CHEMICAL REVIEWS

Review

Cite This: *Chem. Rev.* 2020, 120, 3127−3160

pubs.acs.org/CR

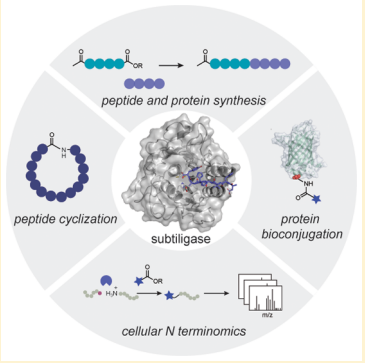
Subtiligase-Catalyzed Peptide Ligation

Amy M. Weeks*,† and James A. Wells*,†,‡

†Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94143, United States

‡Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94143, United States

ABSTRACT: Subtiligase-catalyzed peptide ligation is a powerful approach for site-specific protein bioconjugation, synthesis and semisynthesis of proteins and peptides, and chemoproteomic analysis of cellular N termini. Here, we provide a comprehensive review of the subtiligase technology, including its development, applications, and impacts on protein science. We highlight key advantages and limitations of the tool and compare it to other peptide ligase enzymes. Finally, we provide a perspective on future applications and challenges and how they may be addressed.



CONTENTS

Special Issue: Peptide Chemistry

Received: June 10, 2019
Published: October 30, 2019

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

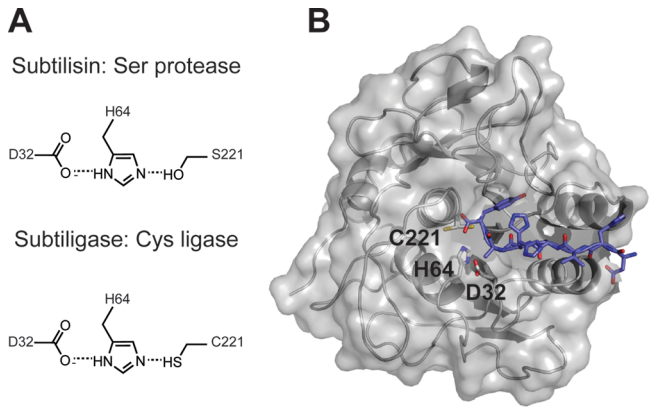


Figure 1. Subtiligase, a rationally designed peptide ligase. (A) Subtiligase was designed starting from subtilisin protease, which uses an Asp-His-Ser catalytic triad. In subtiligase, the catalytic Ser was mutated to Cys (S221C) and a second Pro-to-Ala (P225A) mutation was introduced to better accommodate the larger Cys nucleophile.

1. INTRODUCTION

The modification of proteins and peptides with chemical functionalities not encompassed by the genetically encoded amino acids is vital to advance both research in the biological sciences and the industrial production of protein therapeutics. The ability to site-specifically modify proteins through the formation of native peptide bonds while preserving natural protein sequence, aside from the introduced modification, is of particular interest. Recombinant protein expression in combination with modern molecular biology techniques allows production of full-length proteins but is limited to the incorporation of the natural amino acids and a small number of unnatural amino acids. On the other hand, solid-phase peptide synthesis methods provide access to the full spectrum of chemical functional groups but are generally limited to proteins less than 100 amino acids in length. Enzymatic catalysts for specific peptide ligation combine the advantages of both of these strategies, bridging the gap between recombinant protein expression and total chemical synthesis.

Subtiligase is a peptide ligase that was rationally designed from the broad-specificity serine protease subtilisin BPN′ from *Bacillus amyloliquefaciens* (Figure 1).[1,2] In the parent protease subtilisin, peptide bond hydrolysis strongly predominates over the reverse peptide ligation reaction (Figure 2A). However, subtiligase harbors two key mutations that enable it to efficiently catalyze a ligation reaction between a peptide ester acyl donor and the N-terminal α-amine of a peptide or protein (Figure 2B). Mutation of the catalytic Ser to Cys (S221C) to generate thiolsubtilisin reduces amidase activity to a negligible level but maintains the esterase activity observed in subtilisin, leading to formation of a thioacyl-enzyme intermediate from a peptide ester.[3,4] Introduction of a second Pro to Ala mutation (P225A) to produce subtiligase enhances peptide ligation activity by 2 orders of magnitude compared to thiolsubtilisin, resulting in an efficient peptide ligation catalyst.[2] Importantly, subtiligase-catalyzed peptide ligation occurs with absolute chemoselectivity for N-terminal α-amines over lysine ε-amines and is applicable to fully unprotected peptides and proteins under mild, aqueous conditions. Based on its ability to catalyze peptide bond formation, subtiligase has been applied broadly as a tool in chemistry and biology, enabling peptide and protein synthesis and semisynthesis, site-specific protein bioconjugation, and the chemoproteomic study of cellular N termini. Subtiligase has also been applied in combination with related protein synthesis and bioconjugation technologies to broaden their scope and ease of use.

Subtiligase is highly complementary to other enzymatic technologies that enable ligation of peptides and proteins. While other natural peptide ligase enzymes, such as sortase A[5] and butelase 1,[6] are generally highly sequence specific at one or more positions near the ligation site, subtiligase retains the broad specificity of subtilisin. Subtiligase is therefore applicable in situations where sequence flexibility is desirable, such as peptide ligation with native junctions or chemoproteomic profiling of cellular N termini. In contrast, high-specificity peptide ligases are useful for modification of specific protein sequences in complex protein mixtures or inside living cells, applications in which subtiligase might catalyze many off-target modifications.

This review is focused on the development, application, and impact of the subtiligase technology. We first discuss the rationale of the subtiligase design and advances in protein engineering of subtilisin that enabled its implementation. We next cover practical considerations for the application of subtiligase, including its substrate selectivity, the development and selection of specificity mutants to expand its substrate scope, and strategies for the synthesis of subtiligase substrates. We then review applications of subtiligase to address important chemical problems and biological questions. These include peptide and protein synthesis, site-specific protein bioconjugation, and cellular N terminomics studies. We compare subtiligase-catalyzed peptide ligation to alternative enzymatic peptide ligation strategies, which have unique advantages and limitations. Finally, we present a perspective on future applications and remaining challenges for the use of subtiligase to catalyze peptide bond formation.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

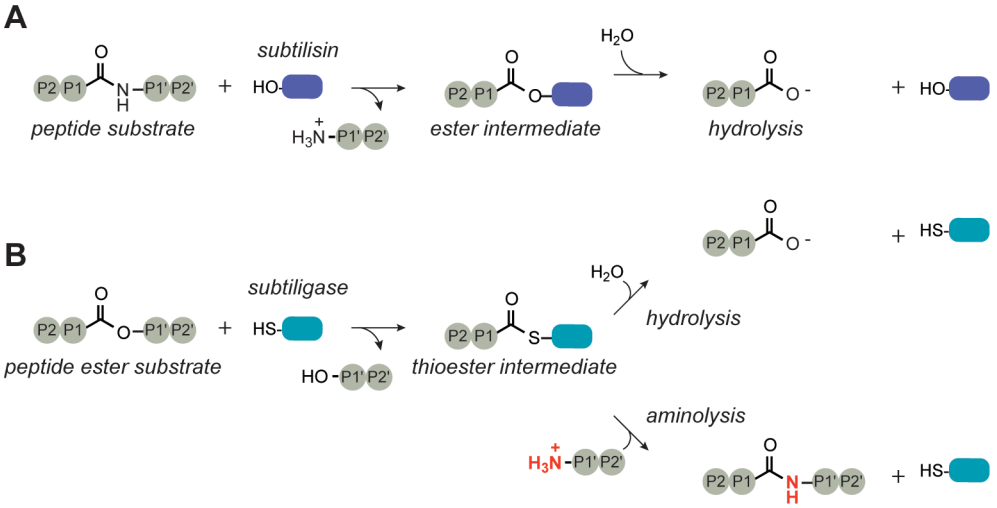


Figure 2. Comparison of the reactions catalyzed by subtilisin and subtiligase. (A) Peptide hydrolysis reaction catalyzed by subtilisin. A peptide bond is attacked by the catalytic serine to form an acyl-enzyme intermediate, which is hydrolyzed. (B) Peptide ligation reaction catalyzed by subtiligase. A peptide ester substrate is attacked by the catalytic Cys to form a thioacyl-enzyme intermediate, which can either be hydrolyzed or intercepted by the α-amine from a peptide or protein nucleophile to form a peptide bond.

2. USING PROTEASES IN REVERSE FOR PEPTIDE BOND FORMATION

2.1. Protease-Catalyzed Peptide Bond Synthesis under Thermodynamic Control

Based on their ability to function under mild, physiological reaction conditions and their high regio- and stereoselectivity, there has been a longstanding interest in using proteases in reverse to catalyze peptide bond formation.[7,8] Like all enzymatic catalysts, proteases obey the principle of microscopic reversibility and therefore enhance the rates of peptide bond hydrolysis and peptide bond formation by the same amount. However, because enzymes alter the rates of these reactions but not the thermodynamic stabilities of the starting materials and product, whether a peptide bond is predominantly formed or hydrolyzed is determined by the position of the equilibrium (Figure 3A). Under physiological conditions, the position of the equilibrium lies strongly in favor of peptide bond hydrolysis ($\Delta G^{\circ\prime} \sim -3$ kcal/mol[9]). Therefore, a key limitation in the application of proteases to catalyze direct reversal of peptide bond hydrolysis is the high thermodynamic stability of the proteolytic products compared to the ligation product.[10] One approach for increasing ligation product yield under equilibrium-controlled conditions is to shift the position of the equilibrium by altering reaction conditions including solvent polarity, temperature, concentration of water, and pH.[7,11−13] These adjustments shift the pK_A of the carboxylate starting material, increasing the concentration of reactive, neutral carboxyl groups (Figure 3A).[11,13] Many examples exist in which proteases have been induced to function in reverse in this way, generally through the use of organic cosolvents, including commercial processes for the production of aspartame[14] and human insulin.[15] However, many practical limitations to this approach remain, including the need to individually optimize the conditions for each reaction or step and the decreased stability and solubility of proteases in organic solvents.

2.2. Protease-Catalyzed Peptide Bond Synthesis under Kinetic Control

An alternative approach for using proteases in reverse is to run the reaction under kinetically controlled conditions, in which the kinetic properties of the enzyme ($\Delta G^{\ddagger}$), rather than the thermodynamic stabilities of the substrates and products, determine the reaction yield (Figure 3B).[10,13] This approach has been applied in the context of serine and cysteine proteases,[16−20] whose reaction mechanisms involve the formation of an acyl-enzyme intermediate.[21,22] In the context of these enzymes, the protease-catalyzed hydrolysis of both esters and amides proceeds through a similar chemical mechanism. The catalytic nucleophile (Ser or Cys) attacks the carbonyl bond of the substrate to form the acyl-enzyme intermediate. This intermediate then breaks down through the attack of water or another nucleophile on the (thio)ester bond. However, although esters and amides generate the same acyl-enzyme intermediate,[21,23] esters are intrinsically more reactive than amides. Acylation of the enzyme by the more activated ester substrate therefore kinetically outcompetes acylation by the ligated peptide product, minimizing product hydrolysis and increasing ligation product yield on short time scales. However, nonproductive hydrolysis rather than aminolysis of the acyl-enzyme intermediate consumes substrate ester and limits reaction yield.[13]

3. PROTEIN ENGINEERING OF SUBTILISIN FOR IMPROVED PEPTIDE BOND SYNTHESIS

3.1. Mutation of the Catalytic Serine to Cysteine

Protein engineering approaches have enabled the manipulation of the kinetic properties of proteases to optimize them for kinetically controlled peptide bond synthesis. Early protease engineering efforts focused on the application of site-directed chemistry to manipulate the catalytic nucleophile of subtilisin, a serine protease with a canonical Ser-His-Asp catalytic triad. Polgar and Bender[4] and Neet and Koshland[3] independently employed a chemical mutagenesis approach to convert the catalytic serine of subtilisin to cysteine, generating the thiolsubtilisin variant. Thiolsubtilisin exhibited a severe

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

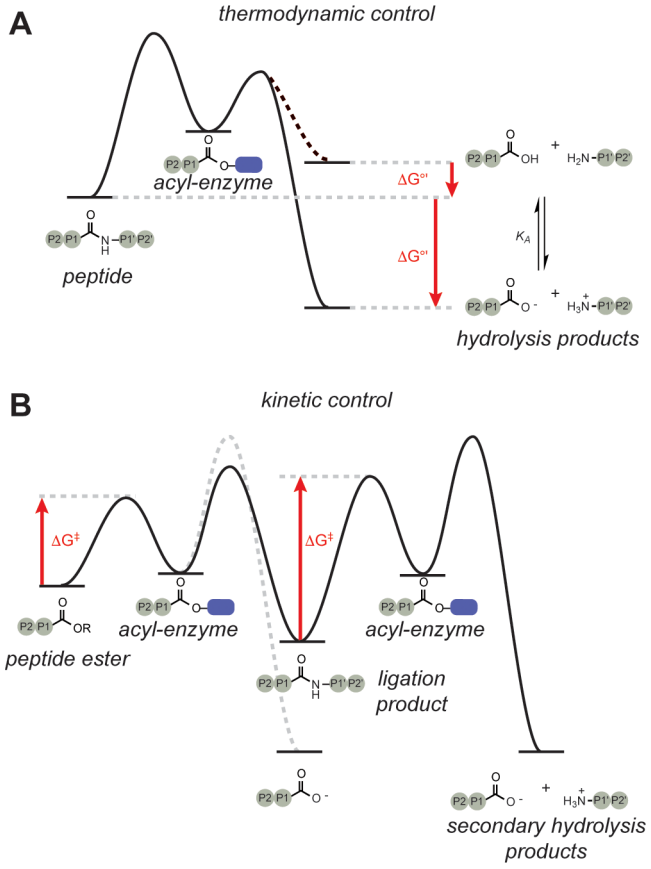


Figure 3. Proteases as catalysts for peptide bond synthesis under thermodynamic or kinetic control. Free energy (G) is plotted on the y-axis, and reaction coordinate is plotted on the x-axis. (A) Under thermodynamic control, the relative yield of peptide ligation products and hydrolysis products is determined by their relative thermodynamic stabilities ($\Delta G^{\circ\prime}$). Reaction conditions can be manipulated to change the position of the equilibrium to increase ligation product yield. (B) Under kinetic control, the relative yield of the ligation and hydrolysis products is determined largely by the energetic barriers ($\Delta G^{\ddagger}$) for acylation of the enzyme by the peptide ester substrate or by the peptide ligation product, which can undergo secondary hydrolysis.

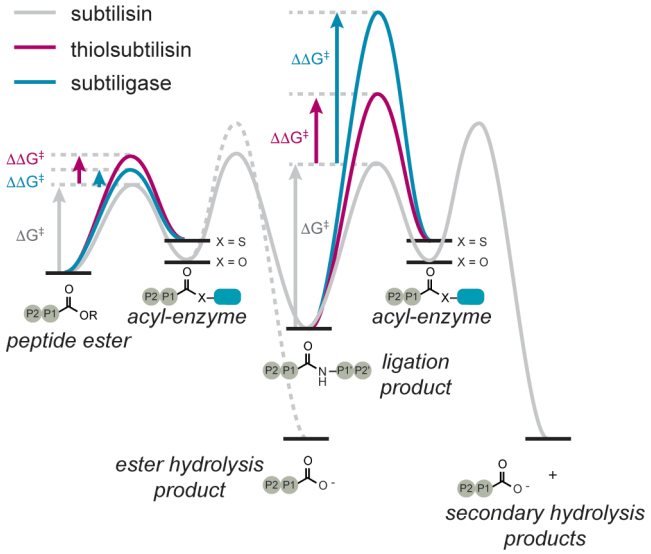


Figure 4. Engineering the kinetic properties of subtilisin for peptide bond synthesis. Introduction of the S221C mutation (thiolsubtilisin) and S221C/P225A (subtiligase) mutations into the subtilisin scaffold changes the energetic barriers ($\Delta G^{\ddagger}$) for specific catalyzed reaction steps. In thiolsubtilisin (magenta), the S221C increases the barrier for acylation of the catalytic nucleophile only slightly for an ester acyl donor (left) but very significantly for an amide acyl donor (right). This results in a favorable ratio of esterase to amidase activity. For subtiligase (cyan), the barrier for acylation of the enzyme by an ester substrate is lower compared to thiolsubtilisin, but the barrier for acylation by a peptide substrate is much higher, resulting in an even more favorable esterase to amidase activity ratio. Free energy (G) is plotted on the y-axis, and reaction coordinate is plotted on the x-axis.

1,000−10,000-fold impairment in amidase activity but retained a relatively high level of esterase activity only 3-fold lower than the wild-type enzyme (Figure 4).

Around the same time, kinetic studies of hydrolysis and aminolysis of esters and thioesters demonstrated that thioesters are much more reactive toward amines than toward water compared to esters.[24,25] Kaiser and co-workers synthesized these observations, recognizing that the kinetic properties of thiolsubtilisin might overcome many of the limitations of proteases in kinetically controlled peptide synthesis.[26] First, because thiolsubtilisin is a poor catalyst for peptide bond hydrolysis, secondary hydrolysis of the ligated peptide product would be diminished compared to conditions under which an active, wild-type protease is used. Second, because thiolsubtilisin maintains the ability to form an acyl-enzyme intermediate from a peptide ester substrate, it can be deacylated through attack of an N-terminal α-amine from another peptide segment to form a peptide bond. Third, because the acyl-enzyme intermediate is a thioester rather than an ester, it is more reactive toward amines, increasing the ratio of aminolysis versus hydrolysis of the intermediate. Using thiolsubtilisin in combination with various p-chlorophenyl esters as acyl donors, it was possible to synthesize small peptides in 80−95% yield. Despite the improved kinetic properties of thiolsubtilisin for peptide bond synthesis, this strategy still required the use of a 20-fold excess of acyl acceptor and high concentrations of organic solvent.

Wu and Hilvert later reported the generation of selenolsubtilisin, a subtilisin variant in which the nucleophilic serine in the protease is replaced by selenocysteine.[27] The selenoenzyme was 14,000 times more efficient than subtilisin and 20 times more efficient than thiolsubtilisin at catalyzing aminolysis. However, selenolsubtilisin is much more susceptible to oxidative inactivation. Therefore, the application of both thiolsubtilisin and selenolsubtilisin in peptide bond synthesis remains limited by the requirement for highly activated esters as acyl donors, the need to use high concentrations of organic solvent, and the low catalytic efficiencies of the mutant enzymes compared to wild-type subtilisin.

3.2. Introduction of the P225A Mutation

The low catalytic efficiency of thiolsubtilisin compared to subtilisin was hypothesized to result in part from steric crowding in the active site based on the larger covalent radius of sulfur (1.03 Å) compared to oxygen (0.65 Å).[2] The development of oligonucleotide-directed site-directed mutagenesis[28] made it possible to evaluate this hypothesis by introducing additional mutations designed to reposition the γ-thiol of the catalytic cysteine. In subtilisin BPN′, the catalytic serine (S221) sits near the N-terminal end of helix α6, which spans residues 220−238 (mature subtilisin BPN′ numbering). Mutation of a highly conserved proline residue (P225) in this

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

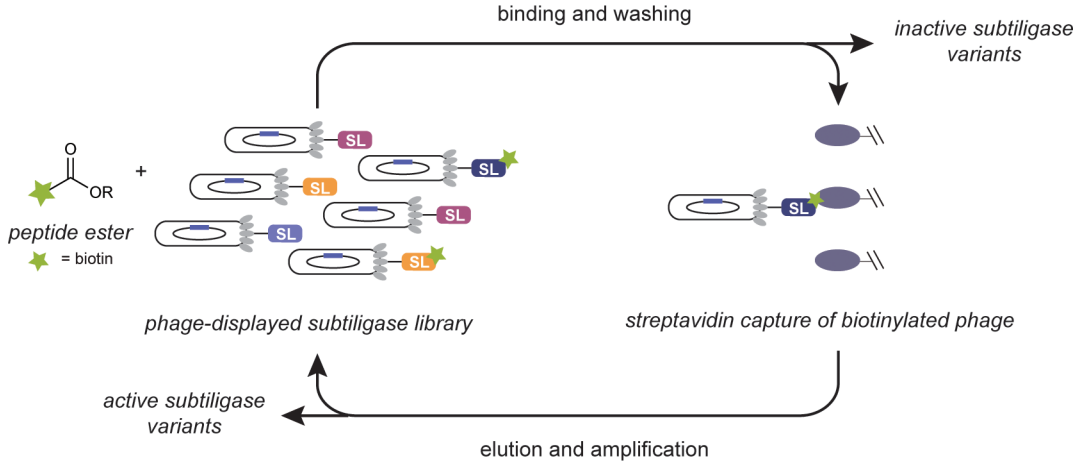


Figure 5. Phage display for screening subtiligase libraries. Subtiligase was displayed on a phage as a fusion with the coat protein pIII. Active subtiligase variants were captured based on their ability to modify their own N termini with a biotinylated peptide ester substrate.

helix was hypothesized to reposition the catalytic nucleophile, moving the γ-hydroxyl of the Ser away from the oxyanion hole and catalytic histidine by 0.5−1.0 Å.[29] Thus, the S221C/P225A double mutant (subtiligase) was designed with the goal of maintaining the improved ratio of aminolysis to hydrolysis observed in thiolsubtilisin while alleviating steric crowding to improve catalytic efficiency.[2] Indeed, X-ray crystallographic analysis of subtiligase compared to subtilisin revealed that the P225A mutation induced a shift of the N-terminal end of helix α6, moving C221 away from the other active site residues by 0.3 Å. This shift in the helix appears to better spatially accommodate the longer carbon−sulfur bond in the introduced catalytic cysteine.

Kinetic analysis of the S221C/P225A double mutant and comparison to the S221C and P225A single mutants demonstrated that the double mutant has improved kinetic properties for peptide bond formation.[2] Each of the single mutants has severely impaired amidase activity, with smaller deficits in esterase activity. In contrast, while the amidase activity of the double mutant is reduced below the detectable limit (a > 10^7-fold decrease compared to wild-type subtilisin), its esterase activity is increased by 10-fold compared to the S221C mutant (Figure 4). At the same time, the S221C/P225A double mutant maintains an aminolysis-to-hydrolysis ratio that is 500-fold improved over that of wild-type subtilisin. Although this ratio is 10-fold below the S221C single mutant, it is sufficient to produce >95% aminolysis without secondary hydrolysis of the ligated product. In the presence of a 10-fold excess of acyl acceptor and using an optimized alkyl ester substrate as the acyl donor, the P225A single mutant quantitatively hydrolyzes the substrate within 1 min. The S221C mutant hydrolyzes about one-third of the substrate, mainly through secondary hydrolysis of the ligated peptide product, resulting in a low (<20%) yield of ligated peptide product in 1 h. In contrast, the S221C/P225A double mutant (subtiligase) rapidly produces the ligated peptide product in high yield (>90%) within 1 h with no detectable secondary hydrolysis. Subtiligase therefore represents a practical tool for peptide segment condensation under kinetic control. Since its development, subtiligase has found widespread application in many areas of chemistry and biology, including peptide synthesis and cyclization,[30,31] total protein synthesis,[32] site-specific protein bioconjugation,[33,34] protein semisynthesis,[33] and mapping proteolytic cleavage sites,[35] among others. The initial S221C/P225A design has also served as a starting point for engineering many of subtiligase's enzymatic properties, including substrate specificity and ligation-to-hydrolysis ratio.[34,36]

3.3. Phage Display Selection of Improved Subtiligase Variants

The success of the P225A mutation in improving the catalytic properties of subtiligase for peptide bond formation led to the hypothesis that additional mutations outside the active site might also favorably reposition the catalytic residues. However, the structure-guided approach used to design the P225A mutant was limited by the need to screen purified enzymes for amidase, esterase, and peptide bond formation activity using laborious, HPLC-based assays. To overcome this limitation, a phage display approach for screening for improved subtiligase variants was developed (Figure 5).[37] Phage display has the advantage of enabling *in vitro* screening of >10^9 enzyme variants if a suitable system can be developed to connect enzyme genotype to catalytic activity.[38] To couple subtiligase activity to phage, the enzyme was displayed as a fusion protein with the phage coat protein pIII and a product capture strategy was employed.[39−41] This approach required several modifications to the original subtiligase sequence. Subtilisin and subtiligase are typically expressed as prepro proteins in which the pre sequence (residues 1−30 of the open reading frame; residues −107 through −78 in mature subtilisin numbering) serves as a signal peptide for secretion and the pro domain (residues 31−107 of the open reading frame; residues −77 through −1 in mature subtilisin numbering) is required for folding of the mature enzyme before its autocatalytic removal.[1,42−45] To avoid the requirement for autocatalytic cleavage to generate the mature enzyme, the pro domain was deleted. To enable subtiligase to fold in the absence of the pro domain, the calcium loop (residues 75−83, mature subtilisin numbering) was also deleted.[42,46,47] Finally, an N-terminal extension of 15 residues was introduced to enable subtiligase to catalyze intramolecular ligation of a biotinylated peptide to its own N terminus, thereby capturing the product of the reaction on the phage particle. This enabled activity-based panning, in which the most active subtiligase mutants could be selected by affinity capture of the biotinylated phage. Using this system, six libraries in which groups of four or five contiguous residues near the active site were randomized were

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

screened for improved activity. Validating the approach, both the catalytic triad and oxyanion hole residues were completely conserved, while randomization at position 225 showed that only Ala or Gly were tolerated. In addition, two new double mutants (M124L/S125A and M124L/L126 V) were identified that had catalytic rates >2-fold improved compared to the original subtiligase design. Interestingly, S125 and L126 are conserved in subtilisins, but mutations at these positions improved subtiligase activity, suggesting that different factors can be adjusted to optimize the designed ligase activity.

3.4. Design of a Calcium- and Pro Domain-Independent Subtiligase Variant

A stabilized variant of subtiligase that folds independently of the pro domain was developed using a variant of the construct that enabled phage display of subtiligase.[48] In addition to the calcium loop deletion that enables pro domain-independent folding,[46,47] 18 stabilizing mutations that were previously identified in subtilisin were introduced. This enzyme, named peptiligase, efficiently catalyzes peptide bond formation and functions in the presence of organic solvents and denaturants.

4. SUBSTRATE SPECIFICITY OF SUBTILIGASE

Much of our understanding of subtiligase substrate specificity is based on the assumption that subtiligase retains the same specificity as the parent protease, subtilisin BPN′. Subtilisin substrate binding has been studied extensively both structurally and biochemically. The standard Schechter and Berger nomenclature for protease substrates defines the scissile bond as the peptide bond linking the P1 (on the N-terminal or acyl side) and P1′ (on the C-terminal or leaving group side) residues.[49] The adjacent residues are numbered outward from the scissile bond. S*n* and S*n*′ denote the corresponding binding pockets on the enzyme that recognize each substrate residue. X-ray crystallographic studies of subtilisin BPN′ in complex with covalent peptide inhibitors, products, and transition state analogs defined a substrate binding cleft that interacts with substrate side chains from P4−P2′.[50−56] Biochemical and kinetic studies of substrate specificity in subtilisin BPN′ have focused largely on the nonprime side of the substrate and substrate binding cleft, presumably due to the ease of using a chromophoric leaving group on the prime side as a readout of activity (Figure 6A). However, more recent work following the development of subtiligase has explored prime-side specificity independently in the context of peptide ligation.

4.1. Substrate Specificity on the Nonprime Side

On the nonprime side, the substrate recognition site in subtilisin BPN′ has been structurally defined in many different crystallographic studies of the protease in complex with protein, peptide, and covalent inhibitors, as well as products and transition state analogs (Figure 7A).[52−54] These studies demonstrate that the enzyme makes numerous contacts with the substrate backbone and also has binding pockets for the side chains of the peptide substrate (Figure 7B). The S4 pocket, which binds the P4 position of the substrate, consists of residues 101−107, Leu 126, and Leu 135. The interior surface of the pocket is primarily hydrophobic, suggesting that it is primed to recognize hydrophobic substrate residues at the P4 position. The S3 binding site forms more of a hydrophobic surface than a pocket, and the side chain of the P3 position of the substrate is generally observed to be oriented facing outward toward solvent. However, contacts are observed between the P3 position and Gly 127, Gly 100, and Leu 126.

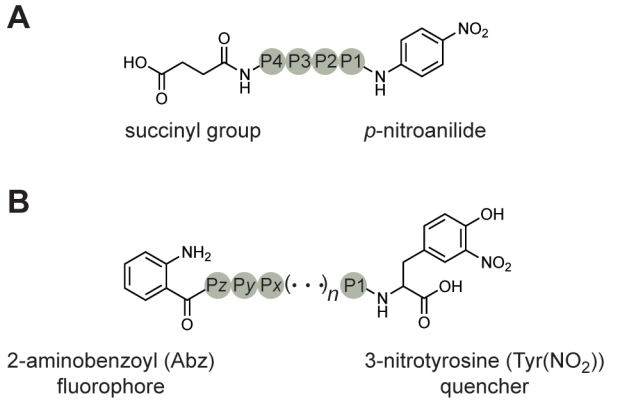


Figure 6. Chromophoric and fluorogenic substrates for subtilisin. (A) Chromophoric substrate for subtilisin. (B) Internally quenched fluorogenic substrate for subtilisin; *x*, *y*, and *z* indicate the distance from the scissile bond.

The S2 binding pocket is comprised of residues 30−35, 60−64, Gly 100, and Leu 126 and is also hydrophobic in nature. The S1 pocket involves primarily Leu 126, Gly 127, Gln 155, and Gly 166. Close contacts are also observed with His 64, Gly 219, Thr 220, and Ser 221.

Internally quenched fluorogenic peptides were used to systematically map amino acid preferences at each subsite in subtilisin BPN′ (Figure 6B).[57] Eight series of substrates were synthesized in which one amino acid in the P5−P3′ positions was systematically varied. For each subsite in subtilisin, the relative preference for each amino acid compared to Gly ($\Delta G_T^{\ddagger}$) was calculated from k_{cat}/K_M. By examining the maximum $\Delta G_T^{\ddagger}$ in each series, it was possible to rank the energetic significance of each P*n*−S*n* interaction. The S4 and S1 pockets are the most discriminating, with smaller contributions from S2, S3, and the prime side. Consistent with crystallographic data that shows that the subtilisin substrate binding cleft spans from P4−P2′,[52,53] variation of the side chain in positions P5 and P3′ had little impact on the efficiency of substrate hydrolysis.

The substrate binding cleft of subtilisin is largely hydrophobic, and biochemical studies demonstrate that the most preferred substrates are hydrophobic as well (Figure 8).[57] In the S4 pocket, there is a strong preference for aromatic residues, although substrates with other large hydrophobic side chains are also hydrolyzed efficiently. Although hydrophobic effects appear to dominate the S4−P4 interaction, there are also steric factors such that P4 Trp substrates are less efficiently hydrolyzed than P4 Phe substrates, an effect that has been attributed to the larger size of Trp. Charged and polar residues are poorly accepted at the P4 position. At the S3 subsite, there is strong discrimination between positively and negatively charged residues, with P3 Arg substrates hydrolyzed the most efficiently. However, there is little discrimination among uncharged residues. Aliphatic side chains in the P2 position are preferentially recognized by the S2 subsite, with negatively charged residues poorly tolerated. The S1 pocket is the most energetically discriminating position in subtilisin BPN′. Hydrophobic side chains are strongly preferred, although β-branched amino acids Val and Ile are poorly accepted. Pro and negatively charged amino acids are also not well tolerated at the P1 position.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

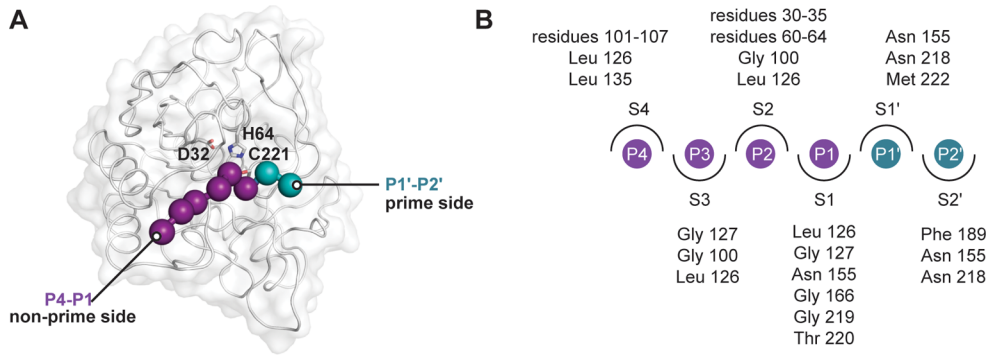


Figure 7. Substrate recognition sites in subtiligase. (A) Subtiligase binds its substrates at the P4 through P2′ positions. (B) Each subtiligase substrate binding site has been structurally identified.

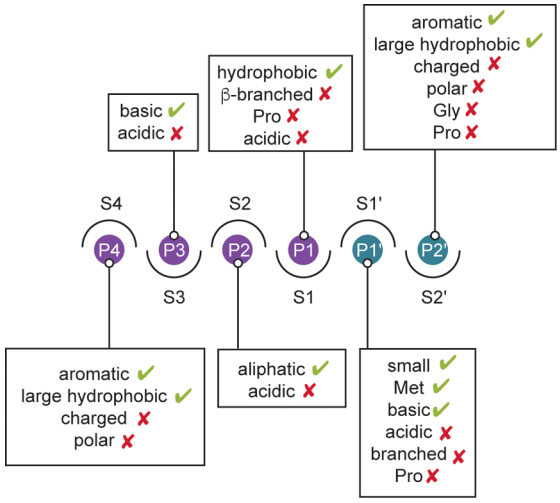


Figure 8. Summary of wild-type subtiligase specificity from P4−P2′.

4.2. Engineering Substrate Specificity on the Nonprime Side

Biochemical studies of the S1−P1 interaction in subtilisin BPN′ show that subtilisin specificity can be engineered in a modular fashion by targeting each of the individual subsites for mutagenesis, enabling the construction of fit-to-purpose enzymes as proteolytic processing tools. A conserved Gly (Gly 166) in the S1 pocket was replaced with 12 nonionic amino acid side chains, and the impact of these mutations on the hydrolysis of substrates with different amino acids at the P1 position was explored.[58] Large changes in substrate specificity were observed in the mutant enzymes. Substrates with small hydrophobic amino acids in the P1 position were generally hydrolyzed more efficiently by mutants with a larger side chain at position 166, while substrates with the large hydrophobic side chains that are preferred by the wild-type enzyme showed a different pattern. Kinetic studies demonstrated that as the volume of the side chain substituted at position 166 increases, the substrate preference of the enzyme shifts from large P1 side chains to small P1 side chains. This shift is proportional to the size of the P1 substrate side chain. Based on analysis of the kinetic data for the different enzyme−substrate pairs, the optimal combined volume of the side chain at position 166 and the P1 substrate side chain is 160 ± 30 Å. This study defined the interplay between steric and hydrophobic effects in substrate recognition and also demonstrated that it is possible to engineer subtilisin variants with altered substrate specificity.

Further protein engineering efforts demonstrated that a small number of amino acid substitutions can be introduced into subtilisin BPN′ to change its substrate specificity.[59] An early substrate specificity engineering study focused on conferring the substrate specificity of *Bacillus licheniformis* subtilisin onto *Bacillus amyloliquefaciens* subtilisin. The two enzymes are 69% identical but differ in their catalytic efficiencies toward various substrates by 10- to 60-fold. In particular, the *B. licheniformis* enzyme is 60-fold more efficient at hydrolyzing substrates with a Glu in the P1 position, and differences in k_{cat}/K_M up to 10-fold are also observed for uncharged substrates. However, only three amino acid substitutions occur proximal to the substrate binding cleft. By introducing the substitutions that occur in the *B. licheniformis* enzyme (Ser 156, Ala 169, and Leu 217) into the *B. amyloliquefaciens* enzyme (Glu 156, Gly 169, and Tyr 217), it was possible to generate a subtilisin BPN′ variant with substrate specificity similar to the *B. licheniformis* enzyme.

In later protein engineering studies, the substrate specificity of subtilisin BPN′ was engineered for efficient hydrolysis of dibasic substrates with the positively charged side chains at the P1 and P2 positions.[60] This was achieved by mutating residues in the S1 and S2 pockets (Gly 166, Ser 33, and Asn 62) to the corresponding amino acids that are observed in the eukaryotic subtilisin homologues Kex2 and PC2, which are known to cleave dibasic substrates.[61] Mutating each position to Glu or Asp led to a shift in substrate preference toward dibasic substrates and away from the hydrophobic substrates preferred by the wild-type enzyme, while mutating multiple positions in combination led to larger, nonadditive shifts in substrate preference. The N62D/G166D and N62D/G166E mutants showed the largest shifts in substrate preference. These mutants were 100- and 500-fold more efficient, respectively, at cleaving substrates with P1−P2 Lys-Lys or Lys-Arg than the wild-type enzyme. For hydrolysis of a P1−P2 Ala-Phe substrate, the mutants were 500- and 100-fold less efficient compared to the wild-type enzyme.

Using the N62D/G166D mutant as a starting point, another subtilisin mutant with additional specificity for a basic residue at the P4 position was engineered.[62] The design of this mutant was based upon the eukaryotic subtilisin family member furin, which naturally recognizes tribasic substrates[63−65] and involved mutating residues in the S4 pocket (Tyr 104, Ile 107, and Leu 126) to the corresponding amino acids in furin. Interestingly, mutating only these positions in subtilisin BPN′ led to the production of inactive enzyme due to poor autocatalytic processing of the cleavage site between the pro-

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

domain and that catalytic domain. However, replacement of the native cleavage site (AHAY) with a furin cleavage site (RHKR) rescued this phenotype and produced an active enzyme with altered specificity. This enzyme, termed furilisin, hydrolyzed succinyl-RAKR-pNA 60,000 times faster than an efficient substrate for the wild-type enzyme, succinyl-AAPF-pNA, with a 360-fold discrimination in favor of Arg compared to Ala observed at P4.

Protein engineering efforts focused on altering subtilisin specificity on the nonprime side have been largely translatable to subtiligase. Using information from mutational studies of subtilisin, subtiligase variants have been designed to recognize specific P1 residues in the donor ester substrate.[2] Like subtilisin, subtiligase prefers hydrophobic or Lys residues at P1. Mutational studies of subtilisin demonstrated that introduction of G166E, E156Q/G166 K, and G166I mutations alters substrate specificity toward P1 Lys or Arg, Glu, and Ala, respectively. Introduction of these same mutations into subtiligase altered substrate specificity as predicted based on their effect on subtilisin substrate hydrolysis. For the substrate succinyl-AAPA-glc-F-amide, the G166I mutant is more efficient at ligation to an Ala-Phe dipeptide, while the E156Q/G166 K mutant efficiently ligates succinyl-AAPE-glc-F-amide. For the succinyl-AAPK-glc-F-amide substrate, the negatively charged G166E mutant catalyzes efficient ligation, while the positively charged E156Q/G166 K double mutant impairs ligation activity. The ability of these mutants to ligate sequences that are not efficiently accepted by wild-type subtiligase provides flexibility in designing ligation junctions, expanding the utility of subtiligase for synthesis of amide bonds. More broadly, these studies demonstrate that specificity-altering mutations identified in the context of subtilisin can be used to generate subtiligase mutants with predictably altered substrate recognition properties.

4.3. Substrate Specificity on the Prime Side

Few structures of subtilisin that illustrate interactions between the enzyme and the prime side of the substrate are available, in part because many structures of subtilisin in complex with covalent inhibitors utilize peptides in which the prime side residues have been replaced with a warhead to modify the catalytic Ser.[66,67] However, many naturally evolved protein inhibitors of subtilisin exist and crystal structures of the enzyme−inhibitor complexes have been used to elucidate the prime side binding pockets.[52−56] Based on the view that proteinaceous inhibitors can be considered as substrates trapped in a potential energy minimum along the reaction coordinate, it is believed that the structures of the enzyme−inhibitor complexes closely resemble the true structure of the enzyme−substrate complex.[52] These structures suggest that the P1′ and P2′ substrate positions are recognized by subtilisin, with P3′ outside the substrate binding cleft (Figure 7). The S1′ pocket involves contacts with His 64, Asn 155, Asn 218, Ser 221, and Met 222. The S2′ pocket involves Phe 189, Asn 155, and Asn 218.

Internally quenched fluorescent peptides have been applied to study the prime side specificity of subtilisin by using an efficient substrate and systematically varying the P1′ and P2′ positions.[57] While the S1′ and S2′ pockets are both less discriminating than any of the nonprime side pockets, the S1′ pocket discriminated more strongly between different amino acids, without a clear pattern of side chain properties in the most preferred substrates. At S2′, aromatic and large hydrophobic residues were preferred. Although the internally quenched peptide strategy allows the P1′ and P2′ positions to be systematically varied, there is a limitation to the approach that complicates interpretation of the data and curbs its translatability from subtilisin to subtiligase. Because production of a fluorescent signal requires only cleavage of the peptide between the fluorophore and the quencher, these experiments do not provide information about the exact cleavage site. It is therefore possible that for substrates with very disfavored residues in the position being varied, subtilisin cleaves in a different position than predicted. In contrast, the use of pNA substrates to probe nonprime side specificity requires cleavage between the P1 residue and the pNA chromophore to produce a signal and therefore provides information about the exact cleavage site. However, this strategy does not enable sampling preferences on the prime side due to the need to replace the P1′ and P2′ residues with pNA.

In the context of subtiligase prime side specificity, the differences in kinetic mechanism between subtilisin and subtiligase are an important consideration (Figure 9). Both

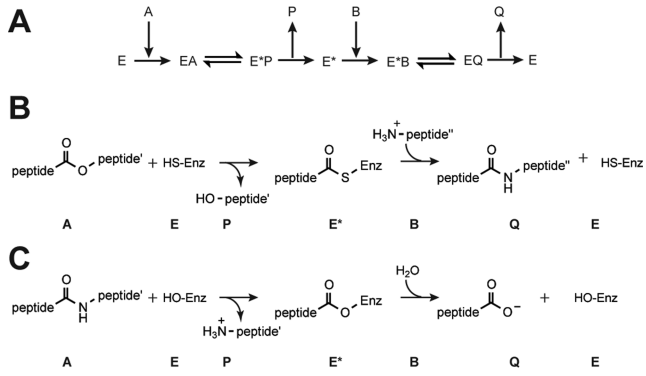


Figure 9. Kinetic mechanisms of subtilisin and subtiligase. (A) A schematic representation of the kinetic mechanism of subtilisin and subtiligase. (B) Kinetic mechanism of subtiligase. (C) Kinetic mechanism of subtilisin.

proteolysis and peptide ligation are bisubstrate reactions that require a peptide or peptide ester substrate to form the acyl−enzyme intermediate and a nucleophile to resolve the intermediate. In both the protease and the ligase, residues on the nonprime and prime sides of the substrate are recognized during the initial substrate binding event prior to formation of the acyl−enzyme intermediate. The residues on the prime side act as a leaving group as the acyl−enzyme intermediate is formed. In the case of the protease reaction, the acyl−enzyme intermediate is hydrolyzed by water as the nucleophile. However, during the peptide ligase reaction, the acyl−enzyme intermediate is intercepted by an α-amine nucleophile from the N terminus of a peptide or protein. This provides a second opportunity for the S1′ and S2′ pockets to discriminate between amino acid side chains independently of the nonprime side of the substrate, which is already covalently attached to the enzyme. Because the nonprime side residues are no longer tethered to the prime side residues in this second discrimination step, energetically significant favorable interactions on the nonprime side can no longer overcome unfavorable interactions on the prime side during substrate binding. Similarly, unfavorable interactions on the nonprime side can no longer lead to the exclusion of otherwise favorable prime-side substrates. For these reasons, a complete under-

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

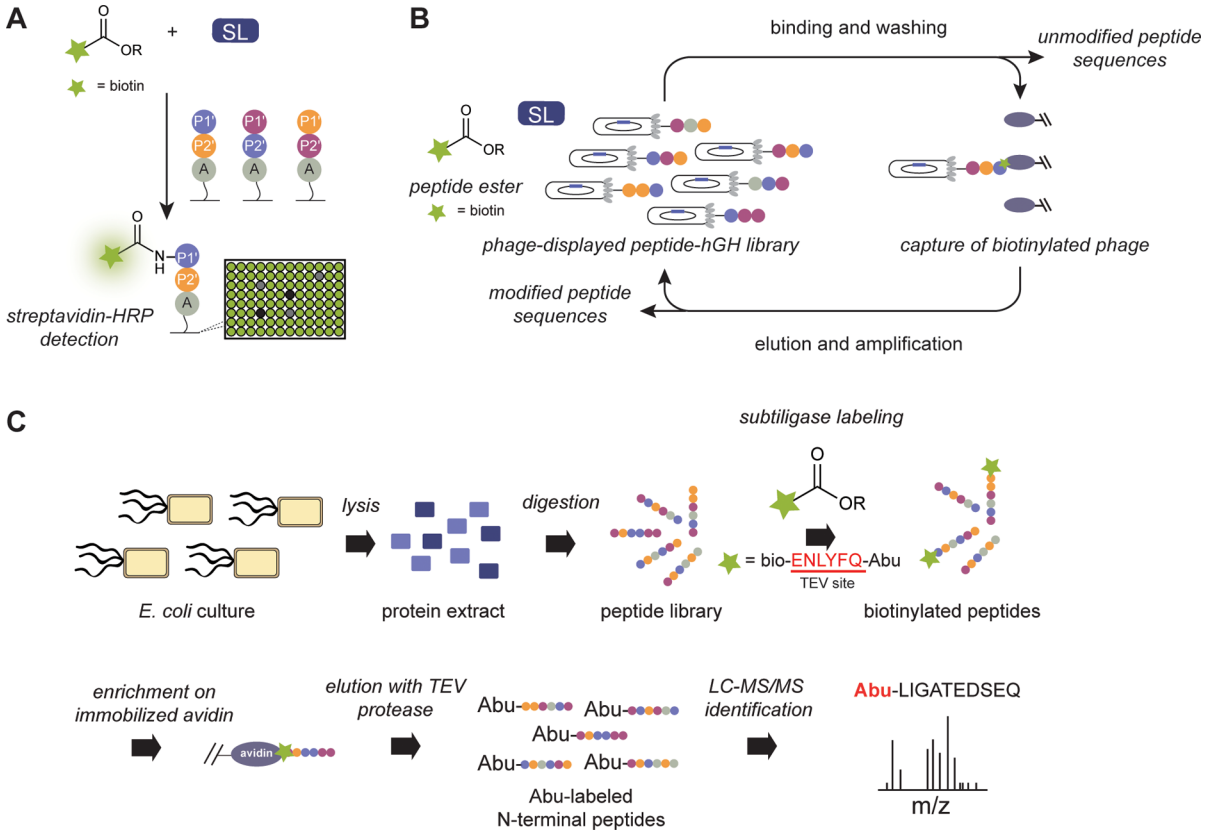


Figure 10. Methods for characterizing subtiligase prime-side specificity. (A) Plate-bound dipeptide library. (B) Substrate phage. (C) Proteomic identification of ligation sites (PILS).

standing of subtiligase prime side specificity required experimental exploration of the ligation efficiency of various prime side substrates, rather than prediction based on kinetic data obtained in the context of subtilisin.

Subtiligase prime side specificity has been characterized using immobilized synthetic peptides, substrate phage, and proteome-derived peptide libraries (Figure 10).[33,34] Using immobilized synthetic tripeptide sequences of the form NH_2-Xaa-Xaa-Ala, subtiligase efficiency for ligation of 400 possible dipeptide sequences was evaluated (Figure 10A).[33] This approach used biotin-KGAAPF-glc-F-amide as the donor peptide ester substrate and used an enzyme-linked immunosorbent assay (ELISA) approach by staining of the immobilized peptides with streptavidin-HRP after a fixed-time incubation with subtiligase. Though somewhat qualitative due to a low dynamic range, this analysis revealed that the most preferred P1′ residues were Met, Phe, Lys, Leu, Arg, Ala, Ser, and His, while the most disfavored residues were Glu, Asp, Pro, Asn, Gln, Thr, Val, and Ile. While the P2′ residue was determined to exert a smaller effect on substrate preference, Cys, Pro, Gly, Gln, Glu, Trp, and Leu were poorly tolerated. However, in many cases, pairing a poor P2′ residue with a good P1′ residue could rescue ligation efficiency (Figure 8).

As a complement to the use of immobilized synthetic peptides, subtiligase substrate specificity was also evaluated using substrate phage (Figure 10B).[33] To display a library of potential subtiligase substrates on the surface of phage, a modified human growth hormone (hGH) with a randomized, three amino acid N-terminal extension was fused to the phage coat protein pIII. After incubation with subtiligase and iminobiotin-KGAAPK-glc-F-amide as the peptide ester donor substrate and three to six rounds of selection, iminobiotin-modified phage was captured on immobilized avidin and sequenced. Although the number of clones sequenced from the input phage library and selectants from rounds three to six was insufficient to statistically sample all possible ligation products, qualitatively, Met, Tyr, Ala, Arg, and Leu were overrepresented at P1′ in the pool of selectants, while Phe, His, Trp, Leu, Ser, Asn, and Tyr were overrepresented at the P2′ position. Based on the common preferred substrates between the synthetic library approach and the substrate phage approach, aromatic, hydrophobic, and positively charged are the most preferred residues in the P1′ position, while acidic residues, Pro, Thr, Asn, Ile, and Val are the least preferred. For analysis at the P2′ position, the results of substrate phage and synthetic peptide analysis did not overlap well, but common preferred residues in this position were Val, His, Tyr, Ser, Ala, Phe, and Asn. Pro and Gly were the least tolerated residues as identified by both methods at P2′ (Figure 8).

In the context of the subtiligase derivative peptiligase, synthetic peptide libraries were used to map prime-side specificity one position at a time.[68] This analysis revealed that peptiligase only accepts Ser, Gly, and Ala at the P1′ position and requires hydrophobic residues at P2′. Mutagenesis studies determined that P1′ specificity is mainly controlled by M213 and L208, analogous to M222 and Y217 in subtiligase.

Proteome-derived peptide libraries[69] have also been deployed to map subtiligase prime side specificity using the proteomic identification of ligation sites (PILS) method (Figure 10C).[34] The PILS method involves digesting the *E. coli* proteome with two digest proteases of orthogonal

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

specificity and using the resulting proteome-derived peptide libraries as a pool of substrates for subtiligase modification. The peptide libraries were incubated with subtiligase and a limiting amount of peptide ester substrate bearing a biotin for affinity enrichment, a TEV protease cleavage site for selective elution, and an aminobutyric acid (Abu) mass tag for substrate identification (biotin-EEENLYFQ-Abu-glc-R-amide). Under these conditions, each peptide sequence is modified according to its abundance in the peptide library and the efficiency with which subtiligase accepts the sequence. Biotinylated substrates were enriched on immobilized streptavidin and selectively eluted by cleaving with TEV protease, leaving behind the Abu mass tag. The input library and the enriched pool of Abu-modified substrates were sequenced by LC-MS/MS to quantify the frequency with which each amino acid occurred in each position in both samples. By evaluating position-specific differences in amino acid frequency, it was possible to quantitatively and comprehensively determine subtiligase specificity. Consistent with previous structural and biochemical studies,[33,52,57] the PILS experiment revealed that there is very little substrate specificity beyond the P2′ position but that there are significant substrate preferences at P1′ and P2′. In the P1′ position, small amino acids (Ala, Ser, and Gly), Met, and Arg were the most favored substrates, while acidic residues (Asp and Glu), branched-chain amino acids (Ile, Leu, Thr, and Val), Pro, and Gln were poorly accepted. At P2′, aromatic (Phe, Trp, and Tyr) and large hydrophobic (Ile, Leu, and Val) residues were enriched among the Abu-modified substrates, while charged (Asp, Glu, Lys, and Arg) and polar (Asn, Gln, and Ser) residues, Gly, and Pro were de-enriched. These results generally agree with the consensus best and worst P1′ and P2′ residues determined by substrate phage and synthetic peptide libraries.[33] However, they also reveal additional substrate discrimination that was not uncovered by the earlier analyses. This is perhaps because the proteome-derived peptide libraries were sufficiently redundant to enable statistical analysis of all 20 amino acids in each position.

Because proteome-derived peptide libraries are sufficiently diverse, the PILS approach also enabled examination of cooperativity between the prime side substrates that could not be examined by varying one amino acid at a time.[34] Measurement of the enrichment or de-enrichment of each of the 400 possible dipeptide sequences following subtiligase modification revealed clear subsite cooperativity in subtiligase. When P2′ is an aromatic residue, any amino acid at the P1′ position except Asp or Glu was efficiently modified by subtiligase. Similarly, when a small amino acid, Arg, or Met is the P1′ residue, a larger number of P2′ residues are well tolerated by subtiligase. Therefore, favorable interactions at either one of the prime side subsites can help to overcome weak or unfavorable interactions at the other, leading to a larger number of sequences that are efficiently ligated by subtiligase.

The PILS method also enabled functional mapping of the S1′ and S2′ pockets by alanine scanning mutagenesis and characterization of the substrate specificity of each mutant.[34] Twenty sites within 7 Å of the catalytic triad were individually mutated to alanine, and the changes in ligation sequence specificity were quantified using PILS. One mutation, F189A, was identified that diminished modification of sequences containing a P2′ residue that was efficiently modified by the wild-type enzyme, while simultaneously enhancing modification of sequences that are poorly tolerated by wild-type subtiligase. This pattern of specificity change suggested that Phe 189 is the primary determinant of P2′ specificity. Another mutation, Y217A, was identified that increased modification of sequences with a P1′ residue that is poorly tolerated by wild-type subtiligase, suggesting that Tyr 217 is the primary determinant of P1′ specificity.

4.4. Engineering Specificity on the Prime Side

Identification of the residues that determine subtiligase prime side specificity enabled the development of engineered variants that efficiently modify P1′−P2′ sequences that are poorly tolerated by wild-type subtiligase.[34] Saturation mutagenesis at position 217 followed by PILS analysis led to the identification of several mutants with altered specificity requirements. The Y217K and Y217R mutants much more efficiently modified sequences with an acidic residue at P1′, while the Y217D and Y217E mutants more efficiently modified sequences with His, Lys, Ser, or Arg at P1′. At position 189, the F189S, F189Q, F189K, and F189R mutants improved modification of peptides with an acidic P2′ residue. Many of the other mutants characterized in this way also led to changes in the efficiency with which individual sequences were modified. Double mutants at positions 189 and 217 were also characterized and resulted in predictable changes in sequence specificity based on the specificity requirements of the corresponding single mutants. In total, 72 subtiligase mutants were characterized, more than doubling the number of the 400 possible dipeptide sequences that can be efficiently modified with subtiligase as measured by the PILS assay. The altered specificity of some of these mutants (Y217K, F189K, and F189R) was validated in the context of a small panel of green fluorescent protein (GFP) variants with altered N-terminal sequences, suggesting that peptide-level specificity results predictably translate to intact proteins. To enable application of this toolbox of subtiligase mutants, a web application, α-Amine Ligation Profiling Informing N-terminal Modification Enzyme Selection (ALPINE) (https://wellslab.ucsf.edu/alpine/) was developed to allow users to algorithmically identify the best subtiligase mutant for modifying a particular N-terminal sequence.[34]

In peptiligase, prime-side specificity was engineered by characterizing the activity of the enzyme toward synthetic peptide libraries that varied one position at a time.[68] Saturation mutagenesis was performed at positions 213 and 208, analogous to 222 and 217 in subtiligase numbering. Replacement of M213 with Ala, Gly, or Pro was determined to broaden the peptiligase substrate profile, and replacement of L208 with Gly, Ala, Ser, or Asn also broadened tolerance of different P1′ amino acids. These mutants were employed for gram-scale synthesis of exenatide, a mimetic of the peptide metabolic hormone incretin, that is used clinically to treat diabetes.[70]

5. PREPARATION OF SUBTILIGASE SUBSTRATES

5.1. Preparation of Peptide Ester Substrates for Subtiligase

5.1.1. Boc/Benzyl Solid Phase Peptide Synthesis. Several methods for the preparation of peptide ester substrates for subtiligase have been reported. In early work, peptide esters were prepared by solid-phase peptide synthesis (SPPS) using Boc chemistry.[71,72] For preparation of a lactate ester, lactic acid sodium salt was incorporated into the resin bound peptide by activation with benzotriazol-1-yloxytris(dimethylamino)-

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

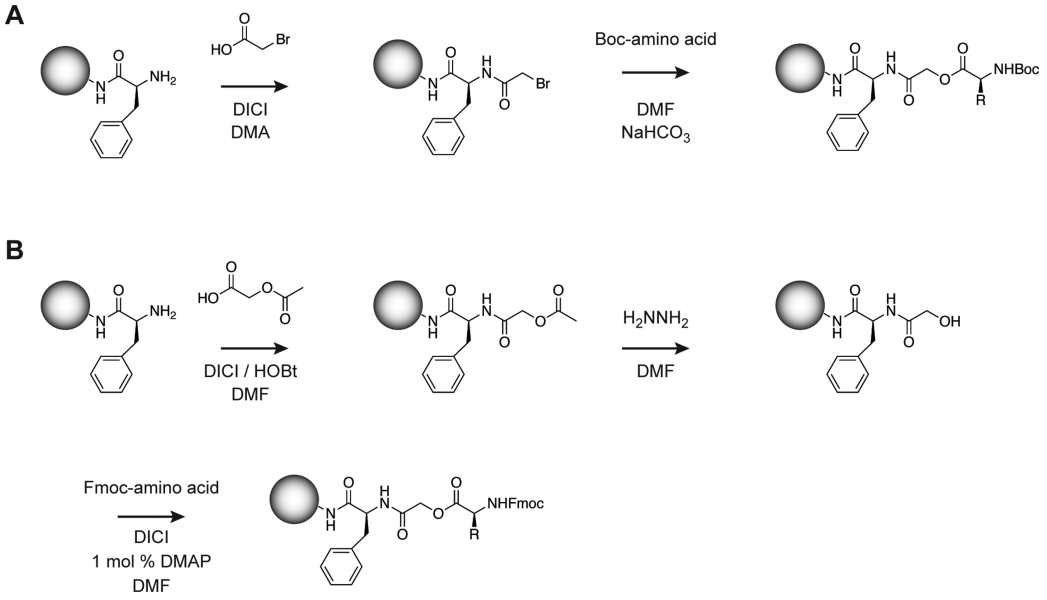


Figure 11. Optimized methods for SPPS of subtiligase ester substrates.

phosphonium hexafluorophosphate (BOP).[73] For preparation of glycolate esters, glycolic acid was incorporated as the *tert*-butyl ether, and the *tert*-butyl group was subsequently removed using 50% trifluoroacetic acid (TFA) in dichloromethane. Synthesis of the ester bond with either the lactic acid or glycolic acid moiety required the use of the nucleophilic catalyst 4-dimethylaminopyridine (DMAP) for coupling of the subsequent amino acid. However, all of the following amino acids could be incorporated using standard Boc SPPS coupling conditions. The synthesis of glycolate esters using SPPS and Boc chemistry was later optimized to eliminate the need for deprotection of the *tert*-butyl ether (Figure 11A).[32] Bromoacetic acid was coupled to the resin-bound peptide using 1,3-diisopropylcarbodiimide (DIC) at 50 °C to generate the bromoacetyl derivative. The subsequent Boc-protected amino acid was then incorporated through nucleophilic displacement of bromide by the carboxylate group in the presence of sodium bicarbonate to form an ester bond. All of the remaining amino acids were then incorporated using standard Boc SPPS coupling conditions. Both methods for Boc SPPS of peptide ester substrates can be carried out either manually or using a peptide synthesizer, and they require the use of anhydrous hydrogen fluoride (HF) for side chain deprotection and cleavage of the peptide from resin.

5.1.2. Fmoc/tBu Solid Phase Peptide Synthesis. While the Boc SPPS method for synthesizing peptide esters is effective, the need to use nonstandard deprotection and coupling conditions, to perform one step at elevated temperature, and to use the hazardous acid HF for peptide cleavage created a demand for development of a more facile approach.[2,32] As a result, two separate methods for using an Fmoc/tBu SPPS scheme for peptide glycolate ester synthesis were reported around the same time.[74,75] One approach relies on solution-phase preparation of a glycolate-modified amino acid before coupling to the resin.[75] Fmoc-L-Tyr(OtBu)-OH was treated with benzyl-2-bromoacetate in the presence of diisopropylethylamine (DIEA) to yield a benzyl ester intermediate. Catalytic reduction of the benzyl ester using H_2 and 10% Pd/C yielded the glycolate-modified amino acid. This modified amino acid was then incorporated into the resin-bound peptide using standard Fmoc SPPS coupling conditions. The second approach for Fmoc SPPS synthesis of peptide glycolate esters relies on direct incorporation of the glycolic acid moiety into the resin-bound peptide (Figure 11B).[74] Following deprotection of the N-terminal Fmoc group, the free amine was coupled to acetoxyacetic acid using standard Fmoc coupling conditions. The acetoxy group was deprotected with hydrazine, and the subsequent amino acid was incorporated in the presence of catalytic DMAP to form the ester bond. Additional amino acids were then incorporated using standard Fmoc SPPS coupling conditions. A similar approach was later optimized that relies on incorporation of iodoacetic acid into the resin-bound peptide, followed by nucleophilic displacement of iodide to form the ester bond.[76]

Although aryl esters and thioaryl esters acylate subtilisin at a faster rate than alkyl esters, they are less stable and more synthetically challenging.[2] Additionally, alkyl esters are closer in structure to natural protease substrates, with glycolate producing an ester linkage analogous to a glycine residue and lactate producing an ester linkage analogous to an alanine residue. Therefore, alkyl esters have been more widely used as acyl donor substrates for subtiligase.[31,33−35,73] To identify the optimal donor ester substrate for subtiligase, a series of glycolate and lactate esters were prepared using Boc SPPS and their kinetics as acyl donors in subtiligase-catalyzed peptide ligation were evaluated.[2] In the glycolate ester series, a systematic decrease in K_M was observed as additional residues were added on the prime side of the ester bond. The peptide ester with glycolate-Phe-Gly-amide as the leaving group acylated subtiligase more efficiently than the substrate with a glycolate-Phe-amide leaving group, which was more efficient than the glycolate-amide substrate. A similar systematic decrease in K_M was observed across the lactate ester series, although the lactate esters acylated subtiligase with k_{cat}/K_M values 5−10-fold lower than those for the corresponding glycolate substrates. The lactate esters contain an additional stereogenic center compared to the glycolate esters; therefore, both the L- and D-lactate stereoisomers were evaluated. The D-lactate ester series acylated subtiligase with k_{cat}/k_M values 40−5000-fold lower than those for the L-lactate ester series,

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

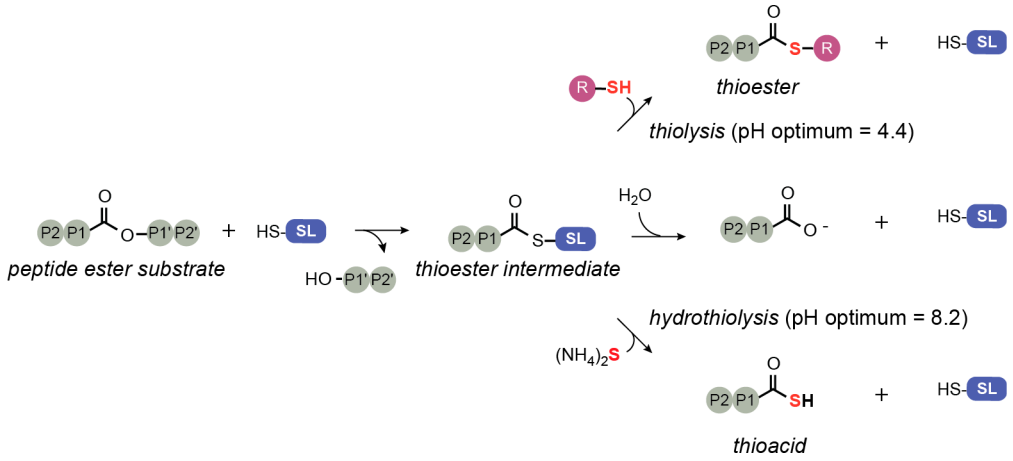


Figure 12. Subtiligase for thioester and thioacid synthesis. The thioacyl−enzyme intermediate in subtiligase can be intercepted by thiols for thioester synthesis or by ammonium sulfide for thioacid synthesis.

suggesting that the lactate ester binds in the S1′ pocket in the same way that a natural amino acid would. Based on the results of these studies, peptide esters with at least one amino acid C-terminal to the ester bond have become the most widely used acyl donor substrates for subtiligase. Although glycolate-phenylalanyl amide substrates were used initially,[32,33,73] subsequent studies have used substrates with glycolate-lysyl amide and glycolate-arginyl amide leaving groups.[77] These substrates function efficiently in subtiligase-catalyzed peptide bond formation but have the advantage of greater solubility as they are positively charged at neutral pH.

5.2. Preparation of Peptide Thioester Substrates for Subtiligase

Subtiligase activity in peptide ligation was originally assessed using the commercially available thiobenzyl ester substrate succinyl-Ala-Ala-Pro-Phe-SBz.[73] Although the k_{cat} value for this substrate in subtiligase-catalyzed peptide ligation was higher than those measured for the alkyl ester substrates, peptide thioester substrates were considered more synthetically challenging than peptide esters at the time subtiligase was first reported. Subsequent studies therefore focused on peptide esters as acyl donors.[33,37,48,73] However, the advent of synthetic methods for peptide thioesters that are compatible with the more convenient Fmoc/tBu SPPS strategy[78−82] made it possible to revisit alkyl thioesters as subtiligase substrates. To compare the efficiency and kinetics of peptide ester and thioester substrates in subtiligase-catalyzed peptide ligation, two model acyl donors of the form acetyl-His-Ala-Ala-Pro-Phe-X-Phe-Gly-amide, where X is either a glycolate or a thioglycolate linkage, were synthesized.[83] The subtiligase-catalyzed kinetics and reaction yields for ligation to an Ala-Phe-Ala-amide substrate were then evaluated. In the case of the thioester substrate, quantitative ligation was observed within 3 min, while the ester substrate required a much longer 65 min reaction time. In the case of a more unfavorable acyl donor substrate sequence, acetyl-LVKEI-X-FG-amide, only 62% of the ester formed the ligation product over 390 min, while the thioester was completely converted to the ligation product within 80 min. Even for a very unfavorable substrate sequence, acetyl-LVKEG-X-FG-amide, use of the thioester substrate resulted in a 10% yield of ligation product, while the ester substrate was completely hydrolyzed. Measurement of the Michaelis−Menten kinetic parameters for both the favorable and unfavorable substrates demonstrated that the catalytic efficiency for ligation of the thioester substrate is 10−20-fold higher than k_{cat}/K_M for the ester substrate. Most of this effect arose in the k_{cat} parameter, suggesting that the enhanced catalytic efficiency is mainly attributable to the intrinsically higher reactivity of the thioester bond. Consistent with the difference in catalytic efficiency arising during the acylation step, use of the thioester as an acyl donor with a poor nucleophile substrate (Gly-Gly-Leu-Gly-amide) could not rescue formation of the ligation product. These results indicate that peptide thioesters could be more useful substrates in cases where reaction time is important and also suggest that acylation is the rate-limiting step in subtiligase-catalyzed peptide bond formation.

An approach to synthesize longer thioester substrates for subtiligase has also been developed.[84] The method relies on the use of a modified intein that can be C-terminally esterified with benzyl mercaptan. This enabled recombinant expression of C-terminally thioesterified proteins that are inaccessible by SPPS because of their length.

6. APPLICATION OF SUBTILIGASE FOR PEPTIDE AND PROTEIN SYNTHESIS

Based on its ability to catalyze peptide bond formation with absolute chemoselectivity for N-terminal α-amines over lysine ϵ-amines, subtiligase has been applied broadly as a tool for synthesis of linear and circular peptides, for site-specific protein bioconjugation, for protein semisynthesis, and for protein total synthesis. Other applications of subtiligase for synthesis of precursors for peptide synthesis, such as thioesters and thioacids, have also been developed. These methods take advantage of the ability of subtiligase to catalyze the formation of an acyl−enzyme intermediate that can be intercepted by nucleophiles other than water or amines, such as thiols. Together, these subtiligase-based methods have enhanced our ability to synthesize complex peptides and uniformly modified proteins as potential therapeutics and probes of biological function.

6.1. Subtiligase-Catalyzed Thioester and Thioacid Synthesis for Peptide and Protein Bioconjugation

Subtiligase has been applied as a tool for enzymatic synthesis of thioesters through an ester-to-thioester transesterification reaction (Figure 12, top).[85] Thioesters are important

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

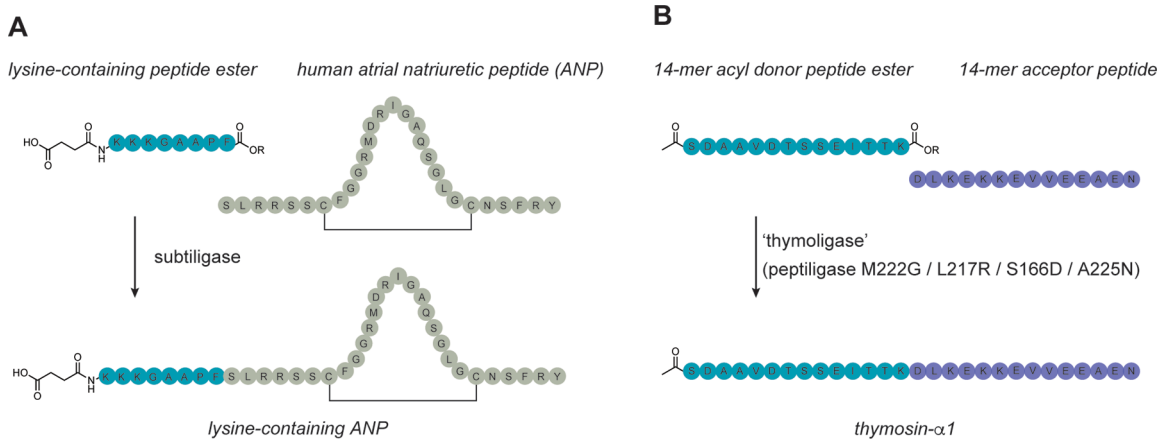


Figure 13. Peptide segment condensation with subtiligase. (A) Ligation of a lysine-containing peptide to human atrial natriuretic peptide for further derivatization. (B) Synthesis of thymosin-α1 from two 14-mer segments using thymoligase, a subtiligase derivative.

intermediates in the synthesis of protein and protein bioconjugates using native chemical ligation and related methodologies.[86] However, peptide thioesters are typically more challenging to access via solid phase peptide synthesis than peptide esters. Subtiligase-catalyzed synthesis of peptide thioesters from peptide esters is therefore an attractive approach to produce these intermediates for use in protein bioconjugation reactions.[85] Thioester synthesis using subtiligase represents a greater challenge than peptide bond synthesis, because rather than resulting in formation of a more stable amide bond, a more labile thioester bond is produced. However, because thiols are better nucleophiles than water, it is possible to achieve catalytic thioester synthesis using subtiligase under kinetic control. Lowering the pH of the reaction medium to 4.4 was observed to decrease the rate of hydrolysis more than that of thiolysis, an effect that was attributed to the lower pK_A of thiols (~9) compared to water (~16). Using these conditions, in combination with a large excess of thiol nucleophile, it was possible to achieve moderate yields (up to 78%) of thioester product. Thiolysis of the subtiligase acyl−enzyme intermediate occurred 800 times faster than the same reaction on a chemically equivalent synthetic peptide substrate, demonstrating that subtiligase catalyzes the transesterification reaction.

In a related application, subtiligase has also been used as a hydrothiolase for the synthesis of peptide thioacids (Figure 12, bottom),[87] important intermediates for peptide segment condensation reactions in which peptide bond formation is driven by entropically favorable intramolecular acylation.[88] Like peptide thioesters, peptide thioacids are challenging to prepare using solid phase peptide synthesis and generally require the use of more hazardous Boc/benzyl-based strategies.[89,90] In the presence of high concentrations of ammonium sulfide, the thioacyl−enzyme intermediate formed on subtiligase from a peptide ester substrate could be captured by sulfide ion to produce the corresponding thioacid.[87] Reaction conditions were carefully optimized to favor the hydrothiolysis product, although the considerations were different from those that governed optimal yields of thioester product. In the hydrothiolysis reaction, the peptide thioacid product was stable toward subtiligase-catalyzed hydrolysis. However, a significant amount of hydrolysis product was formed from direct hydrolysis of the peptide ester substrate. This hydrolysis reaction could not be outcompeted by using a high concentration of ammonium sulfide because of its limited solubility below neutral pH and tendency to escape as hydrogen sulfide. The reaction conditions could therefore not be adjusted to favor hydrothiolysis over hydrolysis by lowering the pH. The optimum ratio of hydrothiolysis product to hydrolysis product was therefore achieved at pH 8.2, producing peptide thioacids in moderate yield. The utility of the peptide thioacid products was demonstrated by using them as precursors in the peptide bioconjugation strategy known as "mini thiol capture ligation".[88] This method relies on capture of a 3-nitro-2-pyridylthio (Npys)-activated cysteine side chain by the thiocarboxylic acid of a peptide thioacid. The resultant acyl disulfide then rapidly undergoes an intramolecular S-to-N acyl shift, producing a peptide bond. The captured cysteine residue can then be thiolytically reduced to generate a native cysteine side chain. A peptide thioacid synthesized from the corresponding glycolate ester using subtiligase, Ac-His-Ala-Ala-Pro-Phe-SH, was incubated with a 3-fold molar excess of H-Cys(Npys)-Phe-Glu-Val-Lys-Gly-NH_2, and the two modified peptides underwent ligation in quantitative yield after 5 min. This demonstrated the utility of peptide thioacids synthesized by subtiligase-catalyzed hydrothiolysis for peptide condensation applications.

6.2. Peptide Segment Condensation

Subtiligase activity has been studied and optimized extensively in the context of ligation of linear peptide segments. As a result, subtiligase-catalyzed peptide ligation has been useful for synthesis of modified and synthetically challenging peptides. Subtiligase was applied to modify atrial natriuretic peptide (ANP), a peptide hormone secreted from the heart to regulate blood pressure and salt balance (Figure 13A).[33] ANP has been applied in the clinic in the context of acute heart failure and kidney disease but is rapidly cleared from serum via kidney filtration.[91] One method for extending serum half-life of peptides and proteins is to modify them with polyethylene glycol (PEG) polymers.[92] This modification is usually achieved by modification of lysine residues with PEG-*N*-hydroxysuccinimide (NHS) ester. However, this method is not applicable to ANP, which does not contain any lysine residues. Using subtiligase, ANP was N-terminally modified with succinyl-KKKGAAPF, enabling subsequent modification of the introduced lysine residues with PEG-NHS as a proof of principle.[33]

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

The subtiligase variant peptiligase has been applied for convergent chemo-enzymatic synthesis of thymosin-α1,[31] an immunoregulatory peptide that is clinically approved for treatment of hepatitis B and C (Figure 13B).[93−95] Synthesis of thymosin-α1, an acetylated 28-mer peptide, via SPPS is challenging because of its tendency to form secondary structure and because a large number of side chain protecting groups are required. A convergent strategy involving the peptiligase ligation of two 14-mer peptides was developed, resulting in a 2-fold increase in yield compared to the current industrial process for thymosin-α1 synthesis.[31] Efficient ligation of the two segments required structure-guided engineering of the substrate-binding cleft for improved recognition of the two peptide segments. In particular, mutation of previously identified residues in the S1 and S1′ pockets that control P1 and P1′ substrate recognition led to increased ligation of the substrates, which contained P1 Lys and P1′ Asp. This led to 55% overall yield of thymosin-α1 with only one HPLC purification step required, demonstrating a chemo-enzymatic peptide synthesis strategy that could increase efficiency and lower costs in industrial peptide production.

6.3. Peptide Cyclization

Cyclic peptides are a therapeutically promising class of molecules that combine many of the advantageous features of small molecule drugs and biologics.[96,97] They are more resistant to degradation than linear peptides, and their conformational rigidity often results in higher binding affinities. Cyclic peptides may be joined head-to-tail, side chain to side chain, or side chain to N or C terminus. While a number of solid-phase and solution chemical methods exist for peptide cyclization, they have some limitations (reviewed in ref 98). The entropic barrier to peptide cyclization is large, making cyclization of peptides larger than 10 residues inefficient. Additionally, intermolecular oligomerization competes with cyclization, necessitating low substrate concentrations that limit the scalability of these reactions. As an alternative to these chemical methods, subtiligase has been applied as a head-to-tail peptide cyclization catalyst (Figure 14).[30] Linear peptide esters ranging in length from 12 to 31 amino acids were incubated with subtiligase to produce cyclic peptides in 30−88% yield.

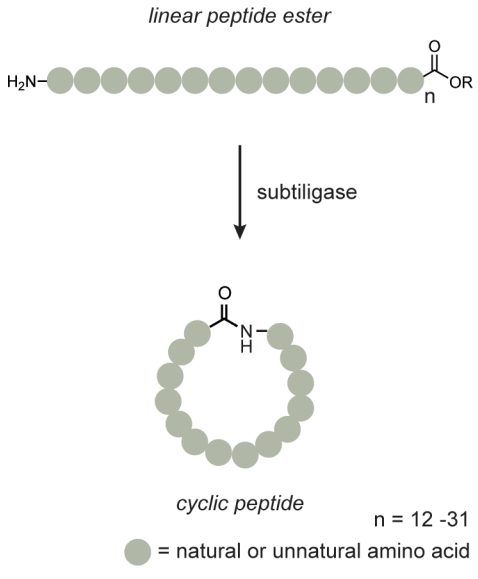


Figure 14. Peptide cyclization with subtiligase.

The efficiency of the ligation reaction was dependent both on the sequence of the ligation junction and on the length of the peptide ester substrate. By using peptide esters with N- and C-terminal sequences that were known to be efficient substrates for subtiligase ligation and varying the linker length between them, it was determined that a minimum of 12 residues are required for cyclization to occur. Peptide esters shorter than 12 residues were intermolecularly oligomerized or hydrolyzed by subtiligase. As the length of the peptide ester substrates increased beyond 12 residues, cyclization efficiency also increased. For peptides longer than 14 residues, only the cyclized product was observed, with no intermolecular ligation occurring. Consistent with an intramolecular step following acylation of the enzyme, yield of the cyclized peptide was independent of substrate concentration. In later work, the subtiligase derivative peptiligase was used in combination with other bioconjugation strategies such as click chemistry, oxime ligation, and chemical linkage of peptides onto scaffolds (CLIPS) to generate bicyclic, tricyclic, and tetracyclic peptides.[99−101] The application of subtiligase to peptide cyclization has the advantages of functioning on fully deprotected peptides, concentration independence, and minimal sequence limitations at the ligation junction compared to the requirements of other enzymes, making this approach scalable and applicable to a wide variety of cyclic peptides.

6.4. Total Protein Synthesis

The incorporation of unnatural amino acids into proteins is an important method for probing protein structure and function.[102−105] While several technologies exist for incorporation of unnatural amino acids into proteins using recombinant expression technologies,[106,107] these methods are limited by the inability to incorporate substitutions at multiple sites, or a lack of site specificity. Total protein synthesis is therefore an attractive alternative approach to enable unnatural amino acid substitutions at multiple defined sites. While SPPS has been applied for this purpose,[108−110] its use is limited to small proteins <100 amino acids in length, and chemical synthesis of proteins of this size remains a significant challenge. Subtiligase has been applied as a tool to catalyze condensation of protein segments produced by SPPS, enabling total synthesis of proteins >100 amino acids in length in good purity with reasonable yields (Figure 15).[32]

To demonstrate the subtiligase-based coupling and deprotection scheme, RNase A (124 amino acids) was chosen as a target for total synthesis.[32] RNase A was an ideal target because it is structurally[111,112] and kinetically[113,114] well characterized, can be refolded *in vitro*,[115,116] and has an interesting catalytic mechanism[117,118] that could be probed by the incorporation of unnatural amino acids. The protein was divided into six fragments 11−31 residues long,[32] which were designed taking into account the optimal P1 and P1′ residues for subtiligase ligation. The fragments were ligated together beginning with the C-terminal peptide and building toward the N terminus of the protein. The fully deprotected C-terminal fragment was coupled to the next fragment, a glycolate-phenylalanyl amide donor peptide protected on its N terminus with an isonicotinyl carbamate (iNoc) group[119] to prevent self-ligation. The iNoc group is advantageous as an N-terminal protecting group because it is compatible with both Fmoc/tBu and Boc/Bzl SPPS and can be removed after completion of the ligation reaction under mildly reducing conditions (Zn dust and acetic acid) to unblock the N terminus for subsequent

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

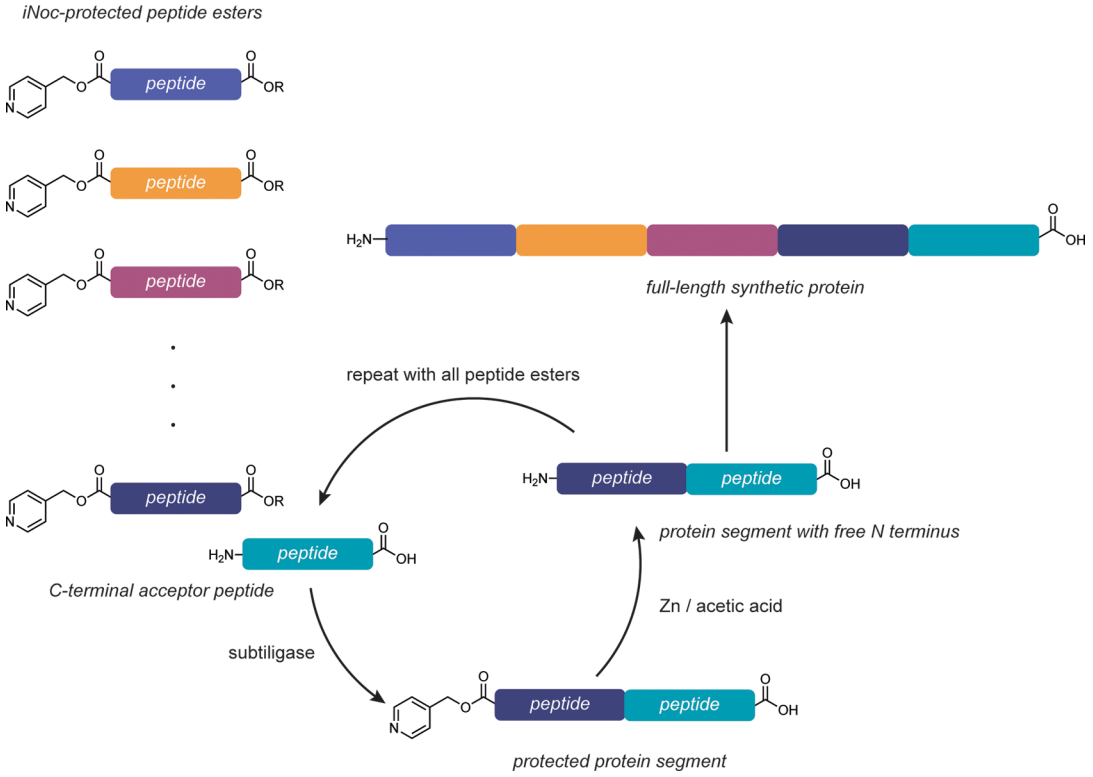


Figure 15. Total protein synthesis with subtiligase. iNoc, isonicotinyl; R = glycolate phenylalanyl ester.

ligations. Each ligation step was carried out with 2−5 mM C-terminal acceptor peptide, 3−5 equiv donor peptide, and 5 μM subtiligase for 1−2 h at room temperature. The average yield per ligation step was 66%, and the average yield per deprotection step was 86%, for a 15% overall yield of synthetic protein before refolding. Full-length RNase A was oxidized and refolded using protein disulfide isomerase and isolated in milligram quantities following HPLC purification (8% overall yield). The isolated RNase A was >98% pure based on SDS-PAGE analysis, and its Michaelis−Menten kinetic parameters were virtually identical to those previously reported for RNase A isolated from natural sources.

Using the same strategy that was used to synthesize wild-type RNase A, three variants were produced in which the catalytic histidines (His 12 and His 119) were substituted either individually or in combination with the unnatural amino acid 4-fluorohistidine (4fHis).[32] Because 4fHis has a lower pK_A (3.5) than that of His (6.8) but is nearly isosteric with His, incorporation of this amino acid enabled the catalytic functions of the two His residues to be probed. In the first step of the RNase A catalytic mechanism, the 2′-hydroxyl attacks the 3′-phosphate to form a 2′,3′-cyclic phosphate intermediate. In the second step, the 2′,3′-cyclic phosphate intermediate is hydrolyzed.[117] In the wild-type enzyme, His 119 functions in general acid catalysis in the first step of the reaction and in general base catalysis in the second step of the reaction, while His 12 functions in general base catalysis in the first step of the reaction and general acid catalysis in the second step of the reaction. For both steps of the reaction, the pH-rate profile of the wild-type enzyme has a maximum at pH 6.5−7, with an acidic inflection point likely reflecting the need for one His to function in general base catalysis and a basic inflection point likely reflecting the need for one His to function in general acid catalysis. As expected based on the pK_A difference between 4fHis and His, substitution of both catalytic His residues with 4fHis results in an acidic shift of the maximal rate of the enzyme for catalysis of both steps from pH 6.5−7 in wild-type RNase A to pH 4−4.5 in the double mutant.[32] Interestingly, the maximal rate of the enzyme is only attenuated by 3-fold in the double mutant, suggesting that matched pK_A's between the two catalytic residues to enable proton transfer is more important than the pK_A's of the catalytic residues relative to that of the substrate.

The subtiligase-catalyzed segment condensation approach has a number of notable advantages. It enables total synthesis of larger proteins than can be accessed using SPPS and relies on fully deprotected peptides that are much more water-soluble than their side chain protected counterparts. Substitutions at multiple sites with unnatural amino acids are possible, and impurities can be removed at every step such that the final product is of high purity. Although yields drop as the number of peptide segments increases, the method can easily be scaled up for synthesis of milligram quantities of protein. Remaining limitations of this method include the need to refold the synthetic protein in vitro and the need to carefully design ligation junctions to match subtiligase substrate specificity.[120]

7. APPLICATION OF SUBTILIGASE FOR SITE-SPECIFIC PROTEIN BIOCONJUGATION

Site-specific protein modification is a key technique that has enabled advances in many areas across the chemical and biological sciences, including the development of chemical probes of enzyme function,[121−123] drug discovery,[124−126] antibody-drug conjugate synthesis,[127,128] and the application of fluorescent imaging techniques.[129] Many protein bioconjugation strategies depend on targeting naturally occurring side chains, such as lysine, which provides poor site specificity and

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

poor control over modification stoichiometry.[130] Other methods target engineered cysteine[127,128] or methionine[131] residues or genetically encoded epitopes,[132−134] requiring genetic engineering of the protein of interest. The N terminus is an attractive site for protein bioconjugation because it naturally occurs once and only once in each polypeptide chain.[135] The application of subtiligase is an attractive method of N-terminal modification because ligation occurs with absolute chemoselectivity for the protein N terminus over lysine ϵ-amines without the requirement for a particular amino acid sequence tag (Figure 16).[2,33,34] The efficiency of ligation

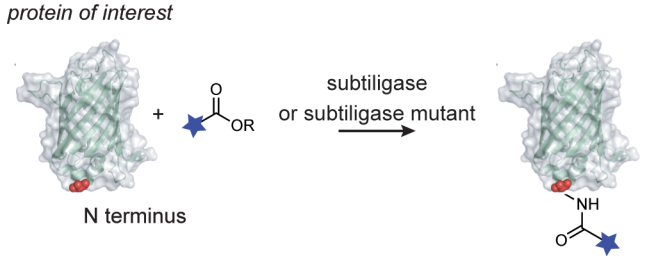


Figure 16. Subtiligase-catalyzed protein bioconjugation.

depends primarily on the suitability of the N-terminal sequence to be modified as a subtiligase substrate and on its accessibility and ability to bind subtiligase in an extended conformation.

7.1. Sequence and Structural Requirements for N-Terminal Modification by Subtiligase

7.1.1. Characterization of Sequence and Structural Requirements. Subtiligase protein bioconjugation was first explored in the context of human growth hormone (hGH) to probe the sequence and structural requirements for efficient modification of an intact protein.[33] Ligation to the native Phe-Pro N terminus was inefficient (2% yield), but extension of the N terminus by one Met to Met-Phe-Pro resulted in 95% ligation yield. This was attributed to the presence of an unfavorable Pro residue at the P2′ position of native hGH which was shifted into the less important P3′ position in the Met-extended variant. Providing support for this hypothesis, mutation of the Pro residue to Ala rescued ligation yield. When subtiligase was used for optimal substrates such as Met-hGH, it was possible to modify the protein efficiently and in high yield with a variety of useful payloads, including a biotin for affinity capture and a peptide bearing a mercurated cysteine residue for applications in X-ray crystallographic studies.

To probe the impact of sequence and N-terminal structure on bioconjugation yields, two truncated variants of hGH (Δ2 hGH and Δ8 hGH) were also tested.[33] The first variant, Δ2 hGH, has an N-terminal sequence of Thr-Ile-Pro, which was predicted based on substrate phage and synthetic peptide library studies to be a poor subtiligase substrate and indeed was ligated in <2% yield. The second variant, Δ8 hGH, has an N-terminal sequence of Leu-Phe-Asp and was predicted to be an efficient substrate for subtiligase. However, the conjugated product was obtained in <5% yield. This was attributed to the observation that helix 1 of hGH begins at position 6, and therefore, residues 9, 10, and 11 may be in a helical conformation that cannot bind the substrate binding cleft of subtiligase, which requires an extended substrate conformation.

Based on the sequence and structural requirements of subtiligase, introduction of extended N-terminal sequences has been employed as a strategy for achieving high modification yields. Introduction of the optimized subtiligase acceptor sequence AFA to the N terminus of the human IgG Fc region (huFc) enabled its efficient modification by subtiligase using either a thioester peptide or an ester peptide as the acyl donor.[83] While both acyl donors were able to modify huFc in high yields, the thioester peptide reacted much faster, facilitating the use of a short (6 min) reaction time.

7.1.2. Engineering Subtiligase To Function in the Presence of Detergents and Chaotropic Agents. To generalize the utility of subtiligase for modification of proteins with unfavorable structural features at the N terminus, a protein engineering effort was undertaken to generate variants that function in the presence of detergents and chaotropic agents.[33] Previous work had identified mutants of subtilisin that enhance its stability to heat, basic conditions, and organic solvents.[136−139] Five of these previously identified stabilizing mutations (M50F, N76D, N109S, K213R, and N218S) were introduced into subtiligase to generate stabiligase.[33] The stabiligase variant retained >50% of its activity for ligation of a peptide ester on Met-hGH even in the presence of 4 M guanidine hydrochloride. For Δ8 hGH, which was previously resistant to subtiligase modification, stabiligase could ligate a peptide ester to the N terminus in 45% yield in the presence of 0.1% SDS, compared to a 15% yield obtained with subtiligase under the same conditions. The development of stabiligase therefore makes it possible to modify proteins that would otherwise be poor subtiligase substrates, expanding the utility of subtiligase for protein bioconjugation.

7.2. Selection of Subtiligase Mutants for Protein Bioconjugation

Although subtiligase has broad sequence specificity on the prime side,[33,34,57] some sequence biases are inherent to the enzyme and placed limitations on its usefulness for protein bioconjugation. To expand the number of N-terminal protein sequences that can be modified using subtiligase, engineered specificity mutants that were identified by the proteomic identification of ligation sites (PILS) approach were applied in protein bioconjugation.[34] As an initial test of whether PILS N-terminal specificity data translates in a predictable way to intact proteins, a small library of N-terminal variants of green fluorescent protein (GFP) variants were tested as substrates for protein bioconjugation with a panel of subtiligase specificity variants (Y217K, F189K, and F189R). The GFP variants contained N-terminal sequences that were either good (Ala-Phe) or poor (Asp-Phe, Glu-Phe, Ala-Asp, and Ala-Glu) substrates for wild-type subtiligase. The specificity mutations were introduced in the context of wild-type subtiligase, stabiligase, and stabiligase-M222A, which contains an additional mutation that enhances the ligation-to-hydrolysis ratio for the peptide ester acyl donor substrate. By testing this panel of mutants, it was possible to identify conditions under which all five GFP variants, including those that were completely recalcitrant to modification with the wild-type enzyme, were modified in >95% yield in two rounds of modification or less. The mutant that was most efficient for modifying any particular GFP variant was predictable based on the specificity profile generated by PILS. Therefore, for future application of subtiligase mutants for N-terminal modification, screening a panel of enzyme variants is unnecessary. Rather, the compiled PILS specificity maps available through the α-Amine Ligation Profiling Informing N-terminal Modification Enzyme Selection (ALPINE) web application (https://wellslab.ucsf.edu/alpine/)

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

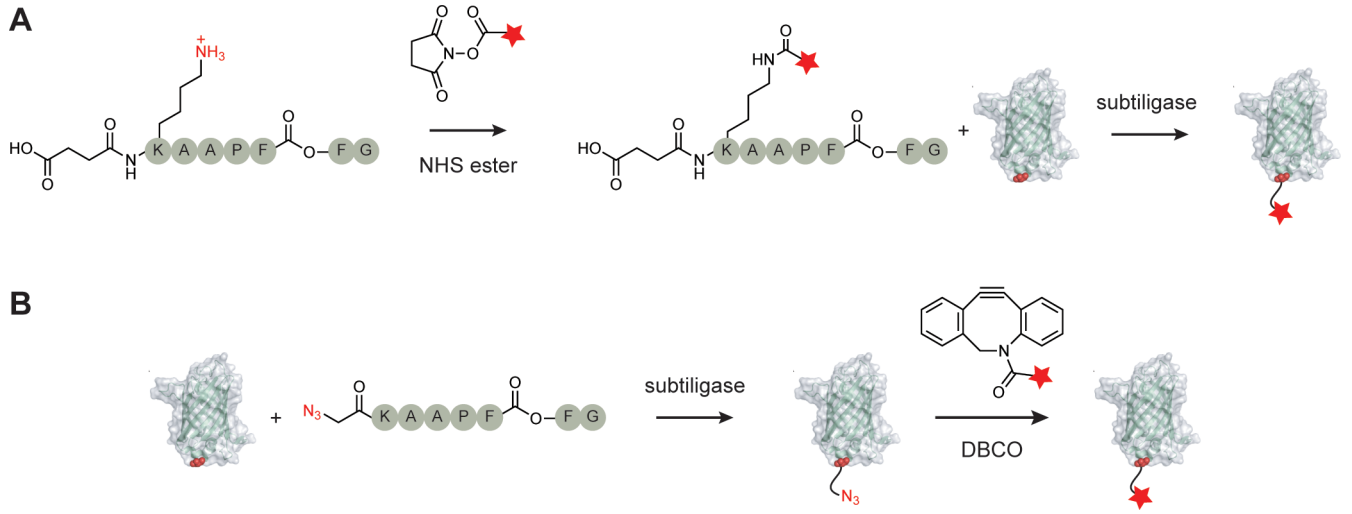


Figure 17. Versatile substrates for subtiligase-catalyzed protein bioconjugation. (A) One-step protein modification following modification of a subtiligase substrate with an *N*-hydroxysuccinimide (NHS) ester. (B) Modular protein modification with an azide-bearing substrate, followed by click chemistry with a dibenzocyclooctyne (DBCO) reagent.

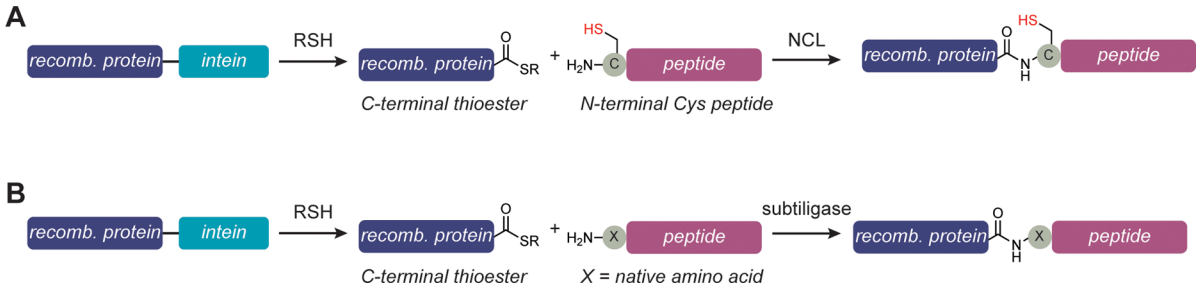


Figure 18. Subtiligase-catalyzed expressed protein ligation. (A) Expressed protein ligation scheme, requiring a C-terminal thioester and a peptide substrate with an N-terminal Cys. (B) Subtiligase-catalyzed expressed protein ligation, requiring a C-terminal thioester and a peptide substrate with any N-terminal residue that is compatible with subtiligase or subtiligase mutant prime-side specificity.

can be used to select the most efficient subtiligase variant for modification of user-specified sequences.

The subtiligase specificity mutant Y217K was applied for high-yield protein bioconjugation to recombinant antibodies,[34] an important class of therapeutic proteins. Recombinant antibodies for a particular antigen can be selected by phage display of a library of Fab (Fragment antigen binding) variants to yield high affinity binding reagents.[140,141] The most commonly applied recombinant antibody phage display libraries are based on the trastuzumab scaffold and therefore have common N-terminal sequences (Glu-Ile on the heavy chain and Ser-Asp on the light chain).[140] Acidic N-terminal residues are also commonly observed in naturally occurring antibodies, making these sequences particularly interesting to target. The Y217K mutant was selected based on PILS specificity data to selectively modify the heavy chain of an αGFP antibody selected by phage display.[34] While wild-type subtiligase was unable to measurably modify this antibody, the Y217K mutant quantitatively modified the N terminus of the heavy chain specifically. Efforts to specifically modify the light chain with an appropriate subtiligase specificity mutant (F189R), however, failed to produce the conjugation product in high yield. However, by extending the Ser-Asp N terminus with a glycine linker, it was possible to rescue modification, suggesting that the light chain is simply inaccessible to bind subtiligase.

7.3. Peptide Ester Substrates for One-Step and Modular Protein Bioconjugation

To further expand the utility of subtiligase and its variants for N-terminal modification, a set of peptide ester substrates were developed to enable both one-step and modular protein bioconjugation (Figure 17).[34] For the one-step approach, a peptide ester substrate, succinyl-KAAPF-glc-F-amide, was synthesized (Figure 17A). Because this substrate has a blocked N terminus and a single free Lys residue, it can be modified with any NHS ester reagent, enabling proteins to be modified site-specifically in one step with a reagent that would otherwise be nonspecific and target all surface-accessible Lys residues.[130] Many NHS ester reagents are commercially available and can be converted to site-specific reagents in this way, making this a general strategy for one-step modification with diverse payloads. This approach was demonstrated by modifying succinyl-KAAPF-glc-F-amide with NHS biotin and then ligating the resultant peptide onto the N terminus of a recombinant antibody to achieve quantitative biotinylation.

For the modular approach, a peptide ester substrate that is N-terminally capped with an azidoacetyl group, N_3Ac-AAPF-glc-F-amide, was synthesized (Figure 17B).[34] This substrate enables modification of proteins with an azide group that can be subsequently derivatized with copper-catalyzed or copper-free click chemistry[142] using alkyne or dibenzocyclooctyne (DBCO) reagents, respectively, or with Bertozzi–Staudinger ligation using phosphine reagents.[143] To demonstrate this

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127–3160

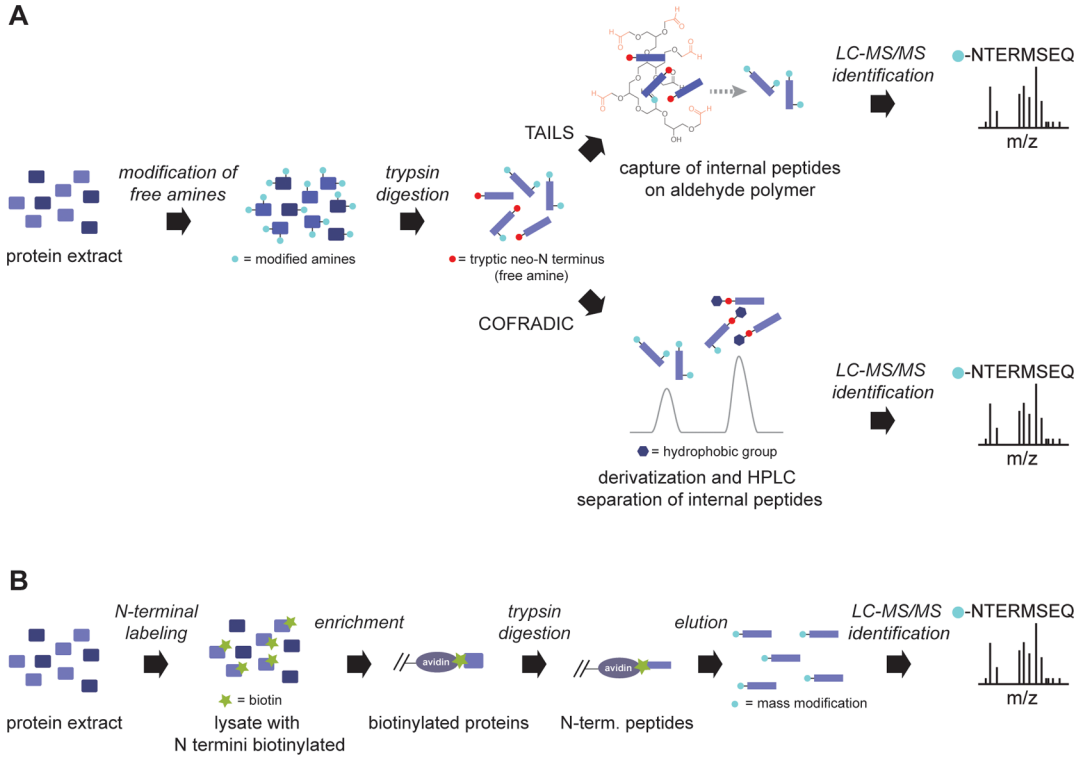


Figure 19. Comparison of N-terminomics strategies. (A) Depletion of internal peptides. Top: TAILS involves modification of all free amines in a protein extract, followed by digestion with trypsin and depletion of tryptic neo-N termini using an aldehyde-derivatized polymer. Bottom: COFRADIC involves modification of all free amines in a protein extract, followed by digestion with trypsin and derivatization of the tryptic neo-N termini with a hydrophobic reagent that changes their HPLC retention time compared to other N-terminal peptides. (B) Positive enrichment N-terminomics. N termini in a protein extract are specifically biotinylated using either subtiligase or a 2-pyridinecarboxaldehyde reagent, enabling their enrichment.

approach, a recombinant antibody was modified with the azide-bearing peptide and then used a starting material for modification with a variety of payloads, including a Cy3 fluorophore, the cytotoxic drug monomethyl auristatin E (MMAE), a DNA barcode, and a PEG polymer.[34] Importantly, the recombinant antibody retained its affinity and specificity following modification.

7.4. Subtiligase-Catalyzed Expressed Protein Ligation

A complementary approach to subtiligase-based methods for protein bioconjugation and protein semisynthesis is expressed protein ligation (EPL), a method in which a protein fragment expressed as a fusion protein with an intein can be converted to a C-terminal thioester (Figure 18A).[144−146] This C-terminal thioester can then undergo thiotransesterification with the side chain thiol of an N-terminal Cys residue on a peptide or protein, followed by an S-to-N acyl shift to generate a native peptide bond. An advantage of EPL is that it enables the N-terminal fragment to be recombinantly expressed. However, a disadvantage is that Cys, one of the least frequently observed residues in proteins, is required at the N terminus of the C-terminal fragment, often requiring the use of a non-native sequence junction. A method for subtiligase-catalyzed EPL was developed to combine the broad sequence specificity of subtiligase with the applicability of EPL to recombinant N-terminal fragments, enabling the synthesis of native proteins without the requirement for Cys (Figure 18B).[147] In this method, a recombinant protein thioester is produced as in EPL. This protein thioester can then be used as the acyl donor substrate for subtiligase in place of a peptide thioester, enabling ligation with any C-terminal fragment whose N-terminal sequence is a good prime-side substrate for subtiligase.

This approach was initially explored by testing the ability of subtiligase to catalyze ligation of ubiquitin (Ub) C-terminal thioester with a synthetic biotinylated peptide.[147] Moderate yields (~50%) of ligation product were obtained using wild-type Ub, which has a C-terminal sequence (P4−P1) of LRGG, and the synthetic decapeptide GLSGRGKGGK(biotin). A number of Ub variants and synthetic peptides were also tested to explore the influence of the identity of the P4, P1, P1′, and P2′ residues on ligation efficiency. While most combinations gave ligation yields in the range of 50−70%, decreases in yield were observed when the P4 or P1′ positions were Asp or Glu and when the P1′ position was Pro, in agreement with previous subtiligase specificity studies. By using subtiligase-E156Q/G166K or subtiligase-Y217K, the ligation efficiency could be improved when the P1 residue was Asp or Glu. Similarly, subtiligase-Y217K gave improved ligation yields with P1′ Glu. Similar results were obtained with C-terminal thioesters of glutathione-S-transferase.

Based on these promising results with model proteins, subtiligase-catalyzed expressed protein ligation was applied to investigate the role of C-terminal phosphorylation on the lipid phosphatase PTEN, a tumor suppressor. PTEN is tetraphosphorylated near its C-terminus, and a tetraphosphorylated variant was previously generated using EPL.[148,149] However, construction of this variant required the introduction of an unnatural Cys mutation, Y379C, which alters PTEN's behavior in cells. Application of subtiligase-catalyzed EPL enabled synthesis of a fully native, tetraphosphorylated PTEN that was

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

used a standard to more accurately measure the stoichiometry of PTEN phosphorylation in cells.[147] This demonstrated the potential of subtiligase-catalyzed EPL as a tool for biology, as native, tetraphosphorylated PTEN is not currently synthetically accessible by other methods.

Subtiligase-catalyzed protein bioconjugation and protein semisynthesis are advantageous because, in principle, they do not require genetic engineering of the protein of interest and can maintain a fully native protein sequence. The main remaining limitation for the application of subtiligase in N-terminal bioconjugation is the need for the N terminus to be accessible and in an extended conformation, a requirement that has also been observed for other N-terminal modification techniques. Advances in protein engineering of subtiligase for altered specificity and stability have made selection of ligation junctions even more flexible and are likely to drive wider application of subtiligase for site-specific protein modification. For example, it has been suggested that subtiligase could be a useful tool for incorporation of isotopically labeled segments into proteins for NMR studies,[150] but so far, this application has not been demonstrated. However, based on its broad sequence compatibility, site-specificity, and ease of use, it is likely that subtiligase will be more widely adopted for this and other protein bioconjugation applications in the future.

8. APPLICATION OF SUBTILIGASE FOR CELLULAR N TERMINOMICS

The protein N terminus is a key site of post-translational modifications that regulate a wide array of biological processes. Modifications such as acetylation, myristoylation, and ubiquitylation regulate protein stability and protein localization,[151] while proteolysis regulates numerous biological processes.[152] Global profiling of the pool of cellular N termini and their modification states can therefore provide critical insights into how proteins are modified to enable or alter their biological functions. Enabled by advances in mass spectrometry-based proteomics,[153] a number of methods have been developed for enrichment of N-terminal peptides to probe their roles in biology. Some approaches rely on depletion or separation of internal peptides following protease digestion (Figure 19A),[154−156] while others rely on specific modification and positive enrichment of unblocked N termini (Figure 19B).[35,157] The combined fractional diagonal chromatography (COFRADIC) approach involves acetylation of all lysine ε-amines and N-terminal α-amines in intact proteins isolated from a sample of interest before digestion with trypsin (Figure 19A, bottom).[154,155] After trypsin digestion, the samples are fractionated by reverse-phase HPLC and modified with 2,4,6-trinitrobenzenesulfonic acid (TNBS). Based on this modification, new, unblocked N termini generated by trypsin digestion acquire a hydrophobic trinitrophenyl group that enables chromatographic separation of internal peptides generated by trypsin digestion from native unblocked protein N-terminal peptides, which were previously acetylated. The isolated N-terminal peptides can then be sequenced by LC-MS/MS to identify the position of the protein N terminus. Another approach, Terminal Amine Isotopic Labeling of Substrates (TAILS), uses a similar chemical modification strategy to separate protein N termini from internal tryptic peptides (Figure 19A, top).[156] In TAILS, the proteome of interest is subjected to reductive dimethylation to block lysine ε-amines and N-terminal α-amines. Proteins are then digested with trypsin, and the resulting peptides are treated with a hyperbranched polyglycerol-aldehyde polymer that reacts with unblocked N termini generated by trypsin cleavage in the presence of sodium cyanoborohydride. N-Terminal peptides can then be collected by filtration and analyzed by LC-MS/MS. Positive enrichment approaches for N terminomics present the challenge that N-terminal α-amines must be selectively targeted for modification over lysine ε-amines, despite the chemical similarity of these functional groups (Figure 19B).[35,135,157,158] Subtiligase meets this challenge by taking advantage of enzyme−substrate molecular recognition to position only the substrate N terminus, and not nearby side chains, for enzymatic modification.[35] Using a biotinylated peptide ester substrate for subtiligase, N termini can be selectively modified and enriched for sequencing by LC-MS/MS. A nonenzymatic strategy for positive enrichment N terminomics, chemical enrichment of protease substrates (CHOPS), that takes advantage of a selective method for N-terminal modification was recently reported.[157,159] In CHOPS, a 2-pyridinecarboxaldehyde-biotin probe reacts with the N-terminal α-amine to form an imine, which then cyclizes with the amide nitrogen from the next peptide bond. Because lysine side chains do not have this amide nitrogen, this method is specific for N-terminal modification. Biotinylated N-terminal peptides can then be enriched and sequenced by LC-MS/MS. The positive enrichment approaches have the advantage of high sensitivity for low-abundance N termini, permitting the identification of proteolytic cleavage events that target only a small fraction of cellular protein but may impact biology by potentiating gain-of-function phenotypes.

8.1. Subtiligase N Terminomics Workflow

The use of subtiligase as a tool for N terminomics was pioneered in the context of global profiling of proteolytic signaling pathways.[35] Proteolysis is a key post-translational modification that regulates many biological processes involved in human health and disease, including viral infection, cancer progression, organismal development, and neurodegeneration, among many others. More than 500 proteases are encoded in the human genome, comprising ∼2% of the human proteome.[160] Although proteolysis can program complex, emergent biological phenotypes, the universal molecular consequence of proteolysis is always the production of a new, unblocked N-terminal α-amine that represents a potential substrate for subtiligase. Complex mixtures such as cell lysate or plasma can therefore be treated with subtiligase to selectively modify free N termini with a peptide ester substrate designed to facilitate isolation of N-terminal peptides and their subsequent identification by LC-MS/MS.

The first step in the subtiligase N terminomics workflow is selective biotinylation of free N termini in a biological sample using a peptide ester substrate designed to facilitate isolation and identification of N-terminal peptides.[35,161] Peptide esters used for N terminomics have typically included three key features: (i) a biotin handle for affinity purification of modified proteins; (ii) a TEV protease cleavage site for selective elution of N-terminal peptides; and (iii) a mass tag such as Ser-Tyr or the unnatural amino acid aminobutyric acid (Abu) that is retained on the N terminus of modified peptides after elution and enables positive identification of substrates that were modified by subtiligase.[35,77] Following subtiligase treatment, biotinylated proteins can be isolated on immobilized avidin and digested with trypsin to remove internal peptides. The captured N-terminal peptides can then be selectively eluted by

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

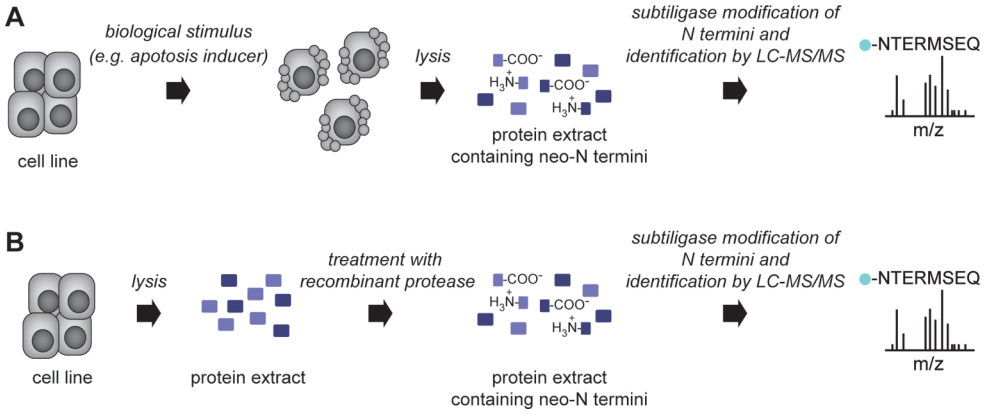


Figure 20. Forward and reverse strategies for subtiligase N terminomics. (A) In a forward N terminomics experiments, a biological stimulus is applied to live cells, and protein is then harvested for subtiligase N terminomics. (B) In a reverse N terminomics experiment, a protein extract is treated with a recombinant protease and then labeled with subtiligase for N terminomics analysis.

cleavage with TEV protease, leaving a mass modification on the N terminus. This pool of eluted peptides can then be sequenced by LC-MS/MS to enable identification of the prime side of a protease cleavage site. The residues on the nonprime side of the cleavage site can be inferred based on the protein sequence preceding the peptide that was modified by subtiligase. To validate this method, subtiligase was used to enable analysis of endogenous N termini in unperturbed Jurkat cells. Using this strategy, 661 unique N termini were sequenced using LC-MS/MS. Of these N termini, 49% either were annotated in SwissProt as native protein N termini or corresponded to cleavage within the first 50 residues of the protein, consistent with a signal or transit peptide processing event. The remaining 51% of N termini were cleaved outside the first 50 residues and potentially correspond to the products of constitutive protein degradation. The subtiligase N terminomics technology therefore enables global sequencing of proteolytic cleavage sites with single amino acid resolution and has been used to probe the role of proteolysis in a number of different biological processes.

Two basic experimental strategies for subtiligase N terminomics, termed "forward" and "reverse", have been applied to identify protease substrates in biological samples (Figure 20).[35,161,162] In forward experiments, a proteolysis-inducing stimulus is applied to a biological system under physiological conditions (Figure 20A).[35] Following the stimulus period, protein is isolated and N-terminally biotinylated with subtiligase to enable isolation and sequencing of N-terminal peptides. N termini identified in the stimulated sample are then compared to those identified in an unstimulated control to determine which N termini may be biologically relevant protease substrates. In reverse experiments, a protein extract derived from a biological system of interest in which endogenous proteases have been inactivated is treated in vitro with an exogenous protease (Figure 20B).[162] After inactivation of the exogenous protease, the protein extract is then N-terminally biotinylated with subtiligase to enable identification of substrates of the exogenous protease. The forward experimental approach has the advantage that proteolysis occurs under in vivo conditions, in which cellular structure and organization are preserved, but has the limitation that it does not allow the specific proteases that catalyze observed cleavage events to be identified. Reverse experiments enable identification of candidate substrates for a protease of interest but are carried out in the context of lysate, in which cellular structure is disrupted; identified substrates therefore may not be physiologically relevant. The forward and reverse approaches are therefore complementary strategies that, when used together, can provide insights into which substrates are biologically relevant and which proteases cleaved them in various biological contexts.

Several different approaches may be taken to identify candidate protease substrates in subtiligase N terminomics data sets. The first approach is nonquantitative and considers candidate substrates to be N-terminal peptides that are identified by LC-MS/MS in the experimental condition but not the control condition.[35] This approach can be combined with information about the sequence motif recognized by the protease of interest, if known, to further filter the list of candidate substrates whose cleavage is induced upon the application of a stimulus. A limitation of this nonquantitative approach is that small amounts of background cleavage (for example, arising from a small number of apoptotic cells in a healthy culture) can lead to identification of *bona fide* substrates in the control condition, which may then be discounted even though they are of biological interest. To overcome this drawback, quantitative proteomics strategies have been applied for both discovery and validation of protease substrates. The stable isotope labeling by amino acids in cell culture (SILAC) approach[163] has been applied in both forward and reverse subtiligase N terminomics experiments for quantitative comparison of the abundance of each N-terminal peptide in the experiment and control conditions, allowing cleavage events that occur at a background level but are induced upon stimulus to be identified.[164,165] Selected reaction monitoring (SRM)[166,167] has similarly been used to validate candidate substrates and quantify known substrates consistently across many samples.[162,168−171] The quantitative approaches in principle have the advantages that (1) no advance knowledge of the sequence motif(s) that are cleaved upon application of a particular biological stimulus is needed, although they have also been combined with this information, and (2) false negatives based on removal of all identified background N-terminal peptides as candidates can be avoided. Similar quantitative approaches, such as isobaric tags and parallel reaction monitoring (PRM), have not yet been widely applied in subtiligase N terminomics, but they are compatible with the method and are likely to be adopted in the future.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

8.2. Biological Applications of Subtiligase N Terminomics

8.2.1. Apoptotic Proteolysis. Apoptosis is a form of programmed cell death that plays a critical role in tissue development, homeostasis, and cellular stress response.[172,173] Based on its function in cellular turnover, apoptosis can act as a check on the uncontrolled cell growth of tumor cells and is therefore of great interest in the cancer biology field.[174] One hallmark of apoptosis is the induction of widespread proteolytic cleavage mediated by the caspases (cysteine aspartate proteases), a family of dimeric proteases that specifically cleave substrates with an Asp in the P1 position.[175−177] Apoptosis may be induced by extracellular ligands via the extrinsic pathway or by environmental stresses such as cytotoxic compounds via the intrinsic pathway. Induction by either pathway leads to activation of the executioner caspases, caspase-3 and caspase-7. These proteases catalyze a large number of proteolytic cleavage events that largely serve to inactivate prosurvival proteins and activate antisurvival proteins, ultimately leading to cell death and removal of the apoptotic bodies by phagocytes. Because of the large number of known apoptotic substrates, as well as the potential for discovery of new substrates, apoptotic proteolysis has served as a "proving ground" for proteomic techniques for global N-terminal sequencing, and subtiligase N terminomics has been an important tool for defining the molecular events of apoptosis.

Subtiligase N terminomics was first applied to profile proteolytic cleavage events that occurred in Jurkat T cells upon induction of apoptosis with the chemotherapeutic drug etoposide,[35] a topoisomerase II inhibitor that induces apoptosis via the intrinsic pathway.[178,179] Following etoposide treatment, N-terminal peptides were isolated from cell lysate using the subtiligase strategy and identified by LC-MS/MS.[35] This analysis revealed a strong induction of cleavage events with an Asp at the P1 position, a signature of caspase cleavage. While 43% of N-terminal peptides in the etoposide-treated sample corresponded to cleavages with Asp at P1, only 3% of cleavages identified in untreated Jurkat cells had P1 Asp. Of this 3%, 55% corresponded to known caspase substrates, suggesting that these substrates were mainly derived from a small fraction of apoptotic cells in the untreated culture condition. The N-terminal data set from the apoptotic sample consisted of 1099 peptides modified with Ser-Tyr dipeptide that is the remnant of TEV protease elution in the subtiligase workflow. Of these, 418 corresponded to cleavage events with Asp at the P1 position, with 333 of these representing unique N termini derived from 282 substrate proteins. Cleavage of 16 of the identified P1 = Asp substrates was orthogonally validated by Western blotting. As further confirmation of the role of caspases in substrate cleavage, proteolysis of a representative panel of the identified substrates could be blocked by treatment of the cells with carboxybenzyl-Val-Ala-Asp(OMe)-fluoromethyl ketone (zVAD(OMe)-fmk), a cell-permeable covalent pan-caspase inhibitor. The large scale of the data set collected enabled bioinformatic analysis of sequence and structural determinants of caspase specificity.[35,180] A sequence logo generated from these results recapitulated the canonical P1-DEVDG-P1′ motif for caspases that was initially derived from peptide-level studies of caspase specificity.[69,181,182] Interestingly, very few of the identified substrates from the subtiligase N terminomics data set contain this exact sequence motif. This result highlights the limited translatability of the efficiency of peptide substrate cleavage for predicting physiological protease substrates and the utility of subtiligase N terminomics for this purpose. Additional bioinformatics analyses revealed that substrates of etoposide-induced proteolysis are often cleaved in surface accessible loop regions and that many of the substrates physically interact with one another and function in the same biological pathways. This result suggests that caspases target specific protein networks to execute apoptosis. Consistent with this hypothesis, subtiligase N terminomics studies of apoptosis in mouse and fly cell lines, as well as in the context of whole nematodes, demonstrated that the pathways that are targeted by proteolytic cleavage are more conserved than the specific protein substrates or exact proteolytic cleavage sites.[183]

A comparative analysis of the subtiligase N terminomics technology with an alternative approach for identifying protease substrates, Protein Topography and Migration Analysis Platform (PROTOMAP), revealed that the two techniques are largely in agreement.[184,185] However, the two methods provide complementary insights into proteolytic cleavage events. Both the subtiligase method and the PROTOMAP method have been applied to profile apoptotic proteolysis in Jurkat T cells, providing data sets that were used for comparative analysis.[35,184] The PROTOMAP approach relies on detection of changes in 1D gel migration of proteins in combination with LC-MS/MS to globally profile proteolysis in biological systems.[186] While the subtiligase technique can be used to systematically identify the exact sites of proteolytic cleavages, exact cleavage sites can only be identified serendipitously using PROTOMAP. Additionally, low-magnitude cleavage events in abundant proteins are sometimes obscured using the PROTOMAP method because abundant proteins tend to be poorly resolved and span multiple gel bands. Therefore, the subtiligase technique has an advantage in identifying proteins that are only cleaved to a small extent. However, the PROTOMAP method has the advantage that it does not rely on detection of a single N-terminal peptide and therefore can often identify protease substrates that are missed by the subtiligase technique because not all peptides are detectable by LC-MS/MS. PROTOMAP also provides information about the extent to which protease substrates are cleaved, which is useful for interpreting the potential functional consequences of proteolytic cleavage in the context of how much of a particular substrate is cleaved. For example, if only a small amount of a particular substrate is cleaved, this could be sufficient to trigger a gain-of-function phenotype but would likely be insufficient for loss of function. The subtiligase N terminomics approach and gel-based methods for mapping proteolysis such as PROTOMAP are therefore highly complementary and can be integrated to gain more comprehensive insights into biological proteolysis.

Forward subtiligase N-terminomics experiments have also been used to profile apoptotic proteolysis induced by a number of stimuli other than etoposide, including staurosporine and the clinically relevant chemotherapeutic agents doxorubicin and bortezomib.[162,168−170,187] While the cellular response to these agents varies widely according to cell type and which agent is used, a largely consistent cohort of caspase substrates was identified in all of these experiments.[168] However, the rates at which the substrates were cleaved and the abundances of the resultant N termini dramatically varied across cell type and apoptotic stimuli. This result supported the hypothesis that there are conserved apoptotic nodes that are critical for the

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

execution of cell death. However, it also suggested that there may be proteolytic fingerprints that correspond to specific drug treatments.[170] Remarkably, some of these caspase-cleaved fragments were found to be elevated in postchemotherapy patient plasma samples using subtiligase N terminomics, indicating that proteins cleaved during apoptosis are released from cells and could potentially be used as novel biomarkers for cell death and chemotherapeutic efficacy. Notably, the subtiligase N terminomics approach reduces sample complexity avoiding the interference from high-abundance plasma proteins that is often problematic in traditional proteomic analysis of plasma.[188]

As a complementary approach to forward N terminomics experiments for profiling apoptosis, a number of reverse subtiligase N terminomics studies of individual apoptotic caspases have been performed.[162,171] Reverse N terminomics studies with subtiligase have relied on the quantitative SRM approach to monitor cleavage of caspase substrates over time, enabling determination of k_{cat}/K_M for each caspase−substrate pair.[162] Reverse N terminomics studies of caspases-3, -7, -8, and -9 using SRM enabled comparison of the rates at which individual caspases cleave substrates to the rates at which they are cleaved when apoptosis is induced with staurosporine (intrinsic pathway) or TRAIL (extrinsic pathway).[162] Comparison of these forward and reverse data sets made it possible to infer protease−substrate pairs and revealed that certain biological pathways are targeted more rapidly than others during apoptosis. A later SRM study of caspases-2 and -6, and comparison to earlier caspase-3, -7, -8, and -9 data sets, showed that many caspases cleave common substrates, but at vastly different rates.[171] This information suggested which caspase−substrate pairs are likely to be relevant in a physiological context.

8.2.2. Inflammatory Proteolysis. The inflammatory caspases-1, -4, -5, and -11 are activated in response to danger signals arising from microbial or viral infection, or other pro-inflammatory stimuli.[172,177,189] Subtiligase N terminomics, in both forward and reverse modes, was applied to identify substrates of caspases-1, -4, and -5.[162] In the forward experiments, a SILAC approach was used to identify proteolytic cleavage events induced by mimics of gout, bacterial infection, or viral infection in THP-1 monocytes. These results were compared to reverse data sets generated by treating THP-1 lysate with recombinant caspases-1, -4, and -5. Interestingly, fewer substrates were identified in forward experiments compared to reverse experiments, suggesting that cellular structure and substrate accessibility play a major role in determining which substrates are cleaved. These studies identified gasdermin D as one of the most efficient substrates of caspase-1. Later studies confirmed that gasdermin D is cleaved in response to inflammation and demonstrated that its N-terminal cleavage product triggers pyroptosis, an inflammatory form of cell death.[190−192] These results highlight the power of subtiligase N terminomics for identifying biologically relevant protease cleavage events.

8.2.3. Viral Protease Substrate Identification. Zika virus is a mosquito-borne virus that is capable of causing birth defects such as microencephaly and brain malformations, and it can lead to Guillain-Barré syndrome in infected adults.[193,194] Subtiligase N terminomics was applied to delineate how Zika virus protease (ZVP) affects host cell pathways using a reverse experimental design. From a cell lysate treated with Zika virus protease, 31 candidate substrates were identified.[165] Among these were well-known viral protease substrates, including eIF4g, a protein that is required for cap-dependent translation and which is cleaved by many viral proteases.[195−200] Additionally, protease substrates that are unique to Zika virus protease were identified, and a subset of these were confirmed in ZVP-treated lysate as well as Zika-infected cells.[165]

8.2.4. Bacterial Protease Substrate Identification. Secreted bacterial proteases have long been known to support bacterial pathogenesis by facilitating adherence to epithelial cells, cleaving extracellular host proteins, permeabilizing the epithelial barrier, and promoting evasion of host innate and adaptive immunity.[201−203] Members of the clostripain-like (C11) protease family are enzymes commonly secreted by commensal bacteria of the human gut microbiome that are implicated in biological and pathogenic processes.[204,205] To begin to elucidate the function of the C11 proteases at the host−microbiome interface, a reverse subtiligase N terminomics approach was applied to identify substrates of the C11 protease PmC11 from *Parabacteroides merdae* in human epithelial cell lysates.[206] This experiment identified 56 candidate PmC11 substrates and confirmed that PmC11 has stringent substrate specificity and cleaves peptides with Gly, Ser, Thr, or Arg at the P2 position and Lys or Arg at the P1 position, in agreement with an in vitro peptide profiling experiment that was performed in parallel. Elucidation of PmC11 specificity enabled design of a specific peptide acyloxymethyl ketone probe for proteases of this class, enabling the study of protease function in complex samples such as the human gut microbiome. Subtiligase N terminomics therefore represents an important tool for understanding and probing protease function in the context of human commensal bacteria.

8.2.5. Mitochondrial N Terminomics. Target peptides are short amino acid sequences that direct protein transport to specific intracellular compartments, including the nucleus, mitochondrion, endoplasmic reticulum, plasma membrane, and peroxisome.[207−209] Many target peptides, such as those encoding secretion and mitochondrial transport, are found at the protein N terminus and are cleaved after the targeted protein has reached its destination. A forward subtiligase N terminomics experiment was performed on mitochondria isolated from mouse liver and kidney to elucidate the mitochondrial N-terminal proteome, shedding light on mitochondrial import and protein processing.[210] This study revealed that mitochondrial targeting peptides are poorly conserved on the sequence level but that features such as net charge (+3 to +6), length (20−60 amino acids), and the presence of stabilizing N-terminal amino acids based on the bacterial N-end rule are highly conserved. These results were largely in agreement with previous N terminomics studies of yeast and human mitochondria that utilized alternative N terminomics techniques.[211,212] Subtiligase therefore represents a generalizable tool for studying cleavage of organellar targeting sequences.

8.2.6. N-Terminal Protein Acetylation. Subtiligase is useful for studying proteolysis based on its ability to modify free N termini in a cellular context. Therefore, in principle, subtiligase could be applied to monitor whether specific N termini are blocked by post-translational modifications based on whether they are substrates for modification. This approach was applied to study the N-terminal acetylation status of a specific group of proteins in the context of overexpression of the antiapoptotic protein Bcl-xL.[213] This assay for N-terminal

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

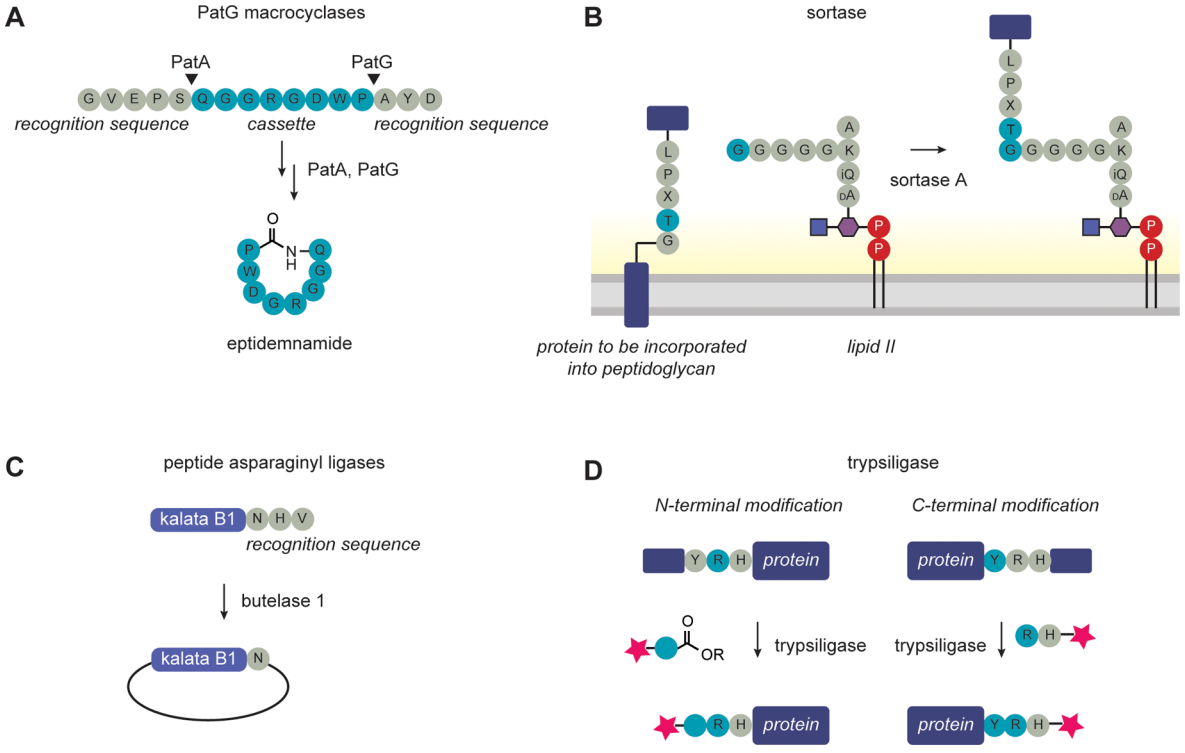


Figure 21. Other natural and engineered peptide ligases. (A) PatG macrocyclases. (B) Sortase. DA, D-Ala; iQ, iso-Gln. (C) Peptide asparaginyl ligases. (D) Trypsiligase.

acetylation is based on the premise that because 80−90% of cytosolic N termini are blocked by post-translational modifications,[214] proteins that are biotinylated by subtiligase following a biological perturbation have specifically lost N-terminal acetyl modifications. This technique was validated for a specific group of proteins that are normally acetylated in the context of RNAi knockdown of the N-terminal acetyltransferase NatA. The method was then used to demonstrate that overexpression of Bcl-xL reduces cellular N-terminal acetylation of specific proteins, connecting apoptotic sensitivity to the available pool of acetyl-CoA and therefore cellular metabolism. In this study, Western blotting for the biotin group introduced by subtiligase modification rather than mass spectrometry-based proteomics was used to monitor the acetylation status of a small number of proteins. Therefore, an interesting future direction will be the application of the subtiligase N terminomics workflow to globally monitor protein acetylation status in this context.

9. COMPARISON OF SUBTILIGASE TO NATURALLY EVOLVED AND ENGINEERED PEPTIDE LIGASES

Since subtiligase was first reported, several classes of naturally evolved and engineered peptide ligases have been discovered. In general, all classes have a catalytic mechanism that is analogous to that of the Ser and Cys proteases.[22] An acyl donor substrate is attacked by the catalytic nucleophile to form an acyl- or thioacyl-enzyme intermediate. This intermediate can then either be hydrolyzed, as would be the case in a protease, or it can be intercepted by an α-amine nucleophile to form a new peptide bond. While the aminolysis reaction is very inefficient under physiological conditions in the proteases, the peptide ligase enzymes support varying ratios of ligation to hydrolysis, enabling efficient synthesis of peptide bonds. Each of these peptide ligases has been developed for technological applications to differing degrees, and each has its own set of advantages and limitations.

9.1. Discovery of Naturally Occurring Subtilisin-like Peptide Ligases

Interestingly, nearly two decades after subtiligase was engineered, a subtilisin-like enzyme that catalyzes peptide ligation in the context of N-to-C peptide macrocyclization, PatG, was discovered in nature (Figure 21A).[215,216] Since this initial discovery, PatG-like enzymes have been implicated in the biosynthesis of several families of ribosomally synthesized and post-translationally modified peptide (RiPP) natural products, including cyanobactins and cyclotides.[217,218] In contrast to subtiligase, PatG macrocyclases retain the canonical Asp-His-Ser catalytic triad characteristic of subtilisin proteases.[215,216,219] These enzymes catalyze proteolytic cleavage of a C-terminal recognition sequence in tandem with peptide macrocyclization. X-ray crystallographic studies of PatG revealed several differences between subtilisin-like macrocyclases and subtilisin proteases.[219] While subtilisin proteases bind their substrates in an extended conformation, PatG macrocyclases bind their substrates in a bent configuration that requires a *cis* conformation of the peptide bond between the P1 and P2 residues. Because this conformation is energetically inaccessible for 19 of the 20 natural L-amino acids, PatG's substrate scope is limited to peptides that contain P1 proline or peptides that have been post-translationally modified by heterocyclization of cysteine, serine, or threonine residues to form P1 thiazole, thiazoline, or oxazoline residues. In contrast to subtiligase, PatG does not employ a serine to cysteine substitution to favor peptide bond formation over hydrolysis. Rather, the PatG enzymes contain a conserved helix−loop−helix insertion that appears to shield the acyl−enzyme intermediate from attack by water. Deletion of this segment

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

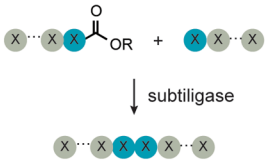
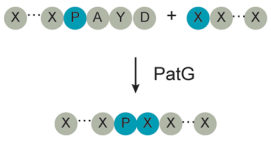
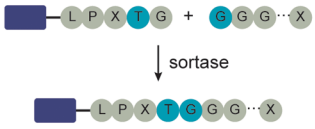
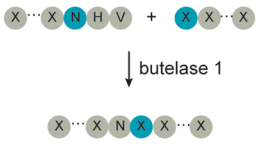
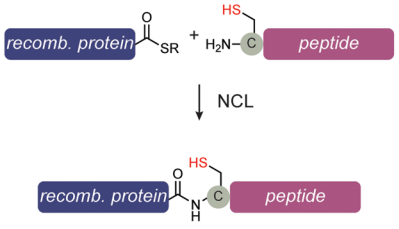
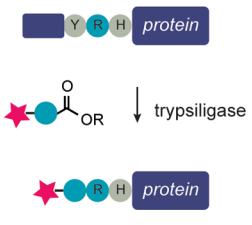
Peptide ligation enzyme / strategy	*Advantages*	*Limitations*
subtiligase 	• recombinantly expressed • fast reaction times • high yields possible • catalytic amount of enzyme required (<1 μM) • broad sequence compatibility for both acyl donor and acyl acceptor	• requires C-terminal ester or thioester • secondary structural requirements • yield depends on N-terminal accessibility
PatG-like enzymes 	• recombinantly expressed	• slow reaction kinetics • secondary structure requirements not well characterized • requires P1 Pro, thiazole, thiazoline, or oxazoline
sortase A 	• recombinantly expressed • fast reaction times • high yields possible	• engineered sequences required • large enzyme amounts required • incompatible with non-recombinant proteins • yield depends on N-terminal accessibility
butelase 1 	• fast reaction times • high yields possible • broad sequence compatibility for acyl acceptor • catalytic amount of enzyme required(<1 μM)	• requires P1 Asn/Asp • requires thiodepsipeptide substrate to achieve high yield • yield depends on N-terminal accessibility • secondary structure requirements not well characterized • enzyme is isolated in low yield from plant material
native chemical ligation / expressed protein ligation 	• high yields possible • broad sequence compatibility for acyl donor • substrates may be genetically encoded	• requires N-terminal Cys on acyl acceptor • requires C-terminal thioester on acyl donor • incompatible with non-recombinant proteins • long reaction times (1-3 d)
trypsiligase 	• can be used for N- or C-terminal protein modification • high yields possible • fast reaction times	• YRH sequence required • secondary structure requirements not well characterized

Figure 22. Comparison of peptide ligation enzymes and strategies.

produces a variant of PatG that maintains protease activity but no longer catalyzes peptide macrocyclization. PatG enzymes and subtiligase therefore represent different modes by which the subtilisin protease scaffold can support peptide bond formation activity. While the PatG-like enzymes represent potentially promising additions to the toolbox of peptide ligation catalysts, so far their application has been limited by low catalytic activity and high P1 sequence specificity (Figure 22). However, PatG suggests ways to further improve the subtiligase ligation-to-hydrolysis ratio by shielding the active site from water.

9.2. Bacterial Transpeptidases

Staphylococcus aureus sortase A is a peptide ligase enzyme that catalyzes transpeptidation of an LPXTG peptide acceptor and the pentaglycine N-terminal extension of lipid II in bacterial cell wall biosynthesis (Figure 21B).[220] Although sortase A is

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

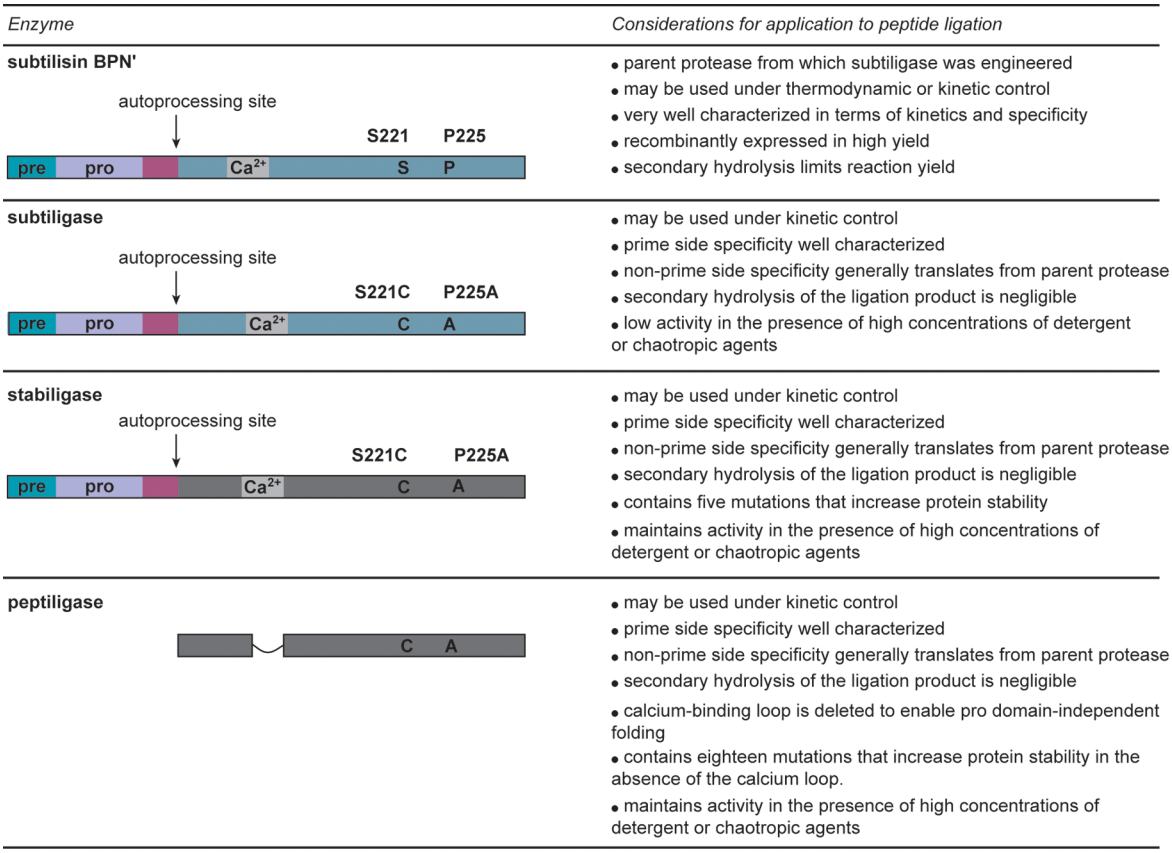


Figure 23. Comparison of subtilisin, subtiligase, stabiligase, and peptiligase.

evolutionarily unrelated to subtilisin and subtiligase, it shares a similar catalytic mechanism and catalytic triad.[221−224] Sortase A first cleaves an LPXTG motif between Thr and Gly to form a thioacyl−enzyme intermediate on an active site cysteine, which is then attacked by the N terminus of pentaglycine to form a peptide bond. Based on its ability to catalyze this reaction, sortase A has been widely applied for site-specific protein modification (Figure 22),[225−228] as well as peptide and protein cyclization.[229,230] While this enzyme has been employed for modification of recombinant proteins, it is particularly useful in the cellular context, because both substrates can be genetically encoded, and its high sequence specificity ensures modification of only the protein of interest.[231−234] This sequence specificity, however, is a limitation to the application of sortase A in situations where it is desirable to preserve native protein sequence. Sortase A has a high ligation to hydrolysis ratio, but this leads to the drawback that a stable acyl−enzyme intermediate can form, limiting the catalytic turnover of the enzyme.[223] However, because sortase A can be recombinantly expressed, protein engineering can be applied to address these limitations.[235,236] Phage display approaches have been applied to enhance sortase A catalytic efficiency by up to 140-fold, with a k_{cat}/K_M of 28,000 M^{-1} s^{-1} (about 10-fold below subtiligase).[235] A similar approach was used to change sortase sequence specificity for recognition of LAXTG and LPXSG motifs.[236] Sortase A and its biological applications were recently reviewed in ref 5.

9.3. Butelase 1 and Asparaginyl Endopeptidase-like Enzymes

Butelase 1 is an Asp/Asx-specific peptide ligase that was recently isolated from *Clitoria ternatea*, a cyclic peptide-producing plant (Figure 21C).[6] Butelase 1 is highly homologous to and shares the same catalytic triad with the asparaginyl endopeptidase (AEP) family of cysteine proteases, but it exhibits no hydrolase activity.[6,237,238] Instead, butelase 1 cleaves an Asn-His-Val motif between Asn and His to form a thioacyl−enzyme intermediate. This intermediate can then be intercepted by the N-terminal α-amine of a peptide to form an amide bond. Although butelase 1 is the best characterized AEP-related peptide ligase, many plant species of the Rubiaceae, Violaceae, Fabaceae, Solanaceae, and Cucurbitaceae families produce cyclotides whose precursor peptides contain a conserved C-terminal Asp or Asn residue, leading to the hypothesis that AEP-like enzymes also catalyze their cyclization.[239,240] Therefore, there is a high potential for discovery of new enzymes in this family. Indeed, a second AEP-like macrocyclase enzyme, OaAEP1, from the cyclotide producing plant *Oldenlandia affinis*, was recently discovered and characterized.[240] While OaAEP1 can be recombinantly expressed, its catalytic effiency is much lower than that of butelase 1. Butelase 1 has been applied as a tool for protein bioconjugation of recombinant proteins,[241] for macrocyclization of peptides,[6,241] for modification of the bacterial cell surface,[242] and for production of peptide dendrimers[243] (Figure 22). Thiodepsipeptides have been developed and applied to drive butelase 1-catalyzed reactions to completion, enabling peptide bond synthesis in high yield.[244] However, butelase 1's utility is limited by its high Asp/Asn specificity, as well as by the need to isolate the enzyme from the native plant source, as it has not been expressed recombinantly at present.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

9.4. Engineered Trypsin Variants

Following reports of thiolsubtilisin, similar experiments were undertaken to alter trypsin's kinetic properties to favor peptide ligation. However, mutation of the catalytic serine of trypsin to cysteine severely attenuated both esterase and amidase activities in the contexts of both rat trypsin[245] and *Streptomyces griseus* trypsin.[246] Subsequent engineering efforts toward converting trypsin to a peptide ligase have focused on the development of highly specific trypsin mutants that retain the native Ser-containing catalytic triad but are activated only the in the presence of specific substrates.[247] The tetramutant trypsiligase contains four mutations (K60E/N143H/E151H/D189K) that result in high specificity for YRH-containing substrates (Figure 21D), with initial cleavage occurring between Tyr and Arg (Figure 21D). Interestingly, this highly specific variant is also highly active for catalysis of peptide ligation through a substrate-assisted activation mechanism. Structural information revealed that trypsiligase adopts a zymogen-like conformation in the absence of ligands that attenuates its hydrolytic activity but that in the presence of a YRH substrate, the zymogen-like conformation is partially shifted toward an active-like conformation which favors ligation over hydrolysis. Tryspiligase can be applied for either N-terminal or C-terminal modification (Figure 22).[247,248] For N-terminal modification, the YRH sequence is introduced at the N terminus of the protein to be modified. Trypsiligase cleaves between Tyr and Arg and then catalyzes ligation of an activated ester substrate to the new N-terminal Arg residue. For C-terminal ligation, YRH is introduced at the C terminus of the protein of interest. Trypsiligase cleaves between Tyr and Arg, forming an acyl−enzyme intermediate that can be intercepted by an Arg-His nucleophile. Limitations of trypsiligase include its high sequence specificity, residual hydrolytic activity, and poorly characterized kinetics.

10. CONCLUSIONS AND FUTURE PERSPECTIVES

Developments in the field of protein engineering over the past three decades enabled the design of subtiligase and its application as an efficient peptide ligase for N-terminal modification.[28,73,139,249,250] The utility of subtiligase has been expanded through further protein engineering efforts for the development of tailor-made subtiligase variants with properties designed for specific applications, including variants with altered sequence specificity and stability under a wider range of reaction conditions.[33,68,73,34] Based on its broad sequence compatibility, chemoselectivity for N termini, short reaction times, ability to function under physiological conditions, and ease of use, subtiligase has had a strong impact on a wide array of problems and questions in chemistry and biology, including protein synthesis,[32] site-specific bioconjugation,[33,34] and the study of biological signaling pathways that involve N-terminal modification, including proteolysis[35,162] and acetylation.[213]

Subtiligase is an efficient peptide ligase that can be recombinantly expressed in high yield, and it has been widely adopted for peptide ligation applications. Several versions of subtiligase, including stabiligase and peptiligase, have been engineered for optimal reactivity under different reaction conditions (Figure 23). However, there are several limitations that remain to be addressed through protein engineering. Although subtiligase can produce high yields of ligated product with a large excess of either the acyl donor or acceptor peptide, hydrolysis of the thioacyl-enzyme intermediate limits its efficiency with lower amounts of substrate.[33,34,73] Optimizing the kinetic properties of ligation versus hydrolysis of the acyl enzyme intermediate is thus an exciting area to be explored further. Although many subtiligase variants with altered substrate specificity have been reported,[34,36,73] most of these do not have broader specificity than the wild-type enzyme. Development of strategies to engineer a single subtiligase variant with broadened sequence compatibility, as opposed to a family with more restricted specificity, would expand subtiligase's applicability to problems in both protein bioconjugation and biological signaling. On the other hand, in some situations it would be desirable to have subtiligase variants that modify only a single, specific sequence, and mutants with this property have yet to be reported. Recent developments in protein engineering, including phage[37] and yeast[235] display strategies for the evolution of bond-forming enzymes, phage-assisted continuous evolution,[251] and microfluidics-[252] and chip-based[253] screening platforms, provide opportunities to meet these challenges.

The coming years are likely to see wider adoption of subtiligase to address important questions in biological signaling. The development of new prime-side specificity mutants will enable the study of proteolytic signaling pathways carried out by proteases whose specificity does not match that of wild-type subtiligase,[34] as is the case for caspases.[181] Subtiligase can be applied broadly to discover new protease substrates, which is likely to enable the study of proteases that lack strong consensus cleavage sites. To date, subtiligase has mainly been applied to study proteolytic signaling pathways,[35,162,164,165,168,170,171,206] with few reports of its application to study other types of N-terminal modifications.[213] Therefore, there is an opportunity to apply subtiligase more broadly to study the implications of other N-terminal modifications, such as acylation, ubiquitination, and lipidation, in biological signaling. Subtiligase N terminomics requires large amounts of sample, which has so far limited its usefulness mainly to cell lines grown in culture. Protein engineering developments that enhance the efficiency of subtiligase would expand the types of samples whose N terminomes can be easily analyzed with subtiligase. Subtiligase N terminomics is also limited by its lack of subcellular spatial resolution and by its inability to measure the extent of modification of proteins. Therefore, future studies are likely to combine the subtiligase approach with other methods that address these important aspects of biological signaling.

Subtiligase has catalyzed significant progress both in protein bioconjugation and in the elucidation of proteolytic signaling pathways. The body of knowledge accumulated about the enzymatic properties of subtiligase, in combination with advances in chemistry, molecular biology, protein engineering, and mass spectrometry, will fuel new applications and advances in the technology that will be impactful in a variety of different scientific fields.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: amweeks@wisc.edu.
*E-mail: Jim.Wells@ucsf.edu.

ORCID

James A. Wells: 0000-0001-8267-5519

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

Author Contributions

The manuscript was written through contributions of both authors.

Notes

The authors declare the following competing financial interest(s): A.M.W., J.A.W., and the Regents of the University of California have filed a patent application (US Provisional Patent Application No. 62/398,898) pertaining to engineered subtiligase variants.

Biographies

Amy Weeks obtained her Bachelor of Science degree in Chemistry in 2007 from the Massachusetts Institute of Technology, where she performed thesis research with Dr. Stuart Licht. She went on to earn her Ph.D. in Chemistry in the Chemical Biology Graduate Program at the University of California, Berkeley, under the guidance of Dr. Michelle Chang. From 2013−2019, she was a Helen Hay Whitney and a Burroughs Wellcome Fund postdoctoral fellow in the laboratory of Dr. James Wells at the University of California, San Francisco. She joined the faculty of the Department of Biochemistry at the University of Wisconsin—Madison as an assistant professor in Fall 2019.

James Wells grew up in the San Francisco Bay Area and earned his B.A. in Biochemistry and Psychology from the University of California, Berkeley, in 1973. He obtained his Ph.D. in Biochemistry in 1979 from Washington State University with Professor Ralph Yount. From 1980−1982, he was a Damon Runyon−Walter Winchell Postdoctoral Fellow at Stanford University Medical School with Professor George Stark. In 1982, he joined Genentech as cofounding member of the Protein Engineering Department. In 1998, he cofounded Sunesis Pharmaceuticals as President and Chief Scientific Officer. In 2005, he joined the faculty of the Department of Pharmaceutical Chemistry and Cellular & Molecular Pharmacology at the University of California, San Francisco as the Harry Wm. and Diana V. Hind Professor in Pharmaceutical Sciences. Wells' group pioneered the engineering of proteins, antibodies, and small molecules that target catalytic, allosteric, and protein−protein interaction sites; and technologies including protein phage display, alanine-scanning, engineered proteases for improved hydrolysis and ligations, bioconjugations, N terminomics, disulfide "tethering" (a novel site-directed fragment based approach for drug discovery), and more recently an industrialized recombinant antibody production pipeline for the proteome. These led to important new insights into protease mechanisms, growth factor signaling, hot-spots in protein−protein interfaces, role of caspases in biology, and more recently how cell surfaces change in health and disease. His team was integral to several protein and small molecule products including Somavert for acromegaly, Avastin for cancer, Lifitegrast for dry eye disease, and engineered proteases sold by Pfizer, Genentech, Shire and Genencor, respectively. He is an elected member of the US National Academy of Sciences, the American Academy of Arts and Sciences, and the National Academy of Inventors.

ACKNOWLEDGMENTS

We thank S. Coyle, M. Ravalin, D. Sashital, J. Byrnes, O. Julien, and members of the Wells group for helpful comments on the manuscript. J.A.W. holds the Harry and Diana Hind Professorship in Pharmaceutical Sciences. A.M.W. is supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (1017065).

REFERENCES

(1) Wells, J. A.; Ferrari, E.; Henner, D. J.; Estell, D. A.; Chen, E. Y. Cloning, Sequencing, and Secretion of *Bacillus Amyloliquefaciens* Subtilisin in *Bacillus Subtilis*. *Nucleic Acids Res.* **1983**, *11*, 7911−7925.
(2) Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. Engineering Subtilisin and Its Substrates for Efficient Ligation of Peptide Bonds in Aqueous Solution. *Biochemistry* **1991**, *30*, 4151−4159.
(3) Neet, K. E.; Koshland, D. E. The Conversion of Serine at the Active Site of Subtilisin to Cysteine: a "Chemical Mutation". *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *56*, 1606−1611.
(4) Polgar, L.; Bender, M. L. A New Enzyme Containing a Synthetically Formed Active Site. Thiol-Subtilisin. *J. Am. Chem. Soc.* **1966**, *88*, 3153−3154.
(5) Pishesha, N.; Ingram, J. R.; Ploegh, H. L. Sortase a: a Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol.* **2018**, *34*, 163−188.
(6) Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 Is an Asx-Specific Ligase Enabling Peptide Macrocyclization and Synthesis. *Nat. Chem. Biol.* **2014**, *10*, 732−738.
(7) Bergmann, M.; Fraenkel-Conrat, H. The Role of Specificity in the Enzymatic Synthesis of Proteins. *J. Biol. Chem.* **1937**, *119*, 707−720.
(8) Kullman, W. *Enzymatic Peptide Synthesis*; CRC Press: Boca Raton, FL, 1987.
(9) Martin, R. B. Free Energies and Equilibria of Peptide Bond Hydrolysis and Formation. *Biopolymers* **1998**, *45*, 351−353.
(10) Morihara, K. Using Proteases in Peptide Synthesis. *Trends Biotechnol.* **1987**, *5*, 164−170.
(11) Homandberg, G. A.; Mattis, J. A.; Laskowski, M. Synthesis of Peptide Bonds by Proteinases. Addition of Organic Cosolvents Shifts Peptide Bond Equilibria Toward Synthesis. *Biochemistry* **1978**, *17*, 5220−5227.
(12) Niekamp, C. W.; Hixson, H. F., Jr; Laskowski, M., Jr. Peptide-Bond Hydrolysis Equilibria in Native Proteins. Conversion of Virgin Into Modified Soybean Trypsin Inhibitor. *Biochemistry* **1969**, *8*, 16−22.
(13) Barbas, C. F.; Matos, J. R.; West, J. B.; Wong, C. H. A Search for Peptide Ligase: Cosolvent-Mediated Conversion of Proteases to Esterases for Irreversible Synthesis of Peptides. *J. Am. Chem. Soc.* **1988**, *110*, 5162−5166.
(14) Isowa, Y.; Ohmori, M.; Ichikawa, T.; Mori, K. The Thermolysin-Catalyzed Condensation Reactions of N-Substituted Aspartic and Glutamic Acids with Phenylalanine Alkyl Esters. *Tetrahedron Lett.* **1979**, *20*, 2611−2612.
(15) Inouye, K.; Watanabe, K.; Morihara, K.; Tochino, Y.; Kanaya, T.; Emura, J.; Sakakibara, S. Enzyme-Assisted Semisynthesis of Human Insulin. *J. Am. Chem. Soc.* **1979**, *101*, 751−752.
(16) Kullmann, W. Protease-Catalyzed Peptide Bond Formation: Application to Synthesis of the COOH-Terminal Octapeptide of Cholecystokinin. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 2840−2844.
(17) Kullmann, W. Protease-Catalyzed Synthesis of Melanocyte-Stimulating Hormone (MSH) Fragments. *J. Protein Chem.* **1983**, *2*, 289−301.
(18) Barbas, C. F.; Wong, C. H. Papain Catalysed Peptide Synthesis: Control of Amidase Activity and the Introduction of Unusual Amino Acids. *J. Chem. Soc., Chem. Commun.* **1987**, 533−534.
(19) Nakatsuka, T.; Sasaki, T.; Kaiser, E. T. Peptide Segment Synthesis Catalyzed by the Semisynthetic Enzyme Thiolsubtilisin. *J. Am. Chem. Soc.* **1987**, *109*, 3808−3810.
(20) West, J. B.; Scholten, J.; Stolowich, N. J.; Hogg, J. L.; Scott, A. I.; Wong, C. H. Modification of Proteases to Esterases for Peptide Synthesis: Methylchymotrypsin. *J. Am. Chem. Soc.* **1988**, *110*, 3709−3710.
(21) Hartley, B. S.; Kilby, B. A. The Reaction of P-Nitrophenyl Esters with Chymotrypsin and Insulin. *Biochem. J.* **1954**, *56*, 288−297.
(22) Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102*, 4501−4524.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

(23) Fersht, A. R.; Fastrez, J. Demonstration of the Acyl-Enzyme Mechanism for the Hydrolysis of Peptides and Anilides by Chymotrypsin. *Biochemistry* **1973**, *12*, 2025−2034.
(24) Bruice, T. C.; Bruno, J. J.; Chou, W.-S. Nucleophilic Displacement Reactions at the Thiol-Ester Bond of Δ-Thiolvalerolactone. *J. Am. Chem. Soc.* **1963**, *85*, 1659−1669.
(25) Chu, S.-H.; Mautner, H. G. Analogs of Neuroeffectors. v. Neighboring-Group Effects in the Reactions of Esters, Thiolesters, and Selenolesters. the Hydrolysis and Aminolysis of Benzoylcholine, Benzoylthiolcholine, Benzoylselenolcholine, and of Their Dimethylamino Analogs 1. *J. Org. Chem.* **1966**, *31*, 308−312.
(26) Nakatsuka, T.; Sasaki, T.; Kaiser, E. T. Peptide Segment Synthesis Catalyzed by the Semisynthetic Enzyme Thiolsubtilisin. *J. Am. Chem. Soc.* **1987**, *109*, 3808−3810.
(27) Wu, Z. P.; Hilvert, D. Conversion of a Protease Into an Acyl Transferase: Selenolsubtilisin. *J. Am. Chem. Soc.* **1989**, *111*, 4513−4514.
(28) Hutchison, C. A.; Phillips, S.; Edgell, M. H.; Gillam, S.; Jahnke, P.; Smith, M. Mutagenesis at a Specific Position in a DNA Sequence. *J. Biol. Chem.* **1978**, *253*, 6551−6560.
(29) Graycar, T. P.; Bott, R. R.; Caldwell, R. M.; Dauberman, J. L.; Lad, P. J.; Power, S. D.; Sagar, I. H.; Silva, R. A.; Weiss, G. L.; Woodhouse, L. R. Altering the Proteolytic Activity of Subtilisin Through Protein Engineering. *Ann. N. Y. Acad. Sci.* **1992**, *672*, 71−79.
(30) Jackson, D. Y.; Burnier, J. P.; Wells, J. A. Enzymic Cyclization of Linear Peptide Esters Using Subtiligase. *J. Am. Chem. Soc.* **1995**, *117*, 819−820.
(31) Schmidt, M.; Toplak, A.; Rozeboom, H. J.; Wijma, H. J.; Quaedflieg, P. J. L. M.; van Maarseveen, J. H.; Janssen, D. B.; Nuijens, T. Design of a Substrate-Tailored Peptiligase Variant for the Efficient Synthesis of Thymosin-α1. *Org. Biomol. Chem.* **2018**, *16*, 609−618.
(32) Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease a with Unnatural Catalytic Residues. *Science* **1994**, *266*, 243−247.
(33) 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.* **1994**, *91*, 12544−12548.
(34) Weeks, A. M.; Wells, J. A. Engineering Peptide Ligase Specificity by Proteomic Identification of Ligation Sites. *Nat. Chem. Biol.* **2018**, *14*, 50−57.
(35) Mahrus, S.; Trinidad, J. C.; Barkan, D. T.; Sali, A.; Burlingame, A. L.; Wells, J. A. Global Sequencing of Proteolytic Cleavage Sites in Apoptosis by Specific Labeling of Protein N Termini. *Cell* **2008**, *134*, 866−876.
(36) Nuijens, T.; Toplak, A.; Quaedflieg, P. J. L. M.; Drenth, J.; Wu, B.; Janssen, D. B. Engineering a Diverse Ligase Toolbox for Peptide Segment Condensation. *Adv. Synth. Catal.* **2016**, *358*, 4041−4048.
(37) Atwell, S.; Wells, J. A. Selection for Improved Subtiligases by Phage Display. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9497−9502.
(38) Clackson, T.; Wells, J. A. In Vitro Selection From Protein and Peptide Libraries. *Trends Biotechnol.* **1994**, *12*, 173−184.
(39) Pedersen, H.; Hölder, S.; Sutherlin, D. P.; Schwitter, U.; King, D. S.; Schultz, P. G. A Method for Directed Evolution and Functional Cloning of Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 10523−10528.
(40) Demartis, S.; Huber, A.; Viti, F.; Lozzi, L.; Giovannoni, L.; Neri, P.; Winter, G.; Neri, D. A Strategy for the Isolation of Catalytic Activities From Repertoires of Enzymes Displayed on Phage. *J. Mol. Biol.* **1999**, *286*, 617−633.
(41) Jestin, J. L.; Kristensen, P.; Winter, G. A Method for the Selection of Catalytic Activity Using Phage Display and Proximity Coupling. *Angew. Chem., Int. Ed.* **1999**, *38*, 1124−1127.
(42) Bryan, P.; Alexander, P.; Strausberg, S.; Schwarz, F.; Lan, W.; Gilliland, G.; Gallagher, D. T. Energetics of Folding Subtilisin BPN'. *Biochemistry* **1992**, *31*, 4937−4945.
(43) Vasantha, N.; Thompson, L. D.; Rhodes, C.; Banner, C.; Nagle, J.; Filpula, D. Genes for Alkaline Protease and Neutral Protease From *Bacillus Amyloliquefaciens* Contain a Large Open Reading Frame Between the Regions Coding for Signal Sequence and Mature Protein. *J. Bacteriol.* **1984**, *159*, 811−819.
(44) Ikemura, H.; Takagi, H.; Inouye, M. Requirement of Pro-Sequence for the Production of Active Subtilisin E in *Escherichia Coli*. *J. Biol. Chem.* **1987**, *262*, 7859−7864.
(45) Zhu, X. L.; Ohta, Y.; Jordan, F.; Inouye, M. Pro-Sequence of Subtilisin Can Guide the Refolding of Denatured Subtilisin in an Intermolecular Process. *Nature* **1989**, *339*, 483−484.
(46) Gallagher, T.; Bryan, P.; Gilliland, G. L. Calcium-Independent Subtilisin by Design. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 205−213.
(47) Strausberg, S. L.; Alexander, P. A.; Gallagher, D. T.; Gilliland, G. L.; Barnett, B. L.; Bryan, P. N. Directed Evolution of a Subtilisin with Calcium-Independent Stability. *Nat. Biotechnol.* **1995**, *13*, 669−673.
(48) Toplak, A.; Nuijens, T.; Quaedflieg, P. J. L. M.; Wu, B.; Janssen, D. B. Peptiligase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Adv. Synth. Catal.* **2016**, *358*, 2140−2147.
(49) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *425*, 497−502.
(50) Wright, C. S.; Alden, R. A.; Kraut, J. Structure of Subtilisin BPN' at 2.5 Angström Resolution. *Nature* **1969**, *221*, 235−242.
(51) Bott, R.; Ultsch, M.; Kossiakoff, A.; Graycar, T.; Katz, B.; Power, S. The Three-Dimensional Structure of *Bacillus Amyloliquefaciens* Subtilisin at 1.8 Å and an Analysis of the Structural Consequences of Peroxide Inactivation. *J. Biol. Chem.* **1988**, *263*, 7895−7906.
(52) 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 Å Resolution. *J. Mol. Biol.* **1991**, *221*, 309−325.
(53) Takeuchi, Y.; Noguchi, S.; Satow, Y.; Kojima, S.; Kumagai, I.; Miura, K.; Nakamura, K. T.; Mitsui, Y. Molecular Recognition at the Active Site of Subtilisin BPN': Crystallographic Studies Using Genetically Engineered Proteinaceous Inhibitor SSI (*Streptomyces* Subtilisin Inhibitor). *Protein Eng., Des. Sel.* **1991**, *4*, 501−508.
(54) Heinz, D. W.; Priestle, J. P.; Rahuel, J.; Wilson, K. S.; Grütter, M. G. Refined Crystal Structures of Subtilisin Novo in Complex with Wild-Type and Two Mutant Eglins. Comparison with Other Serine Proteinase Inhibitor Complexes. *J. Mol. Biol.* **1991**, *217*, 353−371.
(55) Radisky, E. S.; Kwan, G.; Karen Lu, C.-J.; Koshland, D. E. Binding, Proteolytic, and Crystallographic Analyses of Mutations at the Protease-Inhibitor Interface of the Subtilisin BPN'/Chymotrypsin Inhibitor 2 Complex. *Biochemistry* **2004**, *43*, 13648−13656.
(56) Radisky, E. S.; Lu, C.-J. K.; Kwan, G.; Koshland, D. E. Role of the Intramolecular Hydrogen Bond Network in the Inhibitory Power of Chymotrypsin Inhibitor 2. *Biochemistry* **2005**, *44*, 6823−6830.
(57) Grøn, H.; Meldal, M.; Breddam, K. Extensive Comparison of the Substrate Preferences of Two Subtilisins as Determined with Peptide Substrates Which Are Based on the Principle of Intramolecular Quenching. *Biochemistry* **1992**, *31*, 6011−6018.
(58) Estell, D. A.; Graycar, T. P.; Wells, J. A. Engineering an Enzyme by Site-Directed Mutagenesis to Be Resistant to Chemical Oxidation. *J. Biol. Chem.* **1985**, *260*, 6518−6521.
(59) 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. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 5167−5171.
(60) Ballinger, M. D.; Tom, J.; Wells, J. A. Designing Subtilisin BPN' to Cleave Substrates Containing Dibasic Residues. *Biochemistry* **1995**, *34*, 13312−13319.
(61) Mizuno, K.; Nakamura, T.; Ohshima, T.; Tanaka, S.; Matsuo, H. Yeast KEX2 Genes Encodes an Endopeptidase Homologous to Subtilisin-Like Serine Proteases. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 246−254.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

(62) Ballinger, M. D.; Tom, J.; Wells, J. A. Furilisin: a Variant of Subtilisin BPN' Engineered for Cleaving Tribasic Substrates. *Biochemistry* **1996**, *35*, 13579−13585.
(63) Creemers, J. W.; Siezen, R. J.; Roebroek, A. J.; Ayoubi, T. A.; Huylebroeck, D.; Van de Ven, W. J. Modulation of Furin-Mediated Proprotein Processing Activity by Site-Directed Mutagenesis. *J. Biol. Chem.* **1993**, *268*, 21826−21834.
(64) Siezen, R. J.; Creemers, J. W.; Van de Ven, W. J. Homology Modelling of the Catalytic Domain of Human Furin. a Model for the Eukaryotic Subtilisin-Like Proprotein Convertases. *Eur. J. Biochem.* **1994**, *222*, 255−266.
(65) Lipkind, G.; Gong, Q.; Steiner, D. F. Molecular Modeling of the Substrate Specificity of Prohormone Convertases SPC2 and SPC3. *J. Biol. Chem.* **1995**, *270*, 13277−13284.
(66) Robertus, J. D.; Alden, R. A.; Birktoft, J. J.; Kraut, J.; Powers, J. C.; Wilcox, P. E. An X-Ray Crystallographic Study of the Binding of Peptide Chloromethyl Ketone Inhibitors to Subtilisin BPN'. *Biochemistry* **1972**, *11*, 2439−2449.
(67) Matthews, D. A.; Alden, R. A.; Birktoft, J. J.; Freer, S. T.; Kraut, J. X-Ray Crystallographic Study of Boronic Acid Adducts with Subtilisin BPN' (Novo). a Model for the Catalytic Transition State. *J. Biol. Chem.* **1975**, *250*, 7120−7126.
(68) Nuijens, T.; Toplak, A.; Quaedflieg, P. J. L. M.; Drenth, J.; Wu, B.; Janssen, D. B. Engineering a Diverse Ligase Toolbox for Peptide Segment Condensation. *Adv. Synth. Catal.* **2016**, *358*, 4041−4048.
(69) Schilling, O.; Overall, C. M. Proteome-Derived, Database-Searchable Peptide Libraries for Identifying Protease Cleavage Sites. *Nat. Biotechnol.* **2008**, *26*, 685−694.
(70) Sharma, D.; Verma, S.; Vaidya, S.; Kalia, K.; Tiwari, V. Recent Updates on GLP-1 Agonists: Current Advancements & Challenges. *Biomed. Pharmacother.* **2018**, *108*, 952−962.
(71) Merrifield, R. B. Solid Phase Peptide Synthesis. I. the Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149−2154.
(72) Merrifield, R. B. Solid-Phase Peptide Synthesis. *Adv. Enzymol. Relat. Areas Mol. Biol.* **2006**, *32*, 221−296.
(73) Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. Engineering Subtilisin and Its Substrates for Efficient Ligation of Peptide Bonds in Aqueous Solution. *Biochemistry* **1991**, *30*, 4151−4159.
(74) Judice, J. K.; Namenuk, A. K.; Burnier, J. P. Rapid Assembly of "Subtiligase" Substrates to Elucidate Optimal Ligation Junctures. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1961−1966.
(75) Suich, D. J.; Ballinger, M. D.; Wells, J. A.; DeGrado, W. F. Fmoc-Based Synthesis of Glycolate Ester Peptides for the Assembly of De Novo Designed Multimeric Proteins Using Subtiligase. *Tetrahedron Lett.* **1996**, *37*, 6653−6656.
(76) de Beer, R. J. A. C.; Nuijens, T.; Wiermans, L.; Quaedflieg, P. J. L. M.; Rutjes, F. P. J. T. Improving the Carboxyamidomethyl Ester for Subtilisin-Catalysed Peptide Synthesis. *Org. Biomol. Chem.* **2012**, *10*, 6767−6775.
(77) Yoshihara, H. A. I.; Mahrus, S.; Wells, J. A. Tags for Labeling Protein N-Termini with Subtiligase for Proteomics. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6000−6003.
(78) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. Fmoc-Based Synthesis of Peptide-Alpha-Thioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *J. Am. Chem. Soc.* **1999**, *121*, 11684−11689.
(79) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/T-Bu Chemistry. *J. Am. Chem. Soc.* **1999**, *121*, 11369−11374.
(80) Camarero, J. A.; Hackel, B. J.; de Yoreo, J. J.; Mitchell, A. R. Fmoc-Based Synthesis of Peptide Alpha-Thioesters Using an Aryl Hydrazine Support. *J. Org. Chem.* **2004**, *69*, 4145−4151.
(81) Blanco-Canosa, J. B.; Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chem., Int. Ed.* **2008**, *47*, 6851−6855.
(82) Ollivier, N.; Behr, J.-B.; El-Mahdi, O.; Blanpain, A.; Melnyk, O. Fmoc Solid-Phase Synthesis of Peptide Thioesters Using an Intramolecular N,S-Acyl Shift. *Org. Lett.* **2005**, *7*, 2647−2650.
(83) Tan, X.; Yang, R.; Liu, C.-F. Facilitating Subtiligase-Catalyzed Peptide Ligation Reactions by Using Peptide Thioester Substrates. *Org. Lett.* **2018**, *20*, 6691−6694.
(84) 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.* **1999**, *254*, 147−151.
(85) Tan, X.-H.; Wirjo, A.; Liu, C.-F. An Enzymatic Approach to the Synthesis of Peptide Thioesters: Mechanism and Scope. *ChemBioChem* **2007**, *8*, 1512−1515.
(86) Muir, T. W.; Dawson, P. E.; Kent, S. B. Protein Synthesis by Chemical Ligation of Unprotected Peptides in Aqueous Solution. *Methods Enzymol.* **1997**, *289*, 266−298.
(87) Tan, X.-H.; Yang, R.; Wirjo, A.; Liu, C.-F. Subtiligase as a Hydrothiolase for the Synthesis of Peptide Thioacids. *Tetrahedron Lett.* **2008**, *49*, 2891−2894.
(88) Liu, C.-F.; Rao, C.; Tam, J. P. Acyl Disulfide-Mediated Intramolecular Acylation for Orthogonal Coupling Between Unprotected Peptide Segments. Mechanism and Application. *Tetrahedron Lett.* **1996**, *37*, 933−936.
(89) Blake, J.; Li, C. H. New Segment-Coupling Method for Peptide Synthesis in Aqueous Solution: Application to Synthesis of Human [Gly17]-Beta-Endorphin. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 4055−4058.
(90) Yamashiro, D.; Li, C. H. New Segment Synthesis of Alpha-Inhibin-92 by the Acyl Disulfide Method. *Int. J. Pept. Protein Res.* **1988**, *31*, 322−334.
(91) Wong, P. C. Y.; Guo, J.; Zhang, A. The Renal and Cardiovascular Effects of Natriuretic Peptides. *Adv. Physiol. Educ.* **2017**, *41*, 179−185.
(92) Harris, J. M.; Chess, R. B. Effect of Pegylation on Pharmaceuticals. *Nat. Rev. Drug Discovery* **2003**, *2*, 214−221.
(93) Li, J.; Liu, C. H.; Wang, F. S. Thymosin Alpha 1: Biological Activities, Applications and Genetic Engineering Production. *Peptides* **2010**, *31*, 2151−2158.
(94) Tuthill, C. Issues in Pharmaceutical Development of Thymosin Alpha1 From Preclinical Studies Through Marketing. *Ann. N. Y. Acad. Sci.* **2007**, *1112*, 351−356.
(95) Tuthill, C.; Rios, I.; McBeath, R. Thymosin Alpha 1: Past Clinical Experience and Future Promise. *Ann. N. Y. Acad. Sci.* **2010**, *1194*, 130−135.
(96) Morrison, C. Constrained Peptides' Time to Shine? *Nat. Rev. Drug Discovery* **2018**, *17*, 531−533.
(97) Qian, Z.; Dougherty, P. G.; Pei, D. Targeting Intracellular Protein-Protein Interactions with Cell-Permeable Cyclic Peptides. *Curr. Opin. Chem. Biol.* **2017**, *38*, 80−86.
(98) White, C. J.; Yudin, A. K. Contemporary Strategies for Peptide Macrocyclization. *Nat. Chem.* **2011**, *3*, 509−524.
(99) Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; Ippel, H.; Richelle, G. J. J.; Hackeng, T. M.; van Maarseveen, J. H.; Nuijens, T. Omniligase-1: a Powerful Tool for Peptide Head-to-Tail Cyclization. *Adv. Synth. Catal.* **2017**, *359*, 2050−2055.
(100) Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. Rapid and Quantitative Cyclization of Multiple Peptide Loops Onto Synthetic Scaffolds for Structural Mimicry of Protein Surfaces. *ChemBioChem* **2005**, *6*, 821−824.
(101) Streefkerk, D. E.; Schmidt, M.; Ippel, J. H.; Hackeng, T. M.; Nuijens, T.; Timmerman, P.; van Maarseveen, J. H. Synthesis of Constrained Tetracyclic Peptides by Consecutive CEPS, CLIPS, and Oxime Ligation. *Org. Lett.* **2019**, *21*, 2095−2100.
(102) Cornish, V. W.; Mendel, D.; Schultz, P. G. Probing Protein Structure and Function with an Expanded Genetic Code. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 621−633.
(103) Baca, M.; Alewood, P. F.; Kent, S. B. Structural Engineering of the HIV-1 Protease Molecule with a Beta-Turn Mimic of Fixed Geometry. *Protein Sci.* **1993**, *2*, 1085−1091.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

(104) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Nicotine Binding to Brain Receptors Requires a Strong Cation-Pi Interaction. *Nature* **2009**, *458*, 534−537.
(105) Neumann-Staubitz, P.; Neumann, H. The Use of Unnatural Amino Acids to Study and Engineer Protein Function. *Curr. Opin. Struct. Biol.* **2016**, *38*, 119−128.
(106) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-Specific Incorporation of Non-Canonical Amino Acids Into Proteins: Recent Developments and Applications. *Curr. Opin. Chem. Biol.* **2010**, *14*, 774−780.
(107) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413−444.
(108) Denkewalter, R. G.; Veber, D. F.; Holly, F. W.; Hirschmann, R. Studies on the Total Synthesis of an Enzyme. I. Objective and Strategy. *J. Am. Chem. Soc.* **1969**, *91*, 503−504.
(109) Gutte, B.; Merrifield, R. B. The Synthesis of Ribonuclease a. *J. Biol. Chem.* **1971**, *246*, 1922−1941.
(110) Milton, R. C.; Milton, S. C.; Kent, S. B. Total Chemical Synthesis of a D-Enzyme: the Enantiomers of HIV-1 Protease Show Reciprocal Chiral Substrate Specificity. *Science* **1992**, *256*, 1445−1448.
(111) Campbell, R. L.; Petsko, G. A. Ribonuclease Structure and Catalysis: Crystal Structure of Sulfate-Free Native Ribonuclease a at 1.5-a Resolution. *Biochemistry* **1987**, *26*, 8579−8584.
(112) Rico, M.; Santoro, J.; González, C.; Bruix, M.; Neira, J. L.; Nieto, J. L.; Herranz, J. 3D Structure of Bovine Pancreatic Ribonuclease a in Aqueous Solution: an Approach to Tertiary Structure Determination From a Small Basis of 1H NMR NOE Correlations. *J. Biomol. NMR* **1991**, *1*, 283−298.
(113) Ipata, P. L.; Felicioli, R. A. A Spectrophotometric Assay for Ribonuclease Activity Using Cytidylyl-(3′,5′)-Adenosine and Uridylyl-(3“,5”)-Adenosine as Substrates. *FEBS Lett.* **1968**, *1*, 29−31.
(114) Fink, A. L.; Kar, D.; Kotin, R. Ribonuclease Structure and Catalysis: Effects of Crystalline Enzyme, Alcohol Cryosolvents, Low Temperatures, and Product Inhibition. *Biochemistry* **1987**, *26*, 8571−8579.
(115) Lyles, M. M.; Gilbert, H. F. Catalysis of the Oxidative Folding of Ribonuclease a by Protein Disulfide Isomerase: Dependence of the Rate on the Composition of the Redox Buffer. *Biochemistry* **1991**, *30*, 613−619.
(116) Rothwarf, D. M.; Scheraga, H. A. Regeneration of Bovine Pancreatic Ribonuclease A. 1. Steady-State Distribution. *Biochemistry* **1993**, *32*, 2671−2679.
(117) Findly, D.; Herries, D. G.; Mathias, A. P.; Rabin, B. R.; Ross, C. A. The Active Site and Mechanism of Action of Bovine Pancreatic Ribonuclease. *Nature* **1961**, *190*, 781−784.
(118) Cuchillo, C. M.; Nogués, M. V.; Raines, R. T. Bovine Pancreatic Ribonuclease: Fifty Years of the First Enzymatic Reaction Mechanism. *Biochemistry* **2011**, *50*, 7835−7841.
(119) Veber, D. F.; Paleveda, W. J.; Lee, Y. C.; Hirschmann, R. Isonicotinyloxycarbonyl, a Novel Amino Protecting Group for Peptide Synthesis. *J. Org. Chem.* **1977**, *42*, 3286−3288.
(120) Braisted, A. C.; Judice, J. K.; Wells, J. A. Synthesis of Proteins by Subtiligase. *Methods Enzymol.* **1997**, *289*, 298−313.
(121) Liu, Y.; Patricelli, M. P.; Cravatt, B. F. Activity-Based Protein Profiling: the Serine Hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 14694−14699.
(122) Banghart, M.; Borges, K.; Isacoff, E.; Trauner, D.; Kramer, R. H. Light-Activated Ion Channels for Remote Control of Neuronal Firing. *Nat. Neurosci.* **2004**, *7*, 1381−1386.
(123) Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B. D.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative Reactivity Profiling Predicts Functional Cysteines in Proteomes. *Nature* **2010**, *468*, 790−795.
(124) Erlanson, D. A.; Wells, J. A.; Braisted, A. C. Tethering: Fragment-Based Drug Discovery. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 199−223.
(125) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. K. G12C) Inhibitors Allosterically Control GTP Affinity and Effector Interactions. *Nature* **2013**, *503*, 548−551.
(126) Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frödin, M.; Taunton, J. Reversible Targeting of Noncatalytic Cysteines with Chemically Tuned Electrophiles. *Nat. Chem. Biol.* **2012**, *8*, 471−476.
(127) Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; et al. Site-Specific Conjugation of a Cytotoxic Drug to an Antibody Improves the Therapeutic Index. *Nat. Biotechnol.* **2008**, *26*, 925−932.
(128) Lyon, R. P.; Setter, J. R.; Bovee, T. D.; Doronina, S. O.; Hunter, J. H.; Anderson, M. E.; Balasubramanian, C. L.; Duniho, S. M.; Leiske, C. I.; Li, F.; et al. Self-Hydrolyzing Maleimides Improve the Stability and Pharmacological Properties of Antibody-Drug Conjugates. *Nat. Biotechnol.* **2014**, *32*, 1059−1062.
(129) Pleiner, T.; Bates, M.; Trakhanov, S.; Lee, C.-T.; Schliep, J. E.; Chug, H.; Böhning, M.; Stark, H.; Urlaub, H.; Görlich, D. Nanobodies: Site-Specific Labeling for Super-Resolution Imaging, Rapid Epitope-Mapping and Native Protein Complex Isolation. *eLife* **2015**, *4*, No. e11349.
(130) Baslé, E.; Joubert, N.; Pucheault, M. Protein Chemical Modification on Endogenous Amino Acids. *Chem. Biol.* **2010**, *17*, 213−227.
(131) Lin, S.; Yang, X.; Jia, S.; Weeks, A. M.; Hornsby, M.; Lee, P. S.; Nichiporuk, R. V.; Iavarone, A. T.; Wells, J. A.; Toste, F. D.; et al. Redox-Based Reagents for Chemoselective Methionine Bioconjugation. *Science* **2017**, *355*, 597−602.
(132) Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L. Site-Specific C-Terminal and Internal Loop Labeling of Proteins Using Sortase-Mediated Reactions. *Nat. Protoc.* **2013**, *8*, 1787−1799.
(133) Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. Redirecting Lipoic Acid Ligase for Cell Surface Protein Labeling with Small-Molecule Probes. *Nat. Biotechnol.* **2007**, *25*, 1483−1487.
(134) Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. Site-Specific Chemical Modification of Recombinant Proteins Produced in Mammalian Cells by Using the Genetically Encoded Aldehyde Tag. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3000−3005.
(135) Rosen, C. B.; Francis, M. B. Targeting the N Terminus for Site-Selective Protein Modification. *Nat. Chem. Biol.* **2017**, *13*, 697−705.
(136) Cunningham, B. C.; Wells, J. A. Improvement in the Alkaline Stability of Subtilisin Using an Efficient Random Mutagenesis and Screening Procedure. *Protein Eng., Des. Sel.* **1987**, *1*, 319−325.
(137) Pantoliano, M. W.; Whitlow, M.; Wood, J. F.; Dodd, S. W.; Hardman, K. D.; Rollence, M. L.; Bryan, P. N. Large Increases in General Stability for Subtilisin BPN' Through Incremental Changes in the Free Energy of Unfolding. *Biochemistry* **1989**, *28*, 7205−7213.
(138) Wong, C. H.; Chen, S. T.; Hennen, W. J.; Bibbs, J. A.; Wang, Y. F.; Liu, J. L. C.; Pantoliano, M. W.; Whitlow, M.; Bryan, P. N. Enzymes in Organic Synthesis: Use of Subtilisin and a Highly Stable Mutant Derived From Multiple Site-Specific Mutations. *J. Am. Chem. Soc.* **1990**, *112*, 945−953.
(139) Chen, K.; Arnold, F. H. Tuning the Activity of an Enzyme for Unusual Environments: Sequential Random Mutagenesis of Subtilisin E for Catalysis in Dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5618−5622.
(140) Fellouse, F. A.; Esaki, K.; Birtalan, S.; Raptis, D.; Cancasci, V. J.; Koide, A.; Jhurani, P.; Vasser, M.; Wiesmann, C.; Kossiakoff, A. A.; et al. High-Throughput Generation of Synthetic Antibodies From Highly Functional Minimalist Phage-Displayed Libraries. *J. Mol. Biol.* **2007**, *373*, 924−940.
(141) Hornsby, M.; Paduch, M.; Miersch, S.; Sääf, A.; Matsuguchi, T.; Lee, B.; Wypisniak, K.; Doak, A.; King, D.; Usatyuk, S.; et al. A

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

High Through-Put Platform for Recombinant Antibodies to Folded Proteins. *Mol. Cell. Proteomics* **2015**, *14*, 2833−2847.
(142) Jewett, J. C.; Bertozzi, C. R. Cu-Free Click Cycloaddition Reactions in Chemical Biology. *Chem. Soc. Rev.* **2010**, *39*, 1272−1279.
(143) Saxon, E.; Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287*, 2007−2010.
(144) Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed Protein Ligation: a General Method for Protein Engineering. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 6705−6710.
(145) Evans, T. C.; Benner, J.; Xu, M. Q. Semisynthesis of Cytotoxic Proteins Using a Modified Protein Splicing Element. *Protein Sci.* **1998**, *7*, 2256−2264.
(146) Vila-Perelló, M.; Muir, T. W. Biological Applications of Protein Splicing. *Cell* **2010**, *143*, 191−200.
(147) Henager, S. H.; Chu, N.; Chen, Z.; Bolduc, D.; Dempsey, D. R.; Hwang, Y.; Wells, J.; Cole, P. A. Enzyme-Catalyzed Expressed Protein Ligation. *Nat. Methods* **2016**, *13*, 925−927.
(148) Bolduc, D.; Rahdar, M.; Tu-Sekine, B.; Sivakumaren, S. C.; Raben, D.; Amzel, L. M.; Devreotes, P.; Gabelli, S. B.; Cole, P. Phosphorylation-Mediated PTEN Conformational Closure and Deactivation Revealed with Protein Semisynthesis. *eLife* **2013**, *2*, No. e00691.
(149) Chen, Z.; Cole, P. A. Synthetic Approaches to Protein Phosphorylation. *Curr. Opin. Chem. Biol.* **2015**, *28*, 115−122.
(150) Iwai, H.; Züger, S. Protein Ligation: Applications in NMR Studies of Proteins. *Biotechnol. Genet. Eng. Rev.* **2007**, *24*, 129−145.
(151) Varland, S.; Osberg, C.; Arnesen, T. N-Terminal Modifications of Cellular Proteins: the Enzymes Involved, Their Substrate Specificities and Biological Effects. *Proteomics* **2015**, *15*, 2385−2401.
(152) auf dem Keller, U.; Doucet, A.; Overall, C. M. Protease Research in the Era of Systems Biology. *Biol. Chem.* **2007**, *388*, 1159−1162.
(153) Pham, V. C.; Anania, V. G.; Phung, Q. T.; Lill, J. R. Complementary Methods for the Identification of Substrates of Proteolysis. *Methods Enzymol.* **2014**, *544*, 359−380.
(154) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. Exploring Proteomes and Analyzing Protein Processing by Mass Spectrometric Identification of Sorted N-Terminal Peptides. *Nat. Biotechnol.* **2003**, *21*, 566−569.
(155) Van Damme, P.; Martens, L.; Van Damme, J.; Hugelier, K.; Staes, A.; Vandekerckhove, J.; Gevaert, K. Caspase-Specific and Nonspecific in Vivo Protein Processing During Fas-Induced Apoptosis. *Nat. Methods* **2005**, *2*, 771−777.
(156) Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. Isotopic Labeling of Terminal Amines in Complex Samples Identifies Protein N-Termini and Protease Cleavage Products. *Nat. Biotechnol.* **2010**, *28*, 281−288.
(157) Griswold, A. R.; Cifani, P.; Rao, S. D.; Axelrod, A. J.; Miele, M. M.; Hendrickson, R. C.; Kentsis, A.; Bachovchin, D. A. A Chemical Strategy for Protease Substrate Profiling. *Cell Chem. Biol.* **2019**, *26*, 901
(158) Timmer, J. C.; Enoksson, M.; Wildfang, E.; Zhu, W.; Igarashi, Y.; Denault, J.-B.; Ma, Y.; Dummitt, B.; Chang, Y.-H.; Mast, A. E.; et al. Profiling Constitutive Proteolytic Events in Vivo. *Biochem. J.* **2007**, *407*, 41−48.
(159) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. One-Step Site-Specific Modification of Native Proteins with 2-Pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **2015**, *11*, 326−331.
(160) Kappelhoff, R.; Puente, X. S.; Wilson, C. H.; Seth, A.; López-Otín, C.; Overall, C. M. Overview of Transcriptomic Analysis of All Human Proteases, Non-Proteolytic Homologs and Inhibitors: Organ, Tissue and Ovarian Cancer Cell Line Expression Profiling of the Human Protease Degradome by the CLIP-CHIP DNA Microarray. *Biochim. Biophys. Acta, Mol. Cell Res.* **2017**, *1864*, 2210−2219.
(161) Wiita, A. P.; Seaman, J. E.; Wells, J. A. Global Analysis of Cellular Proteolysis by Selective Enzymatic Labeling of Protein N-Termini. *Methods Enzymol.* **2014**, *544*, 327−358.
(162) Agard, N. J.; Mahrus, S.; Trinidad, J. C.; Lynn, A.; Burlingame, A. L.; Wells, J. A. Global Kinetic Analysis of Proteolysis via Quantitative Targeted Proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1913−1918.
(163) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376−386.
(164) Agard, N. J.; Maltby, D.; Wells, J. A. Inflammatory Stimuli Regulate Caspase Substrate Profiles. *Mol. Cell. Proteomics* **2010**, *9*, 880−893.
(165) Hill, M. E.; Kumar, A.; Wells, J. A.; Hobman, T. C.; Julien, O.; Hardy, J. A. The Unique Cofactor Region of Zika Virus NS2B-NS3 Protease Facilitates Cleavage of Key Host Proteins. *ACS Chem. Biol.* **2018**, *13*, 2398−2405.
(166) Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. Full Dynamic Range Proteome Analysis of S. Cerevisiae by Targeted Proteomics. *Cell* **2009**, *138*, 795−806.
(167) Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. High-Throughput Generation of Selected Reaction-Monitoring Assays for Proteins and Proteomes. *Nat. Methods* **2010**, *7*, 43−46.
(168) Shimbo, K.; Hsu, G. W.; Nguyen, H.; Mahrus, S.; Trinidad, J. C.; Burlingame, A. L.; Wells, J. A. Quantitative Profiling of Caspase-Cleaved Substrates Reveals Different Drug-Induced and Cell-Type Patterns in Apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 12432−12437.
(169) Wiita, A. P.; Ziv, E.; Wiita, P. J.; Urisman, A.; Julien, O.; Burlingame, A. L.; Weissman, J. S.; Wells, J. A. Global Cellular Response to Chemotherapy-Induced Apoptosis. *eLife* **2013**, *2*, No. e01236.
(170) Wiita, A. P.; Hsu, G. W.; Lu, C. M.; Esensten, J. H.; Wells, J. A. Circulating Proteolytic Signatures of Chemotherapy-Induced Cell Death in Humans Discovered by N-Terminal Labeling. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7594−7599.
(171) Julien, O.; Zhuang, M.; Wiita, A. P.; O'Donoghue, A. J.; Knudsen, G. M.; Craik, C. S.; Wells, J. A. Quantitative MS-Based Enzymology of Caspases Reveals Distinct Protein Substrate Specificities, Hierarchies, and Cellular Roles. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E2001−E2010.
(172) Ashkenazi, A.; Salvesen, G. Regulated Cell Death: Signaling and Mechanisms. *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 337−356.
(173) Nagata, S. Apoptosis and Clearance of Apoptotic Cells. *Annu. Rev. Immunol.* **2018**, *36*, 489−517.
(174) Letai, A. Apoptosis and Cancer. *Annu. Rev. Cancer Biol.* **2017**, *1*, 275−294.
(175) Thornberry, N. A.; Bull, H. G.; Calaycay, J. R.; Chapman, K. T.; Howard, A. D.; Kostura, M. J.; Miller, D. K.; Molineaux, S. M.; Weidner, J. R.; Aunins, J. A Novel Heterodimeric Cysteine Protease Is Required for Interleukin-1 Beta Processing in Monocytes. *Nature* **1992**, *356*, 768−774.
(176) Cerretti, D. P.; Kozlosky, C. J.; Mosley, B.; Nelson, N.; Van Ness, K.; Greenstreet, T. A.; March, C. J.; Kronheim, S. R.; Druck, T.; Cannizzaro, L. A. Molecular Cloning of the Interleukin-1 Beta Converting Enzyme. *Science* **1992**, *256*, 97−100.
(177) Fuentes-Prior, P.; Salvesen, G. S. The Protein Structures That Shape Caspase Activity, Specificity, Activation and Inhibition. *Biochem. J.* **2004**, *384*, 201−232.
(178) Montecucco, A.; Zanetta, F.; Biamonti, G. Molecular Mechanisms of Etoposide. *EXCLI J.* **2015**, *14*, 95−108.
(179) Mizumoto, K.; Rothman, R. J.; Farber, J. L. Programmed Cell Death (Apoptosis) of Mouse Fibroblasts Is Induced by the Topoisomerase II Inhibitor Etoposide. *Mol. Pharmacol.* **1994**, *46*, 890−895.
(180) Barkan, D. T.; Hostetter, D. R.; Mahrus, S.; Pieper, U.; Wells, J. A.; Craik, C. S.; Sali, A. Prediction of Protease Substrates Using Sequence and Structure Features. *Bioinformatics* **2010**, *26*, 1714−1722.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

(181) Thornberry, N. A.; Rano, T. A.; Peterson, E. P.; Rasper, D. M.; Timkey, T.; Garcia-Calvo, M.; Houtzager, V. M.; Nordstrom, P. A.; Roy, S.; Vaillancourt, J. P.; et al. A Combinatorial Approach Defines Specificities of Members of the Caspase Family and Granzyme B. Functional Relationships Established for Key Mediators of Apoptosis. *J. Biol. Chem.* **1997**, *272*, 17907−17911.
(182) Stennicke, H. R.; Renatus, M.; Meldal, M.; Salvesen, G. S. Internally Quenched Fluorescent Peptide Substrates Disclose the Subsite Preferences of Human Caspases 1, 3, 6, 7 and 8. *Biochem. J.* **2000**, *350*, 563−568.
(183) Crawford, E. D.; Seaman, J. E.; Barber, A. E.; David, D. C.; Babbitt, P. C.; Burlingame, A. L.; Wells, J. A. Conservation of Caspase Substrates Across Metazoans Suggests Hierarchical Importance of Signaling Pathways Over Specific Targets and Cleavage Site Motifs in Apoptosis. *Cell Death Differ.* **2012**, *19*, 2040−2048.
(184) Dix, M. M.; Simon, G. M.; Cravatt, B. F. Global Mapping of the Topography and Magnitude of Proteolytic Events in Apoptosis. *Cell* **2008**, *134*, 679−691.
(185) Simon, G. M.; Dix, M. M.; Cravatt, B. F. Comparative Assessment of Large-Scale Proteomic Studies of Apoptotic Proteolysis. *ACS Chem. Biol.* **2009**, *4*, 401−408.
(186) Dix, M. M.; Simon, G. M.; Cravatt, B. F. Global Identification of Caspase Substrates Using PROTOMAP (Protein Topography and Migration Analysis Platform). *Methods Mol. Biol.* **2014**, *1133*, 61−70.
(187) Julien, O.; Kampmann, M.; Bassik, M. C.; Zorn, J. A.; Venditto, V. J.; Shimbo, K.; Agard, N. J.; Shimada, K.; Rheingold, A. L.; Stockwell, B. R.; et al. Unraveling the Mechanism of Cell Death Induced by Chemical Fibrils. *Nat. Chem. Biol.* **2014**, *10*, 969−976.
(188) Wildes, D.; Wells, J. A. Sampling the N-Terminal Proteome of Human Blood. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4561−4566.
(189) Martinon, F.; Tschopp, J. Inflammatory Caspases: Linking an Intracellular Innate Immune System to Autoinflammatory Diseases. *Cell* **2004**, *117*, 561−574.
(190) Kayagaki, N.; Stowe, I. B.; Lee, B. L.; O'Rourke, K.; Anderson, K.; Warming, S.; Cuellar, T.; Haley, B.; Roose-Girma, M.; Phung, Q. T.; et al. Caspase-11 Cleaves Gasdermin D for Non-Canonical Inflammasome Signalling. *Nature* **2015**, *526*, 666−671.
(191) Shi, J.; Zhao, Y.; Wang, K.; Shi, X.; Wang, Y.; Huang, H.; Zhuang, Y.; Cai, T.; Wang, F.; Shao, F. Cleavage of GSDMD by Inflammatory Caspases Determines Pyroptotic Cell Death. *Nature* **2015**, *526*, 660−665.
(192) Liu, X.; Zhang, Z.; Ruan, J.; Pan, Y.; Magupalli, V. G.; Wu, H.; Lieberman, J. Inflammasome-Activated Gasdermin D Causes Pyroptosis by Forming Membrane Pores. *Nature* **2016**, *535*, 153−158.
(193) Schwartz, D. A. Autopsy and Postmortem Studies Are Concordant: Pathology of Zika Virus Infection Is Neurotropic in Fetuses and Infants with Microcephaly Following Transplacental Transmission. *Arch. Pathol. Lab. Med.* **2017**, *141*, 68−72.
(194) Cao-Lormeau, V. M.; Blake, A.; Mons, S.; Lastere, S.; Roche, C.; Vanhomwegen, J.; Dub, T.; Baudouin, L.; Teissier, A.; Larre, P.; et al. Guillain-Barré Syndrome Outbreak Associated with Zika Virus Infection in French Polynesia: a Case-Control Study. *Lancet* **2016**, *387*, 1531−1539.
(195) Liebig, H. D.; Ziegler, E.; Yan, R.; Hartmuth, K.; Klump, H.; Kowalski, H.; Blaas, D.; Sommergruber, W.; Frasel, L.; Lamphear, B. Purification of Two Picornaviral 2A Proteinases: Interaction with eIF-4 Gamma and Influence on in Vitro Translation. *Biochemistry* **1993**, *32*, 7581−7588.
(196) Lamphear, B. J.; Yan, R.; Yang, F.; Waters, D.; Liebig, H. D.; Klump, H.; Kuechler, E.; Skern, T.; Rhoads, R. E. Mapping the Cleavage Site in Protein Synthesis Initiation Factor eIF-4 Gamma of the 2A Proteases From Human Coxsackievirus and Rhinovirus. *J. Biol. Chem.* **1993**, *268*, 19200−19203.
(197) Etchison, D.; Milburn, S. C.; Edery, I.; Sonenberg, N.; Hershey, J. W. Inhibition of HeLa Cell Protein Synthesis Following Poliovirus Infection Correlates with the Proteolysis of a 220,000-Dalton Polypeptide Associated with Eucaryotic Initiation Factor 3 and a Cap Binding Protein Complex. *J. Biol. Chem.* **1982**, *257*, 14806−14810.
(198) Zamora, M.; Marissen, W. E.; Lloyd, R. E. Multiple eIF4GI-Specific Protease Activities Present in Uninfected and Poliovirus-Infected Cells. *J. Virol.* **2002**, *76*, 165−177.
(199) Haghighat, A.; Svitkin, Y.; Novoa, I.; Kuechler, E.; Skern, T.; Sonenberg, N. The eIF4G-eIF4E Complex Is the Target for Direct Cleavage by the Rhinovirus 2A Proteinase. *J. Virol.* **1996**, *70*, 8444−8450.
(200) Ventoso, I.; Blanco, R.; Perales, C.; Carrasco, L. HIV-1 Protease Cleaves Eukaryotic Initiation Factor 4G and Inhibits Cap-Dependent Translation. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 12966−12971.
(201) Potempa, J.; Pike, R. N. Corruption of Innate Immunity by Bacterial Proteases. *J. Innate Immun.* **2009**, *1*, 70−87.
(202) Imamura, T.; Tanase, S.; Szmyd, G.; Kozik, A.; Travis, J.; Potempa, J. Induction of Vascular Leakage Through Release of Bradykinin and a Novel Kinin by Cysteine Proteinases From Staphylococcus Aureus. *J. Exp. Med.* **2005**, *201*, 1669−1676.
(203) Sumitomo, T.; Nakata, M.; Higashino, M.; Terao, Y.; Kawabata, S. Group a Streptococcal Cysteine Protease Cleaves Epithelial Junctions and Contributes to Bacterial Translocation. *J. Biol. Chem.* **2013**, *288*, 13317−13324.
(204) Guzik, K.; Bzowska, M.; Smagur, J.; Krupa, O.; Sieprawska, M.; Travis, J.; Potempa, J. A New Insight Into Phagocytosis of Apoptotic Cells: Proteolytic Enzymes Divert the Recognition and Clearance of Polymorphonuclear Leukocytes by Macrophages. *Cell Death Differ.* **2007**, *14*, 171−182.
(205) Choi, V. M.; Herrou, J.; Hecht, A. L.; Teoh, W. P.; Turner, J. R.; Crosson, S.; Bubeck Wardenburg, J. Activation of Bacteroides Fragilis Toxin by a Novel Bacterial Protease Contributes to Anaerobic Sepsis in Mice. *Nat. Med.* **2016**, *22*, 563−567.
(206) Roncase, E. J.; Moon, C.; Chatterjee, S.; González-Páez, G. E.; Craik, C. S.; O'Donoghue, A. J.; Wolan, D. W. Substrate Profiling and High Resolution Co-Complex Crystal Structure of a Secreted C11 Protease Conserved Across Commensal Bacteria. *ACS Chem. Biol.* **2017**, *12*, 1556−1565.
(207) Rapoport, T. A. Protein Translocation Across the Eukaryotic Endoplasmic Reticulum and Bacterial Plasma Membranes. *Nature* **2007**, *450*, 663−669.
(208) Mossmann, D.; Meisinger, C.; Vögtle, F.-N. Processing of Mitochondrial Presequences. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2012**, *1819*, 1098−1106.
(209) Marfori, M.; Mynott, A.; Ellis, J. J.; Mehdi, A. M.; Saunders, N. F. W.; Curmi, P. M.; Forwood, J. K.; Bodén, M.; Kobe, B. Molecular Basis for Specificity of Nuclear Import and Prediction of Nuclear Localization. *Biochim. Biophys. Acta, Mol. Cell Res.* **2011**, *1813*, 1562−1577.
(210) Calvo, S. E.; Julien, O.; Clauser, K. R.; Shen, H.; Kamer, K. J.; Wells, J. A.; Mootha, V. K. Comparative Analysis of Mitochondrial N-Termini From Mouse, Human, and Yeast. *Mol. Cell. Proteomics* **2017**, *16*, 512−523.
(211) Vaca Jacome, A. S.; Rabilloud, T.; Schaeffer-Reiss, C.; Rompais, M.; Ayoub, D.; Lane, L.; Bairoch, A.; Van Dorsselaer, A.; Carapito, C. N-Terminome Analysis of the Human Mitochondrial Proteome. *Proteomics* **2015**, *15*, 2519−2524.
(212) Vögtle, F.-N.; Wortelkamp, S.; Zahedi, R. P.; Becker, D.; Leidhold, C.; Gevaert, K.; Kellermann, J.; Voos, W.; Sickmann, A.; Pfanner, N.; et al. Global Analysis of the Mitochondrial N-Proteome Identifies a Processing Peptidase Critical for Protein Stability. *Cell* **2009**, *139*, 428−439.
(213) Yi, C. H.; Pan, H.; Seebacher, J.; Jang, I.-H.; Hyberts, S. G.; Heffron, G. J.; Vander Heiden, M. G.; Yang, R.; Li, F.; Locasale, J. W.; et al. Metabolic Regulation of Protein N-Alpha-Acetylation by Bcl-xL Promotes Cell Survival. *Cell* **2011**, *146*, 607−620.
(214) Martinez, A.; Traverso, J. A.; Valot, B.; Ferro, M.; Espagne, C.; Ephritikhine, G.; Zivy, M.; Giglione, C.; Meinnel, T. Extent of N-Terminal Modifications in Cytosolic Proteins From Eukaryotes. *Proteomics* **2008**, *8*, 2809−2831.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

(215) McIntosh, J. A.; Robertson, C. R.; Agarwal, V.; Nair, S. K.; Bulaj, G. W.; Schmidt, E. W. Circular Logic: Nonribosomal Peptide-Like Macrocyclization with a Ribosomal Peptide Catalyst. *J. Am. Chem. Soc.* **2010**, *132*, 15499−15501.
(216) Lee, J.; McIntosh, J.; Hathaway, B. J.; Schmidt, E. W. Using Marine Natural Products to Discover a Protease That Catalyzes Peptide Macrocyclization of Diverse Substrates. *J. Am. Chem. Soc.* **2009**, *131*, 2122−2124.
(217) Donia, M. S.; Schmidt, E. W. Linking Chemistry and Genetics in the Growing Cyanobactin Natural Products Family. *Chem. Biol.* **2011**, *18*, 508−519.
(218) Barber, C. J. S.; Pujara, P. T.; Reed, D. W.; Chiwocha, S.; Zhang, H.; Covello, P. S. The Two-Step Biosynthesis of Cyclic Peptides From Linear Precursors in a Member of the Plant Family Caryophyllaceae Involves Cyclization by a Serine Protease-Like Enzyme. *J. Biol. Chem.* **2013**, *288*, 12500−12510.
(219) Koehnke, J.; Bent, A.; Houssen, W. E.; Zollman, D.; Morawitz, F.; Shirran, S.; Vendome, J.; Nneoyiegbe, A. F.; Trembleau, L.; Botting, C. H.; et al. The Mechanism of Patellamide Macrocyclization Revealed by the Characterization of the PatG Macrocyclase Domain. *Nat. Struct. Mol. Biol.* **2012**, *19*, 767−772.
(220) Mazmanian, S. K. Staphylococcus Aureus Sortase, an Enzyme That Anchors Surface Proteins to the Cell Wall. *Science* **1999**, *285*, 760−763.
(221) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Purification and Characterization of Sortase, the Transpeptidase That Cleaves Surface Proteins of Staphylococcus Aureus at the LPXTG Motif. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 12424−12429.
(222) Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R. T. Structure of Sortase, the Transpeptidase That Anchors Proteins to the Cell Wall of Staphylococcus Aureus. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 6056−6061.
(223) Frankel, B. A.; Kruger, R. G.; Robinson, D. E.; Kelleher, N. L.; McCafferty, D. G. Staphylococcus Aureus Sortase Transpeptidase SrtA: Insight Into the Kinetic Mechanism and Evidence for a Reverse Protonation Catalytic Mechanism. *Biochemistry* **2005**, *44*, 11188−11200.
(224) Ton-That, H.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus Aureus. Sortase Catalyzed in Vitro Transpeptidation Reaction Using LPXTG Peptide and NH(2)-Gly(3) Substrates. *J. Biol. Chem.* **2000**, *275*, 9876−9881.
(225) Popp, M. W.-L. Site-Specific Labeling of Proteins via Sortase: Protocols for the Molecular Biologist. *Methods Mol. Biol.* **2015**, *1266*, 185−198.
(226) Witte, M. D.; Cragnolini, J. J.; Dougan, S. K.; Yoder, N. C.; Popp, M. W.; Ploegh, H. L. Preparation of Unnatural N-to-N and C-to-C Protein Fusions. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11993−11998.
(227) Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. Sortase-Mediated Protein Ligation: a New Method for Protein Engineering. *J. Am. Chem. Soc.* **2004**, *126*, 2670−2671.
(228) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. Sortagging: a Versatile Method for Protein Labeling. *Nat. Chem. Biol.* **2007**, *3*, 707−708.
(229) van 't Hof, W.; Hansenová Maňásková, S.; Veerman, E. C. I.; Bolscher, J. G. M. Sortase-Mediated Backbone Cyclization of Proteins and Peptides. *Biol. Chem.* **2015**, *396*, 283−293.
(230) Jia, X.; Kwon, S.; Wang, C.-I. A.; Huang, Y.-H.; Chan, L. Y.; Tan, C. C.; Rosengren, K. J.; Mulvenna, J. P.; Schroeder, C. I.; Craik, D. J. Semienzymatic Cyclization of Disulfide-Rich Peptides Using Sortase a. *J. Biol. Chem.* **2014**, *289*, 6627−6638.
(231) Maruyama, T.; Dougan, S. K.; Truttmann, M. C.; Bilate, A. M.; Ingram, J. R.; Ploegh, H. L. Increasing the Efficiency of Precise Genome Editing with CRISPR-Cas9 by Inhibition of Nonhomologous End Joining. *Nat. Biotechnol.* **2015**, *33*, 538−542.
(232) Shi, J.; Kundrat, L.; Pishesha, N.; Bilate, A.; Theile, C.; Maruyama, T.; Dougan, S. K.; Ploegh, H. L.; Lodish, H. F. Engineered Red Blood Cells as Carriers for Systemic Delivery of a Wide Array of Functional Probes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10131−10136.
(233) Popp, M. W.-L.; Karssemeijer, R. A.; Ploegh, H. L. Chemoenzymatic Site-Specific Labeling of Influenza Glycoproteins as a Tool to Observe Virus Budding in Real Time. *PLoS Pathog.* **2012**, *8*, No. e1002604.
(234) Esteban, A.; Popp, M. W.; Vyas, V. K.; Strijbis, K.; Ploegh, H. L.; Fink, G. R. Fungal Recognition Is Mediated by the Association of Dectin-1 and Galectin-3 in Macrophages. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14270−14275.
(235) Chen, I.; Dorr, B. M.; Liu, D. R. A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11399−11404.
(236) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. Reprogramming the Specificity of Sortase Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13343−13348.
(237) Rawlings, N. D.; Morton, F. R.; Barrett, A. J. MEROPS: the Peptidase Database. *Nucleic Acids Res.* **2006**, *34*, D270−D272.
(238) Saska, I.; Gillon, A. D.; Hatsugai, N.; Dietzgen, R. G.; Hara-Nishimura, I.; Anderson, M. A.; Craik, D. J. An Asparaginyl Endopeptidase Mediates in Vivo Protein Backbone Cyclization. *J. Biol. Chem.* **2007**, *282*, 29721−29728.
(239) Poth, A. G.; Colgrave, M. L.; Philip, R.; Kerenga, B.; Daly, N. L.; Anderson, M. A.; Craik, D. J. Discovery of Cyclotides in the Fabaceae Plant Family Provides New Insights Into the Cyclization, Evolution, and Distribution of Circular Proteins. *ACS Chem. Biol.* **2011**, *6*, 345−355.
(240) Harris, K. S.; Durek, T.; Kaas, Q.; Poth, A. G.; Gilding, E. K.; Conlan, B. F.; Saska, I.; Daly, N. L.; van der Weerden, N. L.; Craik, D. J.; et al. Efficient Backbone Cyclization of Linear Peptides by a Recombinant Asparaginyl Endopeptidase. *Nat. Commun.* **2015**, *6*, 10199.
(241) Nguyen, G. K. T.; Qiu, Y.; Cao, Y.; Hemu, X.; Liu, C.-F.; Tam, J. P. Butelase-Mediated Cyclization and Ligation of Peptides and Proteins. *Nat. Protoc.* **2016**, *11*, 1977−1988.
(242) Bi, X.; Yin, J.; Nguyen, G. K. T.; Rao, C.; Halim, N. B. A.; Hemu, X.; Tam, J. P.; Liu, C.-F. Enzymatic Engineering of Live Bacterial Cell Surfaces Using Butelase 1. *Angew. Chem., Int. Ed.* **2017**, *56*, 7822−7825.
(243) Cao, Y.; Nguyen, G. K. T.; Chuah, S.; Tam, J. P.; Liu, C.-F. Butelase-Mediated Ligation as an Efficient Bioconjugation Method for the Synthesis of Peptide Dendrimers. *Bioconjugate Chem.* **2016**, *27*, 2592−2596.
(244) Nguyen, G. K. T.; Cao, Y.; Wang, W.; Liu, C.-F.; Tam, J. P. Site-Specific N-Terminal Labeling of Peptides and Proteins Using Butelase 1 and Thiodepsipeptide. *Angew. Chem., Int. Ed.* **2015**, *54*, 15694−15698.
(245) Higaki, J. N.; Evnin, L. B.; Craik, C. S. Introduction of a Cysteine Protease Active Site Into Trypsin. *Biochemistry* **1989**, *28*, 9256−9263.
(246) Yokosawa, H.; Ojima, S.; Ishii, S. Thioltrypsin. Chemical Transformation of the Active-Site Serine Residue of Streptomyces Griseus Trypsin to a Cysteine Residue. *J. Biochem.* **1977**, *82*, 869−876.
(247) Liebscher, S.; Schöpfel, M.; Aumüller, T.; Sharkhuukhen, A.; Pech, A.; Höss, E.; Parthier, C.; Jahreis, G.; Stubbs, M. T.; Bordusa, F. N-Terminal Protein Modification by Substrate-Activated Reverse Proteolysis. *Angew. Chem., Int. Ed.* **2014**, *53*, 3024−3028.
(248) Meyer, C.; Liebscher, S.; Bordusa, F. Selective Coupling of Click Anchors to Proteins via Trypsiligase. *Bioconjugate Chem.* **2016**, *27*, 47−53.
(249) Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985**, *228*, 1315−1317.
(250) Scott, J. K.; Smith, G. P. Searching for Peptide Ligands with an Epitope Library. *Science* **1990**, *249*, 386−390.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160

(251) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472*, 499−503.
(252) Agresti, J. J.; Antipov, E.; Abate, A. R.; Ahn, K.; Rowat, A. C.; Baret, J.-C.; Marquez, M.; Klibanov, A. M.; Griffiths, A. D.; Weitz, D. A. Ultrahigh-Throughput Screening in Drop-Based Microfluidics for Directed Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4004−4009.
(253) Chen, B.; Lim, S.; Kannan, A.; Alford, S. C.; Sunden, F.; Herschlag, D.; Dimov, I. K.; Baer, T. M.; Cochran, J. R. High-Throughput Analysis and Protein Engineering Using Microcapillary Arrays. *Nat. Chem. Biol.* **2016**, *12*, 76−81.

DOI: 10.1021/acs.chemrev.9b00372
Chem. Rev. 2020, 120, 3127−3160