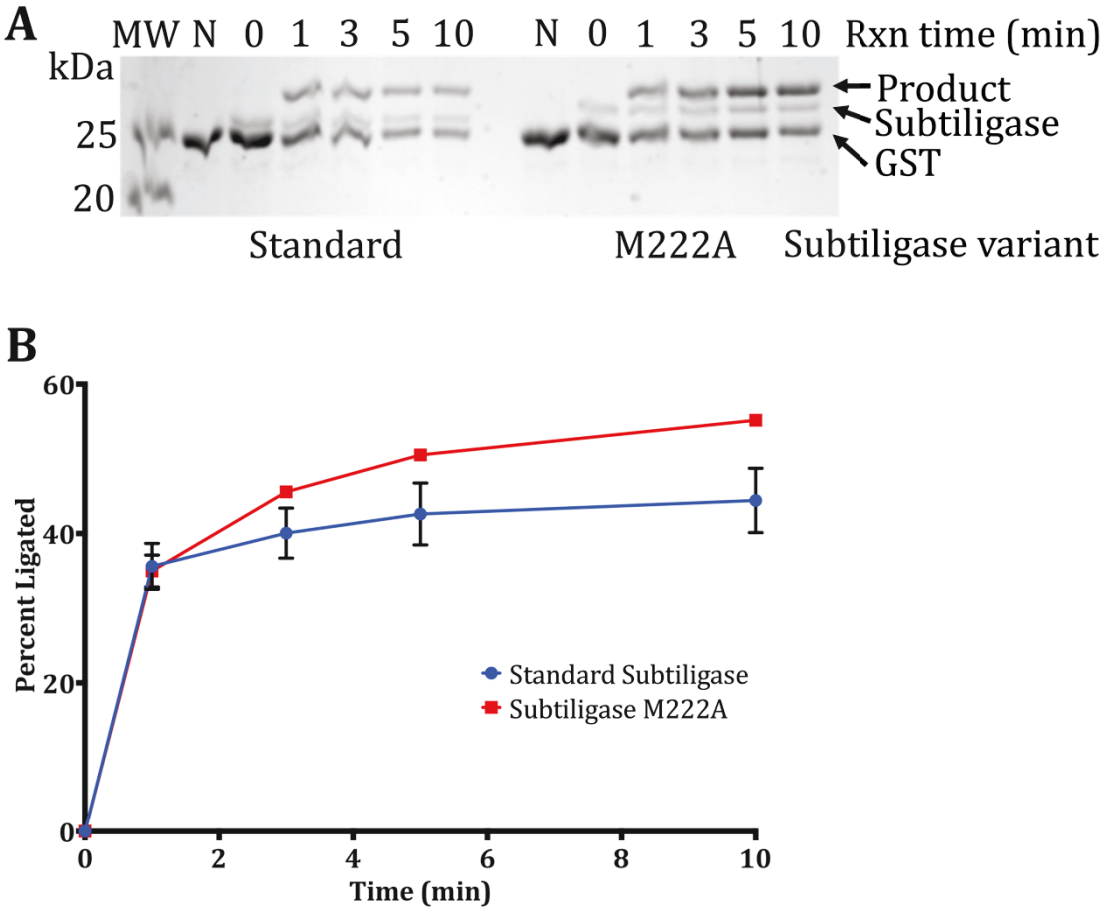


Figure 28. Subtiligase M222A ligations with non-phosphorylated PTEN tail peptides. (**A**) Ligations catalyzed by the indicated subtiligase variant between GST-MESNA (P_4 P_1 = FY) and PTEN non-phosphorylated tail peptides (P_1' P_2' = RY) using the indicated subtiligase. Averages of two replicates ± SEM are shown in (**B**). MW, molecular weight markers; N, no-subtiligase control.

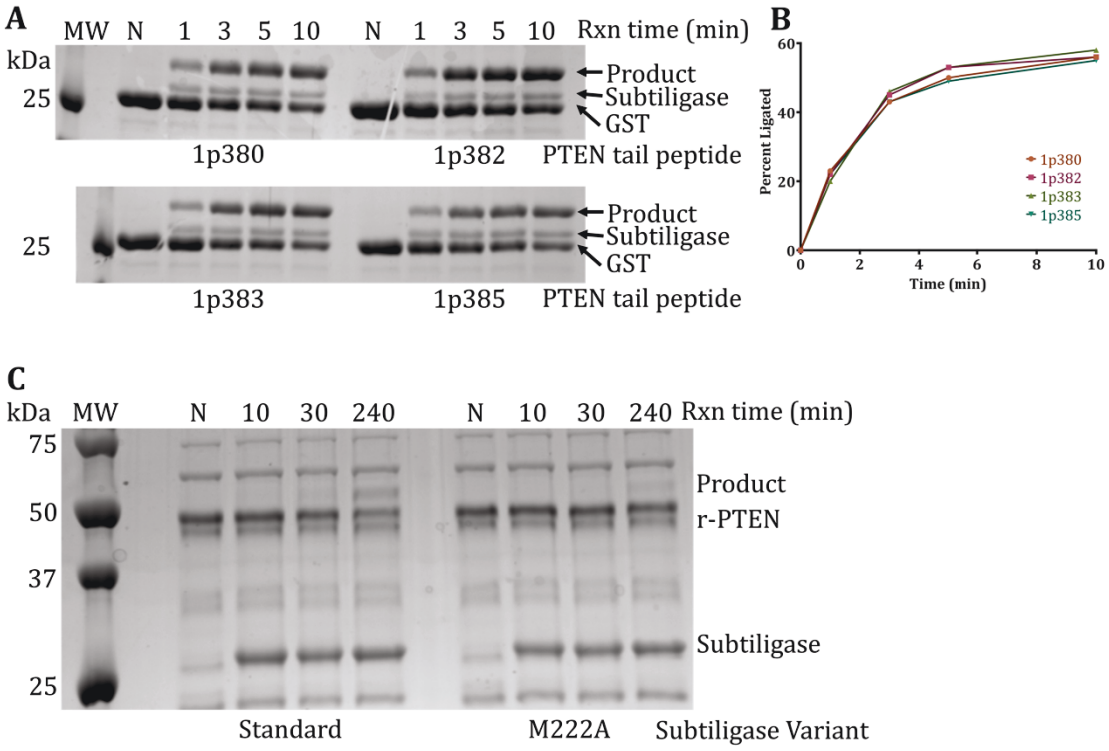


Figure 29. Subtiligase M222A Ligations with phosphorylated PTEN tail peptides. (**A**) Ligations catalyzed by the indicated subtiligase variant between GST-MESNA (P_4 P_1 = FY) and PTEN non-phosphorylated tail peptides (P_1' P_2' = RY). Percent conversions shown in (**B**). (**C**) Results of ligation between r-PTEN-MESNA (P_4 P_1 = PY) and 1p-382 tail peptide (P_1' P_2' = RY) using the indicated subtiligase. MW, molecular weight markers; N, no-subtiligase control.

Western blot analysis of Y379-phospho-PTEN

Western blot of Y379-4p-PTEN with an anti-phospho-PTEN Ab showed a ~4-fold more intense signal for the natural sequence compared with C379-4p-PTEN protein (Figure 30a). We surmise that this is because the Tyr379 residue is an important part of the epitope that is recognized by the commercial anti-phospho-PTEN antibody. This enabled the use of the more natural Y379-4p-PTEN to serve as a standard in determining the level of cellular C-terminal phosphorylation of endogenous PTEN isolated from mammalian cells. By calibrating phospho-PTEN Western blots against total PTEN Western blots, it was found that the stoichiometry of tail phosphorylation of endogenous PTEN in mouse embryonic fibroblasts was ~70% (Figure 30b). The protein kinase CK2 has been reported to be the major kinase responsible for cellular phosphorylation of PTEN,[131] and we tested a CK2 small molecule inhibitor (TBB)[158] to measure how this affects PTEN C-tail phosphorylation. Of note, the overall level of PTEN protein was reduced after CK2 inhibitor treatment, which is consistent with the previously reported suggestion that PTEN phosphorylation stabilizes PTEN protein in cells.[131] However, even accounting for the drop in total PTEN protein, CK2 inhibitor treatment led to reduced C-tail phosphorylation of cellular PTEN (Figure 30b). These results suggest that, although cellular PTEN phosphorylation is near stoichiometric at baseline, it can be kinetically influenced by the dynamic opposing actions of kinases and phosphatases[31]. Previous studies with this same polyclonal antibody showed that it was unable to differentiate between monophosphorylated 1p380-PTEN, and other phospho-PTENs containing phospho-Ser-380.[25] Since this antibody also shows

increased recognition of Tyr-379-containing epitopes relative to those containing Cys-379, we sought to determine if it would also detect Y379-1p380-PTEN as well. We found that it not only recognized Y379-1p380-PTEN, but also Y379-1p383-PTEN, and Y379-1p385-PTEN (Figure 30c). This was surprising, since this antibody is generated against 3p-380,382,383-PTEN tail peptides.

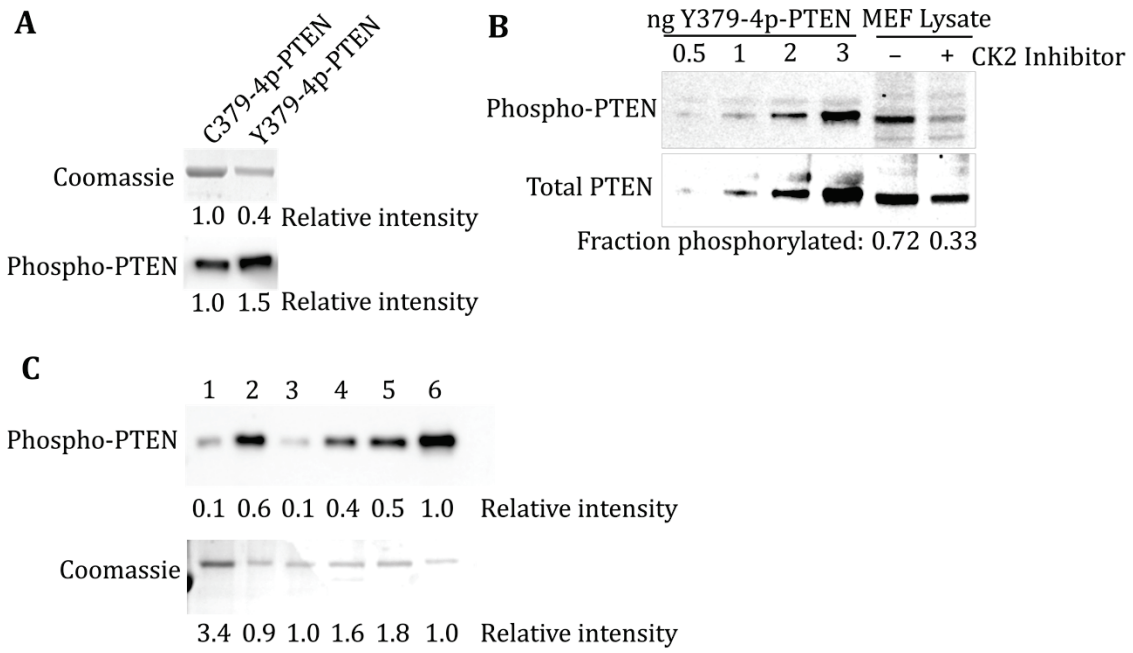


Figure 30. Western blot analyses of phospho-PTEN. (**A**) Anti-phosphorylated-PTEN antibodies respond more strongly to Y379-4p-PTEN than to C379-4p-PTEN. Each image is normalized to the C379-4p-PTEN band. (**B**) Representative Western blots showing a decrease in PTEN phosphorylation following treatment with a CK2 inhibitor. Fraction phosphorylated: Untreated, 0.72 ± 0.06; Treated, 0.33 ± 0.09 (n = 5 biological replicates, p = 0.0071, Student's t-test). (**C**) Anti-phosphorylated-PTEN antibodies also recognize various monophosphorylated PTEN. 1, Y379-n-PTEN; 2, Y379-1p380-PTEN; 3, Y379-1p382-PTEN; 4, Y379-1p383-PTEN; 5, Y379-1p385-PTEN; 6, Y379-4p-PTEN. Each image normalized to the Y379-4p-PTEN band.

Enzymatic activity of Y379-4p-PTEN

We investigated the activity of Y379-4p-PTEN and compared it with C379-4p-PTEN using the soluble PIP_3 substrate di-C6-PIP_3. Prior studies have revealed that C379-4p-PTEN with di-C6-PIP_3 substrate shows reduced catalytic activity versus that of unphosphorylated (n-PTEN) which is matched with truncated (residues 1-379) PTEN[24] , and we confirmed the t-PTEN and C379-4p-PTEN kinetic results here (Figure 31). Such inhibition is consistent with the proposal that the tetraphosphorylated PTEN tail induces an intramolecular conformation in PTEN that impedes catalysis. Interestingly, Y379-4p-PTEN showed a ~3-fold decrease in enzymatic activity compared with C379-4p-PTEN (Figure 31), suggesting that the Tyr379 residue relative to the Cys379 may stabilize the closed PTEN conformation. By comparison, Y379-n-PTEN activity was well-matched with that of t-PTEN (Figure 31) indicating that the Tyr379 containing C-tail in its unphosphorylated form had little impact on PTEN catalytic activity as expected.

We next evaluated the di-C6-PIP_3 phosphatase activity of the PTEN forms at higher ionic strength (Figure 31). When n-PTEN or t-PTEN phosphatase activity is measured with 200 mM NaCl rather than the standard low salt (60 mM) conditions above, both show ~3-fold rate reductions at high salt, whereas C379-4p-PTEN shows only a slight drop (30%) at high salt. We interpret this to mean that high ionic strength suppresses the catalytic activity of the PTEN open form, but this suppression is offset in C379-4p-PTEN because high ionic strength can also shift C379-4p-PTEN to a more conformationally open and active form. In contrast, Y379-4p-PTEN shows different behavior with high salt compared with C379-4p-PTEN.

Y379-4p-PTEN diC6-PIP3 phosphatase activity displayed a sharp 3-fold suppression at high salt, similar to that of n-PTEN and t-PTEN. We deduce from these results that Y379-4p-PTEN is more resistant to conformational opening by high salt compared with C379-4p-PTEN.

Finally, we evaluated the di-C6-PIP_3 phosphatase activity of the four monophosphorylated PTEN forms. All four showed a decrease in catalytic activity relative to truncated or non-phosphorylated PTEN (Figure 32 and Table 9). Y379-1p380-PTEN had showed a 3-fold reduction in k_{cat}/K_M relative to Y379-n-PTEN, which is similar to what had been seen previously with C379-1p380-PTEN.[25] However, Y379-1p382-, Y379-1p383-, and Y379-1p385-PTEN each showed a 7-9-fold reduction, which is greater than the inhibition seen with the corresponding C379-monophosphorylated PTENs.[25] Much of the inhibition of 1p380-, 1p382-, and 1p383-PTEN comes from a reduction in k_{cat}, as these constructs only exhibit a modest increase in K_M. In contrast, 1p385-PTEN exhibited a linear increase in activity with increasing di-C6-PIP_3, which suggests an increase in both k_{cat} and K_M. These data confirm that each phosphorylation event contributes to the overall inhibition of PTEN, but to a greater degree than seen previously. This greater inhibition is presumably due to the presence of Tyr-379. These data also reveal that 1p-385-PTEN is inhibited in a different manner than the other mono-phosphorylated PTENs.

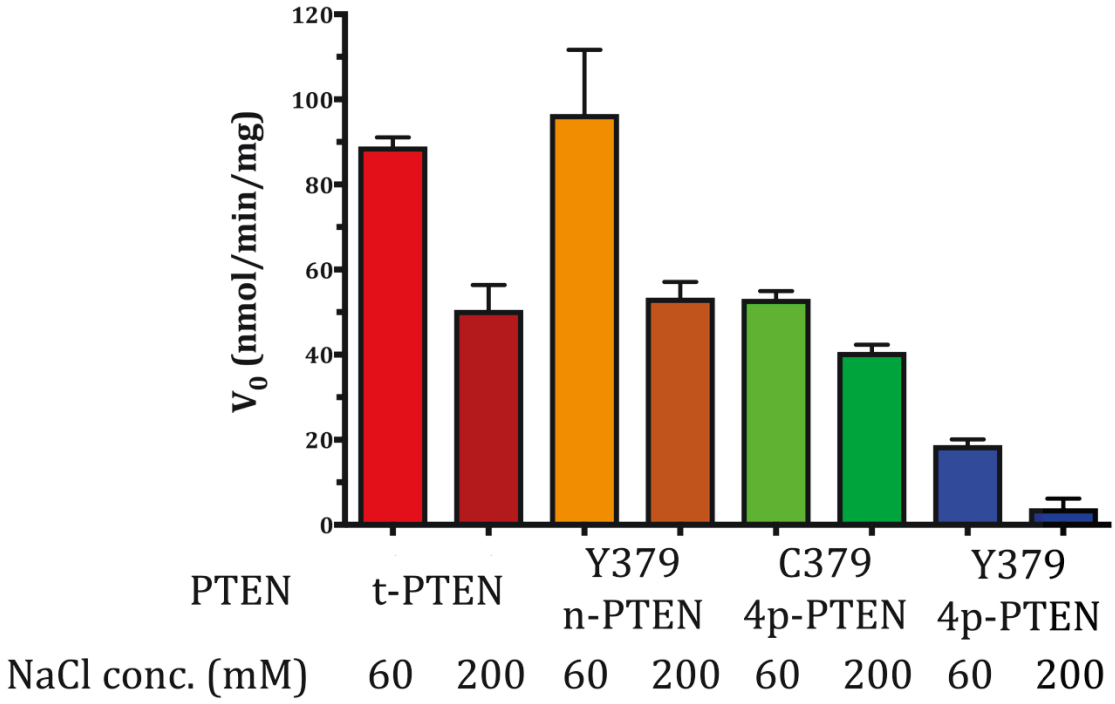


Figure 31. Enzymatic activity of PTEN. Activity of t-, Y379-n-, C379-4p-, and Y379-4p-PTEN toward 160 μM soluble di-C6 PIP3 with 60 mM or 200 mM NaCl. Values are the averages of 2 replicates ± SEM.

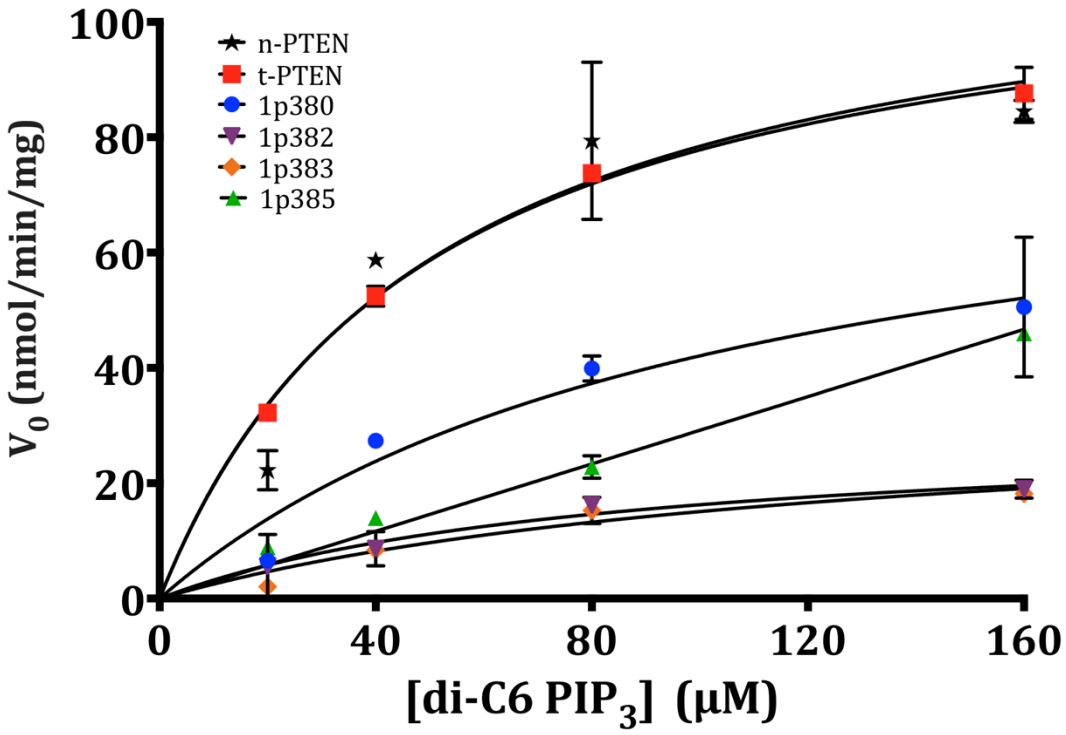


Figure 32. Enzymatic activity of semisynthetic PTEN. Activity of t-PTEN, n-PTEN, and monophosphorylated PTEN toward soluble di-C6-PIP_3 substrate. Values shown are averages of two replicates ± SEM.

Tail phosphate accessibility in semisynthetic phospho-PTENs

Based on the structural model that the 4p-PTEN tail phosphates are engaged with the PTEN C2 domain in the closed conformation, their accessibility to removal by a non-specific alkaline phosphatase is predicted to be hindered[24, 25] (Figure 33a). Indeed, prior studies are consistent with this prediction since denatured C379-4p-PTEN is much more susceptible to dephosphorylation by alkaline phosphatase compared with natively folded C379-4p-PTEN.[24] Moreover, in these earlier studies, mutations of the CBR3 loop in the C2 domain also rendered C379-4p-PTEN more sensitive to alkaline phosphatase-mediated dephosphorylation. Here, we compared alkaline-phosphatase mediated dephosphorylation of natively folded Y379-4p-PTEN to that of C379-4p-PTEN carried out under identical conditions. Detection of phosphorylation status of the PTEN forms was performed using a western blot with a polyclonal anti-4p-PTEN Ab. Since this Ab shows ~4-fold greater signal with Y379-4p-PTEN relative to C379-4p-PTEN, we loaded ~4-fold less Y379-4p-PTEN on SDS-PAGE to equalize blot intensities. Comparison of the pattern of dephosphorylation of C379-4p-PTEN versus Y379-4p-PTEN revealed that the latter had about a 10-fold greater half-life (Figure 33b). These data suggest that Y379-4p-PTEN does indeed reside in a more tightly closed conformation relative to C379-4p-PTEN as was also surmised from the PTEN enzymatic activity studies described above.

We proceeded to examine the sensitivity of Y379-1p380-PTEN, Y379-1p383-PTEN, and Y379-1p385-PTEN to treatment with alkaline phosphatase. All three constructs exhibited half-lives of 1-3 minutes (Figure 34 and Table 9), which is

indicative of a relatively weak interaction between the phosphorylated tail and body of PTEN, and a mostly open conformation. In contrast, under the same conditions C379-4p-PTEN was much more resistant to dephosphorylation, with a half-life of around 21 minutes (Figure 34).

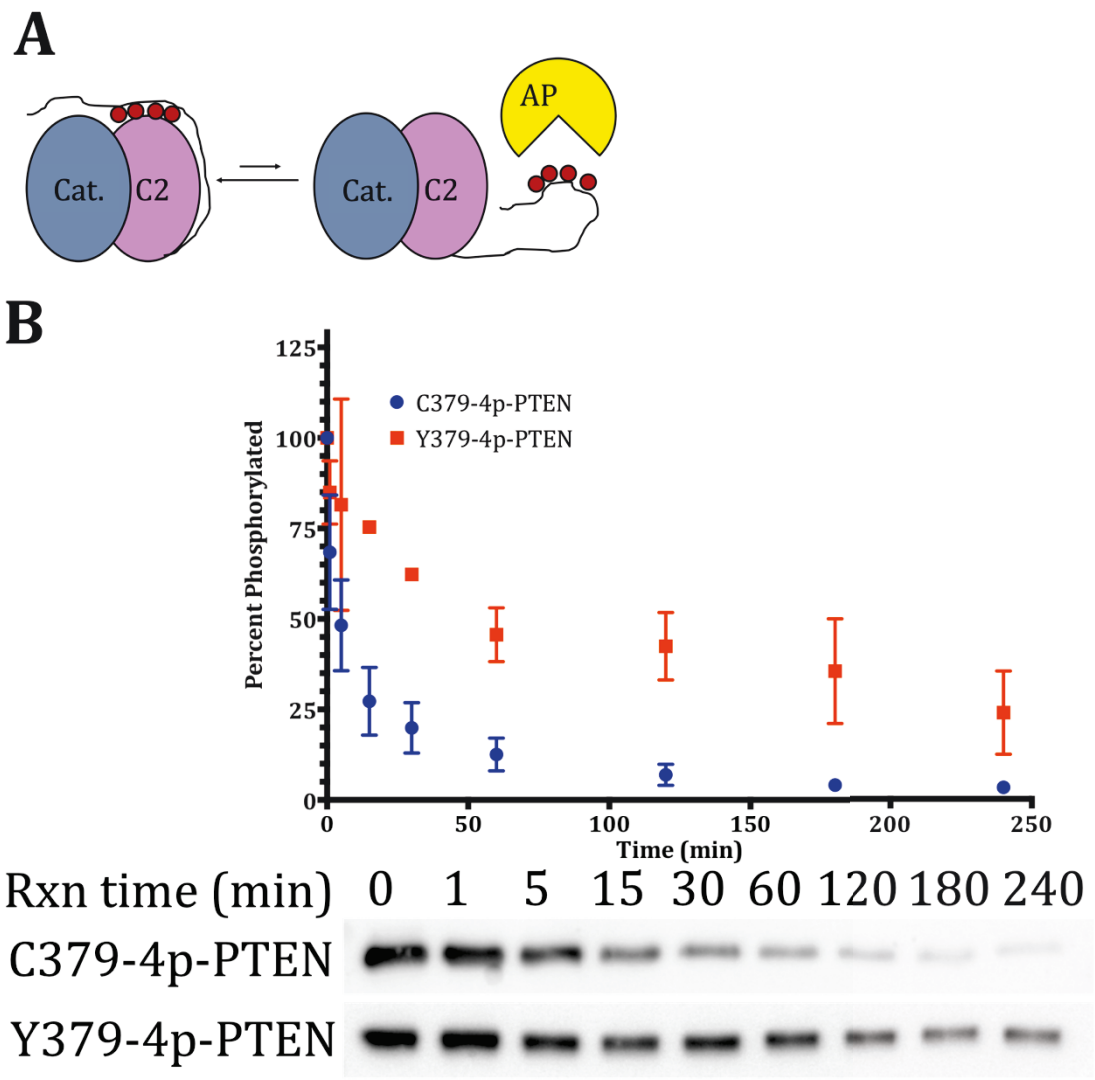


Figure 33. Sensitivity of 4p-PTEN to treatment with alkaline phosphatase. (**A**) Illustration of the open and closed conformations of PTEN and the change in accessibility of the phosphorylation sites to alkaline phosphatase. Red circles indicate phosphorylated residues. Cat., catalytic domain; C2, C2 domain; AP, alkaline phosphatase. (**B**) Y379C-4p-PTEN is more resistant to treatment with alkaline phosphatase than C379-4p-PTEN. Data are the average of two replicates for each PTEN construct. Representative Western blots show the decrease in phosphorylation with longer phosphatase treatment for both PTEN variants. C379-4p-PTEN half-life = 9.0 ± 2.2 min; Y379-4p-PTEN half-life = 112.1 ± 24.5 min.

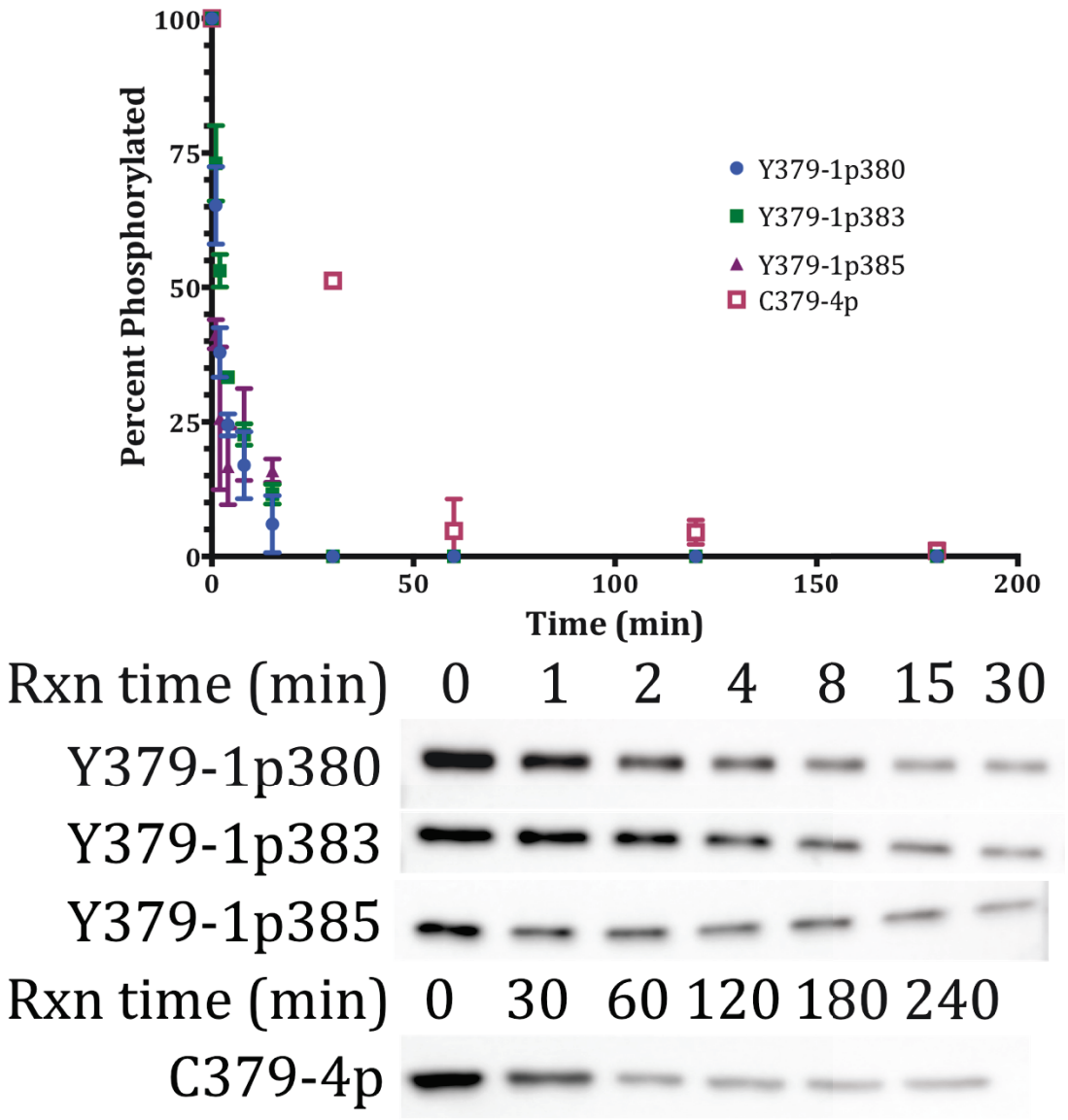


Figure 34. Sensitivity of 1p-PTEN constructs to treatment with alkaline phosphatase. Mono-phosphorylated Y379-1p380-, Y379-1p383-, and Y379-1p385-PTEN are all more sensitive to treatment with alkaline phosphatase than C379-4p-PTEN. Values shown are the averages of two duplicates ± SEM. Representative Western blots show the decrease in phosphorylation with longer phosphatase treatment. Y379-1p380 half-life = 1.7 ± 0.2 min; Y379-1p383 half-life = 2.7 ± 0.2 min; Y379-1p385 half-life = 0.8 ± 0.1 min; C379-4p half-life = 21.2 ± 3 min.

PTEN	k_{cat}/K_M	Tail phosphorylation half-life after treatment with alkaline phosphatase
	x 10^2 M^{-1} s^{-1}	min
t-PTEN	18 ± 2	n.d.
n-PTEN	18 ± 9	n.d.
1p380-PTEN	6 ± 5	1.7
1p382-PTEN	3 ± 2	n.d.
1p383-PTEN	2 ± 1	2.7
1p385-PTEN	2 ± 0.1	0.8

Table 9. Catalytic activity and tail phosphorylation sensitivity to alkaline phosphatase for semisynthetic PTEN constructs.

Discussion

While cysteine-mediated ligation had previously been used to produce Y379C phosphorylated PTEN constructs,[24, 25] it was later discovered that the introduction of a cellular mutant with an unnatural cysteine, or other residues, at this position promoted plasma membrane interaction.[117] There were two logical explanations for Y379C PTEN's cellular behavior. One possibility was that Y379C PTEN failed to undergo protein kinase-catalyzed phosphorylation in cells because of interference by Cys-379. The second possibility for Y379C PTEN's altered cellular localization was that normal CK2-mediated phosphorylation proceeded but Cys-379 antagonized the conformational closure driven by the phospho-cluster. Our biochemical studies with Y379-4p-PTEN and C379-4p-PTEN strongly favor the latter explanation. Both the reduced enzymatic activity and susceptibility to dephosphorylation of Y379-4p-PTEN relative to C379-4p-PTEN support the idea that Tyr379 is important in stabilizing the closed conformation of 4p-PTEN. One can imagine that this may be due to pi stacking or cation-pi interactions mediated by the tyrosine phenol side chain and the PTEN body. Y379-4p-PTEN's resistance to conformational opening at high ionic strength could indicate a conformation which is less driven by electrostatics relative to C379-4p-PTEN. It is interesting to note that a semisynthetic 4p-PTEN mutant, 4p-ePTEN, which was predicted to be in a more open conformation was not, despite its greater membrane association.[25] These results emphasize the importance of biochemical analysis of the individual post-translationally modified proteins in understanding cellular behaviors.

Previous data from studies with C379-phospho-PTEN suggested a correlation between catalytic activity and conformation (as measured by sensitivity to alkaline phosphatase).[24, 25] 3p- or 4p-PTEN was resistant to dephosphorylation by alkaline phosphatase, indicating a tightly closed conformation, and very inhibited. Constructs with fewer phosphorylations were less inhibited, and more susceptible to dephosphorylation. However, the data from Y379-phospho-PTEN that we show here tell a more complicated story. Certainly, Y379-4p-PTEN is more resistant to dephosphorylation and less active than C379-4p-PTEN, but this does not extend to all the monophosphorylated Y379-PTEN constructs. Y379-1p380-, 1p383- and 1p385-PTEN each had very similar, short half-lives, indicating a mostly open conformation, yet despite this similarity, Y379-1p383-, and 1p385-PTEN were more inhibited than Y379-1p380-PTEN.

This suggests that the PTEN activity assays and the alkaline phosphatase assays with monophosphorylated PTEN are not quite measuring the same phenomena. The alkaline phosphatase assay appears to measure the kinetic stability of the interaction between the tail and core of PTEN. More phosphates should form a slow off-rate interaction, and thus exhibit slower dephosphorylation. This is consistent with the 1p-PTEN data. Each construct has a single phosphorylation, and thus show very similar dephosphorylation profiles.

The activity assay, though, gives a different insight, perhaps one that is more thermodynamically driven than kinetically driven. The V versus [s] curve for 1p380-PTEN was different from the curves of 1p382-, and 1p383-PTEN, and these were in turn distinct from the curve for 1p385-PTEN. The data from 1p380-, 1p382-,

and 1p383-PTEN suggest that most of their inhibition relative to n-PTEN arises from a reduction in k_{cat}, as each only showed a modest increase in K_M. 1p385-PTEN showed a different pattern of inhibition, with an increase in both k_{cat}, and K_M, though its k_{cat}/K_M was similar to those of 1p-382- and 1p-383-PTEN. This suggests that there are three distinct interactions between the phosphorylated residues on the tail and the core of PTEN: One involving phospho-Ser-380, another involving phospho-Thr-382 and -383, and the third involving phospho-Ser-385. Cross-linking data[25] suggests that phospho-Ser-385 may be close to the active site of PTEN. If it in fact occupies the active site, or otherwise blocks PIP_3 from accessing the active site, this could explain the observed increase in K_M. When Thr-382/Thr-383 or Ser-385 bind to their respective sites, it could trigger a conformational shift in the active site such that catalysis is reduced, but substrate binding is less affected.

Together, the alkaline phosphatase and activity data suggest that the regulation of PTEN by phosphorylation at Ser-380/Thr-382/Thr-383/Ser-385 may not be a binary question of whether the tail is open or closed. Rather, the regulation may be mediated by specific interactions between tail phosphates and the core of PTEN. This may mean that a partially-phosphorylated PTEN may exhibit a more open conformation and be able to bind the plasma membrane, yet still have reduced catalytic activity. This also suggests that perhaps the observed difference in activity between Y379-4p-PTEN and C379-4p-PTEN is due to a difference in how each tail interacts with the core of PTEN, in addition to the overall stronger interaction seen with the Y379-containing tail.

Y379-4p-PTEN also serves as an accurate measuring stick to quantify levels of phosphorylation in cellular PTEN. Commercial antibody recognition of the phospho-cluster in C379-4p-PTEN is markedly impaired by the presence of the Cys-379, rendering this semisynthetic protein an inaccurate standard for western blots. In addition, we have shown that commercial polyclonal antibody against 3p-380,382,383 PTEN shows cross-reactivity with monophosphorylated 1p380, 1p383, and 1p385 PTEN. This suggests that caution must be taken with signaling analyses that rely on this reagent for Western blotting or immunocytochemistry.

As we have shown, wild type phosphorylated PTEN behaves differently from previously generated Y379C mutants. The monophosphorylated PTEN constructs suggest a more nuanced story of PTEN regulation by tail phosphorylation than was previously demonstrated. It would have been difficult to generate Y379-phospho-PTEN without relying on subtiligase-mediated expressed protein ligation. The relatively broad flexibility of amino acid sequences around the ligation junction, the simplicity of the protocol, and the speed of the process combine to make subtiligase-mediated expressed protein ligation an attractive technique for protein semisynthesis.

Bibliography

1. Qiao, Y., Molina, H., Pandey, A., Zhang, J. & Cole, P. A. Chemical rescue of a mutant enzyme in living cells. *Science* **311**, 1293-1297 (2006).
2. Chen, Z. & Cole, P. A. Synthetic approaches to protein phosphorylation. *Curr Opin Chem Biol* **28**, 115-122 (2015).
3. Hunter, T. Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling. *Cell* **80**, 225-236 (1995).
4. Siman, P. & Brik, A. Chemical and semisynthesis of posttranslationally modified proteins. *Org. Biomol. Chem.* **10**, 5684-5697 (2012).
5. Muir, T. W. Studying protein structure and function using semisynthesis. *Biopolymers* **90**, 743-750 (2008).
6. Zhang, Z., Shen, K., Lu, W. & Cole, P. A. The role of C-terminal tyrosine phosphorylation in the regulation of SHP-1 explored via expressed protein ligation. *J. Biol. Chem.* **278**, 4668-4674 (2003).
7. Lu, W., Shen, K. & Cole, P. A. Chemical dissection of the effects of tyrosine phosphorylation of SHP-2. *Biochemistry* **42**, 5461-5468 (2003).
8. Zheng, W. *et al*. Cellular stabilization of the melatonin rhythm enzyme induced by nonhydrolyzable phosphonate incorporation. *Nat. Struct. Biol.* **10**, 1054-1057 (2003).
9. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **266**, 776-779 (1994).
10. Jackson, D. Y. *et al*. A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **266**, 243-247 (1994).
11. Neumann-Staubitz, P. & Neumann, H. The use of unnatural amino acids to study and engineer protein function. *Curr. Opin. Struct. Biol.* **38**, 119-128 (2016).
12. Park, H. *et al*. Expanding the genetic code of Escherichia coli with phosphoserine. *Science* **333**, 1151-1154 (2011).
13. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4**, 232-234 (2008).
14. Hoffmann, C. & Neumann, H. In Vivo Mapping of FACT-Histone Interactions Identifies a Role of Pob3 C-terminus in H2A-H2B Binding. *ACS Chem. Biol.* **10**, 2753-2763 (2015).
15. Shiota, T. *et al*. Molecular architecture of the active mitochondrial protein gate. *Science* **349**, 1544-1548 (2015).
16. Luo, X. *et al*. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* (2017).
17. Kruize, E. in *Top 2000: volume 2* 1666-1676 (Veen, Amsterdam [u.a.], 2009).
18. M. Chalker, J. *et al*. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chemical Science* **2**, 1666-1676 (2011).
19. Serwa, R. *et al*. Chemoselective Staudinger-phosphite reaction of azides for the phosphorylation of proteins. *Angew. Chem. Int. Ed. Engl.* **48**, 8234-8239 (2009).
20. Nilsson, B. L., Kiessling, L. L. & Raines, R. T. Staudinger ligation: a peptide from a thioester and azide. *Org. Lett.* **2**, 1939-1941 (2000).

21. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6705-6710 (1998).
22. Erlanson, D. A., Chytil, M. & Verdine, G. L. The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem. Biol.* **3**, 981-991 (1996).
23. Wang, Y. *et al*. Regulation of S-adenosylhomocysteine hydrolase by lysine acetylation. *The Journal of biological chemistry* **289**, 31361-31372 (2014).
24. Bolduc, D. *et al*. Phosphorylation-mediated PTEN conformational closure and deactivation revealed with protein semisynthesis. *eLife* **2**, e00691 (2013).
25. Chen, Z. *et al*. Molecular Features of Phosphatase and Tensin Homolog (PTEN) Regulation by C-terminal Phosphorylation. *J. Biol. Chem.* **291**, 14160-14169 (2016).
26. Schwarzer, D. & Cole, P. A. Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr Opin Chem Biol* **9**, 561-569 (2005).
27. Ghosh, I. *et al*. Site-specific protein labeling by intein-mediated protein ligation. *Methods Mol. Biol.* **705**, 87-107 (2011).
28. Ludwig, C. *et al*. Semisynthesis of Proteins Using Split Inteins. *Methods in Enzymology* **462**, 77-96 (2009).
29. Raghavan, R. & Minnick, M. F. Group I introns and inteins: disparate origins but convergent parasitic strategies. *J. Bacteriol.* **191**, 6193-6202 (2009).
30. Chong, S. *et al*. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271-281 (1997).
31. Tarrant, M. K. *et al*. Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. *Nat. Chem. Biol.* **8**, 262-269 (2012).
32. Rawlings, N. D. & Barrett, A. J. Evolutionary families of peptidases. *The Biochemical journal* **290 (Pt 1)**, 205-218 (1993).
33. Masao Fujinaga *et al*. The Molecular Structure and Catalytic Mechanism of a Novel Carboxyl Peptidase from Scytalidium lignicolum. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3364-3369 (2004).
34. Brannigan, J. A. *et al*. A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378**, 416-419 (1995).
35. Rawlings, N. D., Barrett, A. J. & Bateman, A. Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *The Journal of biological chemistry* **286**, 38321-38328 (2011).
36. Polgár, L. The catalytic triad of serine peptidases. *Cell. Mol. Life Sci.* **62**, 2161-2172 (2005).
37. Nelson, D. L., Cox, M. M. & Lehninger, A. L. in *Lehninger principles of biochemistry,* 2017).
38. Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* **216**, 413-417 (1994).
39. Narayanan, S. Multifunctional roles of thrombin. *Ann. Clin. Lab. Sci.* **29**, 275-280 (1999).

40. Janiak, M. C. Digestive enzymes of human and nonhuman primates. *Evol. Anthropol.* **25**, 253-266 (2016).
41. Julien, O. & Wells, J. A. Caspases and their substrates. *Cell Death Differ.* **24**, 1380-1389 (2017).
42. Hough, R., Pratt, G. & Rechsteiner, M. Purification of two high molecular weight proteases from rabbit reticulocyte lysate. *J. Biol. Chem.* **262**, 8303-8313 (1987).
43. Morihara, K., Tsuzuki, H. & Oka, T. Comparison of the specificities of various neutral proteinases from microorganisms. *Archives of Biochemistry and Biophysics* **123**, 572-588 (1968).
44. Morihara, K. & Tsuzuki, H. Comparison of the specificities of various serine proteinases from microorganisms. *Archives of Biochemistry and Biophysics* **129**, 620-634 (1969).
45. Abrahmsén, L. *et al*. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **30**, 4151-4159 (1991).
46. Atwell, S. & Wells, J. A. Selection for Improved Subtiligases by Phage Display. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9497-9502 (1999).
47. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. Recruitment of Substrate-Specificity Properties from One Enzyme into a Related One by Protein Engineering. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 5167-5171 (1987).
48. Polgar, L. & Bender, M. L. A New Enzyme Containing a Synthetically Formed Active Site, Thiol-Subtilisin. *J. Am. Chem. Soc.* **88**, 3153-3154 (1966).
49. Polgar, L. The reactivity of thiol-subtilisin, an enzyme containing a synthetic functional group. *BIOCHEMISTRY US* **6**, 610-620 (1967).
50. Nakatsuka, T., Sasaki, T. & Kaiser, E. T. Peptide Segment Coupling Catalyzed by the Semisynthetic Enzyme Thiolsubtilisin. *Journal of the American Chemical Society* **109**, 3808-3810 (1987).
51. Chu, S. & Mautner, H. G. Neighboring-Group Effects in the Reactions of Esters, Thioesters, and Selenoesters. The Hydrolysis and Aminolysis of Benzoylcholine, Benzoylthiolcholine, Benzoylselenolcholine, and of their Dimethylamino Analogs. *J. Org. Chem.* **3**, 308-312 (1966).
52. Wu, Z. P. & Hilvert, D. Conversion of a protease into an acyl transferase: selenolsubtilisin. *J AMER CHEM SOC* **111**, 4513-4514 (1989).
53. Chang, T. K., Jackson, D. Y., Burnier, J. P. & Wells, J. A. Subtiligase: a tool for semisynthesis of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12544-12548 (1994).
54. Braisted, A. C., Judice, J. K. & Wells, J. A. Synthesis of proteins by subtiligase. *Meth. Enzymol.* **289**, 298-313 (1997).
55. Takeuchi, Y., Satow, Y., Nakamura, K. T. & Mitsui, Y. Refined crystal structure of the complex of subtilisin BPN' and Streptomyces subtilisin inhibitor at 1.8 A resolution. *J. Mol. Biol.* **221**, 309-325 (1991).
56. Yoshihara, H. A. I., Mahrus, S. & Wells, J. A. Tags for labeling protein N-termini with subtiligase for proteomics. *Bioorg. Med. Chem. Lett.* **18**, 6000-6003 (2008).
57. Wildes, D. & Wells, J. A. Sampling the N-terminal proteome of human blood. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4561-4566 (2010).

58. Sefton, B. M. & Hunter, T. in *Protein phosphorylation* (Academic Press, San Diego, Calif., 1998).
59. Chen, M. J., Dixon, J. E. & Manning, G. Genomics and evolution of protein phosphatases. *Sci Signal* **10** (2017).
60. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).
61. Burke, T. R. & Zhang, Z. Y. Protein-tyrosine phosphatases: structure, mechanism, and inhibitor discovery. *Biopolymers* **47**, 225-241 (1998).
62. Vara, J. Á F. *et al*. PI3K/Akt signalling pathway and cancer. *Cancer Treatment Reviews* **30**, 193-204 (2004).
63. Courtney, K. D., Corcoran, R. B. & Engelman, J. A. The PI3K pathway as drug target in human cancer. *J. Clin. Oncol.* **28**, 1075-1083 (2010).
64. Chalhoub, N. & Baker, S. J. PTEN and the PI3-kinase pathway in cancer. *Annu Rev Pathol* **4**, 127-150 (2009).
65. Carnero, A., Blanco-Aparicio, C., Renner, O., Link, W. & Leal, J. F. M. The PTEN/PI3K/AKT signalling pathway in cancer, therapeutic implications. *Curr Cancer Drug Targets* **8**, 187-198 (2008).
66. Carpenter, C. L. *et al*. Phosphoinositide 3-kinase is activated by phosphopeptides that bind to the SH2 domains of the 85-kDa subunit. *J. Biol. Chem.* **268**, 9478-9483 (1993).
67. Foster, F. M., Traer, C. J., Abraham, S. M. & Fry, M. J. The phosphoinositide (PI) 3-kinase family. *J. Cell. Sci.* **116**, 3037-3040 (2003).
68. Vanhaesebroeck, B. *et al*. Synthesis and function of 3-phosphorylated inositol lipids. *Annu. Rev. Biochem.* **70**, 535-602 (2001).
69. Cross, D. A., Alessi, D. R., Cohen, P., Andjelkovich, M. & Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785-789 (1995).
70. Brunet, A. *et al*. Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857-868 (1999).
71. Inoki, K., Li, Y., Zhu, T., Wu, J. & Guan, K. TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nat. Cell Biol.* **4**, 648-657 (2002).
72. Manning, B. D. & Cantley, L. C. AKT/PKB signaling: navigating downstream. *Cell* **129**, 1261-1274 (2007).
73. Manning, B. D. & Toker, A. AKT/PKB Signaling: Navigating the Network. *Cell* **169**, 381-405 (2017).
74. Wise, H. M., Hermida, M. A. & Leslie, N. R. Prostate cancer, PI3K, PTEN and prognosis. *Clin. Sci.* **131**, 197-210 (2017).
75. Hollander, M. C., Blumenthal, G. M. & Dennis, P. A. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nat. Rev. Cancer* **11**, 289-301 (2011).
76. Steck, P. A. *et al*. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat. Genet.* **15**, 356-362 (1997).
77. Maehama, T. & Dixon, J. E. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **273**, 13375-13378 (1998).

78. Myers, M. P. *et al*. The lipid phosphatase activity of PTEN is critical for its tumor supressor function. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13513-13518 (1998).
79. Li, J. *et al*. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**, 1943-1947 (1997).
80. Li, D. M. & Sun, H. TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. *Cancer Res.* **57**, 2124-2129 (1997).
81. Liaw, D. *et al*. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* **16**, 64-67 (1997).
82. Alimonti, A. *et al*. Subtle variations in Pten dose determine cancer susceptibility. *Nat. Genet.* **42**, 454-458 (2010).
83. Myers, M. P. *et al*. P-TEN, the Tumor Suppressor from Human Chromosome 10q23, is a Dual-Specificity Phosphatase. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9052-9057 (1997).
84. Sano, T. *et al*. Differential expression of MMAC/PTEN in glioblastoma multiforme: relationship to localization and prognosis. *Cancer Res.* **59**, 1820-1824 (1999).
85. Gong, L. *et al*. Nuclear PTEN tumor-suppressor functions through maintaining heterochromatin structure. *Cell Cycle* **14**, 2323-2332 (2015).
86. Nosaka, R. *et al*. Role for loss of nuclear PTEN in a harbinger of brain metastases. *J Clin Neurosci* (2017).
87. Nguyen, H. -. *et al*. A new class of cancer-associated PTEN mutations defined by membrane translocation defects. *Oncogene* **34**, 3737-3743 (2015).
88. Shen, W. H. *et al*. Essential role for nuclear PTEN in maintaining chromosomal integrity. *Cell* **128**, 157-170 (2007).
89. He, J. *et al*. PTEN regulates EG5 to control spindle architecture and chromosome congression during mitosis. *Nat Commun* **7**, 12355 (2016).
90. Zhang, Z. *et al*. PTEN regulates PLK1 and controls chromosomal stability during cell division. *Cell Cycle* **15**, 2476-2485 (2016).
91. Chang, C. *et al*. PTEN nuclear localization is regulated by oxidative stress and mediates p53-dependent tumor suppression. *Mol. Cell. Biol.* **28**, 3281-3289 (2008).
92. Song, M. S. *et al*. Nuclear PTEN regulates the APC-CDH1 tumor-suppressive complex in a phosphatase-independent manner. *Cell* **144**, 187-199 (2011).
93. Choi, B. H., Xie, S. & Dai, W. PTEN is a negative regulator of mitotic checkpoint complex during the cell cycle. *Exp Hematol Oncol* **6**, 19 (2017).
94. Whelan, J. T., Forbes, S. L. & Bertrand, F. E. CBF-1 (RBP-J kappa) binds to the PTEN promoter and regulates PTEN gene expression. *Cell Cycle* **6**, 80-84 (2007).
95. Stambolic, V. *et al*. Regulation of PTEN transcription by p53. *Mol. Cell* **8**, 317-325 (2001).
96. Escrivà, M. *et al*. Repression of PTEN phosphatase by Snail1 transcriptional factor during gamma radiation-induced apoptosis. *Mol. Cell. Biol.* **28**, 1528-1540 (2008).
97. Meng, F. *et al*. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* **133**, 647-658 (2007).

98. Salvesen, H. B. *et al*. PTEN methylation is associated with advanced stage and microsatellite instability in endometrial carcinoma. *Int. J. Cancer* **91**, 22-26 (2001).
99. Lee, J. O. *et al*. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell* **99**, 323-334 (1999).
100. Iijima, M., Huang, Y. E., Luo, H. R., Vazquez, F. & Devreotes, P. N. Novel mechanism of PTEN regulation by its phosphatidylinositol 4,5-bisphosphate binding motif is critical for chemotaxis. *J. Biol. Chem.* **279**, 16606-16613 (2004).
101. Iijima, M. & Devreotes, P. Tumor suppressor PTEN mediates sensing of chemoattractant gradients. *Cell* **109**, 599-610 (2002).
102. Al-Khouri, A. M., Ma, Y., Togo, S. H., Williams, S. & Mustelin, T. Cooperative phosphorylation of the tumor suppressor phosphatase and tensin homologue (PTEN) by casein kinases and glycogen synthase kinase 3beta. *J. Biol. Chem.* **280**, 35195-35202 (2005).
103. Miller, S. J., Lou, D. Y., Seldin, D. C., Lane, W. S. & Neel, B. G. Direct identification of PTEN phosphorylation sites. *FEBS Lett.* **528**, 145-153 (2002).
104. Zhang, X. *et al*. Hypo-CpG methylation controls PTEN expression and cell apoptosis in irradiated lung. *Free Radic. Res.* **50**, 875-886 (2016).
105. Salvesen, H. B. *et al*. PTEN methylation is associated with advanced stage and microsatellite instability in endometrial carcinoma. *Int. J. Cancer* **91**, 22-26 (2001).
106. Goel, A. *et al*. Frequent inactivation of PTEN by promoter hypermethylation in microsatellite instability-high sporadic colorectal cancers. *Cancer Res.* **64**, 3014-3021 (2004).
107. Wang, Q., Zhou, Y., Wang, X., Chung, D. H. & Evers, B. M. Regulation of PTEN expression in intestinal epithelial cells by c-Jun NH2-terminal kinase activation and nuclear factor-kappaB inhibition. *Cancer Res.* **67**, 7773-7781 (2007).
108. Hettinger, K. *et al*. c-Jun promotes cellular survival by suppression of PTEN. *Cell Death Differ.* **14**, 218-229 (2007).
109. Cordier, F. *et al*. Ordered phosphorylation events in two independent cascades of the PTEN C-tail revealed by NMR. *J. Am. Chem. Soc.* **134**, 20533-20543 (2012).
110. Rahdar, M. *et al*. A phosphorylation-dependent intramolecular interaction regulates the membrane association and activity of the tumor suppressor PTEN. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 480-485 (2009).
111. Okumura, K. *et al*. PCAF modulates PTEN activity. *J. Biol. Chem.* **281**, 26562-26568 (2006).
112. Ikenoue, T., Inoki, K., Zhao, B. & Guan, K. PTEN acetylation modulates its interaction with PDZ domain. *Cancer Res.* **68**, 6908-6912 (2008).
113. González-Santamaría, J. *et al*. Regulation of the tumor suppressor PTEN by SUMO. *Cell Death Dis* **3**, e393 (2012).
114. Wang, X. *et al*. NEDD4-1 is a proto-oncogenic ubiquitin ligase for PTEN. *Cell* **128**, 129-139 (2007).
115. Maddika, S. *et al*. WWP2 is an E3 ubiquitin ligase for PTEN. *Nat. Cell Biol.* **13**, 728-733 (2011).

116. Trotman, L. C. *et al*. Ubiquitination regulates PTEN nuclear import and tumor suppression. *Cell* **128**, 141-156 (2007).
117. Nguyen, H. N. *et al*. Mechanism of human PTEN localization revealed by heterologous expression in Dictyostelium. *Oncogene* **33**, 5688-5696 (2014).
118. Lee, S. *et al*. Reversible inactivation of the tumor suppressor PTEN by H2O2. *J. Biol. Chem.* **277**, 20336-20342 (2002).
119. McConnachie, G., Pass, I., Walker, S. M. & Downes, C. P. Interfacial kinetic analysis of the tumour suppressor phosphatase, PTEN: evidence for activation by anionic phospholipids. *Biochem. J.* **371**, 947-955 (2003).
120. Chagpar, R. B. *et al*. Direct positive regulation of PTEN by the p85 subunit of phosphatidylinositol 3-kinase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5471-5476 (2010).
121. Wu, X. *et al*. Evidence for regulation of the PTEN tumor suppressor by a membrane-localized multi-PDZ domain containing scaffold protein MAGI-2. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4233-4238 (2000).
122. Sumitomo, M. *et al*. Synergy in tumor suppression by direct interaction of neutral endopeptidase with PTEN. *Cancer Cell* **5**, 67-78 (2004).
123. Vazquez, F., Ramaswamy, S., Nakamura, N. & Sellers, W. R. Phosphorylation of the PTEN tail regulates protein stability and function. *Mol. Cell. Biol.* **20**, 5010-5018 (2000).
124. Li, Z. *et al*. Regulation of PTEN by Rho small GTPases. *Nat. Cell Biol.* **7**, 399-404 (2005).
125. Vemula, S., Shi, J., Hanneman, P., Wei, L. & Kapur, R. ROCK1 functions as a suppressor of inflammatory cell migration by regulating PTEN phosphorylation and stability. *Blood* **115**, 1785-1796 (2010).
126. Yim, E. *et al*. Rak functions as a tumor suppressor by regulating PTEN protein stability and function. *Cancer Cell* **15**, 304-314 (2009).
127. Chen, Z. *et al*. Enzymatic Analysis of PTEN Ubiquitylation by WWP2 and NEDD4-1 E3 Ligases. *Biochemistry* **55**, 3658-3666 (2016).
128. Al-Khouri, A. M., Ma, Y., Togo, S. H., Williams, S. & Mustelin, T. Cooperative phosphorylation of the tumor suppressor phosphatase and tensin homologue (PTEN) by casein kinases and glycogen synthase kinase 3beta. *J. Biol. Chem.* **280**, 35195-35202 (2005).
129. Maccario, H., Perera, N. M., Davidson, L., Downes, C. P. & Leslie, N. R. PTEN is destabilized by phosphorylation on Thr366. *Biochem. J.* **405**, 439-444 (2007).
130. Vazquez, F. *et al*. Phosphorylation of the PTEN tail acts as an inhibitory switch by preventing its recruitment into a protein complex. *J. Biol. Chem.* **276**, 48627-48630 (2001).
131. Torres, J. & Pulido, R. The tumor suppressor PTEN is phosphorylated by the protein kinase CK2 at its C terminus. Implications for PTEN stability to proteasome-mediated degradation. *J. Biol. Chem.* **276**, 993-998 (2001).
132. Ning, K. *et al*. Leptin-dependent phosphorylation of PTEN mediates actin restructuring and activation of ATP-sensitive K+ channels. *J. Biol. Chem.* **284**, 9331-9340 (2009).
133. Moult, P. R. *et al*. Leptin regulates AMPA receptor trafficking via PTEN inhibition. *J. Neurosci.* **30**, 4088-4101 (2010).

134. Tellios, N. *et al*. TGF-β induces phosphorylation of phosphatase and tensin homolog: implications for fibrosis of the trabecular meshwork tissue in glaucoma. *Sci Rep* **7**, 812 (2017).
135. Zhang, X. C., Piccini, A., Myers, M. P., Van Aelst, L. & Tonks, N. K. Functional analysis of the protein phosphatase activity of PTEN. *Biochem. J.* **444**, 457-464 (2012).
136. Georgescu, M. M. *et al*. Stabilization and productive positioning roles of the C2 domain of PTEN tumor suppressor. *Cancer Res.* **60**, 7033-7038 (2000).
137. Das, S., Dixon, J. E. & Cho, W. Membrane-binding and activation mechanism of PTEN. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7491-7496 (2003).
138. Reeck, G. R. & Fisher, L. A statistical analysis of the amino acid compositions of proteins. *Int. J. Pept. Protein Res.* **5**, 109-117 (1973).
139. Jungck, J. R. The genetic code as a periodic table. *J. Mol. Evol.* **11**, 211-224 (1978).
140. Jukes, T. H., Holmquist, R. & Moise, H. Amino acid composition of proteins: Selection against the genetic code. *Science* **189**, 50-51 (1975).
141. Sun, S. B., Schultz, P. G. & Kim, C. H. Therapeutic applications of an expanded genetic code. *Chembiochem* **15**, 1721-1729 (2014).
142. Vila-Perelló, M. & Muir, T. W. Biological applications of protein splicing. *Cell* **143**, 191-200 (2010).
143. Dirksen, A. & Dawson, P. E. Expanding the scope of chemoselective peptide ligations in chemical biology. *Curr Opin Chem Biol* **12**, 760-766 (2008).
144. Kent, S. B. H. The critical role of peptide chemistry in the life sciences. *J. Pept. Sci.* **21**, 136-138 (2015).
145. Shah, N. H. & Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chem Sci* **5**, 446-461 (2014).
146. Schmohl, L. & Schwarzer, D. Sortase-mediated ligations for the site-specific modification of proteins. *Curr Opin Chem Biol* **22**, 122-128 (2014).
147. Liu, W. R., Wang, Y. & Wan, W. Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach. *Mol Biosyst* **7**, 38-47 (2011).
148. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413-444 (2010).
149. Evans, T. C., Benner, J. & Xu, M. Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* **7**, 2256-2264 (1998).
150. Ma, J., Zeng, J. & Wan, Q. Postligation-desulfurization: a general approach for chemical protein synthesis. *Top Curr Chem* **363**, 57-101 (2015).
151. Loibl, S. F., Harpaz, Z. & Seitz, O. A Type of Auxiliary for Native Chemical Peptide Ligation beyond Cysteine and Glycine Junctions. *Angew. Chem. Int. Ed. Engl.* **54**, 15055-15059 (2015).
152. Welker, E. & Scheraga, H. A. Use of benzyl mercaptan for direct preparation of long polypeptide benzylthio esters as substrates of subtiligase. *Biochem. Biophys. Res. Commun.* **254**, 147-151 (1999).
153. Pantoliano, M. W. *et al*. Large increases in general stability for subtilisin BPN' through incremental changes in the free energy of unfolding. *Biochemistry* **28**, 7205-7213 (1989).

154. McPhalen, C. A. & James, M. N. Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* **27**, 6582-6598 (1988).
155. Gallagher, T., Bryan, P. & Gilliland, G. L. Calcium-independent subtilisin by design. *Proteins* **16**, 205-213 (1993).
156. Masson, G. R., Perisic, O., Burke, J. E. & Williams, R. L. The intrinsically disordered tails of PTEN and PTEN-L have distinct roles in regulating substrate specificity and membrane activity. *Biochem. J.* **473**, 135-144 (2016).
157. Van Veldhoven, P. P. & Mannaerts, G. P. Inorganic and organic phosphate measurements in the nanomolar range. *Anal. Biochem.* **161**, 45-48 (1987).
158. Sarno, S. *et al*. Selectivity of 4,5,6,7-tetrabromobenzotriazole, an ATP site-directed inhibitor of protein kinase CK2 ('casein kinase-2'). *FEBS Lett.* **496**, 44-48 (2001).

Curriculum Vitae

Samuel Henager
1214 N Charles St, Baltimore, MD, 21201
shenage1@jhu.edu

Education

Johns Hopkins University, Baltimore, MD **August 2012 – September 2017**
Doctor of Philosophy (candidate): Chemical Biology
Advisor: Dr. Philip Cole, MD, PhD
Dissertation: Enzyme-Catalyzed Expressed Protein Ligation

Seattle Pacific University, Seattle, WA **September 2008 – June 2012**
Bachelor of Science: Biochemistry
Magna Cum Laude

Research and Policy Experience

Johns Hopkins University, Baltimore, MD

- Department of Pharmacology and Molecular Sciences, Lab of Dr. Philip Cole
  Graduate Research Assistant June 2013 – August 2017
  - Developed the enzyme subtiligase as a tool to create semi-synthetic proteins to study protein modifications
  - Created wild-type, modified PTEN, a tumor-suppressor protein often mutated in cancer and studied how modifications to PTEN affect its structure and activity

Federation of American Societies for Experimental Biology (FASEB), Bethesda, MD

- Office of Public Affairs
  Science Policy Fellow September – December 2016
  - Attended and drafted summaries and blog posts of NSF, NIH, and HHS advisory committee meetings for FASEB member societies and scientists
  - Wrote articles for lay audiences on topics related to animals in research
  - Created the first comprehensive national database of model organism stock centers and repositories (faseb.org/Science-Policy-and-Advocacy/Science-Policy-and-Research-Issues/Shared-Research-Resources/Stock-Center-Database.aspx)
  - Collaborated with FASEB policy analysts to draft policy positions related to the use of interim research products (e.g. preprints) in NIH grant applications

Seattle Pacific University, Seattle, WA

- Department of Chemistry and Biochemistry, Lab of Dr. Benjamin McFarland
  Undergraduate Research assistant June 2011 – June 2012
  - Predicted protein mutations using RosettaDesign software to increase the interaction between two proteins

- Studied how predicted mutations affected the interaction between two proteins, and how well software predictions correlated with experimental results

Pacific Northwest National Laboratory, Richland, WA

- Glass Development Laboratory, Dr. Pavel Hrma
  Research Technician II June 2009 – August 2010
  - Studied how silica particle size, nitrogen content, and carbon content affect the expansion of glass during the melting process to determine the optimum composition of high-level nuclear waste glass.

Publications

Henager, S. H., Chu, N., Chen, Z., Bolduc, D., Dempsey, D. R., Hwang, Y. Wells, J. A. and Cole, P. A. Enzyme-catalyzed expressed protein ligation. *Nat. Methods* **13**, 925-927 (2016).

Henager, S. H., Hale, M. A., Maurice, N. J., Dunnington, E. C., Swanson, C. J., Peters, M. J., Ban, J. J., Culpepper, D. J., Davies, L. D., Sanders, L. K., and McFarland, B. J. Combining different design strategies for rational affinity maturation of the MICA-NKG2D interface. *Protein Science* **21**, 1396-1402 (2012).

McFarland B. J., Hale M. A., and **Henager S. H**. Protein-Protein Interactions Rationally Redesigned by Undergraduates: Four Strategies for Enhancing MICA-NKG2D Affinity. 2012. Protein Science **21**, 182-182

Henager S. H., Hrma P. R., Swearingen K. J., Schweiger M. J., Marcial J., TeGrotenhuis N. E.. Conversion of batch to molten glass, I: Volume expansion. 2010. J. Non-Crystalline Solids, **357:3**; 829-835.

Hrma P. R., Marcial J., Swearingen K. J., **Henager S. H.**, Schweiger M. J., TeGrotenhuis N. E. Conversion of batch to molten glass, II: Dissolution of quartz particles. 2010. J. Non-Crystalline Solids, **357:3**; 820-828.

Professional Awards

Department of Pharmacology and Molecular Sciences Scheinberg travel award – 2016

References

Philip A. Cole, M.D., Ph.D. – Graduate Advisor
Director, and E.K. Marshall and Thomas H. Maren Professor of Pharmacology
Johns Hopkins University School of Medicine
pcole@jhmi.edu • 410.614.8849

Anne M. Deschamps, Ph.D. – Supervisor at FASEB
Senior Science Policy Analyst
Federation of American Societies for Experimental Biology
adeschamps@faseb.org • 301.634.7650

Benjamin J. McFarland, Ph.D. – Undergraduate Advisor
Professor of Biochemistry; Chair of Department of Chemistry and Biochemistry
Seattle Pacific University
bjm@spu.edu • 206.281.2749