CHEMICAL BIOLOGY TOOLS FOR PEPTIDE AND PROTEIN RESEARCH

EDITED BY: Yu-Hsuan Tsai, Hideo Iwaï, Klaus Pors and Tao Peng
PUBLISHED IN: Frontiers in Chemistry

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-668-2
DOI 10.3389/978-2-88974-668-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

CHEMICAL BIOLOGY TOOLS FOR PEPTIDE AND PROTEIN RESEARCH

Topic Editors:
Yu-Hsuan Tsai, Shenzhen Bay Laboratory, China
Hideo Iwaï, University of Helsinki, Finland
Klaus Pors, University of Bradford, United Kingdom
Tao Peng, Peking University, China

Citation: Tsai, Y.-H., Iwaï, H., Pors, K., Peng, T., eds. (2022). Chemical Biology Tools for Peptide and Protein Research. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-668-2

Table of Contents

frontiers in Chemistry

EDITORIAL
published: 18 February 2022
doi: 10.3389/fchem.2022.861699

Editorial: Chemical Biology Tools for Peptide and Protein Research

Yu-Hsuan Tsai[1]*, *Hideo Iwaï*[2]* and *Klaus Pors*[3]*

[1]Institute of Molecular Physiology, Shenzhen Bay Laboratory, Shenzhen, China, [2]Institute of Biotechnology, University of Helsinki, Helsinki, Finland, [3]Institute of Cancer Therapeutics, University of Bradford, Bradford, United Kingdom

Keywords: solid-phase peptide synthesis, non-canonical amino acid, bioorthogonal chemistry, genetic code expansion, site-specific modification, protein engineering, chemical biology

Editorial on the Research Topic

Chemical Biology Tools for Peptide and Protein Research

INTRODUCTION

Peptides and proteins are important biological molecules for all living systems. The human body uses different peptides as hormones for signal transduction, whereas proteins are indispensable for cellular structure and function. Indeed, peptides and proteins are closely associated with nearly all human diseases. Not surprisingly, many peptides and proteins have been utilized for disease prevention or treatment. However, of the more than 10,000 different proteins that a human cell can produce, many are not yet fully understood (Wilhelm et al., 2014). Even for peptides and proteins that have been studied, developing them as research tools or clinical agents still faces different obstacles.

The need to address the limitations in peptide and protein research has led to the development of chemical biology tools in several aspects. Solid-phase peptide synthesis (Jaradat, 2018) is particularly useful for the production of peptides with less than 50 amino acid residues. The combination of solid-phase peptide synthesis and chemical ligation methods provides access to longer peptides and proteins without size limitation. Key advantages of these chemical approaches are the capability to include non-canonical amino acids containing diverse functional groups, as well as the production of homogeneous materials, whereas peptides and proteins isolated from cells can exist in a mixture of isoforms with different side-chain or backbone modifications.

Nevertheless, genetic means enable production of the desired proteins in large scale with ease as well as investigation of the target proteins in their native cellular environments. In addition, genetic means allow engineering of proteins with different properties, such as allosteric site for functional regulation or recognition of new substrates. Particularly, advances in analytical techniques facilitate rapid identification of protein mutants with the desired properties, supporting the development of protein engineering.

The ability to site-specifically modify proteins has also opened many opportunities for biological and medical applications. This can be achieved through bioorthogonal reactions with genetically incorporated non-canonical amino acids (de la Torre and Chin, 2021) or enzyme-mediated ligations (Xu et al., 2018). Fluorescent labeling of the target protein allows determination of its subcellular localization. Conjugation of recombinant proteins with targeting motifs may afford a target-specific delivery system. Enzyme-mediated ligation is especially useful for making backbone cyclized proteins, which often have superior stability to their linear counterparts (Hayes et al., 2021).

OPEN ACCESS

Edited and reviewed by:
Olga Avrutina,
Darmstadt University of Technology, Germany

***Correspondence:**
Yu-Hsuan Tsai
Tsai.Y-H@outlook.com
Hideo Iwaï
hideo.iwai@helsinki.fi
Klaus Pors
k.pors1@bradford.ac.uk

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *25 January 2022*
Accepted: *27 January 2022*
Published: *18 February 2022*

Citation:
Tsai Y-H Iwaï H and Pors K (2022) Editorial: Chemical Biology Tools for Peptide and Protein Research. Front. Chem. 10:861699. doi: 10.3389/fchem.2022.861699

Below, we discuss the nine original articles and one review within this series in three categories: chemical synthesis of peptides and proteins, genetic means for protein engineering, and applications of site-specific protein modification.

CHEMICAL SYNTHESIS OF PEPTIDES AND PROTEINS

Chemical synthesis of peptides and proteins is not limited to the reservoir of canonical amino acids and enables access to their derivatives, thus facilitating elucidation of the structure-activity relationship. Towards the synthesis of human selenoprotein F (SelF), which is believed to play important roles in the quality control of the endoplasmic reticulum, Liao and He optimized the synthesis of a SelF homolog through a three-segment two-ligation strategy. Improvement in the synthetic strategy not only enabled the production of multi-milligrams of the folded synthetic protein but also set the stage for the synthesis of the native selenoprotein, of which the exact physiological functions remain largely elusive. An additional outcome of the study was the elucidation of the disulfide pairing mode of SelF. The data provided useful structural insights into the previously unresolved UGGT-binding domain of SelF.

On the other hand, Wu and Chu presented a neat synthetic route to a macrocyclic depsipeptide, pagoamide A, and evaluated its antimicrobial activity. Pagoamide A is a marine natural product, and the authors achieved its total synthesis by incorporating two thiazole building blocks in the solid-phase peptide synthesis. While the antimicrobial activity of pagoamide A was moderate against *Bacillus subtilis*, the synthetic approach was modular and is theoretically applicable to synthesizing molecules with common structural features found in peptide natural products, among them heterocycles, N-methylations, and backbone macrocyclization.

Lastly, Miret-Casals et al. employed solid-phase peptide synthesis to study furan-mediated chemical crosslinking. Although furan has been used as a warhead for chemical crosslinking, its crosslinking partners had not been experimentally characterized in detail. The authors synthesized a series of α-helical peptides, which can form coiled-coil peptide dimers. They then used these peptides to explore furan reactivity and its site-specific crosslinking. Activated furan warheads were shown to react with lysine, cysteine, and tyrosine. This *in vitro* validation demonstrates the versatility of the furan crosslinkers and provides further ground for exploiting furan-technology in developing furan-modified ligands/proteins for covalent tethering to target proteins through various amino acid side chains.

GENETIC MEANS FOR PROTEIN ENGINEERING

Mutagenesis is a common genetic approach to engineer proteins with novel properties. Using mutagenesis, Senoo et al. engineered metabotropic glutamate receptor 1 (mGlu1) with a novel allosteric metal-binding site. Specifically, they identified mGlu1 mutants, of which activity can be artificially induced by a protein conformational change in response to a palladium complex, Cu^{2+} or Zn^{2+}. Notably, the activation of the mutants was mutually orthogonal, resulting in cell-type selective activation demonstrated in differentially transfected HEK293 cells. This direct activation *via* a coordination-based chemogenetic approach allows cell-specific activation of the target receptor, providing a useful way for investigating the physiological function of a protein in different cells.

Mutagenesis also forms the basis of directed evolution, where a protein can be transformed to equip a completely different property, such as novel enzymes that catalyze reactions invented by synthetic chemists (Arnold, 2019). A key challenge in directed evolution is to identify the desired mutants from a pool of millions or more variants. Fu et al. reviewed recent advances on sorting methods of high-throughput droplet-based microfluidics in enzyme-directed evolution. They highlighted successful cases, as well as discussed the advantages and challenges of different sorting methods.

APPLICATIONS OF SITE-SPECIFIC PROTEIN MODIFICATION

Site-specific modification is a powerful strategy for studying the structure and function of peptides and proteins (Tamura and Hamachi, 2019). To study proteins within their natural environment, various bioorthogonal reactions have been developed. Since canonical amino acids lack bioorthogonal functionalities, a bioorthogonal group must be introduced into the target protein to enable site-specific modifications. This can be readily achieved through genetic code expansion, where a non-canonical amino acid can be co-translationally introduced into the desired position of the target protein (de la Torre and Chin, 2021).

In this Research Topic, Meineke et al. expanded the scope of genetic code expansion to incorporate a copper-chelating amino acid, picolyl azide lysine (PazK). They demonstrated efficient incorporation of PazK into proteins in mammalian cells, enabling improvement in copper-catalyzed azide–alkyne cycloaddition and live-cell fluorescent labeling. Results of this work can be combined with their previous finding for simultaneous dual or multiple bioorthogonal labeling on live mammalian cells.

In addition to copper-catalyzed azide–alkyne cycloaddition, azides can also react with strained alkenes and alkynes bioorthogonally. Using genetic code expansion, Johnson et al. generated two monomeric proteins containing either a *p*-azidophenylalanine (AzF) or a cyclooctyne lysine residue and demonstrated the formation of heterodimers by the two proteins. On the other hand, Wang et al. identified a variant of *Methanosarcina mazei* pyrrolysyl-tRNA synthetase for efficient incorporation of AzF. This synthetase variant enabled a generation of a model protein containing up to ten AzF residues in *Escherichia coli*. Using this pyrrolysyl-tRNA synthetase variant, they produced AzF-containing protein nanocages composed of a HER2 receptor recognition peptide

and human heavy chain ferritin. Conjugation of the protein nanocages with doxorubicin by strain-promoted azide-alkyne cycloaddition afforded a cell-specific delivery system, which was demonstrated using $Her2^+$ breast cancer cells and showed prolonged drug release.

Besides bioorthogonal chemistry, enzyme-mediated ligations are also useful for site-specific protein modification. Among different enzymes, asparaginyl endopeptidases (AEPs) have attracted much attention in recent years due to their superior kinetics. However, a major drawback in most enzyme-catalyzed reactions is the need of large excess of a reacting partner to drive the reaction toward completion. This is due to the reversibility and/or hydrolytic nature of the enzymes. Based on their prior knowledge of AEPs, Chen et al. engineered an AEP from *Viola canadensis* so that the enzyme variant mainly functions as a ligase for amide bond formation instead of a protease for cleavage. They achieved this by mutating residues in the non-conserved substrate-binding pockets of the enzyme. The authors also demonstrated the use of this AEP variant in generating a backbone cyclized protein as well as fluorescent labeling on live MCF7 cells.

Head-to-tail backbone cyclization of proteins and peptides has been shown to greatly enhance protein and peptide stability. By taking advantage of enzyme-mediated ligation, Hsu et al. took one step further to generate a backbone cyclized knotted protein that cannot be untied. Over a thousand knotted proteins possessing loose peptide ends (i.e., mathematically linear polypeptide chains) have been identified in nature. A mathematical knotted protein without the loose ends has opened the possibility of investigating the effect of the protein backbone topology. It is generally not possible to control the topology of the unfolded states of knotted proteins with loose ends, posing challenges in their characterization. The authors addressed this problem by performing head-to-tail cyclization of a model knotted protein. In this way, they studied the deeply trefoil-knotted YibK from *Pseudomonas aeruginosa* and investigated the effect of topological knotting in conformational entropy and protein folding.

PERSPECTIVES

This collection of articles showcases different chemical biology tools for peptide and protein research, including solid-phase peptide synthesis, protein engineering, genetic code expansion, bioorthogonal chemistry, and enzyme-mediated ligations. Close collaboration and interplay between chemists and biologists will further expand the scope and applications of these tools. Indeed, the outcomes of this interdisciplinary field are facilitating and will certainly accelerate biological and medical research.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We would like to thank all contributing authors for their hard work in producing the enclosed articles. We hope you enjoy this Research Topic featured as a part of the Chemical Biology section of Frontiers in Chemistry.

REFERENCES

Arnold, F. H. (2019). Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chem. Int. Ed.* 58, 14420–14426. doi:10.1002/anie.201907729

de la Torre, D., and Chin, J. W. (2021). Reprogramming the Genetic Code. *Nat. Rev. Genet.* 22, 169–184. doi:10.1038/s41576-020-00307-7

Hayes, H. C., Luk, L. Y. P., and Tsai, Y.-H. (2021). Approaches for Peptide and Protein Cyclisation. *Org. Biomol. Chem.* 19, 3983–4001. doi:10.1039/d1ob00411e

Jaradat, D. s. M. M. (2018). Thirteen Decades of Peptide Synthesis: Key Developments in Solid Phase Peptide Synthesis and Amide Bond Formation Utilized in Peptide Ligation. *Amino Acids* 50, 39–68. doi:10.1007/s00726-017-2516-0

Tamura, T., and Hamachi, I. (2019). Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems. *J. Am. Chem. Soc.* 141, 2782–2799. doi:10.1021/jacs.8b11747

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., et al. (2014). Mass-spectrometry-based Draft of the Human Proteome. *Nature* 509, 582–587. doi:10.1038/nature13319

Xu, S., Zhao, Z., and Zhao, J. (2018). Recent Advances in Enzyme-Mediated Peptide Ligation. *Chin. Chem. Lett.* 29, 1009–1016. doi:10.1016/j.cclet.2018.05.024

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

MINI REVIEW
published: 23 April 2021
doi: 10.3389/fchem.2021.666867

Recent Advances on Sorting Methods of High-Throughput Droplet-Based Microfluidics in Enzyme Directed Evolution

Xiaozhi Fu[1], *Yueying Zhang*[2,3], *Qiang Xu*[2,3], *Xiaomeng Sun*[2,3] *and Fanda Meng*[2,3,4]*

[1] Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden, [2] Department of Clinical Laboratory Medicine, The First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital, Shandong Medicine and Health Key Laboratory of Laboratory Medicine, Jinan, China, [3] School of Basic Medicine, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, China, [4] Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Gothenburg, Sweden

Droplet-based microfluidics has been widely applied in enzyme directed evolution (DE), in either cell or cell-free system, due to its low cost and high throughput. As the isolation principles are based on the labeled or label-free characteristics in the droplets, sorting method contributes mostly to the efficiency of the whole system. Fluorescence-activated droplet sorting (FADS) is the mostly applied labeled method but faces challenges of target enzyme scope. Label-free sorting methods show potential to greatly broaden the microfluidic application range. Here, we review the developments of droplet sorting methods through a comprehensive literature survey, including labeled detections [FADS and absorbance-activated droplet sorting (AADS)] and label-free detections [electrochemical-based droplet sorting (ECDS), mass-activated droplet sorting (MADS), Raman-activated droplet sorting (RADS), and nuclear magnetic resonance-based droplet sorting (NMR-DS)]. We highlight recent cases in the last 5 years in which novel enzymes or highly efficient variants are generated by microfluidic DE. In addition, the advantages and challenges of different sorting methods are briefly discussed to provide an outlook for future applications in enzyme DE.

Keywords: sorting methods, enzyme directed evolution, microfluidics, droplet, high-throughput

OPEN ACCESS

Edited by:
Yu-Hsuan Tsai,
Cardiff University, United Kingdom

Reviewed by:
Dan Tan,
Xi'an Jiaotong University, China
Zhilian Liu,
University of Jinan, China

***Correspondence:**
Fanda Meng
mengfinder@mail.ipc.ac.cn

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *11 February 2021*
Accepted: *02 March 2021*
Published: *23 April 2021*

Citation:
Fu X, Zhang Y, Xu Q, Sun X and Meng F (2021) Recent Advances on Sorting Methods of High-Throughput Droplet-Based Microfluidics in Enzyme Directed Evolution. Front. Chem. 9:666867. doi: 10.3389/fchem.2021.666867

INTRODUCTION

Nature itself is a great reservoir of various enzymes whose catalysis of substrates makes life and industry possible. Many ancestral enzymes have low catalytic efficiency and low specificity but might go through rounds of mutations and natural selection toward more specific and efficient variants. This process might take millions of years, which is part of natural evolution. Over the recent two decades, scientists are trying to mimic natural selection conditions in the laboratories and accelerate the selection toward desirable properties, which is called directed evolution (DE). The whole process starts from a mutant library of one existing enzyme, or *de novo* synthetic enzyme; and the mutant library could be generated by rational/semi-rational design or random mutagenesis. The variants are expressed *in vivo*/*in vitro* or in cell-free system and then selected for improved properties. An effective assay requires tight linkage of genotype and phenotype, so that promising variants could be subjected to further cycles of optimization (Zeymer and Hilvert, 2018).

High-throughput screening (HTS) is commonly defined as screening no <100,000 samples per day (Attene-Ramos et al., 2014), which equals to 1.16 test per second, i.e., 1.16 Hz. Traditional HTS is performed with microliter plates (MTPs) in 96-, 384-, or 1,536-well formats and agar plates whose throughput is $\sim 10^4$ variants with manual operation or $\sim 10^6$ variants with robots per day (Markel et al., 2020). While libraries of 10^{10} variants could be easily generated by error-prone PCR (You and Percival Zhang, 2012), traditional screening process is still time-consuming and labor-intensive. Automated fluorescence measurement and robotic colony picking lighten the tedious screening workload, but the physical and material constraints associated with spatial separation inherently limit throughput (Packer and Liu, 2015) and always come with significantly increased reagent consumption (Martis et al., 2011). The developments of microfluidics with integrated droplet generation, droplet manipulation, and screening modules make automation possible for the whole enzyme screening process. By inputting a library of enzyme variants, researchers could collect outputs of desired ones from up to 10^8 candidates per day, while consuming 10^6-fold less sample volume (Vallejo et al., 2019).

In a typical microfluidic droplet workflow (**Figure 1**) for enzyme DE, a single cell from the mutant enzyme-expression library is encapsulated in a water-in-oil (w/o) droplet with their substrates. The droplets could be incubated for a specific time for the enzyme to fully react with the substrate. Both on-chip and off-chip incubations are possible. If the enzymes are expressed *in vivo*, cell lysis buffer will also be encapsulated into the droplet in the droplet generation step. Then the droplets could be sorted based on specific detectable signals. According to the signals, dielectrophoresis (DEP) usually drags droplets with active enzymes inside toward a sorting channel by high-voltage electric pulses. The sorting methods determine the selection threshold for DEP at the junction. Many researches have been done in droplet generation step in the last decade, including w/o droplets (Bransky et al., 2009; Tarchichi et al., 2013), w/o-in-water (w/o/w) droplets (Nabavi et al., 2015), and hydrogel beads (Um et al., 2008; Marquis et al., 2015) from picoliter to nanoliter volumes; recently, there emerges rapid development of different sorting methods, along with practical applications in enzyme DE.

Since droplet generation frequency could reach 10∼30 kHz, which means ∼10,000–30,000 droplets per second, the efficiency of droplet sorting limits the efficiency of the whole microfluidics throughput. The sorting method constrains limits for droplet size and substrate concentration and affects the droplet recovery viability. Traditional microfluidic-based droplet sorting mainly relies on laser-induced fluorescence (LIF) detection. However, there have emerged other novel sorting methods for on-chip droplet HTS sorting recently and shows the trend of label-free sorting. In this work, both traditional labeled sorting methods and new label-free approaches in the past 5 years have been reviewed.

LABELED SORTING

Labeled sorting is defined as sorting methods based on characteristics that do not result directly from the enzymatic reaction itself. Chemicals or tagged groups are added into the reaction so that labeled sorting methods could work on them or their derivatives. Fluorescence-activated droplet sorting (FADS) and absorbance-activated droplet sorting (AADS) are two commonly applied labeled sorting methods and also regarded as optical sorting. FADS relies on the fluorophore yield or fluorescent tagging in the droplet. AADS is based on changes in UV or visible light absorption. Its absorption changes are always caused by absorbing reagent yield in the enzymatic reaction or in its coupled assays.

Fluorescence-Activated Droplet Sorting

FADS is based on a very similar principle to fluorescence-activated cell sorting (FACS). FACS has been considered as the gold standard for single cell sorting (Attene-Ramos et al., 2014) and could also be used for droplet sorting. Compared with FACS, FADS is performed on-chip and advantageous in the following aspects. Firstly, only hydrophilic w/o/w droplets could be sorted by FACS, while FADS could detect both w/o or oil in water (o/w) droplets and double emulsions. Secondly, high-speed camera makes it possible to visualize each droplet sorting event, which is not yet possible with FACS. Lastly, FADS devices are much cheaper than the dedicated FACS instrument.

FADS has been used for DE of aldolases (Obexer et al., 2017), DNA polymerases (Vallejo et al., 2019), NAD(P)-dependent oxidoreductases (Oyobiki et al., 2014), xylanase (Ma et al., 2019), lipases (Qiao et al., 2018), and oxidase (Debon et al., 2019) in bacteria (Qiao et al., 2018; Vallejo et al., 2019), yeast (Ma et al., 2019), and even filamentous fungi (Beneyton et al., 2016). It proves to be a powerful tool for enzyme DE.

In most cases, fluorogenic reporter substrate is needed for FADS. Ma et al. (2019) utilized the conversion of fluorogenic substrate 6,8-difluoro-4-methylumbelliferyl b-D-xylobiose (DiFMUX2) to fluorophore DiFMU by xylanase to evolve xylanase-producing *Pichia pastoris* and screened out a 1.3-fold mutant. Fenneteau et al. (2017) synthesized a new sulfonylated rhodamine for more sensitive peptidase activity detection with droplet-based microfluidics, so that absorption and emission ranges of yield product are separated. Larsen et al. (2016) found that Cy3-Iowa Black fluorophore–quencher pair for DNA labeling maintains a higher signal-to-noise ratio than commonly used fluorophore–quencher pairs. Vallejo et al. (2019) further applied 5′-Cy3 DNA labeling in the DE of nucleic acid enzymes, like polymerase, T4 ligase, and restriction enzyme with FADS. Vallejo et al. (2019, 2020) summarized this strategy as droplet-based optical polymerase sorting (DrOPS) (**Figure 2A**). Polymerase variant library is expressed in *Escherichia coli*, and single bacteria cells are encapsulated in microfluidic droplets. The polymerase is released into the droplet microcompartment upon cell lysis and will produce Cy3-based fluorescence signals by disrupting a donor–quencher pair. Fluorescent droplets with active polymerase variants could be sorted by FADS, and the plasmid encoding the variant could be recovered.

Instead of using fluorophore-labeling substrate directly, in other cases, researchers design secondary or tertiary reactions for the generation reporter chemicals of particular excitation/emission patterns. Debon et al. (2019) obtained a cyclohexylamine oxidase (CHAO) variant of 960-fold improvement compared with the wild type. In their case, amines

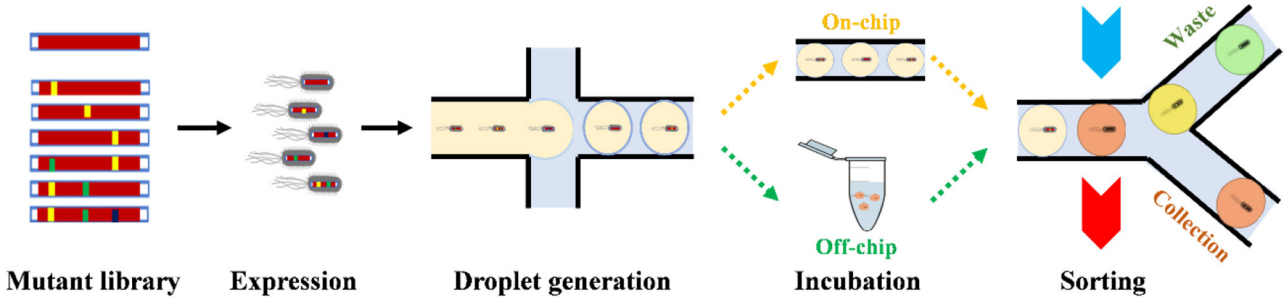


FIGURE 1 | Typical microfluidic droplet workflow for enzyme directed evolution.

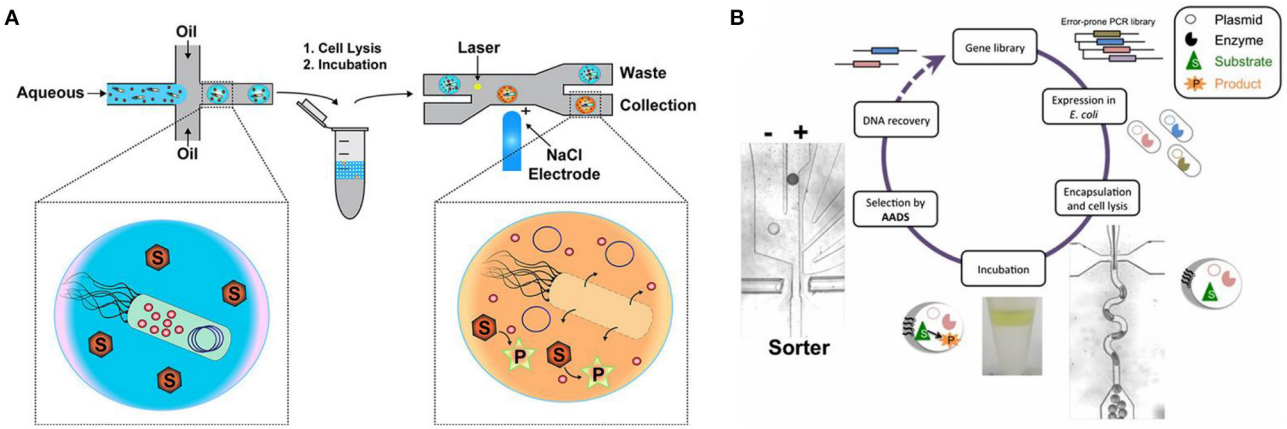


FIGURE 2 | Labeled sorting methods based on fluorescence-activated droplet sorting (FADS) **(A)** [adapted with permission from Vallejo et al. (2019) Copyright © American Chemical Society] and absorbance-activated droplet sorting (AADS) **(B)** [adapted with permission from Gielen et al. (2016) Copyright © National Academy of Sciences].

are converted into imine by CHAO, meanwhile reducing 1 equiv. of flavin adenine dinucleotide (FAD). Oxygen-dependent cofactor recycling produces equimolar amounts of hydrogen peroxide (H_2O_2), which is detected by the downstream oxidation of a fluorogenic dye Amplex by horseradish peroxidase. This coupled-enzyme assay is applicable to other oxidases or relative enzymes, like L-asparaginase (Lim and Gruner, 2020) and glucose oxidase (Prodanović et al., 2020), which produced H_2O_2 and secondary reactions were added so that fluorophore was further produced.

Researchers have developed some upgraded versions of commonly used FADS system. Two fluorescence-activated droplet sorters were set in series along the chip and equipped with two sets of excitation/emission wavelengths (Ma et al., 2018). A double-gated control algorithm could process both fluorescence signals from the same droplets. Two chemical reactions happened simultaneously in one droplet so that esterase mutants from *Archaeoglobus fulgidus* with both high enzymatic activity and high enantioselectivity could be sorted.

Researchers also look into other characteristics of fluorescence and newly developed fluorescence lifetime-activated droplet sorting (FLADS) (Hasan et al., 2019) based on the lifetime of fluorophore. FLADS has been successfully applied to distinguish droplets containing either pyronine or fluorescein, or both, since both chemicals are excited at the wavelength of 470 nm (Hasan et al., 2019) but have different fluorescence lifetime. This might help broaden the scope of fluorescence sorting but presently is limited by the throughput, whose frequency of only up to 50 Hz was achieved till now (Haidas et al., 2019).

Absorbance-Activated Droplet Sorting

AADS helps extend the application of microfluidics, by breaking the exclusive boundary of fluorescent readouts. Gielen et al. (2016) firstly set up an AADS coupled droplet microfluidic for enzyme DE (**Figure 2B**), by embedding two optical fibers aligned face-to-face across the droplet channel. With this device, the activity of phenylalanine dehydrogenase *Rhodococcus* sp. M4 was improved >4.5-fold in lysate and k_{cat} increased >2.7-fold after two rounds of DE. The reduction of the cofactor NAD^+ to NADH in the deamination direction is detected by a coupled assay involving the electron coupling reagent 1-methoxy-5-methylphenazinium methyl sulfate (mPMS) and the reduction of the water-soluble tetrazolium salt 2-(4-iodophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium (WST-1) to

TABLE 1 | Optimal specification of different sorting methods.

Sorting methods	Sensitivity	Highest frequency	Minimal droplet size
FADS	2.5 nM (Colin et al., 2015)	5 kHz (Neun et al., 2019)	2 pl (Colin et al., 2015)
AADS	10 μM (Gielen et al., 2016)	100 Hz (Gielen et al., 2016)	100 pl (Maceiczyk et al., 2017)
ECDS	1 μM (Goto et al., 2020)	10 Hz (Goto et al., 2020)	30 nl (Goto et al., 2020)
MADS	5 μM (Kempa et al., 2020)	35 Hz (Kempa et al., 2020)	0.8 nl (Kempa et al., 2020)
RADS	200 μM (Sobota et al., 2019)	4.3 Hz (Wang et al., 2017)	65 pl (Wang et al., 2017)
NMR-DS	1 mM (Davoodi et al., 2020)	—	130 pl (Hale et al., 2018)

—means not reported.

give the absorbing dye WST-1 formazan. The sorting frequency could achieve ∼1 million droplets per hour. Theoretically, AADS could have much more applications since most small molecules exhibit absorption in the UV and visible regions of electromagnetic spectrum.

AADS has the inert disadvantage of reduced optical pathlength together; its sensitivity is three to four magnitude lower than that of FADS (**Table 1**). To address the sensitivity problem, Maceiczyk et al. (2017) combined differential detection photothermal interferometry (DDPI) with absorbance detection in droplet-based microfluidics. DDPI allows for quantitative, single-point absorbance detection in femtoliter–picoliter volume droplets and is weakly dependent on the optical pathlength. They applied this method to detect 100 pl with as low as 1.4 μM of erythrosine B at 1-kHz frequency and femtoliter droplets at 10-kHz frequency. This method was proved workable for colorimetric assay of HL-60 cells growth and β-galactosidase activity, but it has not been applied in enzyme DE yet.

Instead of improving the sensitivity of AADS itself, Zurek et al. (2021) recently developed a strategy by increasing enzyme molecules in the droplets. They set up a workflow for clonal amplification in droplets and demonstrated that around 400 *E. coli* cells will be in one 100-pl droplet after single-cell cultivation overnight. Through increasing enzyme molecules, the reaction rate of phenylalanine dehydrogenases (PheDH) improved 12-fold as detected in absorbance assay of droplets. The same strategy might also be applied for other less sensitive sorting methods.

LABEL-FREE SORTING

Apart from the two common optical (labeled) sorting approaches, a trend of developing label-free detections emerged, which uses intrinsic physical or chemical biomarkers to separate and sort cells. Several major label-free sorting methods are like electrochemical detection, mass spectrometry (MS), and Raman and nuclear magnetic resonance (NMR) coupling with a microfluidic chip. Different from easy coupling of optical sorting set up with microfluidic devices, the following label-free sorting methods often need a specific design or coupling techniques.

Electrochemical-Based Droplet Sorting

Electrochemical detection is label free and can be applied to complex samples without interference from suspension, autofluorescence, or staining. Due to the limitation of detection electrode surface size, which should not be larger than the droplet, the size of droplets applied is normally limited to nanoliter. The detection device on a microfluidic chip could also be coupled with a DEP sorting device, by converting electrochemical signal into digital signal and then into alternating current (AC) signal to turn on DEP sorter.

Goto et al. (2020) applied a boron-doped diamond (BDD) electrode to measure the current change in 30-nl droplets for the DE of a NAD(P)-dependent oxidoreductase. As low as 1 μM of NADH could be detected, and a 3-fold higher activity of isocitrate dehydrogenase (IDH) mutant from *Streptococcus* mutants was screened (**Figure 3A**). Oyobiki et al. (2014) proved previously that the characteristics of BDD, such as wide potential window, a low background current, and the higher stability against deactivation, are suitable for measuring NADH oxidation.

A Sorting based on Interfacial Tension (SIFT) method was developed even without extra electrodes to differentiate droplets of various interfacial tension due to different pH values in droplets (Abbyad et al., 2019). Low pH leads to high interfacial tension of droplets, and droplets of lower pH tend to flow along a trail on a chip. The method has been developed to distinguish between pH 0.2, and maximum sorting rate is around 30 Hz (Horvath et al., 2019). However, this method has not been applied to any enzyme high-throughput sorting yet.

Mass-Activated Droplet Sorting

As the most widespread and versatile analytical technique to analyze chemical mixtures, MS has been combined with microfluidics in recent 5 years. Both electrospray ionization (ESI) (Mahler et al., 2018; Qiao et al., 2018; Kempa et al., 2020) and matrix-assisted laser desorption ionization (MALDI) (Haidas et al., 2019) have been applied in droplet HTS. ESI could produce charged ions directly from a liquid, which facilitates it to be coupled with microfluidics. However, the online coupling of MALDI with microfluidics is challenging, because MALDI target is under vacuum while the microfluidic droplets are generated under atmospheric pressure. Meanwhile, MALDI cannot be performed as an online technique for analysis, which also restricts its throughput (Payne et al., 2020).

Oil carrier phase usually needs to be removed before directing the aqueous flow into an electrospray emitter, since oil phase interferes with the ESI process by both sequestering charge carriers and preventing stable Taylor cone formation. An orthogonal chip-MS set up (Beulig et al., 2017) was developed by placing a grounded metal plate as a counter-electrode with a specific distance to the spray on the chip, while orthogonal to the MS orifice. The inlet flow to MS orifice could be adjusted by changing the distance from MS orifice to the spray. This setup could avoid sample overloading while maintaining a

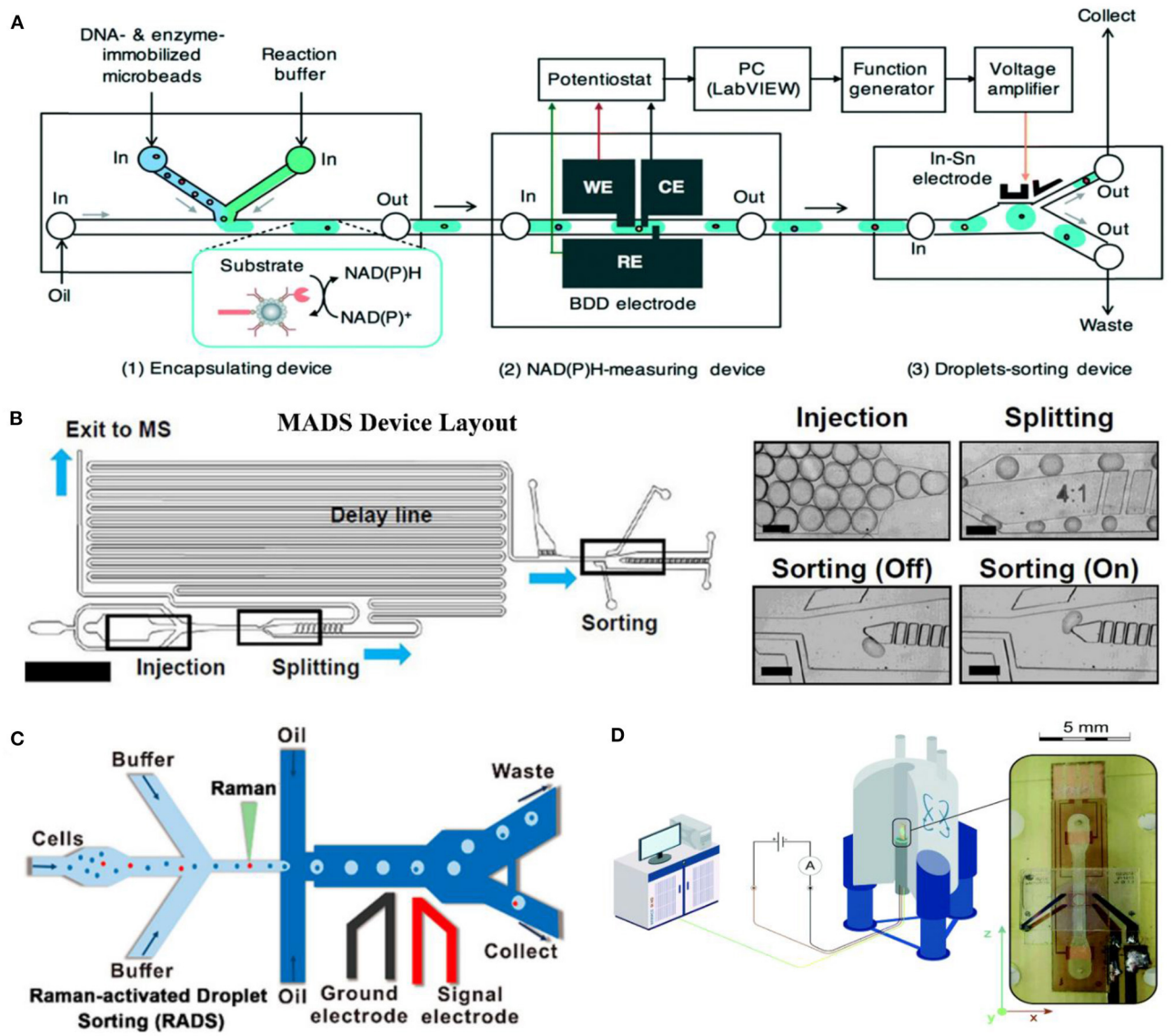


FIGURE 3 | Label-free sorting methods based on electrochemical detection **(A)** [adapted with permission from Goto et al. (2020) Copyright © Royal Society of Chemistry], mass spectrometry **(B)** [adapted with permission from Holland-Moritz et al. (2020) Copyright © John Wiley and Sons, Inc.], Raman **(C)** [adapted with permission from Wang et al. (2017) Copyright © American Chemical Society], and NMR **(D)** [adapted with permission from Davoodi et al. (2020) Copyright © Royal Society of Chemistry].

stable electrospray. Sample overloading might cause transfer line contamination and signal saturation.

Researchers also developed another strategy of applying a specific rate of sheath flow so that the whole w/o droplets could be directly infused into ESI (Diefenbach et al., 2018; Holland-Moritz et al., 2020). Carryover could be eliminated by replacing hydrophilic stainless (SS) needle with Teflon ESI needle to avoid cross contamination between droplets (Diefenbach et al., 2018).

Since MS detection is sample disruptive (Haidas et al., 2019), MS is usually coupled with microfluidic for enzymatic catalysis analysis. Holland-Moritz et al. (2020) developed a mass-activated droplet sorting (MADS) system based on ESI (**Figure 3B**). In their system, 25-nl droplets are split into two portions by a T-split line on a chip almost asymmetrically. A 15-nl daughter droplet flows directly into perfluoroalkoxy alkane (PFA) capillary for quadrupole mass spectrometer via a sheath-flow ESI source. The other daughter droplet is sorted by DEP. Sorting decisions are made based on the MS signals and time delay between these two daughter droplets. This system achieved 0.7 samples per second with 98% accuracy in transaminase ATA-117 cell-free system.

The MS detection rate is normally around 0.5–1.0 s per droplet (Oyobiki et al., 2014; Mahler et al., 2018) and requires nanoliter-sized droplets (Wink et al., 2018; Holland-Moritz et al., 2020). Up to 33-Hz nanosized droplet sorting rate was achieved by Kempa et al. (2020), by coupling traveling-wave ion mobility quadrupole time of flight (TWIMS Q-TOF) with microfluidics. Such on-chip

MS detection has been applied in chemical reaction analysis, either with cells (Diefenbach et al., 2018; Mahler et al., 2018) or not (Beulig et al., 2017).

Raman-Activated Droplet Sorting

Unlike MS, RADS is non-invasive but also label-free. RADS could sort cells at a rate of hundreds of cells per minute [up to 500 cells/h achieved (Lee et al., 2020a)]. Wang et al. (2017) (**Figure 3C**) developed RADS platform by placing single-cell Raman spectrum (SCRS) prior to droplet generation, since Raman background of oil medium might adversely affect the system's accuracy. Single cells are detected by SCRS, encapsulated into w/o droplet and then directly flow to DEP sorting. This system was successfully applied to screen out *Haematococcus pluvialis* with high astaxanthin yield and a relatively high throughput (∼260 cells/min) and high accuracy (∼98%) were achieved.

Most Raman-based sorting assays are based on resonance signals. Mcilvenna et al. (2016) reported continuous sorting of cyanobacteria based on carotenoids with Raman-microfluidic system. By adding ^{13}C-bicarbonate into the culturing medium, shifts in carotenoid bands could be measured, indicating active dissolved-CO_2-fixing cells. Lee et al. (2020b) prepared a D_2O-containing minimal medium supplemented with unlabeled interested compound mucin and detected the changes of cytochrome signal with Raman-based microfluidics. Mucin-utilizing Muribaculaceae strains were successfully screened out from a mouse with this device.

Resonance signals however associate with only a few classes of cellular compounds like pigments and therefore limit the application genotype scope. Actually, Raman spectrum is informative in not only resonance signals but also non-resonance ones. Non-resonance signals associate with more chemicals *in vivo* (e.g., starch, protein, and nucleic acid); nevertheless, they need longer acquisition time, which usually conflicts with throughput.

Recent breakthroughs have just been made in developing non-resonance-based RADS. A significantly improved rate of 120 cells/min and two variants of an unknown enzyme [algal diacylglycerol acyltransferases (DGATs)] were discovered by applying this RADS setup (Wang et al., 2020). A quartz-made chip, with low background signals, could be used for non-resonance signal RADS rather than PDMS.

The major disadvantage of Raman spectrometry is its relative low sensitivity. Researchers make attempts by fabricating surface-enhanced Raman spectrometry (SERS) substrates to improve the sensitivity. Sobota et al. (2019) applied a SERS substrate SK307 and detected as low as 200 μM of 1,2,3-trichloropropane on SERS-coupled microfluidics, which might inspire the screening of haloalkane dehalogenase enzymes.

NMR-Based Droplet Sorting

NMR could give information from all states of matter in a wide range of temperatures. Unlike MS, NMR offers an option of a completely non-invasive metabolomic readout. Theoretically, NMR is obviously disadvantageous in its low intrinsic mass sensitivity, which means normal concentrations below 1 mM are hard to observe (Davoodi et al., 2020). Davoodi et al. (2020) developed an NMR-compatible microfluidic platform by placing the metal tracks in the side walls of a microfluidic channel (**Figure 3D**). NMR radio-frequency excitation performance was found to be actually enhanced without compromising B_0 homogeneity.

For NMR, pre-shimming of samples is significant, so that homogeneity and stability of the magnetic field could be obtained. Shimming is usually performed by applying currents to various shim coils. With the design of highly efficient planar NMR Helmholtz microcoil and transmission line resonators, the problem of the NMR sensitivity on small volume samples could be solved. Van Meerten et al. (2018) simplified the regular shim coils with a series of parallel wires placed perpendicular to B_0 as a Shim-on-Chip shim system, which is particularly suited for microliter samples in capillaries. To further address the challenge of interfacing microcoils with droplets, Lei et al. (2015) firstly reported interface between digital microfluidics and low-field NMR. Later, Swyer et al. (2016) developed the first digital microfluidic system capable of interfacing droplets of analyte with microcoils in high-field NMR, which is appropriate for chemical characterization. The system was successfully applied to monitor a glucose oxidase reaction in 4-μl droplets, but not for sorting yet.

Another challenge that hinders NMR application in microfluidics is the preservation of high spectral resolution, which requires a highly homogenous magnetic field over the sample volume. However, differences in magnetic susceptibility between the chip, the continuous phase, and the droplet phase will lead to a demagnetizing field that varies continuously over the sample volume (Hale et al., 2018). Researchers show that susceptibility difference between the chip and the continuous phase could be mitigated by a combination of structural shimming and doping of the less diamagnetic of the liquid phases with a europium compound, such as Eu^{3+} compound (Hale et al., 2018).

DISCUSSION AND CONCLUSIONS

Droplet-based microfluidics is becoming a powerful toolbox for enzyme DE, especially for a randomly generated mutant library. There are some breakthroughs of highly efficient enzymes screened out by microfluidics. Droplets usually go through "generation," "incubation," "manipulation (optional)," and "sorting" steps on a chip. For the first three steps, there have been many technological advances in the last decade, and there have already been commercial droplet generation toolkits and droplet manipulation devices like a picoinjector (Abate et al., 2010). More generally, droplets could be generated at several to tens of kHz frequency (Zhu and Wang, 2017). The major limitation lies in the sorting step, whose diversity and frequency restrict the sorting efficiency and target enzyme scope.

FADS is the most mature sorting method and reaches the highest sorting rate in all sorting methods. FADS is also of highest sensitivity and could be as low as 2.5 nM of fluorescein at a 2-pl droplet (Hasan et al., 2019), which means less than

a single turnover for all enzyme molecules from a single cell (**Table 1**). Genotype is usually linked with fluorescent phenotype by fluorophore activation (Qiao et al., 2018), quencher release (Nikoomanzar et al., 2019), or coupled assay. In general, FADS is the primary choice if fluorescence could be achieved in the enzymatic reaction. AADS is similar with FADS, even if of lower efficiency and sensitivity, which could offer another option for broader enzyme applications. AADS and FADS could even be combined into one chip to be more informative. Siltanen et al. (2018) designed a FAADS (fluorescence and AADS) device in which a source-coupled fiber for excitation and two fibers for transmittance and emission were embedded on the chip. The excitation fiber is connected with continuous-wave lasers with 405, 473, 532, and 640 nm. Emitted light and transmitted light are collected by the other fibers, so that optical density and fluorescence values are obtained almost simultaneously.

Label-free sorting methods are drawing more and more attention, since they are advantageous in maintaining the integrity and independence of the whole enzymatic reaction system, away from any extra disruptions. What is more, they offer various options for researchers and save the trouble of linking the genotype with fluorescent phenotype. In principle, they all could be used for droplet screening and recycling for further gene recovery and sequencing. Even for MADS, splitting chip made it possible to detect and recycle droplets (Holland-Moritz et al., 2020). Challenges for broad applications of label-free detection lie in low sorting frequency and complex sorting device design. For electrochemical sorting, there has not been a standardized way for the sorting device and might even need to be customized according to various electrochemical characteristics. RADS and NMR-based sorting requires special techniques to set up the sorting system. Standardization in design and convenience for microfluidic coupling will be the future direction for label-free detection setup. Presently, there are just few application cases of label-free detection in enzyme DE.

When choosing among all sophisticated droplet-based sorting methods, we need to consider cell species, enzyme type, enzyme yield, enzymatic efficiency, etc. FADS and AADS would be the primary two choices if optical change could be linked in the enzymatic reaction. The droplet size for AADS needs to be carefully considered, since sufficient enzyme molecules need to be accumulated to stimulate the optical detection due to AADS's relative low sensitivity. Even though there are rare application cases, electrochemical-based sorting is promising in a wider application if the physical electrodes and electrochemical sensors are more versatile. MADS is efficient for low-concentration substrate and could be an option especially for isomers. Raman-activated cell sorting (RACS) would be a good choice if richer information is required apart from the enzymatic reaction itself (Wang et al., 2020); meanwhile, the relative low throughput can be compromised. Moreover, another informative sorting method-NMR-based sorting is molecule-informative while of low throughput (Hale et al., 2018). **Table 1** shows the optimal specification in which different sorting methods have been achieved in recent publications.

Technically speaking, a sorting device includes a detector and a sorter. There are various sorters like acoustic, magnetic, pneumatic, thermal, and electric actuation sorters. Among them, DEP is the most widely used one. The typical DEP sorting setup consists of a sorting junction linked by two-way channels. Without an electric field, droplets would be sent to a waste output channel. Otherwise, an electric field is triggered by the detection part, and positive droplets could be sent to the collection channel by adjusting the hydrodynamic resistance (higher than that of the waste channel) (Frenzel and Merten, 2017).

There are also some attempts to extend traditional DEP sorter to multiple channels, so that the whole system could reach a higher efficiency. Frenzel and Merten (2017) designed a sorting module, in which four channels sequentially branch off from the waste channel and each collection channel has its own electrode pair running parallel to the channel wall. They achieved an ∼100% reliable two-way sorting, largely independent of the relative flow rates in the channels downstream of the sorting junction. In another design (Caen et al., 2018), five channels were designed with the sorter, and droplets could be sorted by different voltages applied by two symmetrical live and ground electrodes. The device was successfully applied to screen droplets of different concentrations of a fluorescent dye sulforhodamine B and reached a sorting rate of up to 200 droplets per second (Caen et al., 2018). Combined with downstream next-generation sequencing, a multichannel could offer more information about variants and help accelerate the DE rate by deep learning. Those developments will help make droplet sorting for enzyme DE more versatile in the future.

OUTLOOK

DE of enzymes toward high specificity and efficiency is significant to both scientific researches and industrial applications. Droplet-based microfluidics paves a cheap and convenient way for enzyme DE with ultra-high throughput, and meanwhile, it is becoming a useful tool for *de novo* synthetic enzyme screening. Sorting methods, as the main step in DE, determine the efficiency of the whole system. Sorting devices are developing toward standardized modules compatible with different instruments and microfluidic chips. Combinations of different sorting methods could help gain multiplex perspectives into enzyme and boost a wide range of application.

AUTHOR CONTRIBUTIONS

XF and FM conceived the concept, conducted literature survey, and drafted and revised the manuscript. All the authors organized the figures and approved them for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (82073911), Shandong Provincial Natural Science Foundation (ZR2020MH389), Medical and Health Science and Technology Project of Shandong Province (2017WS075, 202011000657), and The Innovation Project of Shandong Academy of Medical Sciences.

REFERENCES

Abate, A. R., Hung, T., Mary, P., Agresti, J. J., and Weitz, D. A. (2010). High-throughput injection with microfluidics using picoinjectors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19163–19166. doi: 10.1073/pnas.1006888107

Abbyad, P., Pan, C. W., Horvath, D. G., Braza, S., Moore, T., Lynch, A., et al. (2019). Sorting by Interfacial Tension (SIFT): label-free selection of live cells based on single-cell metabolism. *Lab Chip* 19:1344. doi: 10.1039/C8LC01328D

Attene-Ramos, M. S., Austin, C. P., and Xia, M. (2014). *High Throughput Screening, 3rd Edn*. Amsterdam: Elsevier.

Beneyton, T., Wijaya, I. P. M., Postros, P., Najah, M., Leblond, P., Couvent, A., et al. (2016). High-throughput screening of filamentous fungi using nanoliter-range droplet-based microfluidics. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep27223

Beulig, R. J., Warias, R., Heiland, J. J., Ohla, S., Zeitler, K., and Belder, D. (2017). A droplet-chip/mass spectrometry approach to study organic synthesis at nanoliter scale. *Lab Chip* 17, 1996–2002. doi: 10.1039/C7LC00313G

Bransky, A., Korin, N., Khoury, M., and Levenberg, S. (2009). A microfluidic droplet generator based on a piezoelectric actuator. *Lab Chip* 9, 516–520. doi: 10.1039/B814810D

Caen, O., Schütz, S., Jammalamadaka, M. S. S., Vrignon, J., Nizard, P., Schneider, T. M., et al. (2018). High-throughput multiplexed fluorescence-activated droplet sorting. *Microsys. Nanoeng.* 4:33. doi: 10.1038/s41378-018-0033-2

Colin, P., Kintses, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., et al. (2015). Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.* 6, 1–12. doi: 10.1038/ncomms10008

Davoodi, H., Nordin, N., Bordonali, L., Korvink, J. G., MacKinnon, N., and Badilita, V. (2020). An NMR-compatible microfluidic platform enabling: *in situ* electrochemistry. *Lab Chip* 20, 3202–3212. doi: 10.1039/D0LC00364F

Debon, A., Pott, M., Obexer, R., Green, A. P., Friedrich, L., Griffiths, A. D., et al. (2019). Ultrahigh-throughput screening enables efficient single-round oxidase remodelling. *Nat. Catal.* 2, 740–747. doi: 10.1038/s41929-019-0340-5

Diefenbach, X. W., Farasat, I., Guetschow, E. D., Welch, C. J., Kennedy, R. T., Sun, S., et al. (2018). Enabling biocatalysis by high-throughput protein engineering using droplet microfluidics coupled to mass spectrometry. *ACS Omega* 3, 1498–1508. doi: 10.1021/acsomega.7b01973

Fenneteau, J., Chauvin, D., Griffiths, A. D., and Cossy, J. (2017). Synthesis of new hydrophilic rhodamine based enzymatic substrates compatible. *Chem. Commun.* 53, 5437–5440. doi: 10.1039/C7CC01506B

Frenzel, D., and Merten, C. A. (2017). Microfluidic train station: highly robust and multiplexable sorting of droplets on electric rails. *Lab Chip* 17, 1024–1030. doi: 10.1039/C6LC01544A

Gielen, F., Hours, R., Emond, S., Fischlechner, M., Schell, U., and Hollfelder, F. (2016). Ultrahigh-throughput-directed enzyme evolution by Absorbance-Activated Droplet Sorting (AADS). *Proc. Natl. Acad. Sci. U.S.A.* 113, E7383–E7389. doi: 10.1073/pnas.1606927113

Goto, H., Kanai, Y., Yotsui, A., Shimokihara, S., Shitara, S., Oyobiki, R., et al. (2020). Microfluidic screening system based on boron-doped diamond electrodes and dielectrophoretic sorting for directed evolution of NAD(P)-dependent oxidoreductases. *Lab Chip* 20, 852–861. doi: 10.1039/C9LC01263J

Haidas, D., Bachler, S., Ko, M., Blank, L. M., Zenobi, R., and Dittrich, P. S. (2019). Microfluidic platform for multimodal analysis of enzyme secretion in nanoliter droplet arrays. *Anal. Chem.* 91, 2066–2073. doi: 10.1021/acs.analchem.8b04506

Hale, W., Rossetto, G., Greenhalgh, R., Finch, G., and Utz, M. (2018). High-resolution nuclear magnetic resonance spectroscopy in microfluidic droplets. *Lab Chip*. 18, 3018–3024. doi: 10.1039/C8LC00712H

Hasan, S., Geissler, D., Wink, K., Hagen, A., Heiland, J. J., and Belder, D. (2019). Fluorescence lifetime-activated droplet sorting in microfluidic chip systems. *Lab Chip* 19, 403–409. doi: 10.1039/C8LC01278D

Holland-Moritz, D.A., Wismer, M.K., Mann, B.F., Farasat, I., Devine, P., Guetschow, E.D., et al. (2020). Mass Activated Droplet Sorting (MADS) enables high-throughput screening of enzymatic reactions at nanoliter scale. *Angew. Chem. Int. Ed.* 59, 4470–4477. doi: 10.1002/anie.201913203

Horvath, D. G., Braza, S., Moore, T., Pan, C. W., Zhu, L., Shun, O., et al. (2019). Analytica chimica acta Sorting by Interfacial Tension (SIFT): label-free enzyme sorting using droplet microfluidics. *Anal. Chim. Acta* 1089, 108–114. doi: 10.1016/j.aca.2019.08.025

Kempa, E. E., Smith, C. A., Li, X., Bellina, B., Richardson, K., Pringle, S., et al. (2020). Coupling droplet microfluidics with mass spectrometry for ultrahigh-throughput analysis of complex mixtures up to and above 30 Hz. *Anal. Chem.* 92, 12605–12612. doi: 10.1021/acs.analchem.0c02632

Larsen, A. C., Dunn, M. R., Hatch, A., Sau, S. P., Youngbull, C., and Chaput, J. C. (2016). A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.* 7:11235. doi: 10.1038/ncomms11235

Lee, K. S., Pereira, F. C., Palatinszky, M., Behrendt, L., Alcolombri, U., Berry, D., et al. (2020a). Optofluidic raman-activated cell sorting for targeted genome retrieval or cultivation of microbial cells with specific functions. *Nat. Protoc.* 16, 634–676. doi: 10.1038/s41596-020-00427-8

Lee, K. S., Wagner, M., and Stocker, R. (2020b). Raman-based sorting of microbial cells to link functions to their genes. *Microb. Cell* 7, 62–65. doi: 10.15698/mic2020.03.709

Lei, K., Mak, P., Law, M., and Martins, R. P. (2015). A palm-size MNMR relaxometer using a digital microfluidic device and a semiconductor transceiver for chemical/biological diagnosis. *Lab Chip*. 140, 5129–5137. doi: 10.1039/C5AN00500K

Lim, J., and Gruner, P. (2020). Bacterial expression systems for enzymatic activity in droplet-based microfluidics. *Anal. Chem.* 92, 4088–4916. doi: 10.1021/acs.analchem.9b04969

Ma, C., Tan, Z. L., Lin, Y., Han, S., Xing, X., and Zhang, C. (2019). Gel Microdroplet–based high-throughput screening for directed evolution of xylanase-producing pichia pastoris. *J. Biosci. Bioeng.* 128, 662–668. doi: 10.1016/j.jbiosc.2019.05.008

Ma, F., Chung, M. T., Yao, Y., Nidetz, R., Lee, L. M., Liu, A. P., et al. (2018). Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. *Nat. Commun.* 9, 1–8. doi: 10.1038/s41467-018-03492-6

Maceiczyk, R. M., Hess, D., Chiu, F. W. Y., Stavrakis, S., and Andrew, J. (2017). Differential detection photothermal spectroscopy: detection in picoliter and femtoliter droplets. *Lab Chip* 17, 3654–3663. doi: 10.1039/C7LC00946A

Mahler, L., Wink, K., Julia, R. B., Scherlach, K., Tovar, M., Zang, E., et al. (2018). Detection of antibiotics synthetized in microfluidic picolitre-droplets by various actinobacteria. *Sci. Rep.* 8:13087. doi: 10.1038/s41598-018-34069-4

Markel, U., Essani, K. D., Besirlioglu, V., Schiffels, J., Streit, W. R., and Schwaneberg, U. (2020). Advances in ultrahigh-throughput screening for directed enzyme evolution. *Chem. Soc. Rev.* 49, 233–262. doi: 10.1039/C8CS00981C

Marquis, M., Davy, J., Cathala, B., Fang, A., and Renard, D. (2015). Microfluidics assisted generation of innovative polysaccharide hydrogel microparticles. *Carbohydr. Polym.* 116, 189–199. doi: 10.1016/j.carbpol.2014.01.083

Martis, E. A., Radhakrishnan, R., and Badve, R. R. (2011). High-throughput screening: the hits and leads of drug discovery-an overview. *J. Appl. Pharm. Sci.* 1, 2–10.

Mcilvenna, D., Huang, W. E., Davison, P., and Glidle, A. (2016). Continuous cell sorting in a flow based on single cell resonance raman spectra. *Lab Chip* 16, 1420–1429. doi: 10.1039/C6LC00251J

Nabavi, S. A., Vladisavljević, G. T., Gu, S., and Ekanem, E. E. (2015). Double emulsion production in glass capillary microfluidic device: parametric investigation of droplet generation behaviour. *Chem. Eng. Sci.* 130, 183–196. doi: 10.1016/j.ces.2015.03.004

Neun, S., Kaminski, T. S., and Hollfelder, F. (2019). *Single-Cell Activity Screening in Microfluidic Droplets, 1st Edn*. Amsterdam: Elsevier Inc.

Nikoomanzar, A., Vallejo, D., and Chaput, J. C. (2019). Elucidating the determinants of polymerase specificity by microfluidic-based deep mutational scanning. *ACS Synth. Biol.* 8, 1421–1429. doi: 10.1021/acssynbio.9b00104

Obexer, R., Godina, A., Garrabou, X., Mittl, P. R. E., Baker, D., Grif, A. D., et al. (2017). Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* 9:50. doi: 10.1038/nchem.2596

Oyobiki, R., Kato, T., Katayama, M., Sugitani, A., Watanabe, T., Einaga, Y., et al. (2014). Toward high-throughput screening of NAD(P)-dependent oxidoreductases using boron-doped diamond microelectrodes and micro Fl uidic devices. *Anal. Chem.* 86, 9570–9575. doi: 10.1021/ac501907x

Packer, M. S., and Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16, 379–394. doi: 10.1038/nrg3927

Payne, E. M., Holland-Moritz, D. A., Sun, S., and Kennedy, R. T. (2020). High-throughput screening by droplet microfluidics: perspective into key challenges and future prospects. *Lab Chip* 20, 2247–2262. doi: 10.1039/D0LC00347F

Prodanović, R., Ung, W. L., Đurdić, K. I., Fischer, R., Weitz, D. A., and Ostafe, R. (2020). A high-throughput screening system based on droplet microfluidics for glucose oxidase gene libraries. *Molecules* 25:2418. doi: 10.3390/molecules25102418

Qiao, Y., Zhao, X., Zhu, J., Tu, R., Dong, L., Wang, L., et al. (2018). Fluorescence-activated droplet sorting of lipolytic microorganisms using a compact optical system. *Lab Chip* 18, 190–196. doi: 10.1039/C7LC00993C

Siltanen, C. A., Cole, R. H., Poust, S., Chao, L., Tyerman, J., Kaufmann-Malaga, B., et al. (2018). An oil-free picodrop bioassay platform for synthetic biology. *Sci. Rep.* 8, 1–7. doi: 10.1038/s41598-018-25577-4

Sobota, J., Samek, O., Republic, C., Republic, C., Laboratories, L., and Republic, C. (2019). "Surface-enhanced Raman spectroscopy in microfluidic chips for directed evolution of enzymes and environmental monitoring," in *Proceedings of the 2019 PhotonIcs and Electromagnetics Research Symposium-Spring (PIERS-Spring)* (Rome: IEEE), 17–20.

Swyer, I., Soong, R., Dryden, M. D. M., Fey, M., Maas, W. E., and Wheeler, A. R. (2016). Interfacing digital microfluidics with high-field nuclear magnetic resonance spectroscopy. *Lab Chip*. 16, 4424–4435. doi: 10.1039/C6LC01073C

Tarchichi, N., Chollet, F., and Manceau, J. F. (2013). New regime of droplet generation in a T-shape microfluidic junction. *Microfluid. Nanofluidics* 14, 45–51. doi: 10.1007/s10404-012-1021-8

Um, E., Lee, D. S., Pyo, H. B., and Park, J. K. (2008). Continuous generation of hydrogel beads and encapsulation of biological materials using a microfluidic droplet-merging channel. *Microfluid. Nanofluidics* 5, 541–549. doi: 10.1007/s10404-008-0268-6

Vallejo, D., Nikoomanzar, A., and Chaput, J. C. (2020). Directed evolution of custom polymerases using droplet microfluidics. *Methods Enzymol.* 644, 227–253. doi: 10.1016/bs.mie.2020.04.056

Vallejo, D., Nikoomanzar, A., Paegel, B. M., and Chaput, J. C. (2019). Fluorescence-activated droplet sorting for single-cell directed evolution. *ACS Synth. Biol.* 8, 1430–1440. doi: 10.1021/acssynbio.9b00103

Van Meerten, S. G. J., Van Bentum, P. J. M., and Kentgens, A. P. M. (2018). Shim-on-Chip design for microfluidic NMR detectors. *Anal. Chem.* 90, 10134–10138. doi: 10.1021/acs.analchem.8b02284

Wang, X., Ren, L., Su, Y., Ji, Y., Liu, Y., Li, C., et al. (2017). Raman-activated droplet sorting (RADS) for label-free high-throughput screening of microalgal single-cells. *Anal. Chem.* 89, 12569–12577. doi: 10.1021/acs.analchem.7b03884

Wang, X., Xin, Y., Ren, L., Sun, Z., Zhu, P., Ji, Y., et al. (2020). Positive dielectrophoresis-based raman-activated droplet sorting for culture-free and label-free screening of enzyme function *in vivo*. *Sci. Adv.* 6:eabb3521. doi: 10.1126/sciadv.abb3521

Wink, K., Mahler, L., Beulig, J. R., Piendl, S. K., Roth, M., and Belder, D. (2018). An integrated chip-mass spectrometry and epifluorescence approach for online monitoring of bioactive metabolites from incubated actinobacteria in picoliter droplets. *Anal. Bioanal. Chem* 410, 7679–7687. doi: 10.1007/s00216-018-1383-1

You, C., and Percival Zhang, Y. H. (2012). Easy preparation of a large-size random gene mutagenesis library in *Escherichia coli*. *Anal. Biochem.* 428, 7–12. doi: 10.1016/j.ab.2012.05.022

Zeymer, C., and Hilvert, D. (2018). Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* 87, 131–157. doi: 10.1146/annurev-biochem-062917-012034

Zhu, P., and Wang, L. (2017). Passive and active droplet generation with microfluidics: a review. *Lab Chip* 17, 34–75. doi: 10.1039/C6LC01018K

Zurek, P. J., Hours, R., and Schell, U. (2021). Growth amplification in ultrahigh-throughput microdroplet screening increases sensitivity of clonal enzyme assays and minimizes phenotypic variation. *Lab Chip* 21, 163–173. doi: 10.1039/D0LC00830C

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 24 May 2021
doi: 10.3389/fchem.2021.663241

Tying up the Loose Ends: A Mathematically Knotted Protein

Shang-Te Danny Hsu[1,2]*, *Yun-Tzai Cloud Lee*[1,2‡], *Kornelia M. Mikula*[3‡], *Sofia M. Backlund*[3], *Igor Tascón*[3†], *Adrian Goldman*[4,5] *and Hideo Iwaï*[3]*

[1]*Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan,* [2]*Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan,* [3]*Institute of Biotechnology, University of Helsinki, Helsinki, Finland,* [4]*Division of Biochemistry, Department of Biosciences, University of Helsinki, Helsinki, Finland,* [5]*Astbury Centre for Structural Molecular Biology, School of Biomedical Sciences, University of Leeds, West Yorkshire, United Kingdom*

Knots have attracted scientists in mathematics, physics, biology, and engineering. Long flexible thin strings easily knot and tangle as experienced in our daily life. Similarly, long polymer chains inevitably tend to get trapped into knots. Little is known about their formation or function in proteins despite >1,000 knotted proteins identified in nature. However, these protein knots are not mathematical knots with their backbone polypeptide chains because of their open termini, and the presence of a "knot" depends on the algorithm used to create path closure. Furthermore, it is generally not possible to control the topology of the unfolded states of proteins, therefore making it challenging to characterize functional and physicochemical properties of knotting in any polymer. Covalently linking the amino and carboxyl termini of the deeply trefoil-knotted YibK from *Pseudomonas aeruginosa* allowed us to create the truly backbone knotted protein by enzymatic peptide ligation. Moreover, we produced and investigated backbone cyclized YibK without any knotted structure. Thus, we could directly probe the effect of the backbone knot and the decrease in conformational entropy on protein folding. The backbone cyclization did not perturb the native structure and its cofactor binding affinity, but it substantially increased the thermal stability and reduced the aggregation propensity. The enhanced stability of a backbone knotted YibK could be mainly originated from an increased ruggedness of its free energy landscape and the destabilization of the denatured state by backbone cyclization with little contribution from a knot structure. Despite the heterogeneity in the side-chain compositions, the chemically unfolded cyclized YibK exhibited several macroscopic physico-chemical attributes that agree with theoretical predictions derived from polymer physics.

Keywords: Knotted proteins, NMR spectrocopy, protein *trans*-splicing, enzymatic ligation, protein dynamics, protein stability and folding

OPEN ACCESS

Edited by:
Assaf Friedler,
Hebrew University of Jerusalem, Israel

Reviewed by:
Sophie Elizabeth Jackson,
University of Cambridge,
United Kingdom
Stefan G. D. Rüdiger,
Utrecht University, Netherlands

***Correspondence:**
Shang-Te Danny Hsu
sthsu@gate.sinica.edu.tw
Hideo Iwaï
hideo.iwai@helsinki.fi

†Present Address:
BIOFISIKA Institute (UPV/EHU, CSIC),
Barrio Sarriena S/n 48940 Leioa,
Bizkaia, Basque Country, Spain
‡*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *02 February 2021*
Accepted: *20 April 2021*
Published: *24 May 2021*

Citation:
Hsu S-TD, Lee Y-TC, Mikula KM, Backlund SM, Tascón I, Goldman A and Iwaï H (2021) Tying up the Loose Ends: A Mathematically Knotted Protein. Front. Chem. 9:663241. doi: 10.3389/fchem.2021.663241

INTRODUCTION

Knots always fascinate people and have attracted scientists from all disciplines. Long flexible strings can spontaneously knot themselves upon agitation (Raymer and Smith, 2007). Whereas circular supercoiled DNA in nature can be a true mathematical knot, proteins are linear polymers consisting of 20 different amino acids connected by peptide bonds with open amino (N) and carboxyl (C) termini. Proteins fold into defined three-dimensional (3D) conformations and execute various functions at the molecular level. The apparent complexity of threading events involved in tying a

protein knot made them inconceivable to many structural biologists at first (Mansfield, 1994). Nevertheless, systematic surveys of the protein database have identified more than 1,000 knotted protein structures with different knot types and structural complexities. However, in this context, "knot" does not imply a topological knot, which cannot be undone except by breaking the protein backbone, a knot in the sense of knotted tie, or a sailor's reef knot (Jamroz et al., 2015; Lim and Jackson, 2015). It has been challenging to reconcile experimental and theoretical views of how a polypeptide chain attains an intricately knotted topology (Mallam et al., 2008; Mallam and Jackson, 2012; Beccara et al., 2013; Sulkowska et al., 2013; Lim and Jackson, 2015; Ziegler et al., 2016; Jackson et al., 2017). Now, there have been many experimental studies for better understanding of the protein knotting mechanisms. While the majority of experimental studies showed that knotted proteins fold into the knotted conformations with highly populated folding intermediates along their kinetic folding pathways (Mallam and Jackson, 2007; Andersson et al., 2009; Lim and Jackson, 2015; Wang et al., 2015; Lou et al., 2016; Wang et al., 2016; Dabrowski-Tumanski and Sulkowska, 2017; Jackson et al., 2017; He et al., 2019; Jarmolinska et al., 2019; Rivera et al., 2020), there are also some knotted (or slipknotted) proteins that can fold without populating intermediate states (He et al., 2019; Rivera et al., 2020). Different experimental studies have shown that knotting is rate-limiting (Mallam and Jackson, 2012). Computational approaches have also been used to verify the experimental observations, such as the rugged free energy landscapes of several knotted proteins and multiple intermediates populated along their folding pathways. These computational studies might explain the rate-limiting step of protein knotting (Li et al., 2012; Beccara et al., 2013; Sulkowska and et al., 2013; Faísca, 2015) and lead to various protein knotting mechanisms, such as direct threading, slipknotting, and mousetrapping (Noel et al., 2010; Covino et al., 2014). Only very recently, we combined experimental and computational data to obtain a converged view of how the smallest knotted protein, MJ0366, attains a knotted transition state (Passioni et al., 2021). However, all investigations of protein knots have so far been reported for protein knots with open ends because proteins are synthesized as linear polypeptide chains. As such, proteins are not true mathematical knots but are defined by virtual connections of the N- and C-termini by different mathematical schemes (Taylor, 2000; Lai et al., 2012; Millett et al., 2013).

In this study, we asked whether a backbone knotted protein without open ends could be generated by backbone cyclization. Whereas backbone cyclization of proteins has been widely accepted to stabilize proteins (Iwai and Plückthun, 1999; Scott et al., 1999; Clark and Craik, 2010; Montalbán-López et al., 2012; Borra and Camarero, 2013), disulfide bridges, which were originally considered to stabilize proteins by reducing the entropy of the denatured state as backbone cyclization, have more enthalpic contributions to the stability in the folded state (Mitchinson and Wells, 1989; Betz, 1993). Moreover, backbone cyclization of knotted proteins could create the unique possibility to investigate the unfolded state of proteins with a knot unequivocally. A backbone cyclized knotted protein would conform to the mathematical definition for a truly knotted topology whose path closure does not depend on how the ends are joined together in space. Whereas naturally occurring protein knots have open polypeptide-chain ends and could be disentangled into linear polypeptide chains without knots and entanglements under denaturing conditions, backbone cyclized knotted proteins cannot untie anymore without proteolysis even under denaturing conditions.

Here, we presented the unprecedented characterization of a mathematical backbone protein knot without open peptide-chain ends using various structural and biophysical methods including SAXS, X-ray crystallography, and ^{15}N nuclear relaxation analysis by NMR spectroscopy.

RESULTS

Production of a Knotted Protein With and Without Open Ends

To produce a protein knot without open polypeptide ends, we chose the highly conserved bacterial RNA methyltransferase as a model system, namely YibK from *Pseudomonas aeruginosa* (*Pa*YibK). YibK contains a trefoil (3_1) knotted backbone topology. *Pa*YibK also shares 65% sequence identity with one of the most-studied knotted proteins, YibK form *Haemophilus influenzae* (*Hi*YibK), but contains only one tryptophan (**Supplementary Figure S1**). (Tkaczuk et al., 2007) We first determined the crystal structure of the wild-type *Pa*YibK in its linear form (*Pa*YibK) to confirm the same trefoil knot structure and the dimeric assembly as found in *Hi*YibK (**Figure 1A**; **Supplementary Table S1**; and **Supplementary Figure S1**). The root mean square deviation (RMSD) was only 0.6 Å between the crystal structures of *Pa*YibK and *Hi*YibK, indicating a highly conserved higher-order structure among YibK proteins throughout evolution including the knot structure (**Figure 1A**). The N- and C-termini of *Pa*YibK are separated by ca. 8 Å in the crystal structure, which is sufficiently close for the head-to-tail backbone cyclization without disturbing the backbone knot structure. For backbone cyclization of proteins, various strategies have been established, including intein-mediated protein ligation (or expressed protein ligation), enzymatic protein cyclization using enzymes such as sortaseA (SrtA) and asparagine endopeptidase (AEP), and protein *trans*-splicing (PTS) (Scott et al., 1999; Popp and Ploegh, 2011; Mikula et al., 2017; Iwai et al., 2001). We initially attempted *in vivo* protein cyclization of *Pa*YibK by PTS using the naturally split DnaE intein (*Pa*YibK_Int) (**Figure 1B**) (Aranko et al., 2013), but the *in vivo* spliced product was insoluble (**Figure 1C**). PTS-based backbone cyclization relies on the self-association of intein fragments that brings the N- and C-termini together during the protein folding, followed by spontaneous auto-catalytic removal of the intein fragments, thereby achieving backbone cyclization (Scott et al., 1999; Iwai et al., 2001). We speculated that the rapid self-association of the split intein fragments could interfere with the folding and knotting of YibK, thereby resulting in insoluble

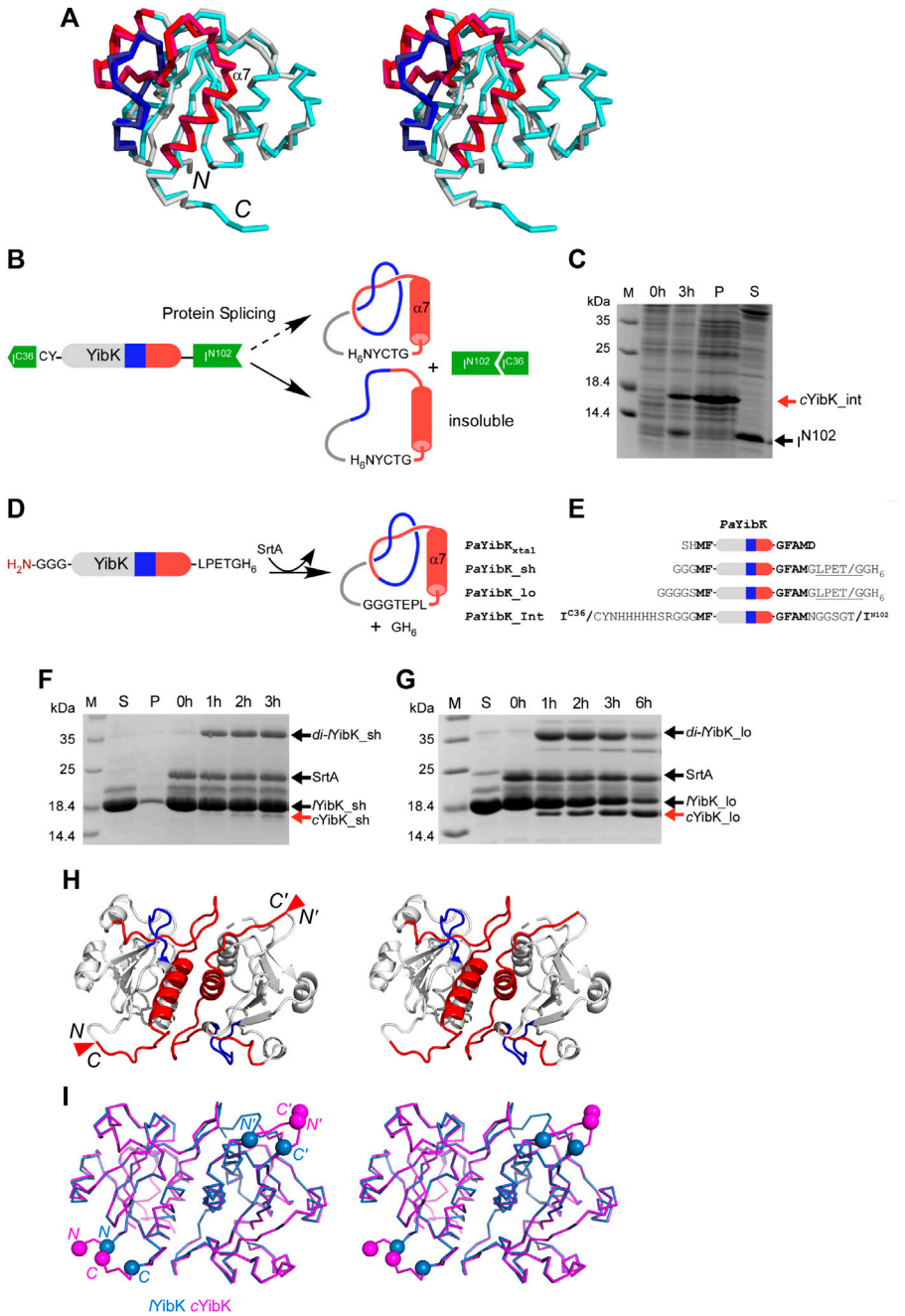


FIGURE 1 | Cyclization and structural analysis of a truly knotted *Pa*YibK. **(A)** A stereo view of the superposition of *Pa*YibK (pdb:6qkv) and *Ha*YibK (pdb:1j85) in ribbon presentation. **(B)** Schematic representation of the experimental strategy used for cyclization of YibK by the split-intein strategy. **(C)** *In vivo* backbone cyclization of YibK by protein *trans*-splicing. Spliced YibK and N-terminal fragment are indicated; M, Molecular weight marker P, insoluble fraction; S, soluble material. **(D)** Backbone cyclization of YibK using sortase A (SrtA). **(E)** Constructs used in backbone cyclization with different lengths of linkers connecting the termini. **(F)** and **(G)** Impact of linker lengths on the *in vitro* sortase ligation efficiency monitored by SDS-PAGE of linear YibK with short **(F)** and long **(G)** poly-glycine linkers annotated as *l*YibK_sh and *l*YibK_lo, respectively; 0–6 h: ligation reactions harvested at specified time points; SrtA, sortase added post-purification; *di*-YibK_sh/lo, undesired dimeric ligation products; *c*YibK_sh/lo, desired cyclized products. The N and C-terminal extensions of YibK for different ligation experiments are shown with their names indicated on the left. **(H)** A stereo-view of the crystal structure of *c*YibK_lo. The structures are shown in cartoon representation with the knotting loop and threading C-terminal helix shown in blue and red, respectively. The positions of the path closures between N- and C-termini are indicated by red triangles. **(I)** Comparison of the crystal structure of linear *Pa*YibK (6qkv; orchid blue) and *c*YibK (6qh8; magenta). The Cα atoms of the S1 and D155 of *Pa*YibK are shown in orchid blue spheres, and the positions of path closures in *c*YibK are indicated by magenta triangles as in **(H)**.

aggregation. The insoluble spliced product could be backbone cyclized YibK without knotting (*c*YibK_Int) (**Figure 1C**). Therefore, we used an alternative enzymatic approach with *S. aureus* sortase A (SrtA), which catalyzes a *trans*-peptidase reaction between the LPETG motif and an N-terminal triglycine peptide (**Figure 1D**). (Popp and Ploegh, 2011) When a seven-residue linker (*Pa*YibK_sh) was used to link the N and C-termini through SrtA-mediated ligation (**Figure 1E**), backbone cyclization was inefficient, and accumulation of a covalent dimer (*di-l*YibK_sh) resulting from intermolecular ligation was more prominent than the monomeric backbone-cyclized form (*c*YibK_sh; **Figure 1F**). This finding underscores the need to further optimize the linker length for the enzymatic ligation. By introducing a longer nine-residue linker to the backbone cyclization (*Pa*YibK_lo), a much more efficient backbone cyclization was achieved to produce a higher amount of monomeric cyclized YibK (*c*YibK_lo hereafter designated as *c*YibK) than the covalent dimer (*di-Pa*YibK_lo; **Figure 1G**). We also produced a linear form of YibK (*l*YibK) as the N-terminal SUMO fusion so that *l*YibK and *c*YibK (*Pa*YibK_lo) have the identical protein sequence for the analysis. The backbone cyclization of *c*YibK manifests in greater mobility during SDS-PAGE as compared with linear YibK (*l*YibK); the complete cyclization was also confirmed by mass spectrometry (**Supplementary Figure S2**).

Structural Comparison of a Knotted Protein With and Without Open Ends

We determined the crystal structures of *c*YibK to a resolution of 2.20 Å (**Figure 1H** and **Supplementary Table S1**). The root mean square deviation (RMSD) between the backbone Cα atoms of the crystal structures of *Pa*YibK and *c*YibK was <0.2 Å, confirming that the backbone cyclization did not perturb the 3D structure of *Pa*YibK in the folded state (**Figure 1I**). The structural similarity between *l*YibK and *c*YibK and their dimeric states were confirmed by small angle X-ray scattering (SAXS) in solution state (*vide infra*).

Functional Assessment of *l*YibK and *c*YibK

Next, we assessed the functional impact of cyclization on the cofactor binding activity, which is essential for the RNA methyltransferase activity of YibK. We used isothermal titration calorimetry (ITC) to determine the dissociation constants (K_d) of *l*YibK and *c*YibK for S-adenosyl-L-homocysteine (SAH), which is the product of the conserved RNA methylation reaction among all SPOUT family members that utilize S-adenosyl-L-methionine (SAM) as the methyl donor (Tkaczuk et al., 2007). The K_d for SAH was 8.80 ± 0.01 and 8.93 ± 0.03 μM for *l*YibK and *c*YibK, respectively, which corroborates our structural analyses showing no appreciable structural perturbation in *Pa*YibK after the backbone cyclization (**Supplementary Table S2**; **Supplementary Figure S4**). In contrast, side-chain disulfide bond-mediated cyclization of *Hi*YibK was reported to decrease SAH binding affinity (K_d increased from 20 to 71 μM) (Mallam et al., 2010). Furthermore, our experimental K_d values for SAH binding to *c*YibK/*l*YibK were 2 to 3-fold smaller than the reported values for other tRNA methyltransferases, namely *Hi*YibK (K_d = 20 μM) (Mallam and Jackson, 2007) and TrmL from *E. coli* (K_d = 25 μM) (Liu et al., 2013). ITC analysis revealed that SAH binding to *c*YibK was enthalpically more favorable than that of *l*YibK (ΔH = −20.1 vs −16.5 kcal mol^{-1}); the difference in enthalpic changes was compensated by the entropic differences (ΔS), resulting in the comparable net free energies of SAH binding (**Figure 2A**). The greater entropic loss in *c*YibK upon SAH binding may be associated with the dimer formation of *c*YibK.

Comparison of Folding of *l*YibK and *c*YibK

As *c*YibK and *l*YibK have the identical primary structure and crystal structure in their native states, we assume their associated free energy levels in the folded states are very similar except for the entropic contribution associated with the decreased degrees of freedom of the fraying ends by closing the ends. To investigate how a path closure to form a truly knotted protein may affect folding stability and kinetics, we assessed the thermal stabilities of *l*YibK and *c*YibK by far-UV circular dichroism (CD) spectroscopy and their chemical stabilities by urea-induced chemical denaturation monitored by intrinsic fluorescence. As expected, the apparent melting temperature (T_m) of *l*YibK was increased by 20 °C for *c*YibK (T_m = 68.7 vs 48.7 °C; **Figure 2B**). This observation is in line with other proteins with cyclized peptide backbones, suggesting that backbone cyclization of a knotted protein reduced the conformational entropy of the unfolded state. As the thermal unfolding was not fully reversible, particularly in the case of *l*YibK, the T_m values derived from the CD analysis could be underestimated. We additionally analyzed the urea-induced equilibrium unfolding of *l*YibK and *c*YibK (**Figures 2C,D**). However, we observed the unexpected loss of intrinsic fluorescence of *l*YibK between 2 and 3 M urea, which we attributed to the aggregation of *l*YibK in the analysis **(Supplementary Figure S4a)**. In contrast, *c*YibK did not show the similar loss of intrinsic fluorescence during urea-induced denaturation (**Supplementary Figure S4b**). The experimental data were fit to a three-state unfolding model by the singular value decomposition approach without considering the contributions of dimerization (Wang et al., 2015; Wang et al., 2016). Although the chemical denaturation of *l*YibK was not fully reversible with *l*YibK, the chemical stability of *c*YibK was clearly higher than *l*YibK by >2 M of the transition urea concentration required to unfold *l*YibK and *c*YibK, supporting the increased apparent thermal stability of *c*YibK (**Supplementary Table S3**; **Figures 2C,D**).

Additionally, we analyzed the folding kinetics of *l*YibK and *c*YibK as a function of urea concentration by monitoring the intrinsic fluorescence of the only endogenous tryptophan residue (W150 according to the nomenclature of *l*YibK construct) lining the dimer interface (**Figure 3**). It is noteworthy that the well-investigated *Hi*YibK contains two tryptophan residues, of which W145 is positioned at the same dimer interface as W150 of *Pa*YibK (**Supplementary Figure S4**). Similar to the reported multiphasic kinetics of *Hi*YibK, *l*YibK also exhibited two unfolding and refolding phases; the faster phase had a very small *m*-value associated with the unfolding arm (**Figure 3**).

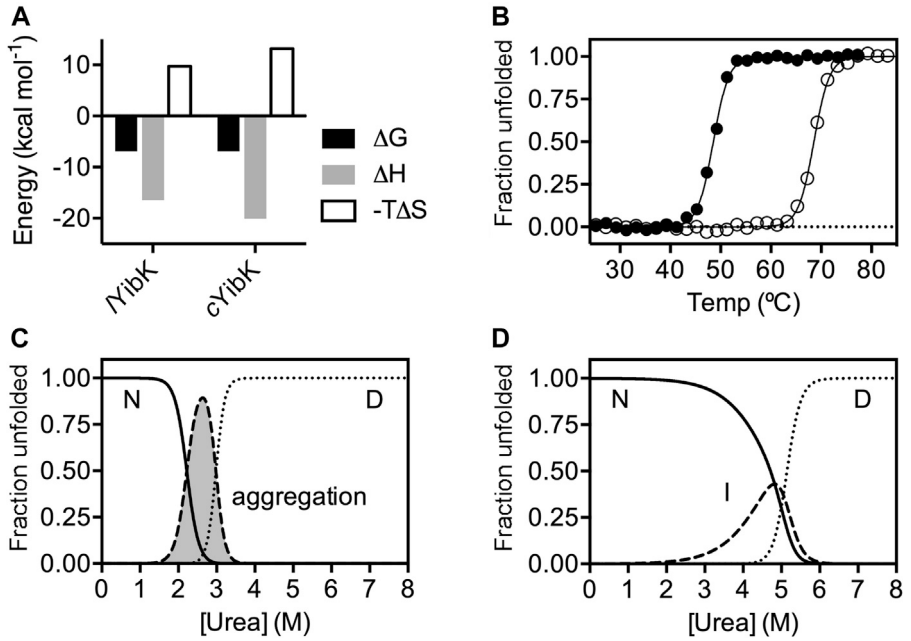


FIGURE 2 | The impacts of cyclization on SAH binding and chemical stability of YibK. **(A)** Thermodynamics parameters associated with SAH binding derived from ITC analyses. **(B)** Normalized fractional unfolded populations of *l*YibK (filled circles) and *c*YibK (open circles) derived from thermal denaturation monitored by far-UV CD spectroscopy. Intrinsic fluorescence-derived normalized fractional populations of native (N), intermediate (I) and denatured (D) states of *l*YibK **(C)** and *c*YibK **(D)** as a function of urea concentration. For *l*YibK, fluorescence signal loss due to aggregation was observed and the corresponding population is shaded in gray and indicated.

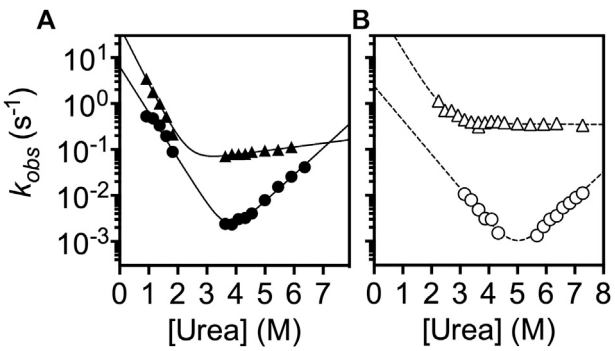


FIGURE 3 | Chevron plot analysis of the folding kinetics of *l*YibK and *c*YibK. The observed folding rates of *l*YibK **(A)** and *c*YibK **(B)** are plotted as a function of urea concentration. Two kinetic phases were observed. Circles and triangles correspond to the slow and fast kinetic phases, respectively. Filled and open symbols are used for *l*YibK and *c*YibK, respectively. The data were fitted to a simple two-state folding model.

The slower intrinsic unfolding rate of *l*YibK ($k_u^{H_2O}$) was 9.5×10^{-6} sec^{-1}, almost 20-fold faster than that of *Hi*YibK ($k_u^{H_2O} = 4.9 \times 10^{-7}$ sec^{-1}) (Mallam et al., 2008). The faster unfolding rate of *l*YibK is presumably associated with the aggregation of *l*YibK we observed. The slower intrinsic unfolding phase of *c*YibK ($k_u^{H_2O} = 6.5 \times 10^{-7}$ sec^{-1}) was about 7-fold slower than that of *l*YibK (**Table 1**). The transition urea concentration $[D]_{50\%}$, associated with the slower kinetic phases of *l*YibK and *c*YibK agreed well with the second transition points derived from equilibrium unfolding (**Figure 2**), suggesting that these kinetic phases are associated with the intermediate-to-denatured state transitions. Consequently, the faster kinetic phases of both *l*YibK and *c*YibK would correspond to the intermediate-to-native state transition. Note that the β-Tanford values (β_T), reporting on the compactness of the transition state with respect to the folded state, were close to 1 for the faster kinetic phase of both *l*YibK and *c*YibK. Thus, the associated transition states (from intermediate to native state; TS_{I-N}) could be as compact as the native state (**Supplementary Table S3**). (Fersht, 1999) In contrast, the β_T values of the slower kinetics of *c*YibK were significantly lower (ca. 0.6) for both *l*YibK and *c*YibK, so the corresponding transition state (from denatured to intermediate state; TS_{D-I}) is highly disordered (**Supplementary Table S3**). Collectively, our kinetic analyses suggest that the intermediate formation is the rate-limiting state, which is consistent with the equilibrium-unfolding analysis finding that the intermediate state of *c*YibK was lowly-populated and that the intermediate state of *l*YibK was aggregation-prone.

Comparison Between *l*YibK and *c*YibK Under a Denaturing Condition by NMR

Considering that the chemical compositions (sequences) and 3D structures of *c*YibK and *l*YibK essentially are identical with the exception of a peptide bond introduced, we presumed that the native states of both YibKs have similar free energies. We assume

TABLE 1 | Kinetic parameters derived from chevron plot analysis of *l*YibK and *c*YibK.

	Kinetic phase	$k_f^{H_2O}$ (s^{-1})	m_f (kcal mol^{-1} M^{-1})	$k_u^{H_2O}$ (s^{-1})	m_u (kcal mol^{-1} M^{-1})	m_{kin}[a] (kcal mol^{-1} M^{-1})	β_T[b]	$[D]_{50\%,kin}$ (M)[c]	ΔG_{kin} (kcal mol^{-1})[c]
*l*YibK	Fast	47.6 ± 3.7	−2.87 ± 0.06	0.036 ± 0.002	0.19 ± 0.01	3.06 ± 0.06	0.94 ± 0.05	2.35 ± 0.06	4.25 ± 0.06
	Slow	6.42 ± 1.28	−2.30 ± 0.13	(9.50 ± 4.64) × 10^{-6}	1.33 ± 0.09	3.63 ± 0.16	0.63 ± 0.06	3.70 ± 0.26	7.94 ± 0.30
*c*YibK	Fast	124 ± 129	−2.28 ± 0.45	0.360 ± 0.063	~0[d]	2.28 ± 0.45[e]	1.00[e]	2.56 ± 0.69[e]	3.46 ± 0.62[e]
	Slow	2.41 ± 1.13	−1.71 ± 0.13	(6.55 ± 3.60) × 10^{-7}	1.35 ± 0.08	3.06 ± 0.15	0.56 ± 0.05	4.94 ± 0.34	8.95 ± 0.43

[a] $M_{kin} = m_u - m_f$.
[b] $\beta_T = -m_f/m_{kin}$.
[c] *The transition points ($[D]_{50\%,kin}$) and free energies of unfolding (ΔG_{kin}) were derived from the kinetic parameters associated with the fast and slow phases.*
[d] *The unfolding arm of the slow kinetic phase of lYibK showed no apparent denaturant concentration-dependency.*
[e] *The results were derived by setting m_u to zero.*

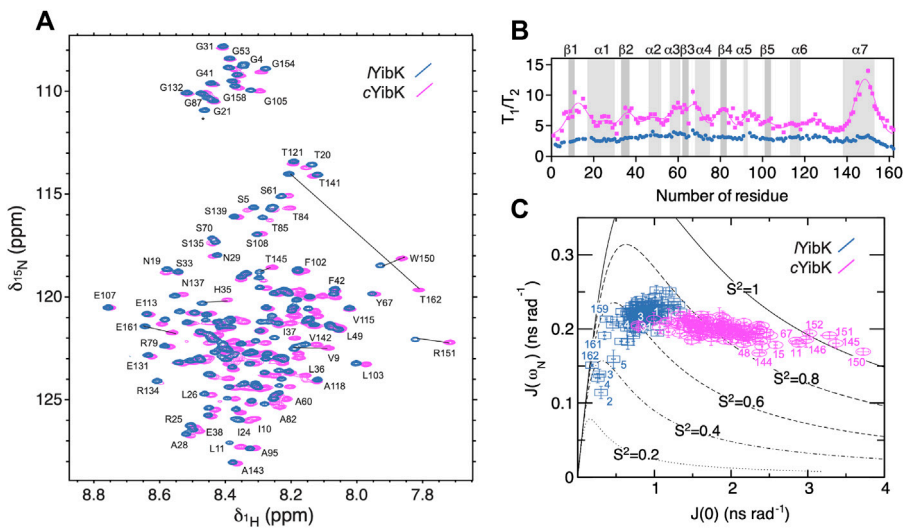


FIGURE 4 | NMR spectroscopy and ^{15}N spin relaxation analysis of the chemically denatured *l*YibK and *c*YibK. **(A)** Superimposition of the [^{15}N-^{1}H] correlation spectra of *l*YibK (orchid blue) and *c*YibK (magenta) in the presence of 7.2 M urea. The spectra were recorded at a ^{1}H Larmor frequency of 850 MHz, and 298 K. Residues with large chemical shift differences upon cyclization are indicated by solid lines that connect the pairs of crosspeaks. **(B)** T_1/T_2 ratios of *l*YibK and *c*YibK as a function of residue number. T_1/T_2 ratios of *c*YibK were fitted to a sum of multiple Gaussian distributions. Regions that correspond to the β-sheets and α-helices in the native structure are highlighted by dark and light gray, respectively, and are indicated above the panel. **(C)** Spectral density mapping of the ^{15}N relaxation data expressed as $J(\omega_N)$ as a function of $J(0)$. Residues that deviate from the cluster distributions are indicated with their residue numbers. Theoretical curves with the assumption of isotropic motions were calculated for different-order parameters, S (Liu et al., 1980), as indicated. All data points are colored using the same scheme as in (**A**).

that the observed changes in the folding/unfolding pathway of YibK could be caused by the unfolded state of *c*YibK, which has a significantly reduced conformational space compared with a linear polypeptide due to the circular backbone peptide chain having closed ends and the presence of a knot structure.

As NMR spectroscopy can investigate proteins under various solution conditions including denaturating ones, we used NMR to characterize the unfolded states into both *l*YibK and *c*YibK in 7.2 M urea to gain structural insights into the denatured states of *l*YibK and *c*YibK (**Figure 4**). The two-dimensional [^{15}N-^{1}H] correlation spectra for both *l*YibK and *c*YibK in 7.2 M urea showed poor chemical shift dispersions along the ^{1}H dimension, characteristic of unfolded and disordered polypeptides (**Figure 4A**). Furthermore, a large number of crosspeaks exhibited major chemical shift differences between the two [^{15}N-^{1}H] SOFAST-HMQC spectra (**Figure 4A**). We could obtain near-complete site-specific NMR assignments of the observed [^{15}N-^{1}H] correlations of *l*YibK and *c*YibK in 7.2 M using a described protocol (Hsu et al., 2009; Hsieh et al., 2014). As expected, the backbone amide of T162 at the C-terminus of *l*YibK showed the largest chemical shift difference from that of *c*YibK owing to the introduction of an additional peptide bond as a result of backbone cyclization. Furthermore, several residues near the N- and C-termini also exhibited significant chemical shift perturbations (**Figure 4A**; **Supplementary Figure S6**).

NMR chemical shift analysis suggested very limited structural differences between urea-denatured *l*YibK and *c*YibK (**Supplementary Figure S6**). Therefore, we performed ^{15}N relaxation analysis to compare their backbone dynamics. Under the denaturing condition, the mean ^{15}N transverse relaxation time (T_2) for *c*YibK was much shorter than that of *l*YibK in 7.2 M (122 vs. 238 ms; **Supplementary Figure S7**). We

also observed that *l*YibK exhibited longer and more uniform T_2 values across the primary sequence with the exception of the fraying ends because of the unrestricted chain dynamics at the open ends **Supplementary Figure S7C**). In contrast, *c*YibK exhibited clusters of fast-relaxing residues across the primary sequence (**Supplementary Figure S7C**), reminiscent of the previously observed long-range non-native interactions in urea-denatured lysozyme (Klein-Seetharaman et al., 2002). For urea-denatured lysozyme, the non-native interactions reflected the clustering of bulky hydrophobic tryptophan residues. However, for *c*YibK, not all rapidly relaxing residues contain aromatic side-chains. The T_2 relaxation clusters may stem from the internal friction imposed by backbone cyclization that restricted the fraying motions of the N- and C-termini. Differences in dynamics or populations were particularly manifested in the dispersed T_1/T_2 values for *c*YibK, whereas a flat profile of a T_1/T_2 values was observed for *l*YibK. The T_1/T_2 values indicated that the differences of backbone dynamics are in the timescale of micro-to milliseconds. In contrast, the heteronuclear $^{15}N\{^1H\}$-NOE of *l*YibK and *c*YibK, which is sensitive to a faster pico-to-nanosecond timescale, did not show apparent differences except for the termini of *l*YibK (**Supplementary Figure S7c**). Thus, the backbone cyclization did not affect the pico-to-nanosecond motions of the individual backbone peptide bonds except for the termini with the conformation restriction imposed by backbone cyclization (**Supplementary Figure S7**).

We also applied the reduced spectral density mapping approach for studying the dynamics of both denatured *l*YibK and *c*YibK (**Figure 4C**). (Farrow et al., 1995; Peng and Wagner, 1995; Shih et al., 2015) The results identified two distinct clusters of spectral density distributions for urea-denatured *l*YibK and *c*YibK, the former located around the theoretical curve for an order parameter *S* (Liu et al., 1980) of <0.7 and the latter exhibiting a cluster around the *S* (Liu et al., 1980) value of 0.8 (**Figure 4C**). Furthermore, several residues located at the C-terminal helix (α7) and some others were located outside the cluster distribution. Collectively, the NMR relaxation dynamics analysis suggested the presence of abundant conformational exchanges. It also implied a broad range of backbone dynamics caused by the backbone cyclization of YibK, restricting the backbone motions and increasing the ruggedness of the free energy landscape of the denatured state of *c*YibK.

*c*YibK_Int Under the Denaturing Condition by NMR

The split-intein approach for cyclization of YibK resulted in the insoluble spliced product that could presumably be the cyclized YibK without knotting (*c*YibK_Int) (**Figure 1C**). Even though *c*YibK_Int has an additional hexahistidine-tag and slightly different amino-acid sequence connecting the N- and C-termini (**Figure 1E**), we decided to purify and investigate *c*YibK_Int under a denaturing condition by NMR spectroscopy (**Supplementary Figure S8**). $^{15}N\{^1H\}$-NOE data for the N- and C-termini of *c*YibK_Int are similar to those of *c*YibK than *l*YibK bearing flexible termini due to the linear polypeptide chain. This observation confirms the backbone cyclization of *c*YibK_Int. The average T_2 relaxation time for *c*YibK_Int was shorter than that of *l*YibK (157 vs. 238 ms) (**Supplementary Figure S8c**). On the other hand, a flat profile of T_1/T_2 values for *c*YibK_Int is closer to *l*YibK, indicating the absence of conformational exchanges observed with *c*YibK, presumably because of the absence of a knot structure under the denaturing condition (**Supplementary Figure S8C**).

Small-Angle X-Ray Scattering Analysis of the Unfolded States of *l*YibK and *c*YibK

We previously used SAXS to demonstrate that chemically denatured knotted proteins with open ends exhibited random coil-like behaviors: their radii of gyration (R_g) scale with their chain lengths to the power of 3/5 (Shih et al., 2015). This observation suggests that backbone knotting with open ends does not necessarily lead to significant compaction of the overall chain dimension under highly denaturing conditions (in good solvent). However, these knotted proteins examined by SAXS are not mathematically knotted because of their free termini. To examine how backbone cyclization may affect the polymer properties of a mathematically knotted protein, we compared the SAXS data of *l*YibK and *c*YibK by using online size-exclusion chromatography-coupled SAXS (SEC-SAXS) apparatus as described (Lee and Hsu, 2018). Under native conditions, *l*YibK and *c*YibK exhibited the same SAXS profiles, with comparable R_g values — 22.17 ± 0.04 and 21.66 ± 0.06 Å—that were in general agreement with the theoretical value based on the crystal structure (19.4 Å; **Figure 5**). The corresponding Kratky plots showed comparable compactness (similar bell-shape profiles; *cf.* black and gray curves for *c*YibK and *l*YibK in **Figure 5B**, respectively). In contrast, 7.2 M urea-denatured *c*YibK showed a significantly smaller R_g value than *l*YibK under the same condition (27.71 ± 0.15 vs 39.26 ± 0.38 Å; **Figures 5A,C**). Urea-denatured *l*YibK exhibited a monotonously increasing Kratky profile typical of a random coil polypeptide, whereas urea-denatured *c*YibK exhibited a distinct bell-shape profile, albeit much smaller than that of folded *c*YibK, which indicates the presence of compact residual structure (blue curve in **Figure 5B**). According to the empirical scaling relationship established from our previous study, the expected R_g value for a chemically denatured *l*YibK is 38.0 Å, assuming a random coil-like behavior (**Figure 5C**). (Shih et al., 2015) A reduction of R_g by >10 Å in chemically denatured *c*YibK equals a decrease of the global dimension by 1/4 and a 3/5 decrease of the excluded volume (assuming that the exclusion volume of the unfolded polypeptide chain is spherical). The substantial conformational compaction further suggests that the conformational entropy of the unfolded state is reduced as a result of the backbone cyclization, which is in line with the stability improvement of *c*YibK.

DISCUSSION

Introduction of a peptide bond between N- and C-termini into a knotted protein unambiguously converts it to a truly mathematical knot without the need to evoke convoluted

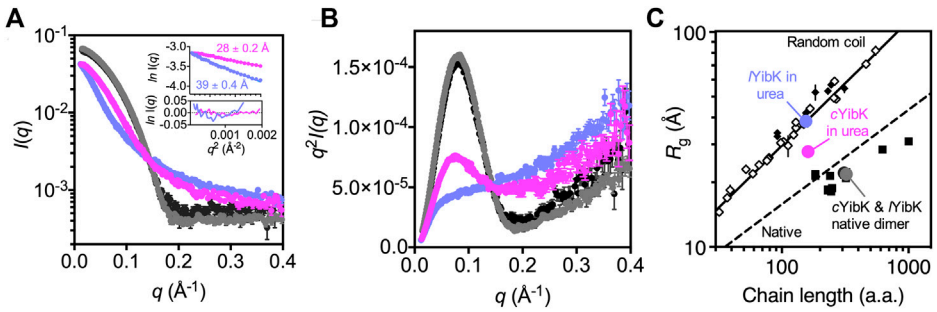


FIGURE 5 | Cyclization significantly reduces the global dimension of the denatured state of *c*YibK. **(A)** SAXS profiles of *l*YibK in 0 M urea (gray), *c*YibK in 0 M urea (black), *l*YibK in 7 M urea (orchid blue), and *c*YibK in 7 M urea (magenta). Inset: Guinier plots of *l*YibK and *c*YibK in 7 M urea with the corresponding R_g values indicated and fitting residues shown below. **(B)** Kratky plots of *l*YibK and *c*YibK in their native and 7 M urea-denatured states, as shown in (a). Urea-denatured *c*YibK exhibits a distinct bell-shape profile, indicative of compact residual structures that are absent in urea-denatured *l*YibK. **(C)** R_g values of *l*YibK and *c*YibK in their native and 7 M urea-denatured states as a function of chain length. The solid and dashed lines indicate the experimentally derived scaling relationship for random-coil and native proteins, respectively. Solid squares and diamonds correspond to reported R_g values of knotted proteins of different topologies and chain lengths. Open diamonds correspond to reported random-coil R_g values of chemically denatured unknotted proteins (Shih et al., 2015).

knot-detecting algorithms, which in some cases have different interpretations of "knots" in proteins. It also eliminates any confusing experimental effects due to fraying of the N- and C-termini in proteins that are "knotted" by the introduction of disulfide bridges, which are largely different due to higher rotational degrees in the side-chains (Betz, 1993; Mallam et al., 2010). Whereas disulfide bonds have a mixture of enthalpic and entropic effects on the protein stability, backbone cyclization connecting the N- and C-termini is generally accepted to stabilize protein by destabilizing the unfolded state (Betz, 1993; Iwai and Plückthun, 1999; Scott et al., 1999; Clark and Craik, 2010; Montalbán-López et al., 2012; Borra and Camarero, 2013).

Here, we produced a trefoil-knotted protein without open ends, a truly mathematical backbone knot in a protein, by post-translational modification by enzymatic ligation (**Figure 1**). Such a backbone modification is irreversible as opposed to disulfide crosslinking between a pair of engineered cysteines at the N- and C-termini. Furthermore, the path closure by a backbone peptide bond essentially removes the origin of the protein sequence, rendering obsolete the conventional definition of a protein folding topology by the hierarchical arrangements of secondary structure elements. As compared with circular permutation, which was recently used to untie the trefoil-knotted *Hi*YibK (Chuang et al., 2019), and *E. coli* YbeA (Ko et al., 2019), SrtA-mediated backbone cyclization allowed us to examine the contribution of a true backbone knot from a completely different perspective with clarity. The apparent melting temperature was increased by 20 °C for the knotted *c*YibK compared with *l*YibK with open ends (**Figure 2**). The path closure by a peptide bond after folding also has seemingly remodeled the protein folding/unfolding pathway of the original *l*YibK and alleviated the aggregation propensity of the folding intermediate observed for *l*YibK, whereas maintaining the native structure and ligand binding affinity (**Figure 2**; **Supplementary Table S2**). Indeed, the backbone cyclization significantly increased the folding rate of the intermediate-to-native state transition, with the corresponding transition state being highly compact and native-like, as evidenced by the β_T value being close to 1 (**Figure 3**; **Supplementary Table S3**). Furthermore, the unfolding rate of intermediate-to-denatured state transition of *c*YibK, $k_u^{H_2O}$ (derived from the slower kinetic phase; **Figure 3**) was >10 times slower than that of *l*YibK, so the denatured state of *c*YibK may have a status with higher Gibbs energy than that of *l*YibK, which could be supported by R_g in 7 M urea estimated by the SAXS data. Note that we have not unambiguously established whether *l*YibK is unknotted or not under urea-denatured state as has been demonstrated earlier (Mallam et al., 2008; Capraro and Jennings, 2016). It is, therefore, possible that the denatured *l*YibK may exist in a mixture of knotted and unknotted structures.

To this end, ^{15}N spin relaxation analysis of *l*YibK showed no appreciable conformational exchange contributions to the three different timescales probed by $J(0)$, $J(\omega_H)$, and $J(\omega_N)$, suggesting the absence of interconversion between knotted and unknotted states of *l*YibK (**Supplementary Figure S7**). *c*YibK_Int, which has closed peptide ends and possibly no knot structure, showed a similar profile of T_1/T_2 to that of *l*YibK in the ^{15}N spin relaxation analysis. In contrast, the enhanced T_2 relaxation observed in *c*YibK, which has significant contributions to the $J(0)$ term, likely reflects the increased internal friction of the cyclized polypeptide chain in the denatured state, thereby leading to destabilization of the unfolded state (**Figure 4**). As observed for many backbone cyclized proteins, we think that the increase in folding stability of *c*YibK could be attributed mainly to the reduced conformational entropy in the denatured state of *c*YibK.

In line with the stability enhancement, the chemically denatured state of *c*YibK was significantly more compact than the random coil-like *l*YibK in 7 M urea (**Figure 5**). The effect of cyclization on the polymer dimension is well understood in the literature. Kramers' polymer model predicts that the R_g values of linear and cyclized polymers follow a simple relationship.

$$\frac{\langle R_g^c \rangle^2}{\langle R_g^l \rangle^2} = \frac{1}{2}$$

where $\langle R_g^c \rangle$ and $\langle R_g^l \rangle$ are the mean R_g values of the cyclized and linear forms of the same polymer, respectively, (Kramers, 1946). Our SAXS analysis of *c*YibK and *l*YibK yielded a ratio of 0.50 ± 0.01, which is in good agreement with Kramers' polymer theory despite the highly heterogeneous amino acid side-chain compositions (Kramers, 1946). The SAXS data implied that a knot formation does not provide any further compactness under a denaturing condition by having a true knot structure after backbone cyclization. In other words, there might be no or little entropic penalty for knotting under highly denaturing conditions due to the high flexibility, suggesting that the stability enhancement of *c*YibK could be attributed mainly to backbone cyclization without additional contribution from the knot structure. The SAXS analysis suggested that the cyclized and knotted YibK under a highly denaturing condition appears to comply with the polymer physics developed for non-self-interacting Gaussian chains, behaving like a long thin string. If this is true for unfolded proteins without denaturants (i.e., intrinsically disordered proteins), some proteins without defined secondary structures could possibly entangle into open knots at a certain probability, as observed for simpler homopolymers (Higgins et al., 1979; Arrighi et al., 2004). An increasing number of intrinsically disordered proteins without fixed conformations have been identified and implicated in many diseases. Our results suggest that protein backbone knots could also be transiently formed without any entropic penalty when polypeptide chains are very flexible, as in the denatured states. Physico-chemical characterizations of even simpler polymers with a well-defined mathematical knot have not been investigated because isolating defined simple polymers with a specific mathematical knot is very challenging. Unlike other simpler polymers, the self-entanglement of proteins into defined knots could be exploited to isolate well-defined mathematical knots for further physicochemical characterizations. The post-translational enzymatic backbone cyclization, as well as the split-intein approach we demonstrated here, could pave the way to investigate other proteins with various knot topologies, which may include transiently formed protein knots, for example, with intrinsically disordered proteins.

MATERIALS AND METHODS

Constructions and Production of Recombinant YibK Variants

For backbone cyclization, *P. aeruginosa* YibK with a sortase recognition sequence LPETG followed by the C-terminal hexa-histidine (H_6) was cloned in pRSF vector as N-terminal SUMO fusion by PCR, resulting in pITRSF1A (*Pa*YibK_sh) and pITRSF3D (*Pa*YibK_lo). *In vivo* cyclization vector for *Pa*YibK was cloned into a pBAD vector containing the genes of split *Npu*DnaE-C intein fragment, H_6-tag, and *Npu*DnaE-N intein fragment by using *Xba*I/*Kpn*I sites, resulting in pJMBAD36(*Pa*YibK_Int) (Iwai et al., 2006). For biophysical characterization of linear *Pa*YibK_lo, plasmid pBHRSF260 was constructed with N-terminally His-tagged SUMO fusion to have the identical sequence as *Pa*YibK_lo construct (Guerrero et al., 2015). Each plasmid was transformed into *E. coli* strain ER2566 cells (New England Biolabs, Ipswich, United States). The cells were cultured in Luria-Bertani medium supplemented with kanamycin at 25 μg·ml^{-1} until OD_{600nm} reached 0.6 at 37°C. The recombinant protein overexpression was induced for 4 h with a final concentration of 1 mM IPTG. The cells were harvested by centrifugation and resuspended in binding buffer (50 mM sodium phosphate buffer, pH 8.0, and 300 mM NaCl). The resuspended cells were lyzed at 15,000 psi for 10 min by using Emulsiflex C3, and the supernatant was separated from cell debris by 1 h centrifugation at 38,465 *g*. The supernatant was loaded onto a pre-packed HisTrap HP column (GE Healthcare Life Sciences, United States). The His-tagged fusion proteins were eluted by a linear gradient of 50–250 mM imidazole and dialyzed against 2 L of 50 mM Tris-HCl buffer, pH 7.5, 0.5 mM EDTA, and 0.5 mM DTT overnight at 8°C (Guerrero et al., 2015). The fusion proteins were digested by Ulp1 protease as described previsouly (Guerrero et al., 2015). The digested fusion proteins were loaded again on pre-equibriated HisTrap HP column and washed to remove the SUMO-tag. The C-terminally His-tagged YibK were eluted by a linear gradient of 50–250 mM imidazole and dialyzed against 50 mM Tris-HCl buffer, pH 7.5 overnight, followed by concentration with a centrifugal device. *c*YibK by the split intein fusion (*c*YibK_Int) was produced from the plasmid pJMBAD36. *E. coli* strain ER2566 cells bearing pJMBAD36 were grown in 2 L of M9 medium supplemented with ampicilin at 100 μg·mL^{-1} at 37°C and induced for 4 h with a final concentration of 0.02% (w/v) arabinose. The cells were harvested and lyzed at 15,000 psi for 10 min using Emulsiflex C3. The insoluble pellet was collected after discarding the supernatant by 1 h centrifugation at 38,465 *g*. The pellet was resolubilized in 25 ml of 8M urea with shaking at 350 rpm overnight. The dissolved solution was cleared by 1 h centrifugation at 38,465 *g*. The supernatant was loaded onto the HisTrap HP column, which was pre-equilibrated with a binding buffer (100 mM sodium phosphate buffer, pH 8.0, 10 mM Tris, and 8 M urea). The His-tagged cyclized YibK (*c*YibK_Int) was eluted by 100 mM sodium phosphate buffer, pH 5.0, 10 mM Tris, and 8 M urea. The precursor protein in the elution fractions was removed by size-exclusion chromatography with a Superdex 75 16/60 column (GE Healthcare, United States) in 20 mM sodium phosphate buffer, pH 5.0, 8M urea. The fractions containing *c*YibK_Int were pooled and concentrated for NMR analysis.

SrtA-Mediated Backbone Cyclization

For the backbone cyclization, sortase (SrtA, from *Staphylococcus aureus*) was added to purified YibK in 1–5 molar ratio and dialyzed against 50 mM Tris-HCl buffer (pH 7.5), 10 mM $CaCl_2$ and 2 mM DTT, for 20 h at room temperature. Finally, the unreacted YibK that contained the His_6-tag at the C-terminus was separated from *c*YibK by incubation with Ni-NTA resin in an open column. The *c*YibK collected from the flow-through fractions was further polished by size-exclusion chromatography with a Superdex 75 16/60 column (GE Healthcare, United States) in buffer A (50 mM Tris-HCl (pH 7.4), 0.5 mM EDTA and 5 mM DTT).

Protein Crystallography

For the crystal structure of *Pa*YibK, the plasmid (pJMRSF13) encoding YibK gene with N-terminal His-tag and the SUMO fusion was produced and purified as described previously (Guerrero et al., 2015). Crystallization was performed with 9.2 mg/ml solution of *Pa*YibK and 11 mg/ml solution of *c*YibK. Drops of 200 nl (100 nl protein solution and 100 nl well solution) were placed in 96-well MRC (Molecular Dimensions) crystallization plates using a Mosquito LCP (TTPLabtech, United Kingdom). Initial hits were obtained from the traditional sparse matrix screens with the local modifications (Cudney et al., 1994). The initial hits were further optimized by grid screening. The final growth conditions for diffracting crystals were 0.15 M ammonium sulfate, 0.9 M lithium sulfate, 0.1 M sodium citrate buffer (pH 5.6) for *Pa*YibK, and 0.3 M ammonium sulfate, 0.1 M MES buffer (pH 6.0), 25% polyethylene glycol monomethyl ether (PEG MME) 5,000 for *c*YibK. 20% glycerol was added on top of the drop of *Pa*YibK for cryoprotection prior to flash-freezing crystals in liquid nitrogen. For *c*YibK, the 25% PEG MME 5000 present in the crystallization drop served as a sufficient cryoprotectant. Diffraction data for the crystals of *Pa*YibK and *c*YibK were collected at the beamline ESRF ID14–4, Grenoble, France and I03 at the Diamond Light Source, Oxfordshire, UK, respectively. The diffraction data were then indexed, integrated, and scaled to 2.0 and 2.2 Å resolution for *Pa*YibK and *c*YibK, respectively, in XDS (Kabsch, 2010). The final crystal parameters and data processing statistics are in **Supplementary Table S1**. The structures of *Pa*YibK and *c*YibK were solved by molecular replacement with MolRep from the CCP4 package (Winn et al., 2011). The structure of *Hi*YibK (PDB ID: 1mxi) was used as a search model for molecular replacement. The model was then built using Coot, followed by rounds of refinement with Refmac5 from the CCP4 package and Phenix (Adams et al., 2010). The final refinement was performed with Phenix, and the quality of the final model was validated by using MolProbity (**Supplementary Table S1**). (Chen et al., 2010) The final refined model of *Pa*YibK was used as a starting model for the molecular replacement to solve the structure of *c*YibK. The structure was solved, refined and validated as mentioned above (**Supplementary Table S1**). The final coordinates were deposited in the Protein Data Bank (PDB) with the accession codes 6qkv and 6qh8 for *Pa*YibK and *c*YibK, respectively.

Chemical Denaturation Monitored by Intrinsic Fluorescence Spectroscopy

Urea-induced equilibrium unfolding of *l*YibK and *c*YibK was monitored by intrinsic fluorescence as described (Wang et al., 2015; Wang et al., 2016). Briefly, 41 aliquots of protein solution (at a final concentration of 2 μM buffered in buffer A) were prepared in a series of urea concentrations (0–7 M) with a linear increment step of 2.5% generated by a two-channel liquid syringe dispenser (Hamilton, United States). The samples were incubated at 25 °C overnight before fluorescence measurements with a fluorimeter (JASCO FP8500, Japan). The samples were excited at 280 nm and emission spectra between 300 and 450 nm were collected. The results underwent singular value decomposition analysis with MatLab (MATLAB and Statistics Toolbox release 2012b; The MathWorks, United States) to determine the number of states associated with the unfolding processes, followed by fitting to a three-state folding equilibrium model with Prism (GraphPad, United States) as described (Wang et al., 2014; Wang et al., 2015; Wang et al., 2016; Lee et al., 2017).

Thermal Denaturation Monitored by Far-UV CD Spectroscopy

The protein solutions were diluted to 10–15 μM in buffer A with a total volume of 0.3 ml, and transferred into a 1 mm path-length quartz cuvette (Hellma, Germany) for far-UV CD measurements. The CD signals between 195 and 260 nm were collected as a function of temperature between 25 and 80 °C with an interval of 2°C by using a CD spectrometer (J-815, JASCO, Japan). The spectra bandwidth was set to 1 nm with a data interval of 0.5 nm, and an averaging time of 1 s. The melting temperatures (T_m) of *l*YibK and *c*YibK were derived by global-fitting the CD spectra as a function of temperature to a two-state model as described (Wang et al., 2014; Lee et al., 2017).

Isothermal Titration Calorimetry

ITC analysis of SAH binding to *l*YibK and *c*YibK was monitored by using MicroCal VP-ITC (Malvern, United Kingdom) as described (Zhao et al., 2015). Stock solutions of YibK variants were dialyzed overnight against buffer B (50 mM Tris-HCl (pH 7.4), 0.5 mM EDTA, and 0.1 mM TCEP) to remove DTT before ITC measurements. The dialysis buffer was used to prepare the stock solution of the titrant, SAH (Sigma-Aldrich, United States) at a concentration of 0.5 mM. An amount of 20 μM *l*YibK or 13 μM *c*YibK was used in the sample cells for ITC measurements. The resulting isotherms were processed by using NITPIC followed by data fitting with SEDPHAT (Zhao et al., 2015).

Folding Kinetics Monitored by Intrinsic Fluorescence Spectroscopy

Chevron plot analyses of the folding kinetics of *l*YibK and *c*YibK involved using a combination of stopped-flow and manual mixing measurements as described (Wang et al., 2015; Wang et al., 2016). Briefly, 10 μM native or 7 M urea-denatured protein stock solution was mixed with 10-fold excess denaturing or refolding buffer (buffer A with different concentrations of urea) and the kinetic traces of total fluorescence emission excited at 280 nm and cutoff by a 320 nm cutoff filter were fit to a linear combination of 2–4 exponential functions depending on the experimental conditions. For the slowest refolding rate of *c*YibK, manual mixing of 7 M urea-denatured *c*YibK with refolding buffer at a 1:10 mixing ratio was performed before fluorescence measurement (excitation 280 nm and emission 325 nm with a bandwidth of 5 nm) with a fluorimeter (JASCO FP8500, Japan).

Small-Angle X-Ray Scattering

SEC-SAXS experiments were performed on beamline BL23A at the National Synchrotron Radiation Research Center (NSRRC, Hsinchu, Taiwan) with the capacity to separate aggregated particles on a silica-based size-exclusion column (Bio SEC-3, Agilent, United States). SAXS signals were detected by using a Pilatus detector (1M-F) and processed by an in-house developed program to obtain the SAXS profiles (Kohn et al., 2004; Jeng et al., 2010; Lee and Hsu, 2018). The SAXS data were collected for momentum transfer q ranging from 0.005 to 0.434 $Å^{-1}$, with X-ray wavelength 1.03 Å and 13 keV. The beam geometry was set to 0.5 × 0.5 mm^2. During the HPLC separation before SAXS measurements, the mobile phase consisted of buffer A (with and without 7 M urea) with the addition of 2% glycerol to prevent radiation damage. The protein solutions were concentrated to 10 mg/ml with the same mobile phase buffer immediately before SAXS measurements.

Nuclear Magnetic Resonance Spectroscopy

Uniformly ^{15}N-labeled and 20% ^{13}C-labeled *c*YibK and *l*YibK were prepared by using M9 medium containing 1 g/L [^{15}N] ammonium chloride as a nitrogen source and a mixture of 0.6 g/L [^{13}C] D-glucose and 2.4 g/L [U-^{12}C] D-glucose as a carbon source, as described (Iwai and Fiaux, 2007; Heikkinen et al., 2021). The NMR samples were fully denatured by 7.2 M urea in buffer A containing 10% D_2O (v/v) at a protein concentration ca. 0.3 mM. A suite of triple resonance experiments in addition to the [^{15}N-^{1}H] band-selective optimized flip-angle short transient heteronuclear multi-quantum correlation (SOFAST-HMQC) were recorded at 298 K on 850 MHz NMR spectrometer equipped with a cryogenic triple resonance probe (Bruker, Germany) for backbone resonance assignments following the strategy described previously (Hsu et al., 2009; Hsieh et al., 2014). Near-complete backbone resonance assignments (H^N, N, C', Cα, Cβ, and Hα) were achieved for both *c*YibK and *l*YibK. The assignments were deposited in the Biological Magnetic Resonance Bank (BMRB) under accession numbers 27685 and 27686 for *c*YibK and *l*YibK, respectively. ^{15}N spin relaxation NMR measurements for longitudinal (R_1) and transverse (R_2) relaxation rates and the heteronuclear Overhauser effect (hetNOE) were as described (Hsu et al., 2009). Eight longitudinal relaxation delays (20, 60, 120, 200, 300, 500, 800, and 1,000 ms) and nine transverse relaxation delays (16, 32, 64, 96, 128, 160, 192, 224, and 256 ms) were used for both *c*YibK and *l*YibK, and the data were collected as pseudo-3D spectra with the relaxation delays incremented in an interleaved manner to minimize heating effects. The R_1 and R_2 rates were extracted by fitting the peak intensities of the individual residues in the ^{15}N-^{1}H correlation spectra to a single exponential decay function by using the relaxation analysis module in Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, United States).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://www.wwpdb.org/, 6qkv, 6qh8; https://bmrb.io/, 27685, 27686.

AUTHOR CONTRIBUTIONS

HI and STDH conceptualized and planned the project. IT, KM, YL established and produced the recombinant proteins. KM and IT crystallized proteins and solved the structures. HI, SB, YL, and STDH recorded and analyzed NMR data. SAXS and biophysical characterizations were performed and analyzed by YL and STDH. HI, AG, and STDH wrote the manuscript with input from all the authors.

FUNDING

This work was supported by the Ministry of Science and Technology, Taiwan (105-2113-M-001-005, 106-2113-M-001-004 and 107-2628-M-001-005-MY3) and Academia Sinica, Taiwan to S.T.D.H; Sigrid Juselius Foundation and Academy of Finland (131413, 137995, 277335) to H.I., and Mobility to Finland under the bilateral agreement of the Academy of Finland (308239) and the National Research Council of Taiwan to S.T.D.H and Y.T.C.L. The Finnish Biological NMR center and crystallization facility are supported by Biocenter Finland, FINStruct, and the Helsinki Institute of Life Science (HiLIFE-INFRA). The Helsinki University Library supported the funding for open access publication.

ACKNOWLEDGMENTS

We thank J. Mutanen, Dr Y.-Q. Yeh, Dr U. Jeng, and T.M. Kauppala for their contribution and assistance, and S. Jääskeläinen and B. Haas for their technical assistance in cloning and protein production. We thank the Biophysics Facility of the Institute of Biological Chemistry, Academia Sinica, the National Synchrotron Radiation Research Center BL23A bioSAXS beamline, crystallization facility at the Institute of Biotechnology, and Finnish Biological NMR center.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.663241/full#supplementary-material

REFERENCES

Beccara, S. a., Skrbic, T., Covino, R., Micheletti, C., and Faccioli, P. (2013). Folding Pathways of a Knotted Protein with a Realistic Atomistic Force Field. *Plos Comput. Biol.* 9, e1003002. doi:10.1371/journal.pcbi.1003002

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Cryst.* 66, 213–221. doi:10.1107/s0907444909052925

Andersson, F. I., Pina, D. G., Mallam, A. L., Blaser, G., and Jackson, S. E. (2009). Untangling the Folding Mechanism of the 5_2-knotted Protein UCH-L3. *FEBS J.* 276, 2625–2635. doi:10.1111/j.1742-4658.2009.06990.x

Aranko, A. S., Oeemig, J. S., Kajander, T., and Iwaï, H. (2013). Intermolecular Domain Swapping Induces Intein-Mediated Protein Alternative Splicing. *Nat. Chem. Biol.* 9, 616–622. doi:10.1038/nchembio.1320

Arrighi, V., Gagliardi, S., Dagger, A. C., Semlyen, J. A., Higgins, J. S., and Shenton, M. J. (2004). Conformation of Cyclics and Linear Chain Polymers in Bulk by SANS. *Macromolecules* 37, 8057–8065. doi:10.1021/ma049565w

Betz, S. F. (1993). Disulfide Bonds and the Stability of Globular Proteins. *Protein Sci.* 2, 1551–1558. doi:10.1002/pro.5560021002

Borra, R., and Camarero, J. A. (2013). Recombinant Expression of Backbone-Cyclized Polypeptides. *Biopolymers* 100, 502–509. doi:10.1002/bip.22306

Capraro, D. T., and Jennings, P. A. (2016). Untangling the Influence of a Protein Knot on Folding. *Biophysical J.* 110, 1044–1051. doi:10.1016/j.bpj.2016.01.017

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Cryst.* 66, 12–21. doi:10.1107/s0907444909042073

Chuang, Y.-C., Hu, I.-C., Lyu, P.-C., and Hsu, S.-T. D. (2019). Untying a Protein Knot by Circular Permutation. *J. Mol. Biol.* 431, 857–863. doi:10.1016/j.jmb.2019.01.005

Clark, R. J., and Craik, D. J. (2010). Invited Reviewnative Chemical Ligation Applied to the Synthesis and Bioengineering of Circular Peptides and Proteins. *Biopolymers* 94, 414–422. doi:10.1002/bip.21372

Covino, R., Skrbic, T., Beccara, S. A., Faccioli, P., and Micheletti, C. (2014). The Role of Non-native Interactions in the Folding of Knotted Proteins: Insights from Molecular Dynamics Simulations. *Biomolecules* 4, 1–19. doi:10.3390/biom4010001

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., and McPherson, A. (1994). Screening and Optimization Strategies for Macromolecular Crystal Growth. *Acta Cryst. D* 50, 414–423. doi:10.1107/s0907444994002660

Dabrowski-Tumanski, P., and Sulkowska, J. (2017). To Tie or Not to Tie? That Is the Question. *Polymers* 9, 454. doi:10.3390/polym9090454

Faísca, P. F. N. (2015). Knotted Proteins: A Tangled Tale of Structural Biology. *Comput. Struct. Biotechnol. J.* 13, 459–468. doi:10.1016/j.csbj.2015.08.003

Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995). Spectral Density Function Mapping Using 15N Relaxation Data Exclusively. *J. Biomol. NMR* 6, 153–162. doi:10.1007/bf00211779

Fersht, A. (1999). *Structure and Mechanism in Protein Science : A Guide to Enzyme Catalysis and Protein Folding*, xxi. New York: W. H. Freeman, 631.

Guerrero, F., Ciragan, A., and Iwaï, H. (2015). Tandem SUMO Fusion Vectors for Improving Soluble Protein Expression and Purification. *Protein Expr. Purif.* 116, 42–49. doi:10.1016/j.pep.2015.08.019

He, C., Li, S., Gao, X., Xiao, A., Hu, C., Hu, X., et al. (2019). Direct Observation of the Fast and Robust Folding of a Slipknotted Protein by Optical Tweezers. *Nanoscale* 11, 3945–3951. doi:10.1039/c8nr10070e

Heikkinen, H. A., Backlund, S. M., and Iwaï, H. (2021). NMR Structure Determinations of Small Proteins Using Only One Fractionally 20 % ^{13}C- and Uniformly 100 % ^{15}N-Labeled Sample. *Molecules* 26, 747. doi:10.3390/molecules26030747

Higgins, J. S., Dodgson, K., and Semlyen, J. A. (1979). Studies of Cyclic and Linear Poly(dimethyl Siloxanes): 3. Neutron Scattering Measurements of the Dimensions of Ring and Chain Polymers. *Polymer* 20, 553–558. doi:10.1016/0032-3861(79)90164-2

Hsieh, S.-J. M., Mallam, A. L., Jackson, S. E., and Hsu, S.-T. D. (2014). Backbone NMR Assignments of a Topologically Knotted Protein in Urea-Denatured State. *Biomol. NMR Assign* 8, 283–285. doi:10.1007/s12104-013-9501-7

Hsu, S.-T. D., Cabrita, L. D., Fucini, P., Dobson, C. M., and Christodoulou, J. (2009). Structure, Dynamics and Folding of an Immunoglobulin Domain of the Gelation Factor (ABP-120) from *Dictyostelium discoideum*. *J. Mol. Biol.* 388, 865–879. doi:10.1016/j.jmb.2009.02.063

Iwai, H., and Fiaux, J. (2007). Use of Biosynthetic Fractional ^{13}C-Labeling for Backbone NMR Assignment of Proteins. *J. Biomol. NMR* 37, 187–193. doi:10.1007/s10858-006-9124-8

Iwai, H., Lingel, A., and Plückthun, A. (2001). Cyclic Green Fluorescent Protein Produced *In Vivo* Using an Artificially Split PI-PfuI Intein from *Pyrococcus Furiosus*. *J. Biol. Chem.* 276, 16548–16554. doi:10.1074/jbc.m011639200

Iwai, H., and Plückthun, A. (1999). Circular β-lactamase: Stability Enhancement by Cyclizing the Backbone. *FEBS Lett.* 459, 166–172. doi:10.1016/s0014-5793(99)01220-x

Iwai, H., Züger, S., Jin, J., and Tam, P.-H. (2006). Highly Efficient Proteintrans-Splicing by a Naturally Split DnaE Intein fromNostoc Punctiforme. *FEBS Lett.* 580, 1853–1858. doi:10.1016/j.febslet.2006.02.045

Jackson, S. E., Suma, A., and Micheletti, C. (2017). How to Fold Intricately: Using Theory and Experiments to Unravel the Properties of Knotted Proteins. *Curr. Opin. Struct. Biol.* 42, 6–14. doi:10.1016/j.sbi.2016.10.002

Jamroz, M., Niemyska, W., Rawdon, E. J., Stasiak, A., Millett, K. C., Sułkowski, P., et al. (2015). KnotProt: a Database of Proteins with Knots and Slipknots. *Nucleic Acids Res.* 43, D306–D314. doi:10.1093/nar/gku1059

Jarmolinska, A. I., Perlinska, A. P., Runkel, R., Trefz, B., Ginn, H. M., Virnau, P., et al. (2019). Proteins' Knotty Problems. *J. Mol. Biol.* 431, 244–257. doi:10.1016/j.jmb.2018.10.012

Jeng, U.-S., Su, C. H., Su, C.-J., Liao, K.-F., Chuang, W.-T., Lai, Y.-H., et al. (2010). A Small/wide-Angle X-Ray Scattering Instrument for Structural Characterization of Air-Liquid Interfaces, Thin Films and Bulk Specimens. *J. Appl. Cryst.* 43, 110–121. doi:10.1107/s0021889809043271

Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Cryst.* 66, 125–132. doi:10.1107/s0907444909047337

Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., et al. (2002). Long-Range Interactions Within a Nonnative Protein. *Science* 295, 1719–1722. doi:10.1126/science.1067680

Ko, K.-T., Hu, I.-C., Huang, K.-F., Lyu, P.-C., and Hsu, S.-T. D. (2019). Untying a Knotted SPOUT RNA Methyltransferase by Circular Permutation Results in a Domain-Swapped Dimer. *Structure* 27, 1224–1233. doi:10.1016/j.str.2019.04.004

Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., et al. (2004). Random-coil Behavior and the Dimensions of Chemically Unfolded Proteins. *Proc. Natl. Acad. Sci.* 101, 12491–12496. doi:10.1073/pnas.0403643101

Kramers, H. A. (1946). The Behavior of Macromolecules in Inhomogeneous Flow. *J. Chem. Phys.* 14, 415–424. doi:10.1063/1.1724163

Lai, Y.-L., Chen, C.-C., and Hwang, J.-K. (2012). pKNOT v.2: the Protein KNOT Web Server. *Nucleic Acids Res.* 40, W228–W231. doi:10.1093/nar/gks592

Lee, Y.-T. C., and Hsu, S.-T. D. (2018). A Natively Monomeric Deubiquitinase UCH-L1 Forms Highly Dynamic but Defined Metastable Oligomeric Folding Intermediates. *J. Phys. Chem. Lett.* 9, 2433–2437. doi:10.1021/acs.jpclett.8b00815

Lee, Y. T. C., Chang, C-Y, Chen, S-Y, Pan, Y-R, Ho, M-R, and Hsu, S.T.D, (2017). Entropic Stabilization of a Deubiquitinase Provides Conformational Plasticity and Slow Unfolding Kinetics Beneficial for Functioning on the Proteasome. *Sci. Rep.* 7, 45174. doi:10.1038/srep45174

Li, W., Terakawa, T., Wang, W., and Takada, S. (2012). Energy Landscape and Multiroute Folding of Topologically Complex Proteins Adenylate Kinase and 2ouf-Knot. *Proc. Natl. Acad. Sci.* 109, 17789–17794. doi:10.1073/pnas.1201807109

Lim, N. C. H., and Jackson, S. E. (2015). Molecular Knots in Biology and Chemistry. *J. Phys. Condens. Matter* 27, 354101. doi:10.1088/0953-8984/27/35/354101

Liu, L., Liu, C., and Alberts, B. (1980). Type II DNA Topoisomerases: Enzymes that Can Unknot a Topologically Knotted DNA Molecule via a Reversible Double-Strand Break. *Cell* 19, 697–707. doi:10.1016/s0092-8674(80)80046-8

Liu, R.-J., Zhou, M., Fang, Z.-P., Wang, M., Zhou, X.-L., and Wang, E.-D. (2013). The tRNA Recognition Mechanism of the Minimalist SPOUT Methyltransferase, TrmL. *Trml.* Nucleic Acids Res. 41, 7828–7842. doi:10.1093/nar/gkt568

Lou, S.-C., Wetzel, S., Zhang, H., Crone, E. W., Lee, Y.-T., Jackson, S. E., et al. (2016). The Knotted Protein UCH-L1 Exhibits Partially Unfolded Forms under Native Conditions that Share Common Structural Features with its Kinetic Folding Intermediates. *J. Mol. Biol.* 428, 2507–2520. doi:10.1016/j.jmb.2016.04.002

Mallam, A. L., and Jackson, S. E. (2007). A Comparison of the Folding of Two Knotted Proteins: YbeA and YibK. *J. Mol. Biol.* 366, 650–665. doi:10.1016/j.jmb.2006.11.014

Mallam, A. L., and Jackson, S. E. (2012). Knot Formation in Newly Translated Proteins Is Spontaneous and Accelerated by Chaperonins. *Nat. Chem. Biol.* 8, 147–153. doi:10.1038/nchembio.742

Mallam, A. L., and Jackson, S. E. (2007). The Dimerization of an α/β-Knotted Protein Is Essential for Structure and Function. *Structure* 15, 111–122. doi:10.1016/j.str.2006.11.007

Mallam, A. L., Morris, E. R., and Jackson, S. E. (2008). Exploring Knotting Mechanisms in Protein Folding. *Proc. Natl. Acad. Sci.* 105, 18740–18745. doi:10.1073/pnas.0806697105

Mallam, A. L., Rogers, J. M., and Jackson, S. E. (2010). Experimental Detection of Knotted Conformations in Denatured Proteins. *Proc. Natl. Acad. Sci.* 107, 8189–8194. doi:10.1073/pnas.0912161107

Mansfield, M. L. (1994). Are There Knots in Proteins?. *Nat. Struct. Mol. Biol.* 1, 213–214. doi:10.1038/nsb0494-213

Mikula, K. M., Tascón, I., Tommila, J. J., and Iwaï, H. (2017). Segmental Isotopic Labeling of a Single-Domain Globular Protein without Any Refolding Step by an Asparaginyl Endopeptidase. *FEBS Lett.* 591, 1285–1294. doi:10.1002/1873-3468.12640

Millett, K. C., Rawdon, E. J., Stasiak, A., and Sułkowska, J. I. (2013). Identifying Knots in Proteins. *Biochem. Soc. Trans.* 41, 533–537. doi:10.1042/bst20120339

Mitchinson, C., and Wells, J. A. (1989). Protein Engineering of Disulfide Bonds in Subtilisin BPN'. *Biochemistry* 28, 4807–4815. doi:10.1021/bi00437a043

Montalbán-López, M., Sánchez-Hidalgo, M., Cebrián, R., and Maqueda, M. (2012). Discovering the Bacterial Circular Proteins: Bacteriocins, Cyanobactins, and Pilins. *J. Biol. Chem.* 287, 27007–27013. doi:10.1074/jbc.r112.354688

Noel, J. K., Sulkowska, J. I., and Onuchic, J. N. (2010). Slipknotting upon Native-like Loop Formation in a Trefoil Knot Protein. *Proc. Natl. Acad. Sci.* 107, 15403–15408. doi:10.1073/pnas.1009522107

Passioni, C., Puri, S., Wang, I., Chen, S. Y., Camilloni, C., and Hsu, S. T. D. (2021). Converging Experimental and Computational Views of the Knotting Mechanism of a Small Knotted Protein. *Biophys. J.* in press. doi:10.1016/j.bpj.2021.03.032

Peng, J. W., and Wagner, G. (1995). Frequency Spectrum of NH Bonds in Eglin C from Spectral Density Mapping at Multiple Fields. *Biochemistry* 34, 16733–16752. doi:10.1021/bi00051a023

Popp, M. W.-L., and Ploegh, H. L. (2011). Making and Breaking Peptide Bonds: Protein Engineering Using Sortase. *Angew. Chem. Int. Ed.* 50, 5024–5032. doi:10.1002/anie.201008267

Raymer, D. M., and Smith, D. E. (2007). Spontaneous Knotting of an Agitated String. *Proc. Natl. Acad. Sci.* 104, 16432–16437. doi:10.1073/pnas.0611320104

Rivera, M., Hao, Y., Maillard, R. A., and Baez, M, (2020). Mechanical Unfolding of a Knotted Protein Unveils the Kinetic and Thermodynamic Consequences of Threading a Polypeptide Chain. *Sci. Rep.* 10, 9562. doi:10.1038/s41598-020-66258-5

Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C., and Benkovic, S. J. (1999). Production of Cyclic Peptides and Proteins In Vivo. *Proc. Natl. Acad. Sci.* 96, 13638–13643. doi:10.1073/pnas.96.24.13638

Shih, P.-M., Wang, I., Lee, Y.-T. C., Hsieh, S.-J., Chen, S.-Y., Wang, L.-W., et al. (2015). Random-coil Behavior of Chemically Denatured Topologically Knotted Proteins Revealed by Small-Angle X-Ray Scattering. *J. Phys. Chem. B* 119, 5437–5443. doi:10.1021/acs.jpcb.5b01984

Sulkowska, J. I., Noel, J. K., Ramírez-Sarmiento, C. A., Rawdon, E. J., Millett, K. C., Onuchic, J. N., et al. (2013). Knotting Pathways in Proteins. *Biochem. Soc. Trans.* 41, 523–527. doi:10.1042/BST20120342

Taylor, W. R. (2000). A Deeply Knotted Protein Structure and How it Might Fold. *Nature* 406, 916–919. doi:10.1038/35022623

Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E., and Bujnicki, J. M. (2007). Structural and Evolutionary Bioinformatics of the SPOUT Superfamily of Methyltransferases. *BMC Bioinformatics* 8, 73. doi:10.1186/1471-2105-8-73

Wang, I., Chen, S.-Y., and Hsu, S.-T. D. (2016). Folding Analysis of the Most Complex Stevedore's Protein Knot. *Sci. Rep.* 6, 31514. doi:10.1038/srep31514

Wang, I., Chen, S.-Y., and Hsu, S.-T. D. (2015). Unraveling the Folding Mechanism of the Smallest Knotted Protein, MJ0366. *J. Phys. Chem. B* 119, 4359–4370. doi:10.1021/jp511029s

Wang, I., Kowalski, M. P., Langley, A. R., Rodriguez, R., Balasubramanian, S., Hsu, S.-T. D., et al. (2014). Nucleotide Contributions to the Structural Integrity and DNA Replication Initiation Activity of Noncoding Y RNA. *Biochemistry* 53, 5848–5863. doi:10.1021/bi500470b

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of theCCP4 Suite and Current Developments. *Acta Crystallogr. D Biol. Cryst.* 67, 235–242. doi:10.1107/s0907444910045749

Zhao, H., Piszczek, G., and Schuck, P. (2015). SEDPHAT - A Platform for Global ITC Analysis and Global Multi-Method Analysis of Molecular Interactions. *Methods* 76, 137–148. doi:10.1016/j.ymeth.2014.11.012

Ziegler, F., Lim, N. C. H., Mandal, S. S., Pelz, B., Ng, W.-P., Schlierf, M., et al. (2016). Knotting and Unknotting of a Protein in Single Molecule Experiments. *Proc. Natl. Acad. Sci. USA* 113, 7533–7538. doi:10.1073/pnas.1600614113

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 29 July 2021
doi: 10.3389/fchem.2021.735149

Chemical Synthesis of the Sec-To-Cys Homologue of Human Selenoprotein F and Elucidation of Its Disulfide-pairing Mode

*Peisi Liao and Chunmao He**

School of Chemistry and Chemical Engineering, South China University of Technology, Guangzhou, China

Herein, we document a highly optimized synthesis of the Sec-to-Cys homologue of the human selenoprotein F (SelF) through a three-segment two-ligation semisynthesis strategy. Highlighted in this synthetic route are two one-pot manipulations, i.e. the first ligation followed by a desulfurization and the second ligation followed by the protein refolding. This way multi-milligrams of the folded synthetic protein was obtained, which set the stage for the synthesis of the natural selenoprotein. Moreover, the disulfide pairing mode of the SelF was elucidated through a combination of site-directed mutagenesis and LC-MS study. It provides not only a criterion to judge the viability of the synthetic protein, and more importantly, useful structural insights into the previously unresolved UGGT-binding domain of SelF.

Keywords: selenoprotein, selenoprotein F, chemical protein synthesis, disulfide bond, thioredoxin-like domain

OPEN ACCESS

Edited by:
Tao Peng,
Peking University, China

Reviewed by:
Gemin Fang,
Anhui University, China
Jia-Bin Li,
Soochow University, China

****Correspondence:***
Chunmao He
hecm@scut.edu.cn

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *02 July 2021*
Accepted: *16 July 2021*
Published: *29 July 2021*

Citation:
Liao P and He C (2021) Chemical Synthesis of the Sec-To-Cys Homologue of Human Selenoprotein F and Elucidation of Its Disulfide-pairing Mode. Front. Chem. 9:735149. doi: 10.3389/fchem.2021.735149

INTRODUCTION

Selenoprotein F (SelF or Selenof), also called the 15-kDa protein (Sep15), is a selenocysteine (Sec, U) containing eukaryotic protein localized to the endoplasmic reticulum (ER) (Gladyshev et al., 1998). As shown in **Scheme 1** (gray color), the N-terminal signal peptide was supposed to lead the expressed SelF to the ER, which is cleaved to form the mature protein (Korotkov et al., 2001; Korotkov et al., 2002). While lacking a typical ER retention peptide sequence, the highly conversed Cys-rich domain of SelF is believed to the key for its localization. This domain is able to bind the UDP-glucose: glycoprotein glucosyltransferase (UGGT)—a large chaperone protein in the ER, it is thus also called the UGGT binding domain (Labunskyy et al., 2005). Further, the C-terminal domain of SelF is identified as a thioredoxin (Trx)-like domain, and the CXU/C redox motif (**Scheme 1**, left) located in a dynamic loop implicates the thiol-disulfide oxidoreductase activity in SelF (Ferguson et al., 2006). As such, SelF is able to play a role in the quality control of the ER (Labunskyy et al., 2007; Labunskyy et al., 2009), but the exact physiological role of the family of protein is yet to be elucidated. A major issue in the study of selenoproteins in general is the lack of reliable recombinant expression techniques, thus most of their biological functions are characterized indirectly, e.g. the knockout assays used in SelF (Joubert et al., 2011; Tsuji et al., 2012; Yim et al., 2018; Zheng et al., 2020). Most of the *in vitro* studies of selenoproteins are thus carried out with the Sec-to-Cys homologue proteins. In this context, the only available structure in the SelF family is reported for a fruit fly Sep15, which contains no Sec residue (**Scheme 1**, right) (Ferguson et al., 2006), and in this case the two Cys residues (Cys80 and Cys82) in the so-called CXC motif forms a disulfide bond. Unfortunately, the Cys-rich UGGT-binding domain in this structure is too flexible to be resolved in the NMR structure, as such the disulfide pairing mode in this domain is currently not known.

Chemical protein synthesis, powered by chemo-selective peptide ligation reactions, has produced almost hundreds of synthetic proteins which conveniently incorporate non-canonical amino acids, among many other purposes (Agouridas et al., 2019; Tan et al., 2020). Although the chemical synthesis of selenoproteins, like SelM and SelW, has been reported (Dery et al., 2017), the synthesis of SelF is, however, supposed to be more challenge as it contains six other Cys residues in the UGGT binding domain, which certainly complicates the refolding process. Moreover, as mentioned, the disulfide pairing mode is currently unknown, making it difficult to determine whether the synthetic SelF is properly folded.

As part of our ongoing research towards the synthesis of SelF, we document herein the chemical synthesis of its Sec-to-Cys homologue—SelF(U65C). The highly optimized synthetic route as well as the refolding strategy developed in the current work would guide the synthesis of the native SelF protein. Further, besides serving a criterion to judge the viability of the synthetic protein, the disulfide pairing mode of SelF(U65C) also provides useful structural insights into the previously unresolved UGGT-binding domain, and it is thus the aim of the current work to elucidate this key information.

MATERIALS AND METHODS

General Reagents and Methods

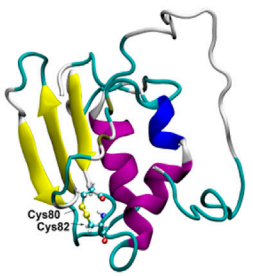
Commercially available materials were obtained from Adamas, Energy Chemicals, or Sigma-Aldrich. Standard Fmoc-amino acids, 2-chlorotrityl chloride (2-Cl-(Trt)-Cl) resin, 1-hydroxybenzotriazole (HOBt), and 2-(1 h-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) were purchased from GL Biochem (Shanghai). The reagents acetylacetone (acac) was purchased from Aladdin. Ethylene diamine tetraacetic acid (EDTA) and 2,2′-(ethylenedioxy) diethanethiol (DODT) were purchased from TCI. 2,2′-azobis [2-(2-imidazolin-2-yl) propane] dihydrochloride (VA-044) and silver acetate (AgOAc) were obtained from J&K Scientific and innochem, respectively. *N*,*N*-diisopropylethylamine (DIPEA), *N*,*N*′-diisopropylcarbodiimide (DIC), dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), trifluoroacetic acid (TFA), 2-methyl-2-propanethiol (t-BuSH), and guanidine hydrochloride (Gn·HCl) were obtained from Adamas. Methoxylamine (CH_3ONH_2·HCl) and triisopropylsilane (TIPS) were purchased from Energy Chemicals. L-arginine hydrochloride (Arg·HCl), tris(hydroxymethyl)aminomethane (Tris), glutathione reduced (GSH) and glutathione oxidized (GSSG) were purchased from Sangon. The reagents *N*,*N*-dimethylformamide (DMF) and dichloromethane (DCM) were purchased from GHTCH (Guangdong). 1,8-diazabicyclo [5.4.0]undec-7-ene (DBU) and 4-mercaptophenylacetic acid (MPAA) were purchased from Alfa Aesar. His_6-Ulp1 was recombinantly expressed using a reported procedure (Malakhov et al., 2004). Acetonitrile (MeCN) used in analytical HPLC and preparative HPLC was obtained from Fisher and Sigma-Aldrich, respectively. Analytical HPLC (Agilent 1,260) was performed on a Phenomenex Jupiter C4 column (4.6 × 250 mm, 300 Å, 5 μm particle size) running at a flow rate of 1 ml/min with UV detection at 214 and 254 nm. Semi-preparative HPLC (Shimadzu AR-20) was performed using a Waters XBridge® peptide BEH C18 OBD™ Prep column, 300 Å, 5 μm, 10 × 250 mm) running at a flow rate of 4.7 ml/min with UV detection at 214 and 254 nm. Preparative HPLC (Ruihe® Tech) was performed using a Welch Ultimate XB-C4 column Prep column, 300 Å, 5 μm, 30 × 250 mm) running at a flow rate of 40 ml/min with UV detection at 214 and 254 nm. Solvent A: 0.1% TFA in water; Solvent B: 0.1% TFA in MeCN. LC-MS was performed on an Agilent LC/MSD (ESI) system on ACE 5 C4 column (150 × 4.6 mm). MALDI-TOF mass spectra (MALDI-8020, Shimadzu) were obtained in the linear positive

```
HUMAN   MVAMAAGPSG CLVPAFGLRL LLATVLQAVS AFGAEFSSEA CRELGFS-SN  49
MOUSE   ---MAAGQGG WLRPALGLRL LLATAFQAVS ALGAEFASEA CRELGFS-SN  46
FLY     ---------- -----MHKCA IFLLLALSCQ QIQAELTAAD CRALGFIKAQ  35

HUMAN   LLCSSCDLLG QFNLLQLDPD CRGCCQEEAQ FETKKLYAGA ILEVCGUKLG  99
MOUSE   LLCSSCDLLG QFNLLPLDPV CRGCCQEEAQ FETKKLYAGA ILEVCGUKLG  96
FLY     LMCSSCEKLD DFGLDTIKPQ CKQCCTLDQQ PAAQRTYAKA ILEVCTCKFR  85

HUMAN   RFPQVQAFVR SDKPKLFRGL QIKYVRGSDP VLKLLDDNGN IAEELSILKW  149
MOUSE   RFPQVQAFVR SDKPKLFRGL QIKYVRGSDP VLKLLDDNGN IAEELSILKW  146
FLY     AYPQIQAFIQ SGRPAKFPNL QIKYVRGLDP VVKLLDASGK VQETLSITKW  135

HUMAN   NTDSVEEFLS EKLERI---- ---------- ---------- ---         165
MOUSE   NTDSVEEFLS EKLERI---- ---------- ---------- ---         162
FLY     NTDTVEEFFE THLAKDGAGK NSYSVVEDAD GDDDEDYLRT NRI         178
```



SCHEME 1 | (Left) Sequence alignment of Selenoprotein F proteins from representative species. Signal peptide sequences are shown in gray color, and all Cys and Sec residues are highlighted. The UGGT-binding domain is boxed (UniProt ID: O60613, Q9ERR7 and Q9VVJ7 for Sep15 from human, mouse and fruit fly, respectively); **(Right)** NMR structure of the fruit fly Sep15, with only the C-terminal Trx-like domain (61–178) resolved and shown (PDB ID, 2A4H). The disulfide bond is shown in ball and stick representation.

mode using a matrix of 10 mg/ml α-Cyano-4-hydroxycinnamic acid (HCCA) in water/MeCN (1: 1, v/v) with 0.1% TFA.

General Procedures for Peptide Synthesis

Preloading of 2-Cl-(Trt)-$NHNH_2$ Resin

2-Cl-(Trt)-Cl (0.9 mmol/g, 1 g) was swollen in DMF for 20 min and then washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml). The resin was treated with freshly prepared 5% hydrazine monohydrate in DMF (2 × 20 ml) for 30 min and then washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml). The resin was treated with freshly prepared 5% MeOH in DMF (20 ml) for 10 min and then washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml). DIPEA (1.2 mmol) was added to a solution of Fmoc-AA-OH (0.6 mmol) and TBTU (0.6 mmol) in DMF (5 ml). After 2 min of pre-activation, the mixture was added to the resin, which was then shaken for 2 h at 25°C. The resin was washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml), and then capped with 20% acetic anhydride in DMF (10 ml) for 20 min, and washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml) again (Zheng et al., 2013).

Estimation of Amino Acid Loading

The resin (10 mg) loaded with the first amino acid was treated with 2% DBU/DMF (2 ml) for 30 min at 25°C to remove the Fmoc group. The deprotection solution (2 ml) was diluted to 10 ml with MeCN, and then 0.8 ml was further diluted to 10 ml with MeCN. The UV absorbance of the resulting piperidine-fulvene adduct solution was measured (λ = 304 nm) to estimate the amino acid loading on the resin.

Fmoc Deprotection

The resin was treated with 20% piperidine in DMF (5 ml, 2 × 10 min) at 25°C and then washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml).

Coupling of General Amino Acids

Peptides were synthesized on a CS Bio 136XT synthesizer using Fmoc solid phase peptide synthesis (SPPS) chemistry. The following Fmoc amino acids with side-chain protecting groups were used: Fmoc-Ala-OH, Fmoc-Arg (Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu (OtBu)-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH. SPPS was performed on 2-Cl-(Trt)-Cl resins. Fmoc deprotections were performed with 20% piperidine in DMF (10 min × 2). Couplings were performed with Fmoc amino acid (4.0 equiv to resin substitution), TBTU (3.9 equiv) and DIPEA (8.0 equiv) in DMF for 60 min (45°C). After coupling, unreacted free amine was capped by treatment with 20% acetic anhydride in DMF and then washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml).

Coupling of Fmoc-Cys(Trt)-OH or Fmoc-Cys(Acm)-OH

A solution of Fmoc-Cys(Trt)-OH or Fmoc-Cys(Acm)-OH (4 equiv), HOBT (3.9 equiv), and DIC (8 equiv) in DMF was added to the resin. The reaction was shaken for 1 h at 45°C. After coupling, unreacted free amine was capped by treatment with 20% acetic anhydride in DMF and then washed with DMF (2 × 5 ml), DCM (2 × 5 ml), and DMF (2 × 5 ml).

Cleavage of the Crude Peptide

The dried resin was treated with TFA/H_2O/TIPS (95:2.5:2.5 v/v/v), or TFA/H_2O/DODT (95:2.5:2.5 v/v/v) (2–3 ml per 100 mg of resin) and shaken for 2 h a t 25°C. After filtration, the filtrate was concentrated by blowing with a gentle flow of N_2. Add the precooled diethyl ether to precipitate crude peptides. The resulted suspension was centrifuged (8,000 rpm, 5 min, 4°C), and the ether layer was decanted. Air-dry the peptide product in the open centrifuge tube for about 30 min. The targeted crude peptide was obtained as the solid.

Synthesis of Thioester Fragment 1

SelF(1–41)-MPAA **(1)** was synthesized on 2-Cl-(Trt)-Cl resin (theoretical loading: 0.9 mmol/g) using Fmoc-Gly-OH with 0.4 mmol/g loading and elongated according to standard Fmoc-SPPS protocols highlighted in *Preloading of 2-Cl-(Trt)-$NHNH_2$ Resin, Estimation of Amino Acid Loading, Fmoc Deprotection, Coupling of General Amino Acids, and Coupling of Fmoc-Cys(Trt)-OH or Fmoc-Cys(Acm)-OH* to afford resin-bound peptide. The peptide was cleaved using TFA/H_2O/DODT (95:2.5:2.5 v/v/v) for 2 h, and according to *Cleavage of the Crude Peptide* to acquire the crude peptide SelF(1–41)-$NHNH_2$ (**1a**). The crude peptide **1a** (400 mg, assumed 100% purity) was dissolved to 40 mg/ml in 6 M Gn·HCl, 0.2 M Na_2HPO_4, pH 3.0, with 10 equiv MPAA, 2.5 equiv acac (from a 0.1 M stock in water) were added to the mixture, and the reaction mixture was stirred for 10 h to form thioester fragment SelF(1–41)-MPAA (**1**). The mixture was centrifuged (8,000 rpm, 5 min, 4°C), filtered and purified by preparative HPLC at 25°C with a gradient of 25–70% MeCN (with 0.1% TFA) in 25 min to obtain 8 mg of segment **1** (1 g resin; 2.5%). Overall, 18 mg of segment **1** was obtained. The purity and exact mass of the peptide was confirmed using analytical HPLC and ESI-MS, respectively (**Supplementary Figure 1**).

Synthesis of Thioester Fragment 2

SelF(42–74)-MPAA (**2**) was synthesized on 2-Cl-(Trt)-Cl resin (theoretical loading: 0.9 mmol/g) using Fmoc-Ala-OH with 0.35 mmol/g loading. Through a similar procedure described in *Synthesis of Thioester Fragment 1* and a preparative HPLC at 25°C with a gradient of 25–50% MeCN (with 0.1% TFA) in 25 min, 25 mg of segment **2** was obtained (1 g resin; 7.4%). Overall, 75 mg of segment **2** was obtained. The purity and exact mass of the peptide was confirmed using analytical HPLC and ESI-MS, respectively (**Supplementary Figure 2**).

Expression of His_6-Sumo-SelF(75-134) and Its Enzymatic Cleavage

The gene encoding for the His_6-SUMO-SelF(75–134) fusion protein was synthesized and codon-optimized for *E. Coli* expression (GenScript Inc., Nanjing). The synthetic gene was cloned into the pET-30a expression vector using the *Nde*I/*Eco*RI

restriction sites. The plasmid was firstly transformed into BL21 (DE3) *E. coli* cells chemically. An overnight culture of the cells harboring an expression vector was inoculated (1:50 dilution) in a 4 L flask containing 30 μg/ml Kanamycin in 2 L LB at 37°C. After reaching an OD_{600} of 0.6–0.8 overexpression of His_6-SUMO-SelF (75–134) was induced by the addition of 2 ml 1 M IPTG stock solution (final conc. 1 mM) at 37°C for 4 h. Cells were harvested by centrifugation (8,000 rpm, 4°C, 15 min). Typically, 6 g cells were resuspended in 50 ml of cell lysis buffer and lysed by ultrasonication (30–40% power, 3 s on 5 s off, 25 min). The crude lysate was centrifuged (16,000 rpm, 4°C, 20 min) and the supernatant was discarded. The precipitate was stirred at 4°C overnight with 10 ml of Ni-NTA binding buffer to extract His_6-SUMO-SelF(75–134). The precipitate was removed by centrifugation (16,000 rpm, 30 min, 4°C, 5 cycles) and the supernatant applied to a HisTrap™ FF column (5 ml) at 2 ml/min with a AKTA pure chromatography system. Absorption was monitored at 280 nm. The column was washed with 20 ml of Ni-NTA binding buffer (see **Supplementary Material** for details). His_6-SUMO-SelF(75–134) was eluted with 25 ml of Ni-NTA eluting buffer (see **Supplementary Material** for details) in fractions of 5 ml. The 10 ml fractions with A280 > 0.1 (**Supplementary Figure 3A**) was gradually dialyzed to the 0.5 L refolding buffer (see **Supplementary Material** for details) to complete the refolding of SUMO domain. 400 μL of a stock solution of His_6-Ulp (A280 = 0.5) were added to the folded His_6-SUMO-SelF(75–134) **(3a)** (10 ml, A280 = 1.8) (Ulp1: protein = 2: 50, v/v) and the reaction was incubated for 2 h at 30°C. Equal volume of the buffer containing 6 M Gn·HCl, 200 mM Na_2HPO_4, 200 mM CH_3ONH_2·HCl, 10 mM TCEP (pH 3), was added to the mixture and the pH was adjusted to 4.0 to remove any C-terminal cyclized byproduct. After overnight incubation, the mixture was centrifuged (8,000 rpm, 5 min, 4°C), filtered and purified by preparative HPLC at 25°C with a gradient of 25–48% MeCN (with 0.1% TFA) in 20 min to obtain 13 mg of SelF(75–134) (**3**) (6.5 mg/L LB) (**Supplementary Figure 3C**). The purity and exact mass of the peptide was confirmed using analytical HPLC and ESI-MS, respectively (**Supplementary Figures 3C-c,D**).

1st Ligation, Desulfurization and Acm Deprotection

1st Ligation and Desulfurization in One-Pot (on a 5 μmol Scale)

The peptide-thioester **2** (6.5 μmol, 1.3 equiv, 26.4 mg) and Cys-peptide **3** (5 μmol, 1 equiv, 34.8 mg) was dissolved in 2.5 ml of ligation buffer of 6 M Gn·HCl and 2.5 M imidazole with 10 mM TCEP at pH 6.5. The solution was incubated at RT for 1 h (confirmed by LC-MS monitoring). Then, to the ligation reaction mixture was added 295 mg TCEP (final conc. 200 mM), t-BuSH (260 μL, 5%, v/v), and an aqueous solution (2.5 ml) of 0.1 M VA-044 (50 equiv, 250 μmol), and the solution (final pH 6.5) was incubated at 42°C (the reaction was monitored by LC-MS). After the desulfurization was completed (24 h), the mixture was centrifuged (8,000 rpm, 5 min, 4°C), filtered and purified by preparative HPLC at 25°C with a gradient of 25–55% MeCN (with 0.1% TFA) in 30 min to collect the desired fractions and immediately lyophilized, affording the desired protein **4** as a white amorphous powder (29.1 mg, 53.8% isolated yield over two steps). The purity and exact mass of the peptide was confirmed using analytical HPLC and ESI-MS, respectively (**Supplementary Figures 4A-g,C**).

Acm Deprotection (on a 1 μmol Scale)

The peptide **4** (1 μmol, 1.0 equiv, 10.8 mg) was dissolved in a 50% aq. acetic acid (5 ml) containing 1% AgOAc, and the mixture was stirred at 50°C for 2 h in the dark. Then 154.2 mg DTT was added to the mixture, and the formed precipitate was separated by centrifugation. The precipitate was repeatedly washed with 6 M Gn·HCl solution and the combined supernatant (ca. 8 ml) was filtered and purified by preparative HPLC at 25°C with a gradient of 25–55% MeCN (with 0.1% TFA) in 30 min to collect the desired fractions and immediately lyophilized, affording the desired protein **5** as a white amorphous powder (5.17 mg, 49.0%). The purity and exact mass of the peptide was confirmed using analytical HPLC and ESI-MS, respectively (**Supplementary Figures 5A-c,B**).

2nd Ligation and Protein Folding in One-Pot

The peptide-thioester **1** (0.76 μmol, 2 equiv, 3.48 mg) and Cys-peptide **5** (0.38 μmol, 1 equiv, 4 mg) was dissolved in 190 μL of the ligation buffer (6 M Gn·HCl, 200 mM Na_2HPO_4, 20 mM TCEP, 50 mM MPAA, pH 6.5). The solution was incubated at RT for 24 h (confirmed by LC-MS monitoring). After the conversion was completed, the ligation reaction mixture was reduced by the addition of 2.86 mg TCEP (final conc. 50 mM) and incubated for 15 min. Then, 200 μL of the ligation mixture (in 50 μL portions) was exchanged into a 150 μL buffer containing 6 M Gn·HCl and 200 mM Na_2HPO_4 at pH 6 (Amicon® Ultra-0.5 concentrator, 0.5 ml, 3 K MWCO, 11,000 rpm, 4°C). The resulting ligation mixture (~600 μL) was added dropwise to a 20 ml refolding buffer (0.4 M Arg·HCl, 0.2 M Tris, 0.1 M $(NH_4)_2SO_4$, 2 mM EDTA, 0.2 mM GSSG, 1 mM GSH, pH 8.2) (final protein conc. was ~0.25 mg/ml) at 4°C in 30 min, and was stirred in a refrigerator at 4°C for 12 h. The mixture was centrifuged (8,000 rpm, 5 min, 4°C), filtered and purified by semi-preparative HPLC with a two-step gradient: 10–30% in 5 min, then 30–55% MeCN (with 0.1% TFA) in 30 min to collect the desired fractions and lyophilized, and the desired folded protein SelF(U65C/Q74A) (**7**) as a white amorphous powder (1 mg, 17.6% isolated yield over two steps). The purity and exact mass of the peptide was confirmed using analytical HPLC and ESI-MS, respectively (**Supplementary Figures 6A-e,C**).

Expression and Purification of Full-Length SelF(U65C) and Its Cys-To-Ser Variants

The gene coding for the His_6-SUMO-SelF(U65C) fusion protein was synthesized and codon-optimized for *E. Coli* expression (GenScript Inc., Nanjing). The synthetic gene was cloned into the pET-30a expression vector using the *Nde*I/*Eco*RI restriction sites. Site-directed mutagenesis using the Sit-directed Mutagenesis kit from Sangon (Shanghai) was carried out to

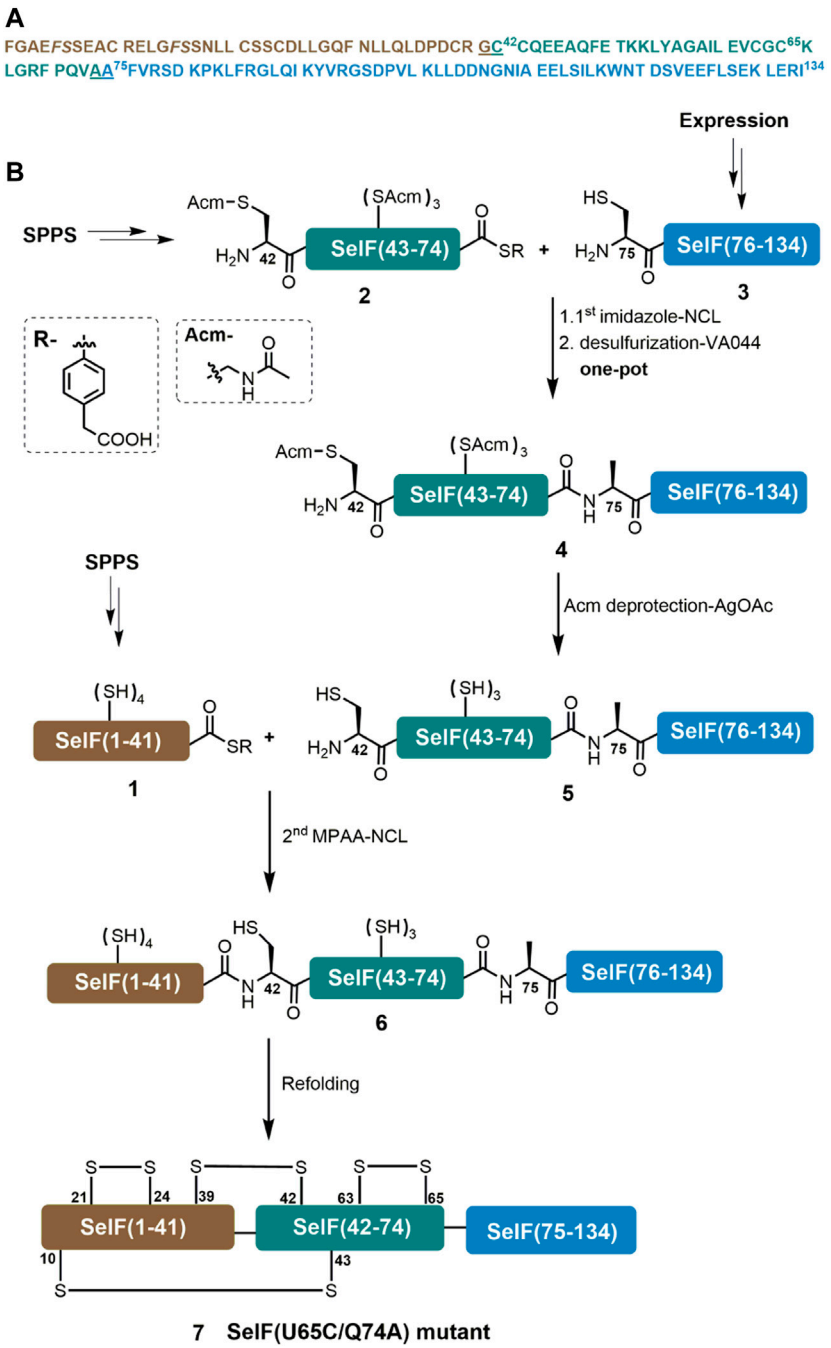


FIGURE 1 | Chemical synthesis of SelF(U65C/Q74A). **(A)** Sequence of the mature form of SelF(U65C/Q74A). The ligation sites are underlined, and the pseudo-proline dipeptide building blocks used during SPPS are shown in italic. **(B)** Synthetic strategy used in the current work.

afford the two Cys-to-Ser variants, U65C/C42S and U65C/C43S respectively. The resulting expression plasmids were transformed into *E. coli* BL21 (DE3). In a typical protein expression experiment, cells were grown in a 2-L LB medium containing 30 ug/mL of kanamycin at 18°C. Expression was induced after reaching an OD_{600} of 0.6 with 1 M IPTG stock solution (final conc. 0.5 mM) and cells were grown for 20 h at 18°C. The cells were harvested by centrifugation (8,000 rpm, 4°C, 15 min). Typically, 4.5 g of cells were resuspended in 30 ml of cell lysis buffer and lysed by ultrasonication (30–40% power, 3 s on 5 s off, 25 min). The crude lysate was centrifuged (16,000 rpm, 4°C, 30 min, 5 cycles) and the supernatant applied to a HisTrap™ FF column (5 ml) at 2 ml/min with a AKTA pure chromatography system. Absorption was monitored at

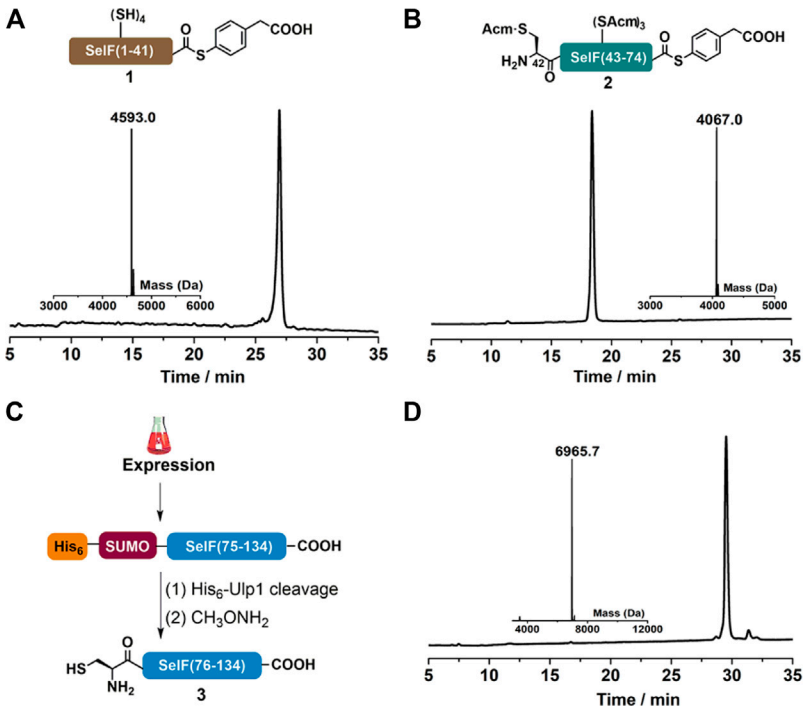


FIGURE 2 | (A–B) Analytical HPLC and mass analysis of SelF(1–41)-MPAA **(1)** and SelF(42–74)-MPAA (**2**). **1**: observed mass 4,593.0 Da, calcd 4,593.0 Da (average isotopes); **2**: obsd. mass 4,067.0 Da, calcd. mass 4,066.9 Da (average isotopes). **(C)** Overexpression and Ulp1 cleavage to afford SelF(75–134) (**3**). **(D)** Analytical HPLC and mass analysis of SelF(75–134) (**3**) with the obsd. mass 6,965.7 Da, calcd. mass 6,965.7 Da (average isotopes). Detailed reaction conditions in the Materials and Methods section.

280 nm. The column was washed with 20 ml of Ni-NTA binding buffer (see supplementary material for details). The desired proteins were eluted with 25 ml of Ni-NTA eluting buffer (see supplementary material for details) in fractions of 5 ml. The 10 ml fractions with A280 > 0.1 (**Supplementary Figures 7, 8, 9A**) were collected. Then, 400 µL of a stock solution of His_6-Ulp (A280 = 0.5) was added to the fusion protein **(8a, 9a or 10a)** (10 ml, A280 = 2.1) (Ulp1/protein = 2/50, v/v) and the reaction was incubated for 2 h at 30°C. An extra alkylation step was carried out for the two variants by adding 100 mg of iodoacetamide (IAM) (10 mg/ml) and incubating for 1 h to protect the free Cys residue with a S-carboxyamidomethyl (CAM) group. The mixture was centrifuged (8,000 rpm, 5 min, 4°C), filtered and purified by preparative HPLC using at 25°C with a gradient of 30–55% MeCN (with 0.1% TFA) in 30 min to obtain the desired proteins: 8–9 mg (4–5 mg/L LB) (**Supplementary Figures 7, 8, 9C**). The purity and exact mass of the protein was confirmed using analytical HPLC and ESI-MS, respectively (**Supplementary Figure 7C–D, 8C–D, 9C–D**).

Circular Dichroism

The secondary structure content of the synthetic SelF(U65C/Q74A) (**7**) was compared to the recombinant SelF(U65C) (**8**) using far-UV CD spectroscopy (200–260 nm). Spectra were recorded on a Chirascan™-Plus Circular Dichroism Spectrometer (Applied Photophysics Ltd., United Kingdom), using a quartz cuvette with a path length of 0.1 cm, and obtained by averaging 3 wavelength scans in 1 nm steps, with a signal averaging time of 1 s and a bandwidth of 1 nm. Each purified protein was dissolved separately in NH_4HCO_3 buffer. Measuring conditions: Protein conc.: ~38 µM; Buffer: 10 mM NH_4HCO_3, pH 8. The folding of the expressed proteins was also confirmed by CD spectroscopy, which gave very similar spectrum for each protein (**Supplementary Figure 10**).

Enzymatic Digestion

Protein was dissolved in a solution of 25 mM NH_4HCO_3 (1 mg/ml, 100 µL). Trypsin (0.5 mg/ml, 10 µL) was firstly added to the mixture to digest the protein (Diemer et al., 2020). The reaction was carried out at 37°C for 2 h. Then, the enzymatic reaction was analyzed by LC-MS directly. Next, chymotrypsin (0.16 mg/ml, 11 µL) was added to the reaction mixture for further digestion. The reaction was carried out at 30°C for 2 h. Finally, the enzymatic reaction was analyzed by LC-MS.

RESULTS AND DISCUSSION

Synthesis Strategy

A major aim of the current work is to establish an efficient synthetic route for the Cys homolog of the mature human SelF protein—which could later be applied for the synthesis of the Sec-containing protein. For this purpose, we initially used a three peptide segments semi-synthetic strategy, in which a short

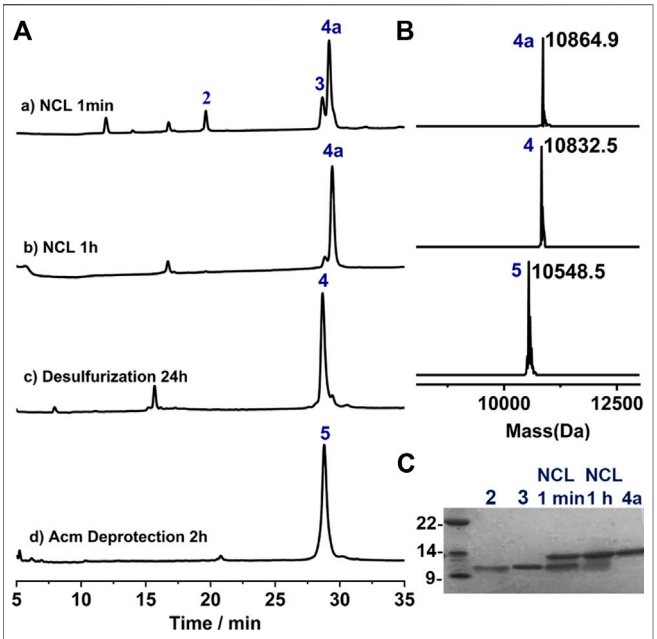


FIGURE 3 | (A) Analytical HPLC traces of one-pot NCL-Desulfurization reaction and the Acm group deprotection reaction. **(B)** Mass analysis of ligation reaction product **4a**, **4** and **5**, respectively. **4a**: observed mass 10,864.9 Da, calcd 10,865.6 Da (average isotopes); **4**: obsd. mass 10,832.5 Da, calcd. mass 10,832.6 Da (average isotopes); **5**: obsd. mass 10,548.5 Da, calcd. mass 10,548.5 Da (average isotopes). **(C)** Analytical SDS-PAGE analysis of the ligation reaction.

peptide hydrazide segment Cys65–Gln74 was synthesized by SPPS. The remaining two segments could be obtained from recombinant expression. This way we hoped to maximize the synthetic yield of the peptide segments and thus the synthetic protein. An extra C-terminal Cys residue was included in the expression plasmid for the first segment—Phe–Gly64, which was supposed to be able to transform to the corresponding thioester via a reported hydrazinolysis procedure (Adams et al., 2013; Pan et al., 2019). However, the effect of hydrazinolysis is not satisfactory, giving only minor product (data not shown). We thus switched to an alternative strategy where disconnection sites at Gly41–Cys42 and Ala74–Ala75 were adopted. It is noted that for the purpose of applying the peptide hydrazide chemistry, the C-terminal Gln74 residue in the middle segment was mutated to an Ala residue, which should not affect the protein folding and function (vide infra). Here, segments **1** and **2** can be obtained directly from SPPS, while segment **3** can be fused N-terminally to a His_6-SUMO tag and recombinantly expressed. As shown in **Figure 1**, the peptide segments were designed to be ligated by native chemical ligation (NCL) (Dawson et al., 1994), and thus an N-terminal Cys residue mutation will be installed in **3** and after its ligation with segment **2**, a desulfurization could lead to the corresponding Ala residue (Wan and Danishefsky, 2007). For this purpose, all Cys sidechains of segment **2** have to be protected with Acm in prior, which after the ligation and desulfurization can be removed. The resulting segment can then be ligated with segment **1** to afford the desired full-length protein **6**.

Peptide Segments Synthesis

As mentioned, peptide segments **1** and **2** were obtained via SPPS, and C-terminal α-hydrazide was used as thioester precursor. The use of pseudoproline dipeptide Fmoc-Phe-Ser($\Psi^{Me,Me}$pro)-OH (Wöhr, et al., 1996) proved to be essential for a reasonable yield (**Figure 1A**). Both segments were cleaved from the resin as peptide hydazides (Fang et al., 2011) and the crude products were transformed into peptide-MPAA thioesters via an acetyl acetone (acac) method before HPLC purification (**Figures 2A,B**) (Flood et al., 2018). It is noted that NCL with purified peptide thioesters gave better yields than *in situ* method applying peptide hydrides (vide infra) Segment **3** was obtained recombinantly with an N-terminal His_6-SUMO tag (**Figure 2C**) (Malakhov et al., 2004; Reif et al., 2014). And after Ulp1-cleavage, the desired segment **3** was obtained with high purity and yield (6.5 mg/L LB, **Figure 2D**). It is noted that before HPLC purification, the Ulp1-cleavage product was treated with methoxylamine to remove any cyclized thiazolidine byproduct at the N-terminus (Reif et al., 2014).

1st Ligation Between Segments 2 and 3

With all the peptide in hand, we preceded to assemble the full-length protein via NCL. The first NCL was carried between peptide segments **2** and **3**, and with the purpose of applying one-pot desulfurization where the use of MPAA as catalyst is problematic, imidazole was instead used as the catalyst for the ligation (Sakamoto et al., 2016). As shown in **Figure 3A,B**, the ligation went almost to completion within 1 h, and the reaction progress is better overviewed from the SDS-page analysis (**Figure 3C**). To this mixture, TCEP and VA-044 were added directly to achieve quantitative desulfurization at Cys75, leading to peptide **4** with excellent isolated yield (29.1 mg, 53.8% over two steps). And after HPLC purification, the Acm protection groups can be conveniently removed using AgOAc (Murar et al., 2020), affording **5** with high isolated yield (14.2 mg, 49.0%, **Figure 3A, d**).

2nd Ligation and Protein Folding

Further, a second ligation between peptide **1** and **5** was carried out to obtain the full-length synthetic protein. In this case, MPAA was added as the catalyst and the ligation proceeded smoothly, albeit with a slow reaction rate compared to the first ligation. The presence of thiolactones (indicated with * in **Figure 4A**, a-b) resulting from the thiol-exchange between the C-thioester and thiol side-chains in peptide **1** was the key for the slow reaction rate. The ligation was allowed to proceed for 24 h and the resulting solution was directly exchanged into refolding buffer, according to the literature (Vetter et al., 2009) (**Figure 4A-C**). Gratifyingly, a sharp peak with high intensity in HPLC appeared after 12 h, which as judged from ESI-MS analysis (~−8 Da compared to the unfolded protein, **6**) corresponds to the folded synthetic protein **7**. Following semi-preparative HPLC purification, **7** was obtained with an isolated yield of 17.6% (3 mg, over two steps, ligation and refolding). Finally, the folded synthetic protein **7** shows identical CD spectrum to the expressed protein counterpart (**Figures 4D,E**), confirming the correct formation of the 3D structure (Dery et al., 2017).

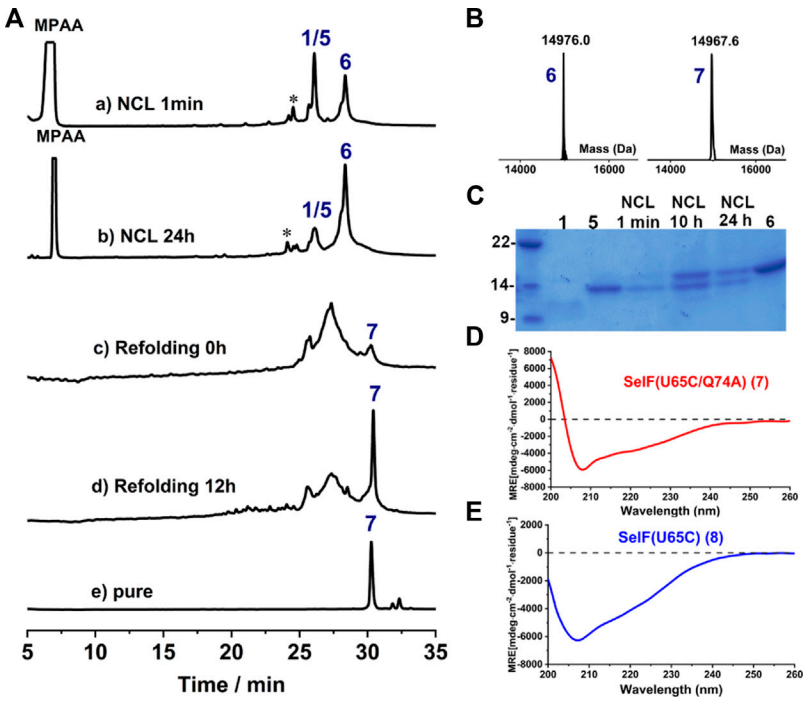


FIGURE 4 | (A) Analytical HPLC traces of ligation and refolding reaction of SelF(U65C/Q74A). **(B)** Mass analysis of ligation reaction product **6** and its folded protein **7**, respectively. **6:** obsd. mass 14,976.0 Da, calcd. mass 14,976.2 Da (average isotopes); **7:** obsd. mass 14,967.6 Da, calcd. mass 14,968.2 Da (average isotopes). **(C)** Analytical SDS-PAGE analysis of the ligation reaction. **(D,E)** CD spectra of the fully synthetic SelF(U65C/Q74A) (**7**) **(D)** and the recombinant SelF (U65C) (**8**) **(E)**.

TABLE 1 | Disulfide bond elucidation through the enzymatic digestion of SelF(U65C) (**8**), SelF(U65C/C42S)-(CAM) (**9**) and SelF(U65C/C43S)-(CAM) (**10**).

Entry	Protein variant	Digestion	Sequences	Molecular weight (Da)	
				Obs	Calc
1	SelF(U65C) (**8**)	Trypsin	LYAGAILEVC^{63}GC^{65}K	1,336.6	1,336.7
2		Trypsin and chymotrypsin	AGAILEVC^{63}GC^{65}K	1,060.7	1,060.5
3			SSNLLC^{21}SSC^{24}DLLGQF	1,583.0	1,583.7
4	SelF(U65C/C42S)-CAM (**9**)	Trypsin	FGAEFSSEAC^{10}R GS^{42}C^{43}QEEAQFETK	2,557.0	2,556.7
5		Trypsin and chymotrypsin	NLLQLDPDC^{39}(CAM)R	1,243.0	1,242.5
6	SelF(U65C/C43S)-CAM (**10**)	Trypsin	FGAEFSSEAC^{10}(CAM)R	1,259.0	1,259.2
7		Trypsin and chymotrypsin	NLLQLDPDC^{39}RGC^{42}S^{43}QEEAQF	2,180.9	2,181.3

Disulfide Pairing Mode Analysis

As mentioned earlier, the disulfide pairing mode of SelF has not been established. The elucidation of such a pairing mode would shed light on its overall structure, especially when the 3D structure of the UGGT binding domain is currently not known. Herein, we identified all disulfide bonds in the SelF(U65C) through a two-step enzymatic digestion of both synthetic and expressed proteins. As shown in **Table 1**, following trypsin treatment of the recombinant protein SelF(U65C), **8**, a peptide fragment bearing the Cys63–Cys65 disulfide can be clearly observed (Entry 1)

(**Supplementary Figure 12A**), which agrees well with the reported NMR structure of a homolog protein—the fruit fly Sep15 (Ferguson et al., 2006). Further treatment of the peptide mixture with chymotrypsin allows the identification of a peptide fragment containing Cys21–Cys24 (**Table 1**, Entry 3) (**Supplementary Figure 12B**). The connectivity mode of the remaining four Cys residues is, however, difficult to resolve as Cys42 and Cys43 are next to each other. In this case, we created two Cys-to-Ser variants, i.e. SelF(U65C/C42S) and SelF(U65C/C43S), and the now-free Cys in each variant was firstly protected with IAM. The resulting proteins—SelF(U65C/C42S)-CAM, **9** and SelF(U65C/C43S)-CAM, **10**—were then processed through a similar sequential trypsin/chymotrypsin digestion, and the remaining two disulfide bonds were established as Cys10–Cys43 and Cys39–Cys42, respectively (**Table 1**, Entries 4 and 7) (**Supplementary Figure 13A, 14B**). Importantly, the synthetic protein showed the same disulfide pairing mode as the expressed protein (**Supplementary Figure 11, 12**), which reassures the viability of the synthetic strategy developed in the current work to provide authentic samples for further biological studies.

CONCLUSION

In summary, we have developed an efficient synthetic strategy affording the Cys homologue of human SelF protein. It highlights the use of two one-pot operations, i.e. the first ligation and desulfurization and the second ligation and refolding, and through this highly optimized strategy, multi-milligram of folded SelF(U65C/Q74A) was obtained. Given that the Sec65 residue is in the middle of the segment **2**, we envision that the synthetic and refolding strategies developed in this study can be directly applied in the chemical synthesis of the native SelF. As such, it would not only provide enough authentic samples for further biological studies, but also set the stage for the synthesis of the selenocysteine-containing protein—SelF. Moreover, the disulfide pairing mode of SelF has been elucidated for the first time, which should provide a good opportunity to understand its unique biological functions, such as its binding to UGGT.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

PL performed the experiments and wrote the manuscript. CH conceptualized the project and revised the manuscript.

FUNDING

This work was supported by the Fundamental Research Funds for the Central Universities of SCUT (No. 2020ZYGXZR056) and partially by the National Natural Science Foundation of China (No. 22077040).

ACKNOWLEDGMENTS

The technical assistance in PCR and protein expression from Zeyuan Mo and Shunzi Huang was gratefully acknowledged. We also thank Yuqi Zhang for helpful discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.735149/full#supplementary-material

REFERENCES

Adams, A. L., Cowper, B., Morgan, R. E., Premdjee, B., Caddick, S., and Macmillan, D. (2013). Cysteine Promoted C-Terminal Hydrazinolysis of Native Peptides and Proteins. *Angew. Chem. Int. Ed.* 52, 13062–13066. doi:10.1002/anie.201304997

Agouridas, V., El Mahdi, O., Diemer, V., Cargoët, M., Monbaliu, J.-C. M., and Melnyk, O. (2019). Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* 119, 7328–7443. doi:10.1021/acs.chemrev.8b00712

Dawson, P., Muir, T., Clark-Lewis, I., and Kent, S. (1994). Synthesis of Proteins by Native Chemical Ligation. *Science* 266, 776–779. doi:10.1126/science.7973629

Dery, L., Reddy, P. S., Dery, S., Mousa, R., Ktorza, O., Talhami, A., et al. (2017). Accessing Human Selenoproteins through Chemical Protein Synthesis. *Chem. Sci.* 8, 1922–1926. doi:10.1039/C6SC04123J

Diemer, V., Ollivier, N., Leclercq, B., Drobecq, H., Vicogne, J., Agouridas, V., et al. (2020). A Cysteine Selenosulfide Redox Switch for Protein Chemical Synthesis. *Nat. Commun.* 11, 2558. doi:10.1038/s41467-020-16359-6

Fang, G.-M., Li, Y.-M., Shen, F., Huang, Y.-C., Li, J.-B., Lin, Y., et al. (2011). Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.* 50, 7645–7649. doi:10.1002/anie.201100996

Ferguson, A. D., Labunskyy, V. M., Fomenko, D. E., Araç, D., Chelliah, Y., Amezcua, C. A., et al. (2006). Nmr Structures of the Selenoproteins Sep15 and Selm Reveal Redox Activity of a New Thioredoxin-like Family. *J. Biol. Chem.* 281, 3536–3543. doi:10.1074/jbc.M511386200

Flood, D. T., Hintzen, J. C. J., Bird, M. J., Cistrone, P. A., Chen, J. S., and Dawson, P. E. (2018). Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed.* 57, 11634–11639. doi:10.1002/anie.201805191

Gladyshev, V. N., Jeang, K.-T., Wootton, J. C., and Hatfield, D. L. (1998). A New Human Selenium-Containing Protein. *J. Biol. Chem.* 273, 8910–8915. doi:10.1074/jbc.273.15.8910

Joubert, M. K., Luo, Q., Nashed-Samuel, Y., Wypych, J., and Narhi, L. O. (2011). Classification and Characterization of Therapeutic Antibody Aggregates. *J. Biol. Chem.* 286, 25118–25133. doi:10.1074/jbc.M110.160457

Korotkov, K. V., Kumaraswamy, E., Zhou, Y., Hatfield, D. L., and Gladyshev, V. N. (2001). Association between the 15-kda Selenoprotein and Udp-glucose: Glycoprotein Glucosyltransferase in the Endoplasmic Reticulum of

Mammalian Cells. *J. Biol. Chem.* 276, 15330–15336. doi:10.1074/jbc.M009861200

Korotkov, K. V., Novoselov, S. V., Hatfield, D. L., and Gladyshev, V. N. (2002). Mammalian Selenoprotein in Which Selenocysteine (Sec) Incorporation Is Supported by a New Form of Sec Insertion Sequence Element. *Mol. Cel. Biol.* 22, 1402–1411. doi:10.1128/MCB.22.5.1402-1411.2002

Labunskyy, V., Hatfield, D., and Gladyshev, V. (2007). The Sep15 Protein Family: Roles in Disulfide Bond Formation and Quality Control in the Endoplasmic Reticulum. *Tbmb* 59, 1–5. doi:10.1080/15216540601126694

Labunskyy, V. M., Ferguson, A. D., Fomenko, D. E., Chelliah, Y., Hatfield, D. L., and Gladyshev, V. N. (2005). A Novel Cysteine-Rich Domain of Sep15 Mediates the Interaction with UDP-glucose:Glycoprotein Glucosyltransferase. *J. Biol. Chem.* 280, 37839–37845. doi:10.1074/jbc.M508685200

Labunskyy, V. M., Yoo, M.-H., Hatfield, D. L., and Gladyshev, V. N. (2009). Sep15, a Thioredoxin-like Selenoprotein, Is Involved in the Unfolded Protein Response and Differentially Regulated by Adaptive and Acute Er Stresses. *Biochemistry* 48, 8458–8465. doi:10.1021/bi900717p

Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D., and Butt, T. R. (2004). Sumo Fusions and Sumo-specific Protease for Efficient Expression and Purification of Proteins. *J. Struct. Func Genom* 5, 75–86. doi:10.1023/B:JSFG.0000029237.70316.52

Murar, C. E., Ninomiya, M., Shimura, S., Karakus, U., Boyman, O., and Bode, J. W. (2020). Chemical Synthesis of Interleukin-2 and Disulfide Stabilizing Analogues. *Angew. Chem. Int. Ed.* 59, 8425–8429. doi:10.1002/anie.201916053

Pan, M., Zheng, Q., Ding, S., Zhang, L., Qu, Q., Wang, T., et al. (2019). Chemical Protein Synthesis Enabled Mechanistic Studies on the Molecular Recognition of K27-linked Ubiquitin Chains. *Angew. Chem. Int. Ed.* 58, 2627–2631. doi:10.1002/anie.201810814

Reif, A., Siebenhaar, S., Tröster, A., Schmälzlein, M., Lechner, C., Velisetty, P., et al. (2014). Semisynthesis of Biologically Active Glycoforms of the Human Cytokine Interleukin 6. *Angew. Chem. Int. Ed.* 53, 12125–12131. doi:10.1002/anie.201407160

Sakamoto, K., Tsuda, S., Mochizuki, M., Nohara, Y., Nishio, H., and Yoshiya, T. (2016). Imidazole-aided Native Chemical Ligation: Imidazole as a One-Pot Desulfurization-Amenable Non-thiol-type Alternative to 4-mercaptophenylacetic Acid. *Chem. Eur. J.* 22, 17940–17944. doi:10.1002/chem.201604320

Tan, Y., Wu, H., Wei, T., and Li, X. (2020). Chemical Protein Synthesis: Advances, Challenges, and Outlooks. *J. Am. Chem. Soc.* 142, 20288–20298. doi:10.1021/jacs.0c09664

Tsuji, P. A., Carlson, B. A., Naranjo-Suarez, S., Yoo, M.-H., Xu, X.-M., Fomenko, D. E., et al. (2012). Knockout of the 15 Kda Selenoprotein Protects against Chemically-Induced Aberrant Crypt Formation in Mice. *PLoS ONE* 7, e50574. doi:10.1371/journal.pone.0050574

Vetter, S. W., Terentis, A. C., Osborne, R. L., Dawson, J. H., and Goodin, D. B. (2009). Replacement of the Axial Histidine Heme Ligand with Cysteine in Nitrophorin 1: Spectroscopic and Crystallographic Characterization. *J. Biol. Inorg. Chem.* 14, 179–191. doi:10.1007/s00775-008-0436-x

Wan, Q., and Danishefsky, S. J. (2007). Free-radical-based, Specific Desulfurization of Cysteine: A Powerful advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem. Int. Ed.* 46, 9248–9252. doi:10.1002/anie.200704195

Wöhr, T., Wahl, F., Nefzi, A., Rohwedder, B., Sato, T., Sun, X., et al. (1996). Pseudo-prolines as a Solubilizing, Structure-Disrupting protection Technique in Peptide Synthesis. *J. Am. Chem. Soc.* 118, 9218–9227. doi:10.1021/ja961509q

Yim, S. H., Everley, R. A., Schildberg, F. A., Lee, S.-G., Orsi, A., Barbati, Z. R., et al. (2018). Role of Selenof as a Gatekeeper of Secreted Disulfide-Rich Glycoproteins. *Cel Rep.* 23, 1387–1398. doi:10.1016/j.celrep.2018.04.009

Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P., and Liu, L. (2013). Chemical Synthesis of Proteins Using Peptide Hydrazides as Thioester Surrogates. *Nat. Protoc.* 8, 2483–2495. doi:10.1038/nprot.2013.152

Zheng, X., Ren, B., Li, X., Yan, H., Xie, Q., Liu, H., et al. (2020). Selenoprotein F Knockout Leads to Glucose and Lipid Metabolism Disorders in Mice. *J. Biol. Inorg. Chem.* 25, 1009–1022. doi:10.1007/s00775-020-01821-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 04 August 2021
doi: 10.3389/fchem.2021.733550

Designed Artificial Protein Heterodimers With Coupled Functions Constructed Using Bio-Orthogonal Chemistry

Rachel L. Johnson[1], *Hayley G. Blaber*[1,2], *Tomas Evans*[1], *Harley L. Worthy*[1,2], *Jacob R. Pope*[1] *and D. Dafydd Jones*[1]*

[1]*Molecular Biosciences Division, School of Biosciences, Cardiff University, Cardiff, United Kingdom,* [2]*The Henry Wellcome Building for Biocatalysis, Exeter University, Exeter, United Kingdom*

The formation of protein complexes is central to biology, with oligomeric proteins more prevalent than monomers. The coupling of functionally and even structurally distinct protein units can lead to new functional properties not accessible by monomeric proteins alone. While such complexes are driven by evolutionally needs in biology, the ability to link normally functionally and structurally disparate proteins can lead to new emergent properties for use in synthetic biology and the nanosciences. Here we demonstrate how two disparate proteins, the haem binding helical bundle protein cytochrome b_{562} and the β-barrel green fluorescent protein can be combined to form a heterodimer linked together by an unnatural triazole linkage. The complex was designed using computational docking approaches to predict compatible interfaces between the two proteins. Models of the complexes where then used to engineer residue coupling sites in each protein to link them together. Genetic code expansion was used to incorporate azide chemistry in cytochrome b_{562} and alkyne chemistry in GFP so that a permanent triazole covalent linkage can be made between the two proteins. Two linkage sites with respect to GFP were sampled. Spectral analysis of the new heterodimer revealed that haem binding and fluorescent protein chromophore properties were retained. Functional coupling was confirmed through changes in GFP absorbance and fluorescence, with linkage site determining the extent of communication between the two proteins. We have thus shown here that is possible to design and build heterodimeric proteins that couple structurally and functionally disparate proteins to form a new complex with new functional properties.

Keywords: genetic code expansion, bioorthogonal chemistry, azide alkyne cycloaddition, artificial protein oligomer, protein design and engineering, fluorescent proteins, energy transfer

OPEN ACCESS

Edited by:
Tao Peng,
Peking University, China

Reviewed by:
Jiantao Guo,
University of Nebraska-Lincoln, United States
Zhenrun J. Zhang,
University of Chicago, United States

***Correspondence:**
D. Dafydd Jones
jonesdd@cardiff.ac.uk

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *30 June 2021*
Accepted: *22 July 2021*
Published: *04 August 2021*

Citation:
Johnson RL, Blaber HG, Evans T, Worthy HL, Pope JR and Jones DD (2021) Designed Artificial Protein Heterodimers With Coupled Functions Constructed Using Bio-Orthogonal Chemistry. Front. Chem. 9:733550. doi: 10.3389/fchem.2021.733550

INTRODUCTION

Protein oligomerisation, commonly referred to as protein quaternary structure, is the association of specific individual polypeptide chains through defined intermolecular interactions to form a single multimeric complex (Goodsell and Olson, 2000; Nooren and Thornton, 2003; Ali and Imperiali, 2005). So prevalent is oligomerisation in nature, protein oligomers are more common than their monomeric counterparts, at least in the protein data bank (Goodsell and Olson, 2000; Ali and

Imperiali, 2005). Protein oligomers can comprise purely of non-covalent intermolecular interactions or also utilise inter-subunit covalent crosslinking, predominantly via disulphide bridges. Oligomerisation is largely seen as beneficial by reducing surface residues' (especially hydrophobics) exposure to solvent, resulting in a lower surface area to volume ratio leading to improved stability against degradation and aggregation (Larsen et al., 1998; Ali and Imperiali, 2005; Gwyther et al., 2019). Crucially, oligomerisation also leads to functional features not available in monomers; they locally concentrate multiple active sites resulting in improved activity and enabling functional cooperativity whereby synergy [communication] between each polypeptide unit can positively or negatively regulate activity or even lead to new properties (Goodsell and Olson, 2000; Gwyther et al., 2019).

Given the benefits of protein oligomerisation, protein designers and engineers have sort to address this area by constructing bespoke, artificial protein oligomeric systems (Oohora and Hayashi, 2014; Norn and Andre, 2016; Kobayashi and Arai, 2017; Ljubetic et al., 2017; Gwyther et al., 2019) ranging from simple dimers (Mou et al., 2015; Fallas et al., 2017) to higher order supramolecular structure (Thomson et al., 2014; Gonen et al., 2015; Butterfield et al., 2017). The main problem with these systems is that they lack functional synergy or even function beyond directing the assembly process. Thus, such systems tend to manifest the basic functional properties of their starting components. Recently, we have demonstrated functional synergy between designed artificial dimers (Worthy et al., 2019; Pope et al., 2020). Using computational approaches, we identified mutually compatible interfaces between various β-barrel fluorescent proteins and stabilised the interaction by genetically encoded click chemistry. Our approach allowed us to generate both symmetrical and non-symmetrical dimers together with homo and heterodimers that displayed either positive or negative functional synergy. While dimers represent the simplest protein oligomeric unit, they are the most frequently observed structural form in nature, with homo-dimers (comprising of the same polypeptide) dominating over hetero-dimer (composed of two different polypeptides) (Goodsell and Olson, 2000; Marianayagam et al., 2004; Mei et al., 2005).

The next challenge arguably involves the generation of heterodimers from structurally and functionally diverse proteins. Heterodimers are potentially rich in new functional features as they have the potential to combine drastically different and disparate functions leading to new emergent properties. Here we aim to take the next step in dimer construction and demonstrate that it is feasible to design and build intimately linked heterodimers comprised of structurally and functionally disparate proteins by linking a helical bundle protein to a β-barrel protein using a combination of computational protein design and bio-orthogonal chemistry. While bioorthogonal chemistry has been used previously to link different proteins together (Hatzakis et al., 2006; Eger et al., 2010; Hudak et al., 2012; Schoffelen et al., 2013; Torres-Kolbus et al., 2014; Kim et al., 2016; White and Bode, 2018), little effort is given to predicting compatible interfaces and residues pairs so the individual proteins generally remain functionally and structurally distinct. Furthermore, extended chemical linkers are routinely used as part of the chemical coupling process resulting in spatially separated protein units and preventing any interactions forming that normally comprise natural protein dimers (Hudak et al., 2012; Kim et al., 2012; Schoffelen et al., 2013; White and Bode, 2018). Thus, there is limited benefit above that of traditional genetic fusion approaches. Some rely on natural amino acid chemistry inherent to one partner protein to facilitate linkage (Hatzakis et al., 2006; Schoffelen et al., 2013; Torres-Kolbus et al., 2014), which can restrict coupling sites between individual units. To address these issues, we recently developed a computational design approach to facilitate the design of dimers linked by two genetically encoded compatible bio-orthogonal reaction handles (azide and alkyne) (Worthy et al., 2019; Pope et al., 2020). We focused on structurally similar proteins [fluorescent proteins with a common β-barrel architecture] whereby mutually compatible symmetrical dimer interfaces can facilitate construction. Such symmetry will not be available when structurally distinct proteins are used as proposed here and thus poses a greater challenge to predicting mutually compatible coupling sites and thus heterodimer construction.

To test our approach, we will use cytochrome b_{562} (cyt b_{562}) as the helical bundle protein and the superfolder version of green fluorescent protein (sfGFP) as the β-barrel protein (**Figure 1A**). Cyt b_{562} is a small 4-helical bundle protein that binds haem tightly but non-covalently (Arnesano et al., 1999) while sfGFP is a directly evolved descendent of the original *Aequorea victoria* GFP (Pedelacq et al., 2006). We have previously linked the function of the closely related homolog of sfGFP, enhanced GFP (EGFP), with cyt b_{562} through a directed evolution domain insertion approach (Edwards et al., 2008; Arpino et al., 2012). The domain insertion approach differs significantly from heterodimerization as a single polypeptide unit contains both original proteins; the GFP primary structure is disturbed by insertion of the cyt b_{562} sequence within it at the genetic level. One variant demonstrated high energy transfer efficiency (close to 100%). Structural analysis revealed that the two original proteins are close in space within the single polypeptide unit but do not form many distinct inter-unit interactions as would be expected of a true oligomeric system. There was also little change in the inherent function of each domain compared to the starting parent protein suggesting limited synergy between the two.

Here, we used computational approaches to predict compatible interfaces between sfGFP and cyt b_{562}. We then used 1-to-1 strain promoted alkyne-azide cycloaddition to covalently link the two proteins through a triazole link by genetically encoded incorporation of the azide group (via non-canonical amino acid azidophenylalanine) into cyt b_{562} and a strained cyclooctyne group (via the non-canonical amino acid strained-cyclooctyne lysine) in sfGFP. The successfully constructed heterodimers lead to enhanced sfGFP molar absorbance coefficients and were capable of energy transfer from sfGFP to cyt b_{562}, with linkage site determining transfer efficiency.

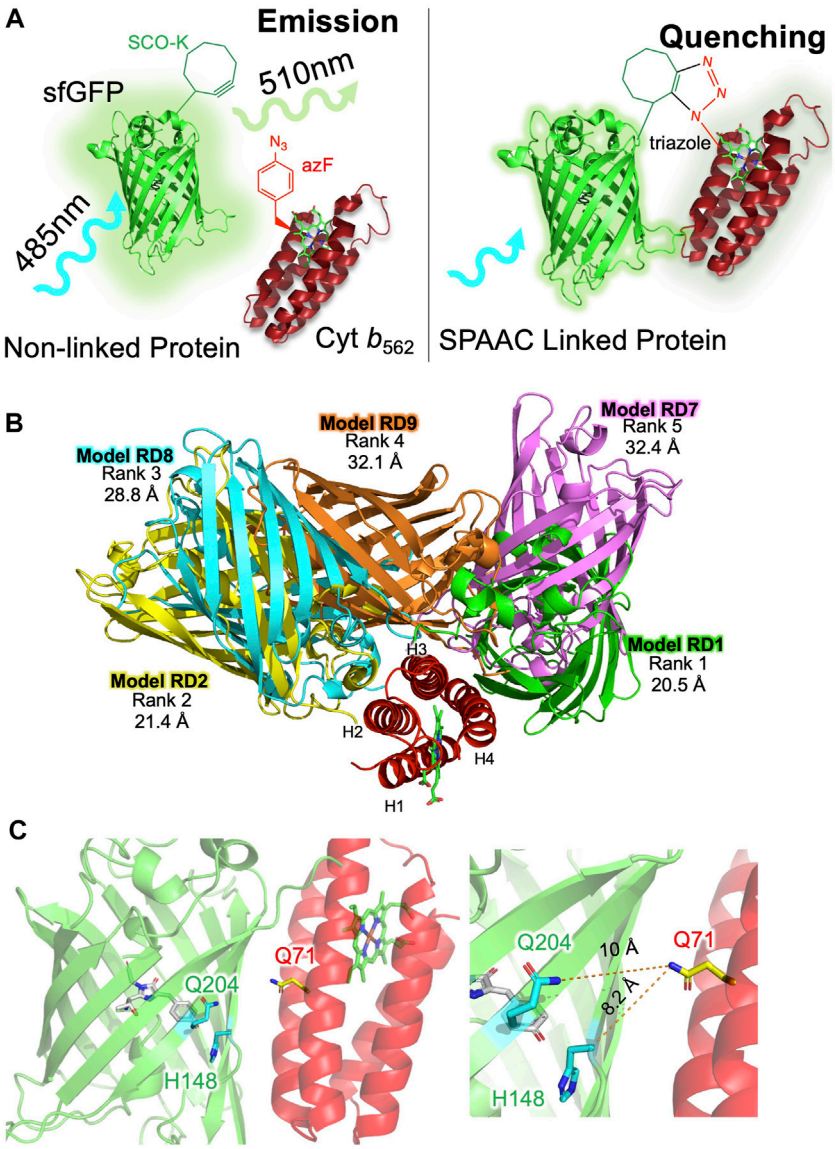


FIGURE 1 | Designing sfGFP-cyt b_{562} SPAAC linked heterodimers. **(A)** Basic strategy for constructing heterodimers using SPAAC. As a monomeric protein, sfGFP (green) will excite at ~485 nm and then emit light at ~510 nm. On forming a heterodimer, sfGFP fluorescence will predominantly be quenched by cyt b_{562} (red). The complex will only persist if SPAAC is successful and a triazole link is formed. **(B)** The top five alignments of sfGFP (various colours) and cyt b_{562} (red) ranked according to total energy of interaction by RosettaDock. Distances measured between each model chromophore pair are shown below the rank. Model RD1 has the closest chromophore proximity and lowest total energy. **(C)** Residue selected for mutation to SCO-K (Q204 and H148; green) in sfGFP or AzF (Q71; red) in cyt b_{562}.

METHODS AND MATERIALS

In Silico Docking

The combined ClusPro (Kozakov et al., 2017) and RosettaDock (Leaver-Fay et al., 2011; Alford et al., 2017) approach has been described previously (Worthy et al., 2019). Briefly, the ClusPro Protein-Protein server (https://cluspro.bu.edu) "Dock" function was utilised to predict potential interface sites between proteins. ClusPro simulation was run with cyt b_{562} (PDB 1qpu (Arnesano et al., 1999)) as the receptor and sfGFP (PDB 2b3p (Pedelacq et al., 2006)) as the freely rotating ligand. The models were ranked by the number of clustered simulations as calculated in balanced simulation mode (recognition of all intermolecular electrostatics), and the top ten ranked interfaces were downloaded as PDBs from the server for further analysis. Dynamic modelling of the molecular interfaces was further refined in ROSETTA's high-resolution docking protocol (Liu and Kuhlman, 2006; Lyskov and Gray, 2008). The RosettaDock protocol ranked the predicted interfaces obtained within ClusPro simulation by both "Total Energy" and "Interface Energy". Both the haem of cyt b_{562} and the chromophore of sfGFP were reintroduced back into each structure using PyMOL. The predicted heterodimer structure

obtained by this method were therefore taken as the most energetically favourable and the top ranked model inspected in PyMOL for suitably close partner residues.

Protein Engineering and Recombinant Production

The SCO-K (see **Supplementary Figure S1A** for chemical structure) containing sfGFP variants were generated and produced as described previously (Worthy et al., 2019; Pope et al., 2020). The gene encoding cyt b_{562} was present in the pBAD plasmid. The cyt b_{562}^{50AzF} variant was generated and produced as described previously (Zaki et al., 2018; Thomas et al., 2020). The cyt b_{562}^{71AzF} variant was generated by introducing a TAG codon in place of the Q71 encoding codon by whole plasmid PCR (Forward primer 5′-C GGT **TAG** ATT GAC GAC G-3′ and reverse primer 5′-AC CAG AAT GTC GAA ACC G-3). Incorporation of AzF (see **Supplementary Figure S1A** for chemical structure) into cyt b_{562} and its subsequent purification was achieved as described previously (Zaki et al., 2018; Thomas et al., 2020) using the pDULE plasmid (Miyake-Stoner et al., 2010). After cell lysis and removal of cellular debris, the soluble lysate was subjected to 30% (w/v) ammonium sulphate precipitation to remove some contaminant protein as precipitant. The soluble supernatant was subjected to a further round of ammonium sulphate precipitation to a concentration of 90% (w/v) to precipitate all protein. The precipitated lysate was resuspended 50 mM Tris pH 8.0 and if desired was mixed with 100x molar excess of haem porphyrin to generate of holo-cyt b_{562}^{71AzF}. Protein was applied to Hiload™ 16/600 Superdex™ S75. Fractions containing cyt b_{562} were applied to a Sepharose Q anion exchange column before final buffer exchange into 50 mM Tris pH 8.0 using a PD-10 desalting column. Pure cyt b_{562}^{71AzF} was concentrated to 100 μM and separated into 100 μL samples before flash freezing and storage at −80°C until use. Apo-cyt b_{562}^{71AzF} was generated by haem extraction essentially as described elsewhere (Jones and Barker, 2004; Bowen et al., 2020).

Heterodimer Formation via SPAAC

Strain promoted azide-alkyne cycloaddition of protein was achieved by mixing of SCO-K containing sfGFP (50 μM) with cyt b_{562}^{71AzF} (50 μM) at 37°C overnight. After incubation, the formation of oligomeric protein was determined by SDS-PAGE. Purification of oligomeric protein was achieved by size exclusion chromatography using Hiload™ 26/600 Superdex™ S200 gel filtration column (Section 2.4.2) and the purity of the resultant protein oligomer was assessed by SDS PAGE. Yield of dimer was estimated by ImageJ (Schindelin et al., 2012) analysis of band intensity of each form after Coomassie staining of polyacrylamide gels.

Absorbance and Fluorescence Spectroscopy

UV-visible absorption spectra were recorded with Cary 60 spectrophotometer using 1 cm pathlength quartz cuvettes. Absorbance was recorded between 200–800 nm at a scan rate of 300 nm/min. The molar absorbance co-efficient (ε) for the sfGFP variants have been determined previously (Worthy et al., 2019). The concentration of cyt b_{562}^{71AzF} was determined using the DC-protein assay (BioRad) using the wild-type cyt b_{562} as the standard. The subsequent molar absorbance co-efficient was calculated using the Beer-Lambert law with the absorbance of a known concentration of cyt b_{562}^{71AzF}. Fluorescence emission spectra were recorded on a Cary Eclipse Fluorimeter, using 5 mm × 5 mm QS quartz cuvette. Samples were excited at the λ_{max} and emission was recorded at every 1 nm from the point of excitation to 700 nm. A scan rate of 120 nm/min was used for all spectra recorded with a 5 nm slit width and voltage set to medium. To measure the emission in reducing conditions protein was first incubated with a 10-fold molar excess of DTT.

RESULTS

In Silico Prediction of sfGFP-cyt b_{562} Interface

The first step in heterodimer design is to predict compatible interfaces between the two starting proteins. Previous work has shown that not all surface exposed residues are amenable to dimer formation via SPAAC (Worthy et al., 2019). We show the same appears true here. Incorporation of strained-cyclooctyne-lysine (SCO-K) in sfGFP at residue 204 is known to promote dimerisation with a mutually compatible interface (Worthy et al., 2019; Pope et al., 2020). Incorporation of azidophenylalanine (AzF) at residue 50 in cyt b_{562} places the non-canonical amino acid (ncAA) within a dynamic extended surface loop (**Supplementary Figure S1B**). Previous work has shown that cyt b_{562}^{50AzF} is amenable to chemical functionalisation with non-biological entities (Zaki et al., 2018; Thomas et al., 2020). As shown in **Supplementary Figure S1C**, protein dimerisation did not occur to any great extent.

To predict potentially compatible interfaces between cyt b_{562} and sfGFP, we used an *in silico* docking approach developed recently for constructing fluorescent protein dimers (Worthy et al., 2019; Pope et al., 2020). The first step uses ClusPro (Kozakov et al., 2017) to generate unbiased docking of sfGFP (PDB 2b3p (Pedelacq et al., 2006)) to cyt b_{562} (1qpu (Arnesano et al., 1999)). Haem and the sfGFP chromophore are automatically removed leaving the core structures intact relative to the starting structure. Of the 30 alignments generated, the top 10 models ranked according to cluster number were further analysed (see **Supplementary Figure S2** for alignments and cluster information). None included an interface involving residue 50 in cyt b_{562}. To provide a more quantitative analysis and refine the docking procedure, each initial model was further assessed and ranked using RosettaDock (Leaver-Fay et al., 2011; Alford et al., 2017) through the generation of an estimated energy of the interface and total energy between the molecules. The top five models are shown in **Figure 1B** with the energies in **Supplementary Table S1**. The distances between the two chromophores varied from 20.5 Å in the highest ranked model (RD1) to 32.4 Å. While model RD9 had the marginally lowest interface energy, RD1 had the

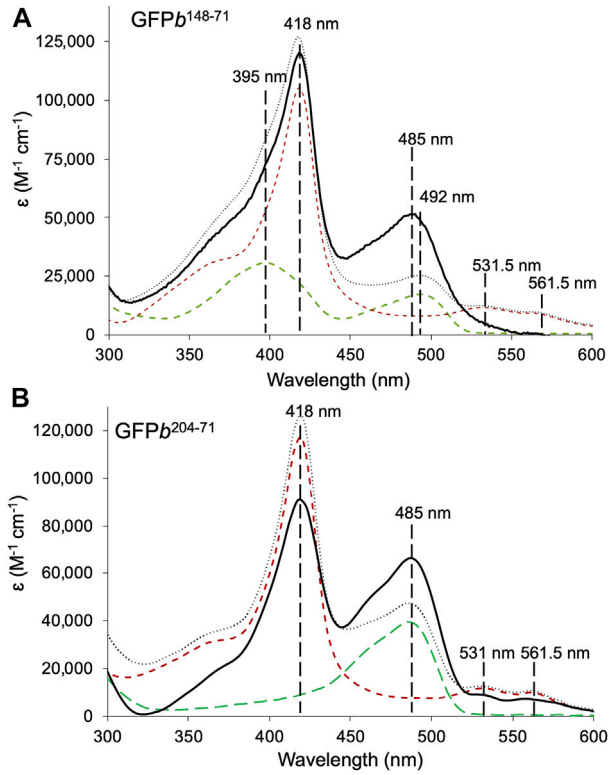


FIGURE 2 | Absorbance spectra of sfGFP-cyt b_{562} heterodimers. **(A)** GFP$b^{148\text{-}71}$ (black solid line), and constituent monomers sfGFP^{148SCO} (green dashed line) and cyt $b_{562}{}^{71azF}$ (red dashed line). **(B)** GFP$b^{204\text{-}71}$ (black solid line), and constituent monomers sfGFP^{204SCO} (green dashed line) and cyt $b_{562}{}^{71azF}$. The combined sum of each monomeric absorbance is shown for comparison (grey dotted line). Data shown as molar extinction coefficients. The wavelengths associated with the λ_{max} for each monomer absorbance peak is shown for reference. The molar absorbance coefficients are shown in **Supplementary Table S2**.

lowest total energy and was derived from the largest cluster number (CP1 in **Supplementary Figure S1**). Thus, RD1 was taken forward as the primary model to base the design of SPAAC linkage sites.

Analysis of RD1 revealed that two previous residues in sfGFP known to successfully promote dimerisation via SPAAC, 148 and 204 (Worthy et al., 2019), were located close to the dimer interface (**Figure 1C**). Residue H148 is critical to sfGFP function as it helps define the fluorescent properties through H-bonding to the chromophore's phenolic group. Changing H148 to a ncAA is tolerated and changes the inherent fluorescence properties (Reddington et al., 2013; Hartley et al., 2016; Worthy et al., 2019). Q204 to is also tolerant to ncAA incorporation (Reddington et al., 2012; Worthy et al., 2019). Thus, these residues were selected within the context of sfGFP. In cyt b_{562} helices three and four comprised the main docking interface; residue Q71 in helix three was chosen as it was close to both residues 148 and 204 in sfGFP (**Figure 1C**). The SCO-K ncAA was previously incorporated into sfGFP residues 148 (sfGFP^{148SCO}) and 204 (sfGFP^{204SCO}) and characterised (Worthy et al., 2019). Incorporation of azF into cyt b_{562} at residue 71 (cyt $b_{562}{}^{71azF}$) in response to a TAG codon has been demonstrated previously as part of a separate directed evolution codon exchange study (Arpino et al., 2015). Cyt $b_{562}{}^{71azF}$ produced here has similar spectral characteristic to that of wild-type cyt b_{562} (**Supplementary Figure S3**).

Cycloaddition of sfGFP and Cyt b_{562} and Its Impact of Absorbance

Analysis by SDS-PAGE revealed that dimerisation of cyt $b_{562}{}^{71azF}$ with sfGFP^{148SCO} or sfGFP^{204SCO} were successful, with yields in the range of 20–35% (see **Supplementary Figure S4** for representative SDS-PAGE gels). The yields are slightly lower compared SPAAC based dimerisation of structurally similar proteins (35–80%) (Worthy et al., 2019). The two new heterodimers termed GFP$b^{148\text{-}71}$ and GFP$b^{204\text{-}71}$ were isolated from their monomeric forms by size exclusion chromatography. The absorbance spectra have characteristics of both constituent proteins with major peaks at 418 nm equivalent to cyt b_{562} and ~485 nm contributed by sfGFP (**Figure 2**). The absorbance spectra of the heterodimers also indicate positive functional changes on dimerisation with respect to the sfGFP unit (**Figure 2** with molar extinctions provided in **Supplementary Table S2**).

In terms of GFP$b^{148\text{-}71}$, compared to the starting monomers, the absorbance peak at ~485 nm associated with sfGFP^{148SCO} increases by just over 3-fold in the heterodimer with a concomitant drop in the shoulder at 395 nm (**Figure 2A**). Such spectral characteristics are associated with a switch in the protonation state of sfGFP chromophore; the neutral phenol chromophore has a peak absorbance of ~400 nm and the phenolate anionic form absorbs at ~485 nm (Remington, 2011). Simple addition of the monomer absorbance spectra confirms the promotion of the anionic sfGFP chromophore form rather than any baseline addition from the cyt b_{562} unit. Promotion of the anionic chromophore has been shown previously for symmetrically arranged sfGFP homodimers linked by residue 148 (Worthy et al., 2019). Thus, modulation of the sfGFP chromophore charged state and hence function is still feasible when linked to a very distinct partner protein.

GFP$b^{204\text{-}71}$ also shows a significant increase in absorbance corresponding the sfGFP chromophore, with molar absorbance at 485 nm almost doubling (**Figure 2B**). The simple monomer addition spectrum confirms that the increase in the 485 nm absorbance is not due to an underlying contribution by cyt b_{562}. Thus, the ability of sfGFP to interact with light has been enhanced on dimerisation. Unlike sfGFP^{148SCO}, sfGFP^{204SCO} exists predominantly in the anionic state so the increase in 485 nm absorbance is not down to change in chromophore ionisation state. We have previously proposed that such a positive synergistic effect is due to reduced water dynamics in channels leading to the sfGFP chromophore when homo-dimerisation occurs via residue 204 (Pope et al., 2020); the same may also be occurring here.

Functional Communication in the Heterodimers

Haem can quench fluorescence by resonance energy transfer, providing the fluorophore is within close proximity (Willis et al.,

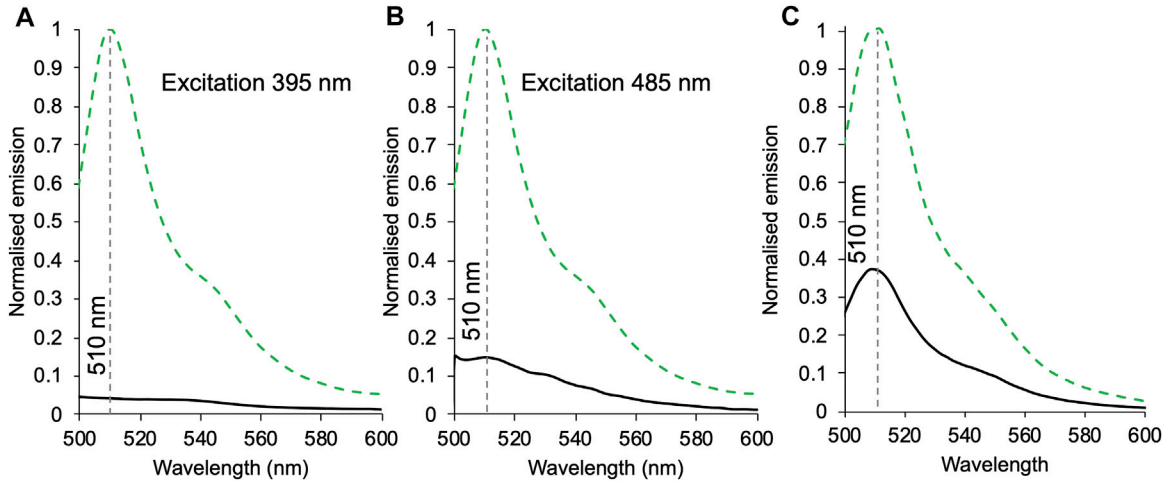


FIGURE 3 | The effect of heterodimerisation on fluorescence emission. Emission of 1 μM of sfGFP^{148SCO} (green dashed) or GFPb^{148-71} (black) excited at either 395 nm **(A)** or 485 nm **(B)**. **(C)** Emission of 1 μM sfGFP^{204SCO} (green dashed) and GFPb^{204-71} (black) on excitation at 485 nm. Spectra were normalised to sfGFP^{148SCO} **(A, B)** or sfGFP^{204SCO} **(C)**.

1990; Takeda et al., 2001; Arpino et al., 2012). The requirement for close proximity is shown in **Supplementary Figure S5**, where free haem or free cyt b_{562} do not quench sfGFP to any appreciable extent. To assess communication between sfGFP and cyt b_{562} in our heterodimers, fluorescence was measured on excitation at the major absorbance peaks of sfGFP (**Figure 3**). As GFPb^{148-71} has two potential absorbance peaks, excitation was performed at both 395 and 485 nm (**Figures 3A,B**). On excitation at 395 nm, emission is reduced by 96% compared to the monomer, and on excitation at 485 nm emission was reduced by 85%. The difference between the two may be due to the relative change in absorbance at each wavelength on conversion from a monomer to heterodimer; cyt b_{562} may also absorb some of the 395 nm light (see **Figure 2A** for spectral overlap). This still represents a major reduction in fluorescence suggesting a high degree of energy transfer and thus communication between sfGFP and cyt b_{562} in the GFPb^{148-71} construct. By comparison, the drop in emission on excitation of GFPb^{204-71} is 67% (**Figure 3C**). This suggests that while energy transfer is still occurring the efficiency is reduced compared to GFPb^{148-71}. SEC clearly resolves the heterodimer from the monomers (**Supplementary Figure S6**) thus we do not believe contaminating monomeric sfGFP^{204SOC} is the cause of the residual fluorescence observed for GFPb^{204-71}.

Haem and Redox-State Dependent Quenching in GFPb^{204-71}

The iron centre of the cyt b_{562} haem group switches between the reduced Fe^{2+} and oxidised Fe^{3+} state that results in changes to the absorbance spectrum and affinity for the protein (Robinson et al., 1997; Wittung-Stafshede et al., 1999; Jones and Barker, 2005). The observed fluorescence emission from GFPb^{204-71} allows us to monitor how fluorescence output can be tuned to both haem binding and redox conditions. Conversion from oxidised to reduced haem was achieved through the addition of the reducing agent dithiothreitol (DTT). The cyt b_{562} unit in GFPb^{204-71} is still capable of redox state switching as shown by the switch in the 418 nm absorbance peak for the oxidised form to 426 nm characteristic of reduced cyt b_{562} with the typical higher molar absorbance (**Figure 4A** and **Supplementary Table S3**). The α/β band peaks also become more prominent as expected on conversion from oxidised to reduced cyt b_{562}, which in turn increases the spectral overlap between the sfGFP emission and cyt b_{562} absorbance (**Figure 4B**). As anticipated, the reducing agent had little impact on sfGFP absorbance (**Figure 4A**). On the addition of reducing agent, fluorescence emission dropped by an additional 31% (**Figure 4C**). Thus, GFPb^{204-71} output can respond to changes in redox conditions through coupling changes in the redox state of haem iron bound to cyt b_{562} to fluorescence output of sfGFP.

We thus attempted to extend this to see if the apo-heterodimer (no haem bound) could bind and respond to haem. Linking haem binding to cyt b_{562} with fluorescent protein output are potentially useful biosensors for this important biological co-factor (Takeda et al., 2003; Arpino et al., 2012; Hanna et al., 2016). The apo-GFPb^{204-71} responded to haem binding and subsequent switch in redox conditions. On addition of haem to apo-GFb^{204-71}, fluorescence emission dropped by 64% and by a further 12% (equivalent to a 33% compared to the oxidised holo-GFb^{204-41}) on addition of reducing agent (**Figure 4D**). These results are comparable to the results observed above for holo-GFb^{204-41} (when compared to monomeric GFP^{204SCO}; **Figure 3C**) suggesting under the conditions used, full haem binding has occurred to the apo-protein, and the heterodimer responds to change in redox conditions.

DISCUSSION

While traditionally protein engineering has focused on converting oligomeric proteins into monomers, especially with

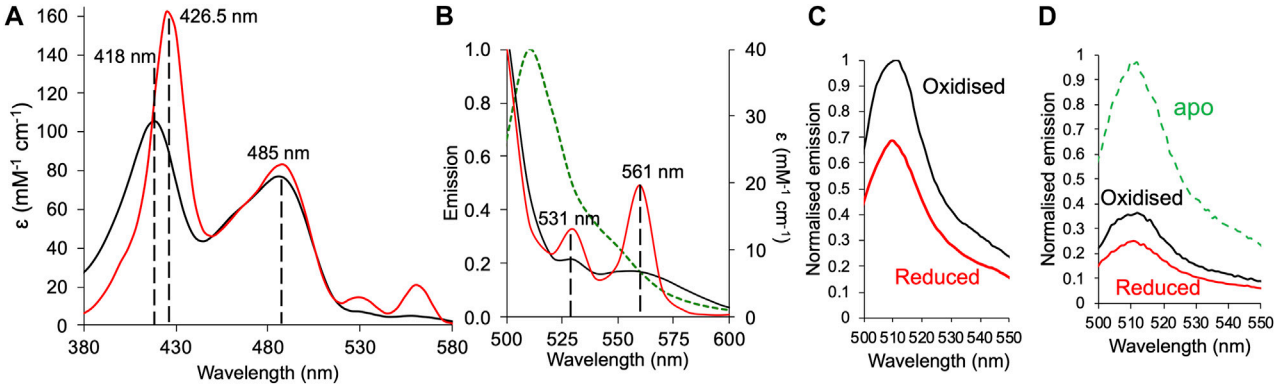


FIGURE 4 | Redox dependent fluorescence emission of GFP$b^{204\text{-}71}$. **(A)** Absorbance spectra of GFP$b^{204\text{-}71}$ under oxidising (black line) and reduced (red line). Reducing conditions stimulated by addition of a 10-fold molar excess of DTT. **(B)** Absorbance spectra under oxidising and reducing conditions shown in **(A)** for the α/β peak region overlaid with the emission spectrum (on excitation at 485 nm) of sfGFP (green dashed line). **(C)** Emission spectra of 1 μM GFP$b^{204\text{-}71}$ in the absence (oxidised; black) and presence (reduced; red) of DTT. Spectra were normalised to oxidised holo-GFP$b^{204\text{-}71}$. **(D)** The emission spectra of apo-GFP$b^{204\text{-}71}$ before the addition of haem under oxidising (black line) or reduced (red line) conditions. Spectra were normalised to apo-GFP$b^{204\text{-}71}$.

regards to fluorescent proteins (Zacharias et al., 2002; Scott et al., 2018), the design and construction of artificial protein oligomers or "supramolecular" systems is currently of great interest (Norn and Andre, 2016; Yeates et al., 2016; Kobayashi and Arai, 2017; Beesley and Woolfson, 2019; Gwyther et al., 2019). Both GFP (Ding et al., 2015; Kim et al., 2015; Leibly et al., 2015; Worthy et al., 2019; Arpino and Polizzi, 2020) and cyt b_{562} (Baldwin et al., 2006; Radford et al., 2011; Brodin et al., 2012; Onoda et al., 2012; Song and Tezcan, 2014; Oohora et al., 2018; Golub et al., 2020) have been central to exemplifying the ability to generate new oligomeric supramolecular protein structures (Oohora and Hayashi, 2014). Artificial protein oligomers offer the same potential impact as oligomerisation does in nature: the generation of complex higher-order structures from a limited monomeric building block repertoire. This in turn allows new structural and thus functional space to be sampled not accessible in monomeric proteins. Whether it is O_2 binding to haemoglobin (Ciaccio et al., 2008) or enzyme catalysis and allosteric regulation (Goodsell and Olson, 2000; Ali and Imperiali, 2005), one of the main benefits of oligomerisation is synergy between individual subunits. Such linked functionality is generally hard to design and engineer into artificial complexes due to the requirement of long-range interactions that link active sites. Thus, most designed protein complexes focus on the interface region. We have shown previously that such networks can be generated between fluorescent protein homo and heterodimers using an approach like that used here, which leads to functional switching and fluorescence enhancement (Worthy et al., 2019).

Here we decided to test our ability to select compatible interfaces between disparate proteins by choosing two monomeric proteins with very different structures: a largely β-sheet protein (sfGFP) and a helical protein (cyt b_{562}). The rationale for using SPAAC to covalently stabilise the heterodimer structure is that classical approaches such as disulphide bridges cannot discriminate between homo-dimers and heterodimers leading to a mixed population; the bioorthogonal nature of SPAAC means only heterodimer will form. Furthermore, the ncAAs used here have longer side chains than the short -CH_2-SH group of cysteine meaning steric clashes between monomers is less likely to inhibit covalent bond formation but still allow an intimate interaction between the individual monomer units. The triazole link is also more stable than a disulphide bridge. While our aim was not to generate a newly designed dimer interface mimicking natural protein-protein interactions, we did need to identify compatible interfaces that will at least persist for a length of time to allow covalent functionalisation. Without identifying compatible interfaces, the chances of stabilising the interface through SPAAC is minimal. Residue 50 in cyt b_{562} is largely surface exposed in a flexible extended loop and would normally be considered an ideal residue to target for covalent coupling with another protein (**Supplementary Figure S1B**) but SPAAC facilitated dimerisation was not possible via this residue (**Supplementary Figure S1C**). Previous work has shown that incorporation of azF at residue 50 is reactive (Zaki et al., 2018; Thomas et al., 2020). Lack of dimerisation is thus likely a result of the individual subunits unable to become spatially localised for long enough to promote SPAAC. This mirrors previous work with fluorescent proteins that suggested compatible interfaces are required for SPAAC facilitate dimerisation even when surface exposed residues are selected (Worthy et al., 2019)). Our computational design approach proved successful both here (**Figure 1** and **Supplementary Figure S4**) and elsewhere (Worthy et al., 2019; Pope et al., 2020) in identifying compatible interfaces. It also provides a rationale for the inability of cyt $b_{562}{}^{50azF}$ to form dimers with sfGFP as residue 50 is not involved in any of the predicted protein interfaces. The step forward here was using proteins with disparate structural folds. The overall total binding energy was higher here (less negative) than for structurally homologues proteins tested previously suggesting a weaker interaction (**Supplementary Figure S2** and **Supplementary Table S2** *versus* data in references (Worthy et al., 2019; Pope et al., 2020)) but still proved useful in generating models for predicting successful coupling sites (**Figure 1**). The model suggests that there are

intermolecular interactions between cyt b_{562} and sfGFP beyond residues involved in SPAAC, but it is not clear if they persist in the SPAAC link dimer. However, our previous structural work with SPAAC linked fluorescent proteins dimers shows that extensive intermolecular non-covalent interactions are formed at the interface, including long range interaction networks linking the active sites (Worthy et al., 2019; Pope et al., 2020). Thus, given the intimate nature of the coupling between sfGFP and cyt b_{562}, additional non-covalent interactions outside the triazole linkage between the two monomers on forming the dimer are highly likely to be present.

Two heterodimers were designed based on the top ranked model (**Figure 1C**). In terms of sfGFP, residues 148 and 204 were selected to host the strained alkyne ncAA (SCO-K). These sites have been shown previously to be amenable to both small molecule (Reddington et al., 2012; Hartley et al., 2016) and protein (Worthy et al., 2019; Pope et al., 2020) attachment via SPAAC. H148 in sfGFP forms a H-bond with the chromophore, which directly impacts on the fluorescence properties of the protein by assisting in deprotonation of the chromophore's phenol group. In many crystal structures, H148's side chain is largely buried but is known to be dynamic (Seifert et al., 2003) with the "flipped out" conformation observed in some crystal structure populations, resulting in the residue becoming solvent exposed (Reddington et al., 2013; Arpino et al., 2014). Changing H148 to a larger ncAA results in breakage of a critical H-bond with the chromophore and exclusion from the core of the protein due to steric clashes (Hartley et al., 2016); the residue now amenable to chemical modification. The result is a mixed population of the protonated and anionic form of the sfGFP chromophore prior to SPAAC modification (**Figure 2A**). We have previously shown that we can control the relative populations of two chromophore states causing switching of the fluorescent properties in terms of excitation wavelength (Reddington et al., 2013; Hartley et al., 2016; Worthy et al., 2019). We have successfully achieved the same here through the formation of a heterodimer with promotion of the anionic form on dimerisation (**Figure 2A**). As mutation of H148 to a ncAA removes the H-bond to the sfGFP chromophore critical to formation of the phenolic anion, we have proposed previously that a structural water molecule replaces the imidazole group and plays the role of the H-bond acceptor that promotes ionisation of the chromophore (Worthy et al., 2019; Pope et al., 2020); the same scenario may also be the case here with a water molecule trapped at the sfGFP-b_{562} interface in a position to H-bond to the chromophore.

With respect to GFP$b^{204\text{-}71}$, we see enhancement of the sfGFP molar absorbance (**Figure 2B**). Unlike H148, Q204 is surface exposed and plays little role in dictating the fluorescence properties of sfGFP, even when replaced by an ncAA (Reddington et al., 2012; Worthy et al., 2019). Water dynamics is again thought to play a major role. Dimerisation of sfGFP with itself or closely related fluorescent proteins enhances molar absorbance and structural analysis revealed ordered water molecules at the dimer interface, including waters comprising a channel through to the chromophore (Pope et al., 2020). The same may be happening with GFP$b^{204\text{-}71}$ with the dimer interface trapping water molecules leading to a sustained water-protein bond network that improves the ability of the sfGFP chromophore to interact with light.

Bringing the sfGFP chromophore within close proximity to haem through formation of the heterodimer should result in energy transfer from sfGFP that is quenched by cyt b_{562}, a feature that has been observed before for single polypeptide systems (Takeda et al., 2003; Arpino et al., 2012; Hanna et al., 2016). The extent of quenching is related to the distance between the two chromophores (Arpino et al., 2012). Quenching in a classical N- or C-terminal fusion of EGFP to cyt b_{562} was less than 65% (Arpino et al., 2012). Functional communication was present in both our heterodimers, but the efficiency of energy transfer differed depending on the SPAAC linkage positions, haem redox state and, to an extent, the excitation wavelength (**Figure 3**). This is despite the two residues being adjacent to each other in the structure. Linkage via sfGFP residue 148 had the highest energy transfer efficiency (>85%) with more apparent efficient energy transfer on excitation 395 nm. Linkage via sfGFP residue 204 resulted in a lower energy transfer (63%), comparable to previous "head-to-tail" fusions with cyt b_{562} where the two chromophores are not anticipated to be close in space (Arpino et al., 2012). Thus, the question arises as to why different linkage sites that are close together in sfGFP (**Figure 1C**) generate very different energy transfer efficiencies? Given that GFP$b^{204\text{-}71}$ can be separated from its monomeric components by size exclusion chromatography (**Supplementary Figure S6**), the residual fluorescence is unlikely to be from contaminating sfGFP^{204SCO} monomer. We have shown previously that linkage via residue 204 in fluorescent protein dimers resulted in lower than predicted energy transfer (Pope et al., 2020). The exact reason was not clear. The calculated R_0 (the Förster radius at which energy transfer is 50% efficient) between EGFP and cyt b_{562} is 46 Å (Takeda et al., 2001). As energy transfer is related to the donor-acceptor chromophore distance (r) and energy transfer (E) through equation $E = 1/[1+(r/R_0)^6]$, we can estimate the distance between the chromophores. Previously, we showed that there was a good correlation between r from energy transfer (99%) and structure in our domain insert EGFP-cyt b_{562} construct (**Figure 5**). Based on energy transfer efficiency, in GFP$b^{148\text{-}71}$ the closest interchromophore distance is predicted to be ~27 Å while for GFP$b^{204\text{-}71}$ it is ~41 Å. The interchromophore distances derived from the original top ranked model suggest that it might provide a realistic representation for GFP$b^{148\text{-}71}$ but not GFP$b^{204\text{-}71}$ (**Figure 5**). As shown in **Figure 5**, the domain arrangement in our heterodimer model compared to the determined structure of the high energy transfer efficient domain insert protein is different; the interchromophore distance is 8 Å longer thus could account for the slightly reduced energy transfer efficiency in GFP$b^{148\text{-}71}$. Residue 204 is close to the interface rather than directly forming the interface, which may result in the two monomers adjusting their relative placement on SPAAC compared to that of the model. There are two isomeric forms for the triazole crosslink between AzF and SCO-K (Worthy et al., 2019; Pope et al., 2020): the *syn* isomer that forms a turn structure and the *anti* form that results in an elongated linkage (**Figure 5B**). The *anti* form may dominate for GFP$b^{204\text{-}71}$ resulting in the two monomers and thus chromophores being separated by a longer

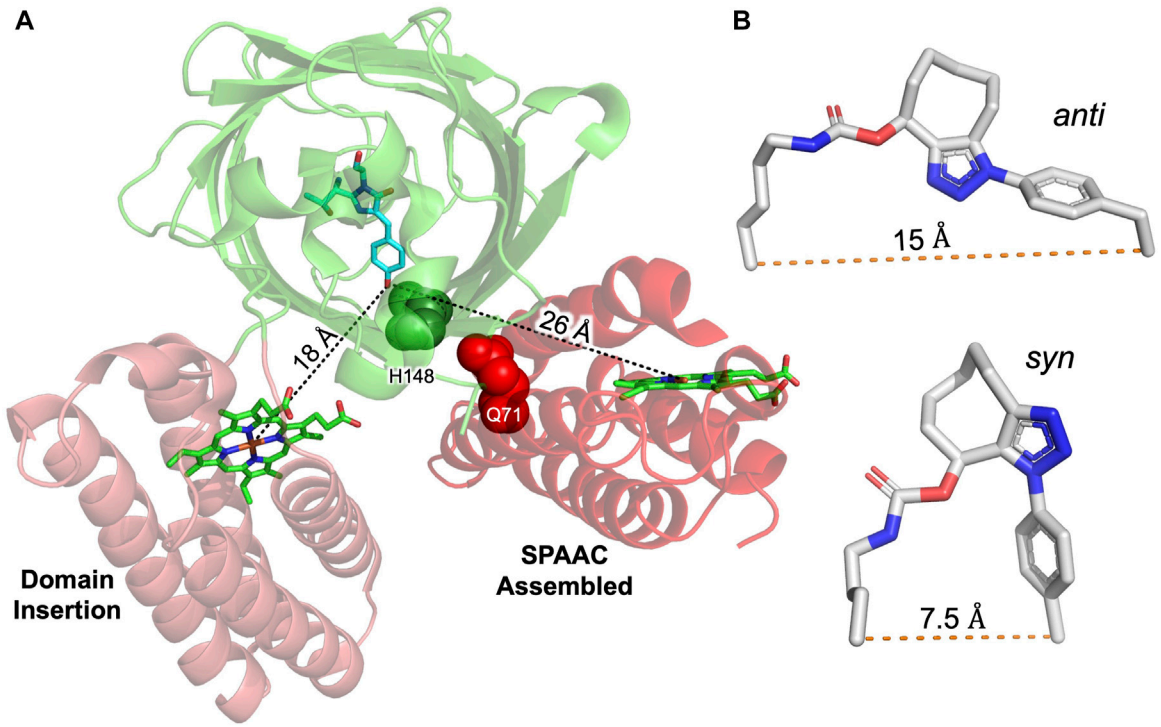


FIGURE 5 | Comparison of chromophore distances between GFP and cyt b_{562} constructed by domain insertion or SPAAC assembly. Alignment of crystal structure (3u8p) of EGFP-cyt b_{562} CG6 domain insert variant (Arpino et al., 2012) with the highest ranked model of sfGFP (2b3p) and cyt b_{562} (1 qpu) obtained by RosettaDock. GFP (green) used as centre of alignment with chromophore shown (cyan sticks). The measured distance between the GFP chromophore phenolic hydroxyl group and central iron of each haem chromophore are shown associated with the black dotted lines. Residues mutated to ncAAs are shown as spheres and labelled. **(B)** The *anti* (derived from PDB 5 nhn (Worthy et al., 2019))and *syn* (derived from PDB 5ni3 (Pope et al., 2020)) configurations around the triazole bond with the relative distances between the C_α of the two original amino acids shown.

distance. Interestingly, energy transfer efficiency was oxidation state-dependent (**Figures 4C,D**). The ability apo-GFP$b^{204\text{-}71}$ to bind haem suggests that the haem binding site is still accessible as predicted by our original modelling and on binding haem under non-reducing conditions exhibits a similar drop in fluorescence to that observed for holo-GFP$b^{204\text{-}71}$ suggesting full haem occupancy (**Figures 3C**, **4D**). On addition of reducing agent, fluorescence decreased by another third suggesting conversion from Fe^{3+} to Fe^{2+} resulted in increased energy transfer efficiency. DTT is not known to reduce the fluorescence emission spectra of sfGFP (**Supplementary Figure S7** and (Reddington et al., 2015)) so the increased energy transfer efficiency may due to inherent change in the protein-bound haem such as increased molar absorbance of the α/β peaks that overlap with the sfGFP emission (**Figure 4B**). Even under reducing conditions, energy transfer efficiency is still ~75% equating to an interchromophore distance of 38 Å. Thus, we cannot rule out other currently unknown events contributing to the reduced energy transfer efficiency in GFP$b^{204\text{-}71}$, as also observed in fluorescent protein dimers linked via residue 204 (Pope et al., 2020).

To conclude, here we have shown that it is feasible to design and construct artificial heterodimers between structurally and functionally disparate proteins linked by a genetically encoded bio-orthogonal link. The *in silico* docking helped identify suitability compatible protein-protein interfaces that where then stabilised by a triazole link formed by SPAAC. In both heterodimer configurations tested, the ability of phenol anion chromophore form of the sfGFP to interact with light was enhanced, with energy transfer to the haem centre of cyt b_{562} demonstrating functional linkage.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://doi.org/10.17035/d.2021.0080088158.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing of this manuscript. RJ contributed to all in silico design, experimental work and analysis on both heterodimer systems. HB performed some experimental work and analysis on the GFP*b*148-71 system. TE performed initial in silico modelling, experimental work and analysis on the GFP*b*204-71. JP undertook the initial dimerisation work with cyt b_{562}^{50AzF}. HW helped with the *in silico* docking. The idea was conceived by DJ. DJ also contributed to data analysis and supervised the project.

FUNDING

We thank BBSRC (BB/H003746/1 and BB/M000249/1) for funding this work. RJ was funded by a KESS studentship - (project code 511113).

ACKNOWLEDGMENTS

We would like to thank Edward Lemke and his group at EMBL Heidelberg for donating the pEVOL-SCO plasmid for incorporating SCO-K and Ryan Mehl for making available pDULE via AddGene for incorporating AzF. We would like to thank the Protein Technology Hub, School of Biosciences, Cardiff University for use of facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.733550/full#supplementary-material

REFERENCES

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., Dimaio, F. P., Park, H., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theor. Comput.* 13, 3031–3048. doi:10.1021/acs.jctc.7b00125

Ali, M. H., and Imperiali, B. (2005). Protein Oligomerization: How and Why. *Bioorg. Med. Chem.* 13, 5013–5020. doi:10.1016/j.bmc.2005.05.037

Arnesano, F., Banci, L., Bertini, I., Faraone-Mennella, J., Rosato, A., Barker, P. D., et al. (1999). The Solution Structure of Oxidized Escherichia coli Cytochrome B562,. *Biochemistry* 38, 8657–8670. doi:10.1021/bi982785f

Arpino, J. A., Baldwin, A. J., Mcgarrity, A. R., Tippmann, E. M., and Jones, D. D. (2015). In-frame Amber Stop Codon Replacement Mutagenesis for the Directed Evolution of Proteins Containing Non-canonical Amino Acids: Identification of Residues Open to Bio-Orthogonal Modification. *PLoS One* 10, e0127504. doi:10.1371/journal.pone.0127504

Arpino, J. A. J., Czapinska, H., Piasecka, A., Edwards, W. R., Barker, P., Gajda, M. J., et al. (2012). Structural Basis for Efficient Chromophore Communication and Energy Transfer in a Constructed Didomain Protein Scaffold. *J. Am. Chem. Soc.* 134, 13632–13640. doi:10.1021/ja301987h

Arpino, J. A. J., and Polizzi, K. M. (2020). A Modular Method for Directing Protein Self-Assembly. *ACS Synth. Biol.* 9, 993–1002. doi:10.1021/acssynbio.9b00504

Arpino, J. A. J., Rizkallah, P. J., and Jones, D. D. (2014). Structural and Dynamic Changes Associated with Beneficial Engineered Single-Amino-Acid Deletion Mutations in Enhanced green Fluorescent Protein. *Acta Cryst. D Biol. Crystallogr.* 70, 2152–2162. doi:10.1107/s139900471401267x

Baldwin, A. J., Bader, R., Christodoulou, J., Macphee, C. E., Dobson, C. M., and Barker, P. D. (2006). Cytochrome Display on Amyloid Fibrils. *J. Am. Chem. Soc.* 128, 2162–2163. doi:10.1021/ja0565673

Beesley, J. L., and Woolfson, D. N. (2019). The De Novo Design of α-helical Peptides for Supramolecular Self-Assembly. *Curr. Opin. Biotechnol.* 58, 175–182. doi:10.1016/j.copbio.2019.03.017

Bowen, B. J., Mcgarrity, A. R., Szeto, J.-Y. A., Pudney, C. R., and Jones, D. D. (2020). Switching Protein Metalloporphyrin Binding Specificity by Design from Iron to Fluorogenic Zinc. *Chem. Commun.* 56, 4308–4311. doi:10.1039/d0cc00596g

Brodin, J. D., Ambroggio, X. I., Tang, C., Parent, K. N., Baker, T. S., and Tezcan, F. A. (2012). Metal-directed, Chemically Tunable Assembly of One-, Two- and Three-Dimensional Crystalline Protein Arrays. *Nat. Chem* 4, 375–382. doi:10.1038/nchem.1290

Butterfield, G. L., Lajoie, M. J., Gustafson, H. H., Sellers, D. L., Nattermann, U., Ellis, D., et al. (2017). Evolution of a Designed Protein Assembly Encapsulating its Own RNA Genome. *Nature* 552, 415–420. doi:10.1038/nature25157

Ciaccio, C., Coletta, A., De Sanctis, G., Marini, S., and Coletta, M. (2008). Cooperativity and Allostery in Haemoglobin Function. *IUBMB Life* 60, 112–123. doi:10.1002/iub.6

Ding, Y., Li, J., Enterina, J. R., Shen, Y., Zhang, I., Tewson, P. H., et al. (2015). Ratiometric Biosensors Based on Dimerization-dependent Fluorescent Protein Exchange. *Nat. Methods* 12, 195–198. doi:10.1038/nmeth.3261

Edwards, W. R., Busse, K., Allemann, R. K., and Jones, D. D. (2008). Linking the Functions of Unrelated Proteins Using a Novel Directed Evolution Domain Insertion Method. *Nucleic Acids Res.* 36, e78. doi:10.1093/nar/gkn363

Eger, S., Scheffner, M., Marx, A., and Rubini, M. (2010). Synthesis of Defined Ubiquitin Dimers. *J. Am. Chem. Soc.* 132, 16337–16339. doi:10.1021/ja1072838

Fallas, J. A., Ueda, G., Sheffler, W., Nguyen, V., Mcnamara, D. E., Sankaran, B., et al. (2017). Computational Design of Self-Assembling Cyclic Protein Homo-Oligomers. *Nat. Chem.* 9, 353–360. doi:10.1038/nchem.2673

Golub, E., Subramanian, R. H., Esselborn, J., Alberstein, R. G., Bailey, J. B., Chiong, J. A., et al. (2020). Constructing Protein Polyhedra via Orthogonal Chemical Interactions. *Nature* 578, 172–176. doi:10.1038/s41586-019-1928-2

Gonen, S., Dimaio, F., Gonen, T., and Baker, D. (2015). Design of Ordered Two-Dimensional Arrays Mediated by Noncovalent Protein-Protein Interfaces. *Science* 348, 1365–1368. doi:10.1126/science.aaa9897

Goodsell, D. S., and Olson, A. J. (2000). Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* 29, 105–153. doi:10.1146/annurev.biophys.29.1.105

Gwyther, R. E. A., Jones, D. D., and Worthy, H. L. (2019). Better Together: Building Protein Oligomers Naturally and by Design. *Biochem. Soc. Trans.* 47, 1773–1780. doi:10.1042/bst20190283

Hanna, D. A., Harvey, R. M., Martinez-Guzman, O., Yuan, X., Chandrasekharan, B., Raju, G., et al. (2016). Heme Dynamics and Trafficking Factors Revealed by Genetically Encoded Fluorescent Heme Sensors. *Proc. Natl. Acad. Sci. USA* 113, 7539–7544. doi:10.1073/pnas.1523802113

Hartley, A. M., Worthy, H. L., Reddington, S. C., Rizkallah, P. J., and Jones, D. D. (2016). Molecular Basis for Functional Switching of GFP by Two Disparate Non-native post-translational Modifications of a Phenyl Azide Reaction Handle. *Chem. Sci.* 7, 6484–6491. doi:10.1039/c6sc00944a

Hatzakis, N. S., Engelkamp, H., Velonia, K., Hofkens, J., Christianen, P. C., Svendsen, A., et al. (2006). Synthesis and Single Enzyme Activity of a Clicked Lipase-BSA Hetero-Dimer. *Chem. Commun.* 19, 2012–2014. doi:10.1039/B516551B

Hudak, J. E., Barfield, R. M., de Hart, G. W., Grob, P., Nogales, E., Bertozzi, C. R., et al. (2012). Synthesis of Heterobifunctional Protein Fusions Using Copper-free Click Chemistry and the Aldehyde Tag. *Angew. Chem. Int. Ed.* 51, 4161–4165. doi:10.1002/anie.201108130

Jones, D. D., and Barker, P. D. (2005). Controlling Self-Assembly by Linking Protein Folding, DNA Binding, and the Redox Chemistry of Heme. *Angew. Chem. Int. Ed.* 44, 6337–6341. doi:10.1002/anie.200463035

Jones, D. D., and Barker, P. D. (2004). Design and Characterisation of an Artificial DNA-Binding Cytochrome. *Chembiochem* 5, 964–971. doi:10.1002/cbic.200300569

Kim, C. H., Axup, J. Y., Dubrovska, A., Kazane, S. A., Hutchins, B. A., Wold, E. D., et al. (2012). Synthesis of Bispecific Antibodies Using Genetically Encoded Unnatural Amino Acids. *J. Am. Chem. Soc.* 134, 9918–9921. doi:10.1021/ja303904e

Kim, S., Ko, W., Sung, B. H., Kim, S. C., and Lee, H. S. (2016). Direct Protein-Protein Conjugation by Genetically Introducing Bioorthogonal Functional Groups into Proteins. *Bioorg. Med. Chem.* 24, 5816–5822. doi:10.1016/j.bmc.2016.09.035

Kim, Y. E., Kim, Y. N., Kim, J. A., Kim, H. M., and Jung, Y. (2015). Green Fluorescent Protein Nanopolygons as Monodisperse Supramolecular Assemblies of Functional Proteins with Defined Valency. *Nat. Commun.* 6, 7134. doi:10.1038/ncomms8134

Kobayashi, N., and Arai, R. (2017). Design and Construction of Self-Assembling Supramolecular Protein Complexes Using Artificial and Fusion Proteins as

Nanoscale Building Blocks. *Curr. Opin. Biotechnol.* 46, 57–65. doi:10.1016/j.copbio.2017.01.001

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., et al. (2017). The ClusPro Web Server for Protein-Protein Docking. *Nat. Protoc.* 12, 255–278. doi:10.1038/nprot.2016.169

Larsen, T. A., Olson, A. J., and Goodsell, D. S. (1998). Morphology of Protein-Protein Interfaces. *Structure* 6, 421–427. doi:10.1016/s0969-2126(98)00044-6

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., et al. (2011). Rosetta3. *Methods Enzymol.* 487, 545–574. doi:10.1016/b978-0-12-381270-4.00019-6

Leibly, D. J., Arbing, M. A., Pashkov, I., Devore, N., Waldo, G. S., Terwilliger, T. C., et al. (2015). A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* 23, 1754–1768. doi:10.1016/j.str.2015.07.008

Liu, Y., and Kuhlman, B. (2006). RosettaDesign Server for Protein Design. *Nucleic Acids Res.* 34, W235–W238. doi:10.1093/nar/gkl163

Ljubetic, A., Gradisar, H., and Jerala, R. (2017). Advances in Design of Protein Folds and Assemblies. *Curr. Opin. Chem. Biol.* 40, 65–71. doi:10.1016/j.cbpa.2017.06.020

Lyskov, S., and Gray, J. J. (2008). The RosettaDock Server for Local Protein-Protein Docking. *Nucleic Acids Res.* 36, W233–W238. doi:10.1093/nar/gkn216

Marianayagam, N. J., Sunde, M., and Matthews, J. M. (2004). The Power of Two: Protein Dimerization in Biology. *Trends Biochem. Sci.* 29, 618–625. doi:10.1016/j.tibs.2004.09.006

Mei, G., Di Venere, A., Rosato, N., and Finazzi-Agrò, A. (2005). The Importance of Being Dimeric. *FEBS J.* 272, 16–27. doi:10.1111/j.1432-1033.2004.04407.x

Miyake-Stoner, S. J., Refakis, C. A., Hammill, J. T., Lusic, H., Hazen, J. L., Deiters, A., et al. (2010). Generating Permissive Site-specific Unnatural Aminoacyl-tRNA Synthetases. *Biochemistry* 49, 1667–1677. doi:10.1021/bi901947r

Mou, Y., Huang, P.-S., Hsu, F.-C., Huang, S.-J., and Mayo, S. L. (2015). Computational Design and Experimental Verification of a Symmetric Protein Homodimer. *Proc. Natl. Acad. Sci. USA* 112, 10714–10719. doi:10.1073/pnas.1505072112

Nooren, I. M. A., and Thornton, J. M. (2003). NEW EMBO MEMBER'S REVIEW: Diversity of Protein-Protein Interactions. *EMBO J.* 22, 3486–3492. doi:10.1093/emboj/cdg359

Norn, C. H., and André, I. (2016). Computational Design of Protein Self-Assembly. *Curr. Opin. Struct. Biol.* 39, 39–45. doi:10.1016/j.sbi.2016.04.002

Onoda, A., Kakikura, Y., Uematsu, T., Kuwabata, S., and Hayashi, T. (2012). Photocurrent Generation from Hierarchical Zinc-Substituted Hemoprotein Assemblies Immobilized on a Gold Electrode. *Angew. Chem. Int. Ed.* 51, 2628–2631. doi:10.1002/anie.201105186

Oohora, K., Fujimaki, N., Kajihara, R., Watanabe, H., Uchihashi, T., and Hayashi, T. (2018). Supramolecular Hemoprotein Assembly with a Periodic Structure Showing Heme-Heme Exciton Coupling. *J. Am. Chem. Soc.* 140, 10145–10148. doi:10.1021/jacs.8b06690

Oohora, K., and Hayashi, T. (2014). Hemoprotein-based Supramolecular Assembling Systems. *Curr. Opin. Chem. Biol.* 19, 154–161. doi:10.1016/j.cbpa.2014.02.014

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and Characterization of a Superfolder green Fluorescent Protein. *Nat. Biotechnol.* 24, 79–88. doi:10.1038/nbt1172

Pope, J. R., Johnson, R. L., Jamieson, W. D., Worthy, H. L., Kailasam, S., Ahmed, R. D., et al. (2020). Association of Fluorescent Protein Pairs and its Significant Impact on Fluorescence and Energy Transfer. *Adv. Sci.* 8, 2003167. doi:10.1002/advs.202003167

Radford, R. J., Lawrenz, M., Nguyen, P. C., Mccammon, J. A., and Tezcan, F. A. (2011). Porous Protein Frameworks with Unsaturated Metal Centers in Sterically Encumbered Coordination Sites. *Chem. Commun.* 47, 313–315. doi:10.1039/c0cc02168g

Reddington, S. C., Baldwin, A. J., Thompson, R., Brancale, A., Tippmann, E. M., and Jones, D. D. (2015). Directed Evolution of GFP with Non-natural Amino Acids Identifies Residues for Augmenting and Photoswitching Fluorescence. *Chem. Sci.* 6, 1159–1166. doi:10.1039/c4sc02827a

Reddington, S. C., Rizkallah, P. J., Watson, P. D., Pearson, R., Tippmann, E. M., and Jones, D. D. (2013). Different Photochemical Events of a Genetically Encoded Phenyl Azide Define and Modulate GFP Fluorescence. *Angew. Chem. Int. Ed.* 52, 5974–5977. doi:10.1002/anie.201301490

Reddington, S. C., Tippmann, E. M., and Dafydd Jones, D. (2012). Residue Choice Defines Efficiency and Influence of Bioorthogonal Protein Modification via Genetically Encoded Strain Promoted Click Chemistry. *Chem. Commun.* 48, 8419–8421. doi:10.1039/c2cc31887c

Remington, S. J. (2011). Green Fluorescent Protein: a Perspective. *Protein Sci.* 20, 1509–1519. doi:10.1002/pro.684

Robinson, C. R., Liu, Y., Thomson, J. A., Sturtevant, J. M., and Sligar, S. G. (1997). Energetics of Heme Binding to Native and Denatured States of Cytochrome B562. *Biochemistry* 36, 16141–16146. doi:10.1021/bi971470h

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an Open-Source Platform for Biological-Image Analysis. *Nat. Methods* 9, 676–682. doi:10.1038/nmeth.2019

Schoffelen, S., Beekwilder, J., Debets, M. F., Bosch, D., and Hest, J. C. M. v. (2013). Construction of a Multifunctional Enzyme Complex via the Strain-Promoted Azide-Alkyne Cycloaddition. *Bioconjug. Chem.* 24, 987–996. doi:10.1021/bc400021j

Scott, D. J., Gunn, N. J., Yong, K. J., Wimmer, V. C., Veldhuis, N. A., Challis, L. M., et al. (2018). A Novel Ultra-stable, Monomeric Green Fluorescent Protein for Direct Volumetric Imaging of Whole Organs Using CLARITY. *Sci. Rep.* 8, 667. doi:10.1038/s41598-017-18045-y

Seifert, M. H. J., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B., et al. (2003). Backbone Dynamics of Green Fluorescent Protein and the Effect of Histidine 148 Substitution†. *Biochemistry* 42, 2500–2512. doi:10.1021/bi026481b

Song, W. J., and Tezcan, F. A. (2014). A Designed Supramolecular Protein Assembly with *In Vivo* Enzymatic Activity. *Science* 346, 1525–1528. doi:10.1126/science.1259680

Takeda, S., Kamiya, N., Arai, R., and Nagamune, T. (2001). Design of an Artificial Light-Harvesting Unit by Protein Engineering: Cytochrome B562-Green Fluorescent Protein Chimera. *Biochem. Biophys. Res. Commun.* 289, 299–304. doi:10.1006/bbrc.2001.5966

Takeda, S., Kamiya, N., and Nagamune, T. (2003). A Novel Protein-Based Heme Sensor Consisting of green Fluorescent Protein and Apocytochrome B562. *Anal. Biochem.* 317, 116–119. doi:10.1016/s0003-2697(03)00096-4

Thomas, S. K., Jamieson, W. D., Gwyther, R. E. A., Bowen, B. J., Beachey, A., Worthy, H. L., et al. (2020). Site-Specific Protein Photochemical Covalent Attachment to Carbon Nanotube Side Walls and its Electronic Impact on Single Molecule Function. *Bioconjug. Chem.* 31, 584–594. doi:10.1021/acs.bioconjchem.9b00719

Thomson, A. R., Wood, C. W., Burton, A. J., Bartlett, G. J., Sessions, R. B., Brady, R. L., et al. (2014). Computational Design of Water-Soluble α-helical Barrels. *Science* 346, 485–488. doi:10.1126/science.1257452

Torres-Kolbus, J., Chou, C., Liu, J., and Deiters, A. (2014). Synthesis of Non-linear Protein Dimers through a Genetically Encoded Thiol-Ene Reaction. *PLoS One* 9, e105467. doi:10.1371/journal.pone.0105467

White, C. J., and Bode, J. W. (2018). PEGylation and Dimerization of Expressed Proteins under Near Equimolar Conditions with Potassium 2-Pyridyl Acyltrifluoroborates. *ACS Cent. Sci.* 4, 197–206. doi:10.1021/acscentsci.7b00432

Willis, K. J., Szabo, A. G., Zuker, M., Ridgeway, J. M., and Alpert, B. (1990). Fluorescence Decay Kinetics of the Tryptophyl Residues of Myoglobin: Effect of Heme Ligation and Evidence for Discrete Lifetime Components. *Biochemistry* 29, 5270–5275. doi:10.1021/bi00474a008

Wittung-Stafshede, P., Lee, J. C., Winkler, J. R., and Gray, H. B. (1999). Cytochrome B562 Folding Triggered by Electron Transfer: Approaching the Speed Limit for Formation of a four-helix-bundle Protein. *Proc. Natl. Acad. Sci.* 96, 6587–6590. doi:10.1073/pnas.96.12.6587

Worthy, H. L., Auhim, H. S., Jamieson, W. D., Pope, J. R., Wall, A., Batchelor, R., et al. (2019). Positive Functional Synergy of Structurally Integrated Artificial Protein Dimers Assembled by Click Chemistry. *Commun. Chem.* 2, 83. doi:10.1038/s42004-019-0185-5

Yeates, T. O., Liu, Y., and Laniado, J. (2016). The Design of Symmetric Protein Nanomaterials Comes of Age in Theory and Practice. *Curr. Opin. Struct. Biol.* 39, 134–143. doi:10.1016/j.sbi.2016.07.003

Zacharias, D. A., Violin, J. D., Newton, A. C., and Tsien, R. Y. (2002). Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells. *Science* 296, 913–916. doi:10.1126/science.1068539

Zaki, A. J., Hartley, A. M., Reddington, S. C., Thomas, S. K., Watson, P., Hayes, A., et al. (2018). Defined Covalent Assembly of Protein Molecules on Graphene Using a Genetically Encoded Photochemical Reaction Handle. *RSC Adv.* 8, 5768–5775. doi:10.1039/c7ra11166e

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 14 September 2021
doi: 10.3389/fchem.2021.741290

Total Synthesis and Antimicrobial Evaluation of Pagoamide A

*Cheng-Han Wu and John Chu**

Department of Chemistry, National Taiwan University, Taipei City, Taiwan

Natural products are often the starting point for drug development and also the testing ground for synthetic methods. Herein we describe the total synthesis and antimicrobial evaluation of a marine natural product, pagoamide A, which is a macrocyclic depsipeptide with two backbone thiazole units and a dimethylated *N*-terminus. The two thiazole building blocks were synthesized from commercially available materials in four or fewer steps and employed directly in solid-phase peptide synthesis (SPPS) to afford pagoamide A. The use of SPPS ensured that the synthetic sequence is operationally straightforward and, if needed, permits modular substitution of building blocks to easily access diverse structural analogs. Our antimicrobial assays showed that pagoamide A has moderate activity against *Bacillus subtilis*.

Keywords: pagoamide A, marine natural products, total synthesis, Hantzsch thiazole synthesis, solid-phase peptide synthesis, antimicrobial activity

OPEN ACCESS

Edited by:
Yu-Hsuan Tsai,
Shenzhen Bay Laboratory, China

Reviewed by:
Manuel Mueller,
King's College London, United Kingdom
Louis Luk,
Cardiff University, United Kingdom

****Correspondence:***
John Chu
johnchu@ntu.edu.tw

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *14 July 2021*
Accepted: *01 September 2021*
Published: *14 September 2021*

Citation:
Wu C-H and Chu J (2021) Total Synthesis and Antimicrobial Evaluation of Pagoamide A. Front. Chem. 9:741290. doi: 10.3389/fchem.2021.741290

INTRODUCTION

Microbial natural products have historically been a fruitful source of bioactive small molecules (Newman and Cragg, 2020). Peptide natural products, including nonribosomal peptides (NRPs) (Fischbach and Walsh, 2006) and ribosomally-synthesized and post-translationally modified peptides (RiPPs) (Arnison et al., 2013), are highly represented in small molecule therapeutics in clinical use today, particularly for the treatment of microbial infections. Both are very attractive classes of natural products that have inspired scientists as they look for starting points for the development of new drugs. Once a natural product demonstrated its potential as a drug candidate, the ability to access it with high efficiency through chemical synthesis from simple starting materials, i.e., total synthesis, is critically important and play an integral role as it progresses through the drug development pipeline.

Peptide natural products like NRP and RiPP are biosynthesized modularly, the former by an enzymatic assembly line and the latter by the ribosome (Fischbach and Walsh, 2006; Arnison et al., 2013). In these two biosynthetic platforms, an amino acid building block is first activated at the expense of an ATP and then added to a growing peptide chain through an intermediary, i.e., the phosphopantetheine arm on the thiolation domain and tRNA for NRP and RiPP, respectively. Solid-phase peptide synthesis (SPPS) is a synthetic platform devised by chemists for the construction of consecutive amide bonds that operates in a similarly modular fashion (Coin et al., 2007). A variety of reagents have been developed for the activation of amino acids (Takayama et al., 2018) to ensure high efficiency in the formation of an amide bond on the solid support (resins). An excess of activated amino acid can be used to drive the reaction to near completion; meanwhile, workup and purification are reduced to simply washing the resins with the appropriate solvents. As such, SPPS has become the method of choice for the synthesis of linear peptides.

However, peptide natural products are much more complex than simple linearly connected strings of amino acids (Fischbach and Walsh, 2006; Arnison et al., 2013). Structural complexity poses

a number of synthetic challenges that SPPS must meet before it can be applied routinely for the total synthesis of diverse peptide natural products (McIntosh et al., 2009). Our laboratory is interested in expanding the scope of SPPS beyond the mere formation of amide bonds. In particular, we see an opportunity to combine the versatility of conventional solution-phase organic chemistry and the high efficiency of SPPS into a multifaceted synthetic platform.

Here we report the total synthesis of pagoamide A (**1**) (Li et al., 2020), a natural product isolated from the culture of green algae *Derbesia* sp. Pagoamide A (**1**) is made of eight L- and three D-amino acids. It contains a number of structural features that prevents it from being synthesized by standard SPPS, including a dimethylated *N*-terminus, two thiazole moieties, and a depsipeptide macrocycle with 19 atoms. Matching spectral characterizations of **1** obtained by total synthesis and culture extracts confirmed its stereochemical assignment. We also evaluated its antimicrobial activities against a panel of common microbial pathogens.

MATERIALS AND METHODS

Chemicals, Glassware, Microbes, and Growth Media

Wang resins were purchased from Novabiochem (Merck). Amino acid building blocks and coupling reagents were purchased from P3 BioSystems and Novabiochem (Merck). Reaction vessels for SPPS were custom made by Tung-Yin Glassware Co., Ltd., and other common glassware were purchased from the same vendor. Reagents and consumables for organic synthesis were purchased from Thermo Fisher Scientific Inc. and Sigma-Aldrich Inc. Growth media were prepared from premixed powders purchased from BioShop Canada Inc.

Fmoc-Gly-NH_2 (5)

HBTU (569 mg, 1.5 mmol) was added to a solution of Fmoc-Gly-OH (297 mg, 1.0 mmol) and DIPEA (348 µl, 2.0 mmol) in THF (5 ml) and stirred at room temperature for 10 min. Ammonium bicarbonate (79 mg, 1.0 mmol) was added slowly, and the mixture was stirred for another 8 h before THF was removed *in vacuo*. The remaining residue was dissolved in EtOAc, washed successively with 1N HCl and 1M NaOH, dried over anhydrous sodium sulfate (Na_2SO_4), concentrated under reduced pressure, and purified by silica gel column chromatography (R_f = 0.24, CH_2Cl_2/MeOH = 95: 5) to afford **5** as a white solid (292 mg, 92%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.90 (d, *J* = 7.4 Hz, 2H), 7.72 (d, *J* = 7.1 Hz, 2H), 7.42 (t, *J* = 7.1 Hz, 2H), 7.34 (t, *J* = 7.3 Hz, 2H), 7.26 (br, 1H), 6.99 (br, 1H), 4.29 (d, *J* = 6.3 Hz, 2H), 4.23 (t, *J* = 6.7 Hz, 1H), 3.55 (d, *J* = 5.8 Hz, 2H); ^{13}C NMR (100 MHz, DMSO-d_6,) δ 171.1, 156.5, 143.9, 140.7, 127.6, 127.1, 125.3, 120.1, 65.7, 46.7, 43.3; HRMS (ESI-TOF) calculated for $C_{17}H_{17}N_2O_3$ $[M^+H]^+$: 297.1234; found: 297.1249.

Fmoc-Glycine Thioamide (6)

Lawesson reagent (404 mg, 1.0 mmol) was added to a solution of **5** (296 mg, 1.0 mmol) dissolved in dimethoxyethane (DME, 15 ml). The reaction was quenched by the addition of saturated sodium bicarbonate ($NaHCO_3$) after stirring at room temperature for 12 h. The mixture was extracted successively with EtOAc, 5% (w/v) $KHSO_4$, and brine. It was then dried over Na_2SO_4, filtered, passed through a short silica gel pad, and then concentrated under reduced pressure. *N*-Fmoc-glycine thioamide (**6**) was obtained as a white powder and used without further purification.

Fmoc-Gly-Thz-OH (2)

Bromopyruvic acid (334 mg, 2.0 mmol) in DME (5 ml) was added dropwise at 0°C into a slurry of $CaCO_3$ (600.5°mg, 3.0 mmol) and **6** (312.4 mg, 1.0 mmol) in DME (5 ml) with stirring. The mixture was allowed to gradually warm to room temperature over 12 h, which was then filtered and washed with DME. The solvent was removed *in vacuuo* and the remaining residue was dissolved in EtOAc, washed successively with 5% (w/v) $KHSO_4$ and brine, and dried over Na_2SO_4. Recrystallization in EtOAc/hexane gave **7** as a white solid (277.8 mg, 73%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 13.0 (br, 1H), 8.35 (s, 1H), 8.29 (t, J = 5.45 Hz, 1H), 7.89 (d, *J* = 7.3 Hz, 2H), 7.70 (d, *J* = 7.3 Hz, 2H), 7.42 (t, *J* = 7.2 Hz, 2H), 7.33 (t, *J* = 7.2 Hz, 2H), 4.48 (d, *J* = 5.8 Hz, 2H), 4.39 (d, *J* = 6.7 Hz, 2H), 4.25 (t, *J* = 6.4 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 170.8, 162.0, 156.4, 146.9, 143.7, 140.8, 128.5, 127.6, 127.1, 125.1, 120.1, 65.8, 46.7, 42.2; HRMS (ESI-TOF) calculated for $C_{20}H_{16}N_2O_4SNa$ $[M^+Na]^+$: 403.0726; found: 403.0723.

Fmoc-L-Valine Thioamide (8)

Compound **8** was synthesized using procedures similar to those described above for Fmoc-glycine thioamide (**6**). See **Supplementary Material** for details.

Fmoc-L-Val-Thz-OEt (9)

Ethyl bromopyruvate (90% solution, 627 µl, 4.5 mmol) was added dropwise at 0°C to a solution of **8** (531 mg, 1.5 mmol) and 2,6-lutidine (524 µl, 4.5 mmol) in DME (5 ml) with stirring. The mixture was allowed to gradually warm to room temperature over 16 h. Trifluoroacetic anhydride (TFAA, 625 µl, 4.5 mmol) and 2,6-lutidine (524 µl, 4.5 mmol) in DME (2 ml) was added at 0°C and stirred at room temperature for 1 h. The reaction mixture was diluted with EtOAc, washed successively with saturated $NaHCO_3$, 5% (w/v) $KHSO_4$, and brine, dried over Na_2SO_4, and purified by silica gel column chromatography (R_f = 0.30, hexanes/EtOAc = 3:1) to yield **11** (648 mg, 96%) as an orange oil. ^{1}H NMR ($CDCl_3$, 400 MHz) δ 8.02 (s, 1H), 7.68 (d, *J* = 7.3 Hz, 2H), 7.54 (br, 2H), 7.32 (t, *J* = 7.3 Hz, 2H), 7.23 (t, *J* = 7.1 Hz, 2H), 5.86 (d, *J* = 9.0 Hz, 1H), 4.94 (t, *J* = 7.1 Hz, 1H), 4.40–4.33 (m, 4H), 4.16 (t, *J* = 5.5 Hz, 1H), 2.42–2.37 (m, 1H), 1.34 (t, *J* = 7.1 Hz, 3H), 0.93–0.88 (m, 6H); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 172.0, 160.9, 155.8, 146.9, 143.4, 143.3, 140.9, 127.3, 126.7, 124.7, 119.6, 66.6, 61.0, 58.3, 46.8, 33.0, 19.1, 17.2, 14.0; HRMS (ESI-TOF) calculated for $C_{25}H_{27}N_2O_4S$ $[M^+H]^+$: 451.1680, found: 451.1686.

Me2-L-Val-Thz-OEt (10)

Compound **9** (450 mg, 1.0 mmol) was taken up in 20% (v/v) diethylamine in CH_2Cl_2 (10 ml), stirred at room temperature for 2 h, concentrated under reduced pressure, and then dissolved in methanol (5 ml). Formaldehyde solution (37% w/w) was added dropwise (245.7 µl, 3.0 mmol). Sodium cyanoborohydride (220 mg, 3.5 mmol) was added 40 min. later, and the resulting mixture was stirred at room temperature for another 15 h. The solvent was removed under reduced pressure and the residue was redissolved in EtOAc, which was then washed successively with sat. $NaHCO_3$ and brine, dried over Na_2SO_4, and purified by silica gel column chromatography (R_f = 0.33, CH_2Cl_2/EtOAc = 9:1) to afford **10** (161 mg, 63%) as an orange oil. ^{1}H NMR ($CDCl_3$, 400 MHz) δ 8.09 (s, 1H), 4.35 (q, *J* = 7.1 Hz, 2H), 3.55 (d, *J* = 8.8 Hz, 1H), 2.17 (s, 1H), 2.14–2.09 (m, 1H), 1.33 (t, *J* = 7.1 Hz, 3H), 0.95 (d, *J* = 6.6 Hz, 3H), 0.76 (d, *J* = 6.6 Hz, 3H); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 170.0, 161.2, 145.9, 127.1, 72.8, 61.1, 41.5, 29.9, 19.8, 18.6, 14.1; HRMS (ESI-TOF) calculated for $C_{12}H_{21}N_2O_2S$ $[M^+H]^+$: 257.1329, found: 257.1318.

Me2-L-Val-Thz-OH (3)

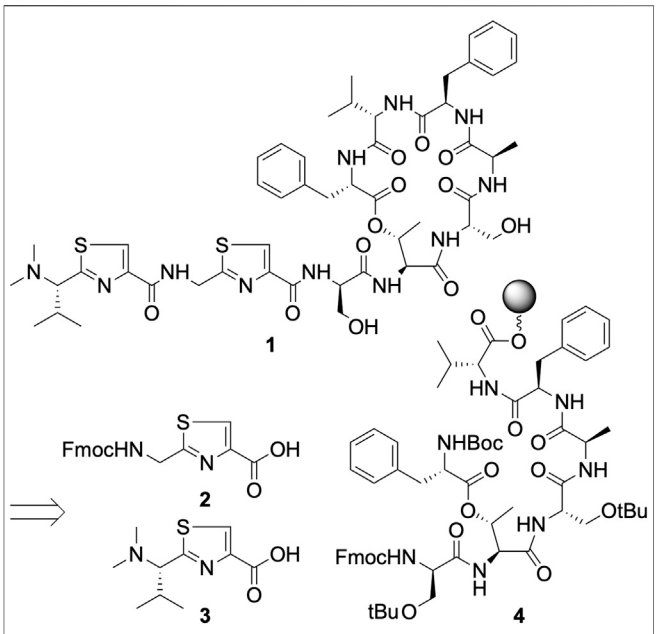
Compound **10** (231 mg, 0.9 mmol) was dissolved in tetrahydrofuran (2 ml) and cooled to 0°C. Lithium hydroxide (2 ml, 1M) was added, and the reaction mixture was stirred at room temperature for 1.5 h. It was neutralized at 0°C with 1N HCl and warmed to room temperature. The crude mixture was concentrated and purified by column chromatography using a manually packed column of C18 resins using a gradient of 0–5% (v/v) methanol. Compound **3** appeared as a colorless foam after solvent removal (213 mg, 92%). ^{1}H NMR (CD_3OD, 400 MHz) δ 8.51 (s, 1H), 4.91 (d, *J* = 7.9 Hz, 1H), 2.93 (s, 6H), 2.72–2.64 (m, 1H), 1.12 (d, *J* = 6.6 Hz, 3H), 0.91 (d, *J* = 6.5 Hz, 3H); ^{13}C NMR (CD_3OD, 100 MHz) δ 164.3, 162.2, 131.2, 71.6, 42.2, 30.3, 20.2, 18.2; HRMS (ESI-TOF) calculated for $C_{10}H_{17}N_2O_2S$ $[M^+H]^+$: 229.1005, found: 229.1016. Compound **3** was then used as a typical amino acid building block in SPPS.

Solid-Phase Peptide Synthesis

Standard Fmoc-based solid-phase peptide synthesis procedures were employed; see **Supplementary Material** for details.

Macrocyclization

Upon completion of peptide synthesis, global deprotection and cleavage was performed by treating the resins with a trifluoroacetic acid (TFA) cocktail supplemented with 2.5% (v/v) of each triisopropylsilane and water. The resins were removed by filtration and TFA was dried under a gentle nitrogen stream. Cold hexane (approx. 50 ml) was added and the linear precursor of pagoamide A (**14**) crashed out as a light-yellow precipitate, which was collected by centrifugation at 4°C. PyAOP (208.5 mg, 0.4 mmol) and DIPEA (174.2 µl, 1.0 mmol) were added to **14** in DCM (25 ml) to effect macrocyclization. The crude product was concentrated under reduced pressure, dissolved in methanol, and purified by reversed-phase HPLC using a dual solvent system with a Hypersil GOLD aQ C18 column (250 (L) × 10 (ID) mm, 5 µm) at 8 ml/min, wherein solvent A and B are water and acetonitrile supplemented with 0.1% (v/v) of formic acid, respectively. The desired macrocyclic product **1** eluted at 40%B. HRMS (ESI-TOF) calculated for $C_{51}H_{68}N_{11}O_{12}S_2$ $[M^+H]^+$: 1,090.449, found: 1,090.451. See **Supplementary Material** for analytical data, including HPLC traces, MALDI-MS (reaction monitoring), and NMR and HRMS spectra.



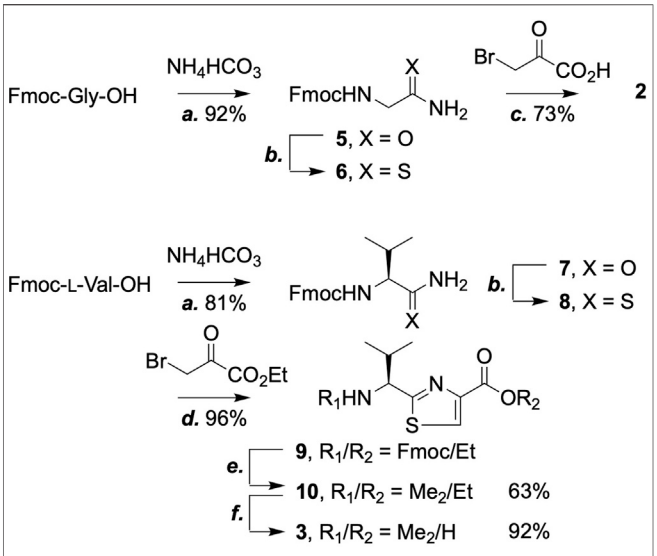
SCHEME 1 | Retrosynthetic analysis of pagoamide A.

Antimicrobial Assays

Microbes were grown from single colonies on agar plates into LB (bacteria) and YPD broths (*Candida albicans*). Assays were carried out in 96-well microtiter plates to determine the minimum inhibitory concentrations (MICs) of **1** against various microbial pathogens by using the broth microdilution method. Briefly, DMSO stock solution of synthetic **1** (12.8 mg/ml) was added to the first well of each row and 1/2× serially diluted across the plate. Wells 11 and 12 were free of **1** (positive control) and free of microbe (negative control), respectively. Overnight microbial cultures were diluted 5,000-fold and used as the inocula, such that the final volume in all wells were 100 µl with the concentrations of **1** ranging from 64 to 0.125 µg/ml in each row. The plate was incubated statically at 30°C for 24 h. The lowest concentration of **1** that led to inhibition of microbial growth was recorded as the MIC.

RESULTS

Pagoamide A (**1**) is a cyclic depsipeptide made of 11 amino acids, three of which are D-amino acids. While SPPS is an obvious choice for the construction of its backbone, as both L- and D-Fmoc building blocks are readily available from commercial venders, a number of features in **1** call for expanding the scope of



SCHEME 2 | Synthesis of thiazole building blocks. Reagents and conditions: **(a)** Ammonium bicarbonate (1 equiv.), HBTU (1.5 equiv.), DIPEA (2 equiv.), THF, 25°C; **(b)** Lawesson reagent (1 equiv.), DME, 25°C; **(c)** Bromopyruvic acid (1 equiv.), $CaCO_3$ (3 equiv.), DME, 25°C; **(d)** Ethyl bromopyruvate (3 equiv.), 2,6-lutidine (3 equiv.), DME, 0–25°C then TFAA (3 equiv.), 2,6-lutidine (3 equiv.), 0–25°C; **(e)** Formaldehyde solution (37% w/w, 3 equiv.), sodium cyanoborohydride (3.5 equiv.), 25°C; **(f)** LiOH (2 ml, 1 M), THF, 25°C.

synthesis beyond simple amino acid coupling in conventional SPPS. These features include two thiazole rings, the dimethylated *N*-terminus, and the macrocyclic structure.

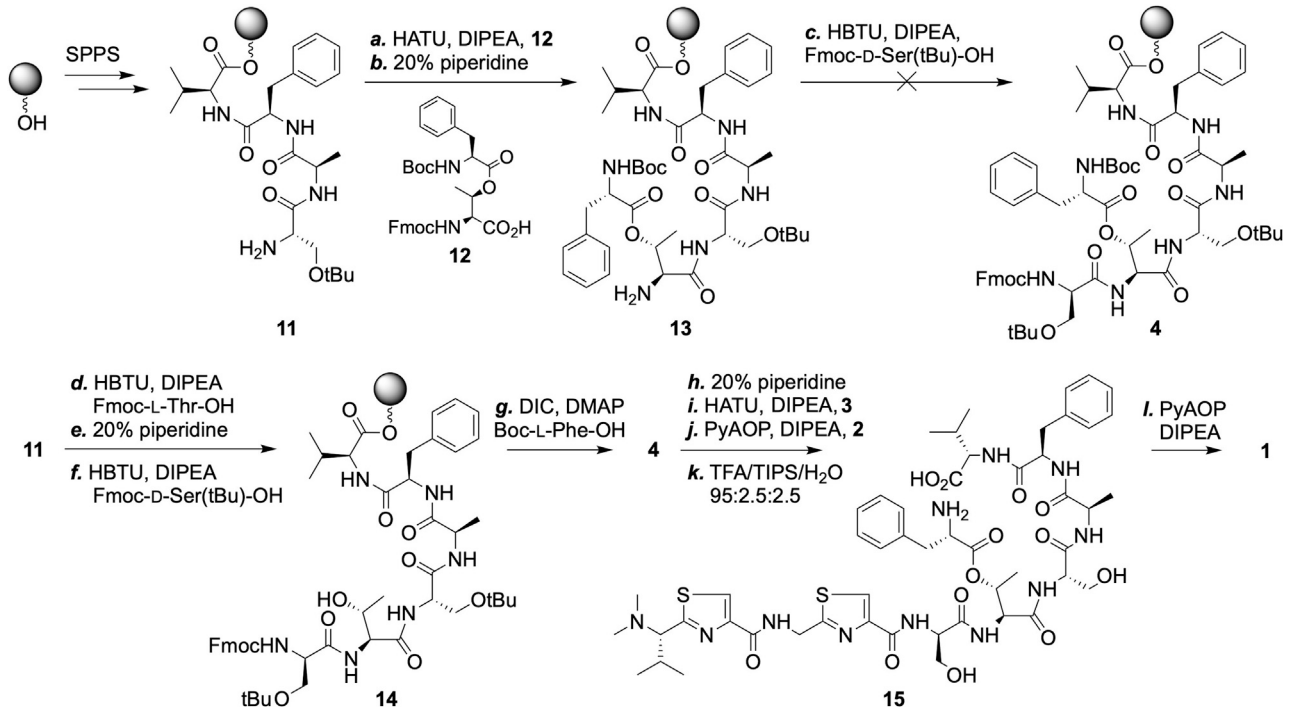
In our retrosynthetic plan (**Scheme 1**), each thiazole moiety will be incorporated as a pre-made building block that spans the length of two amino acids. In this strategy, because the dimethylated *N*-terminus of **1** is part of a thiazole building block, *N*,*N*-dimethylation will be performed after thiazole formation and before the use of this building block in SPPS. Total syntheses of cyclic depsipeptides reported in the literature mostly opt to construct ester bonds early in the synthetic scheme rather than using its formation as the final macrocyclization step (Lam et al., 2013; Lohani et al., 2015). This is because amines, as opposed to the less nucleophilic hydroxyl groups, are generally a superior choice as the site to conclude macrocycle formation. We therefore planned to build **4** as a key intermediate.

We started by constructing the thiazole building blocks **2** and **3** (**Scheme 2**). *N*-Fmoc-protected glycine (Fmoc-Gly-OH) was first converted to the amide (**5**) and then, via the use of Lawesson reagent, (Thomsen et al., 1984) the thioamide (**6**). Hantzsch thiazole synthesis afforded **2** with satisfactory yield (Singh et al., 2012; Fenner et al., 2016; Aihara et al., 2017). The same workflow turned Fmoc-L-valine into the corresponding thiazole (**9**), except for the use of trifluoroacetic anhydride (TFAA) on the condensation product of ethyl bromopyruvate and **8** to promote dehydration. Once the Fmoc protecting group was removed from **9**, two methyl groups were installed onto the *N*-terminus by reductive amination with an excess amount of sodium cyanoborohydride and formaldehyde to give **10**, which was hydrolyzed to afford **3**. Both compounds **2** and **3** are now ready-to-use SPPS building blocks.

A critical part of our plan was to rely on the formation of an ester bond as the final macrocyclization step. With this in mind, we attempted the synthesis of intermediate **4** en route a pre-made threonine-phenylalanine dimer (**12**) (**Scheme 3**). The coupling of **12** to extend the peptide chain on solid-support proceeded without difficulty. The very next coupling, however, was sluggish and saw no improvement in yield after repeated couplings or the use of alternative reagents. This prompted us to explore a different sequence in constructing **4**, wherein a threonine building block without side-chain protection was incorporated as the sixth residue, followed by the coupling of a standard serine building block. Next, *N*-Boc-protected phenylalanine was activated by *N*,*N*'-diisopropylcarbodiimide and installed successfully onto the exposed threonine side-chain hydroxyl group. Subsequent couplings of **3** and **2** proceeded cleanly.

Trifluoroacetic acid cocktail treatment resulted in global deprotection and peptide cleavage off of the resins. The linear precursor of pagoamide A (**14**) was collected after cold hexane precipitation and subjected to macrocyclization without further purification. This reaction was monitored by MALDI-mass spectrometry and, to our delight, no signal of **14** was observed when four equivalents of the phosphonium coupling reagent PyAOP was used (see **Supplementary Material**). The amide bond formation reaction between the amine of the Phe_{11} and the carboxylate of the Val_{10} residues was essentially complete after 24 h. The HPLC purified final product was confirmed by high resolution ESI-MS. ^{1}H NMR spectrum of synthetic **1** matches that obtained from the culture extracts of *Derbesia* sp., thereby confirming the reported stereochemistry assignment (see Supplementary Material for analytical characterizations) (Li et al., 2020).

The isolation of **1** from crude extracts was guided by a MS-based molecular network and a calcium oscillation assay (Yang et al., 2013; Wang et al., 2016). However, once purified, **1** turned out to be inactive in this very assay and showed no cytotoxicity against the H-460 human lung cancer cell line. No other bioactivity was reported. We were able to obtain sufficient amount of **1** by total synthesis for antimicrobial assays and performed a systematic evaluation of its antimicrobial potential (M07-A9, 2012). Our assay included the ESKAPE pathogens (Mulani et al., 2019) – *Enterococcus faecalis*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas baumannii*, and *Enterobacter* spp. – a panel of highly virulent bacteria whose antibiotic resistant strains are the major cause of fatal infections in humans around the world. Also included in our assay are *Bacillus subtilis* and *Escherichia coli*, model organisms for Gram-positive and Gram-negative bacteria, respectively, as well as a common fungal pathogen *Candida albicans*. In this initial round of screening, our results showed that **1** has only moderate activity (64 μg/ml) against *Bacillus subtilis* (**Table 1**).



SCHEME 3 | SPPS based synthesis of pagoamide A (1). Reagents and conditions: Standard SPPS procedures were carried out throughout the synthesis of pagoamide A (see **Supplementary Material**). Amide couplings were carried out in steps ***a***, ***c***, ***d***, ***f***, ***i***, and ***j***, except for the use of varying coupling reagents; Fmoc groups were removed in steps ***b***, ***e***, ***h***, and after step ***j***. ***g.*** DIC (10 equiv.), DMAP (2 equiv.), 25°C; ***k.*** TFA cocktail, 25°C; ***l.*** PyAOP (4 equiv.), DIPEA (10 equiv.), DCM, 25°C.

TABLE 1 | Pagoamide A (1) was tested against a number of common pathogens.

Name of pathogen	MIC[a] (μg/ml)
Bacillus subtilis BCRC 10614	64
Escherichia coli DH5α	>
Enterococcus faecalis BCRC 10789	>
Staphylococcus aureus BCRC 11863	>
Klebsiella pneumoniae BCRC 11546	>
Acinetobacter baumannii BCRC 10591	>
Pseudomonas aeruginosa BCRC 11864	>
Enterobacter cloacae BCRC 10401	>
Candida albicans BCRC 21538	>

[a]*The highest concentration tested was 64 μg/ml.*

DISCUSSION

We reasoned that one could combine the convenience of peptide synthesis (performed on solid support) and the versatility of conventional organic synthesis (performed in solution) into a powerful new synthetic platform. This platform would be particularly useful in the synthesis of natural products of peptide backbones. We put this concept into practice and report the synthesis of **1**, a natural product with complex structural features on top of a peptide backbone, whose thiazole units, one of which was dimethylated, were constructed in solution into ready-to-use building blocks. The entire linear precursor was assembled on-resin and then macrocyclized off-resin to afford **1**. The approach presented herein is modular and broadly applicable to the synthesis of peptide natural products that contain heterocycles, *N*-methylations, and backbone macrocyclization, all of which are commonly seen structural features in peptide natural products.

A separate total synthesis of pagoamide A was reported by the Ye group after we submitted this manuscript (Wu et al., 2021).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

C-HW synthesized pagoamide A (**1**). JC conceived of the project, performed antimicrobial assays, and wrote the manuscript. All authors read and agreed to the published version of the manuscript.

FUNDING

This work was supported by General Research Grant no. 109WFA0110175 of the Ministry of Science and Technology, Taiwan.

ACKNOWLEDGMENTS

We thank the following core facility personnel at the Department of Chemistry, National Taiwan University, for technical support: Shou-Ling Huang, Ph.D. (NMR) and Corina Leung, Ph.D. (HRMS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.741290/full#supplementary-material

REFERENCES

Aihara, K., Inokuma, T., Jichu, T., Lin, Z., Fu, F., Yamaoka, K., et al. (2017). Cysteine-free intramolecular ligation of *N*-sulfanylethylanilide peptide using 4-mercaptobenzylphosphonic acid: synthesis of cyclic peptide trichamide. *Synlett* 28, 1944–1949.

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., et al. (2013). Ribosomally Synthesized and post-translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* 30, 108–160. doi:10.1039/c2np20085f

Coin, I., Beyermann, M., and Bienert, M. (2007). Solid-phase Peptide Synthesis: from Standard Procedures to the Synthesis of Difficult Sequences. *Nat. Protoc.* 2, 3247–3256. doi:10.1038/nprot.2007.454

Fenner, S., Wilson, Z. E., and Ley, S. V. (2016). The Total Synthesis of the Bioactive Natural Product Plantazolicin A and its Biosynthetic Precursor Plantazolicin B. *Chem. Eur. J.* 22, 15902–15912. doi:10.1002/chem.201603157

Fischbach, M. A., and Walsh, C. T. (2006). Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* 106, 3468–3496. doi:10.1021/cr0503097

Lam, H. Y., Zhang, Y., Liu, H., Xu, J., Wong, C. T. T., Xu, C., et al. (2013). Total Synthesis of Daptomycin by Cyclization via a Chemoselective Serine Ligation. *J. Am. Chem. Soc.* 135, 6272–6279. doi:10.1021/ja4012468

Li, Y., Yu, H.-B., Zhang, Y., Leao, T., Glukhov, E., Pierce, M. L., et al. (2020). Pagoamide A, a Cyclic Depsipeptide Isolated from a Cultured marine Chlorophyte, *Derbesia* sp., Using MS/MS-based Molecular Networking. *J. Nat. Prod.* 83, 617–625. doi:10.1021/acs.jnatprod.9b01019

Lohani, C. R., Taylor, R., Palmer, M., and Taylor, S. D. (2015). Solid-phase Total Synthesis of Daptomycin and Analogs. *Org. Lett.* 17, 748. doi:10.1021/acs.orglett.5b00043

M07-A9 (2012). *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically; Approved Standard - Ninth Edition*. Wayne, PA: Clinical and Laboratory Standards Institute.

McIntosh, J. A., Donia, M. S., and Schmidt, E. W. (2009). Ribosomal Peptide Natural Products: Bridging the Ribosomal and Nonribosomal Worlds. *Nat. Prod. Rep.* 26, 537–559. doi:10.1039/b714132g

Mulani, M. S., Kamble, E. E., Kumkar, S. N., Tawre, M. S., and Pardesi, K. R. (2019). Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Front. Microbiol.* 10, 539. doi:10.3389/fmicb.2019.00539

Newman, D. J., and Cragg, G. M. (2020). Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi:10.1021/acs.jnatprod.9b01285

Singh, E. K., Ramsey, D. M., and McAlpine, S. R. (2012). Total Synthesis of Trans,trans-Sanguinamide B and Conformational Isomers. *Org. Lett.* 14, 1198–1201. doi:10.1021/ol203290n

Takayama, R., Hayakawa, S., Hinou, H., Albericio, F., and Garcia-Martin, F. (2018). Further Applications of Classical Amide Coupling Reagents: Microwave-Assisted Esterification on Solid Phase. *J. Pep Sci.* 24, e3111. doi:10.1002/psc.3111

Thomsen, I., Clausen, K., Scheibye, S., and Lawesson, S. O. (1984). Thiation with 2,4-Bis(4-Methoxyphenyl)-1,3,2,4- Dithiadiphosphetane 2,4-Disulfide: N-Methylthiopyrrolidone. *Org. Synth.* 62, 158. doi:10.15227/orgsyn.062.0158

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34, 828–837. doi:10.1038/nbt.3597

Wu, F., Yu, J., Meng, J., Guo, Y., and Ye, T. (2021). Total Synthesis of Pagoamide A. *Molecules* 26, 4224. doi:10.3390/molecules26144224

Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., et al. (2013). Molecular Networking as a Dereplication Strategy. *J. Nat. Prod.* 76, 1686–1699. doi:10.1021/np400413s

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fchem.2021.768854

Site-Specific Protein Modifications by an Engineered Asparaginyl Endopeptidase from *Viola canadensis*

Yu Chen[1,2], *Dingpeng Zhang*[1,2], *Xiaohong Zhang*[1,2], *Zhen Wang*[1,2], *Chuan-Fa Liu*[1,2,3] *and James P. Tam*[1,2,3]*

[1]School of Biological Sciences, Nanyang Technological University, Singapore, Singapore, [2]Synzymes and Natural Products Center, Nanyang Technological University, Singapore, Singapore, [3]Nanyang Institute of Structural Biology, Nanyang Technological University, Singapore, Singapore

Asparaginyl endopeptidases (AEPs) or legumains are Asn/Asp (Asx)-specific proteases that break peptide bonds, but also function as peptide asparaginyl ligases (PALs) that make peptide bonds. This ligase activity can be used for site-specific protein modifications in biochemical and biotechnological applications. Although AEPs are common, PALs are rare. We previously proposed ligase activity determinants (LADs) of these enzymes that could determine whether they catalyze formation or breakage of peptide bonds. LADs are key residues forming the S2 and S1′ substrate-binding pockets flanking the S1 active site. Here, we build on the LAD hypothesis with the engineering of ligases from proteases by mutating the S2 and S1′ pockets of VcAEP, an AEP from *Viola canadensis*. Wild type VcAEP yields <5% cyclic product from a linear substrate at pH 6.5, whereas the single mutants VcAEP-V238A (Vc1a) and VcAEP-Y168A (Vc1b) targeting the S2 and S1′ substrate-binding pockets yielded 34 and 61% cyclic products, respectively. The double mutant VcAEP-V238A/Y168A (Vc1c) targeting both the S2 and S1′ substrate-binding pockets yielded >90% cyclic products. Vc1c had cyclization efficiency of 917,759 $M^{-1}s^{-1}$, which is one of the fastest rates for ligases yet reported. Vc1c is useful for protein engineering applications, including labeling of DARPins and cell surface MCF-7, as well as producing cyclic protein sfGFP. Together, our work validates the importance of LADs for AEP ligase activity and provides valuable tools for site-specific modification of proteins and biologics.

Keywords: asparaginyl endopeptidases, peptide asparaginyl ligases, ligase-activity determinant, *Viola canadensis*, cell surface labeling, protein engineering

OPEN ACCESS

Edited by:
Hideo Iwaï,
University of Helsinki, Finland

Reviewed by:
Thomas Durek,
The University of Queensland, Australia
Francisco Solano,
University of Murcia, Spain

***Correspondence:**
James P. Tam
JPTam@ntu.edu.sg

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *01 September 2021*
Accepted: *06 October 2021*
Published: *22 October 2021*

Citation:
Chen Y, Zhang D, Zhang X, Wang Z, Liu C-F and Tam JP (2021) Site-Specific Protein Modifications by an Engineered Asparaginyl Endopeptidase from Viola canadensis. *Front. Chem. 9:768854. doi: 10.3389/fchem.2021.768854*

INTRODUCTION

Asparaginyl endopeptidases (AEP), also known as legumain and vacuolar processing enzymes (VPE), belong to the C13 family of cysteine proteases, which break peptide bonds after Asn/Asp (Asx) residues (Abe et al., 1993; Bottari et al., 1996; Shimada et al., 2003). AEP is initially synthesized as an inactive precursor which undergoes an acidic auto-activation in the vacuole (plants) or lysosome (mammals) (Müntz and Shutov 2002; Dall and Brandstetter 2016). AEPs are particularly well studied in plants and are multifaceted enzymes that display three distinct enzymatic functions: proteolysis, splicing, and ligation.

AEPs were discovered as endopeptidases in the early 1990s, but were also shown to be splicing enzymes that post-translationally modify the circularly permutated lectin concanavalin A (Min and Jones 1994; Nonis et al., 2021). Recently, certain AEPs were shown to be potent peptide Asx-specific ligases that can serve as bioprocessing enzymes for maturation of cyclic peptides (Nguyen et al., 2014; Bernath-Levin et al., 2015; Harris et al., 2015; Yang et al., 2017; Jackson et al., 2018; Zauner et al., 2018; Harris et al., 2019; Hemu et al., 2019; Hemu et al., 2020). As such, AEPs play important roles in processing seed storage proteins to release nutrients and bioactive peptides, forming seed coats, regulating programmed cell death, and generating host-defense antimicrobial and anti-fungal peptides (Hara-Nishimura, Inoue and Nishimura 1991; Saska et al., 2007; Hatsugai et al., 2015; Dall and Brandstetter 2016; Loo et al., 2016; Yamada et al., 2020).

Of particular interest to chemical biology is the discovery that AEPs have potent and specific ligase activity for forming Asx-peptide bonds. Our laboratory termed AEP-type ligases that can reverse the enzymatic direction from proteolysis to ligation as peptide Asn/Asp-specific ligases (PALs) (Nguyen et al., 2014; Hemu et al., 2019; Hemu et al., 2020; Tam et al., 2020). Butelase-1, the prototypic PAL, was isolated from a cyclotide-producing plant *Clitoria ternatea,* a plant which is known locally as bunga telang from which the name butelase is derived (Nguyen et al., 2014). Butelase-1 is highly efficient in ligating various Asx-containing peptides and proteins at pH 4–6.5, an acidic pH range which generally favors proteolysis by legumains (Nguyen et al., 2015b). Consequently, butelase-1 has become a model for engineering an AEP to a PAL. In addition, butelase1-like PALs are valuable tools for engineering and site-specific modification of proteins, and for theranostics. Such PALs can drive macrocyclization, orthogonal ligation, antibody-drug conjugation, and protein-protein fusion reactions (Nguyen et al., 2015a; Nguyen et al., 2015b; Cao et al., 2015; Cao et al., 2016; Hemu et al., 2016; Nguyen et al., 2016; Bi et al., 2017; Wang et al., 2021; Zhang et al., 2021).

AEPs and PALs share similar structures that include a core domain, linker region, and cap domain, with identical catalytic triads composed of Cys, His and Asn (**Supplementary Figure S1**). To date, 16 unique AEP and PAL structures have been reported in the protein data bank. However, their superimposed structures with <1.2 Å deviations show little difference and yield no clear clues about the enzymatic directionality toward protease or ligase activity (**Supplementary Figures S2, S3**, **Supplementary Tables S1, S2**). To understand the molecular determinants underpinning the directionality of AEP and PAL activity, our laboratory has focused on the substrate-binding sites flanking the S1 pocket. We used the Schechter and Berger's nomenclature of protease, substrate-binding pockets (Sn) which are the sub-sites of protease beside the active site (S1 pocket), to represent the region on the surface of an enzyme that can interact with peptide substrate residues (Pn) with specificity (**Supplementary Figure S4**) (Abramowitz, Schechter and Berger 1967). AEPs and PALs show high amino acid sequence conservation around the catalytic S1 pocket and have minor differences in the S2 and S1′ binding pockets. These sequence differences are termed ligase activity determinants (LADs) (Hemu et al., 2019). According to the LAD hypothesis, critical residues in the S2 and S1′ binding pockets that flank the catalytic S1 site are essential to steer the directionality of a legumain towards hydrolase or ligase activity (Hemu et al., 2019). Modification of these critical residues by site-directed mutagenesis was shown to either affect the enzymatic efficiency or alter the enzymatic directionality of AEPs. Our team performed extensive mutagenesis on OaAEP1b, an AEP originated from *Oldenlandia affinis* (Yang, et al., 2017). We found that the C247A mutation displayed enhanced kinetic efficiency relative to wild type. Accordingly, we coined the term "gate-keeper", a term preceding our use of "LAD1", to describe the importance of this site in controlling ligase activity. Recently, our group discovered asecond determinant for ligase activity which we named LAD2. We showed the importance of LAD2 for controlling ligase activity of AEPs by generating several AEP mutants derived from *Viola yedoensis* (VyPAL3-Y175G) and *Viola canadensis* (VcAEP-Y168A) (Hemu et al., 2019). Also, we successfully engineered butelase-2, which displays dominant protease activity at pH 4–6.5, to exhibit ligase activity, also at acidic pH, by mutagenesis of key LAD residues (Hemu et al., 2020). However, additional validation to strengthen the LAD hypothesis is needed.

Here we report engineering of VcAEP, an AEP from the plant *Viola canadensis*, based on the LAD hypothesis to reverse its enzymatic direction from an asparaginyl endopeptidase to a ligase. VcAEP is a dual-functional AEP that displays both protease and ligase activities. Our previous work has shown that modification of LAD2 site improved ligase activity of VcAEP. In the current study, we performed mutagenesis targeting the VcAEP S2 and S1′ pockets individually and jointly. We show that these mutants displayed enhanced ligase activity and diminished protease activity. Furthermore, we show that a double mutated VcAEP at both LAD sites is a highly efficient ligase and useful as a tool for protein labeling, sfGFP cyclization, and cell-surface labeling.

MATERIALS AND METHODS

Data-Mining Search of AEP Analogs

The core domains of butalase-1 and OaAEP1b were used to perform a data-mining search to discover novel AEPs having ligase activity. Results of this search were further narrowed in a second search that specified Val/Ile/Cys at "gate-keeper" sequences. VcAEP has 68.2% sequence identity and 85.7% similarity with butelase-1 and 65.9% identity and 89.1% similarity with OaAEP1b. VcAEP has a Val at the "gate-keeper" position, and thus is considered to be a ligase.

Expression, Purification, Activation of VcAEP

The full sequence of VcAEP was inserted into pET28a(+) and an N-terminal His-Ub tag was added to facilitate purification. The recombinant construct was transformed into *E. coli* SHuffle T7 cells, which were cultured at 30°C to OD_{600} 0.5. Then, 0.1 mM IPTG was added to induce VcAEP protein expression at 16°C for

~18 h. Cell pellets were collected and lysed in lysis buffer (50 mM HEPES, 300 mM NaCl, 1 mM $MgCl_2$, 1 μl/25 ml DNase I, 0.1% TritonX-100, 1mM PMSF, pH 7) with sonication. After centrifugation, the supernatant was collected and subjected to a three-step purification, including immobilized metal affinity chromatography, ion-exchange chromatography, and size exclusion chromatography. Activation of purified proenzyme was performed under acidic conditions (20 mM phosphate buffer, 0.5 mM N-Lauroylsarcosine, pH 4.5) at 37°C for 15–30 min. In addition, the activation was conducted by adding 1% or 5% acetic acid to adjust the pH to 4–4.5 with 0.5 mM N-Lauroylsarcosine, at 4°C for 10–16 h. The active form of the enzyme was purified by size exclusion chromatography using acidic buffer (20 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, pH 4). The purified active form of the enzyme was stored at −80°C. In-gel digestion followed by MS/MS *de novo* sequencing was performed to characterize autocleavage sites in the C terminus.

pH-Dependent Cyclization

To determine the optimal pH for cyclization activity of VcAEP and mutants, reaction mixtures containing the active enzyme (200 nM for wild type and 100 nM for Vc1a-1c) and substrate (4 μM for wild type, 20 μM for Vc1a, 50 μM for Vc1b, and 100 μM for Vc1c) were incubated at 37°C for 15 min according to the catalytic efficiency. The reaction buffers contain 20 mM phosphate, 1 mM EDTA, 1 mM DTT. The wild-type VcAEP displays slow enzymatic activity compared to the mutants, and no product was observed in 15 min using the same enzyme:substrate ratio as the mutants. Thus, for the wild type VcAEP, we increased the enzyme concentration from 100 to 200 nM. Similarly, we decreased the substrate concentration to 4 μM. Reaction products were analyzed by calculating the peak areas of MALDI-TOF mass spectrometry.

Kinetics Study Using FRET-Mediated Assay

A kinetics study of enzyme cyclization and ligation was conducted using fluorescence resonance energy transfer (FRET) at 37°C and pH 6.5. Fluorescence was recorded with a BioTek Cytation 5 cell imaging multimode plate reader at excitation and emission wavelengths of 390 and 460 nm, respectively. The initial rates V_0 (μM/s) for different substrate concentrations were calculated using the Michaelis-Menten equation. Vmax and Km and consequently kcat and kcat/Km were calculated using GraphPad Prism. All FRET substrates were synthesized by standard Fmoc-based solid-phase peptide synthesis on rink amide resin (Alsina et al., 2000). EDANS was introduced by using Fmoc-Glu(EDANS)-OH. DABCYL was directly coupled to ε-amine of Lys residue in dry DCM containing 2x DIEA of DABSYL for 2 h, at room temperature. All peptides were cleaved in 95% TFA/2.5% H_2O/2.5% TIS cleavage solution and purified by preparative RP-HPLC.

Applications

The synthetic fluorescent peptide EDANS-ANGI contains a signal "NGI" that can be recognized by ligases. DARPin9_26 was recombinantly expressed with an N-terminal His tag and TEV cleavage site. TEV cleavage exposes an N-terminal dipeptide "GL" that can function as an incoming nucleophilic attack group. The reaction was conducted at 37°C and pH 6.5 with 20 nM Vc1c and 4 μM DARPin9_26. DARPins that are successfully ligated with fluorescent peptide produce a green color under UV radiation.

SfGFP is used as an example of Vc1c-mediated protein circularization. Acyclic sfGFP was recombinantly expressed with additional PAL-recognizing tags (N-terminal Met-Ile and C-terminal Asn-Ser-Leu-His_6). The reaction was conducted at 37°C, pH 6.5, in the presence of 25 μM GFP and 0.5 μM Vc1c. The products were determined by high-resolution ESI-MS.

Application of Vc1c for MCF-7 cell surface labeling was performed as follows. First, MCF-7 cells cultured in plates were detached and resuspended in phosphate buffer (PBS, pH 7.4). Then, 50 μM of fluorescein-peptide and 200 nM of ligase were added to the solution and incubated at 37°C for 1 h. The cells were then washed 5 times with PBS to remove residual peptides and were resuspended in PBS buffer. FITC channels were used to detect fluorescein peptide-labeled cells.

RESULTS AND DISCUSSION

Expression, Purification, and Activation of VcAEP

The VcAEP sequence was obtained by data mining using the core domains of butelase-1 and OaAEP1b. VcAEP has 68.2% sequence identity and 85.7% similarity with butelase-1 and 65.9% identity and 89.1% similarity with OaAEP1b. To purify recombinant VcAEP, the cDNA sequence containing a His-Ub (hexa-histidine-ubiquitin) tag was synthesized and cloned into pET28a(+) (**Figure 1A**). The plasmid was expressed in SHuffle T7, an *E. coli* strain with enhanced folding capacity. The affinity-purified proenzyme of VcAEP was obtained using immobilized metal affinity chromatography (IMAC) followed by ion exchange (IEX) and size exclusion chromatography (SEC).

Similar to the previously reported OaAEP1b (Yang et al., 2017), the VcAEP proenzyme was obtained in two forms with or without a His-Ub tag (proform **1** and **2**) (**Figures 1A,B** left, and **Supplementary Figure S5**). Both forms were around 50 kDa on SDS-PAGE (**Figure 1B**, left upper), but only the proform **1** with a His-Ub tag was detected on western blot using an anti-His antibody (**Figure 1B**, left lower). The active form of VcAEP (activated form, **3**) was generated by incubation at 37°C for 30 min with 0.5 mM lauroylsarcosine to promote acidic activation at pH 4.5 (20 mM sodium phosphate buffer, 1 mM EDTA), followed by purification by SEC using pH 4 buffer (**Figure 1B**, middle and right). Autoactivation sites were determined by LC-MS/MS sequencing of the tryptic-digested active forms. The Asn/Asp cleavage sites at both ends of the core domain were N35 in the N-terminal region (NTR) and D343/D349/N360 in the cap domain (**Figure 1A** and **Supplementary Figure S6**). The cleavage sites of the truncated form and cap domain were also determined (**Supplementary Figures S7, S8**). The result implies that a multi-step cleavage is involved in the VcAEP acidic auto-activation. In addition, the processed activated form **3** of VcAEP retains a short cap domain, which might be helpful for the stabilization of activated form.

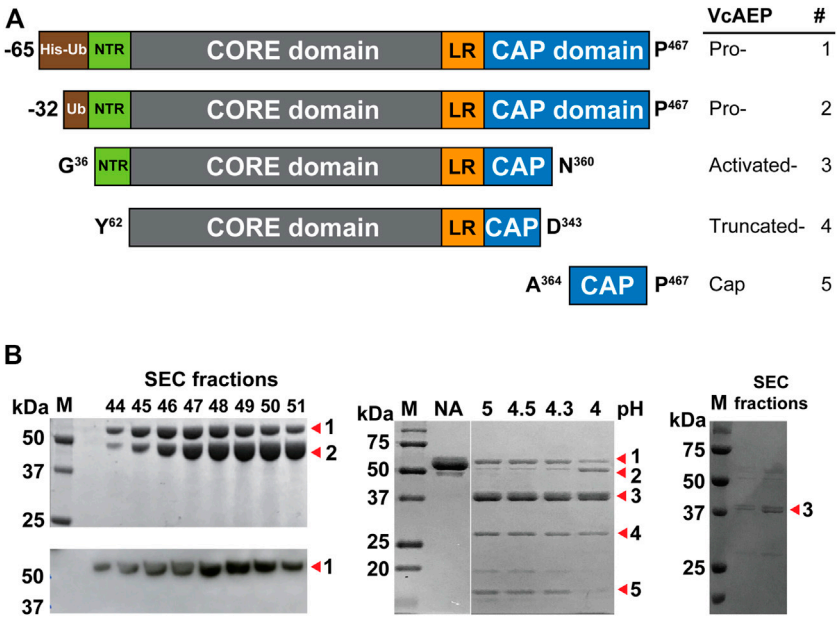


FIGURE 1 | Purification and activation of VcAEP. **(A)** Schematic representation of recombinant VcAEP constructs. Pro-VcAEP 1: full-length proenzyme form expressed with His-Ub-VcAEP (His = His6 tag, Ub = ubiquitin) and ending at P467. Pro-VcAEP 2: full-length VcAEP without the His-tag. Activated-VcAEP 3: active form of VcAEP after acidic autoactivation that begins with G36 and ends with N360. Truncated-VcAEP 4: truncated form of VcAEP, starting with Y62 and ending with D343. Cap 5: cleaved cap domain after acidic autoactivation, starting with A364 and ending with P467. The processing sites were determined by in-gel trypsin digestion followed by MS/MS *de novo* sequencing. NTR, N-terminal region; LR, linker region. **(B)** Purification, expression and activation of VcAEP. Left panel: analysis of purified fractions obtained from SEC chromatography, SDS-PAGE (upper) and western blot (lower). SDS-PAGE shows two bands, Pro-VcAEP 1 and Pro-VcAEP 2 (see panel A), with the lower band 2 being the VcAEP proenzyme without the His tag. Middle panel: SDS-PAGE analysis of the lauroylsarcosine-mediated acidic autoactivation of Pre-VcAEP 1 and Pro-VcAEP 2 and its products Activated-VcAEP 3, Truncated-VcAEP 4, and Cap 5, performed at pH 4–5 at 37°C for 30 min to generate Activated-VcAEP 3 (~37 kDa). NA: no activation. Right panel: SDS-PAGE analysis of purified Activated-VcAEP 3 by SEC chromatography using pH 4 buffer.

Characterization of VcAEP Activity

AEPs require a tripeptide recognition motif Asx-Xaa-Yaa (P1-P1′-P2′, using Schechter and Berger nomenclature) to make peptide bonds (Abramowitz et al., 1967). Generally, at the P1′ position, the substrate requirements are not stringent, although small amino acid residues are typically present. At P2′, bulky and aromatic amino acid residues are required for AEPs to carry out ligations (**Figure 2A**) (Nguyen et al., 2014; Jackson et al., 2020). We synthesized a 16-residue model peptide substrates GN^{14}-GI (GISTKSIPPISYAN^{14}-GI) to characterize the activity of VcAEP. The functional activity of VcAEP at pH 6 was initially assayed by MALDI-TOF mass spectrometry using GN^{14}-GI (MW 1618 Da) to determine its ligation efficiency to yield the 14-residue cyclic product cGN^{14} (MW 1429 Da) or its proteolytic activity to yield the linear product GN^{14} (MW 1447 Da) (**Figure 2B**). Under given conditions, both cyclic (15%) and linear products (5%) were generated, indicating that VcAEP is a dual-function enzyme that has both protease and ligase activities (**Figure 2C**).

The ligase activity of AEPs and PALs is highly dependent on pH (Hemu et al., 2019; Hemu et al., 2020). Thus, we performed the pH-dependent cyclization of VcAEP over the range of pH 4–8 using GN^{14}-GI. The reactions were conducted at 37°C for 15 min at the indicated pH. The yield of cyclic product (MW 1430 Da) and linear product (MW 1448 Da) were quantified in triplicate by calculating the peak areas of MALDI-TOF mass spectrometry. VcAEP displays a low ligase activity between pH 4–7 toward the peptide substrate GN^{14}-GI with a high enzyme concentration of 1:20 molar ratio of enzyme to substrate (**Figure 2D**).

Mutants Targeting S2 and S1′ Substrate-Binding Pockets Display Potent Ligase Activities

Sequence alignment of known AEPs and ligases shows that the S1 and S3 pockets are highly conserved. Meanwhile, the S2′ pocket contains four residues that have diverse sequences and are located relatively far from the catalytic S1 pocket (**Figures 3A,B**). Sequence variations of S2 and S1′ substrate-binding pockets are vital factors to control the ligase activity of AEPs. The S2 pocket, also known as the "gate-keeper" or LAD1, comprises three residues, of which the first residue is an aromatic Trp or Tyr and the third residue is hydrophobic residues Ala/Val or neutral residues Thr/Tyr. The ligases tend to have Val/Ile/Cys as the middle residue, whereas proteases often have Gly. The S1′ pocket, also known as LAD2, has two residues. Ligases often have Gly-Ala or Ala-Ala or Ala-Pro, and AEPs typically have Gly-Pro. It should be noted that the ligase activity of AEP is not solely determined by substrate-binding pockets, but rather results from cooperation of different factors (Haywood et al., 2018; Jackson et al., 2018).

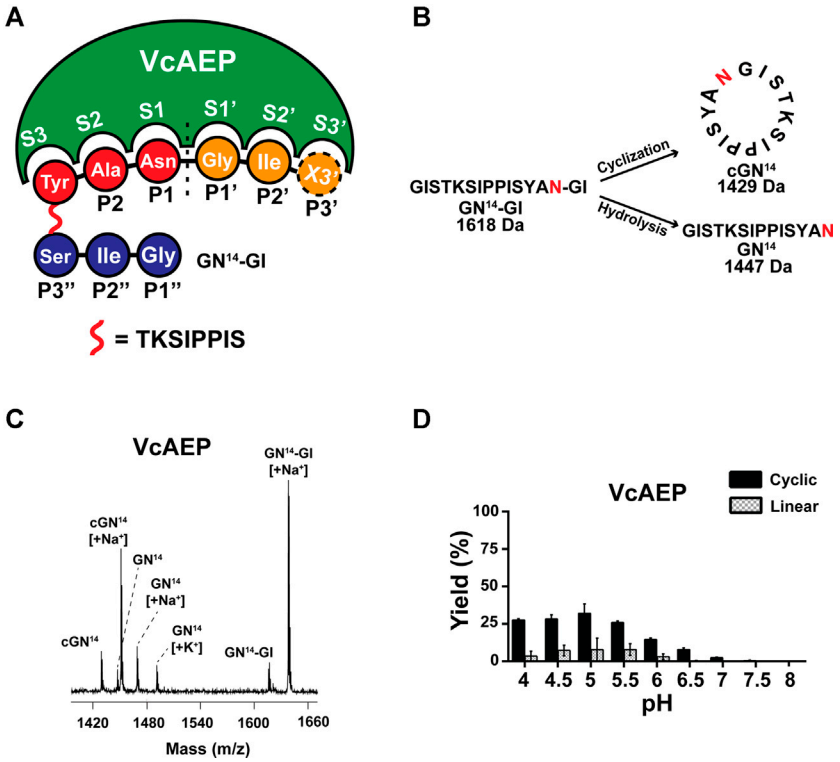


FIGURE 2 | Analysis of VcAEP AEP activity. **(A)** Nomenclature of AEP/PAL and substrate, using the peptide substrate GN^{14}-GI (GISTKSIPPISYAN-GI) as a model. Based on the nomenclature of substrate binding of protease proposed by Schechter and Berger, the cleavage site of the substrate is named P1 (red), and the corresponding binding pocket of the enzyme is S1. After cleavage, the amino acid residues of leaving group are P1′, P2′, P3′ Etc (orange). For PALs to cyclize or ligate substrate, the amino acid residues of the incoming group are P1″, P2″, P3″ etc., (blue). X stands for the 20 amino acids. **(B)** Schematic representations of AEP-mediated cyclization and hydrolysis of the 16-residue peptide substrate GN^{14}N-GI (MW 1618 Da). The P1-Asn position is colored red. The enzymatic reaction of VcAEP yields two products: the 14-residue cyclic product cGN^{14} (1429 Da) and 14-residue linear product GN^{14} (1447 Da) with release of a GI dipeptide. **(C)** Representative MALDI-TOF mass spectrometry of the cyclic product cGN^{14} and the linear product GN^{14} generated by VcAEP together with GN^{14}N-GI as the starting material. The reaction was conducted at pH 6, at 37°C for 15 min with a 1:20 molar ratio of enzyme:substrate. **(D)** Product yields of VcAEP wild type were quantified by calculating the peak areas of MALDI-TOF mass spectrometry. Gray and black bars indicate the percentage yield of linear product and cyclic product, respectively. The reactions were conducted at 37°C for 15 min with a molar enzyme:substrate ratio of 1:20 (VcAEP). Average yield and SDs were calculated from experiments performed in triplicate.

Previously, our laboratory engineered eight PAL-like mutants of butelase-2 which is an AEP with protease activity at acidic pH. These PAL-like mutants targets both the LAD1 and LAD2 sites (see **Figure 3B**). They include Bu2a (V/GP), Bu2b (I/GP), and Bu2c (A/GP), targeting the middle residue of LAD1 motif Gly252 by substituting it with Val, Ile, or Ala, respectively. We also prepared LAD2 mutants targeting Gly182-Pro183 by substituting them with Ala, giving Bu2d (G/GA) (butelase-1-like) and Bu2e (G/AP) (VyPAL2-like). In addition, we engineered double-site-LAD mutants: Bu2f (V/AP) (VyPAL2-like), Bu2g (V/GA which is butelase-1-like), and Bu2h (V/AA whichOaAEP1b-like). The best engineered butelase-2 mutant acting as a butelase-1 was Bu2g which contains mutations in both LAD1 and LAD2 sites (Hemu et al., 2020).

To confirm our mutation study of converting butelase-2 from a protease to a ligase, we herein focus on engineering the S2 and S1′ pockets of VcAEP corresponding to the LAD1 and LAD2 sites, respectively, to promote its ligase activity and demonstrate the essential roles these factors play in AEP ligase activity. According to the LAD hypothesis, VcAEP is a bifunctional AEP with a Trp-Val-Ala sequence in the S2 pocket and thus is considered a ligase. In contrast, the S1′ pocket sequence of VcAEP is Tyr-Pro, which does not fit the criteria for ligases. To validate the LAD hypothesis, we engineered three mutants: VcAEP-V238A or Vc1a that has an Ala substituted for Val in the middle residue in the LAD1 motif at position 238 (V238A); VcAEP-Y168A or Vc1b that has the LAD2 residue Y168 changed to Ala; and VcAEP-V238A/Y168A or Vc1c, a double mutant targeting both LAD1 and LAD2 pockets (**Figure 3B**). We used a similar protocol to express, purify, and activate these mutants as was used for wild type VcAEP (**Supplementary Figure S9**). Although Vc1a and Vc1b differ by a single amino acid, they migrated differently in SDS-PAGE as bands MW <37 and >37 kDa, respectively. We do not have an explanation for this result. However, we found that Vc1c and Vc1a which differ also by a single amino acid migrated similarly in the SDS-PAGE.

To compare the ligase activity of the mutants Vc1a-1c to the wild type VcAEP, we performed the pH-dependent cyclization reactions mediated by Vc1a-1c. The ligation reactions were performed over the pH range of pH 4 to pH 8 using the peptide substrate GN^{14}-GI, at 37°C for 15 min. The yield of

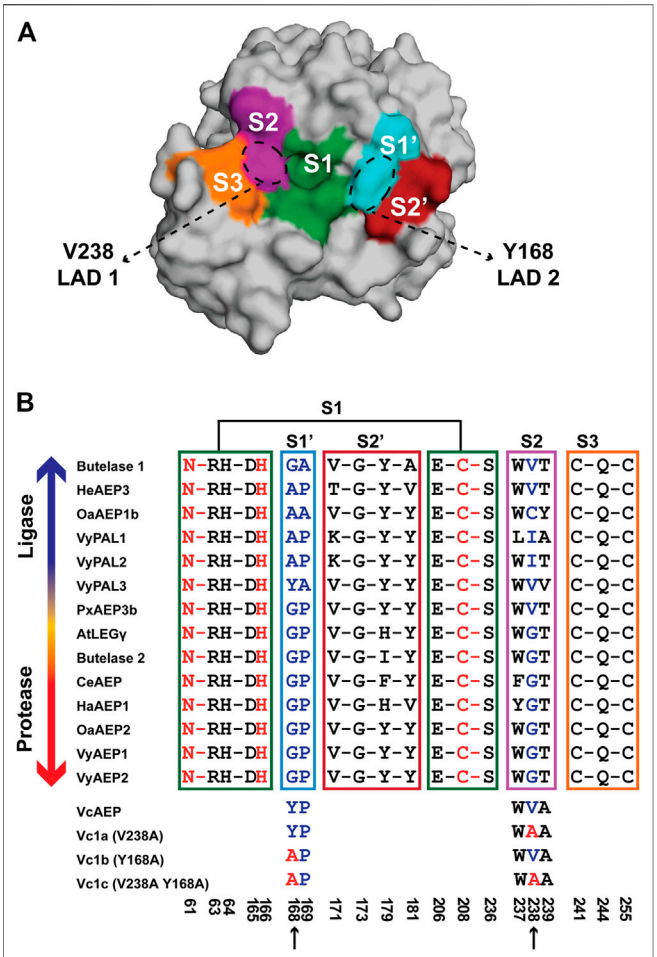


FIGURE 3 | Engineering a ligase from a dual-functional AEP, VcAEP. **(A)** Proposed substrate binding pockets (S3, S2, S1, S1′, S2′) based on the crystal structure of the VcAEP proform (pdb: 5ZBI). The S3 pocket (orange) comprises residues C241, Q244, and C255, S2 (purple) W237, V238, and A239, S1 (green) N61, R63, H64, D165, H166, E206, C208, and S236, S1′ (cyan) Y168 and P169, S2′ (red) V171, G173, Y179, and Y181. **(B)** Sequence alignment of substrate-binding pockets of known AEPs for which protease activity dominates (e.g., CeAEP, butelase-2), and PALs having dominant ligase activity (e.g., butelase-1, VyPAL2). The blue arrow indicates increasing ligase activity and the red arrow indicates increasing protease activity. The three engineered VcAEP mutants having mutations at S2 (V238 using VcAEP numbering) and S1′ (Y168) sites are Vc1a (V238A), Vc1b (Y168A), and Vc1c (V238A Y168A). Color codes are the same as for **(A)** and **(B)** with S3 in orange, S2 in purple, S1 in green, S1′ in cyan, and S2 in red. The catalytic triad N61, H166, and C208 in the oxyanion hole corresponding to S1 is colored red. Key residues in the S1′ and S2 pockets for steering AEP or PAL activities (panel B) are respectively colored cyan for Y168P169, and purple for V238. Residues comprising the catalytic S1, S1′, S2′, S2 and S3 pockets are framed by a green, cyan, red, purple and orange box, respectively. Three VcAEP mutants with increasing ligase activity were engineered that target S2 and S1′ pockets (indicated by black arrows).

cyclic product (MW 1430 Da) and linear product (MW 1448 Da) were quantified in triplicate by MALDI-TOF mass spectrometry.

Vc1a (VcAEP-V238A), a LAD1 mutant that has a V238A mutation in the S2 binding pocket, displayed dominant protease activity at pH < 6, but dominant ligase activity at pH > 6.5 using the GN^{14}-GI peptide substrate at a 1:200 molar ratio of enzyme to substrate. Using the same reaction time and temperature, Vc1a consumed substrate at a much faster rate than wild type VcAEP. As such, we used a 1:200 enzyme:substrate molar ratio for Vc1a and 1:20 for VcAEP because Vc1a is a "faster" enzyme compared to VcAEP (**Figure 4A**).

The process for AEP and PAL to produce linear or cyclic products mainly involves two steps: formation of an S-acyl intermediate that then undergoes nucleophilic attack by a water molecule (hydrolysis) or a free amine of the peptide substrate (ligation). LAD1 is proposed to affect the formation of the S-acyl intermediate, which is the rate-limiting step, by shifting the molecular surface toward the incoming peptide substrate to affect the overall reaction rate (Hemu et al., 2020). Therefore, substitution of Ala for Val at position 238 of LAD1 is likely to accelerate the formation of the S-acyl intermediate and the overall catalysis process, including the production of linear and cyclic products.

Vc1b (VcAEP-Y168A) has a Tyr to Ala mutation in the LAD2 pocket (Y168A) and exhibited dominant ligase activity at pH > 4.5 and no visible hydrolytic product at pH > 6.5 using the peptide substrate GN^{14}-GI at a 1:500 molar ratio of enzyme to substrate. At pH 6.5, about 60% cyclization yield was observed (**Figure 4B**). This result supports the LAD hypothesis that LAD2 plays a more direct role in determining AEP enzyme directionality. The wild type VcAEP has a Tyr-Pro dipeptide in LAD2, and the bulky Tyr would likely accelerate the departure of the leaving group, such that water molecules would have increased opportunity to attack the S-acyl intermediate and in turn generate more hydrolysis product. However, Vc1b with an Ala-Pro dipeptide in LAD2 would retain the leaving group for a longer time to allow selectively an incoming amine group to be ligated.

The double mutant Vc1c (VcAEP-V238A/Y168A) with mutations in both LAD1 and LAD2 showed the fastest activity of all four enzymes tested. Moreover, Vc1c generated >80% cyclic product between pH 5 and 6.5 using the peptide substrate GN^{14}-GI at a 1:1000 molar ratio of enzyme:substrate (**Figure 4C**). Engineering both LAD1 and LAD2 in Vc1c integrated the advantages of substitutions at both sites to enhance catalytic efficiency of peptide bond formation. Together, our results show that both LAD mutants, Vc1b and Vc1c, display significant ligase activity at physiological pH and could be valuable for peptide/protein engineering applications.

Kinetic Studies of VcAEP and Vc1a-1c

To figure out the catalytic potency of the VcAEP and its mutants, we determined their catalytic efficiency using the fluorescence resonance energy transfer (FRET)-base assay. FRET substrates were synthesized by standard Fmoc-based solid-phase peptide synthesis on rink amide resin (Alsina et al., 2000). The fluorophore EDANS was introduced by using Fmoc-Glu(EDANS)-OH. DABCYL was directly coupled to the ε-amine of Lys residue.

We initially performed the kinetics study for the wild type VcAEP and its mutant Vc1a using a 17-residue peptide substrate

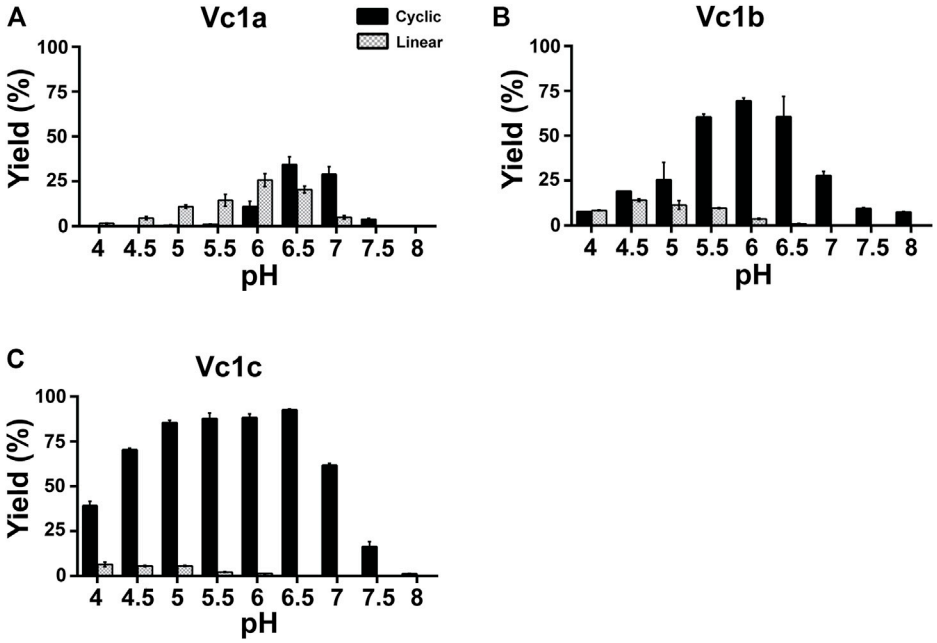


FIGURE 4 | pH-dependent cyclization efficiency of VcAEP and mutants. **(A–C)** Quantitative summary of VcAEP mutant Vc1a-Vc1c product yield analyzed using MALDI-TOF mass spectrometry. Gray and black bars indicate the percentage yield of linear product and cyclic product, respectively. The reactions were conducted at 37°C for 15 min with a molar enzyme:substrate ratio of 1:200 (Vc1a), 1:500 (Vc1b), and 1:1000 (Vc1c). Average yield and SDs were calculated from experiments performed in triplicate.

TABLE 1 | Kinetic parameters of VcAEP and Vc1a-1c.

	Enzyme	kcat (s^{-1})	Km (μM)	kcat/Km (s^{-1} M^{-1})	Substrate sequence	Enzyme concentration (μM)
Cyclization	VcAEP	0.01863	4.173	4,464	GISTKSIPPIE(EDANS)YRN-SLK(DABCYL)	0.438
	Vc1a	0.2025	13.77	14,700	GISTKSIPPIE(EDANS)YRN-SLK(DABCYL)	1.39
	Vc1b	0.03471	1.719	200,192	GISKPE(EDANS)SYAN-GIK(DABCYL)	0.06
	Vc1c	2.31	2.517	917,759	GISKPE(EDANS)SYAN-GIK(DABCYL)	0.044
Ligation	Vc1c	0.3761	33.68	11,166	N-terminal: RE(EDANS)AN-GI C-terminal: GVK(DABCYL)	0.044

GISTKSIPPIE(EDANS)YRN-SLK(DABCYL). The determined kcat/Km of VcAEP and Vc1a using substrate were 4,464 M^{-1} s^{-1} and 14,706 M^{-1} s^{-1}, respectively (**Table 1**). The kcat of Vc1a (0.2025 s^{-1}) was about 11-fold higher than that for VcAEP (0.01863 s^{-1}), indicating that Vc1a can consume substrate at a much faster rate than VcAEP. Vc1a also shows a 3-fold increase in kcat relative to VcAEP, suggesting that altering the S2 pocket influences substrate binding such that Vc1a needs more substrate than VcAEP to perform its activity.

The peptide substrate GISTKSIPPIE(EDANS)YRN-SLK(DABCYL) was difficult to purify. Consequently, we synthesized a 13-residue peptide substrate GISKPE(EDANS) SYAN-GIK(DABCYL) to study the kinetics of Vc1b and Vc1c. It should be pointed out that both substrates give highly favored cyclic peptides of 14 and 10 membered rings. Compared with the turnover rate (kcat) in intra-molecular ligation for wild type VcAEP (0.01863 s^{-1}), the kcat for Vc1b and Vc1c was found to increase by 2- (0.03471 s^{-1}) and 120-fold (2.31 s^{-1}), respectively. Together with the affinity constant Km, the catalytic efficiency (kcat/Km) of Vc1b and Vc1c in intra-molecular ligation was calculated to be 20,192 M^{-1} s^{-1} and 917,759 M^{-1} s^{-1}, respectively (**Table 1**). In this regard, Vc1c shows a comparable cyclization activity to that of butelase-1 (GISTKSIPPYRN-SLAN) (Hemu et al., 2019). As such, Vc1c is one of the most efficient peptide ligases reported thus far.

Site-specific modifications of proteins are generally inter-molecular ligation reactions. To determine the capability of Vc1c in inter-molecular reactions, we synthesized an N-terminal peptide REAN-GI with the fluorophore (EDANS) group and a C-terminal peptide GVKR with the quencher (DABCYL) group. Calculation of Michaelis-Menten kinetics for Vc1c ligation based on FRET assay results showed a kcat/Km of 11,166 M^{-1} s^{-1} (**Table 1**) suggesting that the

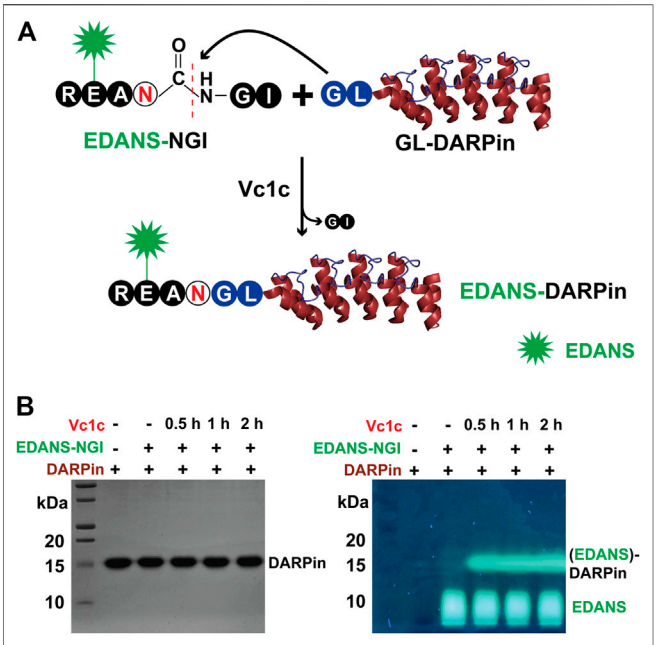


FIGURE 5 | Vc1c-mediated protein labelling. **(A)** Schematic of Vc1c-mediated ligation of DARPin (generated from PDB code 2QYJ) with fluorescent peptide RE(EDANS)AN-GI. A fluorescent peptide RE(EDANS)AN-GI was synthesized and DARPin9_26 was co-expressed with an N-terminal Gly-Leu dipeptide. The ligase can recognize and cleave the C-terminal Gly-Ile of the fluorescent peptide, and ligate it to the GL-DARPin9_26 N terminus. **(B)** Left panel: SDS-PAGE gel after Coomassie blue staining showing bands for DARPin9_26 around 15 kDa; Right: SDS-PAGE gel under UV illumination. The reaction was performed at 37°C, pH 6.5 with 20 nM Vc1c and 4 μM DARPin9_26.

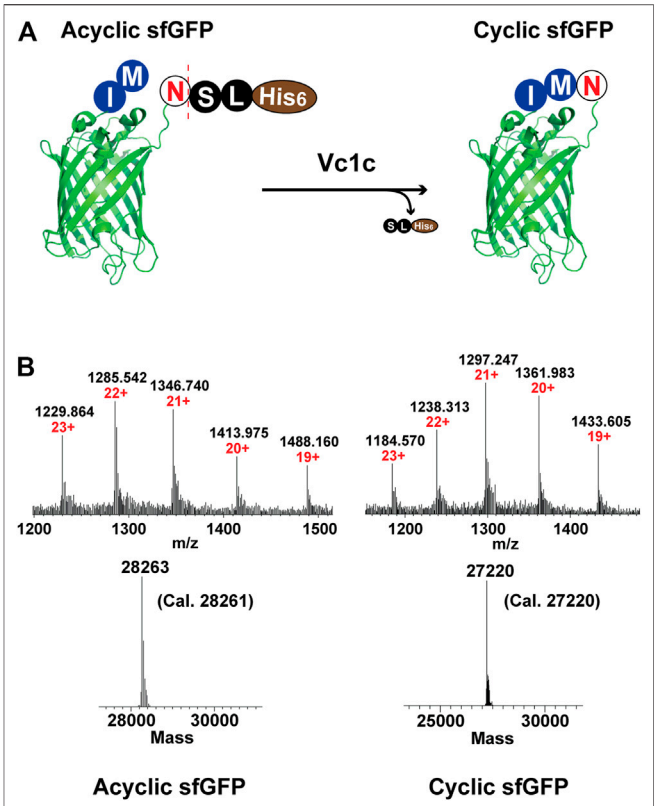


FIGURE 6 | Application of Vc1c in end-to-end cyclization of sfGFP. **(A)** Scheme of ligase-mediated sfGFP (generated from PDB code 2B3P) cyclization. sfGFP was co-expressed with an N-terminal Met-Ile and a C-terminal Asn-Ser-Leu-His6. The ligase can recognize and cleave the C-terminal motif Ser-Leu-His6 and perform end-to-end cyclization of sfGFP. **(B)** Upper: ESI-MS characterization of CM-sfGFP-NSL-His6 starting material or acyclic sfGFP (left) and cyclic sfGFP (right); Lower: Deconvolution of sfGFP cyclization (CM-sfGFP-NSL-His6: Cal. 28261, Obs. 28263; Cyclic sfGFP: Cal. 27220, Obs. 27220). The cyclization reaction was performed at pH 7.0, 37°C for 30 min using 25 μM CM-sfGFP-NSL-His6 and 500 nM Vc1c.

intermolecular ligation is many fold slower than intramolecular ligation which has substantially higher effective molarity than intermolecular reactions. .

Application of Vc1c in Site-Specific Protein Modifications

The catalytic parameters showed that Vc1c displayed a substantially improved catalytic efficiency compared to wild type VcAEP. Next, we showed the versatility of the engineered Vc1c as a useful tool for chemical biology, site-specific protein modification, and cell-surface labeling.

Vc1c-Mediated N-Terminal Labeling of DARPin

We first tested the activity of Vc1c in the labeling of DARPins, a genetically engineering antibody-mimetic protein based on ankyrin repeat proteins (Bork 1993). DARPins conjugated with different molecules have been used for various applications (Stumpp, Binz and Amstutz 2008). DARPin can be selected from a library to bind to any desired target, such as human epidermal growth factor receptor 2 (HER2), a transmembrane protein overexpressed on the surface of some breast cancer cells (Ross 2009). DARPins represent a new generation of protein therapeutics, and the development of various methods for DARPin labeling is of great importance. Vc1c was able to mediate DARPin9_26 N-terminal labeling. First, a fluorescence peptide RE(EDANS)AN-GI was synthesized as an N-terminal peptide. Then, GL-DARPin9_26 having the N-terminal residues "Gly-Leu" was recombinantly expressed (**Figure 5A**). The reaction was performed at 37°C, pH 6.5 with 20 nM Vc1c and 4 μM DARPin9_26 and was complete within 30 min. This result showed that Vc1c efficiently ligated DARPin9_26 to fluorescent peptides (**Figure 5B**).

Vc1c-Mediated GFP Cyclization

Green Fluorescent Protein (GFP) is a widely used reporter for many applications including studies of protein folding, gene translation, and protein-protein interactions (Kalir et al., 2001; Waldo 2003; Magliery et al., 2005). A GFP variant, "super folder" GFP (sfGFP), was developed to reduce aggregation and increase folding efficiency (Pédelacq et al., 2006). In addition, the circular topology of proteins can contribute to improved stability. To demonstrate the feasibility of Vc1c to

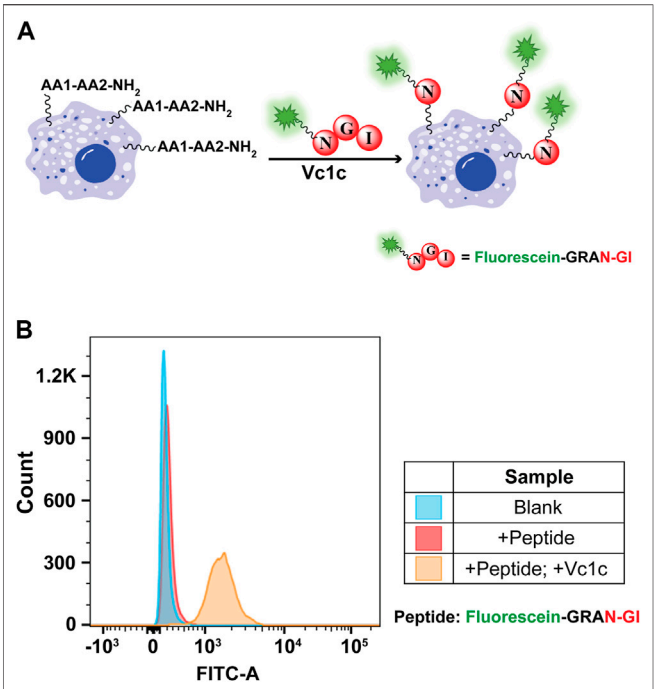


FIGURE 7 | Vc1c-mediated MCF-7 cell labelling. **(A)** Schematic of MCF-7 cell surface amine labelling with the peptide fluorescein-GRANGI using Vc1c. **(B)** Flow cytometry analysis of MCF-7 labelling with the peptide fluorescein-GRANGI. Untreated cells with neither peptide nor ligase added and cells with only peptide added were used as negative controls. The reaction was conducted at 37°C for 1 h using 50 μM fluorescein-peptide and 200 nM Vc1c.

catalyze end-to-end protein cyclization, we expressed sfGFP with an additional N-terminal Met-Ile and a C-terminal Asn-Ser-Leu-His6 (**Figure 6A**). The cyclization reaction was performed in the presence of 25 μM GFP and 0.5 μM Vc1c. The reaction was complete within 30 min, as determined by high-resolution ESI-MS (**Figure 6B**).

Vc1c-Mediated MCF-7 Cell Surface Labeling

AEP-mediated cell surface labeling is a promising method for cell surface modifications (Bi et al., 2017; Harmand et al., 2021). Previous study has shown that AEP can be used for cell surface modification by specifically labeling the C-terminal hTFR-NHV (Bi et al., 2020). However, the labeling efficiency is limited due to the unsatisfactory amount of receptor protein was expressed on the cell surface. To establish a convenient and highly efficient labeling strategy, we explore the use of PAL enzyme for unspecific cell surface N-terminal amine labeling. To perform the reaction, MCF-7 cells (a typical cancer cell line) were detached and suspended in PBS buffer. The reaction was conducted using 50 μM of fluorescein-GRAN-GI and 200 nM of Vc1c at 37°C for 1 h. The cells were then washed and suspended in PBS buffer for flow cytometry with FITC channels used to detect fluorescein peptide-labeled cells (**Figure 7A**). MCF-7 cells were successfully labeled with Vc1c in the presence of fluorescein peptides (**Figure 7B**). These findings show that Vc1c is an efficient ligase for the modification of cell surface proteins, and thus this engineered protein could be valuable for biotechnology applications.

CONCLUSION

In this report, we successfully engineer an AEP from *Viola canadensis* to an AEP-type ligase or PAL by mutating the residues that reside in the substrate-binding pockets flanking the S1 active site. Ligases are useful for site-specific modification proteins and cell-surface labeling which are valuable tools for biotechnology, drug development, and theranostics. A prerequisite requirement is that they can be used under mild conditions, preferably under physiological conditions. PALs fulfill this requirement. Although PALs and AEPs share similar structure and substrate specificity, PALs are rare compared to AEPs. Thus, the ability to engineer a more commonly found AEP to a PAL will provide an additional ligase with a different ligase profile to expand the repertoire of PALs which can be used alone or combined with another ligase for tandem, orthogonal, and one-pot ligation (Harmand et al., 2018; Wang et al., 2021; Zhang et al., 2021). In addition, we have found that these AEP-type ligases have different tripeptide recognition signal, substrate specificity and optimal pH for ligation reactions. Importantly, the mutant Vc1c can be expressed and purified in yields ranging from 5 to 10 mg/L of bacteria culture. The engineered ligase Vc1c is one of the most efficient ligases known to date. In addition, we show that the engineered enzyme Vc1c is a versatile tool in both peptide and protein cyclization, protein and cell-surface labeling, comparable to the natural or recombinant expressed butelase-1 (Nguyen et al., 2014; Hemu et al. 2021). Together, we provide a method to engineer a ligase and valuable tools for site-specific modification of proteins which can be further used in theranostic applications.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.rcsb.org/structure/5ZBI, https://www.rcsb.org/structure/2QYJ, https://www.rcsb.org/structure/2B3P, https://www.rcsb.org/structure/4FGU, https://www.rcsb.org/structure/4NOK, https://www.rcsb.org/structure/4D3Y, https://www.rcsb.org/structure/5H0I, https://www.rcsb.org/structure/5NIJ, https://www.rcsb.org/structure/6DHI, https://www.rcsb.org/structure/6IDV, https://www.rcsb.org/structure/6L4V, https://www.rcsb.org/structure/6L4W, https://www.rcsb.org/structure/6L4X, https://www.rcsb.org/structure/4D3X, https://www.rcsb.org/structure/5LUA, https://www.rcsb.org/structure/6AZT, https://www.rcsb.org/structure/5OBT.

AUTHOR CONTRIBUTIONS

YC, DZ, XZ, C-FL, and JT designed research; YC, DZ, XZ, and ZW performed research; YC, DZ, and JT analyzed data; and YC, DZ, and JT wrote the paper.

FUNDING

This research was supported by Academic Research Grant Tier 3 (MOE2016-T3-1-003) from Singapore Ministry of Education (MOE), and Synzyme and Natural Products Center (SYNC, 001304-00001) at Nanyang Technological University.

ACKNOWLEDGMENTS

We would like to thank Assoc Prof Julien Lescar, Hemu Xinya and Wong Yeehwa at the School of Biological Sciences, NTU, Singapore, for their valuable advices and supports. We also thank Professor Lloyd Ruddock (University of Oulu, Finland) for providing the disulfide-promoting plasmids.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.768854/full#supplementary-material

REFERENCES

Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I., et al. (1993). Asparaginyl Endopeptidase of jack Bean Seeds. Purification, Characterization, and High Utility in Protein Sequence Analysis. *J. Biol. Chem.* 268, 3525–3529. doi:10.1016/s0021-9258(18)53726-1

Abramowitz, N., Schechter, I., and Berger, A. (1967). On the Size of the Active Site in Proteases II. Carboxypeptidase-A. *Biochem. Biophysical Res. Commun.* 29, 862–867. doi:10.1016/0006-291x(67)90299-9

Alsina, J., Yokum, T. S., Albericio, F., and Barany, G. (2000). A Modified Backbone Amide Linker (BAL) Solid-phase Peptide Synthesis Strategy Accommodating Prolyl, N-Alkylamino Acyl, or Histidyl Derivatives at the C-Terminus. *Tetrahedron Lett.* 41, 7277–7280. doi:10.1016/s0040-4039(00)01255-7

Bernath-Levin, K., Nelson, C., Elliott, A. G., Jayasena, A. S., Millar, A. H., Craik, D. J., et al. (2015). Peptide Macrocyclization by a Bifunctional Endoprotease. *Chem. Biol.* 22, 571–582. doi:10.1016/j.chembiol.2015.04.010

Bi, X., Yin, J., Nguyen, G. K. T., Rao, C., Halim, N. B. A., Hemu, X., et al. (2017). Enzymatic Engineering of Live Bacterial Cell Surfaces Using Butelase 1. *Angew. Chem. Int. Ed.* 56, 7822–7825. doi:10.1002/anie.201703317

Bi, X., Yin, J., Zhang, D., Zhang, X., Balamkundu, S., Lescar, J., et al. (2020). Tagging Transferrin Receptor with a Disulfide FRET Probe to Gauge the Redox State in Endosomal Compartments. *Anal. Chem.* 92, 12460–12466. doi:10.1021/acs.analchem.0c02264

Bork, P. (1993). Hundreds of Ankyrin-like Repeats in Functionally Diverse Proteins: Mobile Modules that Cross Phyla Horizontally. *Proteins* 17, 363–374. doi:10.1002/prot.340170405

Bottari, A., Capocchi, A., Galleschi, L., Jopova, A., and Saviozzi, F. (1996). Asparaginyl Endopeptidase during Maturation and Germination of Durum Wheat. *Physiol. Plant* 97, 475–480. doi:10.1111/j.1399-3054.1996.tb00506.x

Cao, Y., Nguyen, G. K. T., Chuah, S., Tam, J. P., and Liu, C.-F. (2016). Butelase-mediated Ligation as an Efficient Bioconjugation Method for the Synthesis of Peptide Dendrimers. *Bioconjug. Chem.* 27, 2592–2596. doi:10.1021/acs.bioconjchem.6b00538

Cao, Y., Nguyen, G. K. T., Tam, J. P., and Liu, C.-F. (2015). Butelase-mediated Synthesis of Protein Thioesters and its Application for Tandem Chemoenzymatic Ligation. *Chem. Commun.* 51, 17289–17292. doi:10.1039/c5cc07227a

Dall, E., and Brandstetter, H. (2016). Structure and Function of Legumain in Health and Disease. *Biochimie* 122, 126–150. doi:10.1016/j.biochi.2015.09.022

Hara-Nishimura, I., and Hatsugai, N. (2011). The Role of Vacuole in Plant Cell Death. *Cell Death Differ* 18, 1298–1304. doi:10.1038/cdd.2011.70

Hara-Nishimura, I., Inoue, K., and Nishimura, M. (1991). A Unique Vacuolar Processing Enzyme Responsible for Conversion of Several Proprotein Precursors into the Mature Forms. *FEBS Lett.* 294, 89–93. doi:10.1016/0014-5793(91)81349-d

Hara-Nishimura, I., Shimada, T., Hiraiwa, N., and Nishimura, M. (1995). Vacuolar Processing Enzyme Responsible for Maturation of Seed Proteins. *J. Plant Physiol.* 145, 632–640. doi:10.1016/s0176-1617(11)81275-7

Harmand, T. J., Bousbaine, D., Chan, A., Zhang, X., Liu, D. R., Tam, J. P., et al. (2018). One-pot Dual Labeling of IgG 1 and Preparation of C-To-C Fusion Proteins through a Combination of Sortase A and Butelase 1. *Bioconjug. Chem.* 29, 3245–3249. doi:10.1021/acs.bioconjchem.8b00563

Harmand, T. J., Pishesha, N., Rehm, F. B., Ma, W., Pinney, W. B., Xie, Y. J., et al. (2021). Asparaginyl Ligase-Catalyzed One-step Cell Surface Modification of Red Blood Cells. *ACS Chem. Biol.* 16(7), 1201–1207. doi:10.1021/acschembio.1c00216

Harris, K. S., Durek, T., Kaas, Q., Poth, A. G., Gilding, E. K., Conlan, B. F., et al. (2015). Efficient Backbone Cyclization of Linear Peptides by a Recombinant Asparaginyl Endopeptidase. *Nat. Commun.* 6, 1–10. doi:10.1038/ncomms10199

Harris, K. S., Guarino, R. F., Dissanayake, R. S., Quimbar, P., McCorkelle, O. C., Poon, S., et al. (2019). A Suite of Kinetically superior AEP Ligases Can Cyclise an Intrinsically Disordered Protein. *Sci. Rep.* 9, 1–13. doi:10.1038/s41598-019-47273-7

Hatsugai, N., Yamada, K., Goto-Yamada, S., and Hara-Nishimura, I. (2015). Vacuolar Processing Enzyme in Plant Programmed Cell Death. *Front. Plant Sci.* 6, 234. doi:10.3389/fpls.2015.00234

Haywood, J., Schmidberger, J. W., James, A. M., Nonis, S. G., Sukhoverkov, K. V., Elias, M., et al. (2018). Structural Basis of Ribosomal Peptide Macrocyclization in Plants. *Elife* 7, e32955. doi:10.7554/eLife.32955

Hemu, X., El Sahili, A., Hu, S., Wong, K., Chen, Y., Wong, Y. H., et al. (2019). Structural Determinants for Peptide-Bond Formation by Asparaginyl Ligases. *Proc. Natl. Acad. Sci. U S A.* 116, 11737–11746. doi:10.1073/pnas.1818568116

Hemu, X., El Sahili, A., Hu, S., Zhang, X., Serra, A., Goh, B. C., et al. (2020). Turning an Asparaginyl Endopeptidase into a Peptide Ligase. *ACS Catal.* 10, 8825–8834. doi:10.1021/acscatal.0c02078

Hemu, X., Qiu, Y., Nguyen, G. K. T., and Tam, J. P. (2016). Total Synthesis of Circular Bacteriocins by Butelase 1. *J. Am. Chem. Soc.* 138, 6968–6971. doi:10.1021/jacs.6b04310

Hemu, X., Zhang, X., Nguyen, G. K., To, J., Serra, A., Loo, S., et al. (2021). Characterization and application of natural and recombinant butelase-1 to improve industrial enzymes by end-to-end circularization. *RSC Advances* 11, 23105–23112.

Ishii, S.-i., Abe, Y., Matsushita, H., and Kato, I. (1990). An Asparaginyl Endopeptidase Purified from Jackbean Seeds. *J. Protein Chem.* 9, 294–295.

Jackson, M. A., Gilding, E. K., Shafee, T., Harris, K. S., Kaas, Q., Poon, S., et al. (2018). Molecular Basis for the Production of Cyclic Peptides by Plant Asparaginyl Endopeptidases. *Nat. Commun.* 9, 2411–2412. doi:10.1038/s41467-018-04669-9

Jackson, M. A., Nguyen, L. T., Gilding, E. K., Durek, T., and Craik, D. J. (2020). Make it or Break it: Plant AEPs on Stage in Biotechnology. *Biotechnol. Adv.* 45, 107651. doi:10.1016/j.biotechadv.2020.107651

Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., et al. (2001). Ordering Genes in a Flagella Pathway by Analysis of Expression Kinetics from Living Bacteria. *Science* 292, 2080–2083. doi:10.1126/science.1058758

Loo, S., Kam, A., Xiao, T., Nguyen, G. K., Liu, C. F., and Tam, J. P. (2016). Identification and Characterization of Roseltide, a Knottin-type Neutrophil

Elastase Inhibitor Derived from Hibiscus sabdariffa. *Sci. Rep.* 6, 39401–39416. doi:10.1038/srep39401

Magliery, T. J., Wilson, C. G. M., Pan, W., Mishler, D., Ghosh, I., Hamilton, A. D., et al. (2005). Detecting Protein−Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism. *J. Am. Chem. Soc.* 127, 146–157. doi:10.1021/ja046699g

Min, W., and Jones, D. H. (1994). *In Vitro* splicing of Concanavalin A Is Catalyzed by Asparaginyl Endopeptidase. *Nat. Struct. Mol. Biol.* 1, 502–504. doi:10.1038/nsb0894-502

Müntz, K., and Shutov, A. D. (2002). Legumains and Their Functions in Plants. *Trends Plant Science* 7, 340–344. doi:10.1016/s1360-1385(02)02298-7

Nguyen, G. K. T., Cao, Y., Wang, W., Liu, C. F., and Tam, J. P. (2015a). Site-Specific N-Terminal Labeling of Peptides and Proteins Using Butelase 1 and Thiodepsipeptide. *Angew. Chem.* 127, 15920–15924. doi:10.1002/ange.201506810

Nguyen, G. K. T., Hemu, X., Quek, J.-P., and Tam, J. P. (2016). Butelase-Mediated Macrocyclization Ofd-Amino-Acid-Containing Peptides. *Angew. Chem. Int. Ed.* 55, 12802–12806. doi:10.1002/anie.201607188

Nguyen, G. K. T., Kam, A., Loo, S., Jansson, A. E., Pan, L. X., and Tam, J. P. (2015b). Butelase 1: a Versatile Ligase for Peptide and Protein Macrocyclization. *J. Am. Chem. Soc.* 137, 15398–15401. doi:10.1021/jacs.5b11014

Nguyen, G. K. T., Wang, S., Qiu, Y., Hemu, X., Lian, Y., and Tam, J. P. (2014). Butelase 1 Is an Asx-specific Ligase Enabling Peptide Macrocyclization and Synthesis. *Nat. Chem. Biol.* 10, 732–738. doi:10.1038/nchembio.1586

Nonis, S. G., Haywood, J., Schmidberger, J. W., Mackie, E. R. R., Soares da Costa, T. P., Bond, C. S., et al. (2021). Structural and Biochemical Analyses of Concanavalin A Circular Permutation by jack Bean Asparaginyl Endopeptidase. *The Plant Cell* 33, 2794–2811. doi:10.1093/plcell/koab130

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and Characterization of a Superfolder green Fluorescent Protein. *Nat. Biotechnol.* 24, 79–88. doi:10.1038/nbt1172

Ross, J. S. (2009). Breast Cancer Biomarkers and HER2 Testing after 10 Years of Anti-HER2 Therapy. *Drug News Perspect.* 22, 93–106. doi:10.1358/dnp.2009.22.2.1334452

Saska, I., Gillon, A. D., Hatsugai, N., Dietzgen, R. G., Hara-Nishimura, I., Anderson, M. A., et al. (2007). An Asparaginyl Endopeptidase Mediates *In Vivo* Protein Backbone Cyclization. *J. Biol. Chem.* 282, 29721–29728. doi:10.1074/jbc.m705185200

Sheldon, P. S., Keen, J. N., and Bowles, D. J. (1996). Post-translational Peptide Bond Formation during Concanavalin A Processing *In Vitro*. *Biochem. J.* 320, 865–870. doi:10.1042/bj3200865

Shimada, T., Yamada, K., Kataoka, M., Nakaune, S., Koumoto, Y., Kuroyanagi, M., et al. (2003). Vacuolar Processing Enzymes Are Essential for Proper Processing of Seed Storage Proteins in Arabidopsis thaliana. *J. Biol. Chem.* 278, 32292–32299. doi:10.1074/jbc.m305740200

Stumpp, M. T., Binz, H. K., and Amstutz, P. (2008). DARPins: a New Generation of Protein Therapeutics. *Drug Discov. Today* 13, 695–701. doi:10.1016/j.drudis.2008.04.013

Tam, J. P., Chan, N.-Y., Liew, H. T., Tan, S. J., and Chen, Y. (2020). Peptide Asparaginyl Ligases—Renegade Peptide Bond Makers. *Sci. China Chem.* 63 (3), 1–12. doi:10.1007/s11426-019-9648-3

Waldo, G. S. (2003). Improving Protein Folding Efficiency by Directed Evolution Using the GFP Folding Reporter. *Directed Enzyme Evol.* 230, 343–359. doi:10.1385/1-59259-396-8:343

Wang, Z., Zhang, D., Hemu, X., Hu, S., To, J., Zhang, X., et al. (2021). Engineering Protein Theranostics Using Bio-Orthogonal Asparaginyl Peptide Ligases. *Theranostics* 11, 5863–5875. doi:10.7150/thno.53615

Yamada, K., Basak, A. K., Goto-Yamada, S., Tarnawska-Glatt, K., and Hara-Nishimura, I. (2020). Vacuolar Processing Enzymes in the Plant Life Cycle. *New Phytol.* 226, 21–31. doi:10.1111/nph.16306

Yang, R., Wong, Y. H., Nguyen, G. K. T., Tam, J. P., Lescar, J., and Wu, B. (2017). Engineering a Catalytically Efficient Recombinant Protein Ligase. *J. Am. Chem. Soc.* 139, 5351–5358. doi:10.1021/jacs.6b12637

Zauner, F. B., Elsässer, B., Dall, E., Cabrele, C., and Brandstetter, H. (2018). Structural Analyses of Arabidopsis thaliana Legumain γ Reveal Differential Recognition and Processing of Proteolysis and Ligation Substrates. *J. Biol. Chem.* 293, 8934–8946. doi:10.1074/jbc.m117.817031

Zhang, D., Wang, Z., Hu, S., Balamkundu, S., To, J., Zhang, X., et al. (2021). pH-Controlled Protein Orthogonal Ligation Using Asparaginyl Peptide Ligases. *J. Am. Chem. Soc.* 143(23), 8704–8712. doi:10.1021/jacs.1c02638

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fchem.2021.768535

A Genetically Encoded Picolyl Azide for Improved Live Cell Copper Click Labeling

Birthe Meineke[1,2], Johannes Heimgärtner[1,2], Alexander J. Craig[3], Michael Landreh[4], Lindon W. K. Moodie[3,5] and Simon J. Elsässer[1,2]**

[1]Science for Life Laboratory, Department of Medical Biochemistry and Biophysics, Division of Genome Biology, Karolinska Institutet, Stockholm, Sweden, [2]Ming Wai Lau Centre for Reparative Medicine, Stockholm Node, Karolinska Institutet, Stockholm, Sweden, [3]Drug Design and Discovery, Department of Medicinal Chemistry, Biomedical Centre, Uppsala University, Uppsala, Sweden, [4]Department of Microbiology, Tumor and Cell Biology, Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden, [5]Uppsala Antibiotic Centre, Uppsala University, Uppsala, Sweden

Bioorthogonal chemistry allows rapid and highly selective reactivity in biological environments. The copper-catalyzed azide–alkyne cycloaddition (CuAAC) is a classic bioorthogonal reaction routinely used to modify azides or alkynes that have been introduced into biomolecules. Amber suppression is an efficient method for incorporating such chemical handles into proteins on the ribosome, in which noncanonical amino acids (ncAAs) are site specifically introduced into the polypeptide in response to an amber (UAG) stop codon. A variety of ncAA structures containing azides or alkynes have been proven useful for performing CuAAC chemistry on proteins. To improve CuAAC efficiency, biologically incorporated alkyne groups can be reacted with azide substrates that contain copper-chelating groups. However, the direct incorporation of copper-chelating azides into proteins has not been explored. To remedy this, we prepared the ncAA paz-lysine (PazK), which contains a picolyl azide motif. We show that PazK is efficiently incorporated into proteins by amber suppression in mammalian cells. Furthermore, PazK-labeled proteins show improved reactivity with alkyne reagents in CuAAC.

Keywords: genetic code expansion, amber suppression, noncanonical amino acid, bioorthogonal chemistry, click chemistry, copper catalyzed azide–alkyne cycloaddition (CuAAC)

OPEN ACCESS

Edited by:
Tao Peng,
Peking University, China

Reviewed by:
Jeffery M Tharp,
Yale University, United States
Zhenrun J. Zhang,
University of Chicago, United States

****Correspondence:***
Lindon W. K. Moodie
lindon.moodie@ilk.uu.se
Simon J. Elsässer
simon.elsasser@scilifelab.se

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *31 August 2021*
Accepted: *13 October 2021*
Published: *11 November 2021*

Citation:
Meineke B, Heimgärtner J, Craig AJ, Landreh M, Moodie LWK and Elsässer SJ (2021) A Genetically Encoded Picolyl Azide for Improved Live Cell Copper Click Labeling. Front. Chem. 9:768535. doi: 10.3389/fchem.2021.768535

INTRODUCTION

Genetic code expansion allows the expression of proteins with distinct chemical handles through the residue- or site-specific introduction of noncanonical amino acids (ncAAs). First established in *Escherichia coli*, genetic code expansion has been adapted to all domains of life (Chin, 2014; Chin, 2017; Brown et al., 2018). When incorporated into proteins, ncAAs can confer a plethora of different functionalities: posttranslational modifications, crosslinking, spectroscopic probes, and also bioorthogonal chemical handles for selective reactions in the cellular context (Sletten and Bertozzi, 2009; Lang and Chin, 2014; Drienovská and Roelfes, 2020). Bioorthogonal chemistries enable endless possibilities for further derivatizing ncAA-containing proteins in or on live cells with fluorophores, lipids, or affinity handles (Lang et al., 2012; Elliott et al., 2014; Lang and Chin, 2014; Peng and Hang, 2016; Li et al., 2020; Meineke et al., 2020). The copper-catalyzed azide–alkyne cycloaddition (CuAAC, also referred to as "click" chemistry), is a Cu(I)-dependent, fast,

biorthogonal, and widely utilized reaction to form covalent bonds between alkyne and azide moieties (Rostovtsev et al., 2002; Tornøe et al., 2002; Hein and Fokin, 2010; Haldón et al., 2015; Li and Zhang, 2016). To circumvent the need for Cu(I) catalysis, strained alkynes have also been realized in so-called strain-promoted azide–alkyne cycloaddition (SPAAC) reactions (Beatty et al., 2010; Jewett et al., 2010). Alkyne and azide ncAAs, e.g., the methionine analogs azidoalanine, 6-azido-norleucine, and homopropargylglycine, can be used for CuAAC-mediated metabolic labeling. These ncAAs are substrates for the endogenous translation machinery, charged onto $tRNA^{Met}$ by methionyl-tRNA-synthetase (or an engineered mutant) and stochastically incorporated into nascent proteins in response to the AUG codon (Saleh et al., 2019). Site-directed incorporation of ncAAs into proteins, on the other hand, requires reprogramming of one codon and introduction of a dedicated, engineered pair of tRNA and aminoacyl-tRNA-synthetase (aaRS) that is orthogonal to, i.e., not interfering with, the translation machinery of the host. A widely used strategy to reprogram a codon is amber suppression, as the amber codon (UAG) is the least abundant of the three stop codons in *E. coli* and mammalian cells. Two different tRNA/aaRS systems have been used to site specifically install azide moieties in eukaryotic cells: AzFRS has been engineered from *E. coli* TyrRS to accept azido-phenylalanine (AzF) (Chin et al., 2002; Liu et al., 2007; Ye et al., 2009; Ye et al., 2010). AzFRS is combined with an amber suppressor mutant of *Bacillus stearothermophilus* TyrT (*Bst* $TyrT^{CUA}$) for AzF incorporation in the mammalian system, which has shown higher expression than the cognate *Eco*TyrT (Sakamoto et al., 2002; Liu et al., 2007). AzF is routinely used for UV-crosslinking studies [reviewed in (Coin, 2018)]; the azide is also reactive in CuAAC or other click reactions (Bundy and Swartz, 2010; Tian et al., 2014). An alternative tRNA/aaRS pair for amber suppression is the versatile pyrrolysine-tRNA (PylT) and pyrrolysine-tRNA-synthetase (PylRS) pair derived from methanogenic archaea, which is orthogonal across bacterial and eukaryotic hosts. *Methanosarcina mazei* PylT/RS (*Mma* PylT/RS)-mediated ncAA incorporation is efficient in mammalian cells, and a large number of active site mutants for incorporation of structurally diverse ncAAs have been described. The lysine-based ncAAs *N*-propargyl-L-lysine (ProK) and *N*-ε-([2-Azidoethoxy]carbonyl)-L-lysine (AzeoK) are efficiently incorporated with the *Mma* PylT/RS pair (Nguyen et al., 2009; Meineke et al., 2020). Thus, genetic incorporation of azides and alkynes has provided facile means to derivatize proteins using bioorthogonal CuAAC chemistry. However, the dependence on Cu(I) for catalysis has provided challenges in performing CuAAC in a cellular environment. Due to the sensitivity of Cu(I) ions toward oxidation in the presence of atmospheric oxygen, Cu(I) is typically generated *in situ* using stoichiometric amounts of sodium ascorbate as a reducing agent. Water-soluble Cu(I) ligands, such as tris(hydroxypropyltriazolylmethyl)amine (THPTA), have greatly improved biocompatibility of CuAAC by effectively complexing Cu(I), enhancing reaction speed at low Cu(I) concentrations, while inhibiting both the reoxidation of Cu(I) to Cu(II) and the production of reactive oxygen species (Hong et al., 2009, 2010). A complementary approach to increase biocompatibility of CuAAC is the use of "copper-chelating azides," such as picolyl azide (Uttamapinant et al., 2012; Kuang et al., 2010; Brotherton et al., 2009). Uttamapinant and others have demonstrated that positioning the azidomethyl group adjacent to the pyridine nitrogen significantly increases its reactivity in the presence of low Cu(I) concentrations, presumably by increasing the local concentration of the catalyst (Uttamapinant et al., 2012). Interestingly, copper-chelating azides improved reaction rates at low Cu(I) concentration synergistically with THPTA; hence, the combination of soluble ligands with picolyl azide allowed CuAAC to be performed on live cells at as low as 40 μM Cu(I) concentration, for which no toxicity was observed (Uttamapinant et al., 2012).

Despite the favorable properties of picolyl azide, genetic incorporation of copper-chelating azide moieties has not been reported in literature. Here, we synthesize a picolyl azide-lysine (PazK) ncAA that is readily incorporated using existing PylT/RS variants. We find that PazK has improved reactivity over simple azides in lysate and on live cells, especially at low Cu(I) concentrations, upgrading the repertoire of genetically encodable CuAAC reagents.

MATERIALS AND METHODS

Chemical synthesis of picolyl azide-lysine

Experimental procedures for the synthesis of PazK can be found in the supporting information.

Commercial Non-canonical Amino Acids

4-Azido-L-phenylalanine (AzF, CAS: 33173-53-4, Santa Cruz Biotechnology) and (S)-2-amino-6-[(2-azidoethoxy) carbonylamino]hexanoic acid (AzeoK, CAS: 1994331-17-7, Iris Biotech) were prepared as 100 mM stock solutions in 200 mM NaOH and 15% DMSO (w/v), and used at the final concentrations indicated.

DNA constructs

The constructs for expression of *Mma* PylT/RS wild type (RRID: Addgene_140009) and AF (RRID: Addgene_140023) variants as well as the sfGFP150TAG reporter constructs (RRID: Addgene_154766) were described previously (Meineke et al., 2018, 2020). We generated analogous constructs for AzFRS with four repeats of *Bst* $TyrT^{CUA}$ (RRID: Addgene_140018 and Addgene_174891). The plasmids share a common architecture and are here collectively referred to as "pAS" (Amber Suppression) plasmids: the aaRS, reporter or gene of interest are controlled by an EF1 promoter and followed by an IRES that allows expression of a downstream selection marker. A cassette with four tandem repeats of the tRNA gene, controlled by 7SK Pol III promoter, is placed upstream of the EF1 promoter in antisense orientation. All DNA constructs were verified by Sanger sequencing.

Cell culture and transfection

HEK293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM, GlutaMAX^{TM}, Thermo) supplemented with 10% (v/v) FBS at 37°C and 5% CO_2 atmosphere. For transient transfection, 1.5–2.0 × 10^5 cells/ml were seeded 24 h before transfection with TransIT-LT1 (Mirus) according to the instructions of the manufacturer. ncAAs were added at the time of transfection, and cells were harvested after 24 h.

Intact mass spectrometry

A modified transfection protocol was used for larger-scale GFP expression for bead purification, increasing the amount of total DNA to 6 μg (*Mma* PylT/RS AF and PylT/sfGFP150TAG at 2 + 8 ratio) per ml culture and transfecting 5.0–8.0 × 10^5 cells/ml with 2 μg of polyethylenimine (PEI) per μg of DNA. PazK was supplemented to 0.5 mM at transfection and until harvest after 6 days. Cells were lysed in RIPA buffer supplemented with 1× cOmplete protease inhibitor (Roche). The insoluble fraction was removed by centrifugation. Expressed GFP was captured on GFP-Trap_MA magnetic beads (ChromoTEK), washed with RIPA buffer and PBS, and eluted in 1% (v/v) acetic acid.

Purified GFP samples were desalted and rebuffered into 100 mM ammonium acetate, pH 7.5, using ZebaSpin columns with a 7-kDa cutoff (Thermo). Samples were directly infused into an Orbitrap Fusion Tribrid mass spectrometer equipped with an offline nanospray source using borosilicate capillaries (Thermo). The capillary voltage was 1.5 kV, and the pressure in the ion-routing multipole was maintained at 0.11 torr. Spectra were acquired in the Orbitrap mass analyzer operated in high mass mode at a resolution of 60.000 between 1,000 and 4,000 m/*z*. Data were analyzed using Excalibur (Thermo).

Live cell imaging for GFP expression

GFP-expressing HEK293T cells were imaged in a ZOE Fluorescent Cell Imager (BioRad).

Bioorthogonal labeling in lysate

HEK293T cells were transfected, cultured in the presence of 0.25 mM ncAA for 24 h and lysed in RIPA buffer with 1× cOmplete protease inhibitor (Roche). The insoluble fraction was removed by centrifugation. CuAAC was carried out on equal volume aliquots in 1 mM $CuSO_4$, 1 mM TCEP, 100 μM THPTA, and 1 μM AF647 dye (AF647-Alkyne or AF647-Picolyl Azide (Jena Bioscience)) for 1 h at 24°C, 450 rpm followed by incubation at 4°C overnight. Samples were separated on 4%–20% Tris-glycine gels (BioRad) and exposed for in-gel fluorescence at 630 nm in a GE AI600 imager and further analyzed by Western blot.

Bioorthogonal labeling of surface receptor proteins on live cells

Transfected HEK293T cells were grown in the presence of 0.25 mM PazK or 0.25 mM AzeoK for 24 h. Cells were washed with PBS and labeled with 5 μM AF647-alkyne dye (Jena Bioscience), 10–50 μM $CuSO_4$, 50–250 μM THPTA in 2.5 mM ascorbic acid (from a freshly prepared 100 mM stock) for 10 min at room temperature (Hong et al., 2010). Cells were collected in cold PBS, spun down, and lysed in PBS 0.1% (v/v) triton X-100 supplemented with 1× cOmplete protease inhibitor (Roche). Aliquots were separated on 4%–20% Tris-glycine gels (BioRad) and exposed for in-gel fluorescence at 460 and 630 nm in a GE AI600 imager and further analyzed by Western blot.

Labeling of surface receptor proteins on live cells for fluorescence microscopy

Transfected HEK293T were grown on poly-L-lysine-coated 18-well imaging slides (Ibidi) in the presence of 0.25 mM PazK or AzeoK for 24 h. Cells were washed with PBS and labeled with 5 μM alkyne dye (AFdye 647 alkyne, Jena Bioscience) in 50 μM $CuSO_4$, 250 μM THPTA, and 2.5 mM ascorbic acid for 10 min at room temperature. Subsequently, the cells were washed with PBS, counterstained with 2 μM Hoechst33342 (Life Technologies) in PBS for 30 min, washed again, and fixed in 4% formaldehyde for 10 min. The cells were washed and imaged in PBS on a Nikon Eclipse Ti2 inverted widefield microscope, using a ×20 (0.75 NA) objective and filter sets for DAPI and Cy5 fluorescence.

SDS-PAGE and Western blot

Aliquots of cell lysates were separated on 4–20% Tris-glycine gels (BioRad) and transferred to nitrocellulose membranes. Expression of GFP reporter and FLAG-aaRS was confirmed by immunoblotting with antibodies against GFP (Santa Cruz, RRID: AB_627695), HA-HRP (Roche, RRID:AB_390917), FLAG-HRP (Sigma, RRID:AB_439702), GAPDH (Millipore, RRID: AB_10615768), and corresponding secondary HRP-conjugated antibodies when needed (BioRad, RRID:AB_11125936 and Invitrogen, RRID:AB_2534727). Quantitative analysis of gel lanes was performed using ImageJ software.

RESULTS

Synthesis of picolyl azide-lysine

To synthesize picolyl azide-lysine (PazK), two building blocks were required, lysine derivative **3** and azide **7** (**Figure 1**). The synthesis of **3** commenced with orthogonally protected *N*α-Boc-*N*ε-Cbz-Lysine **1**. After methylation under standard conditions, hydrogenation afforded **3** (Schnell et al., 2020), which contains the free side chain amine, in 83% yield over two steps. To access azide **7**, dipicolinic acid dimethyl ester **4** was selectively reduced with $NaBH_4$ to alcohol **5** in 65% yield. Installation of the requisite azide functionality was effected using a one-pot process where the hydroxyl group of **5** was converted into the corresponding alkyl bromide (PPh_3 and CBr_4), followed by displacement with sodium azide without isolation of the bromide intermediate. Subsequent ester hydrolysis under basic conditions afforded carboxylic acid **7** (Hanna et al., 2017). With building blocks **3** and **7** in hand, amide formation was performed using standard conditions (EDCI, HOBt, DIPEA in DMF). Lithium hydroxide-mediated ester hydrolysis, followed by Boc deprotection under acidic conditions afforded the desired PazK, as the hydrochloride salt.

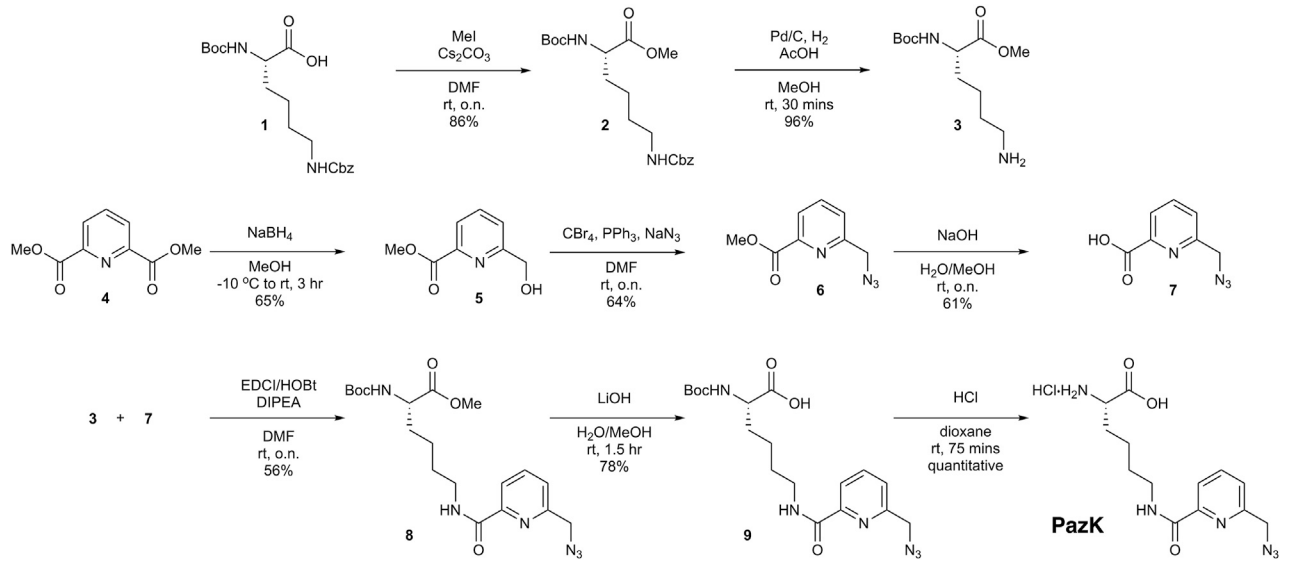


FIGURE 1 | Chemical synthesis of picolyl azide-lysine (PazK).

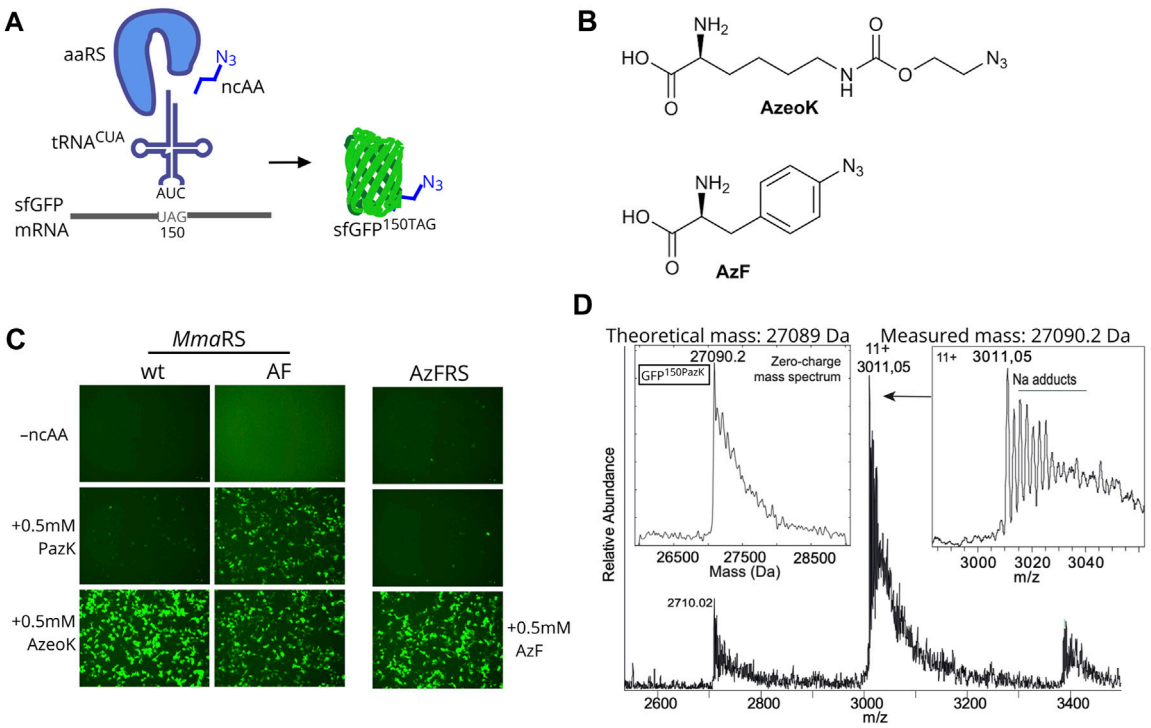


FIGURE 2 | Amber suppression mediated PazK incorporation into protein in mammalian cells. **(A)** Schematic depiction of noncanonical amino acid (ncAA) incorporation into GFP. Expression of an orthogonal tRNA and aaRS pair allows incorporation of ncAAs with azide side chains into the GFP reporter in response to an amber codon (UAG in the mRNA). **(B)** Chemical structures of (S)-2-amino-6-[(2-azidoethoxy)carbonylamino]hexanoic acid (AzeoK) and 4-azido-L-phenylalanine (AzF). **(C)** Live-cell imaging of HEK293T cells transfected with *Methanosarcina mazei* pyrrolysine-tRNA/RS (*Mma* PylT/RS) wt, *Mma* PylT/RS AF, or *Bacillus stearothermophilus* (*Bst*) TyrT/AzFRS and cognate tRNA/GFP150TAG reporter plasmid (1 + 4 ratio) in the absence (–ncAA) or presence of 0.5 mM of the indicated ncAA. Images were taken 24 h posttransfection. **(D)** Intact mass determination of purified GFP containing 150PazK (incorporated with *Mma* PylRS AF in GFP150TAG).

Methanosarcina mazei pyrrolysine-tRNA-synthetase AF active site mutant allows incorporation of picolyl azide-lysine into proteins

Next, we needed to establish that PazK can be accepted as a substrate for tRNA aminoacylation by a $tRNA^{CUA}$/aaRS pair orthogonal in mammalian cells. We used a GFP reporter with an amber codon at position 150, allowing the use of fluorescence as a readout for incorporation efficiency. If the ncAA added to the medium is accepted by the aaRS to aminoacylate the cognate $tRNA^{CUA}$, the amber stop codon is suppressed, and full-length fluorescent GFP bearing PazK at position 150 ($GFP^{150PazK}$) is produced (**Figure 2A**).

We tested the incorporation of PazK by *Mma* PylRS and its variant with mutations Y306A and Y384F: *Mma* PylRS AF (Yanagisawa et al., 2008). Wild-type *Mma* PylRS can accommodate a variety of ncAA substrates in its active site, but the Pyl binding pocket cannot accommodate large or bulky lysine adducts. The *Mma* PylRS AF mutant has been rationally designed to enlarge the ncAA binding pocket (Yanagisawa et al., 2008; Yanagisawa et al., 2019) and has enabled incorporation of lysine derivatives with aromatic and larger hydrocarbon rings (Borrmann et al., 2012; Nikić et al., 2014; Ge et al., 2016).

The *Mma* PylT/RS pairs were cotransfected with a PylT/sfGFP150TAG amber suppression reporter in HEK293T cells. We assayed AzeoK (**Figure 2B**) and PazK against a control with no ncAA, which showed no GFP fluorescence; AzeoK is an excellent substrate for wild-type *Mma* PylRS (Meineke et al., 2020) and, as expected, produced strong GFP fluorescence. PazK only yielded low GFP fluorescence with the same wild-type PylT/RS-transfected cells (**Figure 2C**). By adding AzeoK and PazK to *Mma* PylT/RS AF-expressing cells, we observed similar GFP fluorescence levels for both ncAAs (**Figure 2C**, right). For comparison, we also tested the incorporation of AzF (**Figure 2B**) and 6-azido-lysine (6AzK) (**Supplementary Figure S1**). AzF was efficiently incorporated in *Bst* $TyrT^{CUA}$/AzFRS-expressing cells as judged by GFP fluorescence (**Figure 2C**), while 6AzK was not a substrate for *Mma* PylT/RS (**Supplementary Figure S1**). We further tested *Methanogenic archaeon ISO4-G1* (*G1*) PylT/RS and *G1* $PylT/RS^{Y125A}$ pairs (Meineke et al., 2020) and found that wild-type *G1* PylRS accepted PazK with low efficiency, but *G1* $PylRS^{Y125A}$ showed high incorporation efficiency for PazK (**Supplementary Figure S2**). Hence, we conclude that azide-bearing ncAAs can be incorporated well in mammalian cells with existing tRNA/aaRS pairs.

We further sought to confirm the selective incorporation and chemical stability of PazK in a target protein. Hence, we purified $sfGFP^{150PazK}$ from HEK293T cells transfected with PylT/sfGFP150TAG and *Mma* PylT/RS AF and performed intact mass spectrometry. The calculated mass of 27,089 Da and determined mass of 27,090.2 Da were in agreement, confirming PazK incorporation and the stability of the picolyl-azide moiety in the cellular environment (**Figure 2D**).

Copper-catalyzed azide–alkyne cycloaddition reactivity of GFP containing different azide-bearing non-canonical amino acids

To compare CuAAC labeling of the three azide-containing ncAAs, AzeoK, PazK, and AzF, we reacted $GFP^{150ncAA}$ with fluorescent AF647-alkyne in HEK293T cell lysates after transient transfection of amber-suppressor tRNA/aaRS, using *Mma* PylT/RS AF for AzeoK and PazK and *Bst* TyrT/AzFRS for AzF (**Figure 3A**). In agreement with fluorescent imaging, anti-GFP Western blot confirmed the efficient incorporation of all three ncAAs, in the order AzF > AzeoK > PazK under the conditions used. For assessing the specificity of CuAAC reaction for the three ncAAs, we reacted a fluorescent dye, AF647-alkyne, via CuAAC in whole-cell lysate. Here, we chose traditional *in vitro* conditions with excess alkyne dye, high concentration of copper salt (1 mM), 100 μM THPTA, and long reaction time (1 h at RT followed by overnight incubation at 4°C) to reach a reaction end point. CuAAC AF647-alkyne yielded a single band corresponding to the size of GFP visible with in-gel fluorescence imaging at 630 nm (**Figure 3A**). No other bands are observed, confirming that all ncAA are orthogonal to (i.e., not incorporated by) the endogenous complement of aaRS enzymes. In principle, stoichiometric labeling should be observed under the given reaction conditions for the three azide-modified GFP proteins. However, despite the lower amount of total GFP produced, the signal for AF647-labeled GFP was strongest for PazK and weakest for AzeoK (corresponding to a roughly 7.5-fold higher AF647/GFP ratio for PazK compared to AzeoK) (**Figure 3A**). These results confirm that AzeoK and AzF are more efficiently incorporated, but suggest that the incorporated PazK has a higher CuAAC reactivity. There are several potential explanations for this observation: terminal azides can undergo reduction to amines, and aromatic azides are known to be photolabile; hence, some of the AzeoK and AzF azide moieties may have been eliminated in cellulo or upon lysis (Milles et al., 2012). On the other hand, natural Cu(I) chelating molecules in the crude lysate and reoxidation of Cu(I) to Cu(II) with atmospheric oxygen may deplete Cu(I) available for CuAAC under elongated reaction conditions. As an additional control, we performed an SPAAC reaction with dibenzocyclooctyne (DBCO)-TAMRA fluorescent dye in lysates of all the three azide-bearing GFP species and again observed an improved reactivity of PazK over AzeoK and AzF (**Supplementary Figure S3**). This further hinted at the decomposition of the AzeoK and AzF azide moieties in cellulo or upon cell lysis. In summary, these results, together with the intact mass (**Figure 2D**) suggest that PazK is favorably stable and reactive compared with other available azide ncAAs. Of note, SPAAC labeling with DBCO showed less specific labeling of the azide-bearing GFP and a number of background bands, in line with prior reports that SPAAC

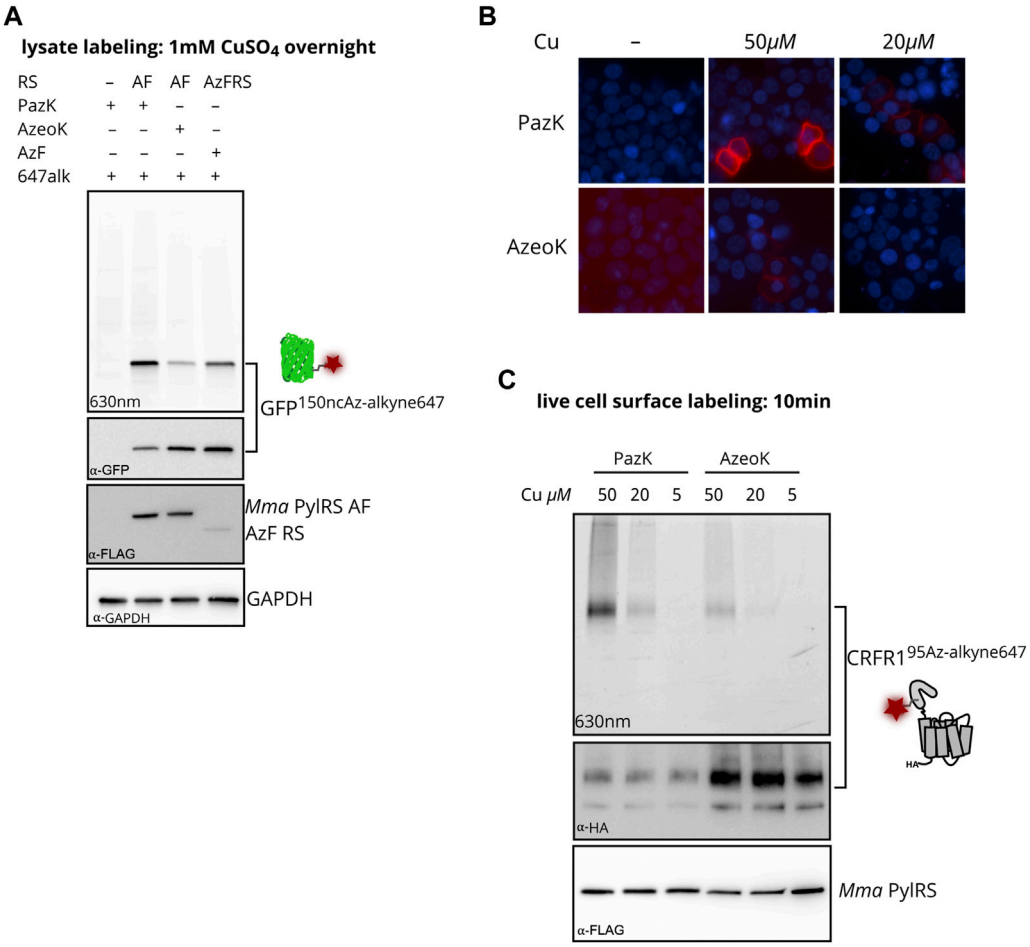


FIGURE 3 | Incorporation of PazK allows copper-catalyzed azide–alkyne cycloaddition (CuAAC) at reduced copper concentrations. **(A)** CuAAC labeling of azide ncAAs in GFP in HEK293T cell lysate. Cells were transfected with *Mma* PylT/RS AF, or *Bst* TyrT/AzFRS and cognate tRNA/GFP150TAG reporter plasmid (1 + 4 ratio) and cultured in the absence (–ncAA) or presence of 0.25 mM of the indicated ncAA for 24 h. CuAAC labeling with 1 mM $CuSO_4$, 1 mM TCEP, 100 μM THPTA, and 1 μM AF647-alkyne in cell lysate. Lysate aliquots were separated by SDS-PAGE and imaged for in-gel fluorescence. Immunostaining for GFP, FLAG-tagged aminoacyl-tRNA synthetase, and GAPDH loading control after membrane transfer of the same gel. **(B, C)** CuAAC labeling of $CRFR1^{95PazK}$ and $CRFR1^{95AzeoK}$ on the surface of live HEK293T cells. Cells were transfected with *Mma* PylT/RS AF and *Mma* PylT/CRFR1 95TAG reporter plasmid (1 + 4 ratio) and cultured in the absence (–ncAA) or presence of 0.25 mM of the indicated ncAA for 24 h. CuAAC labeling with 5–50 μM $CuSO_4$, 25–250 μM THPTA, and 5 μM AF647-alkyne on live cells. **(B)** CuAAC-labeled cells were counterstained with Hoechst 33342, fixed in 4% formaldehyde for fluorescence microscopy. **(C)** Lysate aliquots were separated by SDS-PAGE and imaged for in-gel fluorescence. Immunostaining for HA-tagged CRFR1 and FLAG-tagged aminoacyl-tRNA synthetase after membrane transfer of the same gel.

reactions are not strictly bioorthogonal due to side reactions with thiols (van Geel et al., 2012).

Picolyl azide-lysine labeling with low copper concentrations

We moved on to investigate CuAAC labeling on live cells, where the concentration of added copper and labeling conditions must be optimized to find a compromise between reaction efficiency and adverse side effects to proteins and cells. $CuSO_4$ concentrations of 50 μM in the presence of excess copper chelators have been successfully used for live cell CuAAC labeling, while higher concentrations have been shown to impact cell viability (Hong et al., 2010; Uttamapinant et al., 2012; Meineke et al., 2020). We incorporated PazK and AzeoK with *Mma* PylT/RS AF into an amber mutant of the class B GPCR corticotropin-releasing factor type 1 receptor (CRFR1 95TAG) (Coin et al., 2013; Serfling et al., 2018, Serfling et al., 2019). CuAAC with AF647-alkyne was performed with 5–50 μM $CuSO_4$ and a fivefold excess of THPTA on the surface of live cells expressing $CRFR1^{95AzeoK}$ or $CRFR1^{95PazK}$ (**Figures 3B, C**). $CRFR1^{95AzeoK}$ could be labeled with AF647-alkyne on the surface of live cells with 50 μM $CuSO_4$, while AF647 fluorescence was barely detectable at 20 μM $CuSO_4$ and undetectable at 5 μM $CuSO_4$. $CRFR1^{95PazK}$ yielded much stronger specific AF647 fluorescence at 50 and 20 μM $CuSO_4$ despite the lower expression level. Incorporation efficiency of PazK and AzeoK into CRFR1 can be compared via detection of a

C-terminal HA tag: at the same ncAA concentration, AzeoK addition allows much more efficient amber suppression (**Figure 3C**). Thus, PazK demonstrated greatly improved CuAAC reactivity over AzeoK at copper concentrations as low as 20 μM, while further reducing the copper concentration did not support CuAAC with either ncAA. Thus, we conclude that PazK, in combination with THPTA, allows efficient CuAAC reactions on live cells with minimal expected toxicity (Hong et al., 2010).

DISCUSSION

The discovery of strain-promoted inverse electron-demand Diels–Alder cycloaddition (SPIEDAC) has enabled versatile bioorthogonal reactions that are fast, efficient, and nontoxic in and on live cells (Lang and Chin, 2014; Nikić et al., 2014). As a result, CuAAC has become obsolete for many fluorescent labeling and chemical conjugation applications in cellular environments. However, CuAAC is exquisitely bioorthogonal as well as orthogonal to SPIEDAC and, thus, remains a universal choice for performing two orthogonal chemical conjugations in the same cellular environment (Nikić and Lemke, 2015). We have previously demonstrated orthogonal dual-color labeling of surface receptors on live cells combining SPIEDAC and CuAAC on genetic encoded *trans*-cyclooct-2-en-lysine (TCO*K) and ProK. Because PazK is a substrate for *Mma* PylRS AF and *G1* PylRS Y125A, it cannot be combined with TCO*K to form a second orthogonal ncAA pair for dual labeling. However, we note that ProK and PazK could be incorporated with the orthogonal *Mma* PylT/PylRS and *G1* hybT*/PylRS Y125A pairs (Meineke et al., 2020), hence, providing a route for installing site-specific alkynes and azides that could be employed for orthogonal fluorescent labeling as well as site-specific intramolecular or intermolecular crosslinking.

Currently, CuAAC reactions are limited to the cell surface because low intracellular Cu(I) concentration does not permit catalysis, and artificially raising copper concentrations within cells is likely toxic (Bevilacqua et al., 2014; Li et al., 2017). We determine a lower limit of 20 μM of copper for a successful CuAAC reaction with PazK on live cells. Synthesizing and screening additional structural variants of PazK may, in the future, improve incorporation efficiency and reactivity. For catalysis at even lower free copper concentrations, the copper-chelating properties of the azide ncAA could be enhanced by multivalent chelating ligands. For example, coordinating azides with two or three triazole rings have been shown, in principle, to enable intracellular CuAAC (Bevilacqua et al., 2014; Li et al., 2017). It will, thus, be an interesting challenge if PylT/RS variants can be identified that can accept larger copper-chelating azides and if availability of Cu(I) in the intracellular environment would be sufficient for catalysis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

BM, AC, LM, and SE conceived and planned this study. AC performed chemical syntheses. BM and JH performed biological experiments and analyzed the data. ML performed the intact mass spectroscopy and analyzed the data. BM wrote the first draft of the manuscript. BM, AC, LM, and SE wrote sections of the manuscript. All authors contributed to the manuscript revision, and read and approved the submitted version.

FUNDING

SE acknowledges funding from the Karolinska Institutet SFO Molecular Biosciences, Sweden; Vetenskapsrådet, Sweden (2015–04815, 2020-04313); Ming Wai Lau Center for Reparative Medicine, Sweden; Ragnar Söderbergs Stiftelse, Sweden; and the Knut och Alice Wallenbergs Stiftelse, Sweden (2017–0276). LM acknowledges the Uppsala Antibiotic Centre for funding. AC is supported by funding from the Carl Tryggers Stiftelse. ML is supported by a Karolinska Institutet Faculty-funded Career position, Vetenskapsrådet, Sweden (2019-01961 and 2019-02463), Cancerfonden, Sweden (19 0480 Pj).

ACKNOWLEDGMENTS

We thank all members of the Elsässer lab for input into the concept, experimental design, and manuscript. We would like to thank Dr. Jeffrey Hawkes, Department of Chemistry, Uppsala University, for the mass spectrometry experiments. We thank the other groups in the Division of Genome Biology for their support, specifically the J. Bartek lab for access to the Tecan Infinite M200 Pro Plate Reader, and the O. Fernandez-Capetillo lab for access to the GE AI600 Gel Imager.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.768535/full#supplementary-material

REFERENCES

Beatty, K. E., Fisk, J. D., Smart, B. P., Lu, Y. Y., Szychowski, J., Hangauer, M. J., et al. (2010). Live-cell Imaging of Cellular Proteins by a Strain-Promoted Azide-Alkyne Cycloaddition. *Chem. Eur. J. Chem. Bio.* 11, 2092–2095. doi:10.1002/cbic.201000419

Bevilacqua, V., King, M., Chaumontet, M., Nothisen, M., Gabillet, S., Buisson, D., et al. (2014). Copper-chelating Azides for Efficient Click Conjugation Reactions in Complex media. *Angew. Chem. Int. Ed.* 53, 5872–5876. doi:10.1002/anie.201310671

Borrmann, A., Milles, S., Plass, T., Dommerholt, J., Verkade, J. M. M., Wiessler, M., et al. (2012). Genetic Encoding of a bicyclo[6.1.0]nonyne-Charged Amino Acid Enables Fast Cellular Protein Imaging by Metal-free Ligation. *Chembiochem* 13, 2094–2099. doi:10.1002/cbic.201200407

Brotherton, W. S., Michaels, H. A., Simmons, J. T., Clark, R. J., Dalal, N., and Zhu, L. (2009). Apparent copper(II)-accelerated azide-alkyne cycloaddition. *Org. Lett.* 11, 4954–4957. doi:10.1021/ol9021113

Brown, W., Liu, J., and Deiters, A. (2018). Genetic Code Expansion in Animals. *ACS Chem. Biol.* 13, 2375–2386. doi:10.1021/acschembio.8b00520

Bundy, B. C., and Swartz, J. R. (2010). Site-Specific Incorporation of P-Propargyloxyphenylalanine in a Cell-free Environment for Direct Protein–Protein Click Conjugation. *Bioconjug. Chem.* 21, 255–263. doi:10.1021/bc9002844

Chin, J. W. (2017). Expanding and Reprogramming the Genetic Code. *Nature* 550, 53–60. doi:10.1038/nature24031

Chin, J. W. (2014). Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* 83, 379–408. doi:10.1146/annurev-biochem-060713-035737

Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002). Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia coli. *Proc. Natl. Acad. Sci.* 99, 11020–11024. doi:10.1073/pnas.172226299

Coin, I. (2018). Application of Non-canonical Crosslinking Amino Acids to Study Protein-Protein Interactions in Live Cells. *Curr. Opin. Chem. Biol.* 46, 156–163. doi:10.1016/j.cbpa.2018.07.019

Coin, I., Katritch, V., Sun, T., Xiang, Z., Siu, F. Y., Beyermann, M., et al. (2013). Genetically Encoded Chemical Probes in Cells Reveal the Binding Path of Urocortin-I to CRF Class B GPCR. *Cell* 155, 1258–1269. doi:10.1016/j.cell.2013.11.008

Drienovská, I., and Roelfes, G. (2020). Expanding the Enzyme Universe with Genetically Encoded Unnatural Amino Acids. *Nat. Catal.* 3, 193–202. doi:10.1038/s41929-019-0410-8

Elliott, T. S., Townsley, F. M., Bianco, A., Ernst, R. J., Sachdeva, A., Elsässer, S. J., et al. (2014). Proteome Labeling and Protein Identification in Specific Tissues and at Specific Developmental Stages in an Animal. *Nat. Biotechnol.* 32, 465–472. doi:10.1038/nbt.2860

Ge, Y., Fan, X., and Chen, P. R. (2016). A Genetically Encoded Multifunctional Unnatural Amino Acid for Versatile Protein Manipulations in Living Cells. *Chem. Sci.* 7, 7055–7060. doi:10.1039/c6sc02615j

Haldón, E., Nicasio, M. C., and Pérez, P. J. (2015). Copper-catalysed Azide-Alkyne Cycloadditions (CuAAC): an Update. *Org. Biomol. Chem.* 13, 9528–9550. doi:10.1039/c5ob01457c

Hanna, J., Allan, C., Lawrence, C., Meyer, O., Wilson, N., and Hulme, A. (2017). Optimizing the Readout of Lanthanide-DOTA Complexes for the Detection of Ligand-Bound Copper(I). *Molecules* 22, 802. doi:10.3390/molecules22050802

Hein, J. E., and Fokin, V. V. (2010). Copper-catalyzed Azide-Alkyne Cycloaddition (CuAAC) and beyond: New Reactivity of Copper(I) Acetylides. *Chem. Soc. Rev.* 39, 1302–1315. doi:10.1039/b904091a

Hong, V., Presolski, S. I., Ma, C., and Finn, M. G. (2009). Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chem.* 121, 10063–10067. doi:10.1002/ange.200905087

Hong, V., Steinmetz, N. F., Manchester, M., and Finn, M. G. (2010). Labeling Live Cells by Copper-Catalyzed Alkyne–Azide Click Chemistry. *Bioconjug. Chem.* 21, 1912–1916. doi:10.1021/bc100272z

Jewett, J. C., Sletten, E. M., and Bertozzi, C. R. (2010). Rapid Cu-free Click Chemistry with Readily Synthesized Biarylazacyclooctynones. *J. Am. Chem. Soc.* 132, 3688–3690. doi:10.1021/ja100014q

Kuang, G.-C., Michaels, H. A., Simmons, J. T., Clark, R. J., and Zhu, L. (2010). Chelation-assisted, copper(II)-acetate-accelerated azide-alkyne cycloaddition. *J. Org. Chem.* 75, 6540–6548. doi:10.1021/jo101305m

Lang, K., and Chin, J. W. (2014). Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* 114, 4764–4806. doi:10.1021/cr400355w

Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D. J., Blackman, M. L., et al. (2012). Genetic Encoding of Bicyclononynes and Trans-cyclooctenes for Site-specific Protein Labeling *In Vitro* and in Live Mammalian Cells via Rapid Fluorogenic Diels-Alder Reactions. *J. Am. Chem. Soc.* 134, 10317–10320. doi:10.1021/ja302832g

Li, L., and Zhang, Z. (2016). Development and Applications of the Copper-Catalyzed Azide-Alkyne Cycloaddition (CuAAC) as a Bioorthogonal Reaction. *Molecules* 21, 1393. doi:10.3390/molecules21101393

Li, S., Wang, L., Yu, F., Zhu, Z., Shobaki, D., Chen, H., et al. (2017). Copper-Catalyzed Click Reaction On/in Live Cells. *Chem. Sci.* 8, 2107–2114. doi:10.1039/C6SC02297A

Li, Y., Wang, S., Chen, Y., Li, M., Dong, X., Hang, H. C., et al. (2020). Site-specific Chemical Fatty-Acylation for Gain-Of-Function Analysis of Protein S-Palmitoylation in Live Cells. *Chem. Commun.* 56, 13880–13883. doi:10.1039/d0cc06073a

Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P. G. (2007). Genetic Incorporation of Unnatural Amino Acids into Proteins in Mammalian Cells. *Nat. Methods.* 4, 239–244. doi:10.1038/nmeth1016

Meineke, B., Heimgärtner, J., Eirich, J., Landreh, M., and Elsässer, S. J. (2020). Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells. *Cel Rep.* 31, 107811. doi:10.1016/j.celrep.2020.107811

Meineke, B., Heimgärtner, J., Lafranchi, L., and Elsässer, S. J. (2018). Methanomethylophilus Alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS Chem. Biol.* 13, 3087–3096. doi:10.1021/acschembio.8b00571

Milles, S., Tyagi, S., Banterle, N., Koehler, C., VanDelinder, V., Plass, T., et al. (2012). Click Strategies for Single-Molecule Protein Fluorescence. *J. Am. Chem. Soc.* 134, 5187–5195. doi:10.1021/ja210587q

Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., and Chin, J. W. (2009). Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-tRNA Synthetase/tRNACUA Pair and Click Chemistry. *J. Am. Chem. Soc.* 131, 8720–8721. doi:10.1021/ja900553w

Nikić, I., and Lemke, E. A. (2015). Genetic Code Expansion Enabled Site-specific Dual-Color Protein Labeling: Superresolution Microscopy and beyond. *Curr. Opin. Chem. Biol.* 28, 164–173. doi:10.1016/j.cbpa.2015.07.021

Nikić, I., Plass, T., Schraidt, O., Szymański, J., Briggs, J. A. G., Schultz, C., et al. (2014). Minimal Tags for Rapid Dual-Color Live-Cell Labeling and Super-resolution Microscopy. *Angew. Chem. Int. Ed.* 53, 2245–2249. doi:10.1002/anie.201309847

Peng, T., and Hang, H. C. (2016). Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* 138, 14423–14433. doi:10.1021/jacs.6b08733

Rostovtsev, V. V., Green, L. G., Fokin, V. V., and Sharpless, K. B. (2002). A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* 41, 2596–2599. doi:10.1002/1521-3773(20020715)41:14<2596:AID-ANIE2596>3.0

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., et al. (2002). Site-specific Incorporation of an Unnatural Amino Acid into Proteins in Mammalian Cells. *Nucleic Acids Res.* 30, 4692–4699. doi:10.1093/nar/gkf589

Saleh, A. M., Wilding, K. M., Calve, S., Bundy, B. C., and Kinzer-Ursem, T. L. (2019). Non-canonical Amino Acid Labeling in Proteomics and Biotechnology. *J. Biol. Eng.* 13, 43. doi:10.1186/s13036-019-0166-3

Schnell, S. D., Hoff, L. V., Panchagnula, A., Wurzenberger, M. H. H., Klapötke, T. M., Sieber, S., et al. (2020). 3-Bromotetrazine: Labelling of Macromolecules via Monosubstituted Bifunctional S-Tetrazines. *Chem. Sci.* 11, 3042–3047. doi:10.1039/c9sc06169j

Serfling, R., Lorenz, C., Etzel, M., Schicht, G., Böttke, T., Mörl, M., et al. (2018). Designer tRNAs for Efficient Incorporation of Non-canonical Amino Acids by the Pyrrolysine System in Mammalian Cells. *Nucleic Acids Res.* 46, 1–10. doi:10.1093/nar/gkx1156

Serfling, R., Seidel, L., Bock, A., Lohse, M. J., Annibale, P., and Coin, I. (2019). Quantitative Single-Residue Bioorthogonal Labeling of G Protein-Coupled Receptors in Live Cells. *ACS Chem. Biol.* 14, 1141–1149. doi:10.1021/acschembio.8b01115

Sletten, E. M., and Bertozzi, C. R. (2009). Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* 48, 6974–6998. doi:10.1002/anie.200900942

Tian, H., Naganathan, S., Kazmi, M. A., Schwartz, T. W., Sakmar, T. P., and Huber, T. (2014). Bioorthogonal Fluorescent Labeling of Functional G-Protein-Coupled Receptors. *Chembiochem* 15, 1820–1829. doi:10.1002/cbic.201402193

Tornøe, C. W., Christensen, C., and Meldal, M. (2002). Peptidotriazoles on Solid Phase: [1,2,3]-triazoles by Regiospecific Copper(i)-Catalyzed 1,3-dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* 67, 3057–3064. doi:10.1021/jo011148j

Uttamapinant, C., Tangpeerachaikul, A., Grecian, S., Clarke, S., Singh, U., Slade, P., et al. (2012). Fast, Cell-Compatible Click Chemistry with Copper-Chelating Azides for Biomolecular Labeling. *Angew. Chem. Int. Ed.* 51, 5852–5856. doi:10.1002/anie.201108181

van Geel, R., Pruijn, G. J. M., van Delft, F. L., and Boelens, W. C. (2012). Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide-Alkyne Cycloaddition. *Bioconjug. Chem.* 23, 392–398. doi:10.1021/bc200365k

Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008). Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) Lysine for Site-specific Protein Modification. *Chem. Biol.* 15, 1187–1197. doi:10.1016/j.chembiol.2008.10.004

Yanagisawa, T., Kuratani, M., Seki, E., Hino, N., Sakamoto, K., and Yokoyama, S. (2019). Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-tRNA Synthetase. *Cel Chem. Biol.* 26, 936–949. doi:10.1016/j.chembiol.2019.03.008

Ye, S., Huber, T., Vogel, R., and Sakmar, T. P. (2009). FTIR Analysis of GPCR Activation Using Azido Probes. *Nat. Chem. Biol.* 5, 397–399. doi:10.1038/nchembio.167

Ye, S., Zaitseva, E., Caltabiano, G., Schertler, G. F. X., Sakmar, T. P., Deupi, X., et al. (2010). Tracking G-Protein-Coupled Receptor Activation Using Genetically Encoded Infrared Probes. *Nature* 464, 1386–1389. doi:10.1038/nature08948

Conflict of Interest: The authors confirm that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fchem.2021.779976

Ferritin Conjugates With Multiple Clickable Amino Acids Encoded by C-Terminal Engineered Pyrrolysyl-tRNA Synthetase

Yi-Hui Wang[1†], *Mu-Lung Jian*[1,2†], *Pei-Jung Chen*[1,2], *Jo-Chu Tsou*[1,2], *Le P. Truong*[1] and *Yane-Shih Wang*[1,2*]

[1]Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan, [2]Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei, Taiwan

This study reports the application of expanding genetic codes in developing protein cage-based delivery systems. The evolved Methanosarcina mazei pyrrolysyl-tRNA synthetase (PylRS)•tRNA^{Pyl} pairs derived from directed evolution are examined to probe their recognition for para-substituted phenylalanine analogs. The evolved MmPylRS, AzFRS, harboring a wide range of substrates, is further engineered at the C-terminal region into another variant, AzFRS-MS. AzFRS-MS shows suppression of the elevated sfGFP protein amount up to 10 TAG stop codons when charging p-azido-L-phenylalanine (AzF, **4**), which allows the occurrence of click chemistry. Since protein nanocages used as drug delivery systems that encompass multiple drugs through a site-specific loading approach remain largely unexplored, as a proof of concept, the application of AzFRS-MS for the site-specific incorporation of AzF on human heavy chain ferritin (Ftn) is developed. The Ftn-**4** conjugate is shown to be able to load multiple fluorescence dyes or a therapeutic agent, doxorubicin (Dox), through the strain-promoted azide-alkyne cycloaddition (SPAAC) click reaction. Aiming to selectively target Her2^{+} breast cancer cells, Ftn-**4**-DOX conjugates fused with a HER2 receptor recognition peptide, anti-Her2/neu peptide (AHNP), is developed and demonstrated to be able to deliver Dox into the cell and to prolong the drug release. This work presents another application of evolved MmPylRS systems, whose potential in developing a variety of protein conjugates is noteworthy.

Keywords: pyrrolysyl-tRNA synthetase, ferritin-drug conjugates, strain-promoted azide-alkyne cycloaddition, p-Azido-L-phenylalanine, amber suppression efficiency, drug delivery

OPEN ACCESS

Edited by:
Yu-Hsuan Tsai,
Shenzhen Bay Laboratory, China

Reviewed by:
Chenguang Fan,
University of Arkansas, United States
Weimin Xuan,
Nankai University, China

***Correspondence:**
Yane-Shih Wang
yaneshihwang@gate.sinica.edu.tw

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *20 September 2021*
Accepted: *20 October 2021*
Published: *25 November 2021*

Citation:
Wang Y-H, Jian M-L, Chen P-J, Tsou J-C, Truong LP and Wang Y-S (2021) Ferritin Conjugates With Multiple Clickable Amino Acids Encoded by C-Terminal Engineered Pyrrolysyl-tRNA Synthetase. Front. Chem. 9:779976. doi: 10.3389/fchem.2021.779976

INTRODUCTION

Protein cages consist of a self-assembled protein hollow shell, such as viral capsids, virus-like particles, chaperonins, and human ferritin heavy chain (Ftn) (Aumiller et al., 2018). These biological supermolecules formed by multiple copies of monomer possess symmetric three-dimensional structures with remarkable stability. Protein cages are widely used as delivery platforms in bio-nanotechnology and material science. In the field of drug delivery, protein cages are employed to prolong the half-life of drugs, exert enhanced permeability and retention effect, reduce nonspecific uptake by introducing targeting ligand, and enhance drug solubility or endocytosis effect (Zhang et al., 2021). For instance, Ftn is applied to the delivery of enzymes

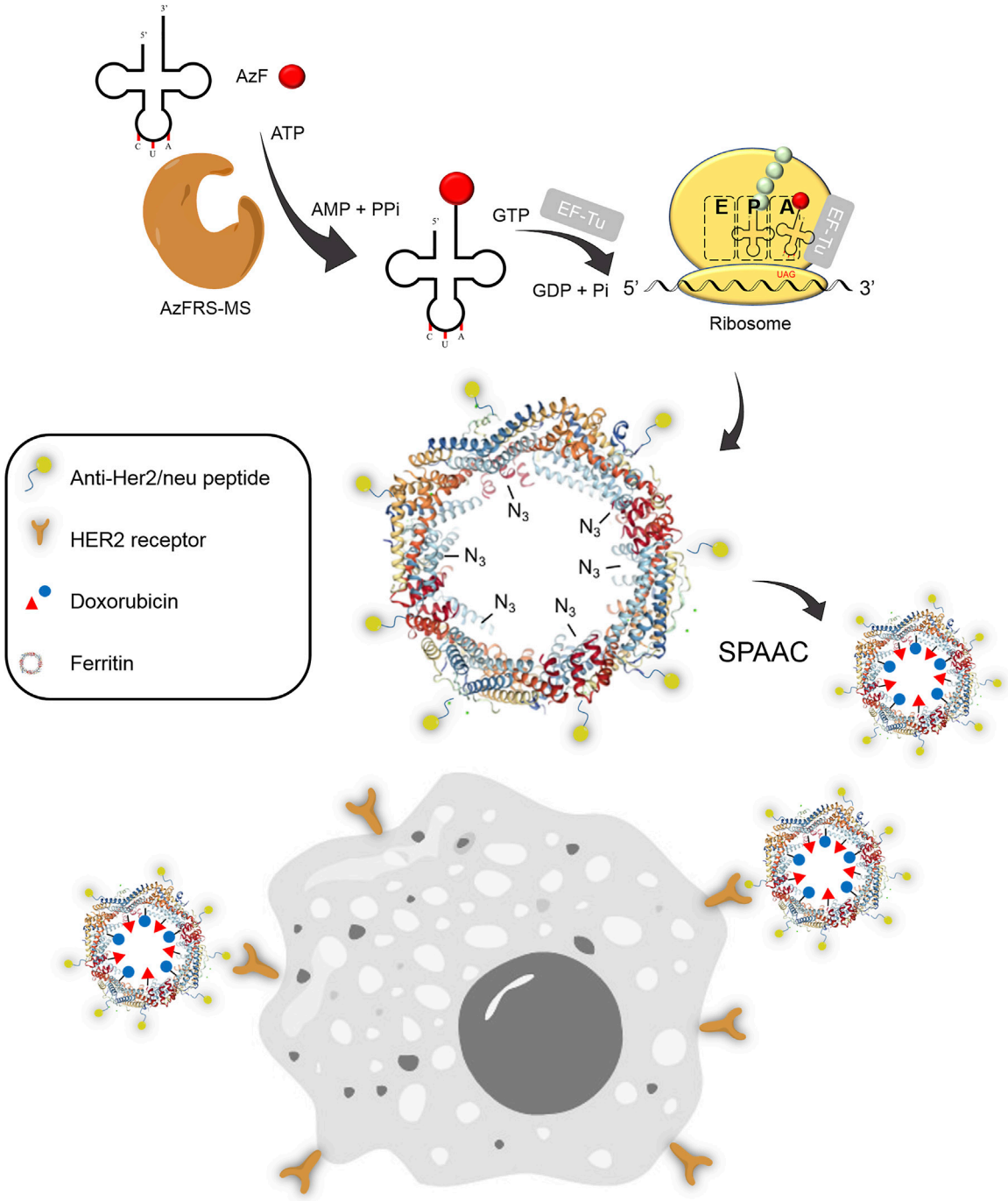
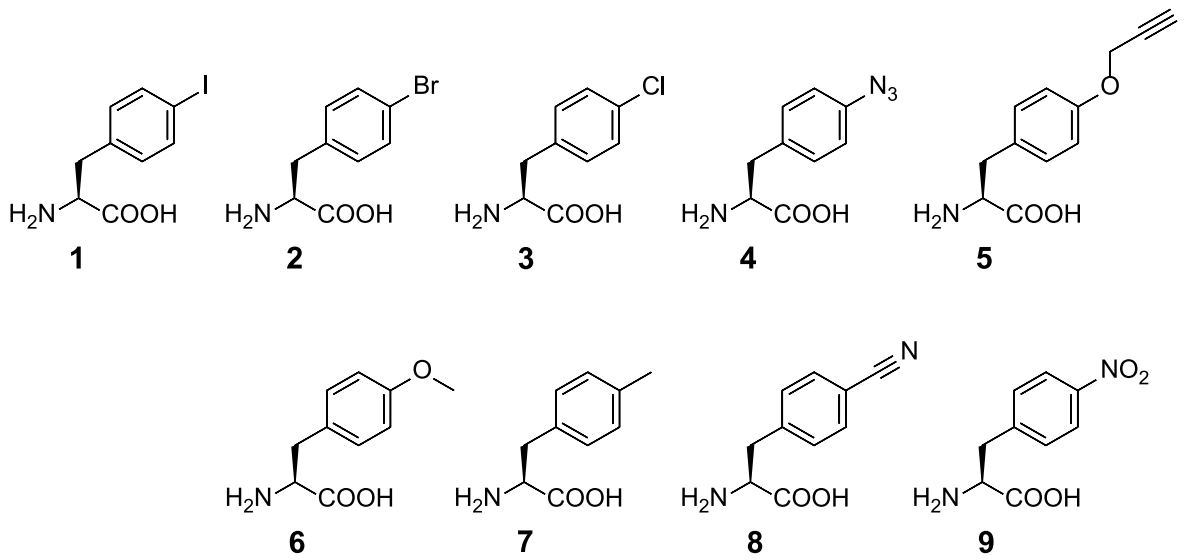


FIGURE 1 | Synthesis of the Ftn–Dox conjugate and the targeting of $Her2^{+}$ breast cancer cells. The Ftn–Dox conjugate is synthesized through the expanding genetic code approach and the SPAAC reaction. The engineered AzFRS-MS●$tRNA^{Pyl}$ pair is used to incorporate AzF (**4**, red filled circle) into Ftn. The installed azido group is shown as $–N_3$ on the Ftn nanocage. The Ftn–Dox conjugate with the installed anti-Her2/neu peptide (AHNP) facilitates the targeting of breast cancer cells, BT474, in this study. SPAAC, strain-promoted alkyne-azide cycloaddition.

(Tetter and Hilvert, 2017; Uchida et al., 2018) and chemotherapy drugs (Ahn et al., 2018; Song et al., 2021).

Ftn was firstly discovered and isolated from horse spleen. It is found to be highly conserved across all kingdoms of life and serves as an iron transporter in biological systems. Ftn consists of 24 subunits that can self-assemble into a spherical structure with an octahedral symmetry and an outer diameter of around 12 nm. The N-terminal region of Ftn is pointed outward, whereas the



SCHEME 1 | Chemical structures of ncAAs used in this study. ncAA **1**: *p*-iodo-L-phenylalanine (IF); **2**: *p*-bromo-L-phenylalanine (BrF); **3**: *p*-chloro-L-phenylalanine (ClF); **4**: *p*-azido-L-phenylalanine (AzF); **5**: *p*-propargyl-L-phenylalanine (PrF); **6**: *p*-methoxy-L-phenylalanine (MeOF); **7**: *p*-methyl-L-phenylalanine (MeF); **8**: *p*-cyano-L-phenylalanine (CNF); **9**: *p*-nitro-L-phenylalanine (NO_2F).

C-terminal E-helix of Ftn is pointed inward (**Figures 1**, **Figures 4A**) (Theil, 1987). For recognition purposes, the embedded targeting motifs are usually designed to be genetically fused to the N-terminus of Ftn, which allows 24 targeting motifs exposed on the Ftn surface to specifically recognize the target expressed on cells. The modified Ftn has been applied in cancer therapies (Song et al., 2021; Zhang et al., 2021).

In addition, Ftn maintains the assembled nanocage structure at the temperature up to 80°C and remains stable at the pH range of 4–12. Ftn can reversibly assemble the cage from disassembled monomers at neutral pH (Kim et al., 2011). Due to its physicochemical robustness, Ftn is an ideal candidate to be covalently modified into a *de novo* drug delivery platform.

The approach installing versatile and bioorthogonal functional groups by expanding genetic codes has become a practical method for the preparation of protein conjugates and is able to incorporate diverse bioorthogonal click chemistry pairs on proteins (Liu and Schultz, 2010). For instance, some potent *in vivo* bioorthogonal chemical reactions achieved through the expansion of genetic codes are copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC), strain-promoted alkyne-azide cycloaddition (SPAAC), and strain-promoted inverse electron-demand Diels–Alder cycloaddition (Lang and Chin, 2014). The bioorthogonal pyrrolysyl-tRNA synthetase (PylRS)•$tRNA^{Pyl}$ pairs from archaea *Methanosarcina mazei (Mm)* (Yanagisawa et al., 2006; Kavran et al., 2007; Chen et al., 2009) and *Methanosarcina barkeri* (Gautier et al., 2010), as well as eubacteria *Desulfitobacterium hafniense (Dh)* (Lee et al., 2008; Nozawa et al., 2009), are widely engineered to incorporate non-canonical amino acids (ncAAs), mainly for lysine (Cigler et al., 2017; Xuan et al., 2017; Abdelkader et al., 2021), phenylalanine (Phe, e.g., ncAAs **1–9**, **Scheme 1**) (Wang et al., 2011; Wang et al., 2012; Wang et al., 2013), tryptophan (Englert et al., 2015; Jiang et al., 2020b), histidine (Xiao et al., 2014; Sharma et al., 2016), cysteine (Nguyen et al., 2014), homoarginine (Mukai et al., 2015), and aspartic acid analogs (Xuan et al., 2018) into proteins *in vivo*. Therefore, clickable ncAAs such as azide, propargyl, *bi*-cyclooctyne, *trans*-cyclooctene, or tetrazine functional groups can be developed through these pairs (Wan et al., 2014; Tsai et al., 2015). The CuAAC and SPAAC reactions on proteins are demonstrated using genetically encoded *p*-azido-L-phenylalanine (AzF, **4**, **Scheme 1**) produced through amber TAG stop codon suppression by evolved *Methanococcus jannaschii* TyrRS•*Mj*$tRNA^{Tyr}$ pair in the *Escherichia coli* system (Chin et al., 2002). A semi-rationally designed *Dh*PylRS•*Dh*$tRNA^{Pyl}$ pair is also found to be able to charge **4** in *E. coli* (Fladischer et al., 2019). In addition, evolved *E. coli* TyrRS (E2AziRS)•*Ec*$tRNA^{Tyr}$ (Sakamoto et al., 2002; Coin et al., 2013) is applied to incorporate **4** onto proteins in mammalian cells. Among these pairs, the *Mm*PylRS•$tRNA^{Pyl}$ pair is shown to offer proper bioorthogonality spanning from prokaryotes to eukaryotes. Structural and engineering studies that contributed in the understanding of interactions among the N-terminal domain (NTD) (Suzuki et al., 2017; Sharma et al., 2018), the active site, and the linker region connecting the NTD and the active site (Suzuki et al., 2017; Jiang et al., 2020a) of *Mm*PylRS have been conducted. *Mm*PylRS harbors a large and dynamic active site; thus, the directed evolution of *Mm*PylRS leads to the development of mutants with higher suppression efficiency and the expansion of substrate ranges. Nevertheless, the *Mm*PylRS•$tRNA^{Pyl}$ pair has not evolved for charging **4** in *E. coli* and mammalian cells yet.

In this article, the application of rationally designed *Mm*PylRS variants in the preparation of a fluorophore- or drug-loaded Ftn delivery system is reported. The *Mm*PylRS variants with mutations on the C-terminal domain (CTD) that enhances

amber suppression for multiple **4** incorporations on Ftn are developed to enable the clickable chemistry, i.e., SPAAC, for the conjugation of fluorophores or drugs. Using the engineered *Mm*PylRS•tRNA^{Pyl} pair, a **4**-incorporated Ftn delivery system is designed to target a breast cancer overly expressed human epidermal growth factor receptor 2 (HER2-positive; Her2^{+}), which is more malignant and has a higher incidence as well as mortality rate among all breast cancer subtypes (Piccart-Gebhart et al., 2005; Romond et al., 2005). The N-terminus of this Ftn delivery system is fused with anti-Her2/neu peptide (AHNP) (Park et al., 2000), a peptide of 12 amino acids derived from trastuzumab which specifically binds to the HER2 receptor with high affinity. The chemotherapeutic agent, doxorubicin (Dox), whose mode of action occurs through its intercalation into dsDNA in the nucleus, is conjugated on this Ftn delivery system *via* the clickable SPAAC reaction. The successful results herein demonstrate the feasibility of similar designs for the development of chemical moiety-loaded protein cage systems and also provide another example for the versatile use of expanded genetic codes.

MATERIALS AND METHODS

General strains and plasmid construction

The ncAAs **1**–**9** were purchased from Chem-Impex Inc. (Wood Dale, IL, USA). Polymerase chain reaction (PCR) was performed using the KOD Hot Start Polymerase kit (Merck, Bedford, MA, USA). The oligonucleotide synthesis and DNA sequencing service were provided by Genomics Inc. (Taipei, Taiwan). The *sfGFP-10ams* gene, amber mutations at 10 phenylalanine positions F8/F27/F46/F71/F100/F114/F130/F145/F165/F223, and human heavy chain ferritin *Ftn* gene synthesis with *E. coli* codon optimization were performed by Integrated DNA Technologies (IDT) Inc. The primer sequences in this study are listed in the **Supplementary Materials**. The construction of plasmids pET-sfGFP-27am, pET-sfGFP-2~8ams, and sfGFP-10ams followed the general cloning protocols. The *sfGFP-27am* gene was generated using the *sfGFP* gene amplified by PCR from pET-pylT-sfGFP (Jiang et al., 2020a) and then prepared by overlapping PCR with two fragments (fragment 1: sfGFP-NdeI-F and 27am-R; fragment 2: 27am-F and sfGFP-SacI-R). The overlapped *sfGFP-27am* gene products and pET-pylT plasmid were double digested with restriction enzymes NdeI and SacI and then ligated using T4 DNA ligase. Similarly, pET-pylT plasmids harboring sfGFP with multiple TAG amber stop codons were constructed with multiple fragments by overlapping PCR. Briefly, the *sfGFP-2ams* gene with F27/F46 amber mutations was synthesized using three fragments: fragment 1, sfGFP-NdeI-F and 27am-R; fragment 2, 27am-F and 46am-R; and fragment 3, 46am-F and sfGFP-SacI-R. The *sfGFP-3ams* gene, with amber mutations at F8/F71/F100, was prepared using the following three fragments by overlapping the PCR method: fragment 1, NdeI-F8am-F and 71am-R; fragment 2, 71am-F and 100am-R; and fragment 3, 100am-F and sfGFP-SacI-R. The *sfGFP-4ams* gene with F27/F84/F100/F165 amber mutations was amplified by overlapping PCR using the following five fragments: fragment 1, sfGFP-NdeI-F and 27am-R; fragment 2, 27am-F and 84am-R; fragment 3, 84am-F and 100am-R; fragment 4, 100am-F and 165am-R; and fragment 5, 165am-F and sfGFP-SacI-R. The *sfGFP-5ams* gene with F27/F84/F100/F114/F165 amber mutations was prepared by two fragments (fragment 1: sfGFP-NdeI-F and 114am-R; and fragment 2: 114am-F and sfGFP-SacI-R), using pET-sfGFP-4ams as the template. The *sfGFP-6ams* gene with F27/F84/F100/F114/F165/F223 amber mutations was prepared by two fragments (fragment 1: sfGFP-NdeI-F and 223am-R; and fragment 2: 223am-F and sfGFP-SacI-R), using pET-sfGFP-5ams as the template. The *sfGFP-7ams* gene with the additional F8 amber mutation on the *sfGFP-6ams* gene was prepared by primers NdeI-F8am-F and sfGFP-SacI-R, using pET-sfGFP-6ams as the template. The *sfGFP-8ams* gene with S2 amber mutation on the *sfGFP-7ams* gene was prepared by primers NdeI-F2am-F and sfGFP-SacI-R, using pET-sfGFP-7ams as the template. The amber codon positions for sfGFP variants are listed in **Supplementary Table S1**.

For the construction of the pET-Ftn plasmid, the synthesized *Ftn* gene was inserted into the pET-pylT plasmid using the same approach as to that for the preparation of pET-sfGFP-27am. For the construction of pET-Ftn-F81am, pET-Ftn-K143am, and pET-Ftn-2ams, respectively, the *Ftn-F81am*, *Ftn-K143am*, and *Ftn-2ams* genes were synthesized from wild-type *Ftn* genes with F81 and/or K143 amber mutations by overlapping PCR. Briefly, the *Ftn-F81am* gene was generated by overlapping two fragments, fragment 1: Ftn-NdeI-F and F81am-R and fragment 2: F81am-F and Ftn-SacI-R, using the template pET-Ftn. The *Ftn-K143am* and *Ftn-2ams* genes were synthesized by the same approach using two fragments, fragment 1: Ftn-NdeI-F and K143am-R; and fragment 2: K143am-F and Ftn-SacI-R, using templates pET-Ftn and pET-Ftn-F81am, respectively. The plasmids pET-A-Ftn and pET-Ftn-A were also constructed by introducing genes encoding the N- or C-terminal AHNP peptide and linker $(GGGGS)_3$ to the *Ftn* gene to generate *A-Ftn* and *Ftn-A* genes. The *A-Ftn* and *Ftn-A* genes were synthesized with two fragments using pET-Ftn as the template: fragment 1 (N-AHNP-F and N-AHNP-R for *A-Ftn* construction; whereas C-ANHP-F and C-ANHP-R for *Ftn-A*) and fragment 2 (3XG4S-Ftn-F and 3XG4S-Ftn-R for *A-Ftn*, whereas Ftn-3XG4S-C-ANHP-F and Ftn-3XG4S-C-ANHP-R for *Ftn-A*). The plasmids pET-A-Ftn-F81am, pET-A-Ftn-K143am, and pET-A-Ftn-2ams were constructed utilizing the same approach, with the same primers for the construction of the *A-Ftn* gene and using pET-Ftn-F81am, pET-Ftn-K143am, and pET-Ftn-2ams as the templates, respectively.

*Mm*PylRS variants, IFRSs, evolved from directed evolution (Wang et al., 2011) which charge *p*-iodo-L-phenylalanine (IF, **1**), were generated with designated mutations on the active site (**Table 1**). Briefly, for the construction of the plasmid of pCDF-IFRS1, the *IFRS1* gene product was generated by overlapping PCR with three fragments using the template pCDF-PylRS (Jiang et al., 2020b), fragment 1: EcoR1-F and IFRS1-MLS-R; fragment 2: IFRS1-MLS-F and IFRS1-SM-R; and fragment 3: IFRS1-SM-F and BamHI-R. The overlapped *IFRS1* gene product and pET-pylT plasmid were then double

TABLE 1 | PylRS variants employed in this study.

PylRS	L305	Y306	L309	N346	C348	W417	K431	D433	A441
IFRS1[a]	M	L	S	S	M				
IFRS2[a]		T	A	S	M				
AzFRS[a]				A	M	L			
AzFRSc				A	M	L	M	G	S
AzFRS-M				A	M	L	M		
AzFRS-G				A	M	L		G	
AzFRS-S				A	M	L			S
AzFRS-MG				A	M	L	M	G	
AzFRS-MS[b]				A	M	L	M		S
AzFRS-GS				A	M	L		G	S

[a]*These three PylRS variants are directed evolved from plasmid library selection against ncAA* ***1****.*
[b]*This variant is used for both ncAA-encoded sfGFP and ferritin protein production.*

digested with restriction enzymes EcoRI and BamHI and then ligated using T4 DNA ligase. pCDF-IFRS2 was constructed by modifying the *IFRS1* gene of pCDF-IFRS1, and the overlapping PCR product was generated with two fragments, fragment 1: EcoR1-F and IFRS2-TA-R and fragment 2: IFRS1-TA-F and BamHI-R. For the construction of pCDF-AzFRS, the *AzFRS* gene product was generated with three fragments using pCDF-PylRS as the template, fragment 1: EcoRI-F and AzFRS-AM-R; fragment 2: AzFRS-AM-F and AzFRS-L-R; and fragment 3: AzFRS-L-F and BamHI-R. Using pCDF-AzFRS as the DNA template, different combinations of C-terminal mutations (listed in the **Supplementary Materials**), K431M, D433G, and A441S, were introduced through the designed primers to generate *AzFRS-M*, *AzFRS-G*, *AzFRS-S*, *AzFRS-MG*, *AzFRS-MS*, *AzFRS-GS*, and *AzFRSc* genes (**Table 1**).

Expression and purification of ncAA-encoded sfGFP and Ftn proteins

To produce ncAA-encoded sfGFP proteins, the pET-sfGFP-27am and pCDF-AzFRS-MS co-transformed into *E. coli* BL21 (DE3), sequentially. After the transformation, the cell suspension was spread onto the agar plate containing ampicillin (Amp) (100 μg/ml) and streptomycin (Sm) (100 μg/ml). A single colony was chosen from the plate and then cultured in 1 ml Luria–Bertani (LB) medium overnight. The cultured cells were then transferred to 50 ml fresh LB medium and incubated at 37°C until OD_{595} reached 0.6–0.8. Followed by centrifugation (10 min, 6,000 rpm, 4°C), the medium is changed to GMML medium. Protein expression was induced with the supplement of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and ncAA and incubated at 37°C for another 12 h. Cells were harvested, resuspended with lysis buffer (1× phosphate-buffered saline [PBS], pH 7.4), and then sonicated to lyse cells. Followed by centrifugation (60 min, 20,000 rpm, 4°C), the supernatant was collected. As sfGFP was designed with a 6× His tag, the supernatant collected was incubated with 0.5 ml Ni^{2+}-NTA resin (Roche, Basel, Switzerland) for protein purification. Five milliliters of lysis buffer and 2.5 ml of washing buffer (1× PBS, 5 mM imidazole, pH 7.4) were used to wash out nonspecific proteins on the resin. The target protein was eluted from the resin by 2.5 ml of elution buffer (1× PBS, 200 mM imidazole, pH 7.4). The buffer of eluted fraction was changed to 1× PBS with Amicon Ultra-15 Centrifugal Filter Units (MWCO 10 kDa). For Ftn proteins, further purification by anion exchange column HiTrap Q HP (GE Healthcare, Chicago, IL, USA) was conducted with NaCl gradient elution buffer (50 mM Tris, 50 mM to 1 M NaCl, pH 8.0) and size exclusion chromatography (SEC) column HiLoad 10/300 Superdex 200 pg (GE Healthcare) with PBS buffer. Purified sfGFP and Ftn were analyzed by 12% SDS-PAGE with Instant Blue (Marvelgent Biosciences, Canton, MA, USA) staining. The Ftn assembling status was analyzed by native gel, 4%–15% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad, Hercules, CA, USA) with a native protein marker (Thermo Fisher Scientific, Waltham, MA, USA) in TG buffer (25 mM Tris–HCl, 192 mM glycine, pH 8.5) under 80 V for 6 h at 4°C. The gel was stained by Instant Blue for 15 min and destained by H_2O.

Western blot analysis

Whole cells were collected and lysed at 100°C with SDS loading dye for 15 min and then subjected to 12% SDS-PAGE analysis. The gels were stained with Instant Blue to visualize the target proteins with the expected molecular weight around 28 kDa corresponding to the protein size of sfGFP. The suppression efficiency of amber codons in producing sfGFP with the C-terminal His tag was visualized by Western blotting against the anti-6X His tag antibody. The Western blots were performed using Trans-Blot Turbo System (Bio-Rad) and RTA transfer kit. The antibodies used for immunoblotting were Anti-His (SignalChem, Richmond, Canada; H99-61M-100) and HRP-conjugated secondary antibody (Cell Signaling Technology, Danvers, MA, USA; 7076P2). After SDS-PAGE analysis, the gel was immersed in the transfer buffer and then blotted onto the polyvinylidene fluoride (PVDF) membrane (25 V/1.3 A, 10 min). After the completion of the transfer process, the PVDF membrane was washed with PBST buffer for 5 min thrice. Next, the membrane was blocked with 5% skim milk for 1 h at room temperature. The membrane was then washed with PBST buffer for 5 min thrice (washing step). A primary antibody (1:1,000 dilution) was added and then incubated with the membrane for 1 h at room temperature following the washing step. Subsequently, an HRP-conjugated secondary antibody (1: 5,000 dilution) was added and the membrane was incubated for

another 1 h at room temperature. The membrane was then washed. Finally, the WesternBright ECL HRP substrate (Advansta, San Jose, CA, USA; K-12045-D50) was mixed and spread onto the membrane to visualize the band signals using ChemiDoc Imaging Systems (Bio-Rad) in bioluminescence detection mode.

ESI-MS characterization of sfGFP and Ftn proteins

The purified protein was diluted with 50% acetonitrile and 1% formic acid. An aliquot corresponding to one pmol of the pure protein was injected *via* an ESI source (Waters LockSpray Exact Mass Ionization Source) with a syringe pump (Harvard Apparatus, Holliston, MA, USA), and a flow rate of 5 μl/min was held throughout the analysis. The mass of intact proteins was determined using a Waters SYNAPT G2 HDMS mass spectrometer (Waters, Milford, MA, USA). The acquired spectra were deconvoluted to the single-charge state using the MaxEnt1 algorithm of the MassLynx 4.1 software (Waters).

Determination for the suppression efficiencies of MmPylRS variants

To understand the suppression of *Mm*PylRS variants, ncAA **1**–**9** screening assay was performed. The plasmids pET-sfGFP-27am and pCDF-PylRS were co-transformed into *E. coli* BL21 (DE3), sequentially. The cell suspension was spread onto the agar plate containing Amp (100 μg/ml) and Sm (100 μg/ml). The plate was incubated at 37°C overnight. Ten colonies were selected and then inoculated and cultured in 3 ml of LB medium at 37°C overnight. Five hundred microliters of cells was transferred to 25 ml of fresh LB medium and incubated at 37°C until OD_{595} reached 0.6–0.8. The cells were harvested, washed twice with M9 salts, and suspended in M9 medium (M9 salts, 1% glycerol, 2 mM $MgSO_4$ and 0.1 mM $CaCl_2$) supplemented with 1 mM IPTG. Aliquots of 50 μl suspended cells were loaded in a 384-well plate supplemented with **1**–**9** ncAAs (1 mM) in designated wells. Cells were incubated in the plate reader (BioTek, Winooski, VT, USA) at 37°C for 12 h, with continuous monitoring of fluorescence emission intensity ($\lambda_{ex}/\lambda_{em}$ = 535/595 nm) as well as OD_{595}. The control experiments used to measure background signals were performed for wells without adding IPTG and ncAAs, or wells adding IPTG only. The control signal (without adding ncAAs but IPTG) was subtracted from the fluorescence emission intensity of sfGFP obtained and then divided by OD_{595} to generate the relative fluorescence intensity.

Synthesis of Ftn-Dye and A-Ftn–DOX conjugates

To synthesize Ftn-dye conjugates, purified Ftn-81-**4**, Ftn-143-**4**, Ftn-2am-**4**, A-Ftn-81-**4**, A-Ftn-143-**4**, and A-Ftn-2am-**4** proteins were conjugated to fluorescence dyes or Dox by the SPAAC reaction. The DBCO-Cy3 and DBCO-Cy5 (Sigma-Aldrich, St. Louis, MO, USA) were dissolved in 100% DMSO to the final concentration of 1 mM. The Ftn-81-**4** protein (20 μl, 20 μM) in PBS buffer at pH 7.4 was mixed with 50 μl of DBCO-Cy3 and/or DBCO-Cy5 (80 μM, 1:5 M ratio) in the buffer of 90% PBS and 10% DMSO, and then the volume was adjusted to 500 μl. Reaction was incubated at 37°C, 300 rpm for 16 h. A-Ftn-81-TAMRA conjugates were synthesized by mixing A-Ftn-81-**4** and DBCO-PEG4-TAMRA (Sigma-Aldrich) under the same condition. The Ftn-dye conjugates were further purified using the SEC column (HiLoad 10/300 Superdex 200 pg, GE Healthcare) and then subjected to SDS-PAGE, native PAGE, gel fluorescence imaging, and ESI-MS analysis. A-Ftn-143-Dox was synthesized by coupling A-Ftn-143-**4** and dibenzocyclooctyne–PEG4–doxorubicin (DBCO–PEG4–DOX) using the same method. DBCO–PEG4–DOX was synthesized by NHS-amine coupling chemistry. DBCO–PEG4–NHS (Sigma-Aldrich) (50 μl, 1.5 mM, 30% DMSO in PBS) and doxorubicin hydrochloride (Dox, TCI) (50 μl, 2 mM, 30% DMSO in PBS) were mixed and reacted at 25°C for 4 h. The chemical reaction progress was monitored by normal-phase thin-layer chromatography.

Dynamic light scattering, transmission electron microscopy, and fluorescence spectroscopic analysis of Ftn variants

To analyze the particle size by dynamic light scattering (DLS), Ftn variants were diluted to the final concentration of 0.5 mg/ml and the aggregates were removed from the solution by centrifuge. The protein samples were loading into the designated 1 ml disposable cuvettes, and the particle size of protein was measured by Malvern Zetasizer Nano ZS. Measurements were repeated three times, and the primary size distribution by intensity value was determined.

To prepare Ftn samples for transmission electron microscope (TEM) analysis, 50 μg/ml of protein was dissolved in 50 mM Tris buffer at pH 8.0. After the support film grid (formvar/carbon 400 mesh copper) discharged with 25 mA for 30 s by Emitech K100X Glow, 5 μl of protein was loaded onto the grid and incubated for 60 s. The sample was gently drained from the grid by filter paper. Distilled deionized water and 1% uranyl acetate (UA) were alternatively applied onto the grid, incubated for 30 s, and drained by filter paper three times for background cleaning. Images of Ftn proteins were collected on FEG-TEM, FEI Tecnai G2 TF20 Super TWIN. Microscopy was operated at an accelerating voltage of 120 kV.

The fluorescence emission spectra were determined by Fluorolog-3 (Jobin Yvon, Edison, NJ, USA). Proteins were diluted to 50 μg/ml in PBS at pH 7.4 and loaded into the designated 10-mm quartz cuvettes. Ftn-143-Cy3, Ftn-143-Cy5, and Ftn-143-Cy3/Cy5 were excited at 550 nm. The emission spectra were recorded from 560 to 700 nm with 1 nm per point scanning resolution.

In vitro BT474 cell targeting study, confocal microscopy imaging, and cytotoxicity assay

Human breast carcinoma (BT-474) cells (BCRC, Taiwan) were cultured in 45% high-glucose DMEM (Corning, Tewksbury, MA, USA) and 45% low-glucose DMEM (Corning, USA)

supplemented with 10% fetal bovine serum (FBS, HyClone, Logan, UT, USA) and 1% penicillin-streptomycin (Gibco, Grand Island, NY, USA) at 37°C with 5% CO_2 in a humid atmosphere in petri dishes (Falcon). Cells were ready for the detachment when the confluency reached above 60%. BT-474 cells were rinsed with PBS twice and then trypsinized with 0.25% trypsin–EDTA (Gibco, USA). An appropriate volume of trypsin-EDTA was added into petri dishes. After gently shaking, cells were incubated with trypsin–EDTA at 37°C for 1 min. The detached cells were collected by pipetting over the cell layer with fresh complete medium. Twenty microliters of suspended cells was taken and mixed with trypan blue for cell density measurement. Cell density was determined by an automated cell counter (Invitrogen, Carlsbad, CA, USA). One-third of cells was added into a new petri dish and diluted to the appropriate volume with fresh complete medium. Poly-D-lysine (Sigma-Aldrich) coated coverslips were used for BT-474 cell attachment. Coverslips were incubated with 50 μg/ml poly-D-lysine for 2 h. Subsequently, the solution was removed and then the coverslips were rinsed with PBS and sterile water three times, alternatively. The coated coverslips were allowed to dry for at least 2 h before introducing cells and medium. Suspended cells (1×10^5) were seeded on 18-mm coverslips within a 12-well plate and incubated for 24 h, allowing cells to attach on coverslips.

After cells were attached on coverslips, the culture medium was replaced by fresh DMEM containing Dox, fluorescence molecules, A-Ftn-dye, or A-Ftn-drug conjugates. The cells were incubated at 37°C for the designated time and washed twice with PBS solution. The cells were fixed on coverslips by incubating with 4% para-formaldehyde (PFA) (Sigma-Aldrich) for 10 min. PFA was removed and the coverslips rinsed twice with PBS. After equilibrating the coverslips with PBS, the cells were ready to be stained. The nucleus of cells was stained by DAPI (GeneCopoeia, Rockville, MD, USA). An appropriate volume of 300 nM DAPI was added into a culture plate with coverslips and incubated for 5–10 min. Then, DAPI was removed and the coverslips alternatively rinsed with PBS and sterile H_2O three times. The coated coverslips were allowed to dry for at least 5 min before mounting. The cells were mounted onto a slide with one drop or 10 μl of mounting medium, ProLong Gold Antifade Mountant (Invitrogen). The cells were observed under the corresponding channel by confocal microscopy (Leica SP5).

The cytotoxicity of the A-Ftn–DOX conjugate, A-Ftn-**4,** and free Dox, respectively, against $Her2^+$ human breast cancer cells, BT-474, was evaluated by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, Thermo) assay. As the confluency of BT-474 reached above 60%, the cells were detached and 1×10^4 cells were seeded onto a 96-well transparent flat plate per well with 100 μl of fresh medium. The cells were further incubated for 24 h to allow the cell to attach onto the plate. Subsequently, the free Dox, A-Ftn-**4**, and A-Ftn–DOX conjugates were added into designated wells with the final concentration ranging from 0 to 5 μM. After the incubation for 72 h, the medium was removed and replaced by 100 μl of 1.2 mM MTT in PBS per well, following the 4-h incubation. The MTT solution was carefully removed, and DMSO was added into wells to dissolve the purple formazan crystals. The absorption at 540 and 570 nm was measured by Synergy HT plate reader (Biotech).

RESULTS

Efficient MmPylRS variants for p-azido-L-phenylalanine (4) incorporation at multiple sites on proteins

To incorporate AzF (**4**), the evolved PylRS•$tRNA^{Pyl}$ pair (Wang et al., 2011) is chosen from the *Mm*PylRS plasmid library (Wang et al., 2010) to be fine-tuned in charging ncAA **1**. The wild-type *Mm*PylRS and three IFRS variants, IFRS1, IFRS2, and AzFRS (**Table 1**), are selected to explore the substrate recognition against nine *para*-substituted phenylalanine analogs, **1**–**9** (**Scheme 1**; **Figure 2A**), respectively. The mutations at the active site of these three *Mm*PylRS variants cover two conserved mutation sites at N346 and C348, together with randomly distributed mutations at other sites (**Table 1**; **Figure 2B**). Both the pCDF vector containing the mutated PylRS gene and the pET vector containing $tRNA^{Pyl}$ and the *sfGFP-27am* gene are co-introduced into *E. coli*. The read-through of the amber codon in response to ncAAs **1**–**9** (**Scheme 1**), respectively, can be indicated through the fluorescence emission resulting from the *sfGFP-27am* gene product. The screening results obtained from both IFRS1 and IFRS2 give the preference in charging IF (**1**) and MeOF (**6**) (**Figure 2A**), indicating that both IFRS1 and IFRS2 show high substrate specificity. All AzFRS variants share W417L/N346A/C348M mutation sites (**Table 1**). The screening results of AzFRS variants give the high preference when charging **6**, whereas mild preferences are observed when charging **1**–**5**. AzF (**4**) and *p*-propargyl-L-phenylalanine (PrF, **5**), from which versatile functional groups can be derived for click reactions such as SPAAC and CuAAC, show noticeable signal enhancement when being charged by AzFRS.

To enhance the activity of *Mm*PylRS variants, the engineering for the second sphere of the active site or the terminal sequence of *Mm*PylRS is conducted. Although the second sphere or global sequence engineering (Suzuki et al., 2017; Sharma et al., 2018; Jiang et al., 2020a) of *Mm*PylRS has been studied, the effects of *Mm*PylRS CTD mutations on amber suppression efficiency remain unclear. Based on the available x-ray crystal structure, mutations at the CTD for rationally designed AzFRS have been introduced and studied (**Figure 2C**). Among the evolved *Mm*PylRS variants, HarRS, which charges homoarginine with K431M/D433G/G444E mutations at HarRS CTD (Mukai et al., 2015), is found to be able to enhance the activity and selectivity for homoarginine incorporation. Likewise, in this study, to develop the hypothesized remodeled pi–pi interactions among Y242, H432, and F434 positions, K431M/D433G double mutations are chosen for the CTD mutation sites due to their strong hydrophobicity (**Figures 2C,D**). In addition, the A441S mutation is performed to affect the adjacent hydrogen bond network built by the main chain of the R439 residue and to indirectly participate in the hydrogen bond network between the side chain of R439 and the neighboring alpha helix (**Figure 2C**,

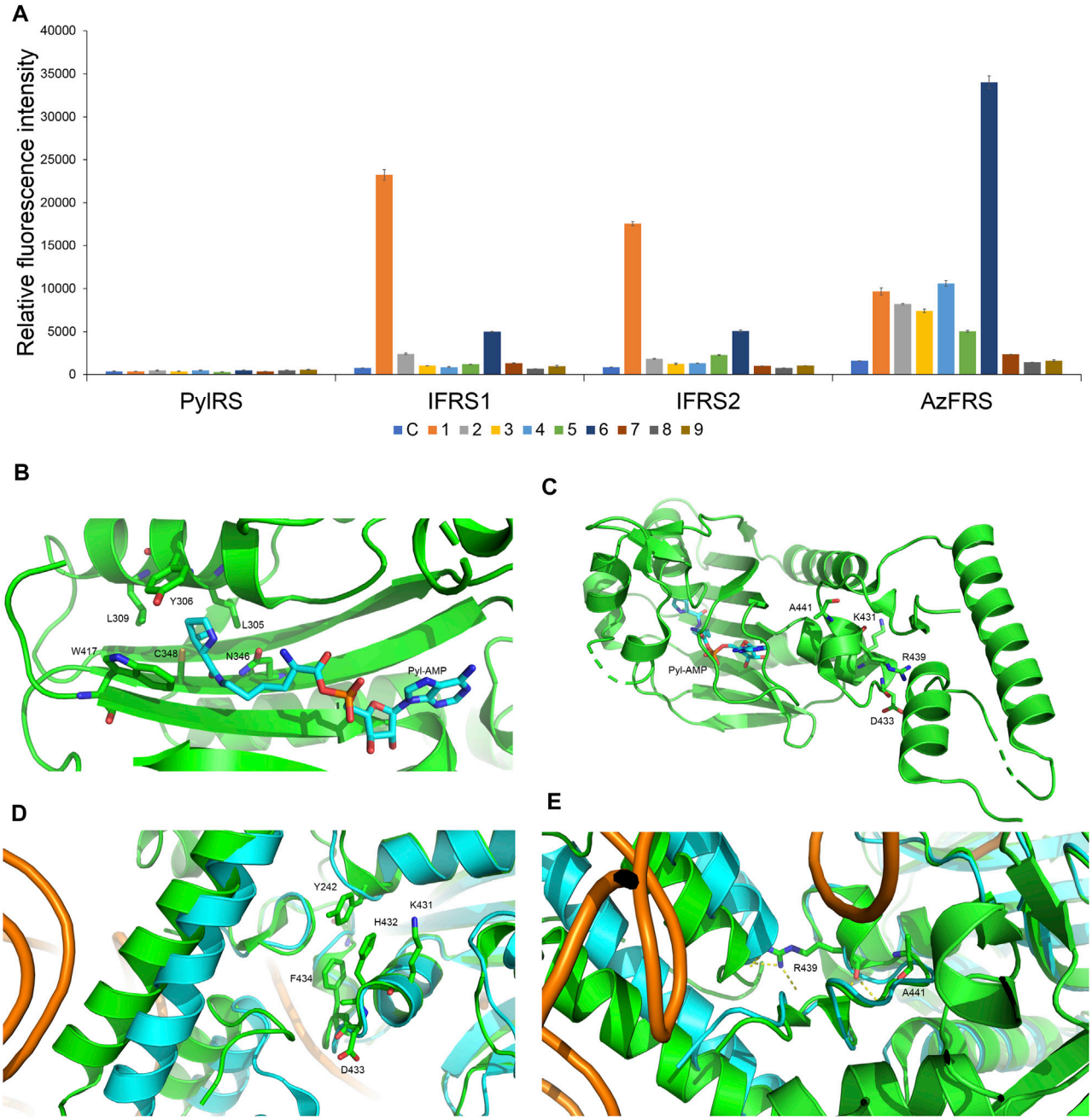


FIGURE 2 | Substrate specificity screening and C-terminal designs of evolved *Mm*PylRS. **(A)** The incorporation efficiencies of PylRS variants, IFRS1, IFRS2, and AzFRS (**Table 1**), are determined by fluorescence emission intensities obtained from *sfGFP-27am* gene products. Proteins are expressed after being supplemented with 1 mM ncAA and IPTG in GMML medium at 37°C for 12 h. Cells are excited at 485 nm, and the fluorescence emission intensities are detected at 535 nm. The cell density is monitored by the UV absorbance at 595 nm. C denotes the control experiment that cells were supplemented with 1 mM IPTG only; **1–9** denote the experiments supplemented with 1 mM IPTG and ncAAs **1–9** (**Scheme 1**). The background signal obtained from cells without adding IPTG is subtracted from the signal obtained from each group. Error bars represent the standard deviation of sfGFP production from three repeated experiments. **(B)** The active site of *Mm*PylRS. **(C)** The structure of *Mm*PylRS CTD in the presence of Pyl-AMP substrate obtained from the X-ray co-crystal structure (PDB code: 2Q7H). The AMP-Pyl and six mutated sites from directed evolution are shown in stick mode. The designed mutation sites, K431, D433, and A441, in this study are labeled. The residue R439 is also labeled to indicate the spatial arrangement toward the first helix (position 191–201). **(D)** and **(E)** The structure of *Mm*PylRS CTD (green) and its superimposition with the *Dh*PylRS CTD (cyan)/$tRNA^{Pyl}$ (orange) complex. The structures are illustrated based on two PDB entries 2Q7H and 2ZNI. The amino acids proposed to interact with the mutated site are labeled. The yellow dotted line indicates the hydrogen bonding.

positions 191–201), whose spatial movement is found in the *Dh*PylRS CTD/tRNA^{Pyl} (cyan) co-crystal structure (**Figures 2D,E**). Those mobile helical regions (green and cyan helices on the left side of **Figure 2E**) are interfaced with the tRNA D-loop region based on the superimposed structures of the *Dh*PylRS CTD/tRNA^{Pyl} co-crystal (cyan, **Figure 2E**) and the *Mm*PylRS CTD (green, **Figure 2E**).

Upon the introduction of three mutations, K431M/D433G/A441S, on *Mm*PylRS to evolve into the AzFRS variant, another three variants with a single mutation at the AzFRS CTD are prepared to examine its effect: AzFRS-M, AzFRS-G, and AzFRS-S (**Table 1**). Through monitoring the fluorescence emitted from the *sfGFP-27am* gene product, it is noted that, compared with AzFRS, AzFRS-M and AzFRS-G give the 1.4- and 1.6-fold enhancement in the fluorescence emission in charging MeOF (**6**), respectively, and give the 1.6- and 1.3-fold enhancement in the fluorescence emission in charging AzF (**4**), respectively (**Figure 3A**). Compared with AzFRS, AzFRS-S yields 2- to 4.3-fold enhancement in the fluorescence emission when charging **1–6**, which shows the improved activity toward those ncAAs. To examine the likelihood whether a combination of those mutations can further elevate the activity of AzFRS, AzFRS variants with double mutations, i.e., AzFRS-MG, AZFRS-MS, and AzFRS-GS, as well as the AzFRS variant with triple mutations, i.e., AzFRSc, are prepared and tested with the same approach. Interestingly, D433G mutation does not seem to contribute to the enhanced activity. The AzFRS-MG variant merely shows a neutral effect, whereas the AzFRS-GS variant even shows a negative effect. Relative to AzFRS, AzFRS-MS and AzFRSc variants yield a 4-fold increase in the fluorescence emission when charging **4**. Based on these results, AzFRS-MS is chosen in charging **4** in the subsequent studies.

The amber suppression efficiencies of AzFRS, AzFRSc, and AzFRS-MS are evaluated through the generation of *sfGFP-2ams*, *sfGFP-3ams*, *sfGFP-4ams*, *sfGFP-5ams*, *sfGFP-6ams*, *sfGFP-7ams*, *sfGFP-8ams*, and *sfGFP-10ams* gene products (Amiram et al., 2015) (**Figure 3B**; **Supplementary Figures S1, S2**). With the similar endogenous protein background, the whole-cell SDS-PAGE and Western blotting analysis of full-length sfGFP indicate an improvement in the suppression efficiency. AzFRS and AzFRS-MS can overly express *sfGFP-2ams*, *sfGFP-3ams*, *sfGFP-4ams*, *sfGFP-5ams*, and *sfGFP-6ams* gene products, with the given protein amount obtained from AzFRS-MS which is greater than that obtained from AzFRS (**Figure 3B**; **Supplementary Figure S1**). Furthermore, the full-length *sfGFP-7ams*, *sfGFP-8ams*, and *sfGFP-10ams* gene products can be expressed after introducing K431M/A441S double mutations to AzFRS CTD (**Figure 3B**; **Supplementary Figure S1**). Although AzFRSc shows slightly better amber read-through efficiency than AzFRS-MS through the observation of *sfGFP-27am* gene product generation, AzFRSc does not generate the full-length *sfGFP-3ams* and *sfGFP-8ams* gene products (**Supplementary Figure S2**). Therefore, AzFRS-MS is selected for the characterization and its Ftn-**4** protein production is examined. To further confirm the ncAA screening results of AzFRS-MS, sfGFP-**1**~**9** are respectively expressed and purified and then subjected to ESI-MS analysis. The observed mass of all nine sfGFP proteins generated by AzFRS-MS matches the calculated mass of the full-length sfGFP (**Table 2**; **Figure 3C**; **Supplementary Figures S3–S11**).

Ferritin conjugate preparation and chemical moiety loading using SPAAC

The Ftn consists of 24 subunits and can self-assemble into a spherical structure with the outer diameter of 12 nm. The sites for the chemical moiety loading, i.e., a fluorophore or a drug with the molecular weight less than 1,000 Da, are selected based on the x-ray crystal structure of Ftn (PDB code: 2FHA, **Figure 4A**). The residues F81 pointed outward (labeled in blue in **Figure 4A**) and K143 pointed inward (labeled in red in **Figure 4A**) on Ftn are chosen to estimate drug loading efficiency based on the penetration of small molecules into the Ftn cavity and the SPAAC reaction. *Ftn-F81am*, *Ftn-K143am*, and *Ftn-2ams* gene products supplemented with AzFRS-MS and AzF (**4**) produce Ftn-81-**4**, Ftn-143-**4**, and Ftn-2ams-**4**, respectively, and these products are characterized by ESI-MS spectrometry. The measured molecular weight of those products matches their corresponding calculated molecular weight (**Table 2**; **Supplementary Figures S12–14**). It is found that, for Ftn-81-**4** and Ftn-2ams-**4**, azide reduction occurs, followed by the conversion to an amino group (**Table 2**).

DBCO-Cy3 and DBCO-Cy5 (molecular weight: 980.28 and 992.28 Da, respectively) are used to probe the labeling efficiency of the SPAAC reaction in loading DBCO-dye (**Figure 4B**). The quantitative labeling of Ftn-81-**4** with DBCO-Cy5 is conducted with the 1:5 M ratio. The SDS-PAGE analysis shows that Ftn is not being labeled after DBCO-Cy5 treatment while Ftn-81-**4** gives the bioorthogonality upon DBCO-Cy5 treatment as a clear band can be visualized (**Figure 4C**). For Ftn-81-**4**, after 4 h treatment with DBCO-Cy5, a species with about 1.0 kDa mass shift is noted from the deconvoluted ESI-MS spectrum, which indicates that Cy5 is labeled through the azido group on Ftn-81-**4**. With the reaction time of 16 h, Ftn-81-**4** (found molecular weight: 22,220 Da; calculated molecular weight: 22,220 Da) is found to be completely converted into Ftn-81-Cy5 (found molecular weight: 23,207 Da; calculated molecular weight: 23,207 Da). A slight azide reduction on Ftn-81-**4** turning into an amino group is observed (22,173 Da); in addition, an oxygen adduct of Ftn-81-Cy5 (23,223 Da) is found (**Table 2**; **Figure 4D**; **Supplementary Figures S12, S15**).

In consideration of structural disruption and disassembling, DBCO-Cy3 labeling on Ftn, Ftn-81-**4**, Ftn-143-**4**, and Ftn-2am-**4** for 4 h is tested; the corresponding labeling efficiency and assembling ratio are then analyzed by SDS-PAGE and native PAGE (**Figure 4E**). It is noted that Ftn is not being labeled after the treatment with DBCO-Cy3 and remains assembled but shows less protein amount after treating with DBCO-Cy3, which might be resulted from the aggregation. Relatively, Ftn-81-**4**, Ftn-143-**4**, and Ftn-2am-**4** proteins give shifted fluorescent bands on SDS-PAGE and clear fluorescent bands of the assembled protein on native PAGE upon being labeled (**Figure 4E**). On the SDS-PAGE, Ftn-2am-**4** gives two bands that represent single- and double-labeled Cy3 adducts.

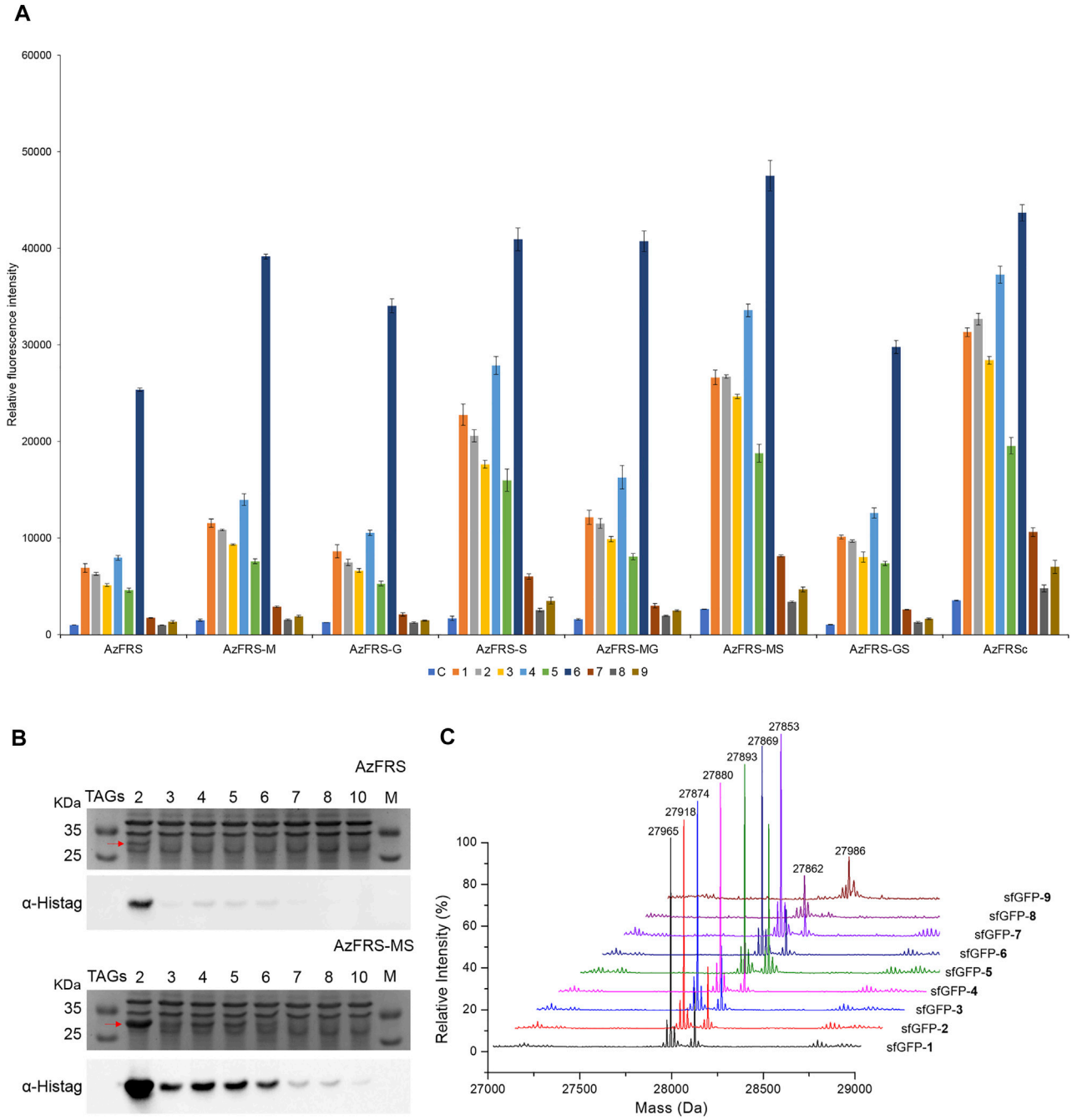


FIGURE 3 | sfGFP production, amber suppression efficiency, and mass characterization obtained from AzFRS variants. **(A)** The incorporation efficiencies of AzFRS variants (**Table 1**) are determined by the fluorescence emission intensities obtained from *sfGFP-27am* gene products, respectively. Proteins are expressed after supplemented with 1 mM ncAA and IPTG in GMML medium at 37°C for 12 h. Cells are excited at 485 nm, and the fluorescence emission intensities are recorded at 535 nm. The cell density is monitored through the UV absorbance at 595 nm. C denotes the control experiment that cells supplemented with 1 mM IPTG only; **1–9** denote the experiments supplemented with 1 mM IPTG and ncAAs **1–9** (**Scheme 1**). The background signal obtained from cells without adding IPTG is subtracted from the signal obtained from each group. Error bars represent the standard deviation of sfGFP production from three repeated experiments. **(B)** Tandem amber codon suppression efficiency analysis for sfGFP production by AzFRS●tRNA^{Pyl} and AzFRS-MS●tRNA^{Pyl} pairs. Amber suppression of *sfGFP-2ams~8ams* and *sfGFP-10ams* genes, containing suppression of 2~10 TAG stop codons, produces full-length sfGFP proteins with the incorporation of multiple **4**. The sfGFP proteins are overexpressed in *E. coli* BL21 (DE3) coding AzFRS●tRNA^{Pyl} or AzFRS-MS●tRNA^{Pyl} pair supplemented with 1 mM IPTG and 1 mM **4** in GMML medium at 37°C for 12 h. The whole-cell lysate is analyzed by SDS-PAGE and Western blotting against the anti-Histag antibody (indicated as α-Histag). The red arrow indicates the sfGFP protein band on the SDS-PAGE. The SDS-PAGE and Western blotting analysis are shown in **Supplementary Figures S1, S2**; **(C)** Mass determination of *sfGFP-27am* gene products coupled with ncAAs **1–9** by ESI-MS. The full-length sfGFP-**1–9** proteins are overexpressed by the AzFRS-MS●tRNA^{Pyl} pair in *E. coli* BL21 (DE3) supplemented with 1 mM IPTG and 1 mM ncAAs **1–9** in GMML medium at 37°C for 12 h. The calculated and observed molecular masses of sfGFP-**1–9** are found in **Table 2**. The detailed electrospray and deconvoluted mass spectrum are shown in **Supplementary Figures S3–S11**.

TABLE 2 | ESI-MS analysis of sfGFP and ferritin proteins that produced by AzFRS-MS●tRNA^{Pyl} pair.

Protein[a]	Calculated Mass (Da)	Found Mass (Da)	ESI-MS[b]
sfGFP-**1**	28,096, 27,965 (–Met)[c]	28,096, 27,965	S3
sfGFP-**2**	28,049, 27,918 (–Met)	28,049, 27,918	S4
sfGFP-**3**	28,005, 27,874 (–Met)	28,005, 27,874	S5
sfGFP-**4**	28,011, 27,880 (–Met)	28,011, 27,880	S6
sfGFP-**5**	28,024, 27,893 (–Met)	28,024, 27,893	S7
sfGFP-**6**	28,000, 27,869 (–Met)	28,000, 27,869	S8
sfGFP-**7**	27,984, 27,853 (–Met)	27,984, 27,853	S9
sfGFP-**8**	27,864 (–Met)	27,862	S10
sfGFP-**9**	28,015	27,986[d]	S11
Ftn-81-**4**	22,200, 22,174[e]	22,199, 22,174[e]	S12
Ftn-143-**4**	22,219	22,219	S13
Ftn-2am-**4**	22,261, 22,235[e]	22,261, 22,233[e]	S14
Ftn-81-Cy5	23,207, 22,174[e]	23,207, 23,223[f], 22,172[e]	S15
A-Ftn-81-**4**	23,448, 23,422[e]	23,451, 23,426[e]	S16
A-Ftn-81-DOX	24,525, 23,422[e]	24,524, 23,421[e]	S17

[a]*All sfGFP proteins are produced from the* sfGFP-27am *gene in GMML medium.*
[b]*Full-length sfGFP electrospray and deconvoluted mass spectra.*
[c]*(–Met) indicates full-length sfGFP protein without N-terminus methionine residues.*
[d]*The found mass of 27,986 Da corresponds to the full-length sfGFP with nitro group reduction, i.e., with an additional amino group, at F27 position and without N-terminal methionine (calculated mass: 27,985 Da).*
[e]*These found masses correspond to the full-length Ftn with azide group reduction, which is converted to an amino group.*
[f]*An oxygen adducts on Ftn-81-Cy5.*

With the successful SPAAC reaction in the DBCO-Cy5 and DBCO-Cy3 labeling on Ftn-81-**4** Ftn-143-**4**, and Ftn-2am-**4**, respectively, the capacity for the incorporation of different and multiple molecules/drugs is subjected to be evaluated. As the inner diameter of Ftn is 8 nm, Cy3 and Cy5 dyes inside Ftn deem to be a suitable fluorescence resonance energy transfer (FRET) pair (Roy et al., 2008) to probe adjacent Cy3/Cy5 within a 70-Å distance. Ftn-143-**4** is selected and treated with DBCO-Cy3 (λ_{ex}/λ_{em} = 555/570 nm) and DBCO-Cy5 (λ_{ex}/λ_{em} = 640/664 nm) dyes at the 1:2.5:2.5 M ratio for 16 h. FRET of the Ftn-143-Cy3 and Ftn-143-Cy5 mixture (1:1; as the control experiment) and Ftn-143-Cy3/Cy5 are measured (Wu and Brand, 1994a; Wu and Brand, 1994b) to ensure that both Cy3 and Cy5 molecules can be co-incorporated into one cage (**Figure 4F**). After the excitation at 550 nm, the mixture of Ftn-143-Cy3 and Ftn-143-Cy5 (1:1) gives an emission peak at 572 nm but lacks the emission peak at 664 nm, which supports this control system. In contrast, after the excitation at 550 nm on Ftn-143-Cy3/Cy5, a low emission peak at 569 nm is observed but an extra high emission peak at 673 nm arises, which demonstrates the successful FRET and proves that both Cy3 and Cy5 are co-incorporated into the Ftn cavity (**Figure 4F**).

Preparation of Ftn targeting Her2$^+$ breast cancer cells

To equip Ftn with a moiety targeting Her2$^+$ breast cancer cells, Ftn is fused with AHNP through a non-structural triple GGGGS peptide at one of the two termini (**Figure 5A**). Since the fused AHNP may affect the self-assembling of Ftn, leading to twisted cavities, the structure analysis of the purified A-Ftn and Ftn-A is conducted using native PAGE, DLS, and TEM, respectively. Native PAGE analysis on A-Ftn and Ftn-A shows that they have similar assembled sizes and patterns, but their sizes are slightly larger than Ftn (**Figure 5B**). DLS analysis indicates that A-Ftn shares a similar outer diameter with Ftn (13.8 and 13.3 nm, respectively) (**Figure 5C**); TEM analysis shows that they both share a spherical shape (**Figure 5D**). However, DLS analysis shows that Ftn-A possesses a larger outer diameter of 18.0 nm (**Figure 5C**) and the TEM image shows that Ftn-A is irregular in its shape (**Figure 5D**). Thus, the A-Ftn construct is selected to target the HER2 receptor as it is more structurally suitable.

A-Ftn-81-**4** reacts with DBCO-PEG4-TAMRA (**Figure 6A**) at 1:5 M ratios for 4 h to form A-Ftn-81-TAMRA. SDS-PAGE analysis is performed after the reaction, which gives a fluorescent band with a larger size, supporting that A-Ftn is being labeled (**Figure 6B**). The effect of A-Ftn on the targeting of the HER2 receptor is evaluated using the BT474 breast cancer cell line. BT474 cells are treated with A-Ftn-81-TAMRA for 1 h and then observed by confocal microscopy. The fluorescence images obtained from confocal microscopy analysis indicate that A-Ftn-81-TAMRA can be found in the cytosol of BT474 cells, suggesting that A-Ftn-81-TAMRA can enter through the cell membrane following the recognition of AHNP by HER2 receptors (**Figure 6B**).

Synthesis of A-Ftn–Dox conjugates used in the treatment with BT474 cells

Upon the successful demonstration for the capability of A-Ftn entering into BT474 cells, a chemotherapeutic drug, Dox, is loaded into A-Ftn to test the cell targeting and the drug release (**Figure 1**). The DBCO-PEG4-DOX is synthesized through the NHS-amine coupling between DBCO-PEG4-NHS and Dox. Using the SPAAC reaction, A-Ftn-81-DOX, A-Ftn-143-DOX, and A-Ftn-2am-DOX are synthesized by mixing A-Ftn-81-**4**, A-Ftn-143-**4**, and A-Ftn-2am-**4** with DBCO-PEG4-DOX, respectively, for 16 h (**Figure 7A**). SDS-PAGE

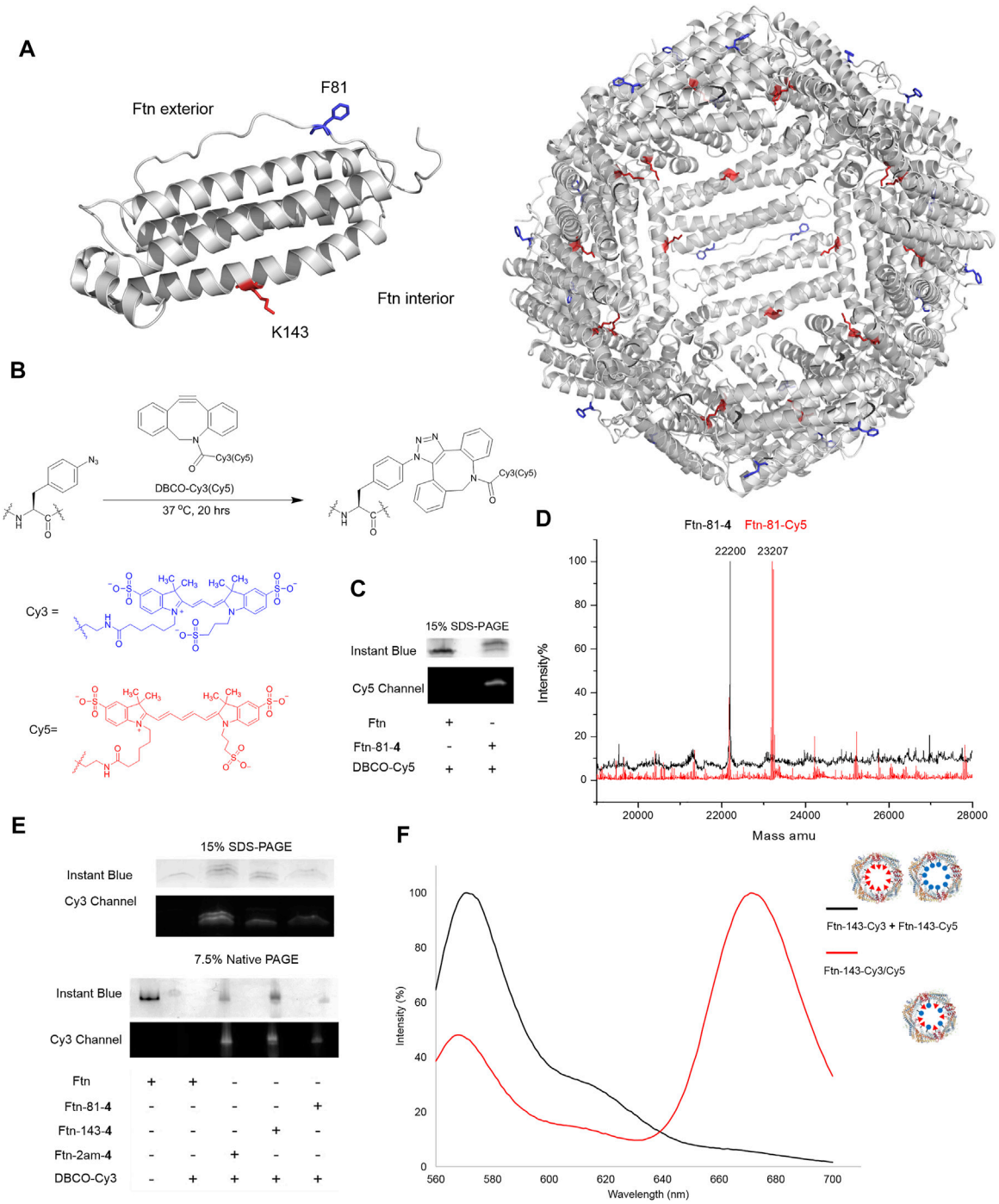


FIGURE 4 | Design of Ftn conjugates and click reaction study for the loading of Cy3 and Cy5. **(A)** Structure of Ftn monomer and the bisection of self-assembled 24mers. F81, facing the exterior of the Ftn nanoparticle, is shown in red stick; K143, facing the interior of the Ftn nanoparticle, is shown in blue stick. The PDB entry is 2FHA. **(B)** Synthesis scheme of SPAAC for loading DBCO-Cy3 ($\lambda_{ex}/\lambda_{em}$ = 555/570 nm) and DBCO-Cy5 ($\lambda_{ex}/\lambda_{em}$ = 640/664 nm) on Ftn-**4** proteins. **(C)** SPAAC of Ftn-81-**4** and DBCO-Cy5. SDS-PAGE is visualized by Instant Blue staining and Cy5. **(D)** The deconvoluted ESI-MS spectra of Ftn-81-**4** and Ftn-81-Cy5. The calculated and found molecular weights are listed in **Table 2**. **(E)** The SDS-PAGE and native PAGE analysis of Ftn-Cy3 conjugates. The gels are visualized by Instant Blue staining and Cy3 fluorescence. Twenty micromolars of Ftn protein and 1 mM DBCO-Cy3 are reacted in room temperature for 4 h. **(F)** FRET analysis of Ftn-dye conjugates with the excitation at 550 nm. The black line denotes Ftn-143-Cy3 and Ftn-143-Cy3 protein mixture, whereas the red line denotes Ftn-143-Cy3/Cy5 protein.

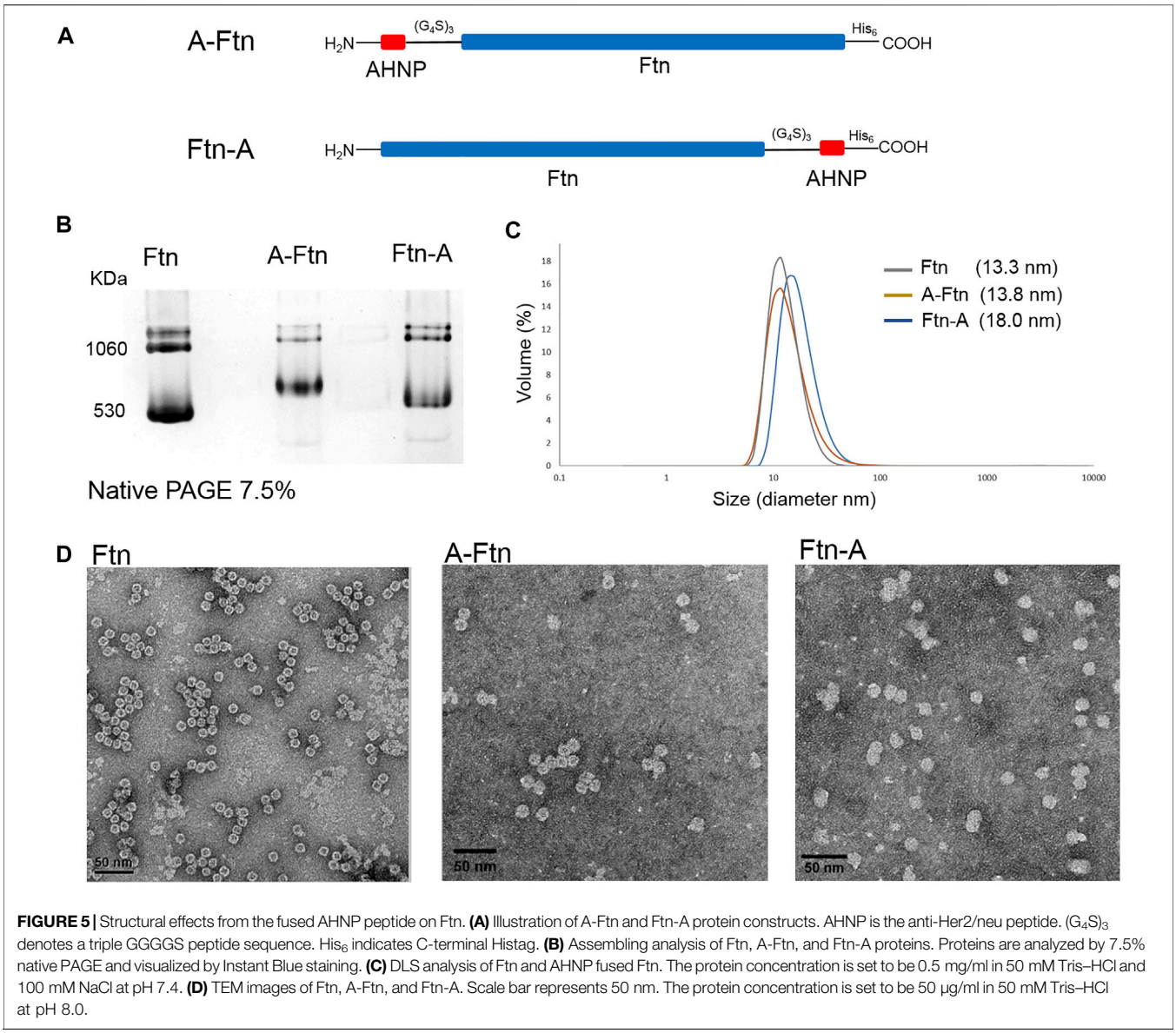


FIGURE 5 | Structural effects from the fused AHNP peptide on Ftn. **(A)** Illustration of A-Ftn and Ftn-A protein constructs. AHNP is the anti-Her2/neu peptide. $(G_4S)_3$ denotes a triple GGGGS peptide sequence. His_6 indicates C-terminal Histag. **(B)** Assembling analysis of Ftn, A-Ftn, and Ftn-A proteins. Proteins are analyzed by 7.5% native PAGE and visualized by Instant Blue staining. **(C)** DLS analysis of Ftn and AHNP fused Ftn. The protein concentration is set to be 0.5 mg/ml in 50 mM Tris–HCl and 100 mM NaCl at pH 7.4. **(D)** TEM images of Ftn, A-Ftn, and Ftn-A. Scale bar represents 50 nm. The protein concentration is set to be 50 μg/ml in 50 mM Tris–HCl at pH 8.0.

analysis shows that all three A-Ftn proteins are successfully conjugated with DOX and yield an almost complete conversion into A-Ftn–DOX products (**Figure 7B**). ESI-MS spectrometry is employed to characterize A-Ftn-81-**4** and A-Ftn-81-DOX, respectively. The results show that the protein mass matches the corresponding calculated protein mass, leading to the confirmation of quantitative labeling with the reaction time of 16 h (**Table 2**, **Supplementary Figures S14, S15**).

Furthermore, an A-Ftn-143-DOX conjugate is chosen for the evaluation of the capability in entering BT474 cells. The short-term (1 h) and long-term (72 h) treatment of A-Ftn-143-DOX with BT474 cells are examined, with the free Dox as the control treatment. MTT cell viability assay is conducted to examine the cell toxicity for BT474 cells after the treatment with free Dox, A-Ftn-143-**4**, and A-Ftn-143-DOX, respectively, for 72 h. A-Ftn-143-DOX gives a compatible trend as free Dox at the concentration from 0.5 to 5 μM while A-Ftn-143-**4** presents a negligible toxicity effect at this concentration range (**Figure 7C**). For the short-term (1 h) treatment of free Dox with BT474 cells, it is noted from the merged DAPI and Dox images that free Dox is found in the nucleus (**Figure 7D**). Nevertheless, for the short-term (1 h) treatment of A-Ftn-143-DOX with BT474 cells, Dox is mainly distributed in the cytosol of BT474 cells (**Figure 7E**), which suggests the elongated drug release. The slight Dox signal is observed in the nucleus, which resulted from free Dox molecules released from A-Ftn-143-DOX. After the long-term (36 h) treatment, free Dox released from A-Ftn-143-DOX is found to reside in the nucleus (**Supplementary Figure S18**).

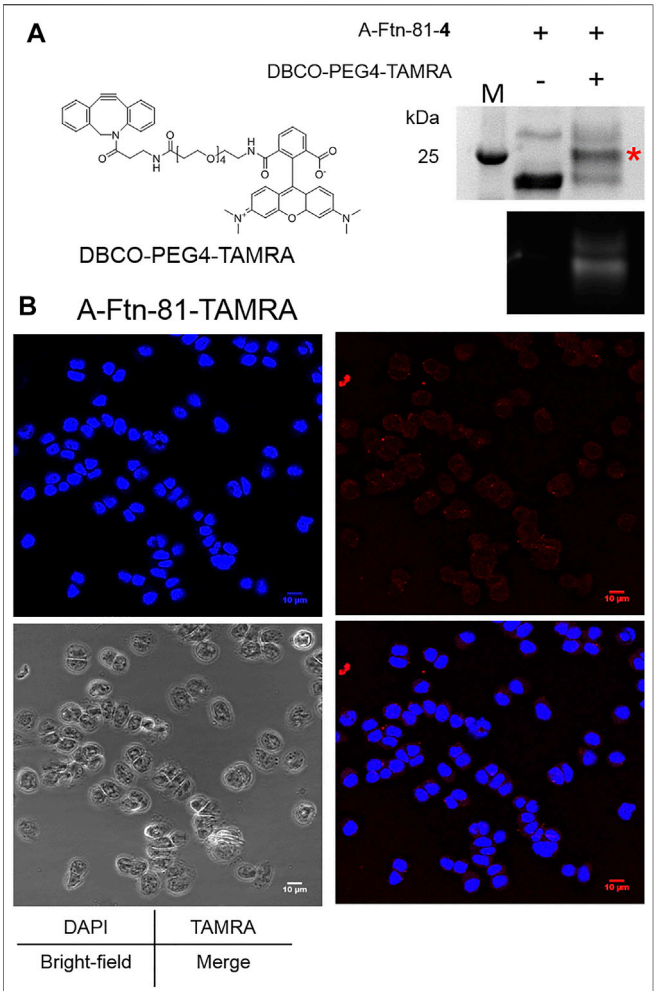


FIGURE 6 | A-Ftn-TAMRA conjugates treatment indicates Her2$^+$ BT474 cell targeting ability. **(A)** Chemical structure of DBCO-PEG4-TAMRA and SPAAC labeling analysis after the treatment of A-Ftn-81-4 with DBCO-PEG4-TAMRA. Proteins are analyzed by 15% SDS-PAGE with Instant Blue staining **(upper panel)** and gel fluorescence emission image **(lower panel)**. The red asterisk indicates the TAMRA-labeled Ftn band. M denotes the protein molecular weight marker. **(B)** Confocal microscopy fluorescence image analysis of BT474 cells incubated with 2 μM A-Ftn-81-TAMRA for 1 h. The cell nuclei are stained by DAPI (blue); A-Ftn-81-TAMRA is detected by TAMRA ($\lambda_{ex}/\lambda_{em}$ = 553 nm/563 nm, red).

DISCUSSION

In this study, efficient *Mm*PylRS is developed to charge a chemical moiety that allows the click reaction to occur on the desired delivery system, *i.e.*, the assembled Ftn. This facilitates the future application in the effective fluorophores/drug loading as well as target-specific chemical moiety release (**Figure 1**). Through the engineering of *Mm*PylRS charging IF (**1**) (**Scheme 1**), three evolved *Mm*PylRS variants, i.e., IFRS1, IFRS2, and AzFRS, obtained from directed evolution (Wang et al., 2011) are tested against nine *para*-substituted phenylalanine ncAAs (**Scheme 1**) for substrate range study (**Figure 2A**). Among these ncAAs, although only **4** that harbors an azide unit is selected for the examination using the SPAAC click reaction herein, some other substrates are also promising in the application in other types of click reaction. For instance, the ncAA **5**, which harbors a propargyl unit, can be employed when using the CuAAC click reaction.

With the conserved mutations at N346 and C348, and combined with other mutation sites, the evolved *Mm*PylRS, IFRS1 and IFRS 2 (**Table 1**), are found to be specific for certain substrates (**Figure 2A**). Another evolved *Mm*PylRS variant, AzFRS, with an additional conserved mutation, W417L, that plays as a gate keeper shows the largely enhanced activity for **6** and the mildly enhanced activity for **1**–**5** (**Figure 2A**). Upon the removal of the bulky indole ring-side chain on W417, the *para*-substituted iodo group can be perfectly harbored in the reshaped binding pocket of the active site (**Figure 2B**). Through the analysis of the substrate range for wild-type aminoacyl-tRNA synthetases from *E. coli* and evolved PylRS variants, the results show that their binding pockets may generate certain diversity as long as natural amino acids are excluded during the evolution process/history (Ko et al., 2013; Fan et al., 2014; Guo et al., 2014). Since AzFRS shows the enhanced activity toward **4** and **5**, the second sphere engineering of *Mm*PylRS is applied to AzFRS. The rationally designed K431M and D433G mutations, which reside at the CTD, are adapted from the evolved *Mm*PylRS, HarRS (**Figures 2C,D**) (Mukai et al., 2015). K431M and D433G are suggested to be able to alter the pi–pi stacking cluster among Y242, H432, and F434 by increasing hydrophobicity that may change the interaction with the nearby α-helices and then indirectly improve the activity (**Figure 2D**). Another A441S mutation at CTD is designed to alter the hydrogen binding network of the R439 guanidino group and to introduce the interaction with the first α-helix of CTD (**Figures 2C,E**). Through the examination of these single mutations on AzFRS, the results show that AzFRS-S can enhance the activity in charging a variety of substrates while AzFRS-M and AzFRS-G only slightly improve the activity for certain substrates (**Figure 3A**). Based on the result obtained from AzFRS with single mutations, double and triple mutations on AzFRS are designed to examine whether the synergic effect can be achieved (**Figure 3A**). Among them, AzFRS-MS and AzFRSc variants are found to be able to further enhance the activity toward miscellaneous substrates. AzFRS-MS can achieve suppression of 10 sequential TAG stop codons when charging **4**, producing the full-length sfGFP (**Figure 3B**; **Supplementary Figures S1, S2**). Compared with AzFRS, perhaps, the AzFRS-MS variant not only enhances the activity in a timely manner but also maintains constant acylated-tRNA concentrations during the protein biosynthesis *in vivo*. Meanwhile, due to the altered local hydrophobic interactions and hydrogen binding network which resulted from K431M and A441S mutations, it is suggested that AzFRS-MS can improve the protein solubility and accelerate the reaction rate. Although only the limited activity can be found on SDS-PAGE when charging **7**–**9** by monitoring the production of the *sfGFP-27am* gene product, ESI-MS spectrometry of sfGFP-**1**~**9** confirms that the observed molecular mass matches the corresponding calculated mass (**Table 2**; **Figure 3C**; **Supplementary Figures S3**–**S11**). Overall, AzFRS-MS presents

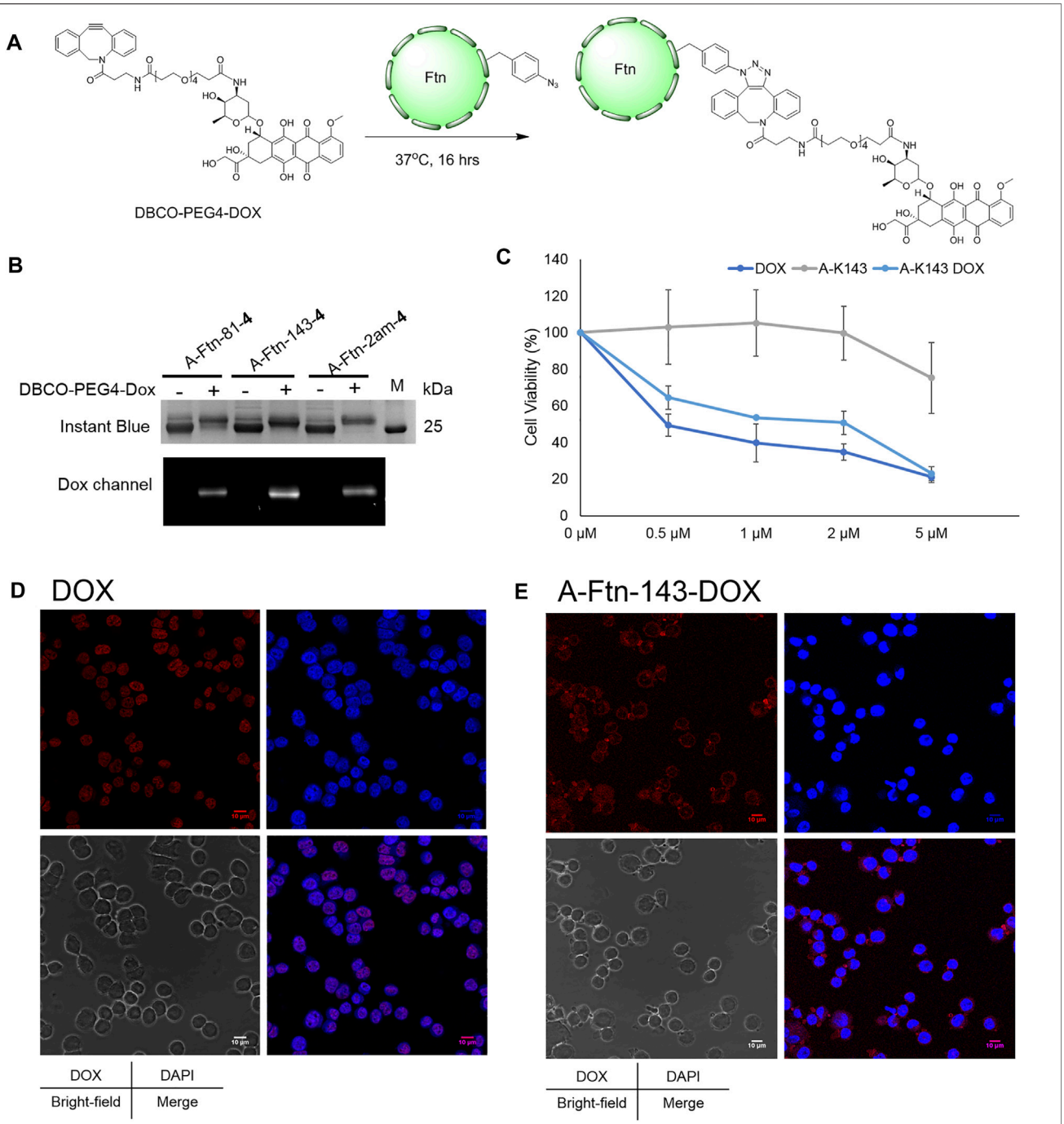


FIGURE 7 | Ftn–Dox conjugates target HER2 overexpressed breast cancer cells. **(A)** Synthesis scheme of Ftn–Dox conjugates. **(B)** SPAAC labeling analysis on A-Ftn-81-**4**, A-Ftn-143-**4**, and A-Ftn-2am-**4** by DBCO-PEG4-DOX. Proteins are analyzed by 15% SDS-PAGE with Instant Blue staining **(upper panel)** and gel fluorescence image **(lower panel)**. M denotes the protein molecular weight marker. **(C)** MTT assay analysis after treating with various concentrations of free Dox, A-Ftn-143-**4** (A-K143), and A-Ftn-143-DOX (A-K143 DOX) for 72 h. Confocal microscopy fluorescence image analysis of BT474 cells incubating with **(D)** 2 μM free Dox and **(E)** 2 μM A-Ftn-143-DOX treating for 1 h. The cell nuclei are visualized by DAPI staining (blue) and by gel fluorescence emission image through DOX ($\lambda_{ex}/\lambda_{em}$ = 480 nm/590 nm, red).

efficient poly-specificity toward all ncAAs employed in this study through the engineering of *Mm*PylRS CTD.

Using the AzFRS-MS variant, the residues F81 pointed outward and K143 pointed inward on Ftn are proved to be suitable for the incorporation of AzF (**4**), either single-incorporated or double-incorporated. The assembled caged structures (**Figure 4A**) are not disrupted by the incorporation of **4**. Interestingly, through the test of DBCO-Cy3 or DBCO-Cy5

(~1.0 kDa) labeling (**Figure 4B**) on Ftn-81-**4**, Ftn-143-**4**, and Ftn-2ams-**4** proteins, DBCO-Cy3 and DBCO-Cy5 are found to easily enter the interior of Ftn cages, not to cause noticeable structural disturbances and disassembling (**Figures 4C,E**). This is also supported by the labeling results observed from A-Ftn-81-**4**, A-Ftn-143-**4**, and A-Ftn-2ams-**4** proteins after the treatment of DBCO-DOX (~1.0 kDa), respectively (**Figure 7B**). These results suggest that the steric hindrance coming from F81 and K143 on the Ftn protein nanostructure for drug labeling through SPAAC chemistry is negligible. The quantitative labeling of Ftn-**4** by DBCO-fluorophores/drugs (**Figures 4A**, **7A**) is conducted through the SPAAC click reaction at the 1:5 M ratio. Ftn-81-Cy5 (**Table 1**; **Figures 4C,D**; **Supplementary Figures S12, S14**) and A-Ftn-81-DOX (**Table 1**; **Figure 7B**; **Supplementary Figures S16, S17**) proteins are characterized by SDS-PAGE, native PAGE, gel fluorescence imaging, and ESI-MS analyses. The Ftn cage might provide a protective space which shields molecules from redox reaction, hydrolysis, or even degradation. The findings herein imply that the labeling at different loading sites might affect the drug stability, and further investigation can be explored. For some drugs with a shorter half-life, the protein cage might prolong or alter the pharmacokinetics. This work shows that Ftn-**4** can be used to load desired drugs up to 48 units per Ftn cage and specifically labeled at F81 or K143 residue (**Figure 4E**). The loading of two different chemical moieties into a single Ftn cage is successfully demonstrated through the generation of Ftn-143-Cy3/Cy5 which is co-labeled with Cy3 and Cy5 (1:1) inside of the cage. This result suggests that a variety of chemical moieties can be loaded into one cage, which poses many potential applications in the future, for instance, loading multiple therapeutic agents in the cage to allow cancer combination therapy.

An example is given in this study to show the application of this designed delivery system. The targeting AHNP that recognizes the HER2 receptor is fused at one of the two termini of Ftn (**Figure 5A**). It is shown that A-Ftn maintains the assembled cage structure, whereas Ftn-A is found to be expanded in size with irregular shapes (**Figures 5C,D**). It is implied that AHNP fused at the C-terminus of Ftn causes the collapse of Ftn cavity and forms fused balls that are observed by TEM images (**Figure 5D**). It is observed from the treatment of synthesized A-Ftn-81-TAMRA with BT474 breast cancer cells that, AHNP on A-Ftn-81-TAMRA probably targets HER2 receptors on BT474 cells. Followed by the recognition, labeled Ftn enters BT474 cells and mainly resides in the cytosol. Some aggregated forms are observed, which may indicate the co-localization on the cell membrane and lysosome.

Breast cancer is the leading cause of cancer death globally, especially the $Her2^+$ subtype, which quickly spreads and potentially develops into metastatic breast cancer. The strategy described here enables the preparation of an Ftn platform conjugated with drugs to allow targeted therapy and chemotherapy for curing malignant breast cancer. DBCO-PEG4-DOX is incorporated on residues of A-Ftn, pointed either inward or outward (**Figures 7A,B**). After the treatment of A-Ftn-143-DOX with BT474 cells for 72 h, the MTT assay is performed to examine the cytotoxicity (**Figure 7C**). A-Ftn-143-DOX is shown to exhibit a similar trend as free Dox. However, free Dox is shown to reside at the nucleus immediately after the incubation for 1 h (**Figure 7D**). In contrast, after the incubation for 1 h, A-Ftn-143-DOX is shown to mainly locate in the cytosol and slightly distribute in the nucleus (**Figure 7E**). It is hypothesized that, after the 1-h incubation, Dox observed in the nucleus is released from A-Ftn-143-DOX and the release might be resulted from acidic hydrolysis and protease digestion in lysosomes. After the incubation for 36 h, most Dox released from A-Ftn-143-DOX is located in the nucleus (**Supplementary Figure S18**). As the hydrolysis and the digestion take time, this might pose the chance for the design of extended-release drugs in the future.

CONCLUSION

In conclusion, an evolved *Mm*PylRS variant, AzFRS-MS, can improve the amber suppression efficiency when charging AzF (**4**), resulting in suppression of up to 10 sequential TAG stop codons through the monitoring for the production of the *sfGFP-27am* gene product. The improved activity of AzFRS-MS is also established when charging the other eight ncAAs tested in this study. The charging of **4** on Ftn by AzFRS-MS allows the SPAAC chemistry for DBCO-mediated conjugation to occur, leading to the application on the preparation of Ftn-fluorophore/drug conjugates with the quantitative conversion. Using this approach, the N-terminal AHNP-fused Ftn–DOX conjugate is prepared and shown to be able to target HER2 receptors on BT474 breast cancer cells. Conjugated Dox is then slowly released into the nucleus. This showcase study demonstrates the potential use of this approach in therapeutic areas. In addition, the capability to co-label Cy3 and Cy5 in a single cage suggests that this Ftn delivery platform can harbor different chemical moieties for various future applications such as combination therapy. Taken together, an example is given herein to support that the evolved *Mm*PylRS systems enabling the chemical conjugation on proteins present the potential versatile applications, and they are worth further exploration.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Biosafety Committee of Academia Sinica. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Conceptualization, Y-SW; methodology, Y-HW, M-LJ, P-JC, J-CT, LT, and Y-SW; validation, Y-SW; formal analysis,

Y-HW, M-LJ, P-JC, J-CT; resources, Y-SW; data curation, Y-HW, M-LJ, P-JC, J-CT, LT, and Y-SW; writing—original draft preparation, Y-SW Y-HW, and M-LJ; writing—review and editing, Y-SW, P-JC, and J-CT; supervision, Y-SW; project administration, Y-SW; funding acquisition, Y-SW.

FUNDING

This research was funded by the Ministry of Science and Technology (MOST 107-2113-M-001-025-MY3 and MOST 110-2113-M-001-044) and Academia Sinica, Taiwan.

ACKNOWLEDGMENTS

We are grateful to Dr. Wen-I Luo for reading the manuscript and insightful suggestions. Mass spectrometry data were acquired at the Academia Sinica (AS) Common Mass Spectrometry Facilities (Dr. Shu-Yu Lin) located at the Institute of Biological Chemistry (IBC), AS, supported by AS Core Facility and Innovative Instrument Project (AS-CFII-108-107). We also acknowledge the use of the DLS analysis in the Biophysics Core Facility (AS-CFII-108-111) and Dr. Meng-Ru Ho for technical assistance in fluorescence and absorbance spectrometry measurement, as well as AS Biological Electron Microscopy Core Facility (AS-CFII-108-119) for EM technical support. We thank the IBC mammalian cell culture facility and Bio-Imaging and Cell Biology Facilities for helping with confocal imaging.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.779976/full#supplementary-material

REFERENCES

Abdelkader, E. H., Qianzhu, H., Tan, Y. J., Adams, L. A., Huber, T., and Otting, G. (2021). Genetic Encoding of N6-(((Trimethylsilyl)methoxy)carbonyl)-L-Lysine for NMR Studies of Protein-Protein and Protein-Ligand Interactions. *J. Am. Chem. Soc.* 143 (2), 1133–1143. doi:10.1021/jacs.0c11971

Ahn, B., Lee, S.-G., Yoon, H. R., Lee, J. M., Oh, H. J., Kim, H. M., et al. (2018). Four-fold Channel-Nicked Human Ferritin Nanocages for Active Drug Loading and pH-Responsive Drug Release. *Angew. Chem. Int. Ed.* 57 (11), 2909–2913. doi:10.1002/anie.201800516

Amiram, M., Haimovich, A. D., Fan, C., Wang, Y.-S., Aerni, H.-R., Ntai, I., et al. (2015). Evolution of Translation Machinery in Recoded Bacteria Enables Multi-Site Incorporation of Nonstandard Amino Acids. *Nat. Biotechnol.* 33 (12), 1272–1279. doi:10.1038/nbt.3372

Aumiller, W. M., Uchida, M., and Douglas, T. (2018). Protein Cage Assembly across Multiple Length Scales. *Chem. Soc. Rev.* 47 (10), 3433–3469. doi:10.1039/c7cs00818j

Chen, P. R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B. H., et al. (2009). A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells. *Angew. Chem. Int. Ed.* 48 (22), 4052–4055. doi:10.1002/anie.200900683

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002). Addition of P-Azido-L-Phenylalanine to the Genetic Code of *Escherichia coli*. *J. Am. Chem. Soc.* 124 (31), 9026–9027. doi:10.1021/ja027007w

Cigler, M., Müller, T. G., Horn-Ghetko, D., von Wrisberg, M.-K., Fottner, M., Goody, R. S., et al. (2017). Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes *In Vitro* and *In Vivo*. *Angew. Chem. Int. Ed.* 56 (49), 15737–15741. doi:10.1002/anie.201706927

Coin, I., Katritch, V., Sun, T., Xiang, Z., Siu, F. Y., Beyermann, M., et al. (2013). Genetically Encoded Chemical Probes in Cells Reveal the Binding Path of Urocortin-I to CRF Class B GPCR. *Cell* 155 (6), 1258–1269. doi:10.1016/j.cell.2013.11.008

Englert, M., Nakamura, A., Wang, Y.-S., Eiler, D., Söll, D., and Guo, L.-T. (2015). Probing the Active Site Tryptophan ofStaphylococcus Aureusthioredoxin with an Analog. *Nucleic Acids Res.* 43 (22), 11061–11067. doi:10.1093/nar/gkv1255

Fan, C., Ho, J. M. L., Chirathivat, N., Söll, D., and Wang, Y.-S. (2014). Exploring the Substrate Range of Wild-type Aminoacyl-tRNA Synthetases. *Chembiochem* 15 (12), 1805–1809. doi:10.1002/cbic.201402083

Fladischer, P., Weingartner, A., Blamauer, J., Darnhofer, B., Birner-Gruenberger, R., Kardashliev, T., et al. (2019). A Semi-Rationally Engineered Bacterial Pyrrolysyl-tRNA Synthetase Genetically Encodes Phenyl Azide Chemistry. *Biotechnol. J.* 14 (3), 1800125. doi:10.1002/biot.201800125

Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A., and Chin, J. W. (2010). Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells. *J. Am. Chem. Soc.* 132 (12), 4086–4088. doi:10.1021/ja910688s

Guo, L.-T., Wang, Y.-S., Nakamura, A., Eiler, D., Kavran, J. M., Wong, M., et al. (2014). Polyspecific Pyrrolysyl-tRNA Synthetases from Directed Evolution. *Proc. Natl. Acad. Sci. USA* 111 (47), 16724–16729. doi:10.1073/pnas.1419737111

Jiang, H.-K., Lee, M.-N., Tsou, J.-C., Chang, K.-W., Tseng, H.-W., Chen, K.-P., et al. (2020a). Linker and N-Terminal Domain Engineering of Pyrrolysyl-tRNA Synthetase for Substrate Range Shifting and Activity Enhancement. *Front. Bioeng. Biotechnol.* 8, 235. doi:10.3389/fbioe.2020.00235

Jiang, H.-K., Wang, Y.-H., Weng, J.-H., Kurkute, P., Li, C.-L., Lee, M.-N., et al. (2020b). Probing the Active Site of Deubiquitinase USP30 with Noncanonical Tryptophan Analogues. *Biochemistry* 59 (24), 2205–2209. doi:10.1021/acs.biochem.0c00307

Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Soll, D., and Steitz, T. A. (2007). Structure of Pyrrolysyl-tRNA Synthetase, an Archaeal Enzyme for Genetic Code Innovation. *Proc. Natl. Acad. Sci.* 104 (27), 11268–11273. doi:10.1073/pnas.0704769104

Kim, M., Rho, Y., Jin, K. S., Ahn, B., Jung, S., Kim, H., et al. (2011). pH-Dependent Structures of Ferritin and Apoferritin in Solution: Disassembly and Reassembly. *Biomacromolecules* 12 (5), 1629–1640. doi:10.1021/bm200026v

Ko, J.-h., Wang, Y.-S., Nakamura, A., Guo, L.-T., Söll, D., and Umehara, T. (2013). Pyrrolysyl-tRNA Synthetase Variants Reveal Ancestral Aminoacylation Function. *FEBS Lett.* 587 (19), 3243–3248. doi:10.1016/j.febslet.2013.08.018

Lang, K., and Chin, J. W. (2014). Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* 114 (9), 4764–4806. doi:10.1021/cr400355w

Lee, M. M., Jiang, R., Jain, R., Larue, R. C., Krzycki, J., and Chan, M. K. (2008). Structure of Desulfitobacterium Hafniense PylSc, a Pyrrolysyl-tRNA Synthetase. *Biochem. Biophys. Res. Commun.* 374 (3), 470–474. doi:10.1016/j.bbrc.2008.07.074

Liu, C. C., and Schultz, P. G. (2010). Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* 79, 413–444. doi:10.1146/annurev.biochem.052308.105824

Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., et al. (2015). Reassignment of a Rare Sense Codon to a Non-canonical Amino Acid inEscherichia Coli. *Nucleic Acids Res.* 43 (16), 8111–8122. doi:10.1093/nar/gkv787

Nguyen, D. P., Mahesh, M., Elsässer, S. J., Hancock, S. M., Uttamapinant, C., and Chin, J. W. (2014). Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *J. Am. Chem. Soc.* 136 (6), 2240–2243. doi:10.1021/ja412191m

Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T., et al. (2009). Pyrrolysyl-tRNA synthetase-tRNAPyl Structure Reveals the

Molecular Basis of Orthogonality. *Nature* 457 (7233), 1163–1167. doi:10.1038/nature07611

Park, B.-W., Zhang, H.-T., Wu, C., Berezov, A., Zhang, X., Dua, R., et al. (2000). Rationally Designed Anti-HER2/neu Peptide Mimetic Disables P185HER2/neu Tyrosine Kinases *In Vitro* and *In Vivo*. *Nat. Biotechnol.* 18 (2), 194–198. doi:10.1038/72651

Piccart-Gebhart, M. J., Procter, M., Leyland-Jones, B., Goldhirsch, A., Untch, M., Smith, I., et al. (2005). Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *N. Engl. J. Med.* 353 (16), 1659–1672. doi:10.1056/NEJMoa052306

Romond, E. H., Perez, E. A., Bryant, J., Suman, V. J., Geyer, C. E., Jr., Davidson, N. E., et al. (2005). Trastuzumab Plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer. *N. Engl. J. Med.* 353 (16), 1673–1684. doi:10.1056/NEJMoa052122

Roy, R., Hohng, S., and Ha, T. (2008). A Practical Guide to Single-Molecule FRET. *Nat. Methods* 5 (6), 507–516. doi:10.1038/nmeth.1208

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., et al. (2002). Site-specific Incorporation of an Unnatural Amino Acid into Proteins in Mammalian Cells. *Nucleic Acids Res.* 30 (21), 4692–4699. doi:10.1093/nar/gkf589

Sharma, V., Wang, Y.-S., and Liu, W. R. (2016). Probing the Catalytic Charge-Relay System in Alanine Racemase with Genetically Encoded Histidine Mimetics. *ACS Chem. Biol.* 11 (12), 3305–3309. doi:10.1021/acschembio.6b00940

Sharma, V., Zeng, Y., Wang, W. W., Qiao, Y., Kurra, Y., and Liu, W. R. (2018). Evolving the N-Terminal Domain of Pyrrolysyl-tRNA Synthetase for Improved Incorporation of Noncanonical Amino Acids. *Chembiochem* 19 (1), 26–30. doi:10.1002/cbic.201700268

Song, N., Zhang, J., Zhai, J., Hong, J., Yuan, C., and Liang, M. (2021). Ferritin: A Multifunctional Nanoplatform for Biological Detection, Imaging Diagnosis, and Drug Delivery. *Acc. Chem. Res.* 54 (17), 3313–3325. doi:10.1021/acs.accounts.1c00267

Suzuki, T., Miller, C., Guo, L.-T., Ho, J. M. L., Bryson, D. I., Wang, Y.-S., et al. (2017). Crystal Structures Reveal an Elusive Functional Domain of Pyrrolysyl-tRNA Synthetase. *Nat. Chem. Biol.* 13 (12), 1261–1266. doi:10.1038/nchembio.2497

Tetter, S., and Hilvert, D. (2017). Enzyme Encapsulation by a Ferritin Cage. *Angew. Chem. Int. Ed.* 56 (47), 14933–14936. doi:10.1002/anie.201708530

Theil, E. C. (1987). Ferritin: Structure, Gene Regulation, and Cellular Function in Animals, Plants, and Microorganisms. *Annu. Rev. Biochem.* 56, 289–315. doi:10.1146/annurev.bi.56.070187.001445

Tsai, Y.-H., Essig, S., James, J. R., Lang, K., and Chin, J. W. (2015). Selective, Rapid and Optically Switchable Regulation of Protein Function in Live Mammalian Cells. *Nat. Chem* 7 (7), 554–561. doi:10.1038/nchem.2253

Uchida, M., McCoy, K., Fukuto, M., Yang, L., Yoshimura, H., Miettinen, H. M., et al. (2018). Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. *ACS Nano* 12 (2), 942–953. doi:10.1021/acsnano.7b06049

Wan, W., Tharp, J. M., and Liu, W. R. (2014). Pyrrolysyl-tRNA Synthetase: an Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta Proteins Proteom.* 1844 (6), 1059–1070. doi:10.1016/j.bbapap.2014.03.002

Wang, Y.-S., Fang, X., Chen, H.-Y., Wu, B., Wang, Z. U., Hilty, C., et al. (2013). Genetic Incorporation of Twelve Meta-Substituted Phenylalanine Derivatives Using a Single Pyrrolysyl-tRNA Synthetase Mutant. *ACS Chem. Biol.* 8 (2), 405–415. doi:10.1021/cb300512r

Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B., and Liu, W. R. (2012). A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum. *J. Am. Chem. Soc.* 134 (6), 2950–2953. doi:10.1021/ja211972x

Wang, Y.-S., Russell, W. K., Wang, Z., Wan, W., Dodd, L. E., Pai, P.-J., et al. (2011). The De Novo Engineering of Pyrrolysyl-tRNA Synthetase for Genetic Incorporation of L-Phenylalanine and its Derivatives. *Mol. Biosyst.* 7 (3), 714–717. doi:10.1039/c0mb00217h

Wang, Y.-S., Wu, B., Wang, Z., Huang, Y., Wan, W., Russell, W. K., et al. (2010). A Genetically Encoded Photocaged Nε-Methyl-L-Lysine. *Mol. Biosyst.* 6 (9), 1557–1560. doi:10.1039/c002155e

Wu, P., and Brand, L. (1994a). Conformational Flexibility in a Staphylococcal Nuclease Mutant K45C from Time-Resolved Resonance Energy Transfer Measurements. *Biochemistry* 33 (34), 10457–10462. doi:10.1021/bi00200a029

Wu, P. G., and Brand, L. (1994b). Resonance Energy Transfer: Methods and Applications. *Anal. Biochem.* 218 (1), 1–13. doi:10.1006/abio.1994.1134

Xiao, H., Peters, F. B., Yang, P.-Y., Reed, S., Chittuluru, J. R., and Schultz, P. G. (2014). Genetic Incorporation of Histidine Derivatives Using an Engineered Pyrrolysyl-tRNA Synthetase. *ACS Chem. Biol.* 9 (5), 1092–1096. doi:10.1021/cb500032c

Xuan, W., Collins, D., Koh, M., Shao, S., Yao, A., Xiao, H., et al. (2018). Site-Specific Incorporation of a Thioester Containing Amino Acid into Proteins. *ACS Chem. Biol.* 13 (3), 578–581. doi:10.1021/acschembio.7b00998

Xuan, W., Shao, S., and Schultz, P. G. (2017). Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angew. Chem. Int. Ed.* 56 (18), 5096–5100. doi:10.1002/anie.201611841

Yanagisawa, T., Ishii, R., Fukunaga, R., Nureki, O., and Yokoyama, S. (2006). Crystallization and Preliminary X-ray Crystallographic Analysis of the Catalytic Domain of Pyrrolysyl-tRNA Synthetase from the Methanogenic archaeonMethanosarcina Mazei. *Acta Cryst. Sect F* 62 (Pt 10), 1031–1033. doi:10.1107/S1744309106036700

Zhang, B., Tang, G., He, J., Yan, X., and Fan, K. (2021). Ferritin Nanocage: A Promising and Designable Multi-Module Platform for Constructing Dynamic Nanoassembly-Based Drug Nanocarrier. *Adv. Drug Deliv. Rev.* 176, 113892. doi:10.1016/j.addr.2021.113892

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 14 January 2022
doi: 10.3389/fchem.2021.825669

Orthogonal Activation of Metabotropic Glutamate Receptor Using Coordination Chemogenetics

Akinobu Senoo[1†], *Yutaro Yamada*[1†], *Kento Ojima*[1,2], *Tomohiro Doura*[1], *Itaru Hamachi*[2,3] *and Shigeki Kiyonaka*[1*]

[1]Department of Biomolecular Engineering, Graduate School of Engineering, Nagoya University, Nagoya, Japan, [2]Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Kyoto, Japan, [3]ERATO (Exploratory Research for Advanced Technology, JST), Tokyo, Japan

Cell-surface receptors play a pivotal role as transducers of extracellular input. Although different cell types express the same receptor, the physiological roles of the receptor are highly dependent on cell type. To understand each role, tactics for cell-specific activation of the target receptor are in high demand. Herein, we developed an orthogonal activation method targeting metabotropic glutamate receptor 1 (mGlu1), a G-protein coupled receptor. In this method, direct activation *via* coordination-based chemogenetics (dA-CBC) was adopted, where activation of mGlu1 was artificially induced by a protein conformational change in response to the coordination of a metal ion or metal-ion complex. Our structure-based protein design and screening approach identified mGlu1 mutants that were directly activated by the coordination of Cu^{2+} or Zn^{2+}, in addition to our previous Pd-complex-sensitive mGlu1 mutant. Notably, the activation of the mutants was mutually orthogonal, resulting in cell-type selective activation in a model system using HEK293 cells.

Keywords: chemogenetics, coordination chemistry, orthogonal activation, class C GPCR, metabotropic glutamate receptor

OPEN ACCESS

Edited by:
Yu-Hsuan Tsai,
Shenzhen Bay Laboratory, China

Reviewed by:
Shinya Tsukiji,
Nagoya Institute of Technology, Japan
Yoshihiro Kubo,
National Institute for Physiological Sciences (NIPS), Japan

***Correspondence:**
Shigeki Kiyonaka
kiyonaka@chembio.nagoya-u.ac.jp

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *30 November 2021*
Accepted: *28 December 2021*
Published: *14 January 2022*

Citation:
Senoo A, Yamada Y, Ojima K, Doura T, Hamachi I and Kiyonaka S (2022) Orthogonal Activation of Metabotropic Glutamate Receptor Using Coordination Chemogenetics. Front. Chem. 9:825669. doi: 10.3389/fchem.2021.825669

INTRODUCTION

Several hundred types of receptors are expressed on the cell surfaces of mammals, each of which plays essential roles in transmitting extracellular information into cells. Moreover, the same receptor is found in different cells or tissues, yet the physiological roles of each different receptor are highly dependent on the cell or tissue type (Vassilatis et al., 2003). Selective activation of receptors of interest in cell lines or primary cell cultures is a useful way to investigate the physiological roles of these receptors. However, chemical methods alone are insufficient to analyze the roles of receptors in a cell-specific manner in particular tissues or organs because of the challenges associated with cell-specific delivery of chemicals (Mondoloni et al., 2019). Thus, the development of new tools for cell-specific activation of target receptors is highly desired.

Chemogenetics, a process in which proteins of interest (POIs) are genetically engineered to selectively interact with designed chemicals, is a potential approach for cell-specific activation of target receptors (Islam, 2015; Atasoy and Sternson, 2018; Tsai et al., 2021). Some representative examples of chemogenetics include the bump-and-hole approach (Bishop et al., 2000; Knight and Shokat, 2007), ligand-induced stabilization (*i.e.*, chemical rescue) (Pratt et al., 2007), and chemically induced dimerization (Schreiber, 1998). Another well-known example of chemogenetic activation of

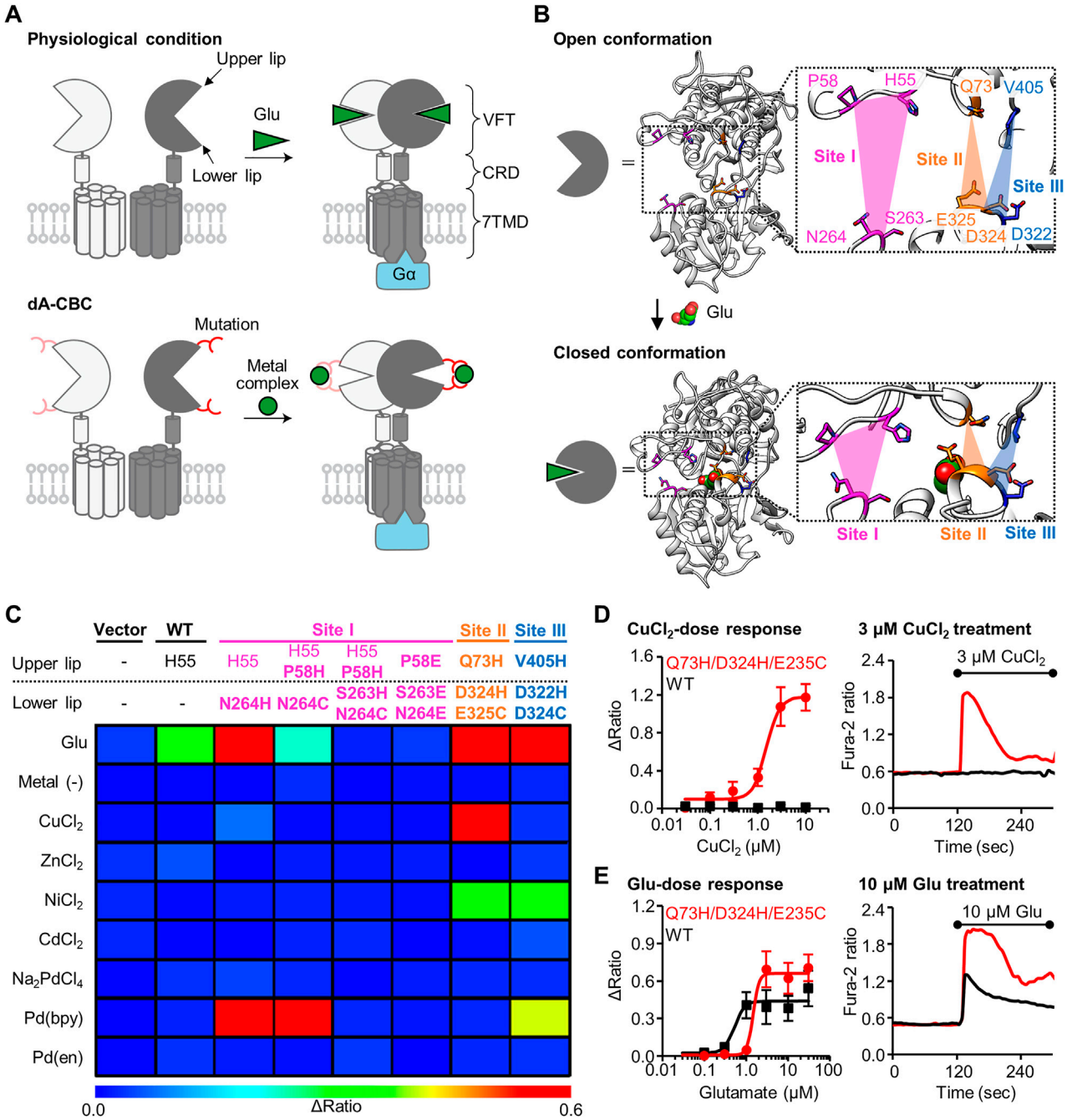


FIGURE 1 | Screening of mGlu1 mutants with sensitivity to metal ion or complex. **(A)** Schematic illustration of activation of mGlu1 in physiological condition or in dA-CBC. In dA-CBC, addition of metal ion or complex induces the conformational change required for signal transduction. **(B)** Crystal structure of VFT domain of mGlu1 (PDB ID; 1EWT, 1EWK). Mutation sites are colored in magenta, orange, or blue, which are named Site I, Site II, or Site III, respectively. **(C)** Heat map of the primary screening. The heat map shows average Δratio values upon the addition of 10 μM glutamate or 3 μM each metal ion or complex. Mutation sites are described separately for upper or lower lip of VFT domain, and the mutation is shown in bold characters. See **Supplementary Figure S2B** for raw data. **(D)** Activation of the hit mutant by Cu^{2+}. *Left*; dose dependency of Cu^{2+}-induced activation of WT mGlu1 (black) or the Q73H/D324H/E325C mutant (red). The EC_{50} value of the mutant was 1.5 μM. (n = 11–15). *Right*; Representative traces of $CuCl_2$ response as a function of time. The black bar represents the period when $CuCl_2$ was added. **(E)** Activation of the hit mutant by glutamate (Glu). *Left*; Dose dependency of Glu-induced activation of WT mGlu1 (black) or Q73H/D324H/E325C mutant (red). The EC_{50} values were 0.52 or 1.4 μM for WT mGlu1 or the mutant, respectively. (n = 12–15). *Right*; Representative traces of glutamate response as a function of time. The black bar represents the period when glutamate was added. Data are represented as mean ± s.e.m.

G-protein coupled receptors (GPCRs) is designer receptor exclusively activated by designer drugs (DREADD) (Armbruster et al., 2007; Urban and Roth, 2015). In DREADD, the designer receptor derived from muscarinic acetylcholine receptor is selectively activated by a designed chemical, clozapine-*N*-oxide, but not endogenous ligands. Similar approaches have been reported using other GPCRs such as the β2 adrenergic receptor (Strader et al., 1991), κ-opioid receptor (Coward et al., 1998; Redfern et al., 1999; Vardy et al., 2015), and free fatty acid receptor 2 (Hudson et al., 2012). Although these methods are powerful for cell-specific activation of the downstream signaling pathways of the targeted GPCRs, they are unsuitable for investigating the physiological roles of receptors of interest that are endogenously expressed in the target cells, because the original ligand-binding properties of the receptors are altered.

Incorporation of a metal-binding site is another potential strategy for chemogenetic activation of POIs. Because of their strictly-defined coordination geometry, incorporation of coordinating amino acid residues at appropriate positions enables metal-induced conformational changes for activation of POIs (Yu et al., 2014; Ghanbarpour et al., 2019). In addition to successful examples of targeted soluble proteins, functional switching of GPCRs by metal coordination was previously demonstrated by Schwartz and co-workers. They introduced histidine (His) or cysteine (Cys) mutations into GPCRs such as tachykinin NK-1 receptor (Elling et al., 1995) and β2 adrenergic receptor (Elling et al., 1999) to yield metal-sensitive GPCRs. However, the mutations were introduced at the ligand-binding site, which resulted in reduced affinity to the endogenous ligand in most cases.

To investigate the physiological roles of receptors of interest using chemogenetics, the target receptors should be mutated in a way that causes a specific response to the designed ligand without affecting the original receptor function. In this context, we have developed a chemogenetic method termed "direct activation *via* coordination-based chemogenetics (dA-CBC)" targeting metabotropic glutamate receptor 1 (mGlu1) by focusing on the structural changes upon glutamate binding (Kiyonaka et al., 2016; Kubota et al., 2019; Ojima et al., 2021). mGlu1, which belongs to class C GPCR, is composed of an extracellular ligand-binding domain called Venus Flytrap (VFT) domain, a cysteine rich domain (CRD), and a 7-transmembrane domain (7TMD) (**Figure 1A**). Glutamate binding to the VFT domain induces closure of the domain, and this signal is then transmitted to the 7TMD to cause receptor activation in mGlu1 (Kunishima et al., 2000; Tsuchiya et al., 2002). Inspired by the signal transduction induced by the structural changes, we introduced His mutations that cause metal-induced structural changes (**Figures 1A,B**). Through dA-CBC of mGlu1, His mutation (N264H) at the N264 position led to Pd-2,2′-bipyridine (Pd (bpy))-induced activation of the mGlu1 mutant (Ojima et al., 2021). In this case, Pd (bpy) coordinates endogenous H55 and N264H on the upper and lower lips of the VFT domain, respectively (**Figure 1B**). More importantly, the original ligand-binding properties were not affected in this method, which allowed us to prepare knock-in mice bearing the N264H mutation in the mGlu1 gene for chemogenetic regulation of endogenous mGlu1(N264H) (Ojima et al., 2021). dA-CBC has the potential for analyzing the physiological roles of mGlu1 endogenously expressed in the human brain. However, mGlu1 is expressed in various brain regions, such as the olfactory bulb, thalamus, hippocampus, and cerebellum (Lavreysen et al., 2004). Thus, orthogonal activation methods are necessary to simultaneously analyze the individual roles of mGlu1 in different brain regions.

In this paper, we identified two mGlu1 mutants that are selectively activated by Cu^{2+} or Zn^{2+} through structure-based design of the mutants and cell-based screening. Notably, these two mutants and the N264H mutant are activated in an orthogonal manner, allowing orthogonal activation of the three mGlu1 mutants by Pd (bpy), Cu^{2+}, and Zn^{2+}. Thus, these mGlu1 mutants could provide a potential platform for analyzing mGlu1 function in a cell-specific manner.

MATERIALS AND METHODS

Construction of Expression Vector of mGlu1 Mutants

Site directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis Kit (NEB) with pBluescript II SK (+) encoding rat mGlu1 by following the manufacture's instruction. The cDNA of mGlu1 was subcloned into pCAGGS vector (Niwa et al., 1991) or pCDM vector to obtain the expression vectors. The pCDM vector were prepared from pcDNA3.1 (+) (Invitrogen), in which neomycin cassette was excised using *Pvu*II.

Culture and Transfection of HEK293 Cells

HEK293 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 100 unit ml^{-1} penicillin and 100 μg ml^{-1} streptomycin and 10% FBS (Sigma) at 37°C in a humidified atmosphere of 95% air and 5% CO_2. HEK293 cells were transiently transfected with plasmids encoding WT mGlu 1, the mGlu1 mutants, or the control vector using Viafect (Promega) following the manufacture's instruction. The cells were co-transfected with pEGFP-F (Clontech), pDsRed monomer-F (Clontech), or iRFP-670 (kindly gifted from Prof. Verkhusha) as transfection markers. For the culture of transfected cells, DMEM supplemented with 10% dialyzed FBS (Sigma) was used to decrease the cytotoxicity. The medium was exchanged 4 h after the transfection, and the cells were used for the experiments after 24–48 h.

Fluorescent Ca^{2+} Imaging

The transfected HEK293 cells were seeded on glass coverslips (Matsunami) coated with poly-L-lysine solution (Sigma) and incubated for 4 h at 37°C in a humidified atmosphere of 95% air and 5% CO_2. The calcium indicator Fura-2 AM (Dojindo) was loaded to the cells at 5 μM for 20–30 min. The imaging experiment was carried out in HBS buffer (20 mM HEPES pH 7.4, 107 mM NaCl, 6 mM KCl, 1.2 mM $MgSO_4$, 2 mM $CaCl_2$, 11.5 mM glucose). The fluorescence images and the Fura-2 ratio

were measured using a fluorescence microscope (IX71, Olympus) equipped with a complementary metal-oxide semiconductor (CMOS) camera (ORCA-flash 4.0, Hamamatsu Photonics) under xenon-lamp illumination, and analyzed with a video imaging system (AQUACOSMOS, Hamamatsu Photonics) following the manufacture's instruction. In imaging experiments, three different HEK293 cells transfected with one of the mGlu1 mutants were co-cultured on a glass coverslip, and each mutant was visually distinguished by the transfection markers. These three different cells were assayed simultaneously. The Δratio value was defined as the difference between the maximum ratio value after adding the reagent (metal ion or complex, or glutamate) and the average ratio before adding the reagent. The Δratio was fitted with KaleidaGraph to calculate the EC_{50} value using the equation: a + (b-a)/(1+(x/c)ᵌd).

Confocal Live Cell Imaging of Cell-Surface mGlu1 in HEK293 Cells

The HEK293 cells transfected with plasmids encoding WT mGlu1, the mGlu1 mutants, or control vector together with iRFP-670 as a transfection marker were seeded on glass bottom dishes coated with poly-L-lysine solution (Sigma) and incubated for 24 h at 37°C in a humidified atmosphere of 95% air and 5% CO_2. After washing the cells with HBS buffer, 100 nM FITM-Cy3 in HBS was added to the dishes and incubated for 30 min at 16°C to suppress endocytosis. Confocal live imaging was performed using a confocal microscope (LSM900, Carl Zeiss) equipped with a 63×, numerical aperture (NA) = 1.4 oil-immersion objective. Fluorescence images were obtained by excitation at 561 or 640 nm derived from diode lasers. Fluorescence intensity from Cy3 on the cell surface was quantified from the line scans of iRFP-670-positive cells and calculated with subtraction of background.

Statistical Analysis

All data are expressed as mean ± s.e.m. We accumulated the data for each condition from at least three independent experiments. We evaluated statistical significance with Student's *t*-test or one-way ANOVA with Dunnet's test. A value of $p < 0.05$ was considered significant.

RESULTS

Design and Screening for Metal-Responsive mGlu1 Mutants

We previously reported that the conformational change of the VFT domain from an open to closed conformation by Pd (bpy) caused the selective activation of the mGlu1 (N264H) mutant. Notably, other metal ions (Cu^{2+}, Zn^{2+}, Ni^{2+}, Cd^{2+}, and Pd^{2+}) failed to activate this mGlu1(N264H) mutant. Thus, identification of new mGlu1 mutants that are selectively activated by metal ions other than Pd (bpy) would allow orthogonal cell-specific activation of mGlu1 by transfecting either the mGlu1(N264H) mutant or the new mGlu1 mutants. To achieve this goal, we first designed a few more potential coordination sites around the N264 residue, which hereafter we call Site I (**Figure 1B**). Based on the mechanisms employed by Cu^{2+}- or Zn^{2+}-binding proteins such as azurin (Tian et al., 2016) or ZntR (Reyes-Caballero et al., 2011), a Cys residue was introduced in place of the His residue at N264 site. We designed the H55/P58H/N264C and H55/P58H/S263H/N264C mutants because Cu^{2+} or Zn^{2+} forms tridentate or tetradentate coordination in the metalloproteins (**Figures 1B,C**). In addition, in the cases of Ni^{2+}- or Cd^{2+}-binding proteins, acidic amino acids are sometimes used for the coordination of those metal ions (Trakhanov et al., 1998; Stegmann et al., 2010). Thus, we also designed P58E/S263E/N264E mutant (**Figures 1B,C**).

In addition to Site I, we focused on different regions in the VFT domain, termed Site II and Site III, as candidate coordination sites. At these sites, the distance between the upper and lower lips drastically changes upon domain closure (**Figure 1B**, **Supplementary Figure S1**). We designed the Q73H/D324H/E325C and V405H/D322H/D324C mutants from Site II and Site III, respectively, for metal ion-induced receptor activation (**Figures 1B,C**). Note that the mutation residues selected here do not participate in glutamate binding (Kunishima et al., 2000), so we expected that the designed mutants would maintain the original activity of mGlu1. Collectively, we performed primary screening of a total of five mutants from Site I through Site III for obtaining metal-responsive mGlu1 mutants.

Fluorescent Ca^{2+} imaging was performed to check whether these mutants are activated by metal ions or complexes (**Supplementary Figure S2A**). Because mGlu1 is a Gq-coupled GPCR, its activation causes elevated intracellular Ca^{2+} concentrations ($[Ca^{2+}]_i$) *via* phospholipase C (PLC) activation. Each mutant was transiently transfected into HEK293 cells, and the $[Ca^{2+}]_i$ changes upon the addition of each metal-ion or complex were monitored by a fluorescent Ca^{2+} indicator, Fura-2. As control experiments, the vector control, wild-type (WT) mGlu1, and the previously reported Pd (bpy)-selective N264H mutant were also examined. As expected on the basis of our prior report, metal ions and metal complexes failed to activate WT mGlu1, whereas Pd (bpy) selectively activated the N264H mutant (**Figure 1C**). We subsequently examined the effects of the metal ions or complexes on the newly designed mGlu1 mutants. Focusing on the mGlu1 mutants from Site I, although the H55/P58H/N264C mutant was activated by Pd (bpy), none of the three newly designed mutants, including the H55/P58H/N264C mutant, were activated by other metal ions. (**Figure 1C**, **Supplementary Figure S2B**). Interestingly, the V405H/D322H/D324C mutant from Site III showed sensitivity to Ni^{2+}. However, this mutant was also activated by Pd (bpy).

In contrast to these mutants from Site I and III, the Q73H/D324H/E325C mutant from Site II was strongly activated by Cu^{2+} while also showing moderate sensitivity to Ni^{2+} (**Figure 1C**). Notably, the Q73H/D324H/E325C mutant was not activated by Pd (bpy), hence allowing orthogonal activation of this mutant and the N264H mutant by Cu^{2+} and Pd (bpy) respectively. The activation of the Q73H/D324H/E325C mutant by Cu^{2+} was further validated by the dose-dependent response of Cu^{2+} in the fluorescent Ca^{2+} imaging. As shown in **Figure 1D**, the EC_{50}

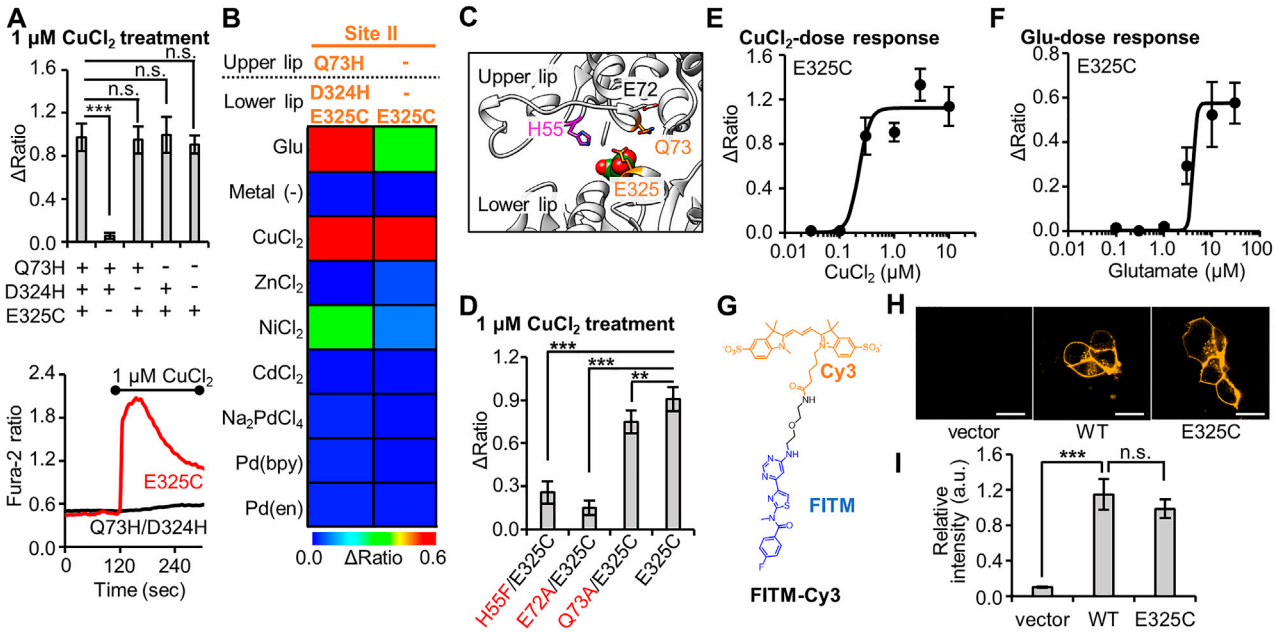


FIGURE 2 | Optimization of Cu^{2+}-responsive mGlu1 mutant and its characterization. **(A)** Mutation study to confirm the importance of His- and Cys-incorporation at the mutation sites. *Upper*; the average Δratio values for Q73H/D324H/E325C, Q73H/D324H, Q73H/E325C, D324H/E325C, or E325C mutant upon the treatment of 1 μM $CuCl_2$. (n = 12–20). *Lower*; representative traces from E325C or Q73H/D324H mutant. The black bar represents the period when $CuCl_2$ was added. **(B)** Heat map showing the metal selectivity of the newly identified Cu^{2+}-responsive E325C mutant. The map shows the average Δratio values upon the addition of 10 μM glutamate or 3 μM metal ion or complex. See **Supplementary Figure S3** for raw data. **(C)** Candidate residues for the coordination of Cu^{2+} with E325C in the VFT domain (PDB ID; 1EWK). **(D)** Mutation study to identify the coordination residues. The average Δratio values upon the treatment of 1 μM $CuCl_2$ are shown. (n = 12–20). **(E)** Activation of the mGlu1 E325C mutant by Cu^{2+}. The EC_{50} value for mGlu1 E325C mutant was 0.22 μM. (n = 12–20). **(F)** Activation of the mGlu1 E325C mutant by glutamate (Glu). The EC_{50} value of the mutant was 4.0 μM. (n = 12–20). **(G)** Chemical structure of FITM-Cy3. **(H)** Evaluation of the expression and distribution of mGlu1 mutant on the cell surface using FITM-Cy3. The representative confocal images are shown. The scale bar shows 20 μm. **(I)** Quantification of signal intensity from FITM-Cy3. (n = 20–31). Data are presented as mean ± s.e.m. ***Significant difference ($p < 0.001$, One-way ANOVA with Dunnet's test). **Significant difference ($p < 0.01$). n.s, not significant ($p > 0.05$).

value was 1.5 μM, while WT mGlu1 was not activated at that concentration of Cu^{2+} at all. Besides, the EC_{50} value of glutamate (1.4 μM) for the mutant slightly shifted to a higher concentration but was comparable to that for the WT mGlu1 (0.52 μM), indicating that this mutant preserved the original ligand-binding properties of mGlu1 (**Figure 1E**). Therefore, we selected the Q73H/D324H/E325C mutant as a Cu^{2+}-responsive mGlu1 mutant from the primary screening.

Optimization and Characterization of Cu^{2+}-Responsive mGlu1 Mutant

We proceeded to evaluate the Q73H/D324H/E325C mutant in further detail. To confirm the importance of the mutated His or Cys residues, we designed the Q73H/D324H, Q73H/E325C, and D324H/E325C mutants, where one of the three mutated residues from the Q73H/D324H/E325C mutant was substituted back with the original amino acid residue. As expected, fluorescent Ca^{2+} imaging revealed a loss in sensitivity to Cu^{2+} for the Q73H/D324H mutant (**Figure 2A**). However, unexpectedly, prominent Cu^{2+}-induced responses were observed in the case of the Q73H/E325C and D324H/E325C mutants. Since all Cu^{2+}-responsive mutants shared the same E325C mutation, we next evaluated the Cu^{2+}-induced response of the single mutant, mGlu1 (E325C). As shown in **Figure 2A**, the Cu^{2+}-induced response was maintained in the single mutant, suggesting that the E325C mutation alone is sufficient for the activation by Cu^{2+}. In addition, the E325C mutant showed a high selectivity to Cu^{2+} (**Figure 2B**, **Supplementary Figure S3**). Considering that the activation mechanism of mGlu1 requires the VFT domain closure, the coordination partner should be found in the upper lip of that domain. According to the crystal structure of the closed conformation of the VFT domain (PDB ID; 1EWK), endogenous residues such as H55, E72, and Q73 are candidate residues for Cu^{2+} coordination in collaboration with the E325C mutation (**Figure 2C**). To check this possibility, these candidate residues were substituted with phenylalanine or alanine, and the Cu^{2+}-induced responses of the H55F/E325C, E72A/E325C, and Q73A/E325C mutants were subsequently examined. As shown in **Figure 2D**, the sensitivity to Cu^{2+} was slightly affected for the Q73A/E325C mutant. In contrast, the Cu^{2+}-induced responses drastically decreased for the H55F/E325C and E72A/E325C mutants, suggesting that H55 and E72 could be potential coordination partners of E325C for Cu^{2+}-induced activation.

Next, we performed a functional characterization of the newly identified E325C mutant. The dose-dependency of Cu^{2+}-induced responses revealed that the EC_{50} value of Cu^{2+} for this mutant was 0.22 μM (**Figure 2E**), which is lower than that for the Q73H/

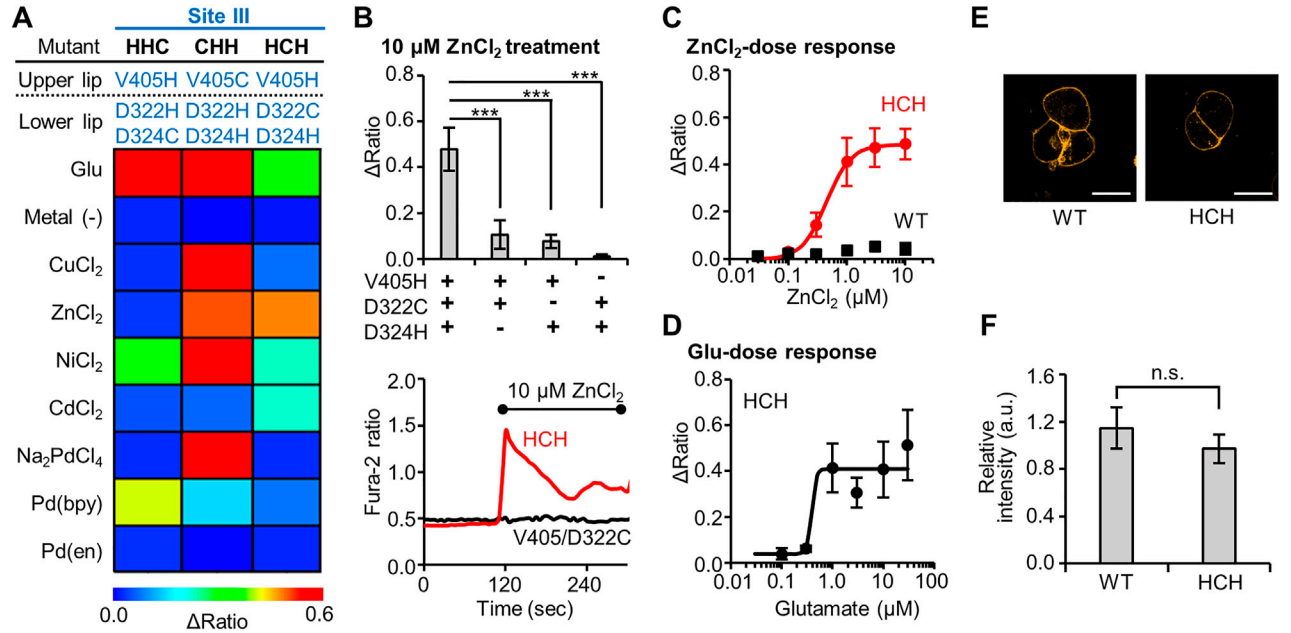


FIGURE 3 | Identification of Zn^{2+}-responsive mGlu1 mutant and its characterization. **(A)** Heat map of the secondary screening. The map shows the average Δratio values upon the addition of 10 μM glutamate or 3 μM metal ion or complex. See **Supplementary Figure S5** for raw data. **(B)** Mutation study to confirm the Zn^{2+}-coordination at the mutation sites. *Upper*; The Δratio values for the HCH, V405H/D322C, V405H/D324H, and D322C/D324H mutants. *Lower*; representative traces of the cells from two mutants. The black bar shows the period when $ZnCl_2$ was added. (n = 12–20). **(C)** Activation of WT mGlu1 (black) or the HCH mutant (red) by Zn^{2+}. The EC_{50} value for the mutant was 0.44 μM. (n = 11–20). **(D)** Activation of the HCH mutant by glutamate (Glu). The EC_{50} value of the mutant was 0.39 μM. (n = 11–20). **(E)** Evaluation of the expression and distribution of mGlu1 mutant on the cell surface using FITM-Cy3. The representative confocal images are shown. The scale bar shows 20 μm. (n = 50–70). **(F)** Quantification of signal intensity from FITM-Cy3. Data are presented as mean ± s.e.m. ***Significant difference ($p < 0.001$, One-way ANOVA with Dunnet's test). n.s, not significant ($p > 0.05$, Student's *t*-test.).

D324H/E325C mutant (**Figure 1D**). The EC_{50} value of glutamate to induce activation of the E325C mutant was 4.0 μM (**Figure 2F**). As was the case with the Q73H/D324H/E325C mutant, the EC_{50} value of glutamate for the E325C mutant slightly increased but was still on the same order of magnitude as that for the WT mGlu1. We also assessed the distribution and expression level of the E325C mutant under live-cell conditions. Here, we used a synthetic fluorescent probe named FITM-Cy3, which is a Cy3-conjugated FITM ligand (**Figure 2G**). FITM is a negative allosteric modulator (NAM) that binds to the 7TMD of mGlu1 (**Supplementary Figure S4**) (Wu et al., 2014), and FITM-Cy3 binds to cell-surface mGlu1 with high affinity (K_d = 6.8 nM) (Ojima et al., 2021). Thus, the distribution of the mGlu1 mutant can be estimated quantitatively using FITM-Cy3. As shown in **Figure 2H**, confocal live imaging of HEK293 cells transfected with either the WT mGlu1 or the E325C mutant revealed the clear localization of the probe on the cell surface, while the fluorescent signal was hardly detectable in vector-transfected cells. Importantly, the fluorescent intensity was not significantly different ($p > 0.05$) between the WT mGlu1 and the E325C mutant (**Figure 2I**), suggesting that the localization and expression level of the E325C mutant was unaffected by the mutation. In summary, through a precise analysis of the hit mutant obtained from the primary screening, we concluded the E325C mutant is a Cu^{2+}-responsive mGlu1 mutant.

Identification of Zn^{2+}-Responsive mGlu1 Mutant

In metal-binding proteins, the coordination geometry has essential roles for the function or metal-selectivity. For instance, the metal ion selectivity of azurin, a Cu^{2+}-binding protein, is affected by the geometry of the coordinating residues (Nar et al., 1992). As described above (see **Figure 1C**), we revealed that the V405H/D322H/D324C mutant from Site III was activated by both Ni^{2+} and Pd (bpy) in the primary screening. As a result, this mutant cannot be applied for orthogonal activation in conjunction with the Pd (bpy)-responsive mutant. Based on examples of prior successes such as that of azurin, we proceeded to perform secondary screening on the V405H/D323H/D324C (HHC) mutant by changing the geometry of the mutated residues.

In the secondary screening, we designed two more mutants, V405C/D322H/D324H (CHH) and V405H/D322C/D324H (HCH), where we altered the position of the Cys residue without changing the mutation sites themselves. Similar to the characterization of the primary screening, metal-induced cellular responses were again evaluated using fluorescent Ca^{2+} imaging. As illustrated in the heat map in **Figure 3A**, the two newly designed mutants (CHH and HCH) showed drastically different metal-selectivity, compared with the original HHC mutant. The CHH mutant showed a broad selectivity since it was strongly

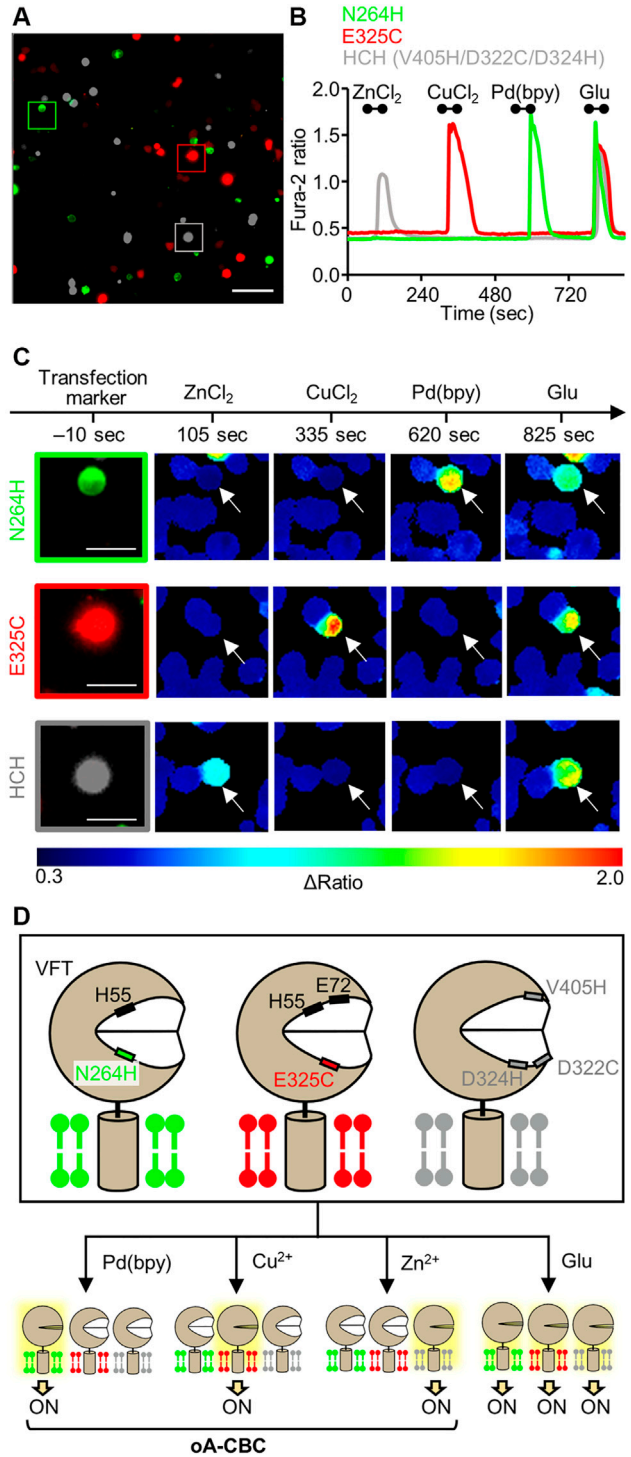


FIGURE 4 | Orthogonal activation of three mGlu1 mutants with Pd(bpy), Cu^{2+}, and Zn^{2+}. **(A)** Merged fluorescent images of HEK293 cells transfected with N264H (green), E325C (red), or HCH mutant (gray). The scale bar shows 100 μm. **(B)** Orthogonal activation of the N264H, E325C, HCH mutant by 3 μM Pd (bpy), 1 μM $CuCl_2$, or 3 μM $ZnCl_2$, respectively. 3 μM glutamate (Glu) was added after the metal ions or complex. The traces of Δratio values as a function of time for the representative cells marked in (A) are shown. The black bars inside the figure show the period when each reagent was

(*Continued*)

FIGURE 4 | added. **(C)** The transfection markers and Δratio images of the cells expressing each mutant. The white arrows show the representative cells marked in (A). The scale bars show 25 μm. **(D)** Schematic illustration of oA-CBC system. Mutations in VFT domain produced three types of mutants orthogonally sensitized to Pd (bpy), Cu^{2+}, and Zn^{2+}. These mutants allow the cell-specific activation by Pd (bpy), Cu^{2+}, and Zn^{2+}, while all of three mutants are activated by Glu, because they keep the original ligand binding property.

activated by several metal ions such as Cu^{2+}, Zn^{2+}, Ni^{2+}, and Pd^{2+}. In contrast, the HCH mutant was activated strongly by Zn^{2+}, moderately by Ni^{2+} or Cd^{2+}, and not activated at all by either Cu^{2+} or Pd (bpy) (**Figure 3A**, **Supplementary Figure S5**). Therefore, the HCH mutant has the potential to be another orthogonal mutant for use along with the Pd (bpy)-responsive N264H mutant and the Cu^{2+}-responsive E325C mutant.

Because Zn^{2+} strongly activated this HCH mutant, the Zn^{2+}-coordination was further analyzed. First, we evaluated whether all three mutations were necessary for the Zn^{2+}-coordination. For this purpose, three mutants, V405H/D322C, V405H/D324H, and D322C/D324H were designed, where each of the three mutated residues in the HCH mutant was individually changed back to the original residue. As shown in **Figure 3B**, the Zn^{2+}-induced responses were significantly impaired in the three mutants, suggesting that all three mutations were required for the Zn^{2+}-induced activation of the mGlu1 mutant.

We then further studied the Zn^{2+}-response of this mutant in detail. The dose-dependent response of Zn^{2+} determined the EC_{50} value to be 0.44 μM for the mutant, while the WT mGlu1 was not activated at all at this Zn^{2+} concentration (**Figure 3C**). To verify whether the original ligand-binding property of mGlu1 is preserved in this mutant, the dose-dependent glutamate response was also measured. As shown in **Figure 3D**, the EC_{50} value of glutamate for the activation of this mutant was 0.39 μM, which was comparable to that of the WT mGlu1 (0.52 μM, **Figure 1E**). In addition, confocal live imaging using FITM-Cy3 revealed that the distribution and the expression level of the mGlu1 mutant were not affected by the introduction of these mutations (**Figures 3E,F**). In summary, our secondary screening identified the HCH mutant as a Zn^{2+}-responsive mGlu1 mutant which was insensitive to Pd (bpy) and Cu^{2+}.

Cell-specific Activation Utilizing Mutually Orthogonal mGlu1 Mutants

Through the primary and secondary screening, we identified two more mutants in addition to the Pd (bpy)-responsive N264H mutant: the Cu^{2+}-responsive E325C and Zn^{2+}-responsive HCH mutants. Notably, these three mutants have sufficient metal-selectivity among Cu^{2+}, Zn^{2+}, and Pd (bpy). Namely, N264H was not activated by Cu^{2+} or Zn^{2+}, E325C was not activated by Zn^{2+} or Pd (bpy), and HCH was not activated by Cu^{2+} or Pd (bpy) (**Supplementary Figure S6**). Besides, the metal-induced activations were repeatedly observed after washing out the bound metals for 2 min, suggesting a minutes-order reversibility of the activation (**Supplementary Figure S7**). However, the activation of the E325C mutant by Cu^{2+} or the

HCH mutant by Zn^{2+} was impaired in the co-presence of Pd (bpy) or Cu^{2+}, respectively (**Supplementary Figure S8**).

Considering these properties of the three mutants, we demonstrated cell-specific mGlu1 activation. Here, we prepared a model experiment for the cell-specific regulation of mGlu1, where HEK293 cells were transfected with a plasmid encoding either the N264H, E325C, or HCH mutant. In this experiment, we regarded each of these transfected cells as different cell types. To visually distinguish these cells, the transfection markers EGFP, DsRed-monomer, and iRFP-670 were co-expressed with the N264H, E325C, and HCH mutant, respectively. After 48 h of transfection, these three types of HEK293 cells were mixed and seeded on a single glass coverslip (**Figure 4A**). Given the impairment on the metal-induced activation of the E325C or HCH mutant in the co-presence of Pd (bpy) or Cu^{2+}, respectively, $ZnCl_2$, $CuCl_2$, and Pd (bpy) were sequentially applied for 60 s onto the glass coverslip *via* perfusion in this order. As shown in **Figures 4B,C**, the fluorescent Ca^{2+} imaging revealed that each metal ion or complex selectively activated the cells that expressed its corresponding mutant. Notably, the glutamate-induced responses were intact after metal-induced activation, suggesting that mGlu1 activity was not affected after metal-induced activation. These results indicate that the HEK293 cells that expressed the three different types of mGlu1 mutants were selectively activated by either Zn^{2+}, Cu^{2+}, or Pd (bpy), thus demonstrating an orthogonal trio for mGlu1 activation. We termed this method orthogonal activation *via* coordination-based chemogenetics (oA-CBC) of mGlu1 (**Figure 4D**).

DISCUSSION

In this study, three mutually orthogonal mutants for the dA-CBC system were developed. Specifically, a N264H mutant sensitized to Pd (bpy), a E325C mutant sensitized to Cu^{2+}, and a HCH mutant sensitized to Zn^{2+} were identified and characterized. Fluorescent Ca^{2+} imaging using HEK293 cells demonstrated that three kinds of metal ions or complexes orthogonally activated the cells expressing a corresponding mGlu1 mutant. This system was termed as the oA-CBC system. Although our current study was limited to demonstrating this orthogonal activation in HEK293 cells, the oA-CBC system has the potential to analyze mGlu1 function in several different human brain regions simultaneously using tissue-selective promoters in the brain, given that mGlu1 is expressed in the olfactory bulb, thalamus, hippocampus, and cerebellum in the brain (Lavreysen et al., 2004).

The mGlu family comprises eight subtypes (mGlu1–8) and is classified into three groups (group I–III) based on the amino-acid sequence and transducing signal properties of each subtype (Niswender and Conn, 2010). mGlu1 and mGlu5 are both "group I" because of their closely homologous sequences that couple to Gq-proteins. Although both are predominantly expressed in the brain, their distribution is different. Unlike mGlu1, mGlu5 is expressed in the caudate-putamen, lateral septum, and cortex in the brain (Romano et al., 1995). Although both subtypes are expressed in the hippocampus, previous studies revealed that the functional contributions of each subtype are different in the region (Neyman and Manahan-Vaughan, 2008). Group I metabotropic glutamate receptor agonists such as 3,5-dihydroxyphenylglycine (DHPG) are frequently used to analyze the physiological roles of these receptors (Fitzjohn et al., 1999). However, this agonist cannot discriminate mGlu1 from mGlu5 because of the high sequence homology. Thus, it is challenging to independently activate and analyze mGlu1 or mGlu5 expressed in the same region of the brain. Therefore, potential future work could be to apply the oA-CBC method to mGlu5, thus creating an orthogonal pair of mGlu1 and mGlu5 mutants. The mutants could be activated by two different metal ions or complexes and utilized for the subtype-selective activation of mGlu proteins.

Another future application of our oA-CBC method is to construct an artificial signal-transducing system using three different metal ions or complexes as its input. Since mGlu1 is coupled to Gq-protein, the activation of mGlu1 by a metal ion or complex leads to the elevation of $[Ca^{2+}]_i$ *via* PLC activation. However, the other mGlu proteins in groups II or III coupled with Gi/o-protein decreases the intracellular cAMP level by inhibiting adenylyl cyclase activity (Niswender and Conn, 2010). The oA-CBC method can also be applied to other types of mGlu proteins to artificially hijack different types of G-protein signals in an orthogonal way.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SK and IH conceived the study. AS, YY, and KO performed the construction of mGlu1 plasmids. AS. and YY performed fluorescent Ca^{2+} imaging in HEK293 cells. AS and TD performed confocal imaging of HEK293 cells using FITM-Cy3. AS and SK wrote the manuscript, and all authors discussed and commented on the manuscript.

FUNDING

This work was funded by Grants-in-Aid for Scientific Research (KAKENHI) (Grant Number 18J22952 to KO, 17H06348 to IH, 19H05778, and 20H02877 to SK.), Daiichi Sankyo Foundation of Life Science (to SK.), the Takeda Science Foundation (to SK.), the Mochida Memorial Foundation for Medical and Pharmaceutical Research (to SK.), and supported by JST ERATO (JPMJER 1802) to IH. This work was partially supported by Nagoya University Research Fund.

ACKNOWLEDGMENTS

The authors thank Haocheng Qianzhu and Kanta Hasegawa for technical supports, and Winnie Shia from Edanz (https://jp.

edanz.com/ac) for editing a draft of this manuscript. The expression plasmid for iRFP-670 was generously gifted by Vladislav Verkhusha, and the pCAGGS vector was provided by the RIKEN BRC through the National Bio-Resource Project of the MEXT, Japan.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.825669/full#supplementary-material

REFERENCES

Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S., and Roth, B. L. (2007). Evolving the Lock to Fit the Key to Create a Family of G Protein-Coupled Receptors Potently Activated by an Inert Ligand. *Proc. Natl. Acad. Sci.* 104, 5163–5168. doi:10.1073/pnas.0700293104

Atasoy, D., and Sternson, S. M. (2018). Chemogenetic Tools for Causal Cellular and Neuronal Biology. *Physiol. Rev.* 98, 391–418. doi:10.1152/physrev.00009.2017

Bishop, A. C., Ubersax, J. A., Petsch, D. T., Matheos, D. P., Gray, N. S., Blethrow, J., et al. (2000). A Chemical Switch for Inhibitor-Sensitive Alleles of Any Protein Kinase. *Nature* 407, 395–401. doi:10.1038/35030148

Coward, P., Wada, H. G., Falk, M. S., Chan, S. D. H., Meng, F., Akil, H., et al. (1998). Controlling Signaling with a Specifically Designed Gi-Coupled Receptor. *Proc. Natl. Acad. Sci.* 95, 352–357. doi:10.1073/pnas.95.1.352

Elling, C. E., Nielsen, S. M., and Schwartz, T. W. (1995). Conversion of Antagonist-Binding Site to Metal-Ion Site in the Tachykinin NK-1 Receptor. *Nature* 374, 74–77. doi:10.1038/374074a0

Elling, C. E., Thirstrup, K., Holst, B., and Schwartz, T. W. (1999). Conversion of Agonist Site to Metal-Ion Chelator Site in the Beta 2-Adrenergic Receptor. *Proc. Natl. Acad. Sci.* 96, 12322–12327. doi:10.1073/pnas.96.22.12322

Fitzjohn, S. M., Kingston, A. E., Lodge, D., and Collingridge, G. L. (1999). DHPG-induced LTD in Area CA1 of Juvenile Rat hippocampus; Characterisation and Sensitivity to Novel mGlu Receptor Antagonists. *Neuropharmacology* 38, 1577–1583. doi:10.1016/S0028-3908(99)00123-9

Ghanbarpour, A., Pinger, C., Esmatpour Salmani, R., Assar, Z., Santos, E. M., Nosrati, M., et al. (2019). Engineering the hCRBPII Domain-Swapped Dimer into a New Class of Protein Switches. *J. Am. Chem. Soc.* 141, 17125–17132. doi:10.1021/jacs.9b04664

Hudson, B. D., Christiansen, E., Tikhonova, I. G., Grundmann, M., Kostenis, E., Adams, D. R., et al. (2012). Chemically Engineering Ligand Selectivity at the Free Fatty Acid Receptor 2 Based on Pharmacological Variation between Species Orthologs. *FASEB j.* 26, 4951–4965. doi:10.1096/fj.12-213314

Islam, K. (2015). Allele-Specific Chemical Genetics: Concept, Strategies, and Applications. *ACS Chem. Biol.* 10, 343–363. doi:10.1021/cb500651d

Kiyonaka, S., Kubota, R., Michibata, Y., Sakakura, M., Takahashi, H., Numata, T., et al. (2016). Allosteric Activation of Membrane-Bound Glutamate Receptors Using Coordination Chemistry within Living Cells. *Nat. Chem.* 8, 958–967. doi:10.1038/nchem.2554

Knight, Z. A., and Shokat, K. M. (2007). Chemical Genetics: Where Genetics and Pharmacology Meet. *Cell* 128, 425–430. doi:10.1016/j.cell.2007.01.021

Kubota, R., Kiyonaka, S., and Hamachi, I. (2019). On-Cell Coordination Chemistry: Chemogenetic Activation of Membrane-Bound Glutamate Receptors in Living Cells. *Methods Enzymol.* 622, 411–430. doi:10.1016/bs.mie.2019.02.033

Kunishima, N., Shimada, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., et al. (2000). Structural Basis of Glutamate Recognition by a Dimeric Metabotropic Glutamate Receptor. *Nature* 407, 971–977. doi:10.1038/35039564

Lavreysen, H., Pereira, S. N., Leysen, J. E., Langlois, X., and Lesage, A. S. J. (2004). Metabotropic Glutamate 1 Receptor Distribution and Occupancy in the Rat Brain: A Quantitative Autoradiographic Study Using [3H]R214127. *Neuropharmacology* 46, 609–619. doi:10.1016/j.neuropharm.2003.11.014

Mondoloni, S., Durand-de Cuttoli, R., and Mourot, A. (2019). Cell-Specific Neuropharmacology. *Trends Pharmacol. Sci.* 40, 696–710. doi:10.1016/j.tips.2019.07.007

Nar, H., Huber, R., Messerschmidt, A., Filippou, A. C., Barth, M., Jaquinod, M., et al. (1992). Characterization and crystal Structure of Zinc Azurin, a By-Product of Heterologous Expression in *Escherichia coli* of *Pseudomonas aeruginosa* Copper Azurin. *Eur. J. Biochem.* 205, 1123–1129. doi:10.1111/j.1432-1033.1992.tb16881.x

Neyman, S., and Manahan-Vaughan, D. (2008). Metabotropic Glutamate Receptor 1 (mGluR1) and 5 (mGluR5) Regulate Late Phases of LTP and LTD in the Hippocampal CA1 Region *In Vitro*. *Eur. J. Neurosci.* 27, 1345–1352. doi:10.1111/j.1460-9568.2008.06109.x

Niswender, C. M., and Conn, P. J. (2010). Metabotropic Glutamate Receptors: Physiology, Pharmacology, and Disease. *Annu. Rev. Pharmacol. Toxicol.* 50, 295–322. doi:10.1146/annurev.pharmtox.011008.145533

Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient Selection for High-Expression Transfectants with a Novel Eukaryotic Vector. *Gene* 108, 193–199. doi:10.1016/0378-1119(91)90434-D

Ojima, K., Kakegawa, W., Ito, M., Miura, Y., Michibata, Y., Kubota, R., et al. (2021). Coordination Chemogenetics for Activation of GPCR-type Glutamate Receptors in Brain Tissue. *bioRxiv*. Available at:. doi:10.1101/2021.10.01.462737

Pratt, M. R., Schwartz, E. C., and Muir, T. W. (2007). Small-molecule-mediated rescue of Protein Function by an Inducible Proteolytic Shunt. *Proc. Natl. Acad. Sci.* 104, 11209–11214. doi:10.1073/pnas.0700816104

Redfern, C. H., Coward, P., Degtyarev, M. Y., Lee, E. K., Kwa, A. T., Hennighausen, L., et al. (1999). Conditional Expression and Signaling of a Specifically Designed Gi-Coupled Receptor in Transgenic Mice. *Nat. Biotechnol.* 17, 165–169. doi:10.1038/6165

Reyes-Caballero, H., Campanello, G. C., and Giedroc, D. P. (2011). Metalloregulatory Proteins: Metal Selectivity and Allosteric Switching. *Biophys. Chem.* 156, 103–114. doi:10.1016/j.bpc.2011.03.010

Romano, C., Sesma, M. A., McDonald, C. T., O'malley, K., van den Pol, A. N., and Olney, J. W. (1995). Distribution of Metabotropic Glutamate Receptor mGluR5 Immunoreactivity in Rat Brain. *J. Comp. Neurol.* 355, 455–469. doi:10.1002/cne.903550310

Schreiber, S. L. (1998). Chemical Genetics Resulting from a Passion for Synthetic Organic Chemistry. *Bioorg. Med. Chem.* 6, 1127–1152. doi:10.1016/S0968-0896(98)00126-6

Stegmann, C. M., Lührmann, R., and Wahl, M. C. (2010). The crystal Structure of PPIL1 Bound to Cyclosporine a Suggests a Binding Mode for a Linear Epitope of the SKIP Protein. *PLoS One* 5, e10013. doi:10.1371/journal.pone.0010013

Strader, C. D., Gaffney, T., Sugg, E. E., Candelore, M. R., Keys, R., Patchett, A. A., et al. (1991). Allele-Specific Activation of Genetically Engineered Receptors. *J. Biol. Chem.* 266, 5–8. doi:10.1016/s0021-9258(18)52392-9

Tian, S., Liu, J., Cowley, R. E., Hosseinzadeh, P., Marshall, N. M., Yu, Y., et al. (2016). Reversible S-Nitrosylation in an Engineered Azurin. *Nat. Chem.* 8, 670–677. doi:10.1038/nchem.2489

Trakhanov, S., Kreimer, D. I., Parkin, S., Ames, G. F. L., and Rupp, B. (1998). Cadmium-induced Crystallization of Proteins: II. Crystallization of the Salmonella Typhimuri Histidine-Binding Protein in Complex with L-histidine, L-arginine, or L-lysine. *Protein Sci.* 7, 600–604. doi:10.1002/pro.5560070308

Tsai, Y.-H., Doura, T., and Kiyonaka, S. (2021). Tethering-based Chemogenetic Approaches for the Modulation of Protein Function in Live Cells. *Chem. Soc. Rev.* 50, 7909–7923. doi:10.1039/d1cs00059d

Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H., and Morikawa, K. (2002). Structural Views of the Ligand-Binding Cores of a Metabotropic Glutamate Receptor Complexed with an Antagonist and Both Glutamate and Gd3+. *Proc. Natl. Acad. Sci.* 99, 2660–2665. doi:10.1073/pnas.052708599

Urban, D. J., and Roth, B. L. (2015). DREADDs (Designer Receptors Exclusively Activated by Designer Drugs): Chemogenetic Tools with Therapeutic Utility. *Annu. Rev. Pharmacol. Toxicol.* 55, 399–417. doi:10.1146/annurev-pharmtox-010814-124803

Vardy, E., Robinson, J. E., Li, C., Olsen, R. H. J., DiBerto, J. F., Giguere, P. M., et al. (2015). A New DREADD Facilitates the Multiplexed Chemogenetic

Interrogation of Behavior. *Neuron* 86, 936–946. doi:10.1016/j.neuron.2015.03.065

Vassilatis, D. K., Hohmann, J. G., Zeng, H., Li, F., Ranchalis, J. E., Mortrud, M. T., et al. (2003). The G Protein-Coupled Receptor Repertoires of Human and Mouse. *Proc. Natl. Acad. Sci.* 100, 4903–4908. doi:10.1073/pnas.0230374100

Wu, H., Wang, C., Gregory, K. J., Han, G. W., Cho, H. P., Xia, Y., et al. (2014). Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator. *Science* 344, 58–64. doi:10.1126/science.1249489

Yu, F., Cangelosi, V. M., Zastrow, M. L., Tegoni, M., Plegaria, J. S., Tebo, A. G., et al. (2014). Protein Design: Toward Functional Metalloenzymes. *Chem. Rev.* 114, 3495–3578. doi:10.1021/cr400458x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past co-authorship with the authors TD, SK.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fchem.2021.799706

Equipping Coiled-Coil Peptide Dimers With Furan Warheads Reveals Novel Cross-Link Partners

Laia Miret-Casals[1], Sander Van De Putte[1], Dorien Aerssens[1], Julien Diharce[2], Pascal Bonnet[2] and Annemieke Madder[1]**

[1]Department of Organic and Macromolecular Chemistry, Organic and Biomimetic Chemistry Research Group, Faculty of Science, Ghent University, Ghent, Belgium, [2]Institut de Chimie Organique et Analytique, Université d'Orléans, UMR CNRS 7311, Orléans, France

Using a coiled-coil peptide dimer as a model system to explore furan reactivity, we describe novel cross-link partners of furan warheads for site-specific cross-linking. We demonstrate that replacement of weak interhelical ionic contacts with a furan moiety and its potential cross-link partner affords covalently connected coiled-coil motifs upon furan activation. We describe for the first time the reaction of the activated furan warhead with cysteine and tyrosine, besides the previously reported lysine, thus enhancing the versatility of the furan cross-link methodology by the possibility to target different amino acids. The present *in vitro* validation of "furan-armed" α-helices provides further grounds for exploiting furan technology in the development of furan-modified ligands/proteins to target proteins in a covalent way through various amino acid side chains.

Keywords: furan oxidation, singlet oxygen, molecular modeling, cross-link (CL), coiled-coil peptide, protein–protein interaction (PPI), peptide–protein interaction

OPEN ACCESS

Edited by:
Yu-Hsuan Tsai,
Shenzhen Bay Laboratory, China

Reviewed by:
Aimee Boyle,
Leiden University, Netherlands
Douglas Root,
University of North Texas, United States
Christopher Wood,
University of Edinburgh, United Kingdom

****Correspondence:***
Laia Miret-Casals
Laia.MiretCasals@UGent.be
Annemieke Madder
Annemieke.Madder@UGent.be

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *21 October 2021*
Accepted: *27 December 2021*
Published: *16 February 2022*

Citation:
Miret-Casals L, Van De Putte S, Aerssens D, Diharce J, Bonnet P and Madder A (2022) Equipping Coiled-Coil Peptide Dimers With Furan Warheads Reveals Novel Cross-Link Partners. Front. Chem. 9:799706. doi: 10.3389/fchem.2021.799706

INTRODUCTION

Peptide–protein and protein–protein interactions play key roles in biological processes. Therefore, approaches to target proteins in a covalent way can provide useful analytical tools for chemical biology and could lead to potential applications in therapeutic development. We recently reported on a novel cross-link technology applied to peptide–protein (Vannecke et al., 2017) and protein–protein (Miret-Casals et al., 2021a) interactions, based on the introduction of an unnatural, furan-containing amino acid in a peptide or protein and its subsequent oxidation into an electrophilic moiety upon the action of reactive oxygen species (ROS) or visible light irradiation in the presence of a photosensitizer (PS, **Figures 1A,B**, respectively) (Vannecke et al., 2017; Miret-Casals et al., 2021a). Initially, a furan modified kisspeptin-10 peptide ligand could be covalently cross-linked to its membrane receptor, GPR54, on live cancer cells with spontaneous oxidation of the furan moiety by endogenous ROS (**Figure 1A**) (Vannecke et al., 2017). The GPR54 (also called KISS1R) is a well-known modulator of the physiology of reproduction in mammals (Pinilla et al., 2012). GPR54 binds kisspeptins, a family of truncated forms of kisspeptin peptides (54, 14, 13, or 10 amino acids) with a common Arg-Phe-amide C terminus responsible for the high-affinity binding and activation of GPR54 (Kotani et al., 2001). We chemically synthesized kisspeptin-10 (KP-10) replacing the Trp3 residue with a furylalanine (Fua) amino acid. We could demonstrate that upon incubation of live cells expressing the GPR54 receptor with the furan-modified peptide, a covalently cross-linked kisspeptin-10/GPR54 complex was formed as demonstrated by extensive analysis of the cell lysate by Western blotting (Vannecke et al., 2017). Due to the complexity of the system

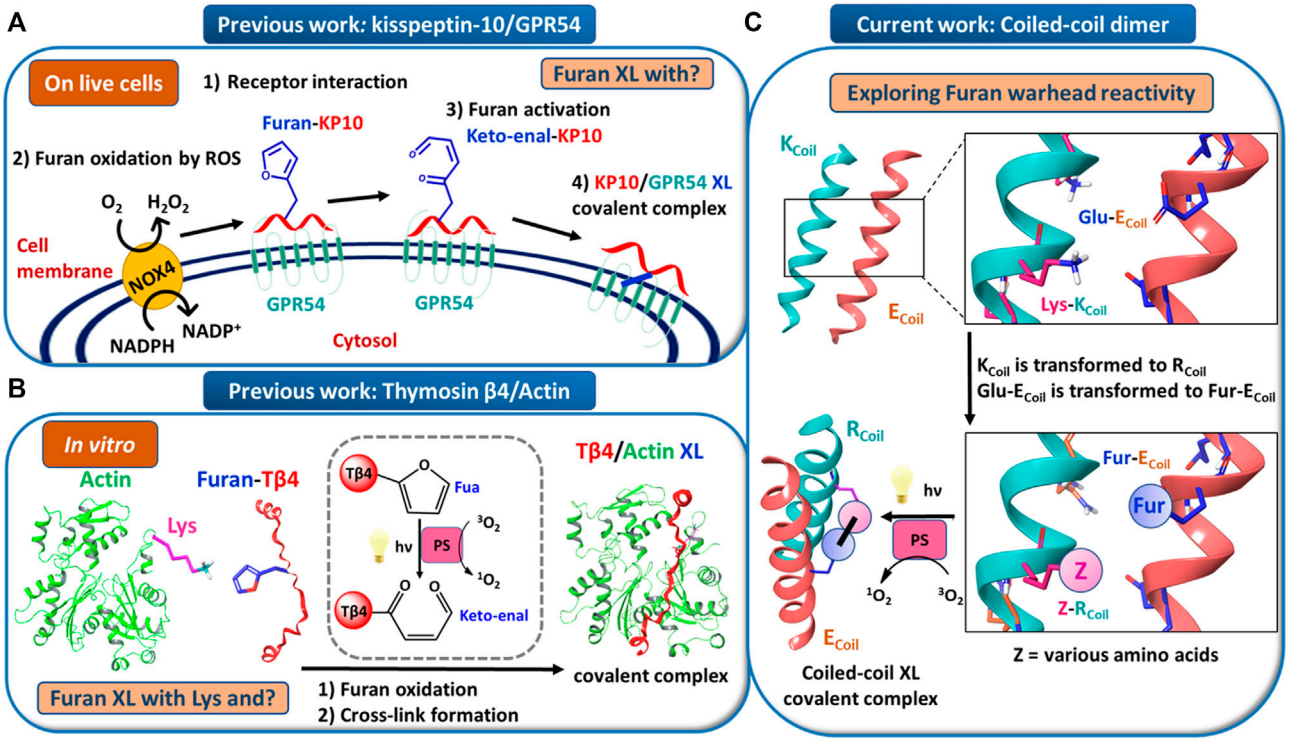


FIGURE 1 | (A) Model for endogenous furan oxidation by reactive oxygen species (ROS). The mature glycosylated GPR54 receptor is present at the cell surface. When the furan-modified kisspeptin-10 peptide (Furan-KP10) is added, it interacts at the cell surface with GPR54 (step 1). Due to the presence of an NADPH oxidase (likely NOX4), present in the plasma membrane, ROS are locally released in the extracellular environment. Then, the Furan-KP10 is oxidized by ROS (step 2) to a keto-enal-KP10 (step 3) and forms a covalent cross-link with the receptor (step 4). **(B)** Model for *in vitro* furan oxidation by singlet oxygen production. Actin and Fua-Tβ4 were first pre-incubated. Then, the furan moiety was activated by oxidation followed by cross-link formation. The oxidation relies on singlet oxygen generation by visible light irradiation in the presence of a photosensitizer (PS), such as Rose Bengal or Rhodamine B. **(C)** Model system to explore furan warhead reactivity towards different amino acid side chains in a coiled-coil peptide heterodimer.

combined with the low abundance of the GPR54 membrane receptor, we were not able to determine the identity of the attacking amino acid residue and the exact structure of the cross-linked species, despite extensive attempts at receptor pulldown and MS-based identification of the complex. In order to shed more light on the exact nature of the nucleophile involved in the covalent bond formation, and since no crystal structure of GPR54 is available, a homology model was built (see ESI 1 and **Supplementary Figure S1**). Three GPCRs were used as templates: the nociceptin opioid receptor, the delta-opioid receptor, and the kappa-opioid receptor. The homology model of GPR54 was built using Modeller software (Webb and Sali, 2016) and 10 models were generated. Molecular dynamics were applied with Amber 16 (Case, 2016) to relax the generated structure and one representative structure was selected to propose two starting points for the prediction of the GPR54/kisspeptin-10 interaction: the rigid and the relaxed state of GPR54. Then, the new *in silico* Iterative Residue Docking and Linking method (Diharce et al., 2019), which allows to efficiently predict peptide–protein interactions, was used to dock kisspeptin-10. The best predicted docking poses of the native kisspeptin-10 ligand into the GPR54 model receptor structure using the rigid (**Figure 2A**; **Supplementary Figure S2A** in ESI 1) and the relaxed (**Supplementary Figure S2B** in ESI 1) conformation of the receptor show that several GPR54-tyrosine (Y) residues, as well as GPR54-tryptophan (W), -cysteine (C), and -lysine (K) are located in sufficient proximity of the KP-10-W3 residue and could cross-link when W3 is replaced by Fua. In a more recent work on the further use of the furan-oxidation strategy for protein–protein cross-linking, we studied the interaction between actin, the major cytoskeletal protein of the cell that forms filaments, and Tβ4, which regulates the polymerization of actin and keeps it in the monomeric form. Starting from the available 3D structure–function information on the Tβ4–actin complex (PDB 4PL7, **Figure 2B**), we determined the optimal position to insert a furan in Tβ4 (Tβ4-E24), and several Actin-lysine (K) residues were identified as potential nucleophiles, proximate enough to engage in covalent bond formation (Miret-Casals et al., 2021a). In addition, several Actin-tyrosine (Y) residues were also observed in close proximity to Tβ4-E24 (**Figure 2B**). The furan-modified Tβ4 analogue was found able to efficiently cross-link to monomeric actin upon singlet oxygen generation by irradiation in the presence of a PS (**Figure 1B**), and the cross-link site in the Tβ4-actin covalent complex was characterized in detail (Miret-Casals et al., 2021a). In this case, we could firmly establish

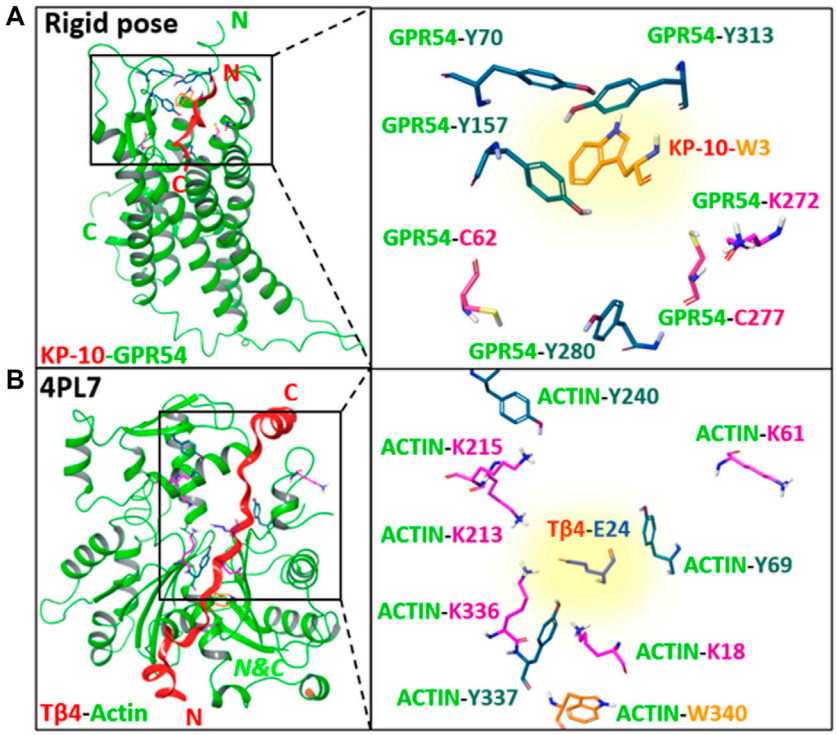


FIGURE 2 | (A) The best predicted docking pose of the native kisspeptin-10 ligand into the GPR54 model receptor structure using the rigid pose conformation of the receptor. **(B)** 3D structure of Tβ4-actin complex based on PDB 4PL7. Left panels: an overview of the complex system; right panels: zoom in on GPR54/ACTIN amino acids side chains (Lys, Cys, Tyr, and Trp) proximate to KP-10-**W3** or Tβ4-**E24**, respectively, without the secondary structure. The amino acid that is replaced by furylalanine (W3 in kisspeptin, E24 in Thymosin β4) is highlighted with a yellow background. N, C = N-, C-termini. Images prepared using MAESTRO (for details, see ESI 1).

that lysine is a target for the furan warhead, and a potential chemical structure of the covalent adduct between lysine and the activated furan was put forward based on the MS data (Miret-Casals et al., 2021a). Even if these cross-linked residues were identified by MS, it cannot be excluded that other amino acids in Actin can also react but the cross-linked products resulting thereof were not detected due to a lack of knowledge on all potential furan-cross-link partners. In both studies, the furan moiety is used as a caged warhead and can be triggered to a keto-enal (electrophilic intermediate) by the production of singlet oxygen (**Figures 1A,B**). Earlier studies on the toxicity of furan derivatives resulting from furan metabolism *in vivo* describe both amine and thiol species as suitable target nucleophiles for activated furan (Peterson, 2013).

In order to further explore the scope and limitations of furan-oxidation-based cross-linking to proteins, we aimed at determining which amino acid other than lysine can be potentially targeted by a furan warhead and lead to a stable cross-linked adduct. For that purpose, we searched for a system able to mimic protein–protein interactions and suited for rational protein design. Coiled-coil systems are described by the interaction between two or more α-helices that coil around each other to form a supercoiled structure composed of identical or different helices, resulting in homomers or heteromers that can be arranged in a parallel or antiparallel manner (Lupas, 1996; Litowski and Hodges, 2002; Mason and Arndt, 2004; Liu et al., 2006; Grigoryan and Keating, 2008; Boyle and Woolfson, 2011; Woolfson et al., 2017; Lapenta et al., 2018).

Coiled-coil helical systems represent ideal scaffolds to bring potential peptide cross-link partners in close proximity and we decided to evaluate the propensity for site-specific cross-linking between two coils, one of which is equipped with a furan moiety (**Figure 1C**). In this work, the well-known heterodimeric E3/K3 coiled-coil (Litowski and Hodges, 2002) was used as a model system (**Figure 1C**), which is composed of a total of 42 residues (**Figure 3A**), comprising two complementary right-handed α-helices wrapped around each other in a left-handed fashion (**Figure 3B** and the helical wheel diagram displayed in **Figure 3C**). Each coil peptide contains three heptad repeats (3 × *gabcdef*) of seven amino acids with hydrophobic contacts at the a and d positions, ionic residues at the e and g positions, and generally polar residues at the surface-exposed positions b, c, and/or f to enhance solubility (**Figure 3A**). The residues at the e and g positions are crucial for ensuring either homo or hetero association in native coiled-coils (Lavigne et al., 1995; Kohn et al., 1998; Krylov et al., 1994), and this has been key to design heterodimeric coiled-coils (Zhou et al., 1994). The E3/K3 coiled-coil with the sequence $(\mathbf{E}IAAL\mathbf{E}K)_3/(\mathbf{K}IAAL\mathbf{K}E)_3$ contains a high number of glutamic acids (Glu, **E**) and lysines (Lys, **K**) at positions e and g to be complementary to one another,

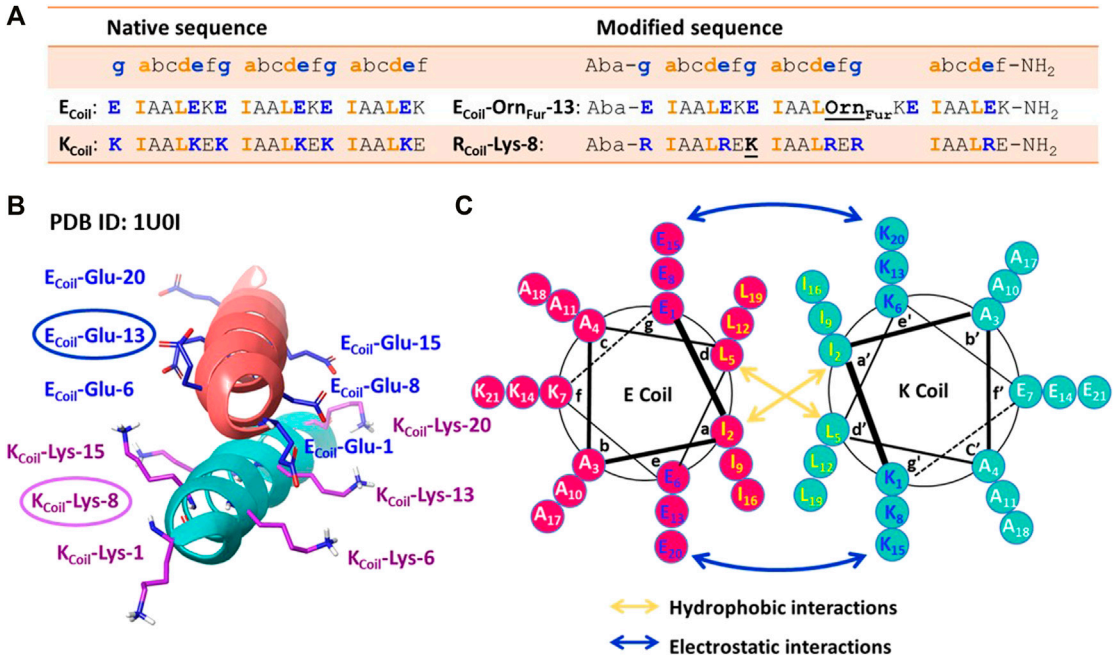


FIGURE 3 | (A) Peptide sequences of the native [(EIAALEK)3/(KIAALKE)3] and modified coiled-coil complex. Peptide sequences are written N-terminus to C-terminus. The four amino acids (IAAL) in the peptide sequence correlate with the positions a, b, c, and d of the coiled-coil heptad repeat (abcdefg), and EKE (ECoil) or KEK/RER (K/RCoil) correlate with the e, f, and g positions. An Aba (aminobenzoic acid) group is introduced at the N-terminus and NH2 indicates the amide group at the C-terminus. **(B)** Crystal structure of the (EIAALEK)3/(KIAALKE)3 coiled-coil heterodimer in which the partner moieties of interest (Ecoil: Glu at position 13 and Kcoil: Lys at position 8) are indicated. PDB ID: 1U0I. **(C)** Helical wheel representation of the (EIAALEK)3/(KIAALKE)3 coiled-coil heterodimer viewed in cross-section. The interhelical hydrophobic interactions are denoted with orange arrows, and interhelical electrostatic interactions are denoted by the blue arrows.

allowing 10 electrostatic interaction pairs at the interface between the two subunits to enhance heterodimer stability and destabilize homodimer formation (Litowski and Hodges, 2002) (**Figure 3**). Folding a sequence with this repeating pattern into an alpha-helical secondary structure causes the hydrophobic residues (isoleucine and leucine at positions a and d, respectively) to be presented in a way that allows interaction with the same residues of the complementary coil (**Figure 3C**). This characteristic "knobs-into-holes" manner of packing (Litowski and Hodges, 2002) establishes the hydrophobic core. The stability of such coiled-coils is directly proportional to the correct complementarity of both the electrostatic and hydrophobic interactions at the interface, resulting in a high-affinity binding with a dissociation constant of 70 nM (Litowski and Hodges, 2002). In the current work, we apply the furan technology for the first time in this coiled-coil peptide heterodimer as a model system to explore the reactivity of oxidized furan moieties towards different amino acid side chains (**Figure 1C**).

RESULTS AND DISCUSSION

Coiled-Coil Structure and Design Rationale for the Current Study

Based on the published NMR structure of the E3/K3 coiled-coil (PBD 1U01) (Lindhout et al., 2004), we determined the optimal position to insert a furan-containing amino acid and the potential nucleophiles on the E3/K3$_{Coil}$, respectively. In the first instance, we tried to reproduce the previously observed cross-linking [*vide supra* (Miret-Casals et al., 2021a)] between an activated furan moiety and a lysine residue located in close proximity in the current coiled-coil context. We envisioned that the replacement of a weak interhelical ionic bond (E$_{Coil}$: Glu at position 13 and K$_{Coil}$: Lys at position 8, **Figure 3B**) respectively with the two partner moieties able to engage in covalent bond formation would not interfere with the stability of the coiled-coil structure. The salt-bridging pairs with the shortest distances (Lindhout et al., 2004), essential for the coiled-coil formation, and the Glu and Lys residues positioned at the N/C terminus with higher flexibility were left untouched. To allow selective cross-linking and avoid problems with the formation of multiple adducts, the K$_{Coil}$ was transformed to an R$_{Coil}$ (**Figure 1C**), where all lysines, except for the Lys residue at position 8, were replaced by arginine (resulting in the sequence R$_{Coil}$-Lys-8), which allows preserving the same number of positive charges in the K$_{Coil}$ but eliminates potential competing nucleophiles (**Figure 3A**). Introduction of the furan moiety for cross-linking involves the precursor unnatural amino acid Orn (ornithine), which is coupled with 2-furanpropionic acid (Fur) through the amino group of the Orn side chain placed at position 13 on the E$_{Coil}$ (**Figure 3A**, resulting in the sequence E$_{Coil}$-Orn$_{Fur}$-13). The ornithine residue, featuring three methylene spacers between the backbone and the amine attachment point for the furan moiety, was chosen to give

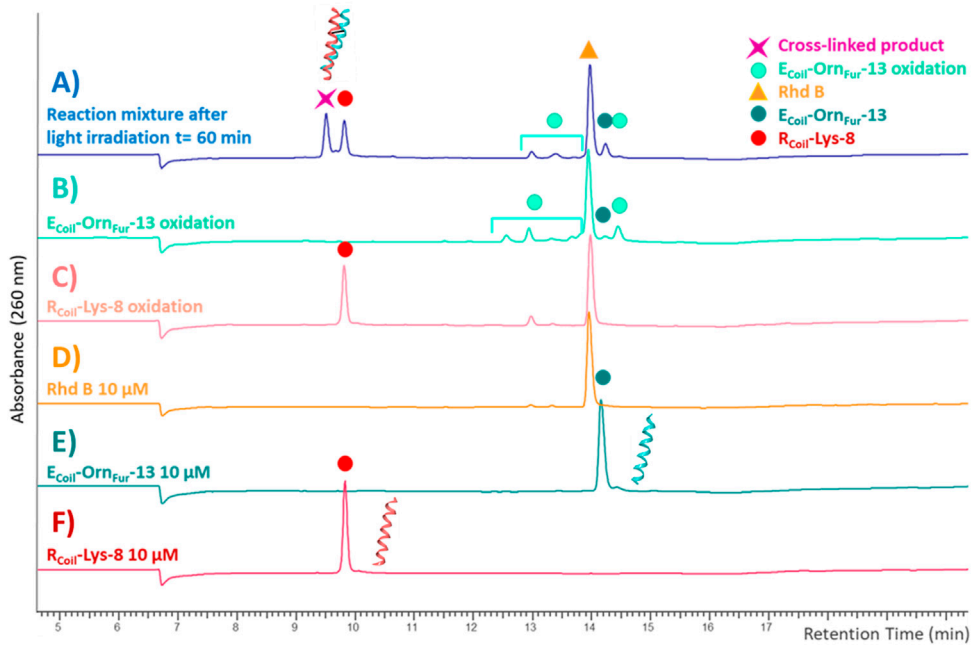


FIGURE 4 | HPLC-UV chromatograms recorded at 260 nm with a XTerra® Shield RP18 column, 125 Å (5 µM 2.1 × 250 mm). The reaction mixture after light irradiation [blue trace **(A)**] is the cross-link reaction between R_{Coil}-Lys-8 [red trace **(F)**] and E_{Coil}-Orn_{Fur}-13 [dark green trace **(E)**] after 60 min of light irradiation with Rhd B [orange trace **(D)**] at 10 µM. The R_{Coil}-Lys-8 oxidation (light red) trace **(C)** and E_{Coil}-Orn_{Fur}-13 oxidation (green) trace **(B)** were generated by exposure to singlet oxygen by light irradiation in the presence of Rhd B at 10 µM for 60 min in absence of the other coil.

enough flexibility to the Fur moiety to scan its R_{Coil} proximate surroundings for potential furan-cross-link partners. In addition, the peptides were additionally modified with an Aba (aminobenzoic acid) group on the N-terminus to increase the UV activity for HPLC detection and with an amide group on the C-terminus to minimize any repulsive electrostatic interactions between the E3/R3 coiled-coil (**Figure 3A**).

The E3/R3 Coiled-Coil System is an ideal Model to Explore Furan Reactivity

To ensure that the chosen coiled-coil model effectively allows cross-linking of the furan warhead with lysine, we thus chemically synthesized E_{Coil}-Orn_{Fur}-13 and R_{Coil}-Lys-8. We assumed that the modified coiled-coil peptides will have similar K_D as the native E3/K3 coiled-coil (**Figure 3A**, ESI 2.1-2.7, and ESI 3.2-3.3). Initially, both coils were pre-incubated to allow coiled-coil formation. At a concentration of 10 µM for each coil peptide, the monomer↔dimer equilibrium is shifted toward the formation of a coiled-coil dimer in 92.0% (see ESI 2.10). Then, the furan moiety can be selectively oxidized to a reactive keto-enal upon generation of singlet oxygen by light irradiation in the presence of a photosensitizer, such as Rose Bengal (RB) or Rhodamine B (Rhd B) (Miret-Casals et al., 2021a) (see ESI 2.8 and 2.9). A site-selective cross-link reaction can then occur between the oxidized furan moiety at the E_{Coil} and the amine group present in the R_{Coil} if proximate.

The reactivity of the oxidized furan moiety towards nucleophiles is followed by HPLC analysis of the corresponding reaction mixtures (**Figure 4**) after light irradiation (trace A) of R_{Coil}-Lys-8 (trace F) and E_{Coil}-Orn_{Fur}-13 (trace E) and concurrent singlet oxygen production with Rhd B (D trace). The individual peptides were also exposed to singlet oxygen using the same conditions but in absence of the complementary coil (traces C and B, respectively). Note that the cross-linked product peak is generated (purple star in trace A) and the area of the signal corresponding to the R_{Coil}-Lys-8 (red dot in trace A) has decreased compared to the red dot in trace C. Importantly, R_{Coil}-Lys-8 does cross-link to activated E_{Coil}-Orn_{Fur}-13 using RB; however, different cross-linked products were observed (see ESI 4.1, **Supplementary Figure S14**). The Rhd B-irradiation protocol (10 µM; 60 min of light irradiation, **Figure 4**) generates one cross-linked product with the expected mass of the covalent product previously described in Miret-Casals et al. (2021a) and the dehydrated form (−18 Da, see ESI 4.2, **Supplementary Figure S15**). The oxidized covalent product (+16 Da) is also detected and becomes the main product when applying the RB-irradiation protocol (2.5 µM; 30 min of light irradiation, see ESI 4.1, **Supplementary Figure S15** in comparison with **Supplementary Figure S16** in 4.2). We have previously described that singlet oxygen produced by PS-irradiation protocols can produce oxidation-induced damage in several amino acids (such as Cys, Met, Tyr, His, and Trp) (Miret-Casals et al., 2021a). However, the E_{Coil}-Orn_{Fur}-13/R_{Coil}-Lys-8 coiled-coil model system does not contain amino acids prone to oxidation. The fact that we observe an oxidized cross-link product in the current system thus indicates that the oxidation takes place at the level of the newly formed

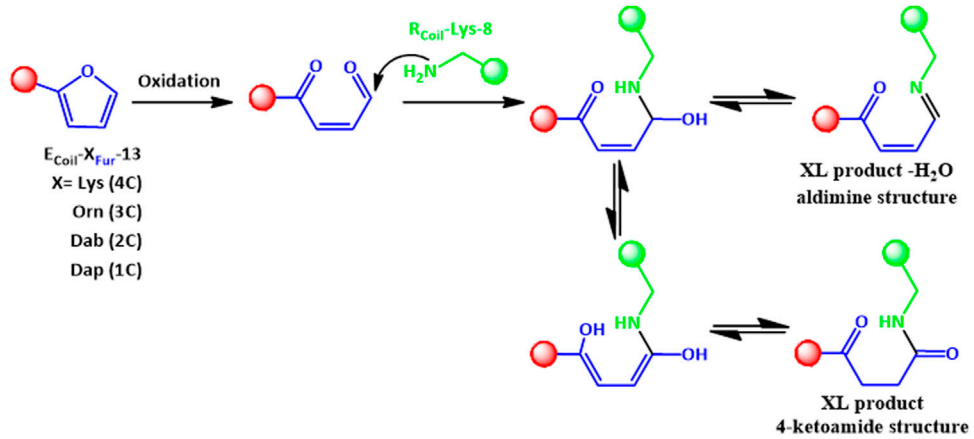
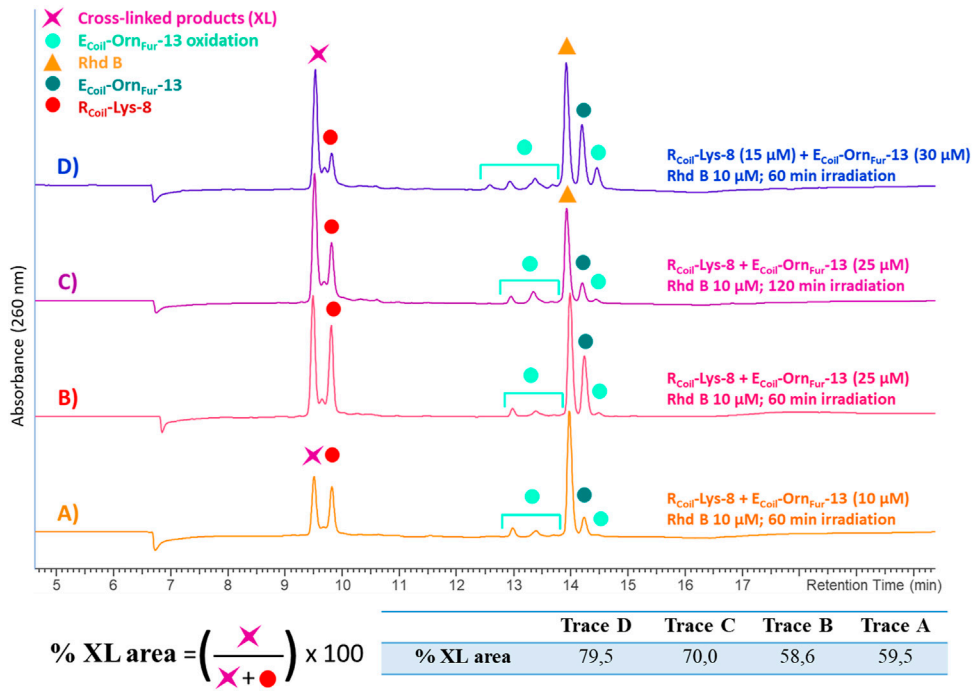


FIGURE 5 | The 4-ketoamide structure is proposed as the end product of the reaction between E_{Coil}-X_{Fur}-13 with R_{Coil}-Lys-8 after furan activation, with the same exact mass as the structure described previously in Miret-Casals et al. (2021b). An aldimine species (XL product—H_2O) can also be formed. XL refers to cross-linked species.



	Trace D	Trace C	Trace B	Trace A
% XL area	79,5	70,0	58,6	59,5

FIGURE 6 | All the chromatograms show the cross-link reaction product (purple star) between R_{Coil}-Lys-8 (red dot) and E_{Coil}-Orn_{Fur}-13 (dark green dot) after light irradiation with Rhd B (orange triangle) at 10 µM. The E_{Coil}-Orn_{Fur}-13 oxidation products (light green dots) are produced after activation of the furan moiety by singlet oxygen generation. **(A)** R_{Coil}-Lys-8 (10 µM) and E_{Coil}-Orn_{Fur}-13 (10 µM) and 60 min of light irradiation. **(B)** R_{Coil}-Lys-8 (25 µM) and E_{Coil}-Orn_{Fur}-13 (25 µM) and 60 min of light irradiation. **(C)** R_{Coil}-Lys-8 (25 µM) and E_{Coil}-Orn_{Fur}-13 (25 µM) and 120 min of light irradiation. **(D)** R_{Coil}-Lys-8 (15 µM) and E_{Coil}-Orn_{Fur}-13 (30 µM) and 60 min of light irradiation. The cross-link yield was quantified as the percentage of the cross-linked product area and the values are indicated in the table.

covalent adduct. Earlier work from Zhu and Sayre (2007) describes a 4-ketoamide as the end product of the reaction of a 4-oxo-2-enal species with a primary amine in pH 7.4 HEPES buffer. Such a 4-ketoamide structure results in the same exact mass as the structure described in Miret-Casals et al. (2021a) (**Figure 5**, XL product). The oxidized cross-linked product (+16 Da) can be explained by oxidation of the 4-ketoamide by oxygen-derived free radical reactions due to singlet oxygen production. In addition, R_{Coil}-Lys-8 can react with the keto-enal-E_{Coil} forming an aldimine (cross-link product −18 Da, **Figure 5**). To further increase the level of cross-linking (ensuring more complete furan-activation) while minimizing

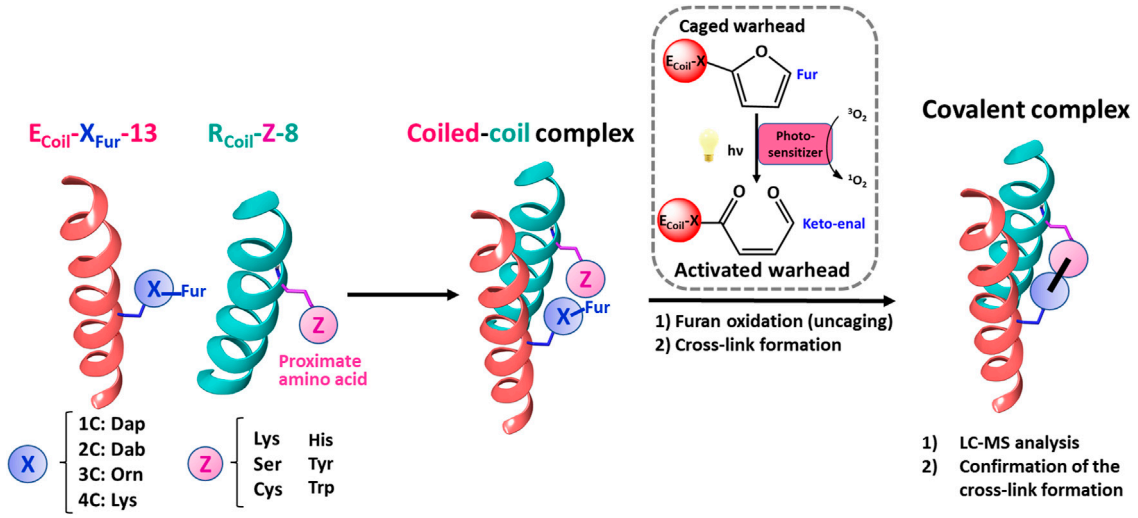


FIGURE 7 | Schematic representation of the strategy followed to explore furan reactivity in an E3/R3 coiled-coil peptide dimer. To scan the furan warhead reactivity towards additional cross-link partners, E_{Coil}-X_{Fur}-13 and R_{Coil}-Z-8 were chemically synthesized. Dap, Dab, Orn, and Lys were introduced at position 13 on the E_{Coil} as handles to introduce the Fur moiety. These four amino acids featuring varying lengths of methylene spacers were used to study the optimal distance between the furan moiety and the peptide backbone to increase the efficiency of cross-linking between the coiled-coil motifs. Ser, Cys, His, Tyr, and Trp, besides Lys, were chosen as potential new partners for the furan warhead and were placed at position 8 on the R_{Coil}. All possible E/R coiled-coil partners were pre-incubated to allow coiled-coil complex formation, and then PSs (RB or Rhd B) were added and irradiated with light to afford covalently connected coiled-coil motifs upon furan activation.

oxidation of the cross-linked product, several conditions were tested using the Rhd B-irradiation protocol [i.e., time dependency of $^1O_2$ production, dose dependency of the coiled-coil interaction, and E_{Coil}:R_{Coil} ratio (**Figure 6**)]. To allow a more quantitative analysis of the cross-link yield and to compare between different experiments, the percentage of the cross-linked product was calculated integrating the different signals in the HPLC chromatograms taking into account that the ratio between photosensitizer concentration and coil peptide concentration is changing, as well as the E/R_{Coil} ratio (see ESI 5.1-5.2 for a more detailed explanation of how the respective yields were calculated). The concentration of both coils was increased from 10 to 25 μM (see **Figures 6A,B**, respectively) to study the influence of the coiled-coil interaction on the cross-link yield, but no differences were observed in terms of the cross-linked product area as indicated in the table of **Figure 6**. Note that cross-link efficiency at 10–25 μM is independent of concentration because there is almost complete association of monomers to the coiled-coil dimer (see ESI 2.10). Next, the $^1O_2$ production time was increased up to 120 min, and the cross-link efficiency improved from 59% to 70% (**Figure 6C**). The highest level of cross-linked product (79.5%) was achieved when changing the E_{Coil}:R_{Coil} ratio from 1:1 to 2:1 (**Figures 6B,D**, respectively); however, single and double oxidized covalent products (+16, +32 Da, respectively, which probably relate to free radical oxidation products of the 4-ketoamide) were also detected (**Supplementary Figures S19–S20** in 5.4).

To further evaluate the influence of chain length and positioning of the furan moiety on the cross-link reaction, Lys (4C chain), Dab (2,4-diaminobutyric acid, 3C chain), and Dap (2,4-diaminopropionic acid, 1C chain) were introduced at position 13 on the E_{Coil} as handles to couple the Fur moiety (**Figure 7** and ESI 3.4-3.6). These three amino acids together with Orn, featuring varying lengths of methylene spacers between the terminal amino group of the side chain and the α-carbon atom, were used to study the optimal distance between the furan moiety and the peptide backbone. All HPLC and MS analyses of reaction mixtures between R_{Coil}-Lys-8 and E_{Coil}-X_{Fur}-8 (with X = Lys, Orn, Dab, or Dap) have been included in Supplementary Materials (see ESI, section 5.3–5.6). In all cases, high cross-link levels were achieved although the ratio between the expected covalent product (4-ketoamide) and the −18 Da form (aldimine) differs, the last one becoming the main product when using E_{Coil}-Dap_{Fur}-13 (see ESI 5.6, **Supplementary Figures S23–S24**). It is important to mention that high levels of doubly oxidized covalent product (+32 Da) were observed when using E_{Coil}-Lys_{Fur}-13 and E_{Coil}-Dab_{Fur}-13 (see ESI 5.3 and 5.5, **Supplementary Figures S18, S22**); however, the oxidation level of the cross-linked product can be tuned by the type of PS and by the PS properties (i.e., dose and time dependency of $^1O_2$ production). The secondary structure of the R_{Coil} and E_{Coil} peptides can be monitored experimentally by circular dichroism (CD). The CD spectra of the R_{Coil}-Lys-8, E_{Coil}-Dap_{Fur}-13, and E_{Coil}-Lys_{Fur}-13, as well as 1:1 mixture of E and R coil were recorded (see ESI 2.10, 5.7 and 5.8, **Supplementary Figures S25, S26**). E_{Coil}-Dap_{Fur}-13 and E_{Coil}-Lys_{Fur}-13 with the shortest and longest furan-side chain length were chosen for CD analysis. The CD spectra are reported as the mean residue molar ellipticity ([θ]) and demonstrate that R_{Coil}-Lys-8 forms a coiled-coil domain with both E_{Coils} with the characteristic minima at 208 and 222 nm as previously reported for the heterodimeric E3/K3 coiled-coil model system (Litowski and Hodges, 2002). The

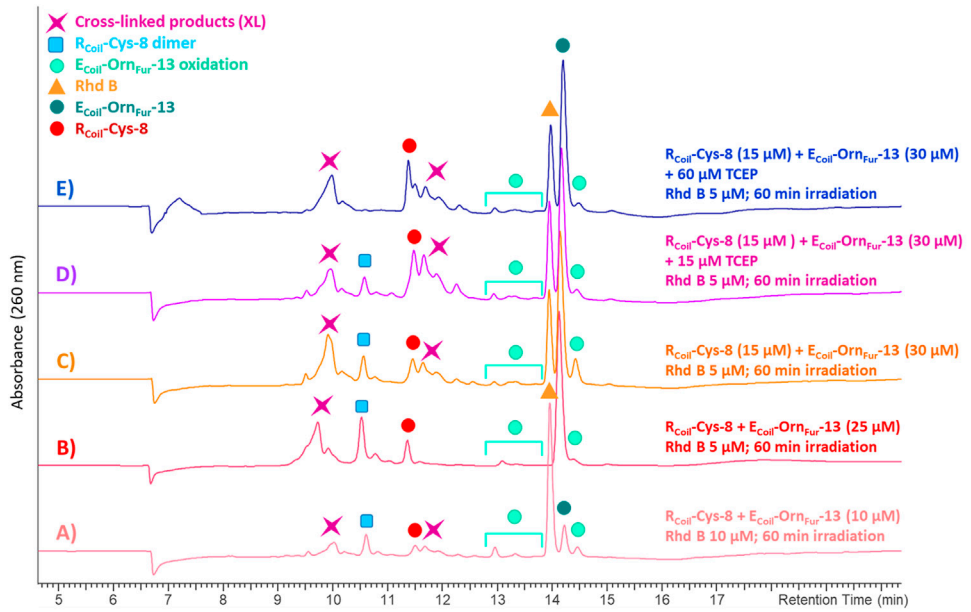
negative molar ellipticity ([θ]) at 222 nm is directly proportional to the amount of helical structure and the $(\theta)_{222}/(\theta)_{208}$ ratio is typically ˃1.0 for E3/K3 coiled-coil helical heterodimers and 0.66–0.72 for the single α-helices E3 and K3 coils as previously reported by Litowski and Hodges (2002) (see **Supplementary Table S1** in ESI 2.10). The 1:1 mixture of both E_{Coils} with R_{Coil}-Lys-8 had the largest (θ) at 222 nm (see ESI 5.7-5.8, tables in **Supplementary Figures S25, S26**), indicating that the heterodimeric coiled-coils have the maximum helical structure. Note that transformation of the K_{Coil} to an R_{Coil} as well as the introduction of the furan-side chain at position 13 (residue on position e, which is not important for the hydrophobic core) on the E_{Coil} do not interfere in coiled-coil formation. Furan cross-linking is described for the first time to constrain coiled-coil helical dimers and demonstrates that the E3/R3 coiled-coil system is an ideal model to explore furan reactivity.

Scanning of Furan Warhead Towards Novel Cross-Link Partners in a Coiled-Coil Peptide Dimer

To evaluate whether other amino acids can cross-link with the furan warhead, three nucleophilic amino acids, cysteine (Cys), histidine (His), and serine (Ser), and two amino acids that can potentially engage in electrophilic aromatic substitution, tyrosine (Tyr) and tryptophan (Trp), were selected and incorporated at position 8 on the R_{Coil} (**Figure 7**). After chemical synthesis (see ESI 3.7-3.11), the secondary structure of the peptides was evaluated by CD spectroscopy for all R_{Coils} and E_{Coil}-Lys_{Fur}-13, as well as a mixture containing both (see ESI 6). All the R_{Coils} formed heterodimeric coiled-coils with E_{Coil}-Lys_{Fur}-13 with the characteristic minima at 208 and 222 nm and the largest [θ] at 222 nm (see ESI 6, **Supplementary Figures S27–S31**). Note that the introduction of different amino acids at position 8 (residue on position g) on the R_{Coil} does not alter the coiled-coil nature. Next, the efficiency of the E_{Coil}-Orn_{Fur}-13 to cross-link to the new R_{Coils} was examined using the two PS-irradiation protocols as described above, and the reaction mixtures after light irradiation were investigated by HPLC analysis (see ESI 7). Rhd B and RB-irradiation protocols show evidence of cross-linked product for Cys and Tyr-R_{Coil} to E_{Coil}-Orn_{Fur}-13 (see ESI 7, **Supplementary Figures S33, S35, S38, S40**). In addition, R_{Coil}-Trp-8 seems to cross-link to E_{Coil}-Orn_{Fur}-13 using Rhd B albeit with a lower cross-link yield (see ESI 7.1.5, **Supplementary Figure S36**). Note that the GPR54/KP-10 homology model allowed visualization of GPR54-C62 and C277 in close proximity to KP-10-W3 residue (**Figure 2A**; **Supplementary Figure S2** in ESI 1). In addition, several GPR54-Tyr residues, as well as several Actin-Tyr residues were observed in sufficient proximity of the KP-10-W3 (see **Figure 2A**; **Supplementary Figure S2** in ESI 1) or Tβ4-E24 (**Figure 2B**), respectively. All these residues were identified as potential furan-cross-link partners when KP-10-W3 or Tβ4-E24 are replaced by Fua, and we indeed prove site-specific cross-linking of Cys and Tyr side chains towards the furan warhead after activation (see below).

Cys was the only nucleophilic amino acid, besides Lys, to cross-link with the Fur moiety. To further increase the level of cross-link efficiency between R_{Coil}-Cys-8 and E_{Coil}-Orn_{Fur}-13 (ensuring furan-activation) while minimizing oxidation of the cross-linked products, Rhd B was chosen as a PS in view of the lower oxidation levels of R_{Coil}-Cys-8 when $^1O_2$ is produced in the absence of the Fur-E_{Coil} compared to RB (see ESI 7, **Supplementary Figures S33, S38**). The concentration of both coils was increased from 10 to 25 μM (**Figures 8A,B**, respectively) to study the dose dependency of the coiled-coil cross-link yield. A higher UV absorbance was observed in the chromatograms, related to the increase in E3/R3 coil concentration, but there were no differences in terms of cross-link yields (69.2% and 67.2%, respectively, see table in **Figure 8**; **Supplementary Table S4** in ESI 8.2). This confirms the almost complete association of monomers to the coiled-coil dimer at 10–25 μM as previously described for R_{Coil}-Lys-8 and E_{Coil}-Orn_{Fur}-13.

We note that the R_{Coil}-Cys-8 cross-links by itself to give a disulfide bond dimer when singlet oxygen is produced (see ESI 8.3, **Supplementary Figures S43, S44**). Disulfides formed from oxidation of two Cys residues by singlet oxygen production *via* the thiyl radical intermediates and subsequent dimerization of these species has been previously reported (Schöneich, 2016; Trujillo et al., 2016). Moreover, the disulfides can further react with singlet oxygen to form zwitterionic peroxides, thiosulfinates, and thiosulfonates (Clennan et al., 1997; Clennan, 2001). We envisaged that by changing the E_{Coil}:R_{Coil} ratio from 1:1 to 2:1 (**Figures 8B,C**, respectively), R_{Coil}–R_{Coil} interactions would become less important, as well as the level of Cys-mediated dimerization, and the level of E3/R3 coiled-coil cross-linking would improve (**Figure 8C**). However, the excess of E_{Coil}-Orn_{Fur}-13 was not enough to completely prevent R_{Coil}-Cys-8 dimerization. R_{Coil}-Cys-8 was pre-treated with tris(2-carboxyethyl)phosphine (TCEP) in a 1:1 ratio for 10 min before adding E_{Coil}-Orn_{Fur}-13 and starting the Rhd B-light irradiation protocol (**Figure 8D**). TCEP is known as a reducing agent and is used to break the disulfide bonds and keep cysteines in their reduced form. A total of 4 equivalents (equiv.) of TCEP were needed to stop R_{Coil}-Cys-8 dimerization and to obtain a high level of coiled-coil cross-linked product (81.3%, **Figure 8E**, **Supplementary Table S4** in ESI 8.2, and ESI 8.5). The optimized protocol (RCoil: 15 μM + 60 μM TCEP, ECoil: 30 μM, Rhd B: 5 μM, and 60 min of light irradiation) was used to cross-link R_{Coil}-Cys-8 with all E_{Coil}-X_{Fur}-13 and the reaction mixtures were investigated by HPLC and MS analysis (all chromatograms and spectra can be found in ESI 8.4-8.7). As previously reported for Lys, high cross-link levels were also achieved with R_{Coil}-Cys-8 with all Fur-E_{Coil} (E_{Coil}-X_{Fur}-13, see ESI 3.1, **Supplementary Table S2**). The MS analyses (see ESI 8.4-8.7, **Supplementary Figures S45, S47, S49, S51**) show peaks at different retention times with the same cross-linked product mass. The mass of the cross-linked product corresponds to the Michael-type addition of the cysteine thiol group of R_{Coil}-Cys-8 to the α,β-unsaturated double bond of the activated furan moiety (keto-enal-E_{Coil}) after generation of singlet oxygen. The addition of the thiol may occur at the α- or β-position relative to the ketone group of the activated furan, and possibly all cross-linked



$$\%\ \text{XL area} = \left(\frac{\times}{\times + \bullet + \square}\right) \times 100$$

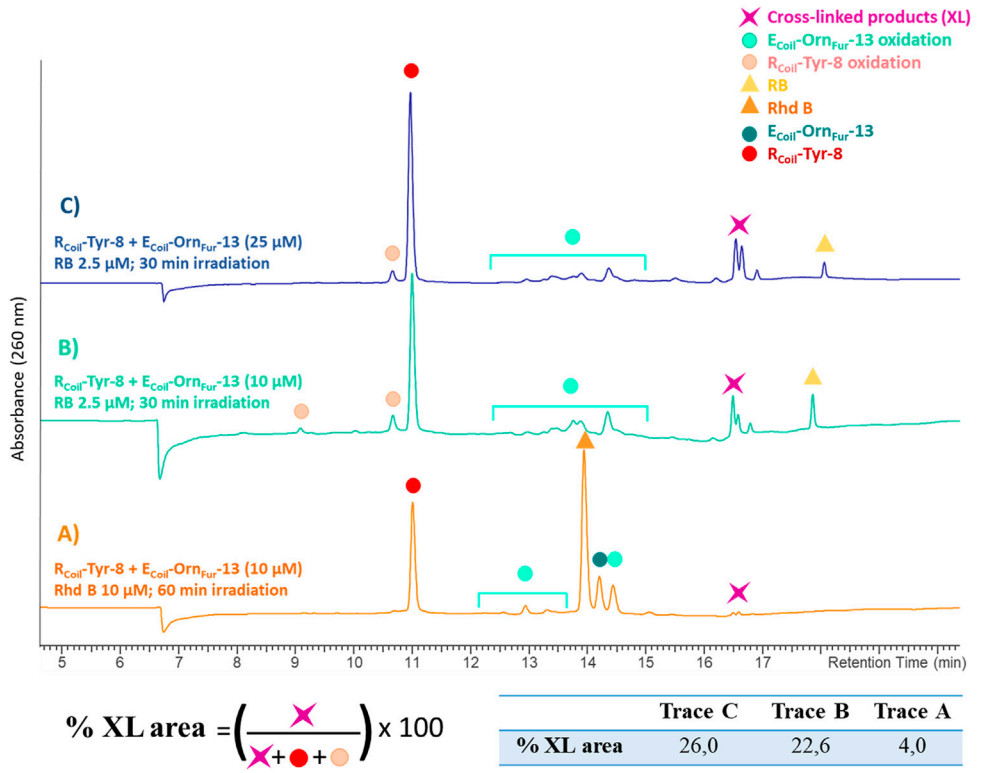
	Trace E	Trace D	Trace C	Trace B	Trace A
% XL area	81,3	73,2	72,9	67,2	69,2

FIGURE 8 | All the chromatograms show the cross-linking reaction product (purple star) between R_{Coil}-Cys-8 (red dot) and E_{Coil}-Orn_{Fur}-13 (dark green dot) after light irradiation with Rhd B (orange triangle). The E_{Coil}-Orn_{Fur}-13 oxidation products (green light dots) are produced after activation of the furan moiety by singlet oxygen generation. **(A)** R_{Coil}-Cys-8 (10 μM) and E_{Coil}-Orn_{Fur}-13 (10 μM), Rhd B 10 μM and 60 min of light irradiation. **(B)** R_{Coil}-Cys-8 (25 μM) and E_{Coil}-Orn_{Fur}-13 (25 μM), Rhd B 5 μM and 60 min of light irradiation. **(C)** R_{Coil}-Cys-8 (15 μM) and E_{Coil}-Orn_{Fur}-13 (30 μM), Rhd B 5 μM and 60 min of light irradiation. **(D)** R_{Coil}-Cys-8 (15 μM) + 15 μM TCEP and E_{Coil}-Orn_{Fur}-13 (30 μM), Rhd B 5 μM and 60 min of light irradiation. **(E)** R_{Coil}-Cys-8 (15 μM) + 60 μM TCEP and E_{Coil}-Orn_{Fur}-13 (30 μM), Rhd B 5 μM and 60 min of light irradiation. The cross-link yield was quantified as the percentage of the cross-linked product area and the values are indicated in the table.

products are mixtures of R and S enantiomers (see ESI 8.1, **Supplementary Figure S42**). That explains why various cross-linked product peaks are observed in the HPLC chromatograms. In addition to the Michael-type addition cross-linked product, the dehydrated and the oxidized form of the cross-linked product (−18 Da and +16 Da, respectively) can be observed in small quantities, as well as R_{Coil}-Cys-8 dimer.

In addition, the activated Fur moiety cross-links with amino acids that can engage in electrophilic aromatic substitution: Tyr and Trp; however, the cross-linking efficiency for Trp (only observed when using Rhd B) is lower as compared to Tyr (cross-linked product is observed for both PS; however, cross-link yield is higher when using RB, see **Figures 9A,B** and ESI 7.1.4, **Supplementary Figure S35** in comparison with 40 in 7.2.4). The level of cross-linking between R_{Coil}-Tyr-8 and E_{Coil}-Orn_{Fur}-13 was slightly increased by increasing the concentration of both coils up to 25 μM using RB (from 22.6% to 26.0%, see **Figures 9B,C** and ESI 9.2). No differences were observed in terms of cross-link efficiency upon increasing the E_{Coil}-to-R_{Coil} ratio from 1:1 to 2:1 and increasing the $^1O_2$ production time (data not shown). Note that the RB-irradiation protocol produces oxidation-induced damage to Tyr, but cross-link yields are higher compared to Rhd B (See **Figures 9A,B**). E_{Coil}-Dab_{Fur}-13 (2 methylene spacers) and E_{Coil}-Dap_{Fur}-13 (1 methylene spacer) with a shorter side chain did not cross-link or cross-linked in much lower yields to R_{Coil}-Tyr-8, respectively (see ESI 9.5-9.6, **Supplementary Figures S58, S59**). The E_{Coil}-Dap_{Fur}-13 mass analysis reveals a mini-peak with a high molecular mass that could correspond to a cross-linked product; however, more detailed characterization was not possible (see ESI 9.6, **Supplementary Figure S60**). E_{Coil}-Lys_{Fur}-13 cross-links to R_{Coil}-Tyr-8 with similar efficiency to the E_{Coil}-Orn_{Fur}-13 (see ESI 9.3-9.4, **Supplementary Figure S54** in comparison with **Supplementary Figure S56**). We propose that R_{Coil}-Tyr-8 cross-links to E_{Coil}-Lys_{Fur}-13 through an electrophilic aromatic substitution, after which aromatic rearrangements and loss of water occur to give the final product (see ESI 9.1, **Supplementary Figure S53** for details of the proposed structure). The calculated mass of this structure correlates with the found mass for the cross-linked product (see ESI 9.3, **Supplementary Figure S55**); however, when shortening the connecting chain to the furan (cfr. cross-link experiments of R_{Coil}-Tyr-8 with E_{Coil}-Orn_{Fur}-13), the corresponding adduct could not be elucidated in detail. These results indicate that the longest chains (E_{Coil}-Lys_{Fur}-13 and E_{Coil}-Orn_{Fur}-13) with more flexibility can accommodate the cross-linked product with the most thermodynamically stable structure.

The R_{Coil}-Trp-8 cross-links to E_{Coil}-Orn_{Fur}-13 with low cross-link yield. The best cross-link efficiency was achieved using an E_{Coil}:R_{Coil} ratio of 1:1 with Rhd B at 10 μM and light irradiation



$$\%\ \text{XL area} = \left(\frac{\times}{\times + \bullet + \bullet}\right) \times 100$$

	Trace C	Trace B	Trace A
% XL area	26,0	22,6	4,0

FIGURE 9 | All chromatograms show the cross-link reaction product (purple star) between R_{Coil}-Tyr-8 (red dot) and E_{Coil}-Orn_{Fur}-13 (dark green dot) after light irradiation with Rhd B (orange triangle) and RB (yellow triangle). The E_{Coil}-Orn_{Fur}-13 oxidation products (green light dots) are produced after activation of the furan moiety by singlet oxygen generation. **(A)** R_{Coil}-Tyr-8 (10 µM) and E_{Coil}-Orn_{Fur}-13 (10 µM), Rhd B 10 µM and 60 min of light irradiation. **(B)** R_{Coil}-Tyr-8 (10 µM) and E_{Coil}-Orn_{Fur}-13 (10 µM), RB 2.5 µM and 30 min of light irradiation. **(C)** R_{Coil}-Tyr-8 (25 µM) and E_{Coil}-Orn_{Fur}-13 (25 µM), RB 2.5 µM and 30 min of light irradiation. The cross-link yield was quantified as the percentage of the cross-linked product area and the values are indicated in the table.

for 120 min (see ESI 10.2, **Supplementary Figures S62, S63**). No cross-link product was observed for the other E_{Coil}-X_{Fur}-13 coil peptides (see ESI 10.1 and 10.3).

CONCLUSION

In previous work, the furan-cross-link technology was applied to KP-10-GPR54 and Tβ4-Actin complex systems, and peptide–protein and protein–protein cross-link products were observed. The formed cross-link complex was shown to result from the reactivity of the furan warhead (incorporated in a peptide or protein) towards a Lys (in close proximity) in the target protein. However, the KP-10-GPR54 homology model as well as the crystal structure of Actin-Tβ4 (4PL7) (**Figure 2**) indicate that besides Lys, other side-chain amino acids can be located in close proximity to the activated furan moiety. The E3/R3 coiled-coil model system used here, now enabled firm identification of Cys and Tyr as new nucleophilic partners, able to react and form a covalent bond. At the same time, the system allowed to study the optimal distance between the furan moiety and the peptide backbone.

In conclusion, we have reported that the replacement of weak interhelical ionic contacts with suitable precursors for furan-oxidation-based cross-linking affords stable dimeric coiled-coil structures that can be covalently cross-linked upon visible light irradiation. We describe for the first time reaction of the oxidized furan moiety with cysteine and tyrosine residues in the target strand, in addition to the earlier reported lysine-oriented furan warhead (Miret-Casals et al., 2021a). It is important to note that unnatural furan-containing amino acids can be incorporated in peptides and proteins by either flexible *in vitro* translation (Decoene et al., 2018) or genetic encoding of noncanonical amino acids, in *Escherichia coli* (Schmidt and Summerer, 2013) and in human cells (Schmidt et al., 2014), respectively. This study suggests that furan warheads, when incorporated into peptides or proteins, can be used for oxidation-induced cross-linking to a variety of amino acid side chains in the target protein when spatial proximity is ensured to discover weak and/or transient protein–ligand and protein–protein interactions.

MATERIALS AND METHODS

Synthesis, Purification, and Characterization of Coil Peptides

Chemicals, general procedures, solid phase peptide synthesis, purification, and characterization of the peptides are described

in the Supplementary Materials section of this paper (see ESI 2 and 3).

Circular Dichroism Spectroscopy

CD spectra of the E/R_{Coils} analogues, as well as a mixture containing both, were obtained using a JASCO J7100 instrument (Tokyo, Japan), equipped with a HAAKE cryostat temperature-controlled cell holder at 25°C. CD spectra are reported as the mean residue molar ellipticity ([θ]) with units of degrees square centimeter per decimole (deg x cm^2/dmol), calculated by the equation:

$$[\theta] = (\theta_{obs} \times MRW)/(10lc)$$

where θ_{obs} is the ellipticity in millidegrees, MRW is the mean residue molecular weight (molecular weight of the peptide divided by the number of amino acid residues), l is the path length of the cuvette in centimeters, and c is the peptide concentration in milligrams per milliliter. The CD spectra were recorded at 50 nm/min scan rate, a bandwidth of 1 nm, a data pitch of 0.1 nm, a response of 0.5 s, a wavelength range of 200–260 nm, and a 1-cm path length cell. Each spectrum was an average of nine scans. Baselines were corrected by subtracting the solvent contribution [phosphate-buffered saline (PBS) 1 × buffer]. The CD spectra for all R_{Coil}-Z-8 and E_{Coil}-Lys_{Fur}-13 were measured at 5 μM in PBS (pH 7.4) for each coil peptide, as well as a 1-to-1 mixture containing both; see Supplementary Materials for further details (see ESI 2.10).

Cross-Linking Experiments and Singlet Oxygen Production

The concentrated R_{Coil}-Z-8 and E_{Coil}-X_{Fur}-13 peptides (100 μM) were diluted in air-saturated PBS (pH 7.4). The used concentration of coil peptides is indicated in figures or figure legends. Cross-linking experiments took place in 2-ml Eppendorf vials in a total volume of 300 μl placed. R_{Coil}-Z-8 and E_{Coil}-X_{Fur}-13 peptides were incubated for 5 min at room temperature (binding step), and then the photosensitizers, Rhodamine B or Rose Bengal, were added to the mixture to a final concentration of 10 or 2.5 μM, respectively (or as indicated in figures). The lamp was then placed on top of the Eppendorf and the samples were irradiated with a Euromex Illuminator Eκ-1 lamp (110 W, 12 V, halogen lamp LE.5210) coupled with an optical fiber arm (Euromex LE.5214 dual-arm light conductor) at room temperature for 60 and 30 min, respectively (or as indicated in figures). The light intensity of the lamp was kept in between 6.8 and 7 Klux. After irradiation, the reaction mixture was left to react for 1 h at 25°C, and the samples were submitted to HPLC-UV and HPLC-MS analysis; see Supplementary Materials for further details (see ESI 2.9).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: SP and DA synthesized the R_{Coil} and E_{Coil} peptide analogues. SP performed the initial cross-link experiments. JD and PB developed the GPR54 homology model and did the kisspeptin-10 docking into the model receptor. LM-C designed, supervised, performed, and optimized the cross-link experiments. LM-C and AM discussed the results, provided critical feedback, and conceived the presented idea. LM-C and AM contributed to the final version of the manuscript. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Marleen Van Troys for CD technical assistance. Johan Winne is thanked for insightful discussion on the nature of the cross-linked species.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.799706/full#supplementary-material

REFERENCES

Boyle, A. L., and Woolfson, D. N. (2011). De Novo designed Peptides for Biological Applications. *Chem. Soc. Rev.* 40, 4295–4306. doi:10.1039/c0cs00152j

Case, D. A. E. A. (2016). *AMBER 2016*. San Francisco: University of California.

Clennan, E. L. (2001). Persulfoxide: Key Intermediate in Reactions of Singlet Oxygen with Sulfides. *Acc. Chem. Res.* 34, 875–884. doi:10.1021/ar0100879

Clennan, E. L., Wang, D., Clifton, C., and Chen, M.-F. (1997). Geometry-dependent Quenching of Singlet Oxygen by Dialkyl Disulfides. *J. Am. Chem. Soc.* 119, 9081–9082. doi:10.1021/ja9720568

Decoene, K. W., Vannecke, W., Passioura, T., Suga, H., and Madder, A. (2018). Pyrrole-Mediated Peptide Cyclization Identified through Genetically Reprogrammed Peptide Synthesis. *Biomedicines* 6, 99. doi:10.3390/biomedicines6040099

Diharce, J., Cueto, M., Beltramo, M., Aucagne, V., and Bonnet, P. (2019). In Silico Peptide Ligation: Iterative Residue Docking and Linking as a New Approach to Predict Protein-Peptide Interactions. *Molecules* 24, 351. doi:10.3390/molecules24071351

Grigoryan, G., and Keating, A. (2008). Structural Specificity in Coiled-Coil Interactions. *Curr. Opin. Struct. Biol.* 18, 477–483. doi:10.1016/j.sbi.2008.04.008

Kohn, W. D., Kay, C. M., and Hodges, R. S. (1998). Orientation, Positional, Additivity, and Oligomerization-State Effects of Interhelical Ion Pairs in α-helical Coiled-Coils. *J. Mol. Biol.* 283, 993–1012. doi:10.1006/jmbi.1998.2125

Kotani, M., Detheux, M., Vandenbogaerde, A., Communi, D., Vanderwinden, J.-M., Le Poul, E., et al. (2001). The Metastasis Suppressor Gene KiSS-1 Encodes

Kisspeptins, the Natural Ligands of the Orphan G Protein-Coupled Receptor GPR54. *J. Biol. Chem.* 276, 34631–34636. doi:10.1074/jbc.m104847200

Krylov, D., Mikhailenko, I., and Vinson, C. (1994). A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J.* 13, 2849–2861. doi:10.1002/j.1460-2075.1994.tb06579.x

Lapenta, F., Aupič, J., Strmšek, Ž., and Jerala, R. (2018). Coiled Coil Protein Origami: from Modular Design Principles towards Biotechnological Applications. *Chem. Soc. Rev.* 47, 3530–3542. doi:10.1039/c7cs00822h

Lavigne, P., Kondejewski, L. H., Houston, M. E., Jr., Sönnichsen, F. D., Lix, B., Sykes, B. D., et al. (1995). Preferential Heterodimeric Parallel Coiled-Coil Formation by Synthetic Max and C-Myc Leucine Zippers: a Description of Putative Electrostatic Interactions Responsible for the Specificity of Heterodimerization. *J. Mol. Biol.* 254, 505–520. doi:10.1006/jmbi.1995.0634

Lindhout, D. A., Litowski, J. R., Mercier, P., Hodges, R. S., and Sykes, B. D. (2004). NMR Solution Structure of a Highly Stablede Novo Heterodimeric Coiled-Coil. *Biopolymers* 75, 367–375. doi:10.1002/bip.20150

Litowski, J. R., and Hodges, R. S. (2002). Designing Heterodimeric Two-Stranded α-Helical Coiled-Coils. *J. Biol. Chem.* 277, 37272–37279. doi:10.1074/jbc.m204257200

Liu, J., Zheng, Q., Deng, Y., Cheng, C.-S., Kallenbach, N. R., and Lu, M. (2006). A Seven-helix Coiled Coil. *Proc. Natl. Acad. Sci.* 103, 15457–15462. doi:10.1073/pnas.0604871103

Lupas, A. (1996). Coiled Coils: New Structures and New Functions. *Trends Biochem. Sci.* 21, 375–382. doi:10.1016/s0968-0004(96)10052-9

Mason, J. M., and Arndt, K. M. (2004). Coiled Coil Domains: Stability, Specificity, and Biological Implications. *Chembiochem* 5, 170–176. doi:10.1002/cbic.200300781

Miret-Casals, L., Vannecke, W., Hoogewijs, K., Arauz-Garofalo, G., Gay, M., Díaz-Lobo, M., et al. (2021). Furan Warheads for Covalent Trapping of Weak Protein-Protein Interactions: Cross-Linking of Thymosin Beta4 to Actin. *Chem. Commun.* 57, 6054. doi:10.1039/d1cc01731d

Miret-Casals, L., Vannecke, W., Hoogewijs, K., Arauz-Garofalo, G., Gay, M., Díaz-Lobo, M., et al. (2021). Furan Warheads for Covalent Trapping of Weak Protein-Protein Interactions: Cross-Linking of Thymosin β4 to Actin. *Chem. Commun.* 57, 6054–6057. doi:10.1039/d1cc01731d

Peterson, L. A. (2013). Reactive Metabolites in the Biotransformation of Molecules Containing a Furan Ring. *Chem. Res. Toxicol.* 26, 6–25. doi:10.1021/tx3003824

Pinilla, L., Aguilar, E., Dieguez, C., Millar, R. P., and Tena-Sempere, M. (2012). Kisspeptins and Reproduction: Physiological Roles and Regulatory Mechanisms. *Physiol. Rev.* 92, 1235–1316. doi:10.1152/physrev.00037.2010

Schmidt, M. J., and Summerer, D. (2013). Red-light-controlled Protein-RNA Crosslinking with a Genetically Encoded Furan. *Angew. Chem. Int. Ed.* 52, 4690–4693. doi:10.1002/anie.201300754

Schmidt, M. J., Weber, A., Pott, M., Welte, W., and Summerer, D. (2014). Structural Basis of Furan-Amino Acid Recognition by a Polyspecific Aminoacyl-tRNA-Synthetase and its Genetic Encoding in Human Cells. *Chembiochem* 15, 1755–1760. doi:10.1002/cbic.201402006

Schöneich, C. (2016). Thiyl Radicals and Induction of Protein Degradation. *Free Radic. Res.* 50, 143–149. doi:10.3109/10715762.2015.1077385

Trujillo, M., Alvarez, B., and Radi, R. (2016). One- and Two-Electron Oxidation of Thiols: Mechanisms, Kinetics and Biological Fates. *Free Radic. Res.* 50, 150–171. doi:10.3109/10715762.2015.1089988

Vannecke, W., Ampe, C., Van Troys, M., Beltramo, M., and Madder, A. (2017). Cross-Linking Furan-Modified Kisspeptin-10 to the KISS Receptor. *ACS Chem. Biol.* 12, 2191–2200. doi:10.1021/acschembio.7b00396

Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* 54, 5–37. doi:10.1002/cpbi.3

Woolfson, D. N. (2017). "Coiled-Coil Design: Updated and Upgraded," in *Fibrous Proteins: Structures and Mechanisms*. Editors D. A. D. Parry and J. M. Squire (Cham: Springer International Publishing), 35–61. doi:10.1007/978-3-319-49674-0_2

Zhou, N. E., Kay, C. M., and Hodges, R. S. (1994). The Role of Interhelical Ionic Interactions in Controlling Protein Folding and Stability. *J. Mol. Biol.* 237, 500–512. doi:10.1006/jmbi.1994.1250

Zhu, X., and Sayre, L. M. (2007). Long-lived 4-Oxo-2-Enal-Derived Apparent Lysine Michael Adducts Are Actually the Isomeric 4-ketoamides. *Chem. Res. Toxicol.* 20, 165–170. doi:10.1021/tx600295j

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership