Traceless enzymatic protein synthesis without ligation sites constraint

Ruifeng Li  
Institute of Microbiology

Marcel Schmidt  
Fresenius Kabi

Tong Zhu  
Institute of Microbiology, CAS  https://orcid.org/0000-0001-7857-8564

Xinyu Yang  
Institute of Microbiology, CAS

Jing Feng  
Institute of Microbiology, CAS

Tian Yue  
Institute of Microbiology, Chinese Academy of Sciences

Yinglu Cui  
Institute of Microbiology  https://orcid.org/0000-0002-2147-5879

Timo Nuijens  
Fresenius Kabi

Bian Wu  (wub@im.ac.cn)  
State Key Laboratory of Microbial Resources & CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences  https://orcid.org/0000-0002-6524-2049

Article

Keywords: enzymatic protein synthesis, biochemistry

DOI: https://doi.org/10.21203/rs.3.rs-143176/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Protein synthesis and semisynthesis offer immense promise for life science and have impacted pharmaceutical innovation. Nevertheless, the absence of a generally applicable method for traceless peptide conjugation with flexible choices of junction sites remains a bottleneck to accessing many important synthetic targets. Here we introduce the protein activation and ligation with multiple enzymes (PALME) platform designed for the sequence-unconstrained synthesis and modification of biomacromolecules. The upstream activating modules accept and process easily accessible synthetic peptides and recombinant proteins, avoiding the challenges associated with the preparation and manipulation of activated peptide substrates. Cooperatively, the downstream coupling module provides comprehensive solutions for sequential peptide condensation, cyclization, and protein N/C-terminal or internal functionalization. This methodology’s practical utility was showcased by synthesizing a series of bioactive targets, ranging from pharmaceutical ingredient to synthetically challenging protein. Together, the modular PALME platform exhibits unprecedented broad accessibility for the traceless protein synthesis and functionalization and holds enormous potential to extent the scope of protein chemistry and synthetic biology.

Introduction

Breaking away from the central dogma, protein total synthesis and semisynthesis are powerful strategies for generating and functionalizing naturally inaccessible proteins, which have enabled groundbreaking applications driving life science advances and impacted the industrial production of biomolecular therapeutics. Classic solid-phase peptide synthesis (SPPS) and emerging automated fast-flow peptide synthesis (AFPS) have provided effective means of incorporating the full spectrum of chemical functional groups at desired locations while assembling amino acids into a peptide chain, but synthesizing longer peptides and proteins has been considered an arduous task due to the exponential decrease in overall yield as the number of residues increases. To acquire a larger-sized protein, it is generally more efficient to divide the whole polypeptide chain into several fragments and couple them sequentially or to ligate synthesized peptides to recombinant protein segments. Therefore, for the past two decades, the development of protein synthesis has centered on how to perform the selective ligation of unprotected peptides in aqueous phase in the presence of the full diversity of reactive functionalities within peptide side chains. In this context, following the principle of tandem chemoselective capture and intramolecular rearrangement, ingenious chemical methods such as native chemical ligation (NCL), ketoacid-hydroxylamine ligation (KAHA), serine/threonine ligation (STL) and diselenide selenoester ligation (DSL) have been devised. In addition to chemical methods, biocatalytic strategies have gained increased attention due to their inherent properties like excellent regio- and chemoselectivity. Alongside split intein tools, genomic mining and protein engineering have led to the discovery and refinement of Sortase-A, Butelase-1, and other transpeptidases bearing similar mechanism. These peptide ligation strategies have created a flourishing field of protein synthesis and functionalization with
thousands of cases\textsuperscript{14}. Nevertheless, the possible retrosynthetic disconnections are still limited (Fig. 1a), remaining demanding challenges for accessing bountiful biomacromolecules of interest.

For releasing ligation site restrictions, subtilisin-derived ligase-catalyzed peptide coupling is deemed the most promising solution to date\textsuperscript{15} because this class of enzymes enables traceless amide bond formations with broad sequence compatibility (Fig. 1b). In the 1990s, Wells and coworkers created an enzyme called Subtiligase by introducing S221C and P225A mutations into subtilisin BPN'. Subtiligase was mechanistically altered to favor aminolysis over hydrolysis, laying the foundation for the development of this ligation technology\textsuperscript{16}. Recently, we developed another cation-independent and exceptionally robust subtilisin-derived ligase, termed Peptiligase\textsuperscript{17}, to make this approach industrially viable. Peptiligase has remarkable catalytic efficiency and offers significantly high average ligation yields (up to 98\% in < 1 h). Further engineering efforts have resulted in a series of Peptiligase variants with either broad specificity that would maximize utility or tight specificity that would allow selective ligation\textsuperscript{18}. As demonstrated by the hundred-gram-scale synthesis of active pharmaceutical ingredient (API) in industrial settings, Peptiligase has proven to be applicable for the practical manufacturing of therapeutic peptides in a cost-efficient and environmentally sustainable manner\textsuperscript{19}.

However, the great potential of Peptiligase for further widespread applications in protein chemistry was severely restricted regarding the problems on preparation and manipulation of reactive handles (Fig. 1c). First, the required acyl donors bearing an active ester on the C-termini are sometimes challenging to prepare through SPPS. In addition, sequential enzymatic ligations for large synthetic targets are almost impractical because C-terminal protection of the acyl acceptor is necessary to prevent repeated condensation. More importantly, although several C-terminal functionalization strategies that rely on intein\textsuperscript{20} or specific sequences\textsuperscript{21–23} have been developed, the convenient activation of recombinant proteins for traceless ligation with broad sequence compatibility remains a longstanding challenge. Consequently, a Peptiligase-compatible method that enables regioselective C-terminal activation in the aqueous phase using easily accessible peptides and proteins is highly desired to overcome the constraints on realizing extensively practicable traceless protein synthesis and functionalization.

Nature brings forth sophisticated biomolecule systems for assembling free amino acids into proteins. In the course of evolution, the elegant collaboration of aminoacyl-tRNA synthetases (amino acid activation), tRNAs (intermediate ester formation), and ribosomes (amide-forming ligation) has allowed the precise assembly of diverse L-amino acids with unmodified α-amino groups and side chains. Revisiting the fascinating principle of nature and possessing the Peptiligase family for peptide ligation, we envisioned that sequence-independent assembly of native peptides might also be feasible through iterative activation and ligation processes using multiple enzymes that present both strict regioselectivity and broad substrate specificity. To this end, we sought to design a multienzyme cooperative activation and ligation strategy (Fig. 1d) for traceless protein synthesis and functionalization, and we now present this platform, termed protein activation and ligation with multiple enzymes (PALME). By unifying the acyl-shift chemistry from the enzymatic and chemical protein synthesis, the PALME platform accepts SPPS
products, i.e., peptide carboxylic acids, amides or hydrazides, and recombinant proteins as the input and presents proteins synthesized via sequential condensation, cyclization, protein N/C-terminal or internal functionalization as the output (Fig. 1e). Our results highlight that enzymes with diverse functions can be rationally harnessed to offer traceless protein synthesis and functionalization with remarkable flexibility in choice of the ligation sites and peptide substrates, providing unprecedentedly broad application potential.

Results And Discussion

Establishing compatibility between peptide-activating and peptide-coupling enzymes

At the outset of our studies, we searched for a broadly applicable enzyme for C-terminal peptide esterification, to provide accessible reactive handles for Peptiligase. Accordingly, we explored the peptide amidase (PAM) from *Stenotrophomonas maltophilia*, which affords sequence-independent C-terminal peptide modification with absolute regioselectivity. Using computational redesign, we significantly expanded the synthetic utility of PAM. However, after exhausting different protein engineering strategies, our surveys for a mutant that catalyzes direct esterification reactions in aqueous solution were unfruitful, leading to the requirement of a bridge joining the two biocatalysts. Thus, we began to consider hydrazide chemistry, which was implemented by the Liu group and has been one of the most widely used extensions to NCL. In this method, the thioester functionality of the acyl donor is initially masked in the form of a C-terminal hydrazide and sequentially retrieved via a combination of nitrite oxidation and thiolysis. We envisioned that this strategy might be adapted for Peptiligase-catalyzed ligation, while it was unclear if there is an appropriate alcohol reagent for peptide acyl shifting.

We initially explored the feasibility of using 2-hydroxyacetamide, which would afford the peptide carboxamidomethyl (Cam) ester (standard Peptiligase substrate), for intermediate ester formation. The model peptide hydrazide Ac-DFSKL-N₂H₃ was oxidized using sodium nitrite in an acidic buffer solution at -15 °C, producing a peptide azide. Subsequently, 2-hydroxyacetamide was added to form the corresponding peptide Cam ester. Finally, the acyl acceptor ALKKA-NH₂ (1.5 equiv.) and Omniligase-1 (0.003 equiv., a commercially available enzyme from the Peptiligase family) were added to the reaction mixture, and the ligation was allowed to proceed for 30 min at pH 8.5 and room temperature. To our delight, the desired ligation product Ac-DFSKLALKKA-NH₂ was formed (20% yield). This preliminary result demonstrated the possibility of using peptide hydrazide in Peptiligase-catalyzed ligation, while it was unclear if there is an appropriate alcohol reagent for peptide acyl shifting.

We initially explored the feasibility of using 2-hydroxyacetamide, which would afford the peptide carboxamidomethyl (Cam) ester (standard Peptiligase substrate), for intermediate ester formation. The model peptide hydrazide Ac-DFSKL-N₂H₃ was oxidized using sodium nitrite in an acidic buffer solution at -15 °C, producing a peptide azide. Subsequently, 2-hydroxyacetamide was added to form the corresponding peptide Cam ester. Finally, the acyl acceptor ALKKA-NH₂ (1.5 equiv.) and Omniligase-1 (0.003 equiv., a commercially available enzyme from the Peptiligase family) were added to the reaction mixture, and the ligation was allowed to proceed for 30 min at pH 8.5 and room temperature. To our delight, the desired ligation product Ac-DFSKLALKKA-NH₂ was formed (20% yield). This preliminary result demonstrated the possibility of using peptide hydrazide in Peptiligase-catalyzed peptide ligation in a one-pot approach, albeit with low ligation efficiency. Detailed analysis of intermediates in the cascade reactions revealed that multiple side products were formed during the esterification process, most likely due to Curtius rearrangement of the peptide azide. Therefore, for a high ligation yield and a clean reaction, it is crucial to rapidly convert the peptide azide into the corresponding ester with a strong nucleophile. The formed peptide ester is also expected to be stable in the weakly alkaline solution and fit the binding sites of Peptiligase. Besides, the alcohol should be a good leaving group for the enzymatic S-O exchange.
With these stipulations in mind, we investigated a panel of alcohols for their efficiency in the model 5 + 5 reaction (Fig. 2), including aliphatic alcohols, aromatic alcohols, fluoroalcohols, and 2-hydroxyacetamide analogues. Most of the tested alcohols were able to mediate peptide ligation with moderate overall yields (10%-50%). In particular, in phenols-mediated reactions, the peptide phenolic ester performed well in the ligation step. Accordingly, we further examined a series of phenol derivatives containing polar and electron-withdrawing substituents on the aromatic ring to improve their nucleophilicity and aqueous solubility. Gratifyingly, 4-hydroxybenzoic acid (HBA, e6), 4-hydroxyphenylacetic acid (e7), and 4-hydroxyphthalic acid (e8) gave almost quantitative conversion of the peptide ester within 5 min. The resulting peptide esters demonstrated good chemical stability under ligation conditions, and their ligation efficiencies (> 90%) were comparable to those of the standard peptide Cam esters. Considering its overall performance and commercial availability at scale, HBA was selected for further studies.

The scope of the enzymatic peptide hydrazide ligation was then investigated. To map the substrate profile of six binding pockets of Omniligase-1 in HBA-mediated peptide hydrazide ligations, we performed an extensive series of reactions with four acyl donor and two acyl acceptor site-saturation peptide libraries under identical reaction conditions (Fig. 3a). We were pleased to observe that the desired ligation products were obtained in all tested reactions of this thorough substrate scan, and a majority of ligations proceeded smoothly with high coupling yields (> 70%) in 1 h. Remarkably, using this activation-ligation approach, we were even able to ligate peptides that usually represent poor substrates for Omniligase-1, for example, P1 = Pro peptide, with over 80% coupling efficiency. These results indicated that the broad sequence compatibility of Omniligase-1 was well maintained and even partially improved in this newly devised HBA-mediated ligation process. In principle, the small portion of less efficient ligations may be accomplished using Peptiligase variants with different substrate profiles.

Having demonstrated the utilization of peptide hydrazides in Peptiligase-catalyzed ligation, we continued to investigate whether the peptide modifying enzyme PAM is feasible for converting the most basic SPPS products (peptide amides/carboxylic acids) into the corresponding peptide hydrazides, by using computationally redesigned PAM12B\textsuperscript{25}. Preliminary experiments showed that the hydrazidation of the model peptide Ac-DFSKL-NH\textsubscript{2} proceeded smoothly in aqueous solution at room temperature through kinetic control. In the presence of 0.5 M hydrazine and 0.00001 equiv. PAM12B, the conversion of the peptide amide was complete after 45 min, giving the hydrazidation product with 96% efficiency. Based on the initial success, we next evaluated the substrate scope of this peptide-activating enzyme. Structural analysis revealed that the peptide substrates interact with the enzyme mainly via the last two C-terminal residues. Accordingly, we investigated the substrate sequence preference of the two terminal residue binding pockets of PAM12B by performing hydrazidation reactions with two site-saturation peptide amide libraries. In most reactions, hydrazidation products (except for P1 = Pro or Asn peptides) could be obtained in over 90% yield in 1 h (Fig. 3b), demonstrating the desired versatility of the peptide-activating enzyme.

However, PAM12B-catalyzed modification of peptide carboxylic acids is practical only in organic environments (H\textsubscript{2}O < 10%) due to the thermodynamic barrier, which restricts the direct activation of
recombinant proteins. To overcome this severe limitation, we sought to recruit an additional biocatalytic module for the selective functionalization of the peptide or protein carboxyl terminus. In the animal kingdom, many secreted peptides are processed by a peptidyl-glycine oxidation system. During the transformation, peptidyl-glycine hydroxylating monooxygenase (PHM) catalyzes the stereospecific hydroxylation of the α-carbon of the terminal glycine with oxygen and ascorbate, and sequentially, one molecule of glyoxylate is removed by peptidyl-α-hydroxyglycine amidating lyase (PAL) to form the des-glycine peptide amide, which is the ideal substrate for PAM. With the expectation that the oxidative enzymes and hydrolytic enzymes might work cooperatively, we prepared PHM from Rattus norvegicus and PAL from Exiguobacterium sp. by recombinant expression and tested their activities. Since both PHM and PAL interact mainly with the last two residues of peptide substrates, we investigated the substrate scope of this oxidative system at the penultimate position. In the presence of PHM and PAL, almost all 20 model peptides DLSYG-OH (1 mM) were quantitatively converted to the corresponding peptide amides in 15 min under identical reaction conditions (4 h for P1 = Cys peptide, Fig. 3c). Generally, these afforded products can be swiftly (< 30 min) and efficiently (> 90% yield) hydrazinolyzed by PAM12B in a one-pot reaction as expected (Fig. 3c).

Having acquired all of activating modules for supplying appropriate substrates for the coupling module, we finally tested the complete reaction route utilizing all catalytic modules with the model peptide Ac-DFSKVG-OH (1). Briefly, this native peptide carboxylic acid was successively converted to Ac-DFSKV-NH₂ (2) by PHM and PAL and to Ac-DFSKV-N₂H₃ (3) by PAM12B. Then, excessive HNO₂ was utilized to remove residual hydrazine and oxidize 3 at -15 °C. Upon the addition of HBA, the corresponding peptide ester (4) was obtained and subsequently conjugated with equivalent ALKKA-NH₂ by Omniligase-1 to produce Ac-DFSKVALKKA-NH₂ (5, Fig. 3d). The whole process was conducted in one pot in 3 h with only trace amounts of enzymes, exhibiting excellent catalytic efficiency, chemoselectivity, and regioselectivity in the presence of a multitude of side-chain reactive functionalities. These results demonstrated that all catalytic modules exhibited broad substrate spectrum and functioned well in series, and the PALME platform was ready for further investigation.

**Broad application scope of the PALME platform for protein synthesis and functionalization**

We next examined the PALME's utility for practical applications. Considering the synthetic availability and cost, we were poised to prepare long peptide/protein hydrazides (> 10 residues) using peptide activating enzymes and synthesize short peptide hydrazides directly via SPPS. Initially, we tested the feasibility of enzymatic peptide N-to-C sequential condensation and employed this strategy to synthesize exenatide, the API in the antidiabetes drugs Byetta® and Bydureon®. We divided exenatide into three segments, which were prepared in the form of peptide hydrazide (N-part) or peptide amide (middle part and C-part). The N-terminal segment was transformed into its HBA ester and ligated with the middle segment, giving the conjugation product with a 48% isolated yield. The obtained peptide hydrazide was then activated and conjugated with the C-terminal segment to produce 3.0 mg of purified exenatide with a 63% isolated yield. (Fig. 4a). Compared to multifragment condensation in the C-to-N direction, our strategy avoids the
multiple intractable metal-mediated deprotection processes at N-termini, providing a more efficient protocol for chemoenzymatic total synthesis of biomacromolecules.

In addition to intermolecular conjugation, we investigated intramolecular ligation that could generate more rigid cyclic peptides than their linear substrates\(^3\). We selected sheep myeloid antimicrobial peptide (SMAP)\(^3\), which does not contain Cys/Ser/Asp/Asn residues for sequence-limited chemical ligation or transpeptidation in aqueous solution, as the tested object. After esterification and ligation processes, SMAP was cyclized with 86% efficiency according to HPLC analysis (Fig. 4b), illustrating that our activation and ligation strategy was a well supplement to the current peptide cyclization methodologies.

Encouraged by the successes of peptide sequential condensation and cyclization, we asked whether this strategy could be applied on the object prepared by recombinant expression, hence Cys-free 4-oxalocrotonate tautomerase (4-OT) was selected as the target for semisynthesis. 4-OT is a fascinating enzyme that promiscuously catalyzes various important synthetic reactions, including Michael addition\(^3\), aldol condensation\(^3\), and epoxidation\(^3\). Since 4-OT is considered to lie at the interface between organocatalysis and biocatalysis, this protein scaffold serves as an excellent template for chemical engineering to further broaden the synthetic versatility of biomacromolecules. We prepared the C-terminal part of 4-OT by recombinant expression followed by removal of the His-SUMO-tag and then attempted to conjugate it with the SPPS-synthesized N-terminal fragment. The full-length protein was obtained after HPLC purification, and the Michaelase activity of the refolded semisynthesized enzyme resembled that of recombinant 4-OT (Fig. 4c). Subsequently, we performed N-terminal functionalization of much bulkier recombinant proteins. In 10 molar equivalents of biotin/FITC-modified peptide hydrazides, the 10 kDa ubiquitin-like modifier FAT10\(^3\) and the 12 kDa rationally designed HIV-1 immunogen C4S3\(^3\) could be labeled with an efficiency of up to 92% (Fig. 4c). The modification process also successfully worked on the 248-mer enhanced green fluorescent protein (EGFP), which implied that our strategy is highly promising for the functioning of the majority of human proteins (those with a mass of up to ~30 kDa).

Next, we tested whether the proteins bearing post-translational modification in the C-terminal region are accessible by the PALME platform. As one of the most widely investigated regulatory proteins\(^3\), ubiquitin (Ub) has hitherto been a classic target of chemical protein synthesis and semisynthesis, despite the tedious desulfurization process after NCL. We divided Ser65-phosphorylated ubiquitin into two segments, and the majority part of the targeted protein can be obtained via recombinant expression. Accordingly, recombinant Ub(1–59)-Gly was amidated by PHM and PAL, followed by PAM12B-mediated hydrazidation. Afterward, the purified protein hydrazide was esterified and ligated with the synthetic 17-mer phosphorylated peptide, successfully producing the full-length phosphorylated Ub (Fig. 4d). Overall, by harnessing multiple activating and coupling enzymes that present both strict regioselectivity and broad substrate specificity, we demonstrated that the designed PALME platform should be able to cover the expected full spectrum of applications.

**Semisynthesis of intractable proteins through the PALME platform**
Having verified the PALME platform’s broad application scope, we next attempted to apply it to the currently intractable targets. Recombinant proteins bearing multiple adjacent Cys residues are tricky to handle because cysteine-based chemical methods require pretreatment of the native proteins to reduce disulfide bonds\(^4\). When applying thiol-dependent chemical methods to synthesize these targets, native Cys residues were often mutated or protected to avoid side reactions. With the thiol-free and native Cys-independent activation and ligation strategy in hand, we were poised to activate and functionalize NrdH-redoxin, an electron donor bearing a CXXC catalytic motif at the active site that forms a disulfide bond in the oxidation state. This protein is a promising drug target, since it functions cooperatively with prokaryote-specific class Ib ribonucleotide reductase and is essential for cell metabolism\(^4\). By utilizing PHM, PAL, PAM12B, and Omniligase-1 together, we converted the recombinant NrdH-Gly to the corresponding protein hydrazide and labeled it with biotin/FITC at the C-terminus (Fig. 5a). The vital disulfide bond was not disturbed throughout the process, suggesting that the PALME platform could be a suitable supplement for handling intractable multiple-Cys proteins without protection/deprotection processes.

Finally, we examined the PALME’s utility in protein semisynthesis applied to internal regions, which is one of the most profoundly demanded methodologies in protein synthesis\(^3\). We chose Lys56-acetylated human mitochondrial heat shock protein 10 (mHSP10), which participates in cellular protein folding by composing a chaperonin symmetrical football complex with mitochondrial heat shock protein 60 (mHSP60)\(^4\), as the demonstrating target. The location of the desired modification site within the protein sequence most often determines whether semisynthesis is viable. Since the modified Lys56 is located in the internal region of mHSP10, a multistep ligation strategy involving the assembly of three segments was expected to be adopted. Unfortunately, there is no Cys or even Ala residue for conventional NCL protocols between Lys56 and Asp102 at C-terminus. Owing to the PALME platform’s unprecedentedly broad substrate spectrum in terms of both the sequence and the C-terminal functionality, we could design a synthetic scheme that requires chemical synthesis of only one 16-mer peptide amide. First, we converted the synthetic peptide amide to the almost equivalent amount of the corresponding peptide hydrazide. Next, the protein hydrazide of N-part was produced from the recombinant protein glycine smoothly. Afterward, two rounds of esterification, ligation, and HPLC purification were performed following a general protocol for sequential fragment condensation, giving out the full-length acetylated mHSP10 (Fig. 5b). Overall, the platform’s modular nature could provide researchers with flexible selections of input substrates and output functions and their combinations, which would create plentiful retrosynthetic disconnections for disassembling hard-to-access proteins.

**Conclusion**

In summary, we have designed and built a robust, modulated, efficient, and operationally simple multienzyme platform (PALME) for traceless total/semi protein synthesis and functionalization. While the versatility of each module was evaluated via hundreds of model reactions, we also demonstrated the utility of PALME by synthesizing a series of real-case targets. This platform offers a comprehensive range
of solutions for chemical protein synthesis because of its unique features. First, the peptide-activating modules endow the platform with an unprecedentedly broad substrate spectrum in terms of both the sequence and the C-terminal functionality. While experienced protein chemists can directly synthesize the desired peptide hydrazides, biological researchers have flexible choices of using more easily accessible peptide amides (provided by peptide synthesis companies) or native peptides/proteins (obtained from in-house recombinant expression) as assembly materials. Second, by introducing new activating reagents in the coupling module, previously challenging applications of Peptiligase, such as multifragments sequential ligation and recombinant protein C-terminal ligation/functionalization, are readily feasible. Thus the full potential of Peptiligase technology is assuredly realized. Third, the catalysts used in this system are proteins whose performances can be further improved by directed evolution or computational engineering. The stabilization of PHM and PAL in high concentrations of guanidine hydrochloride, enhancement of the hydrazine affinity of PAM, and further broadening of the substrate scope and synthetic efficiency of the Peptiligase family would be particularly interesting. Finally and promisingly, the platform's modular nature is likely to mesh well with flow chemistry, which offers more efficient multistep synthesis than traditional batch methods. Taken together, the PALME platform is highly complementary to other modern techniques of chemical protein synthesis and provides viable solutions to challenges that previous strategies could not address. We anticipate that this report will serve as a blueprint for the future development of a simple and widely applicable protocol to access synthetic proteins, and facilitate artificial biomacromolecule design and applications.

Declarations

Conflicts of interest

Marcel Schmidt and Timo Nuijens are employed by Fresenius Kabi. Omniligase-1 is a proprietary product of Fresenius Kabi.

Author contributions

B.W. and T.N. conceptualized the research. J.F. expressed and purified PHM. T.Z. expressed and purified PAL. Y.T. expressed and purified PAM12B. M.S. expressed and purified Omniligase-1. R.L., M.S., T.Z., and Y.C. optimized synthesis conditions, tested substrate spectrum and built the reaction scheme of the PALME platform. X.Y. expressed, and purified recombinant proteins for semisynthesis. R.L. and T.Z. performed total/semi synthesis of proteins, analyzed and performed biological evaluation of the synthetic proteins. B.W. and T.Z. wrote the manuscript with input of all coauthors.

Acknowledgement

This work is supported by National Key R&D Program of China (No. 2019YFA0905100) the National Natural Science Foundation of China (Grant Nos. 31822002, 31870055), and the Biological Resources Programme (KFJ-BRP-017-58) and Key Research Program of Frontier Sciences (ZDBS-LY-SM014) from the Chinese Academy of Sciences.
We declare the following competing interests: Marcel Schmidt and Timo Nuijens are employed by Fresenius Kabi. Omniligase-1 is a proprietary product of Fresenius Kabi.

References

1. Kent, S. B. H. Total chemical synthesis of proteins. *Chem. Soc. Rev.* **38**, 338–351 (2009).
2. Boutureira, O. & Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **115**, 2174–2195 (2015).
3. Thompson, R. B. & Muir, T. W. Chemoenzymatic semisynthesis of proteins. *Chem. Rev.* **120**, 3051–3126 (2020).
4. Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
5. Hartrampf, N., Saebi, A., Poskus, M., Gates, Z. P., Callahan, A. J., Cowfer, A. E., Hanna, S., Antilla, S., Schissel, C. K., Quartararo, A. J., Ye, X., Mijalis, A. J., Simon, M. D., Loas, A., Liu, S., Jessen, C., Nielsen, T. Et. & Pentelute, B. L. Synthesis of protein by automated flow chemistry. *Science* **368**, 980–987 (2020).
6. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
7. Bode, J. W., Fox, R. M., & Baucom, K. D. Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamines and α-ketoacids. *Angew. Chem. Int. Ed.* **45**, 1248–1252 (2006).
8. Zhang, Y., Xu, C., Lam, H. Y., Lee, C. L. & Li, X. Protein chemical synthesis by serine and threonine ligation. *Proc. Natl. Acad. Sci.* **110**, 6657–6662 (2013).
9. Mitchell, N. J., Malins L. R., Liu X., Thompson R. E., Chan B., Radom L. & Payne R. J. Rapid additive-free selenocystine–selenoester peptide ligation. *J. Am. Chem. Soc.* **137**, 14011–14014 (2015).
10. Shah N. H. & Muir T. W. Inteins: nature's gift to protein chemists. *Chem. Sci.* **5**, 446 (2014).
11. Thompson R. E., Stevens A. J. & Muir T. W. Protein engineering through tandem transamidation. *Nat. Chem.* **11**, 737–743 (2019).
12. Pishesha, N., Ingram, J. R. & Ploegh, H. L. Sortase A: A model for transpeptidation and its biological applications. *Annu. Rev. Cell Dev. Biol.* **34**, 163–188 (2018).
13. 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.* **10**, 732–738 (2014).
14. Agouridas, V., El Mahdi, O. E., Diemer, V., Cargobët, M., Monbaliu, J. C. M. & Melnyk, O. Native chemical ligation and extended methods: Mechanisms, catalysis, scope, and limitations. *Chem. Rev.* **119**, 7328–7443 (2019).
15. Henninot, A., Collins, J. C. & Nuss, J. M. The current state of peptide drug discovery: Back to the future? *J. Med. Chem.* **61**, 1382–1414 (2018).
16. Weeks, A. M. & Wells, J. A., Subtiligase-catalyzed peptide ligation. *Chem. Rev.* **120**, 3127–3160 (2020).

17. 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.* **358**, 2140–2147 (2016).

18. 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.* **358**, 4041–4048 (2016).

19. Pawlas, J., Nuijens, T., Persson, J., Svensson, T., Schmidt, M., Toplak, A., Nilsson, M. & Rasmussen, J. H. Sustainable, cost-efficient manufacturing of therapeutic peptides using chemo-enzymatic peptide synthesis (CEPS). *Green Chem.* **21**, 6451–6467 (2019).

20. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci.* **95**, 6705–6710 (1998).

21. Cao, Y., Nguyen, G. K. T., Tam, J. P. & Liu, C. F. Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation. *Chem. Commun.* **51**, 17289–17292 (2015).

22. Li, Y., Li, Y., Pan, M., Kong, X., Huang, Y., Hong, Z. & Liu, L. Irreversible site-specific hydrazinolysis of proteins by use of sortase. *Angew. Chem. Int. Ed.* **53**, 2198–2202 (2014).

23. Qiao, Y., Yu, G., Kratch, K. C., Wang, X. A., Wang, W. W., Leeuwon, S. Z., Xu, S., Morse, J. S. & Liu W. R. Expressed protein ligation without intein. *J. Am. Chem. Soc.* **142**, 7047–7054 (2020).

24. Arif, M. I., Toplak, A., Szymanski, W., Feringa, B. L., Nuijens, T., Quaedflieg, P. J. L. M., Wu, B. & Janssen, D. B. One-step C-terminal deprotection and activation of peptides with peptide amidase from *Stenotrophomonas maltophilia* in neat organic solvent. *Adv. Synth. Catal.* **356**, 2197–2202 (2014).

25. Wu, B., Wijma, H. J., Song, L., Rozeboom, H. J., Poloni, C., Tian, Y., Arif, M. I., Nuijens, T., Quaedflieg, P. J. L. M., Szymanski, W., Feringa, B. L. & Janssen, D. B. Versatile peptide C-terminal functionalization via a computationally engineered peptide amidase. *ACS Catal.* **6**, 5405–5414 (2016).

26. Zhu, T., Song, L., Li, R & Wu, B. Enzymatic clickable functionalization of peptides via computationally engineered peptide amidase. *Chin. Chem. Lett.* **29**, 1116–1118 (2018).

27. Fang, G., Li, Y., Shen, F., Huang, Y. C., Li, J., Lin, Y., Cui, H. K. & Liu, L. Protein chemical synthesis by ligation of peptide hydrazides. *Angew. Chem. Int. Ed.* **50**, 7645–7649 (2011).

28. 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.* **359**, 2050–2055 (2017).

29. Prigge, S. T., Mains, R. E., Eipper, B. A. & Amzel, L. M. New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. *Cell. Mol. Life. Sci.* **57**, 1236–1259 (2000).

30. Heine, R. J., van Gaal, L. F., Johns, D., Mihm, M. J., Widel, M. H. & Brodows, R. G. Exenatide versus insulin glargine in patients with suboptimally controlled type 2 diabetes: A randomized trial. *Ann. Intern. Med.* **143**, 559–569 (2005).
31. 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* **266**, 243–247 (1994).

32. Chow, H. Y., Zhang Y., Matheson E. & Li X. Ligation technologies for the synthesis of cyclic peptides. *Chem. Rev.* **119**, 9971–10001 (2019).

33. Giacometti, A., Cirioni, O., Ghiselli, R., Mocchegiani, F., D’Amato, G., Circo, R., Orlando, F., Skerlavaj, B., Silvestri, C., Saba, V., Zanetti, M. & Scalise, G. Cathelicidin peptide sheep myeloid antimicrobial peptide-29 prevents endotoxin-induced mortality in rat models of septic shock. *Am. J. Respir. Crit. Care. Med.* **171**, 275–281 (2005).

34. Zandvoort, E., Geertsema, E. M., Baas, B. J., Quax, W. J. & Poelarends, G. J. Bridging between organocatalysis and biocatalysis: asymmetric addition of acetaldehyde to β-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angew. Chem. Int. Ed.* **51**, 1240–1243 (2012).

35. van der Meer, J., Poddar, H., Baas, B. J., Miao, Y., Rahimi, M., Kunzendorf, A., van Merkerk, R, Tepper, P. G., Geertsema, E. M., Thunnissen, A. M. W. H., Quax, W. J. & Poelarends, G. J. Using mutability landscapes of a promiscuous tautomerase to guide the engineering of enantioselective Michaelases. *Nat. Commun.* **7**, 10911 (2016).

36. Xu, G., Crotti, M., Saravanan, T., Kataja, K. M. & Poelarends, G. J. Enantiocomplementary epoxidation reactions catalyzed by an engineered cofactor-independent non-natural peroxygenase. *Angew. Chem. Int. Ed.* **132**, 10460–10464 (2020).

37. Raais, S., Schmidtke, G. & Groettrup, M. The ubiquitin-like protein FAT10 forms covalent conjugates and induces apoptosis. *J. Biol. Chem.* **276**, 35334–35343 (2001).

38. Zhu, C., Dukhovlinova, E., Council, O., Ping, L., Faison, E. M., Prabhu, S. S., Potter, L, Upton, S. L., Yin G., Fay, J. M., Kincer, L. P. Speilvogel, E., Campbell, S. L., Benhabbour., S. R., Ke, H., Swanstrom, R. & Dokholyan, N. V. Rationally designed carbohydrate-occluded epitopes elicit HIV-1 Env-specific antibodies. *Nat. Commun.* **10**, 948 (2019).

39. Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).

40. Vamisetti G. B., Satish G., Sulkshane P., Mann G., Glickman M. H. & Brik A. On-demand detachment of succinimides on cysteine to facilitate (semi)synthesis of challenging proteins. *J. Am. Chem. Soc.* **142**, 19558–19569 (2020).

41. van Laer, K., Dzwewulska, A. M., Fislage, M., Wahni, K., Hbeddou, A., Collet, J. F., Versees, W., Mateos, L. M., Dufe, V. T. & Messens, J. NrdH-redoxin of *Mycobacterium tuberculosis* and *Corynebacterium glutamicum* dimerizes at high protein concentration and exclusively receives electrons from thioredoxin reductase. *J. Biol. Chem.* **288**, 7942–7955 (2013).

42. Nisemblat S., Yaniv O., Parnas A., Frolow F. & Azem A. Crystal structure of the human mitochondrial chaperonin symmetrical football complex. *Proc. Natl. Acad. Sci.* **112**, 6044–6049 (2015).
Figure 1

Overview of the PALME platform. (a) Reported chemical and enzymatic methods for site-restricted peptide ligation. Residues left at the ligation junction (scar) are colored grey. (b) Site-independent peptide ligation via an engineered subtilisin-derived ligase termed Peptiligase. (c) The challenges that Peptiligase-catalyzed peptide ligation faced to date. (d) Our solution reported in this manuscript. (e) Flexible multi-input and multi-output options of the PALME platform.
Figure 2

Screening of the alcohols for Peptiligase-compatible esterification. Overall yield was calculated by integration of the peak areas of ligation, hydrolysis and esterification products monitored by HPLC (220 nm). Gn·HCl: guanidine hydrochloride. More details can be found in the supplementary methods and results chapter 4.1.

| Aliphatic alcohols | 2-Hydroxyacetamide analogues | Fluoroalcohols | Aromatic alcohols | Phenol derivatives |
|-------------------|-----------------------------|----------------|------------------|------------------|
|                   | Nucleophile | Overall yield (%) | Nucleophile | Overall yield (%) | Nucleophile | Overall yield (%) | Nucleophile | Overall yield (%) | Nucleophile | Overall yield (%) |
| a1                | CH₃         | 9               | b1            | 20               | c1            | 38              | d1            | 0               | e1            | 3               |
| a2                | CH₃CH₂      | 18              | b2            | 0                | c2            | 49              | d2            | 16              | e2            | 22              |
| a3                | CH₃         | 5               | b3            | 15               |              |                 | d3            | 11              | e3            | 29              |
| a4                | CH₃         | 8               | b4            | 13               |              |                 | d4            | 5               | e4            | 8               |
|                   |             |                 |               |                  |              |                 | d5            | 34              | e5            | 0               |
|                   |             |                 |               |                  |              |                 |               |                  | e6            | 86              |
|                   |             |                 |               |                  |              |                 |               |                  | e7            | 86              |
|                   |             |                 |               |                  |              |                 |               |                  | e8            | 86              |
|                   |             |                 |               |                  |              |                 |               |                  | e9            | 55              |
|                   |             |                 |               |                  |              |                 |               |                  | e10           | 37              |
Figure 3

The scope and compatibility of activating and coupling modules (a) Six site-saturation peptide hydrazide and peptide amide libraries (Ac-DXSKL-N2H3, Ac-DFXKL-N2H3, Ac-DFSXL-N2H3, Ac-DFSXK-N2H3, XLKKA-NH2 and AXKKA-NH2) were utilized to investigate Omniligase-1 substrate profiles. Ligation efficiency (blue band) was calculated by integration of the peak areas of ligation and hydrolysis products monitored by HPLC (220 nm). (b) Two site-saturation peptide amide libraries (Ac-DFSXL-NH2 and Ac-...
DFSKX-NH2) were utilized to investigate PAM12B substrate profiles. The hydrazidation yield (blue band) was calculated by integration of the peak areas of hydrazidation and hydrolysis products monitored by HPLC (220 nm) except for P1 = Pro or Asn. (c) Amidation and hydrazidation reactions were performed on one site-saturation peptide glycine library (DLSYXG-OH). The amidation yield (blue band) was calculated by integration of the peak areas of amidation products and residual substrates monitored by HPLC (220 nm). The hydrazidation yield was determined as described above. (d) One-pot activation and ligation of Ac-DFSKV-G and ALKKA-NH2. Analytical HPLC traces (220 nm) of the reaction mixture (up to down): substrate, 90 min after the addition of PHM and PAL, 10 min after the addition of PAM12B, before the addition of Omniligase-1, 30 min after the addition of Omniligase-1. More details can be found in the supplementary methods and results chapter 4.2 (Omniligase), 4.3 to 4.5 (PAM12B, PHM and PAL) and 4.6 (One-pot conjugation of Ac-DFSKV-G and ALKKA-NH2).
Figure 4

Broad applications of the PALME platform. (a) Total synthesis of exenatide (PDB: 1JRJ) by N-to-C sequential peptide condensation. (b) Cyclization of SMAP. (c) (left) Semisynthesis of 4-OT monomer. The cartoon demonstrates the structure of catalytically active 4-OT hexamer (PDB: 4X19, A33D mutant was introduced to raise Michael-type addition activity). The activity of the semisynthesized and recombinant 4-OT was determined according to the consumption of (E)-2-nitroethenyl benzene (7), which was
monitored by HPLC (320 nm). (right) N-terminal modification of the ubiquitin-like modifier FAT10 (PDB: 6GF2), the rationally designed immunogen C4S3 (PDB: 6CBU) and EGFP (PDB: 2Y0G) with biotin/FITC-modified FSKL-N2H3. (d) Semisynthesis of Ser65-phosphorylated ubiquitin. The protein structures were constructed with PyMOL (Version 1.7 Schrödinger, LLC.). More details can be found in the supplementary methods and results chapter 4.7 (exenatide), 4.8 (SMAP), 4.9 (4-OT), 4.10 (FAT10, C4S3 and EGFP) and 4.11 (Ub).

**Figure 5**

Semisynthesis of intractable proteins. (a) C-terminal modification of NrdH-redoxin. The ESI-MS results showed that the observed molecular weights of substrate, intermediate and final products were in accord with the theoretical weight of the corresponding biomacromolecules in the oxidation state. (b) Semisynthesis of Lys56-acetylated mHSP10. Additional details can be found in the supplementary methods and results chapter 4.12 (NrdH-redoxin) and 4.13 (mHSP10).

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymethodsandresults.docx
- SupplementaryHPLCtracesandMSfigures.docx