Bioconjugate Chemistry

pubs.acs.org/bc

Communication

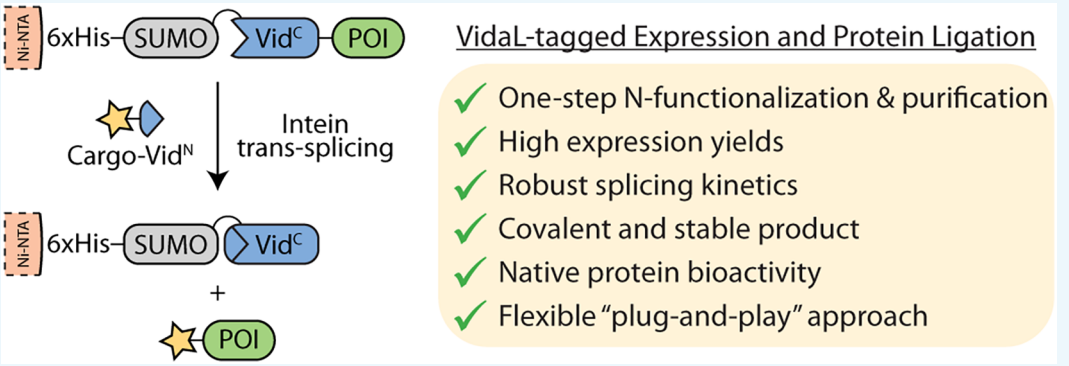
One-Step Purification and N-Terminal Functionalization of Bioactive Proteins via Atypically Split Inteins

Ryan Gharios, Annabella Li, Irina Kopyeva, Ryan M. Francis, and Cole A. DeForest*

Cite This: https://doi.org/10.1021/acs.bioconjchem.4c00223

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Site-specific installation of non-natural functionality onto proteins has enabled countless applications in biotechnology, chemical biology, and biomaterials science. Though the N-terminus is an attractive derivatization location, prior methodologies targeting this site have suffered from low selectivity, a limited selection of potential chemical modifications, and/or challenges associated with divergent protein purification/modification steps. In this work, we harness the atypically split VidaL intein to simultaneously N-functionalize and purify homogeneous protein populations in a single step. Our method—referred to as VidaL-tagged expression and protein ligation (VEPL)—enables modular and scalable production of N-terminally modified proteins with native bioactivity. Demonstrating its flexibility and ease of use, we employ VEPL to combinatorially install 4 distinct (multi)functional handles (e.g., biotin, alkyne, fluorophores) to the N-terminus of 4 proteins that span three different classes: fluorescent (Enhanced Green Fluorescent Protein, mCherry), enzymatic (β-lactamase), and growth factor (epidermal growth factor). Moving forward, we anticipate that VEPL's ability to rapidly generate and isolate N-modified proteins will prove useful across the growing fields of applied chemical biology.

INTRODUCTION

The creation of well-defined protein conjugates, whether wholly genetically encoded or semisynthetic in nature, is crucial for studies in chemical biology and beyond.[1] The expression and assembly of different protein chimeras has, in fact, led to advances in spaces as diverse as biomolecule localization and tracking[2] and the creation of well-defined antibody-drug conjugates for therapeutic targeting.[3] Beyond these applications, site-specific protein modification is also key for the creation of functional biomaterials.[4−7] In recent years, researchers in this space have steadily moved away from stochastic modification methodologies (e.g., based on activated esters or maleimides) that, while convenient, often lead to heterogeneous protein populations and significant batch-to-batch variability. To address this issue, multiple semisynthetic methods have been developed in order to afford a rich armamentarium of site-specific protein modification techniques.[8−11]

In particular, N-terminal modification of proteins has been a long-standing goal for the field partly to recapture the diversity of proteoforms that result from various N-terminal modifications.[12,13] Moreover, N-termini are typically solvent-exposed and fall outside of the final fold, implying that modifications incorporated at this site preserve protein bioactivity better than those installed at internal residues.[14] As such, the N-terminus is particularly suitable for installing synthetic handles onto protein constructs that are otherwise fragile. Hampering this end-goal is the lack of broad applicability of typical N-terminal modification strategies, as these usually suffer from overall suboptimal selectivity and propensity toward multiple side-reactions, the need for different sequential purification and derivatization steps, and/or a constriction to a limited subset of

Received: May 15, 2024
Revised: May 20, 2024
Accepted: May 21, 2024

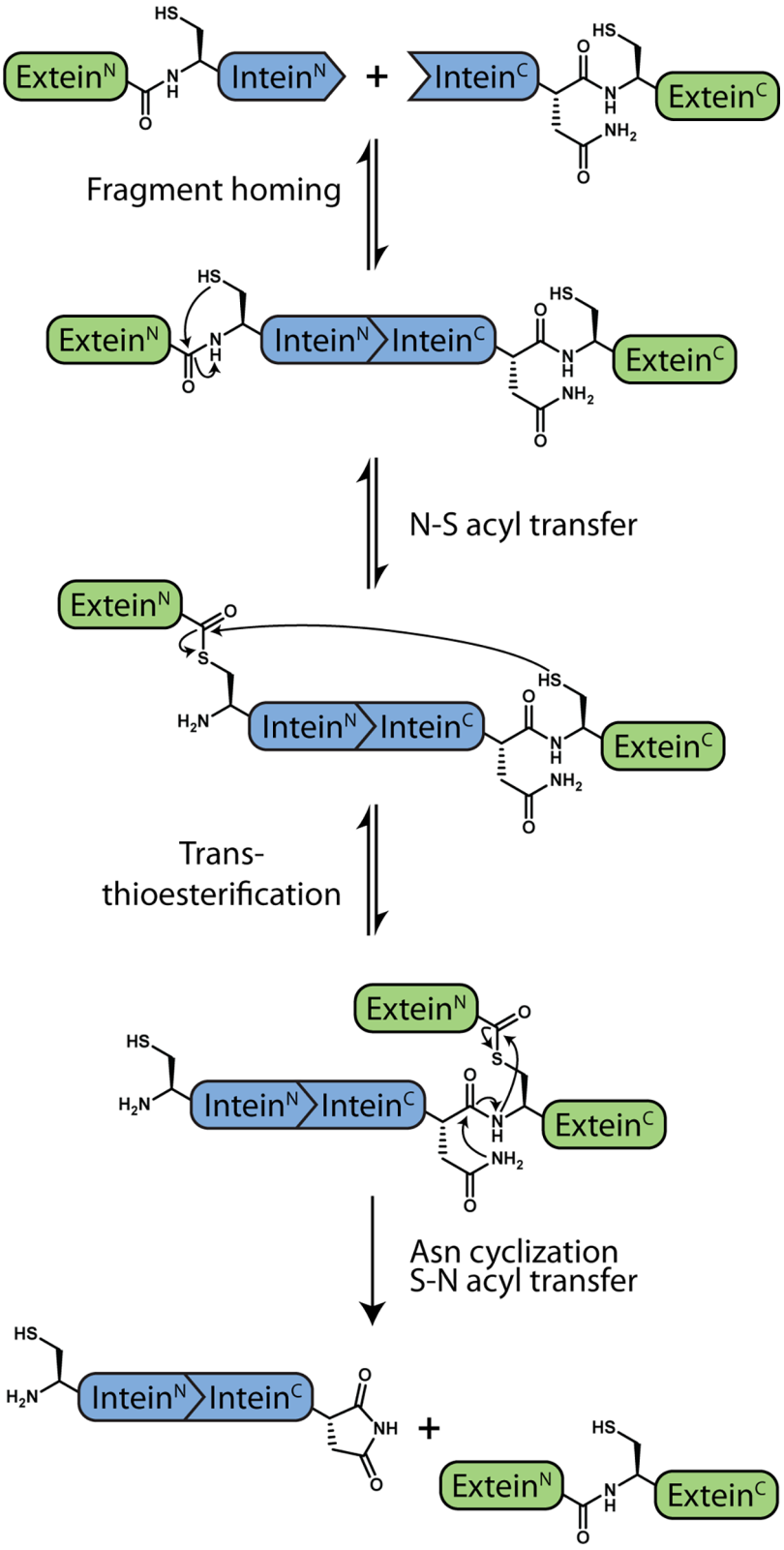


Figure 1. Intein trans-splicing reaction cascade. Intein fragments assemble in solution, ligating their concomitant protein cargos, exteins, in a series of highly choreographed reaction steps.

installable cargos. For example, a seminal early report by the Francis group outlined a powerful transamination-based strategy to install reactive carbonyls at protein N-termini through the use of a pyridoxal-5-phosphate (PLP) reagent; however, this strategy is constricted to N-terminal residues that do not react with aldehydes.[15] Another method developed by the group deployed potassium ferricyanide-based coupling of oxidized o-phenols or o-catechols to N-terminally functionalize proteins, potentially in a single step; this method, however, relies heavily on the presence of an N-terminal proline when

https://doi.org/10.1021/acs.bioconjchem.4c00223
Bioconjugate Chem. XXXX, XXX, XXX−XXX

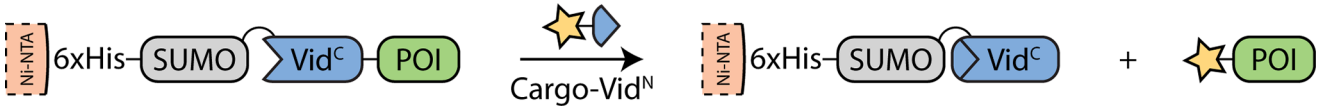


Figure 2. One-step protein biofunctionalization and purification of N-modified bioconjugates via VEPL. In VidaL-tagged expression and protein ligation (VEPL), proteins of interest (POI) are expressed as a C-fusion with Vid^C, SUMO, and a 6xHis tag. After chromatographic isolation on the Ni-NTA resin, a functionalized Vid^N peptide promotes trans-splicing, concomitantly N-functionalizing the POI while displacing it from the resin.

applied to full-length proteins, which may limit its scope of applicability.[16] Another approach developed by the Bordusa group engineered a "trypsiligase" as a ligase derived from trypsin to modify proteins at their N-termini. While robust, it holds the risk of context-specific hydrolysis side-reactions, hampering straightforward off-the-shelf use for different proteins.[17] Analogously, the Wells group pioneered the deployment of subtiligase-based peptide ligation to target protein N-termini.[18,19] However, this methodology, given its targeting of reactive amines, is also liable to inadvertently functionalize lysines throughout the target polypeptide chain; this lack of specificity limits it from achieving pristine regioselectivity. Lastly, a reliable suite of methods includes relatively recent fragment-reconstitution-based approaches such as the SpyTag-SpyCatcher chemistry and related platforms that rely on isopeptide bond formation between split pairs. While these may be convenient for protein−protein ligation at termini, such strategies typically result in the formation of bulky protein assemblies at the ligation site, thereby hampering native protein reconstitution, and require the sequential expression and purification of multiple constructs separately prior to assembly.[20−25]

In an effort to overcome these limitations, we identified split inteins as a uniquely powerful platform for protein−protein ligation.[26,27] Split inteins are protein elements that are able to self-assemble with high affinity upon mixing, ligating their fused polypeptides (termed exteins) with a native amide bond while "splicing" themselves out of the final fold in a series of highly choreographed reaction steps (Figure 1).[28] In comparison to other systems, inteins hold the benefit of being a virtually scarless modality, theoretically leaving as little as a single amino acid scar (typically a cysteine or serine) in the final product depending on the specific intein adopted. Moreover, split inteins are categorized as either canonical splits—wherein the N-terminal intein is larger than its corresponding C-intein—or atypical splits, where the reverse is true. This relative difference in sizes affords a powerful benefit to the incorporation of synthetic handles onto proteins; the larger fragment (N or C, depending on the type of split intein) can be expressed heterologously, while the shorter fragment may be accessible to solid-phase peptide synthesis, obviating the limitations of metabolic incorporation of non-natural groups and making possible the installation of virtually any synthetic handle(s) or probe post-translationally.[29] Owing to these powerful advantages, split inteins can be harnessed as a base platform toward the purification and functionalization of diverse proteins.[30,31] Specifically, early pioneering work by the Wood group led to the development of a pH-activatable intein cleavage system (a miniaturized *Mtu-recA* with a deleted endonuclease domain) for the purification of tag-free proteins,[32] a thrust which has since improved substantially in scope and reaction profile with the discovery and evolution of newer and better-behaved canonically split intein systems (e.g., *Npu* DnaE).[33,34] In the same vein, such systems have also lent themselves well to a growing body of work targeting the C-terminal modification of different proteins with manifold handles, either through regular splicing[35] or controlled cleavage through expressed protein ligation.[36,37] Summarily, while the literature is now rich in intein-based methods that enable the tagless purification and/or C-terminal modification of protein cargos, no such robust methodology has been developed for the single-step purification and N-terminal functionalization of proteins. This is an essential unmet need for cases where the C-terminus is necessary for protein function (e.g., Ubiquitin,[38] Interferon-gamma,[39] among others) or proves persistently refractory to post-translational modification.

RESULTS AND DISCUSSION

In looking to develop a generalizable and modular method for the single-step N-terminal modification of proteins, we were encouraged by a recent report by the Muir lab characterizing an atypical split intein isolated from Lake Vida in Antarctica, termed VidaL.[40] It holds the benefit of being a very short naturally occurring split intein; the N fragment of VidaL (Vid^N) has only 16 residues, rendering it readily accessible to solid-phase peptide synthesis in good yield. Moreover, it is among the fastest splicing intein systems discovered to date, with a reported reaction half-life of approximately a minute. The requisite extein load for VidaL is also an advantage compared to other intein systems and semisynthesis approaches more broadly: its N-terminal extein sequence is an E-S-G tripeptide and its corresponding C-terminal sequence is S-G-K. Functionally, this results in a final splicing "scar" that is a flexible serine-glycine linker, often adopted in the design of fusion chimeras as an innocuous spacer motif.[41] Lastly, while broader use of atypically split inteins has thus far been limited by their increased propensity toward C-terminal cleavage,[42] we were drawn to VidaL for its reported negligible tendency to undergo this side-reaction.

Specifically, we sought to exploit VidaL's splicing efficiency to both generate and purify homogeneous populations of monofunctionalized proteins in a single step. Our method—referred to as VidaL-tagged expression and protein ligation (VEPL)—relies on expressing proteins of interest as C-terminal fusions to a polyhistidine-tagged and SUMO-fused C-fragment of VidaL (Vid^C) protein and binding these onto immobilized metal affinity chromatography supports (in this case, a Ni-NTA matrix) (Figure 2). Introduction of the synthetic N-fragment modified with a cargo of interest then "cleaves" the target modified protein off the matrix via trans-splicing, thereby enabling clean retrieval of pure, modified product from the column flow-through and subsequent washes. The highly folded structure of Vid^C makes it a suitable fusion partner to different proteins-of-interest. Additionally, the small size of Vid^N (1,783.2 Da) paves the way for virtually any synthetic handle to be installed through routine peptide synthesis. As proposed, VEPL should prove broadly applicable regardless of the specific affinity chromatography system being used.

https://doi.org/10.1021/acs.bioconjchem.4c00223
Bioconjugate Chem. XXXX, XXX, XXX−XXX

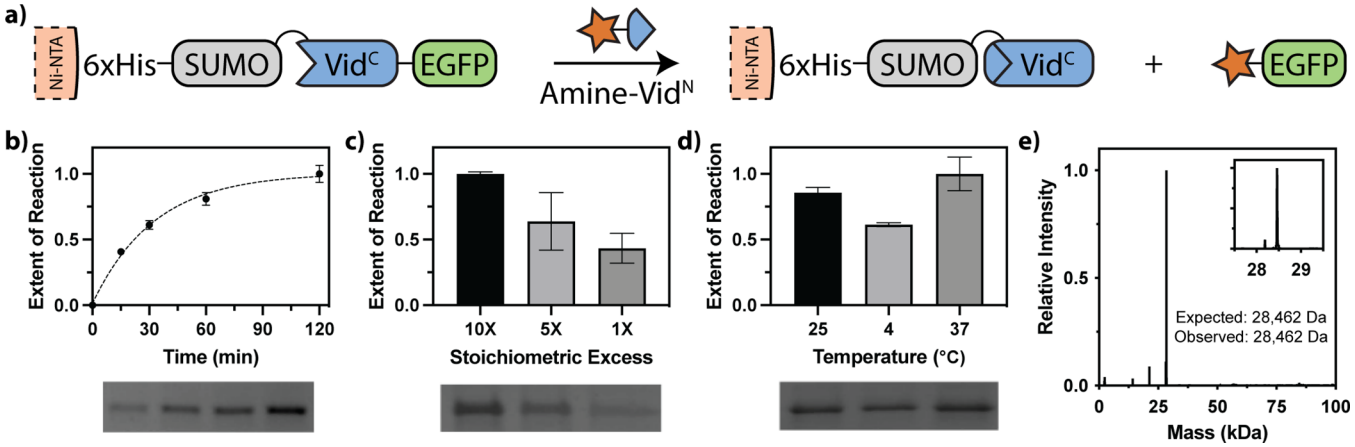


Figure 3. VEPL modification kinetics depend on stoichiometry and temperature. (a) EGFP was modified with an amine-functionalized Vid^N peptide via VEPL. (b) Trans-splicing proceeds quickly on resin, here at 25 °C with 10× stoichiometric excess of the amine-Vid^N peptide, as quantified through SDS-PAGE densitometry. Representative Coomassie-stained gel outlining reaction progress (15, 30, 60, 120 min from left to right). (c) Holding reaction time (120 min) and temperature (25 °C) constant, extent of trans-splicing scales with amine-Vid^N peptide excess. Representative Coomassie-stained gel depicts 10×, 5×, 1× excess from left to right. (d) With constant peptide excess (10×) and trans-splicing time (120 min), extent of trans-splicing varied with temperature. Coomassie-stained gel depicts 25, 4, and 37 °C from left to right. (e) N-terminal amine-tagged EGFP was obtained in high purity and with the expected molecular weight.

A known shortcoming of intein systems generally and atypically split inteins in particular is their varying propensity toward nonspecific N- and/or C-terminal cleavage over splicing. To assess whether VidaL fragment assembly and splicing occur with minimal C-terminal cleavage, we applied two different approaches. The first entailed the expression of a Vid^C-fused EGFP protein (Supplementary Method S1), which was then incubated with Ni-NTA and bound to a fritted column. No cleaved protein is obtained after 2 h on-resin, demonstrating complete absence of C-terminal cleavage (Supplementary Figure S1). Going further, and in order to ensure that cleavage of the system is not caused by a recapture of the intein domain fold,[43] we expressed a maltose-binding protein (MBP)-fused Vid^N and purified it using amylose chromatography (Supplementary Method S2). After expression of a Vid^C-EGFP fusion and incubation on column with Ni-NTA, we introduced the MBP-Vid^N protein and allowed the system to react for 2 h with varied temperatures prior to collecting reaction products in the column flow-through. Encouragingly, no C-terminally cleaved product could be discriminated via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), proving that no C-terminal cleavage takes place even after formation of the assembled intein domain under three different temperature conditions (i.e., 4 °C, 25 °C, 37 °C) (Supplementary Figure S2).

Heartened by these results, and as a prelude to establishing the versatility of the VEPL platform, we aimed to characterize its efficiency under various other reaction conditions: namely, by investigating splicing kinetics, temperature-dependence, and sensitivity to excess peptide stoichiometric ratios (Figure 3). Using EGFP as our model protein and an amine-Vid^N (Supplementary Methods S3−S6) as cognate peptide cargo for proof-of-concept characterization studies, we first conducted a trans-splicing time-course analysis, using a 10× excess of Vid^N under 25 °C ambient temperature. Encouragingly, despite the steric hindrances potentially introduced by protein-resin binding, the splicing reaction proceeded to completion within 2 h. Next, we investigated VEPL's temperature-dependence by running a splicing time-course for 2 h and a 10× excess of Vid^N to Vid^C at different temperatures (4 °C, 25 °C, 37 °C). We found that optimal splicing yield occurs at 37 °C but that meaningful extents of on-resin functionalization (i.e., 50% that of optimal yield) was achieved at 4 °C, highly encouraging for the system's use for thermal-sensitive proteins. Lastly, the effect of peptide excess was interrogated by varying the relative molar ratios of peptide-to-protein (1×, 5×, 10×) while keeping a constant reaction time of 2 h at 25 °C. While typical enzymatic routes to protein modification often require large excesses of the nucleophilic peptide substrate, VidaL can achieve high splicing yields even under equimolar conditions. Optimal stoichiometry is a 10× ratio of Vid^N peptide to the Vid^C-tagged protein, but maintaining as low as a 1:1 ratio still enables roughly 50% yield, which can prove fruitful for more esoteric peptide cargos or those in short supply. Uncropped gels showcasing these assays are provided in Supplementary Figure S3.

To further establish and demonstrate VEPL's broad utility, we next expressed four proteins spanning three different categories: Enhanced Green Fluorescent Protein (EGFP) and mCherry fluorescent proteins, β-lactamase (bla) enzyme, and epidermal growth factor (EGF), each as Vid^C fusions (Figure 4, Supplementary Method S3). Additionally, we exploited solid-phase peptide synthesis to prepare four peptides harboring different N-terminal cargos: a primary amine, a biotin affinity ligand, an alkyne click handle, and a dual cargo modified with both biotin and fluorescein (FAM), each fused to Vid^N (Supplementary Methods S3−S6). Using broadly optimal on-resin splicing conditions identified with Vid^C-EGFP and amine-Vid^N, we found that each protein proved amenable to VEPL with our introduced library of synthetic handles; correctly modified species were obtained in each case and verified via mass spectrometry and Coomassie-stained SDS-PAGE (Supplementary Tables S1 and S2, Supplementary Figures S4−S5). Importantly, all SUMO-Vid^C-POI constructs were expressed directly in soluble form in aqueous buffers, in good purity, and without harsh denaturing agents and subsequent refolding steps, which had proven to be key impediments to more widespread use of prior split intein-based systems. Functionalized proteins were also purified and isolated in good yield; EGFP-, mCherry-, and bla-based

https://doi.org/10.1021/acs.bioconjchem.4c00223
Bioconjugate Chem. XXXX, XXX, XXX−XXX

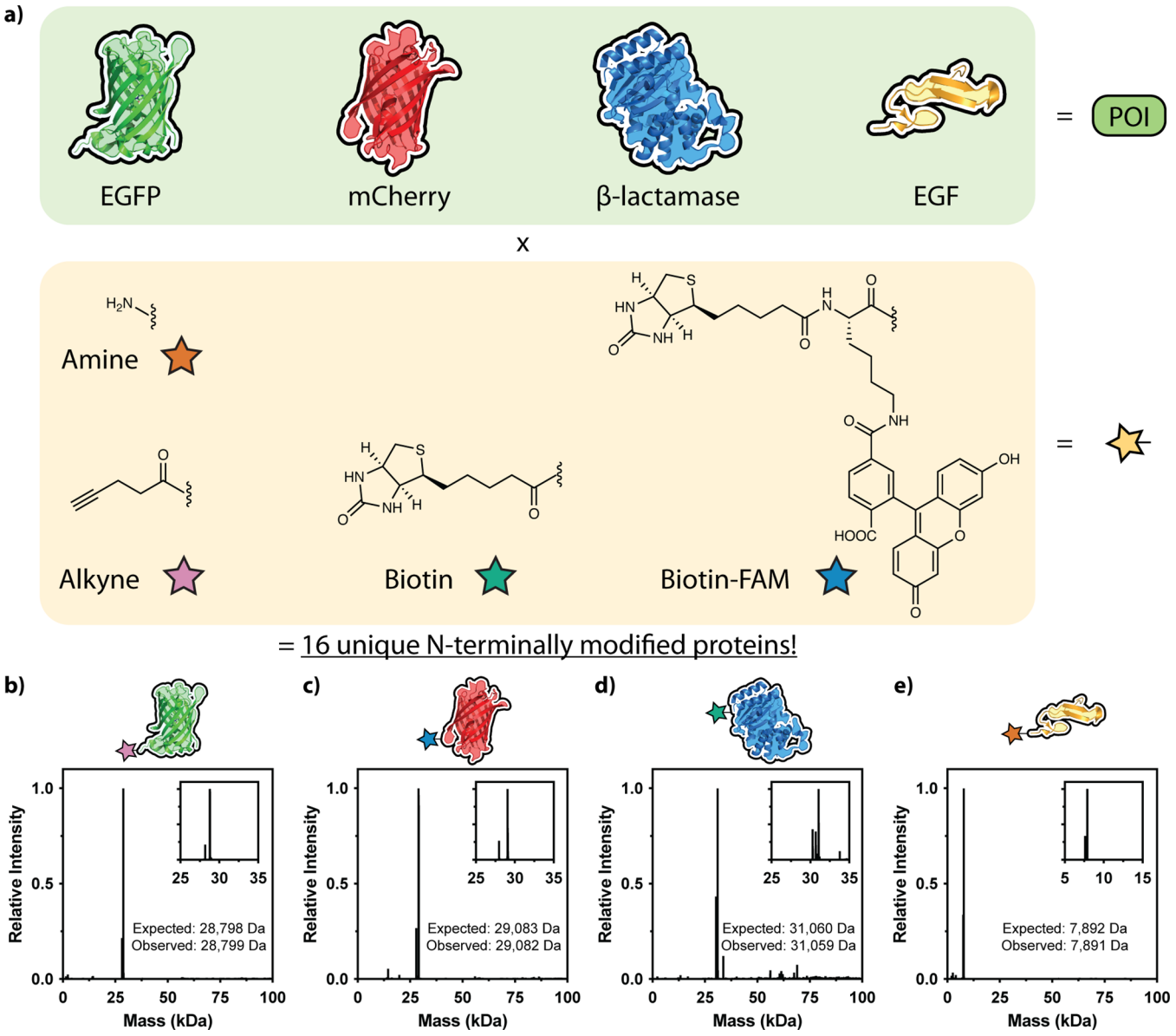


Figure 4. VEPL is compatible with a wide range of proteins and cargos. (a) VEPL was demonstrated using 4 different proteins (i.e., EGFP, mCherry, bla, EGF) and 4 distinctly functionalized Vid^N peptide cargos, yielding 16 unique N-terminal modified proteins. Species were obtained in high purity and with the expected molecular weight, with mass spectrometry analysis shown for (b) alkyne-EGFP, (c) biotin-FAM-mCherry, (d) biotin-bla, and (e) amine-EGF.

species were obtained at approximately 10 mgs/L of culture, and EGF-based species were obtained at 1−2 mgs/L of culture.

With the full library of N-modified proteins in hand, we next sought to assess their bioactivity. For EGFP and mCherry, we quantified fluorescence as a surrogate measure of bioactivity for the amine-, alkyne-, and biotin-modified constructs; studies involving Biotin-FAM modification of EGFP and mCherry were excluded due to fluorescence crosstalk (Figure 5a-b). EGFP fluorescence was assessed using $\lambda_{excitation}$ = 490 nm, and $\lambda_{emission}$ = 525 nm. Similarly, mCherry fluorescence was determined with $\lambda_{excitation}$ = 575 nm, and $\lambda_{emission}$ = 610 nm. Encouragingly, each of the distinctly N-functionalized proteins displayed statistically indistinguishable fluorescence, consistent with our starting hypothesis.

Having observed unaltered activity in fluorescent proteins, we then shifted our attention to the more sensitive bioconstructs. Since VEPL does not require harsh denaturants or other additives, we reasoned that this method could be readily extended to any polypeptide class with no loss of translational promise, particularly those that often prove refractory to functionalization.

For bla, taken as model enzyme, we applied a colorimetric assay based on the hydrolysis of nitrocefin (Supplementary Method S7), a chromogenic substrate which changes color from yellow to red upon cleavage of its beta-lactam bond (Figure 5c). Time-course colorimetric read-outs established that statistically indistinguishable bla activity was retained regardless of the N-terminal cargo installed (Figure 5c).

Lastly, we applied VEPL to the model growth factor EGF, as growth factors are often both of prime interest and challenging to functionalize and post-translationally modify. EGF bioactivity was assessed based on its capacity to stimulate the proliferation of HeLa cells through the RealTime-Glo MT Cell

https://doi.org/10.1021/acs.bioconjchem.4c00223
Bioconjugate Chem. XXXX, XXX, XXX−XXX

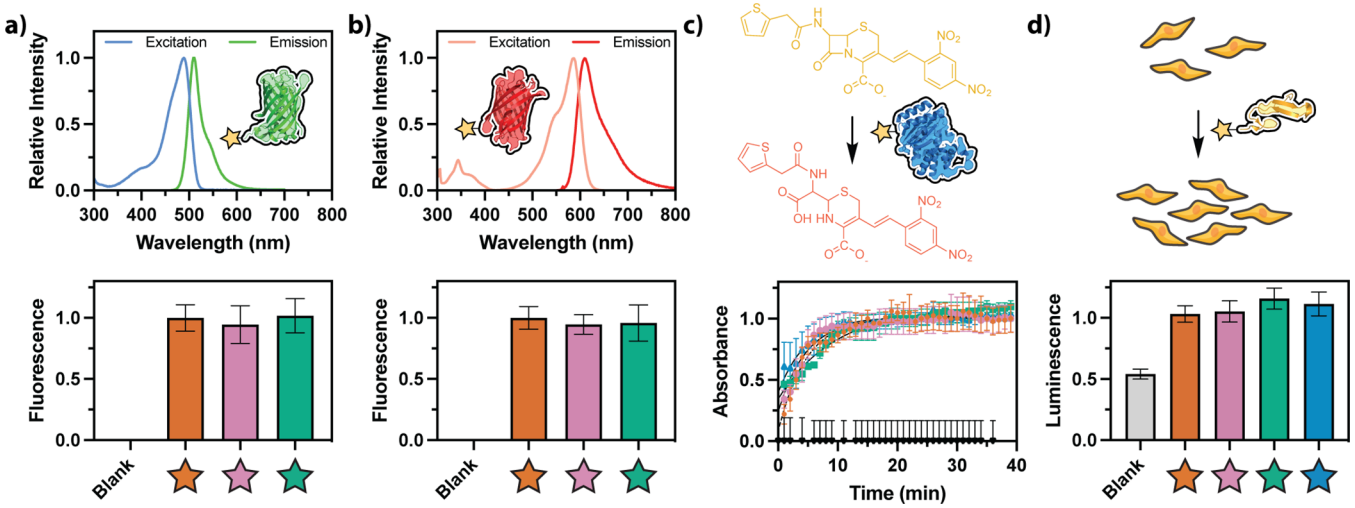


Figure 5. VEPL-modified proteins exhibit native bioactivity. The activity of proteins N-modified via VEPL with amine, alkyne, biotin, or biotin-FAM was assessed (denoted respectively with orange, pink, green, and blue stars and/or individual data point). (a,b) EGFP and mCherry fluorescence was used as a surrogate of their respective activity. (c) Bla activity was determined by its ability to degrade a chromogenic nitrocefin substrate, whereby β-lactam cleavage yields a yellow-to-red color change. (d) EGF activity was assayed by its ability to stimulate HeLa cell proliferation, quantified using a RealTime-Glo MT cell viability assay. In all cases, N-modified protein bioactivity was statistically indistinguishable from native levels. Studies involving Biotin-FAM modification of EGFP and mCherry were excluded due to fluorescence crosstalk. Error bars correspond to the SE about the mean for biological replicate experiments (n = 3 for EGFP, mCherry, and Bla studies; n = 5 for EGF).

Viability Assay (Promega) (Supplementary Methods S8 and S9), which provides a metabolism-based count of viable cells through a bioluminescent assay. Bioactivity of the N-terminally modified EGFs proved unchanged, which is highly encouraging given its relative fragility (Figure 5d). We note that the "blank" treatment displays roughly half the luminescence of the EGF-treated samples, reflecting a smaller number of metabolically active cells present in this condition.

CONCLUSION

In this study, we have introduced the VEPL platform to generate and purify a suite of N-terminally modified proteins in a single step. Taking advantage of the atypically split VidaL intein whose N-fragment is readily accessible to solid-phase synthesis, VEPL permits diverse and multifunctional cargos to be N-terminally installed on a wide range of proteins all while maintaining native levels of bioactivity. Provided a protein-of-interest can be expressed as a Vid^C fusion, VEPL may N-terminally functionalize any POI regardless of the specific amino acid sequence present at the N-terminus. As such, we expect this method to prove broadly useful in the creation of semisynthetic protein chimeras for a host of applications that interface materials and polypeptide-based assemblies.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.bioconjchem.4c00223.

Experimental methods, nucleotide sequences, SDS-PAGE validation, mass spectra, supplementary figures, and tables. (PDF)

AUTHOR INFORMATION

Corresponding Author

Cole A. DeForest − *Department of Chemical Engineering, University of Washington, Seattle, Washington 98105, United States; Department of Bioengineering, Department of Chemistry, Institute of Stem Cell & Regenerative Medicine, Molecular Engineering & Sciences Institute, and Institute for Protein Design, University of Washington, Seattle, Washington 98105, United States;* orcid.org/0000-0003-0337-3577; Email: profcole@uw.edu

Authors

Ryan Gharios − *Department of Chemical Engineering, University of Washington, Seattle, Washington 98105, United States*

Annabella Li − *Department of Chemical Engineering, University of Washington, Seattle, Washington 98105, United States;* orcid.org/0009-0006-1808-0389

Irina Kopyeva − *Department of Bioengineering, University of Washington, Seattle, Washington 98105, United States*

Ryan M. Francis − *Department of Chemical Engineering, University of Washington, Seattle, Washington 98105, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.bioconjchem.4c00223

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Tom Muir and Dr. Anthony Burton for helpful discussions on intein chemistry and for providing us with different intein plasmid constructs. We would also like to thank Dr. Drew Sellers for expert guidance on peptide synthesis and modification. This work was supported by a Faculty Early Career Development CAREER Award (DMR 1652141, C.A.D.) and a grant (DMR 1807398, C.A.D.) from the National Science Foundation (NSF), and a Maximizing Investigators' Research Award (R35GM138036, C.A.D.) from the National Institutes of Health (NIH). Student support was further provided through a John C. Berg Endowed Fellowship (R.G.), an NSF Graduate Student Research

Fellowship (DGE 1762114, I.K.), an NIH Interdisciplinary Training Grant (T32CA080416, I.K.), and an ITHS-TL1 Fellowship (R.M.F.).

REFERENCES

(1) Tamura, T.; Hamachi, I. Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems. *J. Am. Chem. Soc.* **2019**, *141* (7), 2782−2799.
(2) Peng, T.; Hang, H. C. Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* **2016**, *138* (43), 14423−14433.
(3) Freedy, A. M.; Matos, M. J.; Boutureira, O.; Corzana, F.; Guerreiro, A.; Akkapeddi, P.; Somovilla, V. J.; Rodrigues, T.; Nicholls, K.; Xie, B.; Jiménez-Osés, G.; Brindle, K. M.; Neves, A. A.; Bernardes, G. J. L. Chemoselective Installation of Amine Bonds on Proteins through Aza-Michael Ligation. *J. Am. Chem. Soc.* **2017**, *139* (50), 18365−18375.
(4) Fisher, S. A.; Baker, A. E. G.; Shoichet, M. S. Designing Peptide and Protein Modified Hydrogels: Selecting the Optimal Conjugation Strategy. *J. Am. Chem. Soc.* **2017**, *139* (22), 7416−7427.
(5) Spicer, C. D.; Pashuck, E. T.; Stevens, M. M. Achieving Controlled Biomolecule-Biomaterial Conjugation. *Chem. Rev.* **2018**, *118* (16), 7702−7743.
(6) Shadish, J. A.; DeForest, C. A. Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter* **2020**, *2* (1), 50−77.
(7) Gharios, R.; Francis, R. M.; DeForest, C. A. Chemical and Biological Engineering Strategies to Make and Modify Next-Generation Hydrogel Biomaterials. *Matter* **2023**, *6* (12), 4195−4244.
(8) Thompson, R. E.; Muir, T. W. Chemoenzymatic Semisynthesis of Proteins. *Chem. Rev.* **2020**, *120* (6), 3051−3126.
(9) Taylor, R. J.; Geeson, M. B.; Journeaux, T.; Bernardes, G. J. L. Chemical and Enzymatic Methods for Post-Translational Protein-Protein Conjugation. *J. Am. Chem. Soc.* **2022**, *144* (32), 14404−14419.
(10) Warden-Rothman, R.; Caturegli, I.; Popik, V.; Tsourkas, A. Sortase-Tag Expressed Protein Ligation: Combining Protein Purification and Site-Specific Bioconjugation into a Single Step. *Anal. Chem.* **2013**, *85* (22), 11090−11097.
(11) Shadish, J. A.; Benuska, G. M.; DeForest, C. A. Bioactive Site-Specifically Modified Proteins for 4D Patterning of Gel Biomaterials. *Nat. Mater.* **2019**, *18* (9), 1005−1014.
(12) Marino, G.; Eckhard, U.; Overall, C. M. Protein Termini and Their Modifications Revealed by Positional Proteomics. *ACS Chem. Biol.* **2015**, *10* (8), 1754−1764.
(13) Rosen, C. B.; Francis, M. B. Targeting the N Terminus for Site-Selective Protein Modification. *Nat. Chem. Biol.* **2017**, *13* (7), 697−705.
(14) Jacob, E.; Unger, R. A Tale of Two Tails: Why Are Terminal Residues of Proteins Exposed? *Bioinformatics* **2007**, *23* (2), e225−e230.
(15) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angew. Chem., Int. Ed.* **2006**, *45* (32), 5307−5311.
(16) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. N-Terminal Modification of Proteins with o-Aminophenols. *J. Am. Chem. Soc.* **2014**, *136* (27), 9572−9579.
(17) 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* (11), 3024−3028.
(18) Weeks, A. M.; Wells, J. A. Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* **2020**, *120* (6), 3127−3160.
(19) Weeks, A. M.; Wells, J. A. N-Terminal Modification of Proteins with Subtiligase Specificity Variants. *Curr. Protoc. Chem. Biol.* **2020**, *12* (1), No. e79.
(20) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (12), E690−E697.
(21) Veggiani, G.; Nakamura, T.; Brenner, M. D.; Gayet, R. V.; Yan, J.; Robinson, C. V.; Howarth, M. Programmable Polyproteams Built Using Twin Peptide Superglues. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (5), 1202−1207.
(22) Keeble, A. H.; Banerjee, A.; Ferla, M. P.; Reddington, S. C.; Anuar, I. N. A. K.; Howarth, M. Evolving Accelerated Amidation by SpyTag/SpyCatcher to Analyze Membrane Dynamics. *Angew. Chem., Int. Ed.* **2017**, *56* (52), 16521−16525.
(23) Keeble, A. H.; Turkki, P.; Stokes, S.; Anuar, I. N. A. K.; Rahikainen, R.; Hytönen, V. P.; Howarth, M. Approaching Infinite Affinity through Engineering of Peptide-Protein Interaction. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (52), 26523−26533.
(24) Buldun, C. M.; Jean, J. X.; Bedford, M. R.; Howarth, M. SnoopLigase Catalyzes Peptide-Peptide Locking and Enables Solid-Phase Conjugate Isolation. *J. Am. Chem. Soc.* **2018**, *140* (8), 3008−3018.
(25) Keeble, A. H.; Yadav, V. K.; Ferla, M. P.; Bauer, C. C.; Chuntharpursat-Bon, E.; Huang, J.; Bon, R. S.; Howarth, M. DogCatcher Allows Loop-Friendly Protein-Protein Ligation. *Cell Chem. Biol.* **2022**, *29* (2), 339−350.
(26) Shah, N. H.; Muir, T. W. Split Inteins: Nature's Protein Ligases. *Isr. J. Chem.* **2011**, *51* (8−9), 854−861.
(27) Shah, N. H.; Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chem. Sci.* **2014**, *5* (2), 446−461.
(28) Mootz, H. D. Split Inteins as Versatile Tools for Protein Semisynthesis. *ChemBioChem.* **2009**, *10* (16), 2579−2589.
(29) Thiel, I. V.; Volkmann, G.; Pietrokovski, S.; Mootz, H. D. An Atypical Naturally Split Intein Engineered for Highly Efficient Protein Labeling. *Angew. Chem. - Int. Ed.* **2014**, *53* (5), 1306−1310.
(30) Prabhala, S. V.; Gierach, I.; Wood, D. W. The Evolution of Intein-Based Affinity Methods as Reflected in 30 Years of Patent History. *Front. Mol. Biosci.* **2022**, *9*, 857566.
(31) Wood, D. W.; Belfort, M.; Lennon, C. W. Inteins—Mechanism of Protein Splicing, Emerging Regulatory Roles, and Applications in Protein Engineering. *Front. Microbiol.* **2023**, *14*, 1305848.
(32) Wood, D. W.; Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M. A Genetic System Yields Self-Cleaving Inteins for Bioseparations. *Nat. Biotechnol.* **1999**, *17* (9), 889−892.
(33) Guan, D.; Ramirez, M.; Chen, Z. Split Intein Mediated Ultra-Rapid Purification of Tagless Protein (SIRP). *Biotechnol. Bioeng.* **2013**, *110* (9), 2471−2481.
(34) Xia, H.-F.; Zhou, T.-J.; Du, Y.-X.; Wang, Y.-J.; Shi, C.-H.; Wood, D. W. Improved Protein Purification System Based on C-Terminal Cleavage of Npu DnaE Split Intein. *Bioprocess Biosyst. Eng.* **2020**, *43* (11), 1931−1941.
(35) Volkmann, G.; Liu, X.-Q. Protein C-Terminal Labeling and Biotinylation Using Synthetic Peptide and Split-Intein. *PLoS One* **2009**, *4* (12), No. e8381.
(36) Graulus, G.-J.; Ta, D. T.; Tran, H.; Hansen, R.; Billen, B.; Royackers, E.; Noben, J.-P.; Devoogdt, N.; Muyldermans, S.; Guedens, W.; Adriaensens, P. Site-Selective Functionalization of Nanobodies Using Intein-Mediated Protein Ligation for Innovative Bioconjugation. In *Bioconjugation: Methods and Protocols*; Massa, S.; Devoogdt, N., Eds.; Springer: New York, NY, 2019; pp 117−130. DOI: 10.1007/978-1-4939-9654-4_9.
(37) Roller, R. F.; Malik, A.; Carillo, M. A.; Garg, M.; Rella, A.; Raulf, M.-K.; Lepenies, B.; Seeberger, P. H.; Varón Silva, D. Semisynthesis of Functional Glycosylphosphatidylinositol-Anchored Proteins. *Angew. Chem., Int. Ed.* **2020**, *59* (29), 12035−12040.
(38) Reyes-Turcu, F. E.; Horton, J. R.; Mullally, J. E.; Heroux, A.; Cheng, X.; Wilkinson, K. D. The Ubiquitin Binding Domain ZnF UBP Recognizes the C-Terminal Diglycine Motif of Unanchored Ubiquitin. *Cell* **2006**, *124* (6), 1197−1208.

https://doi.org/10.1021/acs.bioconjchem.4c00223
Bioconjugate Chem. XXXX, XXX, XXX−XXX

(39) Szente, B. E.; Soos, J. M.; Johnson, H. M. The C-Terminus of IFN-γ Is Sufficient for Intracellular Function. *Biochem. Biophys. Res. Commun.* **1994**, *203* (3), 1645−1654.
(40) Burton, A. J.; Haugbro, M.; Parisi, E.; Muir, T. W. Live-Cell Protein Engineering with an Ultra-Short Split Intein. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (22), 12041−12049.
(41) Ceballos-Alcantarilla, E.; Merkx, M. Chapter One - Understanding and Applications of Ser/Gly Linkers in Protein Engineering. In *Methods in Enzymology*; Merkx, M., Ed.; Linkers in Biomacromolecules; Academic Press, 2021; Vol. *647*, pp 1−22. DOI: 10.1016/bs.mie.2020.12.001.
(42) Friedel, K.; Popp, M. A.; Matern, J. C. J.; Gazdag, E. M.; Thiel, I. V.; Volkmann, G.; Blankenfeldt, W.; Mootz, H. D. A Functional Interplay between Intein and Extein Sequences in Protein Splicing Compensates for the Essential Block B Histidine. *Chem. Sci.* **2019**, *10* (1), 239−251.
(43) Ludwig, C.; Schwarzer, D.; Zettler, J.; Garbe, D.; Janning, P.; Czeslik, C.; Mootz, H. D. Chapter 4 Semisynthesis of Proteins Using Split Inteins. In *Methods in Enzymology*; Muir, T. W.; Abelson, J. N., Eds.; Non-Natural Amino Acids; Academic Press, 2009; Vol. *462*, pp 77−96. DOI: 10.1016/S0076-6879(09)62004-8.

https://doi.org/10.1021/acs.bioconjchem.4c00223
Bioconjugate Chem. XXXX, XXX, XXX−XXX