EXPANSION OF THE GENETIC CODE: UNNATURAL AMINO ACIDS AND THEIR APPLICATIONS

EDITED BY: Subhendu Sekhar Bag, Ishu Saraogi and Jiantao Guo

frontiers Research Topics

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

EXPANSION OF THE GENETIC CODE: UNNATURAL AMINO ACIDS AND THEIR APPLICATIONS

Topic Editors:
Subhendu Sekhar Bag, Indian Institute of Technology Guwahati, India
Ishu Saraogi, Indian Institute of Science Education and Research, Bhopal, India
Jiantao Guo, University of Nebraska-Lincoln, United States

Table of Contents

Check for updates

Editorial: Expansion of the Genetic Code: Unnatural Amino Acids and their Applications

Subhendu Sekhar Bag[1]*, Ishu Saraogi[2]* and Jiantao Guo[3]*

[1]Department of Chemistry, Indian Institute of Technology Guwahati, Guwahati, India, [2]Department of Chemistry, Indian Institute of Science Education and Research, Bhopal, India, [3]Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE, United States

Editorial on the Research Topic

Expansion of the Genetic Code: Unnatural Amino Acids and their Applications

Every living organism encodes proteins using the same 20 natural amino acid building blocks, possessing a limited number of functional groups, in response to selective triplet codons. However, evidence of a large number of post-translational modifications in the protein cofactors and the rare occurrence of two unnatural amino acids, namely selenocysteine and pyrrolysine, clearly indicates the need for additional functional groups. A milestone achieved by (Offord, 1987; Kaiser, 1989) in the development of synthetic and semi-synthetic methods for the incorporation of unnatural amino acids into desired peptide or protein sequences paved the way for Schultz et al., for developing a strategy towards the site-specific incorporation of unnatural amino acids into proteins, thereby expanding the genetic code. In 1989, *Schultz and coworkers* reported the introduction of an amber stop codon at the desired site, in a gene of interest, via site-directed mutagenesis (Noren et al., 1989). Translation of the corresponding mRNA was possible by means of amber-stop codon suppression (Guo et al., 2009). Since then, an increase in interest in pursuing research on expanding the genetic code has prompted scientists globally to design unnatural amino acids. Although many unnatural amino acids were found to be unsuitable for imparting novel functionalities in desired proteins, the expansion of genetic code has dramatically increased the functional potential of a wide variety of proteins and peptides alike. Thus, the translation of an expanded genetic code has the potential to produce semi-synthetic organisms with increased biodiversity and functionalities.

This special issue mainly focuses on the recent achievements at the junction of organic chemistry and molecular biology, especially towards expanding the genetic code. Here we highlight the synthesis of several unnatural amino acids, the generation of non-canonical tRNA synthetase libraries and tRNAs, site-specific incorporation of uAA into target proteins and the study of synthetic protein functions.

Expansion of the genetic code is known to increase or modify the functionality of proteins. Towards this end, Patrick O'Donoghue et al. reported the production of a site-specifically acylated variant of *Thioredoxin Reductase 1* (*TrxR1*), containing selenocysteine as the unnatural amino acid. The authors showed that the modified *TrxR1* enzymatic protein could provide resistance against oxidative damage even under oxidising conditions. It was further demonstrated that the acetylation of *TrxR1* could enhance the redox activity facilitating *TrxR1* to resist oxidative damage even in the presence of very reactive oxygen species.

In an effort toward engineering the active site of (R)-*amine transaminase* enzyme, Hyngdon Yun and his group reported the site-specific incorporation of non-canonical amino acid, *p*-benzoyl phenylalanince (pBpA), following the strategy of expansion of the genetic code. Thus, the replacement of a single phenylalanine residue at the 88th position of the active site resulted in a 15 times enhancement in activity for 1-phenylpropan-1-amine and an 8 times enhancement for benzaldehyde. In another instance, F86A/F88pBpA, a 30% increment in thermostability of the enzyme at 55°C was reported without any change in enzymatic activity. Therefore, Hyngdon Yun and coworkers clearly showed the enhanced functionalities of enzymes having expanded genetic codes.

The application of expansion of the genetic code in the study of acetylation was demonstrated by Chenguang Fan and coworkers. The enzyme aconitase has been known to catalyse the reversible conversion of citrate and isocitrate. The model organism *E. coli* is known to possess two isoforms of the enzyme, namely AcnA and AcnB. The authors showed that the acetylation of AcnA K684 resulted in a decrease in enzymatic activity, while the acetylation of AcnB K567 led to an increase in activity. This report provided the opportunity of generating proteins of diverse and altered functionality depending on the site-specific incorporation of an expanded genetic code.

Scientists have been trying to improve the interaction between codons and anticodons at in-frame amber-stop codons in order to improve yields of proteins with an expanded genetic code (Mala and Saraogi, 2022). However, reports of site-specific incorporation of unnatural amino acids at the N-termini of proteins are rare. Towards this end, Tae Hyeon Yoo et al. developed an orthogonal translation initiation system for site-specific incorporation of unnatural amino acids at the N-termini of proteins. Thus, a *Methanococcus jannaschii* tRNA^{Tyr} was engineered into an initiator tRNA by introducing identity elements of *E. coli* initiator tRNA. Using this system, they were able to site-specifically incorporate an unnatural amino acid *O*-propargyl-L-tyrosine (OpgY) at the translation initiation position by means of amber codon suppression. Further results indicated that this system was inactive towards the incorporation of uAA in response to any other internal TAG codon. To avoid misincorporation of Gln amino acid, the study was carried out only in the presence of *O*-propargyl-L-tyrosine (OpgY) amino acid.

Over the years, scientists have mainly introduced unnatural amino acids in response to a suppressed amber-stop codon. However, the strategy of sense codon reassignment leads to the incorporation of a wide variety of non-canonical amino acids, which may find their use in the advancement of protein engineering and bioorganic chemistry. In this line, John D. Fisk and his group reported the reassignment of an arginine sense codon, AGG, for the incorporation of tyrosine amino acid. The most efficient *M. jannaschii* tyrosyl tRNA synthetase variants capable of incorporating tyrosine in place of arginine were later selected and transplanted onto another *M. jannaschii* aaRS, evolved for the incorporation of an unnatural amino acid *p*-azidophenylalanine. This study opened up a new gateway of incorporation of newer amino acids by reassigning sense codons for the generation of novel recombinant proteins.

Besides protein engineering, unnatural amino acids find potential applications in designing peptidomimetic therapeutics with enhanced pharmacology, diagnostics and supra-molecular self-assembly/organogel. Over the years, several types of designer amino acids, including fluorescent triazolyl amino acids and scaffold amino acids (Bag et al., 2015) have been developed to synthesise peptidomimetics. Aminopyrazonolyl amino acid scaffold having hydrogen-bonded supramolecular self-assembling properties has been exploited by Nagendra Sharma et al. to showcase organogellation. Thus they have shown that a type of aminopyrazolone amino acid, namely O-alkylated ampyrone containing hybrid peptides, formed organogels after sonication in ethyl acetate: hexane solvent (1:3) under set parameters. Such APA-peptides also showcased the potential formation of hydrogen bonding, and are promising candidates to function as peptidomimetics.

The detailed study of oxidative post-translational modifications of proteins is very important. Such modifications are often reported in disease pathology and can serve as potential tools for predicting the disease. Methods to selectively detect proteins with expanded genetic code, mimicking oxPTM (oxidative post-translational modification) scenarios, are very essential because such oxPTM proteins have been known to result in functional changes that can eventually contribute to disease pathology. The accumulation of oxPTM nitrotyrosine is reported in over 100 proteins associated with diseases. This prompted Ryan A. Mehl and his coworkers to develop a nanobody selective for 3-nitrotyrosine modified 14-3-3 signalling protein. The same nanobody was reportedly less specific for nitrotyrosine present in other proteins. Using this strategy, the authors demonstrated the selectivity of a single nanobody to oxidative post-translationally modified protein targets, even in highly proteinaceous solutions. Nanobodies have immense potential to serve as powerful tools to study the complex intracellular dynamics of oxPTMs and their adverse pathological impacts.

Genetically encoded designer amino acids with electrophilic moieties are often found as excellent candidates for studying protein-protein interactions through cross-linking. Towards this end, an alkyl bromide-based unnatural amino acid (BprY) has been utilised by Xin Shu et al. to investigate protein-protein interactions under both *in vitro* and *in vivo* conditions. The BprY was found to target not only cysteine but also a broad range of nucleophilic amino acids to form protein-protein cross-links. The results were supported by the broad reactivity of BprY with the Affibody/Z protein complex. The same unnatural amino acid was also used to study the interaction of SUMO2 (Small Ubiquitin-like Modifier 2) and RNF111, which was devoid of any cysteine residues at the binding site. A total of 264 SUMO2 interacting proteins were captured and identified at a whole proteome level using the same uAA. Thus, the authors demonstrated that BprY and relevant alkyl halide unnatural amino acids could serve as excellent candidates to map protein-protein interactions. Additionally, Xuyu Liu and his group reviewed the use of synthetic thiol, and selenol-derived unnatural amino acids for

the expansion of chemical protein synthesis. This review is an excellent resource for researchers seeking insights into the chemical synthesis of various analogues of thiolated and selenylated amino acid molecules. The expanded scope of native chemical ligations to assess homogeneously modified proteins in a highly efficient manner has also been highlighted. They have also shown the selected amino acids in the applications of the chemically synthesised thiolated and selenylated amino acids for the chemical synthesis of selected post-translationally modified peptides and proteins. Several strategies, such as native chemical ligation, have been explored for the synthesis of post-translationally modified proteins. The native chemical ligation strategy offers an exciting opportunity to construct libraries of protein therapeutics.

Overall, this thematic issue highlights the potential impact of unnatural amino acids in bioorganic chemistry and chemical biology. Reports on exploring unnatural amino acids toward the expansion of genetic code, generation of proteins with structural and functional diversity, and modulation of enzymatic activities have been demonstrated. Genetically encoded unnatural amino acids have also been showcased in designing peptidomimetic therapeutics and the study of protein-protein interactions. This special issue, thus, would have a significant impact on the advancement of research toward the expansion of the genetic code leading to the generation of proteins/enzymes with enhanced and diverse functionalities.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Bag, S. S., Jana, S., Yashmeen, A., and De, S. (2015). Triazolo-β-aza-ε-amino Acid and its Aromatic Analogue as Novel Scaffolds for β-turn Peptidomimetics. *Chem. Commun.* 51 (25), 5242–5245. doi:10.1039/c4cc08414d

Guo, J., Melançon, C. E., III, Lee, H. S., Groff, D., and Schultz, P. G. (2009). Evolution of Amber Suppressor Trnas for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angew. Chem.* 121 (48), 9312–9315. doi:10.1002/ange.200904035

Kaiser, E. T. (1989). Synthetic Approaches to Biologically Active Peptides and Proteins Including Enzymes. *Acc. Chem. Res.* 22 (2), 47–54. doi:10.1021/ar00158a001

Mala, P., and Saraogi, I. (2022). Enhanced Codon–Anticodon Interaction at In-Frame UAG Stop Codon Improves the Efficiency of Non-Natural Amino Acid Mutagenesis. *ACS Chem. Biol.* 9, 1051–1060. doi:10.1021/acschembio.1c00782

Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989). A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science* 244 (4901), 182–188. doi:10.1126/science.2649980

Offord, R. E. (1987). Protein Engineering by Chemical Means? *Protein Eng. Des. Sel.* 1 (3), 151–157. doi:10.1093/protein/1.3.151

Acetylated Thioredoxin Reductase 1 Resists Oxidative Inactivation

David. E. Wright[1], Nikolaus Panaseiko[1] and Patrick O'Donoghue[1,2]*

[1]Departments of Biochemistry, The University of Western Ontario, London, ON, Canada, [2]Departments of Chemistry, The University of Western Ontario, London, ON, Canada

Thioredoxin Reductase 1 (TrxR1) is an enzyme that protects human cells against reactive oxygen species generated during oxidative stress or in response to chemotherapies. Acetylation of TrxR1 is associated with oxidative stress, but the function of TrxR1 acetylation in oxidizing conditions is unknown. Using genetic code expansion, we produced recombinant and site-specifically acetylated variants of TrxR1 that also contain the non-canonical amino acid, selenocysteine, which is essential for TrxR1 activity. We previously showed site-specific acetylation at three different lysine residues increases TrxR1 activity by reducing the levels of linked dimers and low activity TrxR1 tetramers. Here we use enzymological studies to show that acetylated TrxR1 is resistant to both oxidative inactivation and peroxide-induced multimer formation. To compare the effect of programmed acetylation at specific lysine residues to non-specific acetylation, we produced acetylated TrxR1 using aspirin as a model non-enzymatic acetyl donor. Mass spectrometry confirmed aspirin-induced acetylation at multiple lysine residues in TrxR1. In contrast to unmodified TrxR1, the non-specifically acetylated enzyme showed no loss of activity under increasing and strongly oxidizing conditions. Our data suggest that both site-specific and general acetylation of TrxR1 regulate the enzyme's ability to resist oxidative damage.

Keywords: acetylation, enzymology, genetic code expansion, oxidation, post-translational modification, redox biology, selenocysteine

INTRODUCTION

Human cells actively eliminate reactive oxygen species (ROS) and resolve oxidative damage to proteins using multiple pathways, including the glutathione or thioredoxin (Trx) systems (Schweizer and Fradejas-Villar, 2016). The Trx system includes the selenocysteine-containing protein (selenoprotein) thioredoxin reductase (TrxR1). TrxR1 is a disulfide reductase with specificity for the redox mediator Trx. TrxR reduces a disulfide bond in Trx by catalyzing the oxidation of nicotinamide adenine dinucleotide phosphate (NADPH). The reduced Trx transfers electrons to oxidatively damaged proteins or ROS directly. For example, a pathway that resolves oxidation of methionine residues uses Trx-dependent enzymes to protect the proteome (Kim and Gladyshev, 2007). The resulting oxidized Trx can then be reduced again by TrxR1.

The Trx system is involved in regulating gene expression, embryonic development, cell proliferation, apoptosis, and many other cellular processes (Lu and Holmgren, 2012). In addition to Trx, TrxR1 can also directly reduce other cellular proteins, such as p53, protein disulfide isomerase, glutathione peroxidase, and NK-lysin (Arner et al., 1999; Mustacich and Powis, 2000) as well as low molecular weight ROS, including hydrogen peroxide, lipoic acid,
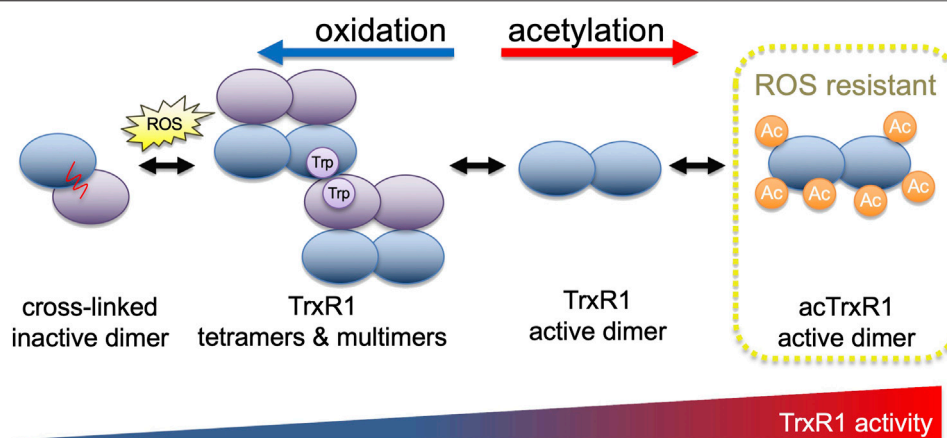
FIGURE 1 | TrxR1 activity is regulated by oxidation and acetylation. With increasing levels of reactive oxygen species, TrxR1 shows increased propensity to form low activity tetramers and higher order multimers (Wright et al., 2018). Oxidation ultimately leads to the formation of a covalent linkage between non-productive TrxR1 monomers, forming inactive cross-linked dimers (Xu et al., 2015). Acetylation of TrxR1, on the other hand, is associated with increased TrxR1 activity, reduced cross-linked dimer formation (Wright et al., 2018), and here we show that acetylated TrxR1 is resistant to oxidation and peroxide-induced multimerization.

selenite, 15-HPETE, and more (Arner, 2009). The Trx system provides a defense mechanism against ROS generated during oxidative stress, and consequently, alterations in the Trx system are associated with various diseases. TrxR1 is over-active in many aggressive cancers and is an early diagnostic marker (Selenius et al., 2010; Dong et al., 2016). TrxR1 is also an established anti-cancer drug target (Wang et al., 2012), and increased TrxR1 production or activity provides chemotherapeutic resistance to treatments that rely on the production of ROS to kill cells (Roh et al., 2017).

TrxR1 exists in an equilibrium of several different quaternary structures (Rengby et al., 2009; Selenius et al., 2010; Xu et al., 2015; Wright et al., 2018). TrxR1 can exist as inactive monomers, or as low active tetramers or higher order oligomers (Rengby et al., 2009; Xu et al., 2015; Wright et al., 2018; Shu et al., 2020). Catalytically competent dimers are the most active form of TrxR1, while inactive cross-linked dimers form because of covalent linkage between opposing subunits in associations between TrxR1 tetramers (**Figure 1**) (Xu et al., 2015; Wright et al., 2018). In cells, oxidative stress generated by Reactivating p53 and Inducing Tumor Apoptosis (RITA) induces TrxR1 tetramerization and covalent linkage, resulting in reduced activity or inactivation of TrxR1 (Xu et al., 2015). Thus, as oxidative stress increases, one of the cell's major oxidative stress defense mechanisms is prone to become increasingly ineffective.

Multiple reports document regulation of the Trx system without changes in total protein levels. These studies implicate post-translational modifications as potent regulators of the activity of different Trx system components (Choudhary et al., 2009; Folami Lamoke et al., 2011; Lamoke et al., 2011; Banerjee Mustafi et al., 2014). Three members of the Trx system can be acetylated to increase their activity, including TrxR1 (Wright et al., 2018), Trx1 (Folami Lamoke et al., 2011), and peroxiredoxin (Prx1) (Parmigiani et al., 2008). Acetylation inhibits super-oxidation of Prx1, preventing its oligomerization into higher molecular weight complexes with low peroxidase activity (Parmigiani et al., 2008). In

the context of disease, TrxR1 acetylation levels correlated positively with the level of oxidized cellular proteins in a mouse model of cardiomyopathy (Banerjee Mustafi et al., 2014). The oxidative stress generating anti-cancer compound RITA reduces TrxR1 activity in cell cultures by altering the oligomerization status of TrxR1 and increasing levels of inactive and cross-linked dimers (Xu et al., 2015). These reports suggest a relationship between acetylation, oligomerization, and oxidative damage in components of the Trx system. Although TrxR1 is known to be acetylated in response to oxidative stress (Banerjee Mustafi et al., 2014), the function of TrxR1 acetylation under oxidizing conditions is unknown.

Proteomic studies in Jurkat T lymphocytes, A549 cells and related non-small cell lung cancer cell lines have identified acetylation of TrxR1 at five distinct sites (Choudhary et al., 2009; Hornbeck et al., 2015; Wu et al., 2015). We showed that single acetylation at three of these sites in TrxR1 resulted in a 1.5 to 3-fold increase in enzyme activity (Wright et al., 2018). Because acetylated TrxR1 shows increased activity, we hypothesized that the acetylation of TrxR1 may serve as a mechanism to maintain TrxR1 activity under oxidizing conditions associated with increased ROS levels. To test this hypothesis, we used protein biochemistry to precisely measure the activity of acetylated TrxR1 variants over a broad range of peroxide concentrations that models the relevant range of ROS levels encountered by cells. We also generated a non-specifically acetylated TrxR1 using aspirin as a model acetyl donor. Together our findings suggest that acetylation is a potent mechanism to regulate TrxR1 activity that also allows the enzyme to evade oxidative inactivation.

MATERIALS AND METHODS

Plasmids and Strains

The plasmid pET-pylT-TrxR1 contains a His-tagged Human TrxR1 (isoform 4) with an in-frame UGA codon (Sec551) followed by the *E. coli* selenocysteine insertion sequence

FIGURE 2 | Genetic code expansion to incorporate acK and Sec in TrxR1. Expressed from a plasmid in *E. coli*, **(A)** AcKRS ligates the UAG-decoding tRNA^Pyl-opt with acK, **(C)** allowing insertion of acK at UAG codons. **(B)** In the endogenous Sec incorporation pathway (reviewed in (Serrao et al., 2018)), *E. coli* SerRS ligates tRNA^Sec with Ser, selenophosphate synthetase (SelD) produces selenophosphate (pSe), and selenocysteine synthase (SelA) uses the products of these reactions to convert Ser-tRNA^Sec to Sec-tRNA^Sec. An endogenous *E. coli* elongation factor (SelB) recruits Sec-tRNA^Sec to a UGA codon by also binding to a selenocysteine insertion sequence (SECIS) included in the 3′ untranslated region of the TrxR1 mRNA, allowing Sec incorporation at a specific UGA codon **(C)**. Together the acK and Sec systems can be used to produce active human TrxR1 protein with 22 genetically encoded amino acids.

(SECIS) RNA-hairpin loop (derived from the *E. coli* FdhF gene) in the 3′ untranslated region (3' UTR), which directs Sec-insertion at the UGA551 codon in recombinantly produced human TrxR1 variants from *E. coli* (**Figure 2**) (Wright et al., 2018). UAG stop codons inserted at positions 141, 200, or 307 in the TrxR1 gene allows for site-specific insertion of N_ε-acetyl-L-lysine (AcK) when co-expressed with plasmids bearing a pyrrolysyl-tRNA synthetase mutant specific for Ack (AckRS) Guo et al. (2014) and an optimized (Fan et al., 2015) UAG-decoding tRNA^Pyl (pTech-acKRS-tRNA^Pyl-opt) (**Figure 2**); (Wright et al., 2018).

TrxR1 and acTrxR1 Protein Purification

E. coli BL21 (DE3) (Invitrogen) was co-transformed with pET and pTech vectors. Preparative cultures (1 l) for each transformed strain were incubated with shaking at 37°C in lysogeny broth (LB) supplemented with 5 mM AcK (A4021-5G Sigma), 10 μM sodium selenite (10102-18-8, AlfaAesar), and appropriate antibiotics (100 μg/ ml ampicillin (BP1760-25, Fisher) for pET and 34 μg/ ml chloramphenicol (02930-100G, Ampresco) for pTech). We employed a previously optimized protocol for production of the selenoproteins in *E. coli* (Wright et al., 2018). At A_{600} = 1.2, the temperature was reduced to 20°C. At A_{600} = 1.5, 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) (BP1755-10, Fisher) was added and the cells then produced protein for a further 16–24 h. Cells were harvested by centrifugation and stored at −80°C until further use. TrxR1 variants were purified as previously (Wright et al., 2018). Briefly, cell pellets were resuspended in 30 ml phosphate buffer (100 mM potassium phosphate (PB0445, Biobasic), pH 7.2, 10% glycerol (CA97063-892, VWR)) supplemented with lysozyme (1 mg/ml) (12650-88-3, Biobasic) and disrupted by sonication. Following centrifugation at 6,250 × g, cell lysate was purified by affinity chromatography using Ni^{2+}-Nitrilotriacetic acid (NTA) resin (HisPur™ Ni-NTA Resin, PI88222, Fisher), as previously (Wright et al., 2018). Purified TrxR1 variants were stored in 100 mM potassium phosphate, pH 7.2, 50% glycerol at −80°C until use.

TrxR1 Activity Assays

TrxR1 activity was assessed using 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB) (D8130-5G, Sigma) also known as Ellman's reagent to detect the rate and level of reductive activity from TrxR1. The colorimetric reaction is followed by measuring reduction of DTNB to 2-nitro-5-thiobenzoate (TNB), which absorbs at 412 nm (A_{412}). Each reaction contains 250 nM TrxR1, 300 μM Nicotinamide adenine dinucleotide phosphate (NADPH) (N5130-25 MG, Sigma) and 5 mM DTNB in buffer containing 100 mM potassium phosphate, 1 mM Ethylenediaminetetraacetic acid (EDTA) (E4378-25G, Sigma), pH 7.0. Reactions were started by the addition of DTNB to a solution containing TrxR1 and NADPH, for a final volume of 100 μl in a 96 well plate. Measurements were taken in a Biotek Synergy H1 microplate reader every 1 min over a 1-h time course. All assays were performed using three independent enzyme reactions for each condition tested.

Peroxide and Aspirin Incubations

TrxR1 variants were also incubated with increasing concentrations of peroxide and then assessed as above for activity. Each TrxR1 and acTrxR1 enzyme variant (1 µM) was incubated in phosphate buffer (100 mM potassium phosphate, 1 mM EDTA, pH 7.0) with increasing peroxide concentrations (0–500 µM H_2O_2 (16,911-250ML-F, Sigma)) for 1 h in a total volume of 200 µl at 37°C. To generate non-site-specifically acetylated TrxR1, 500 nM TrxR1 was incubated with increasing concentrations (0–15 mM) of aspirin (A5376-250G, Sigma) in phosphate buffer for 1 h at 37°C. Activity assays with aspirin-treated and untreated TrxR1 were conducted exactly as above.

Western Blotting

Purified protein samples were suspended in 1 × sodium dodecyl sulfate (SDS) loading buffer (250 mM Tris-HCl pH 6.8, 40% glycerol (v/v), 10% SDS (w/v) (SB0485, Biobasic), 0.05% bromophenol blue (w/v) (0449-25G, Amresco), 5% 2-mercaptoethanol (M6250-100ML, Sigma)) and heated for 5 min at 95°C. Samples were then loaded in 15% SDS-polyacrylamide gels and electrophoresed. Following this, a PVDF membrane was soaked in methanol for 1 min. Both the SDS gel and membrane were soaked in transfer buffer (0.025 M Tris-HCl (TRS001.5, Bioshop Canada), pH 9.5, 0.192 M Glycine (56-40-6, Fisher), 20% (v/v) Methanol (CA71007-742, VWR), 0.5% (w/v) SDS) for 15 min. The blot was carried out with a TransBlot Turbo Transfer System (BioRad) at 15 V with 1.3 A for 15 min. The membrane was incubated in blocking solution (5% (w/v) skim milk powder (LP0031, Oxoid), 0.1% (v/v) Tween20 (9,005–64–5, Amresco), 1x PBS (137 mM NaCl (BP358-212, Fisher), 0.027 mM KCl (7447-40-7, Anachemia), 10 mM Na_2HPO_4 (SDB0487, Biobasic), and 2 mM KH_2PO_{4FF}) for 1 h, shaking, at room temperature. Then the membrane was incubated overnight with the primary antibody (anti-acetyl Lysine antibody from rabbit, Abcam ab80178; or anti-TrxR1 antibody, Santa Cruz Biotechnology sc28321) at 1:1,000 in blocking solution at 4°C overnight. Three 10-min washes with wash solution (0.5% (w/v) skim milk powder, 0.1% Tween20, 1x PBS) were conducted, shaking, at room temperature, followed by a 2 h incubation with a secondary antibody (Rabbit IgG HRP Linked F (ab′)2, GENA9340; Sigma) at 1:5,000 in wash solution at room temperature shaking. Next, three 10-min washes with PBS-tween (0.1% (v/v) Tween20, 1x PBS) were conducted, followed by a final wash with 1x PBS, shaking at room temperature. Clarity Western ECL Substrate (1,705,061; Biorad) was used for signal detection and chemiluminescent imaging was performed on a Chemidoc XRS + (Biorad).

Mass Spectrometry

For the MS/MS analysis of the aspirin acetylated and untreated TrxR1, gels were loaded with 0.31 ug of purified TrxR1 protein. Following SDS-PAGE, a 1 mm circular slice was picked from the gel using an Ettan Robotic Spot-Picker and submitted for proteolytic digestion (Trypsin) and peptide extraction at the Functional Proteomics Facility at the University of Western Ontario. Liquid chromatography and tandem mass spectrometry (LC-MS/MS) analyses of TrxR1 and aspirin-acetylated TrxR1 were performed at the Biological Mass Spectrometry Laboratory at The University of Western Ontario. Gel slices were de-stained with 50 mM ammonium bicarbonate (09830, Sigma) and 50% acetonitrile (00687, Sigma). The protein samples were reduced with 10 mM dithiothreitol (BP25641, Fisher), alkylated with 55 mM acrylamide (BP1406-1, Fisher), and digested with 5 ng/ µl trypsin (Promega). LC-MS/MS was performed using a Q-Tof Micro mass spectrometer (Waters) equipped with a Z-spray source in positive ion mode (+ 0.1% formic acid) or using Thermo Scientific LTQ-Orbitrap XL mass spectrometer. The data were analyzed and visualized using PEAKS software (Bioinformatic Solutions, Inc, Waterloo, Ontario).

Statistical Analysis

All activity assays were conducted in at least three independent enzyme reactions, including enzymes from independent preparations. A no enzyme (-TrxR1) control was subtracted from all reactions. All error bars represent one standard deviation, and p-values were calculated from a one-way analysis of variance.

RESULTS

Purification of AcK and Selenocysteine-Containing TrxR1 Variants

We used genetic code expansion to produce TrxR1 variants in *E. coli* with 22 different genetically encoded amino acids, including the non-canonical amino acids (ncAAs) acetyl-lysine (AcK) and selenocysteine (Sec) (**Figure 2**). We previously described purification, biochemical characterization and mass spectrometry analysis of each of TrxR1 variants to determine activity and verify incorporation of AcK at K141, K200, or K307 as well as Sec incorporation at the key active reside Sec551 (Wright et al., 2018). Briefly, we used a genetic code expansion system based on a PylRS mutant with activity for ligating AcK to tRNA^{Pyl}, which decodes amber (UAG) stop codons. Thus, by placing a UAG codon at the positions 141, 200 or 307 in our expression construct (**Figure 2**), we generated site-specifically acetylated TrxR1 variants.

Our TrxR1 expression construct also has a *E. coli* SECIS appended to the 3-untranslated region of our recombinant human TrxR1 gene. The SECIS element recruits the endogenous Sec-incorporation machinery to recode a UGA codon at position 551 from stop to Sec (**Figure 2**). Bacterial cultures must be supplement with sodium selenite to enable Sec formation (see Materials and Methods). The data generated below are based on multiple independent enzyme preparations, of which representative purified samples were visualized by SDS-PAGE (**Supplementary Figure S1**).

Resistance of Site-specifically Acetylated TrxR1 to Oxidative Damage

Because our work showed increased activity of site-specifically acetylated TrxR1 Wright et al. (2018) and previous studies implicated increased TrxR1 acetylation as a response to
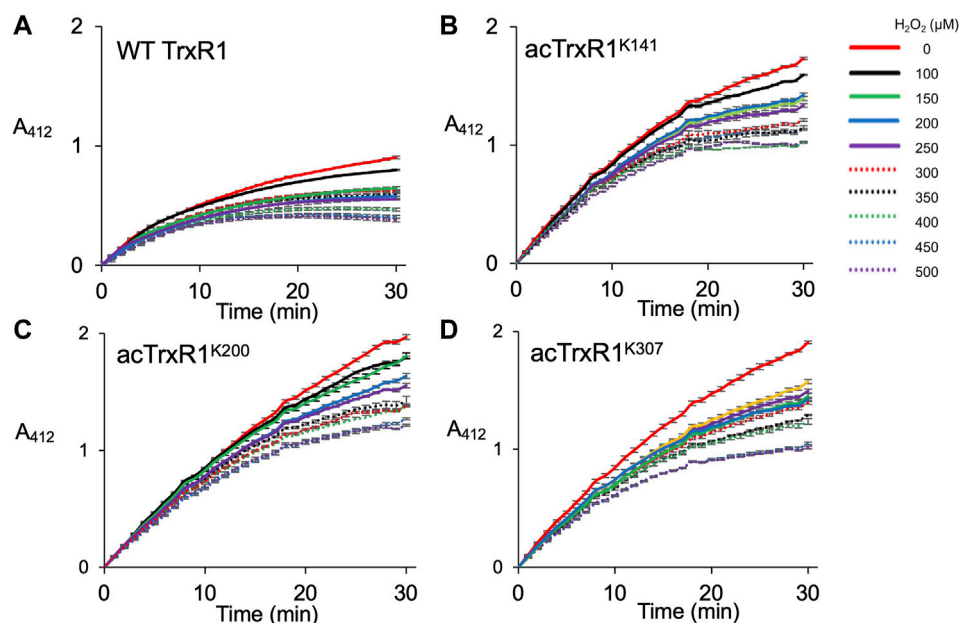
FIGURE 3 | Activity of un-modified and acTrxR1 variants with increasing peroxide levels. Purified TrxR1 variants were incubated with varying concentrations of buffer or H_2O_2 ranging from 0 to 500 µM for 1 h at 37°C. Following incubation, enzyme activity with the TrxR1 substrate DTNB was determined by following absorbance at 412 nm (A_{412}). Activity was measured for **(A)** wild-type TrxR1 and site-specifically acetylated variants **(B)** acTrxR1^{K141}, **(C)** acTrxR1^{K200}, and **(D)** acTrxR1^{K307}. Error bars represent ± 1 standard deviation about the mean of three independent enzyme reactions.

oxidative damage (Banerjee Mustafi et al., 2014), we hypothesized that acTrxR1 may be more active under oxidizing conditions or even resistant to oxidative damage. In cells, there are many sources of ROS, including superoxide, hydroxy, and nitric oxide radicals as well as H_2O_2, which is also an important signaling molecule (Bae et al., 2011).

In mammalian cells, physiological concentrations of H_2O_2 usually range from 0.1 to 10 µM, while cell stress responses are associated with greater peroxide concentrations in range of 10–500 µM (Sies, 2017). In apoptotic or necrotic human melanoma cells, peroxide levels can rise to more than 500 µM (Clement et al., 1998). To mimic oxidative damage that occurs in cells, we designed a series of experiments to measure the activity and initial velocity of purified TrxR1 and acTrxR1 variants under a range of peroxide concentrations (0–500 µM) encountered by normal as well as stressed cells. We first measured the catalytic activity of site-specifically acetylated TrxR1 (acTrxR1) variants as well as the wild-type TrxR1 (**Figure 3**). As previously (Wright et al., 2018), we found that under normal conditions (0 µM H_2O_2), each of the acTrxR1 variants showed substantially more activity (~1.5-fold) than the un-modified TrxR1 (**Figure 3**). Increased activity of acTrxR1 variants is evident in both the maximal level of DTNB reduced (**Figure 3**) as well as the initial velocity observed during the reaction time course (**Figure 4**).

All TrxR1 variants showed reduced catalytic activity with increasing concentrations of hydrogen peroxide (H_2O_2). At all concentrations of H_2O_2, we observed significantly more absolute activity of each acTrxR1 variant compared to wild type (**Figure 4A**). Even at the highest peroxide concentrations, the

absolute maximal activity level of acTrxR1 variants was still higher than that we observed for un-modified TrxR1 even without peroxide treatment (**Figure 3**). The data suggest that the acTrxR1 variants are resistant to oxidation.

Based on the initial velocity observed in the DTNB reduction reactions, we then calculated the absolute (**Figure 4A**) and relative (**Figure 4B**) reduction in catalytic rate observed in wild type compared to acTrxR1 variants with increasing peroxide concentration. The absolute reaction rates responded to oxidation similarly as noted above. Namely, the acTrxR1s showed reduced initial velocities with increasing peroxide that was always significantly greater than the observed for the wild type TrxR1 (**Figure 4A**). The acTrxR1 variants also show robust or partial resistance to oxidation even when the maximal catalytic rates for TrxR1 and acTrxR1 are both normalized to 1.0. After normalization to the un-modified TrxR1, both acTrxR1^{K141} and acTrxR1^{K200} showed significantly less relative activity reduction at each peroxide concentration tested (**Figure 4B**). The data demonstrate that acTrxR1^{K141} and acTrxR1^{K200} are significantly more resistant to oxidation than the un-modified enzyme. For the acTrxR1^{K307} variant, we found similar resistance to oxidation in the normalized relative activity rates at about half of the peroxide concentrations tested from 200–450 µM H_2O_2. The data indicate that site-specific acetylation of TrxR1 is protective against the effects of oxidative damage.

We also monitored the TrxR1 proteins in the above reactions using SDS-PAGE. We visualized the TrxR1 variants under each condition using Coomassie staining and Western blotting with an anti-TrxR1 antibody (**Supplementary Figure S2**). At the highest peroxide
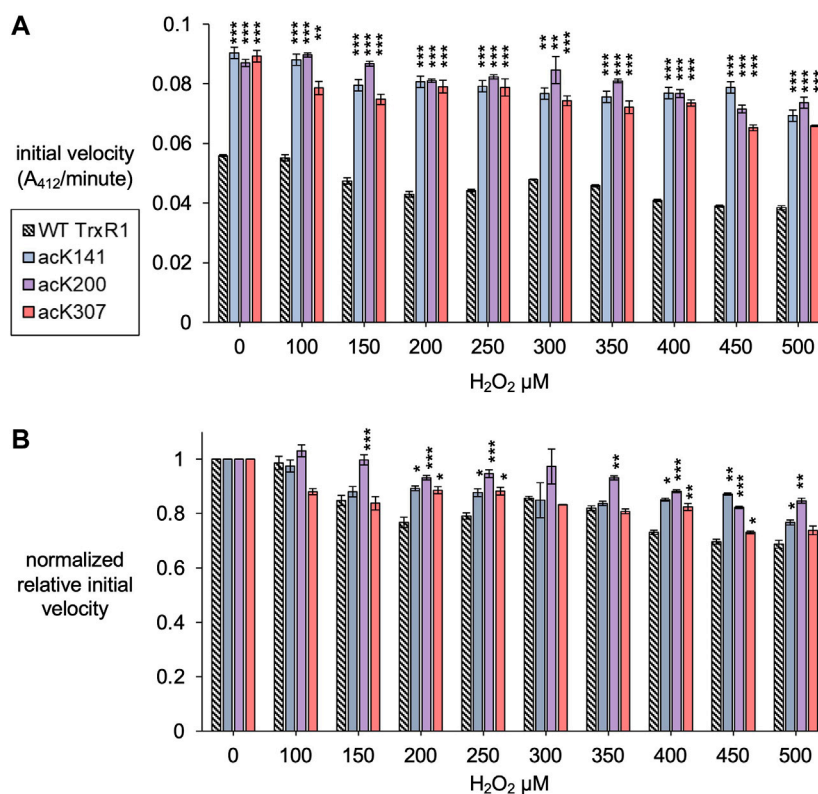
FIGURE 4 | Initial velocity of un-modified and specifically acetylated TrxR1 variants with increasing oxidative damage. **(A)** Initial velocity of each TrxR1 variants under the indicated H_2O_2 concentrations was calculated from the kinetic data (**Figure 3**). **(B)** To show relative changes in activity, the initial velocities were normalized by the activity of each variant under normal conditions (0 μM H_2O_2). Error bars represent ± 1 standard deviation about the mean of three independent enzyme reactions. Significant differences are annotated (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$) to compare acTrxR1 variants to WT TrxR1 in **(A)** or to compare the activity at each peroxide concentration to the 0 μM peroxide condition in (B).

concentration, we observe some degradation of the TrxR1 protein. The western blot revealed the presence of covalently linked high molecular weight TrxR1 oligomers following oxidation with H_2O_2 (**Supplementary Figures S2, S3**). For unmodified TrxR1, the high molecular weight complexes are visible already at 100 μM peroxide (**Supplementary Figure S2A**). Each of the acTrxR1 variants showed less accumulation of the TrxR1 multimers at each peroxide concentration. For the acTrxR1 variants, we observed a similar level of oligomerization to un-modified TrxR1 only at 500 μM peroxide (**Supplementary Figures S2B–S2D**). Quantification of the western blots confirmed that all acTrxR1 variants had a statistically significant reduction in the accumulation of higher molecular weight TrxR1 complexes compared to un-modified TrxR1 at all peroxide concentrations tested for acTrxR1[K307], from 100 to 400 μM H_2O_2 for acTrxR1[K200], and from 200 to 400 μM H_2O_2 for acTrxR1[K141] (**Supplementary Figure S3**).

Aspirin Acetylates TrxR1 and Provides Robust Resistance to Oxidative Damage

Protein acetylation can occur specifically in cells resulting from the activity of acetyltransferases, but also non-specifically

through interactions with acetyl donors including acetyl-CoA or certain drugs, such as aspirin (Tatham et al., 2017). We hypothesized that like specific or programmed acetylation of TrxR1, general or non-specific acetylation of the enzyme may also provide resistance to oxidative damage.

A previous report used incubated HeLa cells with a range of aspirin concentrations from 0.5 to 20 mM to generate aspirin-mediated and non-specific acetylation of many cellular proteins, including TrxR1 (Tatham et al., 2017). Thus, we used western blotting to detect acetylation of purified TrxR1 resulting from incubation with aspirin over a similar concentration range (**Figure 5**). To estimate the level of acetylated TrxR1, we used our site-specifically modified acTrxR1[K141] as control. The untreated and un-modified TrxR1 showed no reactivity with an anti-acK antibody, while robust detection was evident with the acetylated control sample (acTrxR1[K141]). Lower concentrations of aspirin did not lead to detectable acetylation, but at higher concentrations significant acTrxR1 is clearly present in the blot (**Figure 5A**). We then used the Coomassie stained gel to normalize the amount of protein loaded and determined the relative intensity of bands in the anti-acK immunoblot. We showed previously that acTrxR1[K141] has stoichiometric

FIGURE 5 | Aspirin acetylates TrxR1. TrxR1 was incubated with varying concentrations of aspirin (0–15 mM) for 1 h at 37°C. Following incubation, 0.31 μg of protein samples were separated on a 15% SDS-PAGE and visualized **(A)** after immunoblotting with an anti-acetyl-lysine antibody (α-AcK) or staining with Coomassie blue dye. The acTrxR1[K141] produced with genetic code expansion served as a positive control. Based on these data, we calculated the relative level of TrxR1 acetylation compared to the acTrxR1[K141] control **(B)**, which increased linearly with increasing aspirin concentration.

aspirin. Following incubation, the purified protein samples were digested with trypsin and analyzed by LC-MS/MS. We searched the spectra for the possibility of multiple modifications, including lysine acetylation. In the un-treated sample, we identified acetylation only at K307 according to a single peptide hit (**Supplementary Table S1**), perhaps due to acetylation by acetyl-phosphate in *E. coli* (Weinert et al., 2013). Based on our

FIGURE 6 | TrxR1 activity following aspirin incubation. **(A)** TrxR1 variants were incubated with 0–15 mM aspirin for 1 h at 37°C, followed by TrxR1 activity assays using DTNB. Negative controls containing no TrxR1 (-TrxR1) with 0 or 15 mM aspirin was also incubated for 1 h at 37°C, followed by the DTNB assay. **(B)** Initial velocity was calculated for TrxR1 activity based on the data in **(A)**. Statistical analysis comparing the activity at each aspirin concentration **(B)** to the untreated enzyme showed significant reductions in activity only at 10 and 15 mM aspirin (*$p < 0.05$; **$p < 0.005$). **(C)** TrxR1 was incubated with 15 mM aspirin for 1 h at 37°C, followed by incubation with 0–500 μM H_2O_2 for 1 h at 37°C. Following incubations, the initial velocity of TrxR1 activity was determined from TrxR1 activity assays and normalized to the initial velocity with 0 μM H_2O_2. Statistical analysis comparing the activity at each peroxide concentration show significant reductions in activity in the un-modified TrxR1 only and not in the aspirin treated enzyme (*$p < 0.05$). Error bars represent ± 1 standard deviation about the mean of three independent enzyme reactions.

incorporation of acK at the 141 site (Wright et al., 2018). Thus, based on comparison to the control, the relative level of acetylation following aspirin incubation reached a level ~65% of that observed in acTrxR1[K141] (**Figure 5B**).

We next compared the activity of TrxR1 and non-specifically acetylated TrxR1 following incubation with aspirin. In contrast to site-specifically acTrxR1 variants, aspirin incubation results in a slight decrease to ~70% of wild-type TrxR1 activity that was significantly lower only at the highest aspirin concentrations (>5 mM). Controls lacking enzyme (-TrxR1) and with or without aspirin showed no ability to catalytically reduce DTNB (**Figure 6A**). To test the ability of non-specifically acetylated TrxR1 to resist oxidative damage, unmodified TrxR1 was incubated with or without aspirin, followed by a second incubation of 1 h with increasing peroxide concentrations (0–500 μM H_2O_2) before DTNB reduction activity assays were conducted. In contrast to un-treated TrxR1, TrxR1 incubated with aspirin showed no statistically significant decrease in relative activity in response to any of the H_2O_2 concentrations tested (**Figure 6C**). The data suggest that aspirin mediated TrxR1 acetylation provides robust resistant to oxidative damage.

Location of TrxR1 Acetylation Sites Following Aspirin Treatment

We used tandem mass spectrometry to identify acetylation sites in purified wild-type TrxR1 following incubation with or without

FIGURE 7 | Mapping acetylation sites in TrxR1 to the quaternary structure. **(A)** Coverage map showing acetylation sites identified by LC-MS/MS following incubation of TrxR1 with the non-specific acetyl donor aspirin. **(B)** Lysine acetylation sites were mapped onto the TrxR1 tetramer structure (PDB codes 4KPR and 3EAN (Xu et al., 2015)): K176 (gold), K307 (red), K360 (green), and K307 (red). Sec551 (yellow) was modeled based on data from the structure 3ean (Rengby et al., 2009). The FAD co-factor is highlighted (grey). Close-up views show interactions with **(C)** K176, **(D)** K307 **(E)**, and K360. Structures were drawn using VMD (Humphrey et al., 1996).

observations in the acK immunoblot (**Figure 5**), we can conclude that the untreated TrxR1 has low levels of acetylation at this site.

In contrast, mass spectrometry on aspirin treated TrxR1 identified many new and different lysine acetylation sites supported with multiple, high quality peptides hits to the spectra (**Figure 7**; **Supplementary Figure S4**; **Supplementary Table S1**). These data provide strong evidence that aspirin incubation with TrxR1 leads to acetylation at K28, K31, K52, K88, K176, K307, K351, and K360 (**Figures 7A**; **Supplementary Figures S4**; **Supplementary Table S1**).

DISCUSSION

Interplay Between Acetylation and Oxidation in the Trx System

Many proteins are acetylated or hyper-acetylated in conditions of oxidative stress characterized by elevated levels of ROS (Santo-Domingo et al., 2020). Increased acetylation of lysine residues is a common response to oxidative stress documented in diverse proteins, including histones (Gu et al., 2013), FoxO Daitoku et al. (2011) and zinc-finger Wu et al. (2018) transcription factors, tRNA synthetases (Cao et al., 2017), and superoxide dismutase (Ozden et al., 2011). The acetylation status of the Trx system in conditions of oxidative stress is not yet completely characterized; however, increased acetylation of TrxR1 has been linked to oxidative stress in a mouse model of cardiomyopathy (Banerjee Mustafi et al., 2014).

Different components of the Trx system, including TrxR1, Trx, and Prx are acetylated at multiple lysine residues in mammalian cells (Choudhary et al., 2009; Hornbeck et al., 2015; Wu et al., 2015). We and others have found that acetylation increases the activity of TrxR1 (Wright et al., 2018), Trx1 (Folami Lamoke et al., 2011), and Prx1 (Parmigiani et al., 2008). Data from these studies suggest an emerging theme for how acetylation regulates the activity of the Trx system. Our work documented the ability of site-specific acetylation to increase TrxR1 activity by reducing oligomerization (Wright et al., 2018), and here we found acetylation also provided robust resistance to oxidative damage and peroxide-induced TrxR1 oligomer formation.

In studies of Trx, TrxR1, and Prx (Parmigiani et al., 2008), oxidation leads to the formation of low activity oligomers (Parmigiani et al., 2008; Xu et al., 2015), while acetylation provide resistance to oxidative damage and oligomerization (Parmigiani et al., 2008; Wright et al., 2018). Trx acetylation was linked to increased Trx activity in post-mortem diabetic retinas (Folami Lamoke et al., 2011). Much like Prx and TrxR1, Trx oligomerization is induced by oxidation, whereby Trx forms inactive dimers (Ren et al., 1993). The potential for Trx acetylation to counteract oxidation is not yet known. Acetylation of Prx prevents overoxidation and oligomerization (Parmigiani et al., 2008), while TrxR1 acetylation also provides resistance to oxidative inactivation **Figures 4**, **5** and prevents oligomerization (**Supplementary Figures S2, S3**); (Wright et al., 2018). Thus, acetylation may play a similar for Trx, which is an interesting area of future study.

The peroxiredoxins, Prx1 and Prx2, are both acetylated by histone acetyltransferase (HAT) *in vitro*, leading to increased activity and resistance to oxidation (Parmigiani et al., 2008). Human prostate cancer cells (LAPC4) readily acetylate Prx1 as Lys197, and these cells as well as another prostate line (LNCaP4) exposed to 25–100 μM H_2O_2 produced increasing amounts of high molecular mass Prx oligomers (Parmigiani et al., 2008), which are associated with reduced Prx peroxidase activity (Moon et al., 2005). In order to measure the ability of the acetylated enzyme to resist oxidative stress, the activity of acetylated recombinant Prx1 was measured in up to 2 mM H_2O_2. In the presence of 100 μM H_2O_2, a 20% increase in Prx1 activity was

associated with acK197 compared to the unmodified enzyme. Compared to our measurements with TrxR1 at the same peroxide concentration, we found 60–70% increased rate of activity in the acTrxR1 variants compared to the unmodified enzyme.

Here, we presented novel biochemical evidence that acetylation prevents TrxR1 activity loss in response to oxidative damage. Similarly to findings regarding Prx (Parmigiani et al., 2008), our studies suggest that site-specific acetylation increases TrxR1 activity Wright et al. (2018) and regulates TrxR1 oligomerization to prevent oxidative inactivation due to covalently linked multimer formation. TrxR1 forms these inactive cross-linked dimers due to oxidation at Trp114 in the dimer-dimer interface (Xu et al., 2015). Taken together, these studies highlight acetylation as a potent regulatory mechanism for Trx system activity that may provide protection against oxidative damage in cells at a time when the Trx system is most in need. The crosstalk between acetylation and oxidation is an open area of research for future work in the Trx system and beyond.

Non-specific Acetylation of Proteins in the Context of Oxidative Stress

We also demonstrated acetylation by the non-specific acetyl-donor aspirin provides TrxR1 with resistance to oxidative damage. We did not detect non-specific acetylation at the K141 or K200 sites, indicating these sites may not be accessible to aspirin. We previously found that acK307 reduces the formation of low activity tetramers and cross-linked inactive dimers (Wright et al., 2018). K176 is also in the dimer-dimer interface and participates in interactions with E300 of the opposing subunit (**Figure 7C**). K307 is localized close to K176 and in the dimer-dimer interface (**Figure 7D**). Both K176 and K307 participate in salt bridge interactions with the opposing subunit, and acetylation of these sites will weaken or eliminate key interactions that stabilize the TrxR1 tetramers. Other aspirin-mediated acetylation sites, K351 and K360, are located near the FAD cofactor binding site close to the active site of TrxR1 (**Figure 7**). Acetylation of K351 may disrupt interactions with the FAD co-factor, while acetylation of K360 may disrupt salt bridge interactions between K360 and E215 and E393, which are in the vicinity of the FAD co-factor (**Figure 7E**). Because the position and environment of FAD is critical for TrxR1 function, acetylation at K351 and K360 may be responsible for the marginal loss of TrxR1 activity we observed at high aspirin concentrations (**Figure 6B**).

Various metabolic compounds can non-enzymatically acetylate proteins, including acetyl-CoA (Weinert et al., 2014). Mitochondrial proteins show increased non-enzymatic acetylation by acetyl-CoA produced alongside energy metabolism (Wagner and Payne, 2013). In addition to metabolites, certain drugs, such as aspirin, can lead to acetylation of many proteins (Tatham et al., 2017). Indeed, the ability of aspirin to reduce inflammation and relieve pain is due to acetylation of the cyclooxygenases COX-1 and COX-2 in their active sites (Loll et al., 1995). Studies in mammalian cells demonstrated aspirin-mediated acetylation of p53 and many other cellular proteins using radiolabeled aspirin and

immunodetection (Alfonso et al., 2009). These studies provide a direct link between non-enzymatic acetylation of proteins under conditions that generate ROS or in response to certain medications. Non-enzymatic acetylation may play an important role in TrxR1 and other proteins that maintain the balance between oxidative damage and antioxidant defense.

Relevance of Trx System Acetylation to Disease

Resistance to oxidative damage is of particular interest in cancer biology and therapeutics (Klaunig, 2018). The Trx system is overactive in many cancer cells (Selenius et al., 2010; Dong et al., 2016; Roh et al., 2017). The reductive power of the Trx system provides resistance to certain chemotherapies relying on the generation of oxidative stress (Roh et al., 2017). For example, the anti-cancer agent RITA induces oxidative stress in HCT116 colon cancer cells leading to increased TrxR1 tetramer and cross-linked dimer formation (Xu et al., 2015). Newly developed direct inhibitors of TrxR1 have shown efficacy in combination with the cancer chemotherapeutic cisplatin in colon cancer cells (Zhang et al., 2019). Various other small compounds targeting TrxR1 activity are also of interest for cancer treatments (Zhang et al., 2017). Fascinatingly, pre-treatment of the cells with a ROS scavenger and acetate source (Raftos et al., 2007), N-acetyl-l-cysteine, reduced ROS generation, cell death and DNA damage associated with TrxR1 inhibition.

Higher levels of TrxR1 in the serum correlates with shortened overall survival in patients with non-small cell lung cancer (Chen et al., 2017). Proteomic studies of similar cell lines have consistently identified acTrxR1 in HeLa cells, A549 cells and related non-small cell lung cancer cell lines (Choudhary et al., 2009; Weinert et al., 2013; Hornbeck et al., 2015; Wu et al., 2015). Together our data and other studies suggest that TrxR1 acetylation is a potential mechanism to maintain higher Trx system activity. These findings have relevance for the role of TrxR1 acetylation in cancers with poor survival Chen et al. (2017) as well as chemotherapeutic resistance (Iwasawa et al., 2011; Raninga et al., 2016).

CONCLUSION

We demonstrated both site-specific and non-specific acetylation of TrxR1 provides robust resistance to oxidative damage even under oxidizing conditions that represent the full range of ROS

experienced by TrxR1 in cells. Because TrxR1 plays a vital role in resolving oxidative damage on Trx and other target proteins, TrxR1 and the activity of the Trx system is most important to the cell under conditions of oxidative stress or chemotherapeutic assault. Acetylation of TrxR1 provides a route to increase redox activity, enabling TrxR1 to resist oxidative damage associated with the very reactive oxygen species that the enzyme is tasked to resolve.

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 at: PRIDE Proteome Xchange, accession number PXD027882.

AUTHOR CONTRIBUTIONS

DW and NP performed experiments and collected data. DW and PO'D designed the study, analysed the data, and wrote and edited the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.747236/full#supplementary-material

REFERENCES

Arnér, E. S. J. (2009). Focus on Mammalian Thioredoxin Reductases - Important Selenoproteins with Versatile Functions. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1790 (6), 495–526. doi:10.1016/j.bbagen.2009.01.014

Arnér, E. S. J., Zhong, L., and Holmgren, A. (1999). Preparation and Assay of Mammalian Thioredoxin and Thioredoxin Reductase. *Methods Enzymol.* 300, 226–239. doi:10.1016/s0076-6879(99)00129-9

Bae, Y. S., Oh, H., Rhee, S. G., and Yoo, Y. D. (2011). Regulation of Reactive Oxygen Species Generation in Cell Signaling. *Mol. Cell* 32 (6), 491–509. doi:10.1007/s10059-011-0276-3

Banerjee Mustafi, S., Grose, J. H., Zhang, H., Pratt, G. W., Sadoshima, J., Christians, E. S., et al. (2014). Aggregate-prone R120GCRYAB Triggers Multifaceted Modifications of the Thioredoxin System. *Antioxid. Redox Signaling* 20 (18), 2891–2906. doi:10.1089/ars.2013.5340

Bhat, L. F., Srivenugopal, K. S., and Bhat, G. J. (2009). Does Aspirin Acetylate Multiple Cellular Proteins? (Review). *Mol. Med. Rep.* 2 (4), 533–537. doi:10.3892/mmr_00000132

Cao, X., Li, C., Xiao, S., Tang, Y., Huang, J., Zhao, S., et al. (2017). Acetylation Promotes TyrRS Nuclear Translocation to Prevent Oxidative Damage. *Proc. Natl. Acad. Sci. USA.* 114 (4), 687–692. doi:10.1073/pnas.1608488114

Chen, G., Chen, Q., Zeng, F., Zeng, L., Yang, H., Xiong, Y., et al. (2017). The Serum Activity of Thioredoxin Reductases 1 (TrxR1) Is Correlated with the Poor

Prognosis in EGFR Wild-type and ALK Negative Non-small Cell Lung Cancer. *Oncotarget* 8 (70), 115270–115279. doi:10.18632/oncotarget.23252

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., et al. (2009). Lysine Acetylation Targets Protein Complexes and Co-regulates Major Cellular Functions. *Science* 325 (5942), 834–840. doi:10.1126/science.1175371

Clément, M. V., Ponton, A., and Pervaiz, S. (1998). Apoptosis Induced by Hydrogen Peroxide Is Mediated by Decreased Superoxide Anion Concentration and Reduction of Intracellular Milieu. *FEBS Lett.* 440 (1-2), 13–18. doi:10.1016/s0014-5793(98)01410-0

Daitoku, H., Sakamaki, J.-i., and Fukamizu, A. (2011). Regulation of FoxO Transcription Factors by Acetylation and Protein-Protein Interactions. *Biochim. Biophys. Acta (Bba) - Mol. Cell Res.* 1813 (11), 1954–1960. doi:10.1016/j.bbamcr.2011.03.001

Dong, C., Zhang, L., Sun, R., Liu, J., Yin, H., Li, X., et al. (2016). Role of Thioredoxin Reductase 1 in Dysplastic Transformation of Human Breast Epithelial Cells Triggered by Chronic Oxidative Stress. *Sci. Rep.* 6, 36860. doi:10.1038/srep36860

Fan, C., Xiong, H., Reynolds, N. M., and Söll, D. (2015). Rationally Evolving tRNAPylfor Efficient Incorporation of Noncanonical Amino Acids. *Nucleic Acids Res.* 43 (22), e156. doi:10.1093/nar/gkv800

Folami Lamoke, C. S., Maraschi, A., Fan, F., Montemari, A., Marcus, D. M., Martin, P. M., et al. (2011). Epigenetic Regulation of Endogenous Antioxidants in the Diabetic Retina. *Invest. Opthalmology Vis. Sci.* 52 (14), 262–272. doi:10.1002/jcp.21852

Gu, X., Sun, J., Li, S., Wu, X., and Li, L. (2013). Oxidative Stress Induces DNA Demethylation and Histone Acetylation in SH-Sy5y Cells: Potential Epigenetic Mechanisms in Gene Transcription in Aβ Production. *Neurobiol. Aging* 34 (4), 1069–1079. doi:10.1016/j.neurobiolaging.2012.10.013

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

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* 43 (Database issue), D512–D520. doi:10.1093/nar/gku1267

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graphics* 14 (1), 3327–3388. doi:10.1016/0263-7855(96)00018-5

Kim, H.-Y., and Gladyshev, V. N. (2007). Methionine Sulfoxide Reductases: Selenoprotein Forms and Roles in Antioxidant Protein Repair in Mammals. *Biochem. J.* 407 (3), 321–329. doi:10.1042/BJ20070929

Klaunig, J. E. (2019). Oxidative Stress and Cancer. *Cpd* 24 (40), 4771–4778. doi:10.2174/1381612825666190215121712

Lamoke, F., Stampley, C., Maraschi, A., Fan, F., Montemari, A., Marcus, D. M., et al. (2011). Epigenetic Regulation of Endogenous Antioxidants in the Diabetic Retina. *Invest. Ophthalmol. Vis. Sci.* 52 (4448), 262–272. doi:10.1002/jcp.21852

Loll, P. J., Picot, D., and Garavito, R. M. (1995). The Structural Basis of Aspirin Activity Inferred from the crystal Structure of Inactivated Prostaglandin H2 Synthase. *Nat. Struct. Mol. Biol.* 2 (8), 637–643. doi:10.1038/nsb0895-637

Lu, J., and Holmgren, A. (2012). Thioredoxin System in Cell Death Progression. *Antioxid. Redox Signaling* 17 (12), 1738–1747. doi:10.1089/ars.2012.4650

Moon, J. C., Hah, Y.-S., Kim, W. Y., Jung, B. G., Jang, H. H., Lee, J. R., et al. (2005). Oxidative Stress-dependent Structural and Functional Switching of a Human 2-Cys Peroxiredoxin Isotype II that Enhances HeLa Cell Resistance to H2O2-Induced Cell Death. *J. Biol. Chem.* 280 (31), 28775–28784. doi:10.1074/jbc.M505362200

Mustacich, D., and Powis, G. (2000). Thioredoxin Reductase. *Biochem. J.* 346, 1–8. doi:10.1042/bj3460001

Ozden, O., Park, S.-H., Kim, H.-S., Jiang, H., Coleman, M. C., Spitz, D. R., et al. (2011). Acetylation of MnSOD Directs Enzymatic Activity Responding to Cellular Nutrient Status or Oxidative Stress. *Aging* 3 (2), 102–107. doi:10.18632/aging.100291

Parmigiani, R. B., Xu, W. S., Venta-Perez, G., Erdjument-Bromage, H., Yaneva, M., Tempst, P., et al. (2008). HDAC6 Is a Specific Deacetylase of Peroxiredoxins and Is Involved in Redox Regulation. *Proc. Natl. Acad. Sci.* 105 (28), 9633–9638. doi:10.1073/pnas.0803749105

Raftos, J. E., Whillier, S., Chapman, B. E., and Kuchel, P. W. (2007). Kinetics of Uptake and Deacetylation of N-Acetylcysteine by Human Erythrocytes. *Int. J. Biochem. Cell Biol.* 39 (9), 1698–1706. doi:10.1016/j.biocel.2007.04.014

Raninga, P. V., Di Trapani, G., Vuckovic, S., and Tonissen, K. F. (2016). TrxR1 Inhibition Overcomes Both Hypoxia-Induced and Acquired Bortezomib Resistance in Multiple Myeloma through NF-Κβ Inhibition. *Cell Cycle* 15 (4), 559–572. doi:10.1080/15384101.2015.1136038

Ren, X., Bjoernstedt, M., Shen, B., Ericson, M. L., and Holmgren, A. (1993). Mutagenesis of Structural Half-Cystine Residues in Human Thioredoxin and Effects on the Regulation of Activity by Selenodiglutathione. *Biochemistry* 32 (37), 9701–9708. doi:10.1021/bi00088a023

Rengby, O., Cheng, Q., Vahter, M., Jörnvall, H., and Arnér, E. S. J. (2009). Highly Active Dimeric and Low-Activity Tetrameric Forms of Selenium-Containing Rat Thioredoxin Reductase 1. *Free Radic. Biol. Med.* 46 (7), 893–904. doi:10.1016/j.freeradbiomed.2008.12.017

Roh, J.-L., Jang, H., Kim, E. H., and Shin, D. (2017). Targeting of the Glutathione, Thioredoxin, and Nrf2 Antioxidant Systems in Head and Neck Cancer. *Antioxid. Redox Signaling* 27 (2), 106–114. doi:10.1089/ars.2016.6841

Santo-Domingo, J., Dayon, L., and Wiederkehr, A. (2020). Protein Lysine Acetylation: Grease or Sand in the Gears of β-Cell Mitochondria? *J. Mol. Biol.* 432 (5), 1446–1460. doi:10.1016/j.jmb.2019.09.011

Schweizer, U., and Fradejas-Villar, N. (2016). Why 21? the Significance of Selenoproteins for Human Health Revealed by Inborn Errors of Metabolism. *FASEB j.* 30 (11), 3669–3681. doi:10.1096/fj.201600424

Selenius, M., Rundlöf, A.-K., Olm, E., Fernandes, A. P., and Björnstedt, M. (2010). Selenium and the Selenoprotein Thioredoxin Reductase in the Prevention, Treatment and Diagnostics of Cancer. *Antioxid. Redox Signaling* 12 (7), 867–880. doi:10.1089/ars.2009.2884

Serrão, V. H. B., Silva, I. R., da Silva, M. T. A., Scortecci, J. F., de Freitas Fernandes, A., and Thiemann, O. H. (2018). The Unique tRNASec and its Role in Selenocysteine Biosynthesis. *Amino Acids* 50 (9), 1145–1167. doi:10.1007/s00726-018-2595-6

Shu, N., Cheng, Q., Arnér, E. S. J., and Davies, M. J. (2020). Inhibition and Crosslinking of the Selenoprotein Thioredoxin Reductase-1 by P-Benzoquinone. *Redox Biol.* 28, 101335. doi:10.1016/j.redox.2019.101335

Sies, H. (2017). Hydrogen Peroxide as a central Redox Signaling Molecule in Physiological Oxidative Stress: Oxidative Eustress. *Redox Biol.* 11, 613–619. doi:10.1016/j.redox.2016.12.035

Tatham, M. H., Cole, C., Scullion, P., Wilkie, R., Westwood, N. J., Stark, L. A., et al. (2017). A Proteomic Approach to Analyze the Aspirin-Mediated Lysine Acetylome. *Mol. Cell Proteomics* 16 (2), 310–326. doi:10.1074/mcp.O116.065219

Uzawa, S., Yamano, Y., Takiguchi, Y., Tanzawa, H., Tatsumi, K., and Uzawa, K. (2011). Upregulation of Thioredoxin Reductase 1 in Human Oral Squamous Cell Carcinoma. *Oncol. Rep.* 25 (3), 637–644. doi:10.3892/or.2010.1131

Wagner, G. R., and Payne, R. M. (2013). Widespread and Enzyme-independent Nε-Acetylation and Nε-Succinylation of Proteins in the Chemical Conditions of the Mitochondrial Matrix*. *J. Biol. Chem.* 288 (40), 29036–29045. doi:10.1074/jbc.M113.486753

Wang, L., Yang, Z., Fu, J., Yin, H., Xiong, K., Tan, Q., et al. (2012). Ethaselen: a Potent Mammalian Thioredoxin Reductase 1 Inhibitor and Novel Organoselenium Anticancer Agent. *Free Radic. Biol. Med.* 52 (5), 898–908. doi:10.1016/j.freeradbiomed.2011.11.034

Weinert, B. T., Iesmantavicius, V., Moustafa, T., Schölz, C., Wagner, S. A., Magnes, C., et al. (2014). Acetylation Dynamics and Stoichiometry in S Accharomyces Cerevisiae. *Mol. Syst. Biol.* 10, 716. doi:10.1002/msb.134766

Weinert, B. T., Schölz, C., Wagner, S. A., Iesmantavicius, V., Su, D., Daniel, J. A., et al. (2013). Lysine Succinylation Is a Frequently Occurring Modification in Prokaryotes and Eukaryotes and Extensively Overlaps with Acetylation. *Cell Rep.* 4 (4), 842–851. doi:10.1016/j.celrep.2013.07.024

Wright, D. E., Altaany, Z., Bi, Y., Alperstein, Z., and O'Donoghue, P. (2018). Acetylation Regulates Thioredoxin Reductase Oligomerization and Activity. *Antioxid. Redox Signaling* 29 (4), 377–388. doi:10.1089/ars.2017.7082

Wu, C.-C., Lee, P.-T., Kao, T.-J., Chou, S.-Y., Su, R.-Y., Lee, Y.-C., et al. (2018). Upregulation of Znf179 Acetylation by SAHA Protects Cells against Oxidative Stress. *Redox Biol.* 19, 74–80. doi:10.1016/j.redox.2018.08.001

Wu, Q., Cheng, Z., Zhu, J., Xu, W., Peng, X., Chen, C., et al. (2015). Suberoylanilide Hydroxamic Acid Treatment Reveals Crosstalks Among Proteome, Ubiquitylome and Acetylome in Non-small Cell Lung Cancer A549 Cell Line. *Sci. Rep.* 5, 9520. doi:10.1038/srep09520

Xu, J., Eriksson, S. E., Cebula, M., Sandalova, T., Hedström, E., Pader, I., et al. (2015). The Conserved Trp114 Residue of Thioredoxin Reductase 1 Has a Redox Sensor-like Function Triggering Oligomerization and Crosslinking upon Oxidative Stress Related to Cell Death. *Cell Death Dis.* 6, e1616. doi:10.1038/cddis.2014.574

Zhang, J., Li, X., Han, X., Liu, R., and Fang, J. (2017). Targeting the Thioredoxin System for Cancer Therapy. *Trends Pharmacol. Sci.* 38 (9), 794–808. doi:10.1016/j.tips.2017.06.001

Zhang, T., Zheng, P., Shen, X., Shao, R., Wang, B., Shen, H., et al. (2019). Curcuminoid WZ26, a TrxR1 Inhibitor, Effectively Inhibits colon Cancer Cell Growth and Enhances Cisplatin-Induced Cell Death through the Induction of ROS. *Free Radic. Biol. Med.* 141, 93–102. doi:10.1016/j.freeradbiomed.2019.06.005

Towards Engineering an Orthogonal Protein Translation Initiation System

Byeong Sung Lee[1], Woon Jong Choi[1], Sang Woo Lee[1], Byoung Joon Ko[2] and Tae Hyeon Yoo[1,3]*

[1]Department of Molecular Science and Technology, Ajou University, Suwon, South Korea, [2]School of Biopharmaceutical and Medical Sciences, Sungshin Women's University, Seoul, South Korea, [3]Department of Applied Chemistry and Biological Engineering, Ajou University, Suwon, South Korea

In the last two decades, methods to incorporate non-canonical amino acids (ncAAs) into specific positions of a protein have advanced significantly; these methods have become general tools for engineering proteins. However, almost all these methods depend on the translation elongation process, and strategies leveraging the initiation process have rarely been reported. The incorporation of a ncAA specifically at the translation initiation site enables the installation of reactive groups for modification at the N-termini of proteins, which are attractive positions for introducing abiological groups with minimal structural perturbations. In this study, we attempted to engineer an orthogonal protein translation initiation system. Introduction of the identity elements of *Escherichia coli* initiator tRNA converted an engineered *Methanococcus jannaschii* tRNA[Tyr] into an initiator tRNA. The engineered tRNA enabled the site-specific incorporation of O-propargyl-l-tyrosine (OpgY) into the amber (TAG) codon at the translation initiation position but was inactive toward the elongational TAG codon. Misincorporation of Gln was detected, and the engineered system was demonstrated only with OpgY. We expect further engineering of the initiator tRNA for improved activity and specificity to generate an orthogonal translation initiation system.

Keywords: translation initiation, non-canonical amino acid, initiator tRNA, amber codon, Methanococcus jannaschii tRNA

INTRODUCTION

Site-specific modification of the N-terminus of a protein is an attractive strategy to introduce unnatural groups with minimal effects on original protein functions. Protein termini tend to be exposed and flexible but not buried in the core (Christopher and Baldwin 1996; Jacob and Unger 2007). The N-terminal α-amino group is a good nucleophile, and its reactivity has been utilized to modify the N-terminus of proteins specifically. While the pKa of a lysine side-chain is around 10.5, the pKa values of the N-terminal amines have been reported between 6.8 and 9.1, with an average of 7.7 ± 0.5 (Serada and Mant, 1993). The reactivities of N-terminal and lysine side chain amines can be different at a given pH, and several chemical reactions have been reported as methods to modify the N-terminal amine groups specifically (Rosen and Francis 2017). Enzymes have been found to catalyze the peptide-bond formation reaction, and some of them have been used to modify proteins at their N-termini, including sortase (Mao et al., 2004, Proft 2010), intein (Evans and Xu 1999; Xu and Evans 2001), and proteases (Liebscher et al., 2014; Weeks and Wells 2020).

Biological methods for incorporating non-canonical amino acids (ncAAs) into proteins have become invaluable tools for protein engineering (Wang and Schultz 2004). In particular, ncAAs with

reactivities can provide sites for protein modification (Wang et al., 2009; Wals and Ovaa 2014; Chin 2017). *Escherichia coli* wild-type and variant methionyl-tRNA synthetase (MetRS) can charge the *E. coli* initiator methionyl-tRNA (*Ec*-tRNA^{fMet}) with isostructural analogs of methionine (Met), and several analogs with orthogonal reactivities have been successfully installed at the N-termini of proteins (Link et al., 2003; Johnson et al., 2010; Agostini et al., 2017; Tharp et al., 2020; Pagar et al., 2021). However, they were also incorporated into the internal Met positions, and thus, site-specific modification at the N-terminal position was not achieved, besides cases where there was no internal Met residue. Interestingly, *Ec*-tRNA^{fMet} charged with ncAA is translationally active for protein synthesis initiation in eukaryotes (Ngo et al., 2013). The system was demonstrated to tag the proteome N-termini but was unsuitable for producing homogeneous recombinant proteins because either ncAA or Met occupied the N-terminal positions. Budisa et al. proposed a clever strategy to incorporate azidohomoalanine at the protein synthesis initiation position by deleting the elongator tRNA^{Met} genes from the *E. coli* genome and introducing an orthogonal pair of MetRS/tRNA^{Met} from *Sulfolobus acidocaldarius* (Simone, Acevedo-Rocha et al., 2016). However, the system was partially successful because of the incomplete orthogonality of the system.

Orthogonal pairs of aminoacyl-tRNA synthetase (aaRS) and tRNA have been developed to introduce diverse ncAAs into proteins in a site-specific manner in various organisms (Wang and Wang 2012; Xiao et al., 2013; Smolskaya and Andreev 2019). These methods were designed based on codons that are not generally used for coding amino acids (such as stop codons (Liu and Schultz 2010) and four-base codons (Hohsaka et al., 2001; Anderson et al., 2004; Wang et al., 2012; Lee et al., 2017)). Strategies have also been reported to incorporate ncAAs into sense codons by modifying orthogonal pairs (Lee et al., 2015; Mukai et al., 2015). However, all orthogonal tRNAs are elongator tRNAs, and thus, cannot be used to install ncAAs at the N-termini of proteins. Recently, Söll et al. reported an engineered *Ec*-tRNA^{fMet} that enabled the initiation of protein synthesis with ncAAs (Tharp et al., 2020). The initiator tRNA, named itRNATy2, had an identity element in *Methanococcus jannaschii* tRNA^{Tyr} (*Mj*-tRNA^{Tyr}) by mutations of A72G and the anticodon for the amber codon (CAU→CUA). The initiator tRNA was a substrate for *M. jannaschii* tyrosyl-tRNA synthetase (*Mj*-TyrRS), and the tRNAs charged with ncAAs could be used for *E. coli* protein synthesis initiation. Despite this success, the initiator tRNA had the C1:G72 base pair and could not distinguish between the initiation and elongation amber codons.

It has been shown that the introduction of the main determinants of *Ec*-tRNA^{fMet} into elongator tRNAs, such as glutaminyl-tRNA and methionyl-tRNA, could convert them into tRNAs, enabling initiation of protein synthesis in *E. coli* (Varshney and RajBhandary 1990; Lee et al., 1991; Varshney et al., 1993). Based on the achievements of these studies, we aimed to engineer *Mj*-tRNA^{Tyr} into an initiator tRNA, which was intended to work only for an amber codon at the translation initiation position, but not for an internal amber codon

(**Figure 1**). Identity elements for interactions with factors involved in the initiation of protein syntheses, such as methionyl-tRNA transformylase (MTF), the P-site of the 30S ribosomal subunit, and initiation factor-2 (IF-2), were introduced into *Mj*-tRNA^{Tyr}. The engineered *Mj*-tRNA was active toward the initiation amber codon, but did not support the coding of internal amber codons. We believe that the results obtained in this study open a new route for developing an orthogonal translation initiation system to incorporate ncAAs at the N-termini of proteins.

MATERIALS AND METHODS

Plasmid Construction

The plasmids, primers, and proteins used in this study are shown in **Table 1**; **Supplementary Tables S1, S2**, respectively. The synthesized initiator tRNA genes, including the ProK promoter and terminator (*Mj*-itRNA-1 and *Mj*-itRNA-2; the DNA sequences are shown in **Supplementary Table S3**) were cloned into the modified pEVOL (Young et al., 2010) plasmid, in which one copy of the azidophenylalaninyl-tRNA synthetase gene (AzF-RS) is located under the AraBAD promoter (Lee et al., 2015), using the *Apa*LI and *Xho*I sites. The resulting plasmids were named pSEPL773 for *Mj*-itRNA-1 and pSPEL541 for *Mj*-itRNA-2. The methionyl-tRNA transformylase (MTF) gene was amplified (primers 1 and 6) from the chromosome of *E. coli* DH10β and then cloned into the plasmid containing itRNA-1 (pSPEL773) or itRNA-2 (pSPEL541) using the *Nde*I and *Pst*I sites. The two *Pst*I sites present in the MTF gene were removed by assembly PCR using primer pairs (primers 2 and 3 for the first *Pst*I site; primers 4 and 5 for the second *Pst*I site). The MTF gene was located under the glnS promoter. The resulting plasmids were named pSPEL527 (MTF/*Mj*-itRNA-1) and pSPEL528 (MTF/*Mj*-itRNA-2), respectively. The initiator factor-2 (IF-2) gene was amplified (primers 7 and 14) from the chromosome of *E. coli* DH10β and then cloned into the plasmid containing itRNA-1 (pSPEL773) or itRNA-2 (pSPEL541) using the *Nde*I and *Pst*I sites; the three *Pst*I sites present in the IF-2 gene were removed by assembly PCR using two pairs of primers (primers 8 and 9 for the first *Pst*I site; primer 10 and 11 for the second *Pst*I site; primers 12 and 13 for the third *Pst*I site). The resulting plasmids were named pSPEL780 (IF-2/*Mj*-itRNA-1) and pSPEL562 (IF-2/*Mj*-itRNA-2), respectively. Plasmids expressing both MTF and IF-2 were constructed by cloning the IF-2 gene with a ribosome-binding sequence (Kinneer et al., 2019) into the *Pst*I site of pSPEL527 or pSPEL528, and the orientation of the RBS-IF-2 genes was confirmed by DNA sequencing. The resulting plasmids were named pSPEL781 (MTF/IF-2/*Mj*-itRNA-1) and pSPEL563 (MTF/IF-2/*Mj*-itRNA-2), respectively.

A DNA double helix for a multiple cloning site (*Nco*I-*Bam*HI-*Hind*III-*Not*I-*Xho*I-His$_6$ Tag) prepared by annealing primers 15 and 16 was phosphorylated using T4 polynucleotide kinase and then ligated into the pBbE6a (Lee et al., 2011) plasmid digested with *Nde*I and *Xho*I. The RBS-GFP gene with the TAG codon at the translation initiation site was generated by PCR amplification

FIGURE 1 | Schematic representation of the strategy for site-specific incorporation of a non-canonical amino acid (ncAA) into the start position of protein translation. An engineered initiator, tRNA$_{CUA}$ charged with a ncAA is active toward the protein translation initiation with the amber (TAG) codon, but inactive for the elongation process.
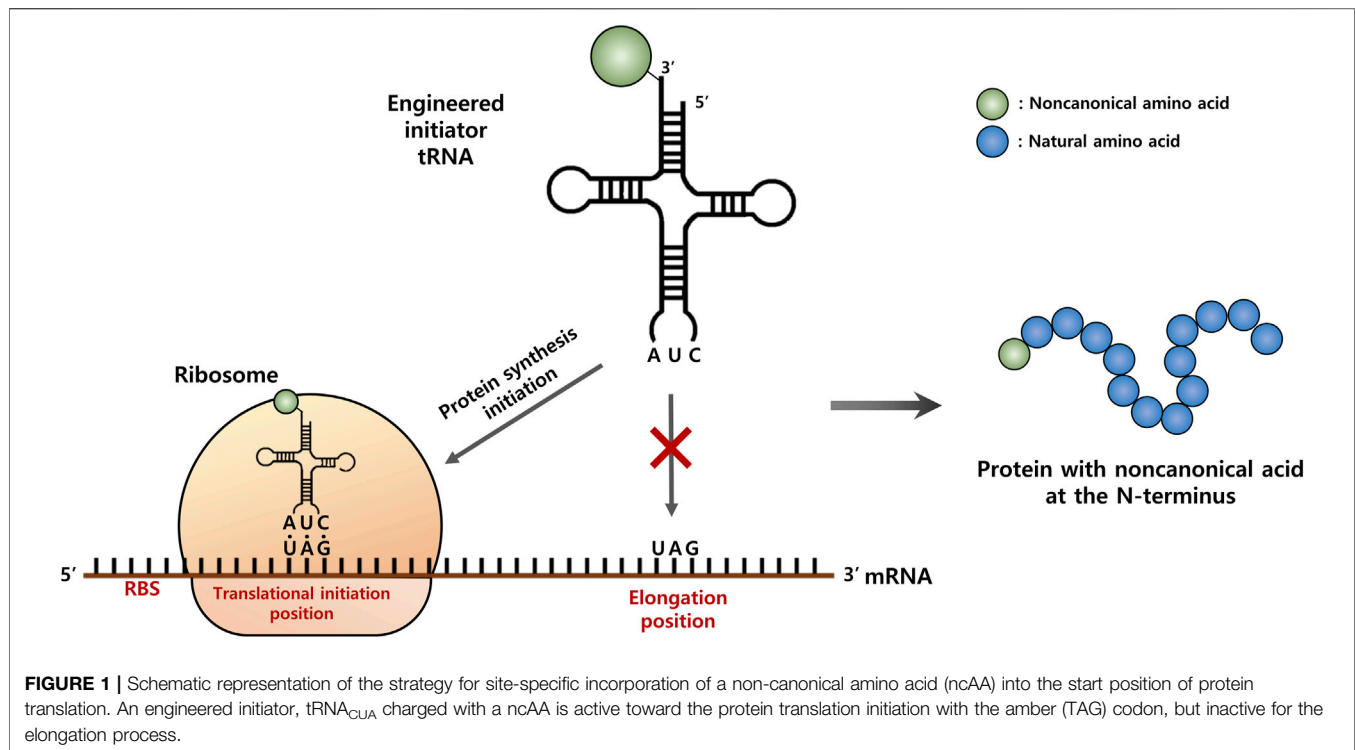
TABLE 1 | Plasmids used in this study.

Name	Characteristics[a]	Source
pEVOL	*Mj*-tRNA^{Tyr}, AzFRS, Cm^R, p15A ori	Young et al. (2010)
pSPEL773	pEVOL-*Mj*-itRNA-1, AzFRS, Cm^R, p15A ori	This study
pSPEL541	pEVOL-*Mj*-itRNA-2, AzFRS, Cm^R, p15A ori	This study
pSPEL527	pEVOL-*Mj*-itRNA-1, AzFRS, MTF, Cm^R, p15A ori	This study
pSPEL528	pEVOL-*Mj*-itRNA-2, AzFRS, MTF, Cm^R, p15A ori	This study
pSPEL780	pEVOL-*Mj*-itRNA-1, AzFRS, IF-2, Cm^R, p15A ori	This study
pSPEL562	pEVOL-*Mj*-itRNA-2, AzFRS, IF-2, Cm^R, p15A ori	This study
pSPEL781	pEVOL-*Mj*-itRNA-1, AzFRS, MTF, IF-2, Cm^R, p15A ori	This study
pSPEL563	pEVOL-*Mj*-itRNA-2, AzFRS, MTF, IF-2, Cm^R, p15A ori	This study
pQE-80L	Expression vector, Amp^R, ColE1 ori	Qiagen
pSPEL542	pQE-80L-TAG-Z domain, Amp^R, ColE1 ori	This study
pBbE6a	Expression vector, Amp^R, ColE1 ori	Lee et al. (2011)
pSPEL530	pBbE6a-TAG-11.3.3, ColE1 ori	This study
pGEX-4T-1	Expression vector with GST, Amp^R, pBR322 ori	Amersham Bioscience
pSPEL236	pGEX-4T-1-GST-TAG-Z domain, Amp^R, pBR322 ori	This study
pBbS2K-ProRS	pBbS2K-Prolyl tRNA synthetase, Kan^R, SC101 ori	Lee et al. (2015)

[a]*Mj, Methanococcus jannaschii; AzFRS, azidophenylalanyl tRNA synthetase; Cm, chloramphenicol; Amp, ampicillin; Kan, kanamycin; MTF, methionyl-tRNA transformylase; IF-2, initiation factor-2.*

using the 11.3.3 GFP gene (Yoo et al., 2007) as a template with primers 17 and 18, and the product was cloned into the modified pBbE6a mentioned above using the *Eco*RI and *Hind*III sites, resulting in pSPEL530. The synthesized gene of the Z domain (the amino acid sequence in **Supplementary Table S2**) with an amber (TAG) codon at the translation initiation site was cloned into pQE-80L (Qiagen) using the *Eco*RI and *Hind*III sites, resulting in pSEPL542. The Z domain gene was amplified using pSPEL542 as a template with primers 19 and 20; the product was cloned into pGEX-4T-1 (Amersham Bioscience)

*Eco*RI and *Xho*I sites. To introduce an amber (TAG) codon between the glutathione-S-transferase (GST) and Z domain, an annealed DNA double helix of primers 21 and 22 was phosphorylated using T4 polynucleotide kinase and then cloned into the pGEX-4T-1 with the Z domain gene using the *Bam*HI and *Eco*RI sites, resulting in pSPEL236.

Non-Canonical Amino Acid Incorporation

The plasmid coding Z domain (pSPEL542) or GFP (pSPEL530) with an amber (TAG) codon at the translation initiation position

was co-transformed with one of the pEVOL derivatives (pSPEL731, 541, 527, 528, 780, or 562) and pBbS2K-ProRS (Lee et al., 2015) encoding *E. coli* prolyl-tRNA synthetase under the tetracycline-inducible promoter into *E. coli* DH10β. The cells were grown in 2xYT media containing 200 μg/ml ampicillin, 34 μg/ml chloramphenicol, and 35 μg/ml kanamycin at 37°C until the OD_{600} reached 0.5. To induce AzF-RS and prolyl-tRNA synthetase expression, 0.2% L-arabinose and 50 nM anhydrotetracycline were added to the culture. When OD_{600} reached 1.0, 1 mM isopropyl β-D-thiogalatoside and 1 mM ncAA were added to express the model protein. After 3 h, the cells were harvested by centrifugation at 14,000 × g at 4 C for 20 min, and the cell pellets were stored at −20°C until use.

Copper (I)-Catalyzed Click Reaction and Western Blotting

The cell pellets were resuspended in 1% SDS solution and heated at 95°C for 5 min for lysis. The supernatant obtained after centrifugation at 14,000 × g for 5 min was subjected to a copper (I)-catalyzed click reaction with biotin-PEG_3-azide (Click Chemistry Tools) following the method described (Hong et al., 2009). The reaction mixtures were analyzed by western blotting using a streptavidin-horseradish peroxidase (HRP) conjugate (Click Chemistry Tools) and an anti-His_5-HRP conjugate (Sigma-Aldrich, St. Louis, MO).

Protein Purification

The Z-domain protein with an N-terminal His6 tag was purified using a Ni-immobilized resin (Clontech, Mountain View, CA) under native conditions, following the manufacturer's instructions. The cell pellets stored at −20 C were resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, pH 7.4) and incubated on ice for 30 min with 50 μg/ml lysozyme. After sonication, the lysed cells were centrifuged at 9,300 × g at 4°C for 1 h. The supernatant was incubated at 4°C for 1 h with Ni resin, pre-equilibrated with lysis buffer. The resin was loaded onto a gravity flow column (Thermo Fisher Scientific, Waltham, MA) and washed three times with washing buffer (50 mM sodium phosphate, 300 mM sodium chloride, 40 mM imidazole, pH 7.4). The protein was eluted using elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 300 mM imidazole, pH 7.4). The protein solutions were buffer-exchanged with phosphate buffer saline (PBS) solution (10 mM KH2PO4, 150 mM NaCl, pH 7.4) using a centrifugal filter unit (Millipore, 3000 MWCO).

Mass Spectrometry

The intact masses of the proteins were analyzed using a Waters ACQUITY I class UPLC system (Milford, MA) with an ACQUITY UPLC Protein BEH C4 column (2.1 mm × 100 mm, 1.7 μm particle size; Waters). The mobile phases were 0.1% formic acid in water (eluent A) and 0.1% formic acid in

acetonitrile (eluent B). The gradient applied was: 0–3 min, 5% eluent B; 3–13 min, linear increase to 50% eluent B at 0.2 ml/min. The eluent was injected into a Thermo Orbitrap Elite (Thermo Fisher Scientific, Waltham, MA) and ionized with an electrospray source. MS spectra were acquired in the mass range of 400–2,000 m/z and 120,000 resolution at m/z 200. The deconvoluted mass spectra were generated using Protein Deconvolution 2.0 (Thermo Fisher Scientific, Waltham, MA).

N-Terminal Sequencing

The purified Z-domain protein was transferred to a PVDF membrane (0.45 μm pore size; Pall), and the membrane was washed with ultrapure water and dried in air. The membrane was subjected to an N-terminal sequencing analysis using Procise® LC492 Protein Sequencing System (Applied Biosystems, Waltham, MA). The retention times of phenylthiohydantoin (PTH)-amino acids were compared with those of the standards (Tokyo Chemical Industry, Tokyo). PHT-OpgY was synthesized following a previously published method (Steiman et al., 1985).

RESULTS

Introduction of the identity elements of Ec-tRNA^{fMet} converted Mj-tRNA^{Tyr} into an initiator tRNA

The two main identity elements of the *E. coli* initiator tRNA (*Ec*-tRNA^{fMet}) are 1) the absence of a Watson-Crick base pair between positions 1 and 72 in the acceptor stem (pink in **Figure 2A**) and 2) three consecutive G:C base pairs in the anticodon stem (green in **Figure 2A**). The C1:A72 mismatch plays a critical role in the interaction with MTF, and the 3G: C pairs are important for targeting the initiator tRNA to the P-site of the 30S ribosomal subunit (Kozak 1999; Laursen et al., 2005; Louise. et al., 2009). It was previously demonstrated that an *E. coli* glutaminyl-tRNA (*Ec*-tRNA^{Gln}) could be converted into an initiator tRNA by introducing these two determinants. The engineered tRNA with the CUA anticodon for the amber (TAG) nonsense codon was activated by the endogenous glutaminyl-tRNA synthetase (GlnRS), and the initiator tRNA charged with Gln enabled the synthesis of recombinant proteins with a TAG codon at their initial position (Varshney et al., 1993). Based on these results, we introduced the two elements into an engineered *Mj*-tRNA^{Tyr} (Wang and Schultz 2001) (**Figure 2B**) by mutations of G72→A72 for the C1:A72 mismatch and A31: U39→G31:C39 for the three consecutive G:C pairs in the anticodon stem. In addition to the C1:A72 mismatch, several other elements in the acceptor and the D stem (blue in **Figure 2A**) have been reported to play a role in the formylation by MTF (RajBhandary 1994, Louise, L. et al., 2009). Based on a previously published report in which *Mj*-tRNA^{Tyr} was engineered for its orthogonality to the 20 endogenous aminoacyl-tRNA synthetases of *E. coli* (Guo et al., 2009), several additional mutations were introduced into the acceptor stem. Since any changes in the A11: U24 pair significantly inhibited charging the tRNA in our previous experiments (unpublished results), we decided not to introduce the element in the D stem into *Mj*-tRNA^{Tyr}, and the
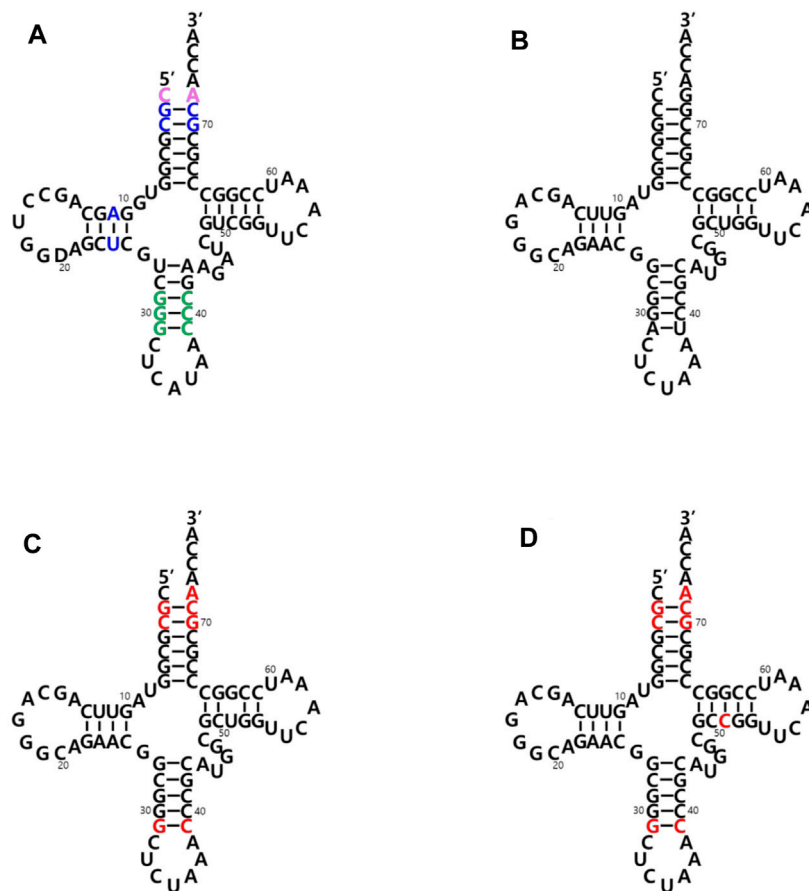
FIGURE 2 | Cloverleaf depiction of tRNAs used in this study. **(A)** *E. coli* initiator tRNA (Ec-tRNA^{fMet}). The two main identity elements of *Ec*-tRNA^{fMet} are the absence of a Watson-Crick base pair between positions C1 and A72 in the acceptor stem (in pink) and three consecutive G:C base pairs in the anticodon stem (in green). Some additional elements playing a role in the formylation reaction are shown in blue. **(B)** Engineered *M. jannaschii* tyrosyl-tRNA$_{CUA}$ (*Mj*-tRNA^{Tyr}). **(C)** *Mj* initiator tRNA$_{CUA}$-1 (*Mj*-itRNA-1). **(D)** Mj initiator tRNA$_{CUA}$-2 (*Mj*-itRNA-2). The changed nucleotides in *Mj*-itRNA-1 and *Mj*-itRNA-2 compared with *Mj*-tRNA^{Tyr} are marked by red letters.

resulting molecule was named *Mj*-itRNA-1 (**Figure 2C**). The C51G52 motif was reported to interact with IF-2 (Louise, L. et al., 2009), although it is weak, and the element was additionally introduced into *Mj*-itRNA-1 by a mutation of U51→C51, resulting in *Mj*-itRNA-2 (**Figure 2D**).

The start codon of a GFP gene (Yoo et al., 2007) was changed to the TAG amber codon, and the gene was cloned into pBbE6a (Lee et al., 2011). The tRNA gene in the pEVOL-AzF plasmid (Young et al., 2010) was replaced with *Mj*-itRNA-1 or *Mj*-itRNA-2. pEVOL-AzF encodes an *Mj*-TyrRS variant (AzF-RS) that is active toward both AzF and O-propargyl-l-tyrosine (OpgY) (Lee et al., 2015). The resulting plasmids were transformed into *E. coli* DH10β with pBbS2k-ProRS, which overexpresses *E. coli* prolyl-tRNA synthetase (*Ec*-ProRS) and suppresses the mischarging of *E. coli* prolyl-tRNA with ncAAs (Lee et al., 2015). Expression of AzF-RS and *Ec*-ProRS was induced by L-arabinose and anhydrotetracycline, respectively, and the GFP protein with the TAG codon at its N-terminus was expressed by isopropyl β-D-thiogalatopyranoside in the presence of OpgY (**Figure 3A**). The cell lysate reacted with biotin-azide via the Cu(I)-catalyzed

click reaction, and the products were analyzed by western blotting using a HRP conjugate (**Figure 3B**). The signal indicates that the protein has an alkyne group of OpgY. Only when all *Mj*-itRNA-2, AzF-RS, and OpgY were present, the western signal was detected; which suggests that OpgY was incorporated into the TAG codon located at the initial position of GFP. That is, *Mj*-itRNA-2 was charged with OpgY by AzF-RS, and the α-amine group of OpgY in the aminoacylated tRNA was probably formylated by MTF, and then the resulting tRNA, formyl-OpgY-*Mj*-itRNA-2, supported the translation initiation with the amber codon.

Overexpression of MTF Increased the Efficiency of the Protein Synthesis Initiation with ncAA

Even though *Mj*-itRNA-2 enabled the initiation of protein synthesis at the amber codon, its interactions with the factors involved in forming the 30S initiation complex might be relatively weak compared with those of *Ec*-tRNA^{fMet}. An approach to address this would be to increase the concentrations of these

FIGURE 3 | Incorporation of non-canonical amino acids into the translation initiation position. **(A)** O-propargyl-l-tyrosine (OpgY). **(B)** Comparison of *Mj*-itRNA-1 and *Mj*-itRNA-2 for the OpgY incorporation into the TAG codon located at the N-terminus of GFP. **(C)** Effects of additional expression of *E. coli* methionyl-tRNA transformylase (MTF) or *E. coli* initiation factor-2 (IF-2) on the protein synthesis initiation at the TAG codon. Each cell lysate was subjected to the Cu(I)-catalyzed click reaction with biotin-azide, and then the blot was probed with a streptavidin-horse radish peroxidase (HRP) conjugate **(B, C)**. The uncropped raw data were shown in **Supplementary Figures S2B, S2C** . **(D)** Deconvoluted electrospray ionization mass spectrum of the purified Z domain. The protein has the TAG codon at the translation initiation position. The calculated mass when OpgY is incorporated into the TAG codon is 8,364.06 Da, and that with Gln is 8,290.96 Da. **(E)** Edman sequencing result for the first residue of the purified Z domain expressed with OpgY. The results for the second to the fifth residue are shown in **Supplementary Figure S4**.

factors instead of engineering the tRNA further. The three interactions of MTF, 30S ribosome, and IF-2 with *Mj*-itRNA-2 aminoacylated with OpgY could be considered to increase translation initiation with OpgY. We decided to explore the overexpression of MTF or IF-2; increasing the concentration of the 30S ribosome could affect various aspects of physiology and thus was not tested in this study. The genes for MTF or IF-2 were cloned under the constitutive Gln promoter of pEVOL-AzF; a biscistronic gene was used to express MTF and IF-2 simultaneously. Additional expression of MTF significantly improved the incorporation of OpgY into the amber codon at the initiation position of GFP, while the effect of overexpressing IF-2 was marginal (**Figure 3C**). This observation suggested that the formylation of the α-amine group of OpgY-*Mj*-itRNA-2 was the rate-determining step in translation initiation with the TAG codon. Expressing IF-2 in addition to MTF via the biscistronic construct decreased the signal compared with the case of expressing only MTF. This might be because the expression of IF-2 decreased the level of MTF due to the limited resources for transcription and translation in expressing multiple genes. The aminoacyl-tRNA synthetase of AzF-RS was originally engineered to activate AzF (Chin et al., 2002) and incorporate AzF into the

initiation TAG codon was attempted using the same *E. coli* strain. However, unlike OpgY, *Mj*-itRNA-2 did not support protein translation initiation with AzF (data not shown). In the case of *Mj*-itRNA-1, additional expression of MTF or IF-2 did not result in detectable incorporation of OpgY into the initiation TAG codon (data not shown).

The TAG Start Codon was Encoded with Either ncAA or Gln by Mj-itRNA-2

A small protein (Z domain (Nilsson et al., 1987)) was used to evaluate OpgY incorporation using a mass spectrometry. The start codon of Z domain was changed to TAG codon, and the resulting gene was cloned into pQE-80 L. The Z domain was expressed using *E. coli* cells expressing *Mj*-itRNA-2, AzF-RS, MTF, and ProRS in the presence of OpgY. The purified protein was analyzed by liquid chromatography-mass spectrometry. The calculated mass of the Z domain with OpgY at its N-terminus was 8,364.06 Da, and a mass of 8,364.17 was detected (**Figure 3D**). This result, along with the western blot results (**Figures 3B,C**), indicated that OpgY was incorporated into the TAG codon of the translation initiation position. However, another mass peak
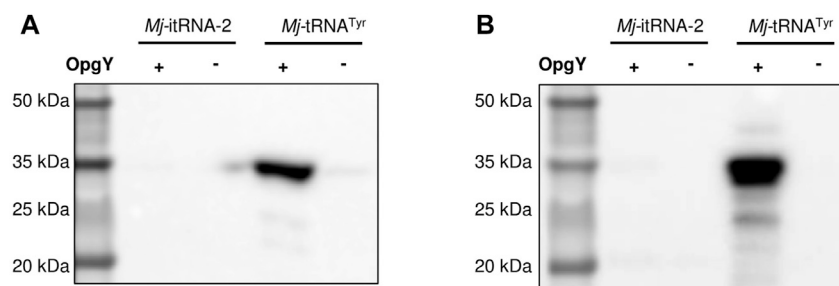
FIGURE 4 | Suppression of an elongational TAG codon with *Mj*-itRNA-2. There is a TAG codon between GST and the Z domain-His$_6$, and the fusion protein was expressed with *Mj*-itRNA-2 or *Mj*-tRNA^{Tyr}. Each cell lysate was subjected to the Cu(I)-catalyzed click reaction with biotin-azide, and then the blot was probed with an anti-His$_5$-HRP conjugate **(A)** or a streptavidin-HRP conjugate **(B)**. The uncropped raw data were shown in **Supplementary Figures S5B, S5C**.

(8,291.15 Da) was observed in the Z-domain sample. Since the mass of the wild-type Z domain (Met at its N-terminus) is 8,294.03 Da (**Supplementary Figure S3**: the observed mass was 8,294.12), it was suspected that a canonical amino acid besides Met was incorporated into the TAG codon. The N-terminal residue of the Z domain was determined by the Edman degradation analysis, and two residues, OpgY and Gln, were detected (**Figure 3E**). The peaks for Gln and OpgY were integrated (597 for Gln and 1477 for OpgY), and the ratio of OpgY to Gln incorporated into the TAG position was calculated as 2.55. The mass of 8,291.15 Da was consistent with the calculated mass (8,290.96 Da) when Gln was incorporated into the TAG codon.

Mj-itRNA-2 Did Not Support the Incorporation of ncAA into an Internal Amber Codon

Next, we examined whether *Mj*-itRNA-2 could function as an elongator tRNA. The two main elements of *Ec*-tRNA^{fMet}, the C1: A72 mismatch, and the three consecutive G:C pairs in the anticodon stem, are known to distinguish the initiator tRNA from elongator tRNAs and to prevent it from being involved in the translation elongation process (Laursen et al., 2005). In addition to the two elements, *Mj*-itRNA-2 had mutations in the acceptor and the D stem; the T stem has also been reported to interact with *E. coli* elongation factor Tu (Guo et al., 2009). A plasmid was constructed by introducing the TAG codon between GST and the Z domain, and the protein had a C-terminal His$_6$-tag. The GST-TAG-Z protein was expressed with either *Mj*-tRNA^{Tyr} (**Figure 2B**) or *Mj*-itRNA-2, and the cell lysates were analyzed by western blotting after the Cu(I)-catalyzed click reaction with biotin-azide. While OpgY was efficiently incorporated into the TAG codon with the elongator *Mj*-tRNA^{Tyr}, *Mj*-itRNA-2 could not suppress the internal nonsense codon (**Figure 4**). These results indicate that *Mj*-itRNA-2 1) is charged with ncAA, 2) initiates protein translation at the TAG codon, and 3) is not active in the elongation process of protein synthesis. *Mj*-itRNA-2 is the first reported tRNA enabling the orthogonal incorporation of ncAA into the amber codon only at the translation initiation position.

DISCUSSION

Installation of ncAAs into specific positions of the protein has provided invaluable tools for not only studying proteins but also engineering them. Orthogonal pairs of tRNA/aaRS have been engineered for these purposes, and novel proteins with various unnatural functional groups have been created. These systems have been developed to target the elongation process of protein synthesis. However, a method to incorporate ncAAs at the translation initiation position in a site-specific manner has rarely been reported. This may be attributed to the difficulty in engineering functional initiator tRNAs compared with elongation tRNAs. The protein synthesis initiation process of *E. coli* is distinct from the elongation process, and the initiator tRNA (*Ec*-tRNA^{fMet}) plays an important role in this process. *Ec*-tRNA^{fMet} interacts with MTF, 30S ribosome, and IF-2(RajBhandary 1994; Laursen et al., 2005); the identity elements for these interactions also prevent association with the elongation process. In this study, we introduced the identity elements of *Ec*-tRNA^{fMet} into an elongator *Mj*-tRNA^{Tyr} variant. One of the designed initiator tRNAs, *Mj*-itRNA-2, incorporated OpgY into the TAG codon at the translation initiation position, but was not active for an internal TAG codon.

Overexpression of MTF significantly improved the efficiency of translation initiation with OpgY, and the result implied that OpgY-*Mj*-itRNA-2 was not an efficient substrate for MTF. The same system failed to initiate protein translation with AzF, even though AzF-RS is also active toward ncAA. The identity of amino acids linked to an initiator tRNA affects the activity of MTF (Schmitt et al., 1998; Laursen et al., 2005). It is plausible that AzF-*Mj*-tRNA-2 was not formylated, considering the structural difference between OpgY and AzF. The elements present in the acceptor stem were introduced for interaction with MTF, but some other elements such as A11: U24 in the D stem (Lee et al., 1991; Ramesh et al., 1997) were not tested in this study because of the detrimental effects of mutations in this region in our previous experiments. In addition to OpgY, Gln was incorporated into the start TAG codon, which suggests that *Mj*-itRNA-2 was charged with Gln,

probably by the endogenous glutaminyl-tRNA synthetase (*Ec*-GlnRS). These results suggest that further engineering of the initiator tRNA, particularly for enhancing the interaction with MTF and preventing the aminoacylation with Gln by Ec-GlnRS, could improve translation initiation efficiency with ncAAs. The newly engineered initiator tRNAs would enable developing an orthogonal translation initiation system active only for the TAG start codon. Furthermore, with many *Mj*-TyrRS variants engineered toward various ncAAs, a repertoire of functional groups would be available for site-specific incorporation at the N-termini of proteins.

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.

REFERENCES

Agostini, F., Völler, J. S., Koksch, B., Acevedo-Rocha, C. G., Kubyshkin, V., and Budisa, N. (2017). Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed.* 56 (33), 9680–9703. doi:10.1002/anie.201610129

Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S., and Schultz, P. G. (2004). An Expanded Genetic Code with a Functional Quadruplet Codon. *Proc. Natl. Acad. Sci.* 101 (20), 7566–7571. doi:10.1073/pnas.0401517101

Chin, J. W. (2017). Expanding and Reprogramming the Genetic Code. *Nature* 550 (7674), 53–60. doi:10.1038/nature24031

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

Christopher, J. A., and Baldwin, T. O. (1996). Implications of N and C-Terminal Proximity for Protein Folding. *J. Mol. Biol.* 257 (1), 175–187. doi:10.1006/jmbi.1996.0154

Evans, T. C., Jr., and Xu, M.-Q. (1999). Intein-mediated Protein Ligation: Harnessing Nature's Escape Artists. *Biopolymers* 51 (5), 333–342. doi:10.1002/(sici)1097-0282(1999)51:5<333:aid-bip3>3.0.co;2-#

Guo, J., Melançon, C. E., 3rd, Lee, H. S., Groff, D., and Schultz, P. G. (2009). Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angew. Chem. Int. Edition* 48 (48), 9148–9151. doi:10.1002/anie.200904035

Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H., and Sisido, M. (2001). Incorporation of Nonnatural Amino Acids into Proteins by Using Various Four-Base Codons in an *Escherichia coli In Vitro* Translation System. *Biochemistry* 40 (37), 11060–11064. doi:10.1021/bi0108204

Hong, V., Presolski, S. I., Ma, C., and Finn, M. â. G. (2009). Analysis and Optimization of Copper-Catalyzed Azideâ€"Alkyne Cycloaddition for Bioconjugation. *Angew. Chem. Int. Ed.* 48 (52), 9879–9883. doi:10.1002/anie.200905087

Jacob, E., and Unger, R. (2007). A Tale of Two Tails: Why Are Terminal Residues of Proteins Exposed. *Bioinformatics* 23 (2), e225–e230. doi:10.1093/bioinformatics/btl318

Johnson, J. A., Lu, Y. Y., Van Deventer, J. A., and Tirrell, D. A. (2010). Residue-specific Incorporation of Non-canonical Amino Acids into Proteins: Recent Developments and Applications. *Curr. Opin. Chem. Biol.* 14 (6), 774–780. doi:10.1016/j.cbpa.2010.09.013

Kinneer, K., Flynn, M., Thomas, S. B., Meekin, J., Varkey, R., Xiao, X., et al. (2019). Preclinical Assessment of an Antibody-PBD Conjugate that Targets BCMA on

AUTHOR CONTRIBUTIONS

BL, WC, and TY designed research; BL, WC, and SL performed the experiments and analyzed the data; BK performed mass analyses; BL, WC, and TY wrote the article. All authors approved the final version of the article.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.772648/full#supplementary-material

Multiple Myeloma and Myeloma Progenitor Cells. *Leukemia* 33 (3), 766–771. doi:10.1038/s41375-018-0278-7

Kozak, M. (1999). Initiation of Translation in Prokaryotes and Eukaryotes. *Gene* 234 (2), 187–208. doi:10.1016/s0378-1119(99)00210-3

Laursen, B. S., Sørensen, H. P., Mortensen, K. K., and Sperling-Petersen, H. U. (2005). Initiation of Protein Synthesis in Bacteria. *Microbiol. Mol. Biol. Rev.* 69 (1), 101–123. doi:10.1128/mmbr.69.1.101-123.2005

Lee, B. S., Kim, S., Ko, B. J., and Yoo, T. H. (2017). An Efficient System for Incorporation of Unnatural Amino Acids in Response to the Four-Base Codon AGGA in *Escherichia coli*. *Biochim. Biophys. Acta Gen. Subj* 1861 (11 Pt B), 3016–3023. doi:10.1016/j.bbagen.2017.02.017

Lee, B. S., Shin, S., Jeon, J. Y., Jang, K.-S., Lee, B. Y., Choi, S., et al. (2015). Incorporation of Unnatural Amino Acids in Response to the AGG Codon. *ACS Chem. Biol.* 10 (7), 1648–1653. doi:10.1021/acschembio.5b00230

Lee, C. P., Seong, B. L., and RajBhandary, U. L. (1991). Structural and Sequence Elements Important for Recognition of *Escherichia coli* Formylmethionine tRNA by Methionyl-tRNA Transformylase Are Clustered in the Acceptor Stem. *J. Biol. Chem.* 266 (27), 18012–18017. doi:10.1016/s0021-9258(18)55230-3

Lee, T. S., Krupa, R. A., Zhang, F., Hajimorad, M., Holtz, W. J., Prasad, N., et al. (2011). BglBrick Vectors and Datasheets: A Synthetic Biology Platform for Gene Expression. *J. Biol. Eng.* 5, 12. doi:10.1186/1754-1611-5-12

Liebscher, S., Schöpfel, M., Aumüller, T., Sharkhuukhen, A., Pech, A., Höss, E., et al. (2014). N-terminal Protein Modification by Substrate-Activated Reverse Proteolysis. *Angew. Chem. Int. Ed.* 53 (11), 3024–3028. doi:10.1002/anie.201307736

Link, A. J., Mock, M. L., and Tirrell, D. A. (2003). Non-canonical Amino Acids in Protein Engineering. *Curr. Opin. Biotechnol.* 14 (6), 603–609. doi:10.1016/j.copbio.2003.10.011

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

Louise, C. V. R, Brain, S. L., and Hansu, U. S. (2009). *Initiator tRNAs in Bacteria and Eukaryotes. Encyclopedia of Life Science*. (New York: John Wiley and Sons).

Mao, H., Hart, S. A., and Pollack, B. A. (2004). Sortase-Mediated Protein Ligation - A New Method for Protein Engineering. *J. Am. Chem. Soc.* 126, 2670–2671. doi:10.1021/ja039915e

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

Ngo, J. T., Schuman, E. M., and Tirrell, D. A. (2013). Mutant Methionyl-tRNA Synthetase from Bacteria Enables Site-Selective N-Terminal Labeling of Proteins Expressed in Mammalian Cells. *Proc. Natl. Acad. Sci.* 110 (13), 4992–4997. doi:10.1073/pnas.1216375110

Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E., et al. (1987). A Synthetic IgG-Binding Domain Based on Staphylococcal Protein A. *Protein Eng. Des. Sel* 1 (2), 107–113. doi:10.1093/protein/1.2.107

Pagar, A. D., Patil, M. D., Flood, D. T., Yoo, T. H., Dawson, P. E., and Yun, H. (2021). Recent Advances in Biocatalysis with Chemical Modification and Expanded Amino Acid Alphabet. *Chem. Rev.* 121 (10), 6173–6245. doi:10.1021/acs.chemrev.0c01201

Proft, T. (2010). Sortase-mediated Protein Ligation: an Emerging Biotechnology Tool for Protein Modification and Immobilisation. *Biotechnol. Lett.* 32 (1), 1–10. doi:10.1007/s10529-009-0116-0

RajBhandary, U. L. (1994). Initiator Transfer RNAs. *J. Bacteriol.* 176 (3), 547–552. doi:10.1128/jb.176.3.547-552.1994

Ramesh, V., Gite, S., Li, Y., and RajBhandary, U. L. (1997). Suppressor Mutations in *Escherichia coli* Methionyl-tRNA Formyltransferase: Role of a 16-amino Acid Insertion Module in Initiator tRNA Recognition. *Proc. Natl. Acad. Sci.* 94 (25), 13524–13529. doi:10.1073/pnas.94.25.13524

Rosen, C. B., and Francis, M. B. (2017). Targeting the N Terminus for Site-Selective Protein Modification. *Nat. Chem. Biol.* 13 (7), 697–705. doi:10.1038/nchembio.2416

Schmitt, E., Panvert, M., Blanquet, S., and Mechulam, Y. (1998). Crystal Structure of Methionyl-tRNAfMet Transformylase Complexed with the Initiator Formyl-Methionyl-tRNAfMet. *EMBO J.* 17 (23), 6819–6826. doi:10.1093/emboj/17.23.6819

Simone, A. D., Acevedo-Rocha, C. G., Hoesl, M. G., and Budisa, N. (2016). Towards Reassignment of the Methionine Codon AUG to Two Different Noncanonical Amino Acids in Bacterial Translation. *Croat. Chem. Acta* 89 (2), 243–253. doi:10.5562/cca2915

Smolskaya, S., and Andreev, Y. A. (2019). Site-Specific Incorporation of Unnatural Amino Acids into *Escherichia coli* Recombinant Protein: Methodology Development and Recent Achievement. *Biomolecules* 9 (7). doi:10.3390/biom9070255

Steiman, D. M., Ridge, R. J., and Matsueda, G. R. (1985). Synthesis of Side Chain-Protected Amino Acid Phenylthiohydantoins and Their Use in Quantitative Solid-phase Edman Degradation. *Anal. Biochem.* 145 (1), 91–95. doi:10.1016/0003-2697(85)90331-8

Serada, T. J., and Mant, C. (1993). Effect of the α-amino Group on Peptide Retention Behaviour in Reversed-phase Chromatography Determination of the pKa Values of the α-amino Group of 19 Different N-Terminal Amino Acid Residues. *J. Chromatogr.* 646, 17–30.

Tharp, J. M., Ad, O., Amikura, K., Ward, F. R., Garcia, E. M., Cate, J. H. D., et al. (2020). Initiation of Protein Synthesis with Non-Canonical Amino Acids *In Vivo. Angew. Chem. Int. Ed.* 59 (8), 3122–3126. doi:10.1002/anie.201914671

Tharp, J. M., Krahn, N., Varshney, U., and Söll, D. (2020). Hijacking Translation Initiation for Synthetic Biology. *Chembiochem* 21 (10), 1387–1396. doi:10.1002/cbic.202000017

Varshney, U., Lee, C. P., and RajBhandary, U. L. (1993). From Elongator tRNA to Initiator tRNA. *Proc. Natl. Acad. Sci.* 90 (6), 2305–2309. doi:10.1073/pnas.90.6.2305

Varshney, U., and RajBhandary, U. L. (1990). Initiation of Protein Synthesis from a Termination Codon. *Proc. Natl. Acad. Sci.* 87 (4), 1586–1590. doi:10.1073/pnas.87.4.1586

Wals, K., and Ovaa, H. (2014). Unnatural Amino Acid Incorporation in *E. coli*: Current and Future Applications in the Design of Therapeutic Proteins. *Front. Chem.* 2, 15. doi:10.3389/fchem.2014.00015

Wang, K., Schmied, W. H., and Chin, J. W. (2012). Reprogramming the Genetic Code: from Triplet to Quadruplet Codes. *Angew. Chem. Int. Ed.* 51 (10), 2288–2297. doi:10.1002/anie.201105016

Wang, L., and Schultz, P. G. (2004). Expanding the Genetic Code. *Angew. Chem. Int. Ed. Engl.* 44 (1), 34–66. doi:10.1002/anie.200460627

Wang, L., and Schultz, P. G. (2001). A General Approach for the Generation of Orthogonal tRNAs. *Chem. Biol.* 8 (9), 883–890. doi:10.1016/s1074-5521(01)00063-1

Wang, Q., Parrish, A. R., and Wang, L. (2009). Expanding the Genetic Code for Biological Studies. *Chem. Biol.* 16 (3), 323–336. doi:10.1016/j.chembiol.2009.03.001

Wang, Q., and Wang, L. (2012). Genetic Incorporation of Unnatural Amino Acids into Proteins in Yeast. *Methods Mol. Biol.* 794, 199–213. doi:10.1007/978-1-61779-331-8_12

Weeks, A. M., and Wells, J. A. (2020). Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* 120 (6), 3127–3160. doi:10.1021/acs.chemrev.9b00372

Xiao, H., Chatterjee, A., Choi, S.-h., Bajjuri, K. M., Sinha, S. C., and Schultz, P. G. (2013). Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells. *Angew. Chem. Int. Ed.* 52 (52), 14080–14083. doi:10.1002/anie.201308137

Xu, M.-Q., and Evans, T. C., Jr. (2001). Intein-mediated Ligation and Cyclization of Expressed Proteins. *Methods* 24 (3), 257–277. doi:10.1006/meth.2001.1187

Yoo, T. H., Link, A. J., and Tirrell, D. A. (2007). Evolution of a Fluorinated green Fluorescent Protein. *Proc. Natl. Acad. Sci.* 104 (35), 13887–13890. doi:10.1073/pnas.0701904104

Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010). An Enhanced System for Unnatural Amino Acid Mutagenesis in *E. coli. J. Mol. Biol.* 395 (2), 361–374. doi:10.1016/j.jmb.2009.10.030

frontiers
in Chemistry

Synthetic Thiol and Selenol Derived Amino Acids for Expanding the Scope of Chemical Protein Synthesis

Ivy Guan[1,2], Kayla Williams[1,2], Joanna Shu Ting Liu[2,3] and Xuyu Liu[1,2]*

[1]School of Chemistry, Faculty of Science, The University of Sydney, Sydney, NSW, Australia, [2]The Heart Research Institute, The University of Sydney, Sydney, NSW, Australia, [3]School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

Cells employ post-translational modifications (PTMs) as key mechanisms to expand proteome diversity beyond the inherent limitations of a concise genome. The ability to incorporate post-translationally modified amino acids into protein targets via chemical ligation of peptide fragments has enabled the access to homogeneous proteins bearing discrete PTM patterns and empowered functional elucidation of individual modification sites. Native chemical ligation (NCL) represents a powerful and robust means for convergent assembly of two homogeneous, unprotected peptides bearing an N-terminal cysteine residue and a C-terminal thioester, respectively. The subsequent discovery that protein cysteine residues can be chemoselectively desulfurized to alanine has ignited tremendous interest in preparing unnatural thiol-derived variants of proteogenic amino acids for chemical protein synthesis following the ligation-desulfurization logic. Recently, the 21st amino acid selenocysteine, together with other selenyl derivatives of amino acids, have been shown to facilitate ultrafast ligation with peptidyl selenoesters, while the advancement in deselenization chemistry has provided reliable bio-orthogonality to PTMs and other amino acids. The combination of these ligation techniques and desulfurization/deselenization chemistries has led to streamlined synthesis of multiple structurally-complex, post-translationally modified proteins. In this review, we aim to summarize the latest chemical synthesis of thiolated and selenylated amino-acid building blocks and exemplify their important roles in conquering challenging protein targets with distinct PTM patterns.

Keywords: peptide ligation, unnatural amino acid, cysteine, chemical protein synthesis, semisynthesis, protein modification, desulfurization, deselenization

INTRODUCTION

In the past two decades, native chemical ligation (NCL) technology has been broadly utilized for the convergent building of peptide and protein targets. As described in the seminal report by Kent and coworkers (Dawson et al., 1994), this technique involves chemical ligation of two unprotected peptide fragments bearing an N-terminal cysteine (Cys) and a C-terminal thioester respectively, and operates in neutral denaturing buffer to furnish the ligation product. The reaction starts with a reversible *trans*-thioesterification reaction through nucleophilic attack on the peptide thioester by the N-terminal Cys thiolate, leading to the formation of a bridged thioester intermediate that subsequently undergoes irreversible, intramolecular S-to-N acyl shift to afford a native peptide

bond (**Scheme 1A**). Kent also demonstrated that sequential NCL reactions can be achieved by harnessing the differential reactivity of peptidyl thioesters (Bang et al., 2006): a bifunctional peptide fragment possessing an N-terminal Cys and a dormant alkyl thioester on the C terminus was ligated with a peptide phenyl thioester without the concern of head-to-tail peptide cyclization;

aryl thiol additives [i.e., thiophenol and 4-mercaptophenylacetic acid (MPAA)] were added subsequently to activate the inert thioester which allowed the next NCL reaction to proceed *in situ* (**Scheme 1B**). This sequential ligation concept has been furthered by the use of 2,2,2-trifluoroethanethiol (TFET) to enhance the rate of peptide ligation through *in situ* generation of a productive

SCHEME 1 | (A) Native chemical ligation; **(B)** Kinetically-controlled ligation; **(C)** Transforming peptide hydrazide to peptide thioester via oxidation and thiolysis; **(D)** Protein thioesterification via thiolysis of intein tag; **(E)** Sortase-mediated peptide ligation; **(F)** One-pot selenocystine-selenoester ligation-deselenization chemistry.
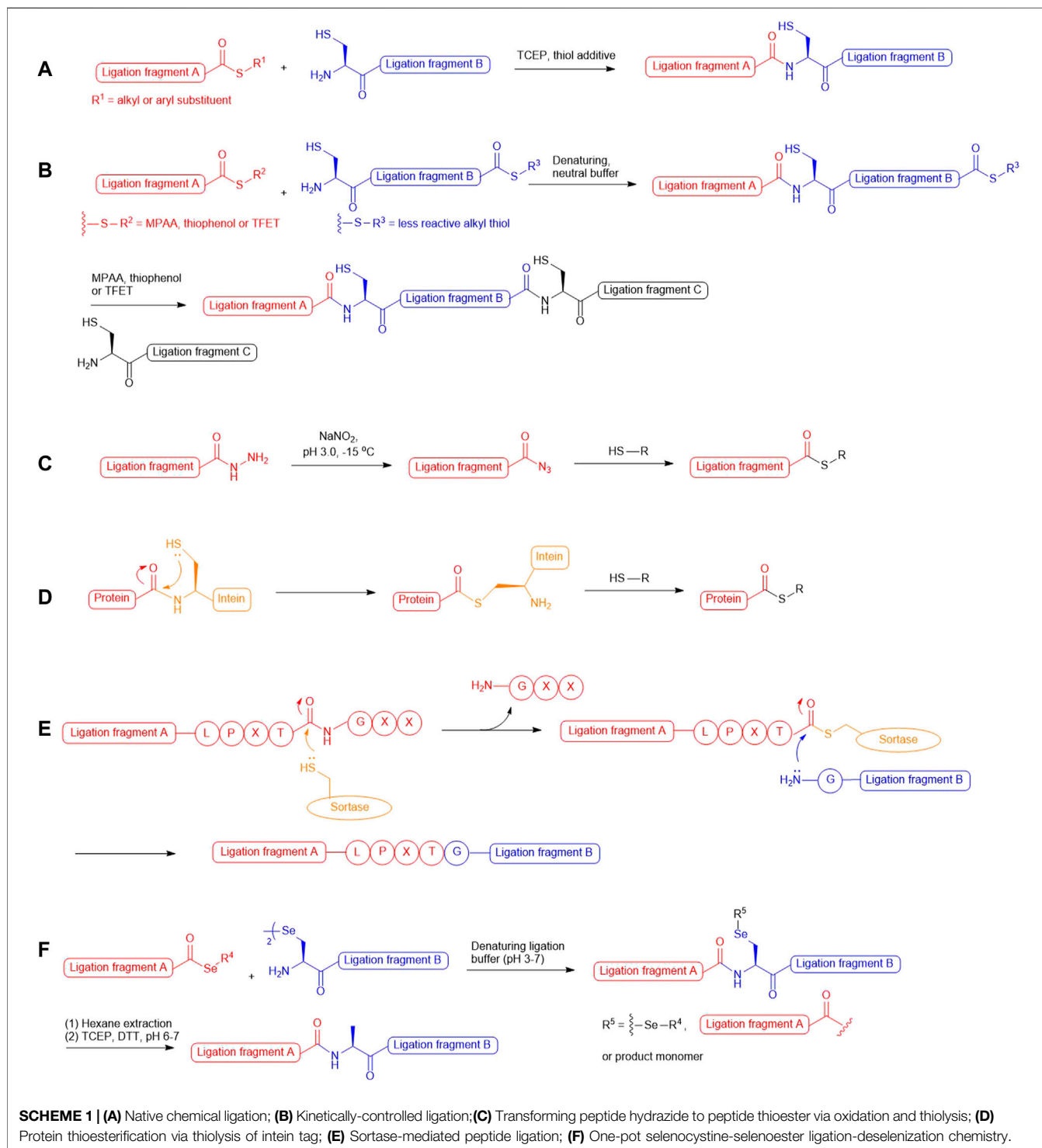
TFET thioester (Thompson et al., 2014; Thompson et al., 2017). Peptide hydrazides have also been extensively used in iterative peptide ligation strategies as latent thioesters (Li et al., 2018; Fang et al., 2011). Importantly, the acyl hydrazide moiety is unresponsive to NCL conditions and acts as a protecting group for the C-terminal carboxylate, which can be converted to acyl azide by $NaNO_2$ followed by thiolysis to unleash the capacity for NCL (Fang et al., 2011; Qi et al., 2019) (**Scheme 1C**). Recently, Flood *et al.* demonstrated that peptide hydrazides can be activated by a stoichiometric amount of acetyl acetone to produce peptide acyl-pyrazoles that will readily undergo transesterification reactions with aryl thiols to afford activated peptide thioesters for peptide ligation (Flood et al., 2018).

These powerful methodologies have been implemented in total synthesis of numerous protein targets and have demonstrated excellent compatibility with those bearing delicate post-translational modifications (PTMs), such as phosphorylation (Shogren-Knaak et al., 2003; Pan et al., 2019; Kilic et al., 2018; Jbara et al., 2016), sulfation (Watson et al., 2018; Hsieh et al., 2014), ubiquitination (Li et al., 2018; Yang et al., 2009; Ajish Kumar et al., 2009; Kumar et al., 2010; Kumar et al., 2011; Siman et al., 2013) and glycosylation (Wang et al., 2013; Wilkinson et al., 2012; Murakami et al., 2016). Whilst recombinant expression technologies remain important means to access large and structurally-complex proteins, they cannot afford the same level of target precision and chemical diversity that chemical protein synthesis provides. In particular, the enzymatic nature of PTM machineries in cells leads to target protein production as a heterogeneous mixture of different PTM forms that are unresolvable using chromatographic techniques. Examples that highlight the ability to access defined PTM forms of natural protein targets via NCL chemistry include the total synthesis of homogeneously glycosylated erythropoietin (Wang et al., 2013), different modified variants of histone proteins (Jbara et al., 2016; Siman et al., 2013; Seenaiah et al., 2015; Jbara et al., 2017) and sulfated anticoagulant proteins (Thompson et al., 2017; Watson et al., 2018; Hsieh et al., 2014; Watson et al., 2019). Despite our ability to perform chemical synthesis of post-translationally modified proteins using suitably protected Cys and acyl donors, the number of iterative ligation steps that can be performed on a large scale to provide the final target in sufficient quantity is often limited. This limitation has spurred tremendous interest in converging NCL with intein-mediated protein splicing technology referred to as Expressed Protein Ligation (EPL) (Muir et al., 1998). Harnessing the thiolytic nature of intein-based protein tags, this method allows for the generation of recombinant proteins possessing C-terminal thioesters (**Scheme 1D**), which permits further engineering with exogenous peptides bearing an N-terminal Cys residue. Readers are advised to refer to a plethora of excellent reviews (Conibear et al., 2018; Thompson and Muir, 2020; Wang and Cole, 2020; Muir, 2003; Flavell and Muir, 2009; Hofmann and Muir, 2002) that summarize the tremendous progress in the area of EPL. The NCL method has also been combined with chemoenzymatic ligation (Thompson and Muir, 2020; Nuijens et al., 2019; Schmidt et al., 2017; Weeks and Wells, 2020; Arnold,

2018) to access protein targets with residue-specific modifications, as exemplified by the recent semisynthesis of β2-adrenergic receptors reported by Liu and coworkers, which combines sortase-catalyzed ligation (**Scheme 1E**) with NCL to access discrete phosphorylation and ubiquitination forms (Li et al., 2021a).

The presence of an N-terminal Cys residue in one of two peptide fragments was a prerequisite for NCL chemistries, which significantly limited the possible disconnection routes in retrosynthetic analysis, particularly given the low occurrence of Cys among all the proteogenic amino acids (1.8%) (Moura et al., 2013). This synthetic barrier has attracted global research endeavours to expand the scope of NCL by developing new cysteine surrogates empowering the access to Cys-free protein targets and those having Cys residues at positions inadequate for peptide ligation. Initial attempts included the use of N-terminal thiol auxiliaries to mediate NCL reactions (Yan and Dawson, 2001; Offer, 2010), and recently, Li et al. also exploited the potential of glycyl auxiliary in isopeptide ligation with a view to precisely introduce ubiquitin modifications to the protein of interest (i.e., H2A) (Li et al., 2018). However, such peptide ligation strategies typically require long reaction times and additional steps to remove the auxiliaries, whereby hydrolyzed and epimerized by-products can develop over time and undermine the overall ligation yield. In this exploratory journey, one of the most effective approaches is to incorporate a thiol moiety on the side chain of the N-terminal amino acid that enables peptide ligation in a way analogous to that at Cys residues. The scope of such NCL methodology was expanded by Yan and Dawson through the development of a post-ligation desulfurization chemistry mediated by either Raney Ni or Pd, whereby traceless removal of the thiol appendage could be achieved (Yan and Dawson, 2001). However, the large excess of Ni or Pd used in the desulfurization reaction can result in undesirable tryptophan hydrogenation or methionine demethylthiolization. Later, Danishefsky and coworkers furthered this ligation−desulfurization concept through applying milder, metal-free radical desulfurization conditions to remove the thiol moiety, which consists of a water-soluble radical initiator 2,2′-azobis [2-(2-imidazolin-2-yl)propane] dihydrochloride (VA-044), TCEP and a sulfhydryl donor in neutral ligation buffer (Wan and Danishefsky, 2007). Importantly, this radical desulfurization chemistry is orthogonal to PTMs on the side chains of proteogenic amino acids and peptidyl thioesters, which facilitates rational design of sequential, multi-component ligation strategies to access large proteins.

In parallel to the development of ligation-desulfurization methods, there have been considerable efforts to address other practical limitations, including the sluggish reaction rates at sterically encumbered ligation junctions and the lack of discrimination between unnatural thiolated amino acids and unprotected cysteine residues under desulfurization conditions. In 2001, three research groups demonstrated the utility of selenocysteine (Sec) in NCL-like reactions via nucleophilic attack on peptide thioesters followed by Se-to-N acyl shift to

afford large selenopeptides and selenoproteins (Hondal et al., 2001; Quaderer et al., 2001; Gieselman et al., 2001). Another landmark report in Sec-mediated NCL methodology from Dawson and coworkers described chemoselective deselenization chemistry that proceeded smoothly to completion in the presence of free Cys residues when using excess amounts of TCEP and dithiothreitol (Metanis et al., 2010). The observed chemoselectivity has been attributed to the weaker C–Se bond favouring homolytic cleavage via the reaction between phosphine and a selenium-centered radical, which results in the generation of a phosphine selenide intermediate and alanyl radical that undergoes hydrogen atom abstraction from a donor to regenerate an alanine residue. The resulting ligation–deselenization method opened a new pathway to access proteins bearing structurally and functionally important Cys residues. However, the selenol moiety in Sec possesses a strong reducing power and is oxidized rapidly in air to form the diselenide species (selenocystine), which necessitated the use of excess reducing agents (i.e., TCEP, MPAA and thiophenol) to maintain a productive level of selenol for NCL reactions. Recently, Payne and coworkers reported that utilizing aryl selenoesters as the ligation counterparts of selenocystine can provide substantial enhancement in ligation rates, leading to clean ligation within minutes even at the sterically demanding junctions (Mitchell et al., 2015). Remarkably, this method involves the simple mixing of peptide selenoesters with the peptide diselenide dimer in ligation buffer without the need of reducing agents (**Scheme 1F**). The additive-free reaction dubbed diselenide–selenoester ligation (DSL) represents an unprecedented reactivity of Sec that gateways to rapid generation of polypeptide and protein therapeutic leads bearing unique and defined modifications (Watson et al., 2018; Watson et al., 2019; Agten et al., 2021). These pioneer works have also fuelled the development of synthetic routes to assess suitably protected β/γ/δ-selenyl derived amino acids that can be employed in traceless selenium-mediated peptide ligation for total synthesis of protein targets bearing distinct modifications. This review will highlight the contributions of thiolated and selenylated amino acid synthesis in the advancement of ligation methodologies for detailed studies on protein structure and function influenced by PTMs.
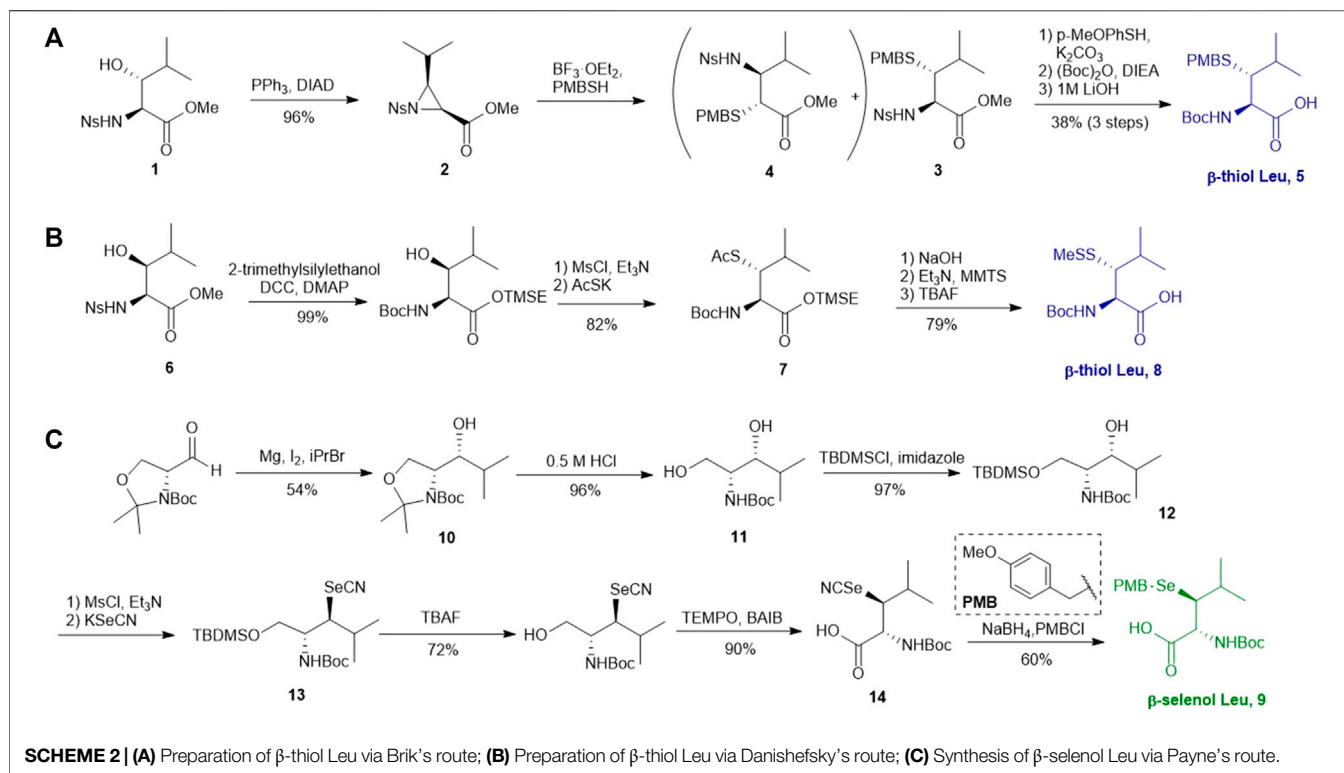
SYNTHESIS OF THIOLATED/ SELENYLATED AMINO ACIDS

β-Thiol/selenol Derived Leucine

Several β/γ-hydroxyl amino acids are commercially available as ideal starting materials for thiolated amino acid synthesis. For example, β-hydroxyleucine has been implemented in the synthesis of β-thiol Leucine (Leu) by Brik's and Danishefsky's groups (Harpaz et al., 2010; Tan et al., 2010) (**Scheme 2A**). Brik and coworkers introduced a nosyl group to the α-amine moiety of a hydroxyl Leu to form **1**, facilitating the aziridine formation via an intramolecular Mitsunobu reaction to afford **2** (Harpaz et al., 2010). Ring-opening by PMB-SH in

the presence of BF$_3$ as Lewis acid afforded the β-thiolated Leu intermediate (**3**) and its regioisomer (**4**) in a 6:4 ratio, which are separable by column chromatography. This reaction was stereospecific, directed by the α-chiral center to produce *anti*-product exclusively. After three steps of protecting group manipulation, the desired *p*-methoxybenzyl (PMB)-protected β-thiol Leu (**5**) was furnished in a 38% yield. The PMB protecting group is stable under fluorenylmethoxycarbonyl-based solid phase peptide synthesis (Fmoc-based SPPS) conditions, acidolytic deprotection and cleavage conditions; it can be readily removed using Hg (OAC)$_2$ and DTT in trifluoroacetic acid (TFA) before ligation.

Danishefsky and coworkers demonstrated a different route to β-thiol Leu, where the hydroxyl group on the side chain of the protected β-hydroxy-Leu (**6**) was activated through mesylation and displaced by thioacetate (**Scheme 2B**) (Tan et al., 2010). Interestingly, direct thiol substitution using thioacetic acid or PMB-SH as nucleophile led to substantial β-elimination product, while the conjugate base of thioacetic acid can mitigate the formation of the elimination by-product. This approach gave the *anti*-product **7** exclusively (in 82% yield over two steps) due to the intrinsic β-chiral center of the amino acid. Next, the thioacetate moiety was transformed to methyl disulfide after saponification and oxidation with methane methylthiosulfonate (MMTS), and the carboxylic acid was unmasked using tetra-n-butylammonium fluoride (TBAF) to provide the desired β-thiol Leu building block (**8**) for SPPS. This approach served as a synthetic template to access β/γ-thiol amino acids, allowing one to access a single stereoisomer to study the impact of the β/γ-chiral centers on peptide ligation efficiency. For example, peptides possessing an *anti*-β-thiolated Leu at the N-terminus can efficiently react with peptide thioesters terminated at Phe residues, however sluggish ligation reactions were observed when using the *syn*-isomer (Tan et al., 2010). The underlying mechanism has been proposed: the *syn*-conformation formed in the process of transesterification causes steric congestion between the isopropyl side chain and peptide backbone that disfavors the S-to-N acyl shift. The β-thiol Leu **7** has also been masked by a thiazolidine protecting group via acidolysis followed by paraformaldehyde incubation. The resulting β-thiol Leu has been used in the convergent synthesis of the ATAD2 bromodomain region (Creech et al., 2014).

Wang *et al.* demonstrated the use of Garner's aldehyde to access selenylated Leu **9** (**Scheme 2C**) (Wang et al., 2017). The synthesis began with the Grignard addition of iPrMgBr to Garner's aldehyde resulting in the formation of oxazolidine **10** as a single diastereomer. Subsequent nucleophilic displacement of the *syn*-mesylate generated from **10** by KSeCN was unsuccessful due to the unfavourable conformation imposed by the oxazolidine ring. The isopropylidene moiety was thereby removed to provide the diol **11** followed by selective protection of the primary alcohol with TBDMS ether (**12**). The resulting secondary alcohol underwent mesylation and substitution smoothly to afford the *anti*-selenylated product **13**. After protecting group manipulation and TEMPO-BAIB

SCHEME 2 | (A) Preparation of β-thiol Leu via Brik's route; **(B)** Preparation of β-thiol Leu via Danishefsky's route; **(C)** Synthesis of β-selenol Leu via Payne's route.

oxidation to yield the amino acid **14**, the selenocyanate moiety was reduced and subjected to PMB protection to furnish the desired selenyl Leu **9** in a 60% isolated yield.
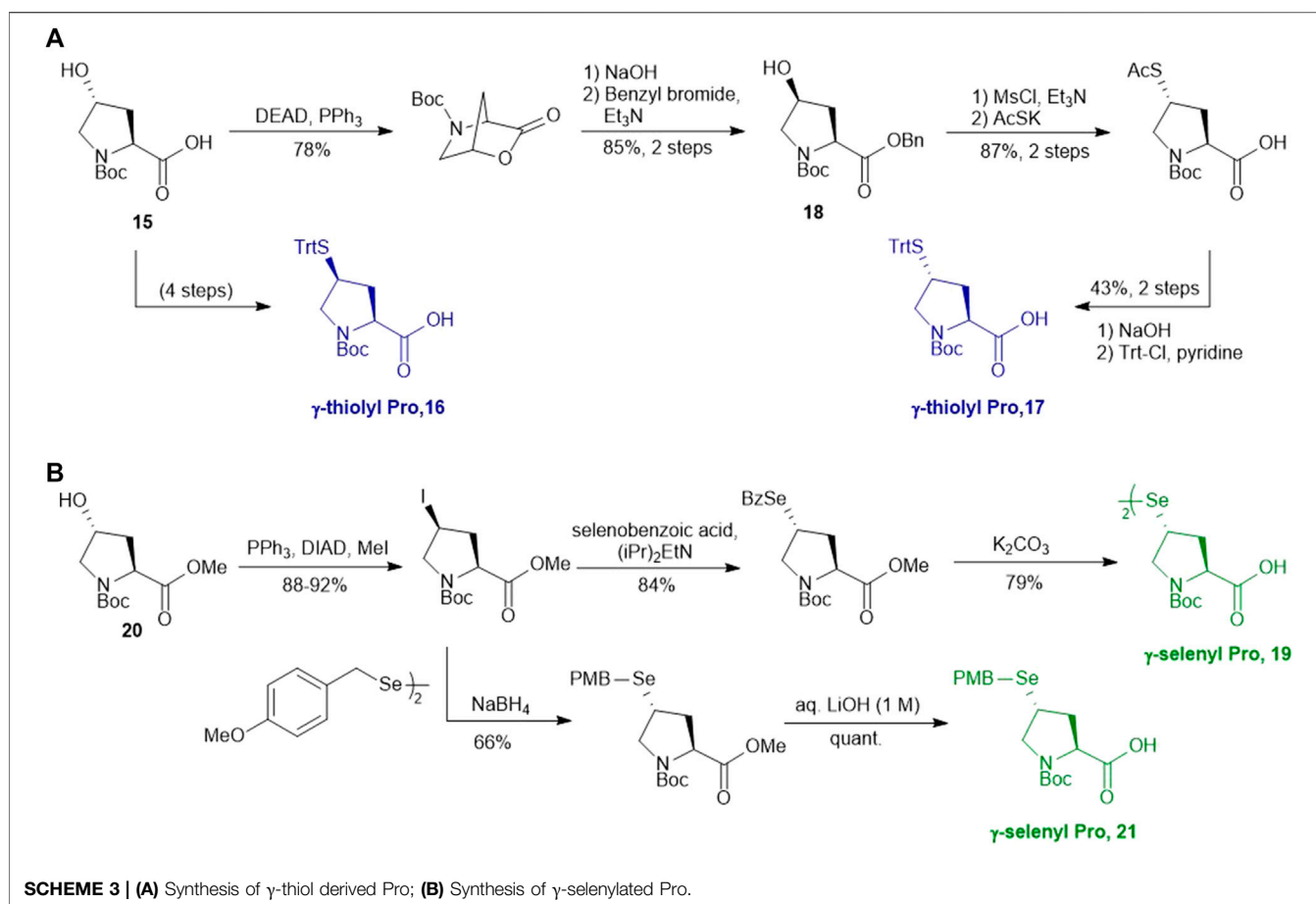
γ-Thiol/selenol Derived Proline

Compared to β-hydroxyl Leu, γ-hydroxyl proline (Pro) building blocks are more readily available, i.e., USD $3 per Gram of the Boc-protected *trans*-γ-hydroxyl Pro (**15**). Through mesylation and displacement by thioacetate, a *syn*-thiol Pro (**16**) can be readily prepared at a large scale (Ding et al., 2011). To afford *anti*-thiol Pro **17**, *syn*-hydroxyl Pro **18** was firstly produced via an intracellular Mitsunobu reaction followed by saponification (**Scheme 3A**). The carboxylic acid was then protected before thioacetate displacement to afford *anti*-thiolated Pro **16**. Interestingly, only the *anti*-isomer manifested satisfying reactivity in NCL reactions. The rat neuromedin U had been successfully prepared through *anti*-thiol Pro-mediated ligation followed by desulfurization.

Though the thiolated Pro **17** had been applied to the chemical protein synthesis of complex glycosylated proteins, its ligation efficiency was severely compromised by the presence of sterically hindered thioesters. In order to enhance the rate of inter- and intramolecular acyl transfer in Pro ligation, *anti*-selenylated Pro **19** was also prepared by Danishefsky and coworkers (**Scheme 3B**) (Townsend et al., 2012). The synthesis began with Mitsunobu inversion of the hydroxyproline **20** to provide a *syn*-iodoproline intermediate followed by nucleophilic displacement with selenobenzoic acid and hydrolysis to afford the desired

diselenide **19** in a 85% yield over two steps. The authors first examined the ligation efficiency of **19** under standard NCL conditions with excess MPAA as a thiol additive. MPAA served to activate the alkyl thioester through transthioesterification; it also reduced the selenium dimer, liberating a small amount of productive selenol to mediate the peptide ligation. Under these conditions, model ligation reactions were completed within 3 h in good to excellent yields. The relative ease of deselenization compared to desulfurization in model peptides were also demonstrated using Dawson's conditions (DTT, TCEP, ligation buffer, pH 5.0–6.0) (Townsend et al., 2012).

Recently, Sayers *et al.* demonstrated the capacity of *anti*-γ-selenyl Pro **21** in rapid DSL reactions (Sayers et al., 2018a) (See **Section 4**). Initially, they followed Danishefsky's route to prepare γ-selenyl Pro building block for SPPS. However, to overcome the issue of having only one Pro unit within the diselenide dimer being coupled to resin-bound peptides, monomeric γ-selenyl Pro **21** was prepared instead, wherein the γ-selenol moiety was masked with a PMB group to prevent oxidative dimerization. This building block can be efficiently incorporated into the N-termini of resin-bound model peptides using only a slight excess of **21** (1.2 equiv.) together with hydroxyazabenzotriazole (HOAt) (1.2 equiv.), N,N′-diisopropylcarbodiimide (DIC) (1.2 equiv.). Acidolysis of side-chain protecting groups with concomitant peptide cleavage from the resin was followed by deprotecting the PMB group in 20% DMSO, 5% iPr₃SiH in TFA. Purification by HPLC provided the desired peptide diselenide dimer in 50% overall yield.

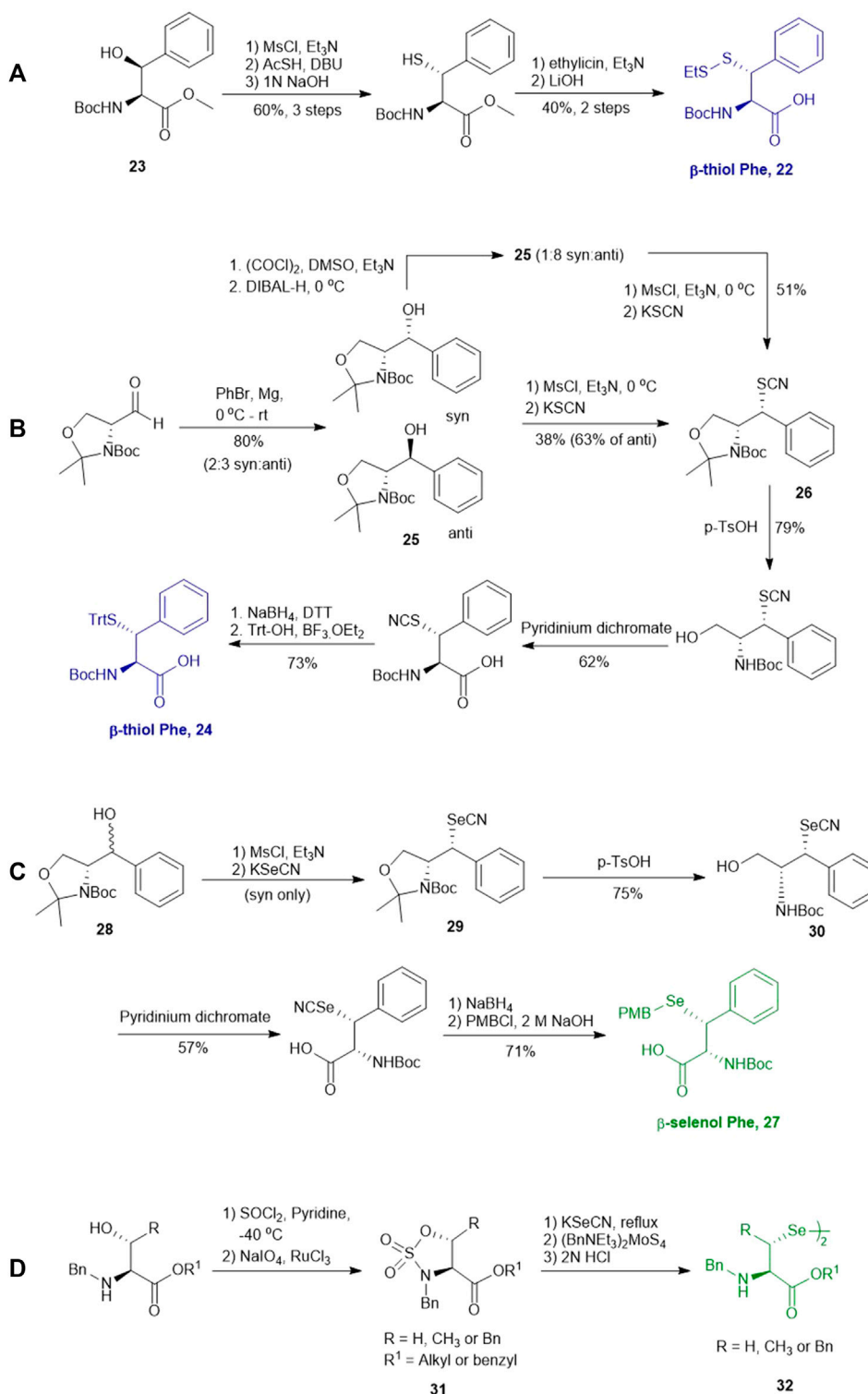**SCHEME 3 | (A)** Synthesis of γ-thiol derived Pro; **(B)** Synthesis of γ-selenylated Pro.

β-Thiol/selenol Derived Phenylalanine

In cases where β/γ-hydroxyl amino acids are not commercially available, the thiolated amino acid can be prepared from the corresponding proteogenic amino acids. Banerjee and coworkers began the synthesis of β-thiol phenylalanine (Phe) **22** by preparing the (2S,3S)-hydroxyl Phe **23** according to Easton's protocol (**Scheme 4A**) (Easton et al., 1994). The following synthesis was performed in analogy to the sulfenylation route mentioned above (via a mesylation–nucleophilic displacement manifold), affording the desired *anti*-β-thiol Phe **22**. The utility of **22** in chemical protein synthesis has also been demonstrated by the authors (Crich and Banerjee, 2007).

Payne and coworkers harnessed Garner's aldehyde to build a hydroxylated Phe precursor in order to access β-thiol Phe **24** (**Scheme 4B**) (Malins et al., 2015a). Firstly, the phenyl alcohol (**25**) was formed in the reaction between a Grignard reagent of bromobenzene and Garner's aldehyde as a mixture of two diastereomers in an excellent yield (80%). The moderate stereoselectivity is due to two competing mechanisms for asymmetric induction: the *anti*-configured isomer is postulated to proceed through the Felkin-Ahn transition state, whereas in the presence of Mg cations, the reaction can proceed via a Cram's chelation transition state resulting in the *syn* product (Passiniemi and Koskinen, 2013). Next, the alcohol

was mesylated followed by nucleophilic displacement by KSCN to afford the desired product in a yield of 38%. Only the *anti*-isomer of **25** could proceed to completion, which provided an opportunity to resolve the diastereomeric mixture by chromatographic purification. Notably, the unreactive *syn*-isomer of **25** was inverted via Swern oxidation followed by asymmetric reduction with DIBAL to provide more of the *anti*-isomer (Malins et al., 2015a). The oxazolidine moiety in **26** was removed using *p*-toluenesulfonic acid in dioxane to obtain the free alcohol in a good yield (79%). This alcohol was then oxidized with pyridinium dichromate to afford the carboxylic acid in a reasonable yield (62%). Finally, the thiocyanate was reduced and masked with a trityl (Trt) protecting group to afford the desired β-thiolated amino acid in 73% yield (Malins et al., 2015a).

Malins *et al.* reported the use of a similar strategy to access β-selenylated Phe (**27**), which started with the Grignard addition of phenylmagnesium bromide to Garner's aldehyde to provide **28** as an inseparable mixture of *syn* and *anti*-diastereoisomers (**Scheme 4C**) (Malins and Payne, 2012). This mixture was used in the next step, which involved mesylation of the secondary alcohol followed by nucleophilic displacement with potassium selenocyanate to afford the oxazolidine **29**. Notably, the isomers could be resolved by chromatography at this stage;

SCHEME 4 | (A) Synthesis of β-thiolated Phe via Easton's route. **(B)** Synthesis of β-thiolated Phe via Payne's route. **(C)** Synthesis of β-selenylated Phe reported by Malins et al.; **(D)** Asymmetric synthesis of β-selenylated Phe via a cyclic sulfamidate intermediate reported by Baig et al.

only the *anti*-mesylate was active in the substitution reaction leading to the desired *syn*-selenocyanate (**29**). Acid deprotection of the oxazolidine **29** followed by oxidation with pyridinium

dichromate yielded the carboxylic acid **30** in moderate yield without any sign of selenium oxidation. Finally, the selenocyanate moiety was reduced with sodium borohydride

followed by protection with 4-methoxybenzyl chloride to afford the desired β-selenylated Phe (**27**) in a good yield (72% over two steps).

Baig *et al.* reported the asymmetric synthesis of several cyclic sulfamidate from the respective β-hydroxyl amino acids (**Scheme 4D**) (Rashid Baig et al., 2010). A benzyl sulfamidate (**31**) was transformed to β-selenylated Phe **32** through nucleophilic displacement with KSeCN followed by reductive dimerization mediated by tetrathiomolybdate.

Thiol/Selenol Derived Lysine

γ-thiol lysine (Lys) was first synthesized by Liu and coworkers, using Asp as the precursor (Yang et al., 2009; Marin et al., 2004). Subsequently, they proposed a general, diastereoselective approach to prepare γ-thiol Lys derivatives (**Scheme 5A**) (Pasunooti et al., 2009). The side-chain carboxylate group of Asp **33** was first reduced to aldehyde, which allowed for asymmetric Reformatsky reaction to extend the side chain, leaving a hydroxyl group on the γ-carbon, followed by protection with *tert*-butyldiphenylsilyl group. The ester moiety in **34a** was reduced to alcohol that was subsequently activated and displaced by sodium azide to form **35**. The resulting azide group was reduced to form the ε-amine of Lys and protected with benzyl chloroformate. Finally, the γ-hydroxyl group was deprotected for sulfenylation through the mesylation−nucleophilic displacement manifold. This synthesis required 21 steps overall (including Asp protection), but a γ-thiol Lys building block (**36**) with defined stereochemistry could be achieved. The utility of **36** was demonstrated in the synthesis employing dual native chemical ligation to assemble ubiquitinated peptides, which will be further discussed in **Section 4**.

Brik and coworkers introduced a thiol functionality to the δ-carbon of Lys through conjugate addition (**Scheme 5B**) (Ajish Kumar et al., 2009). The side-chain carboxylate of Boc-Glu-OMe was first reduced to aldehyde followed by a Henry reaction to afford a *syn* and *anti* mixture of nitro alcohol **37**. The alcohol was acetylated and eliminated to provide a mixture of *trans*- and *cis*-nitro olefin as the Michael acceptor, which underwent electrophilic addition to *tert*-butylthiol to form a δ-thiolated Lys precursor (**38**) as a mixture of *syn*- and *anti*-isomers. The nitro moiety was reduced to the primary amine using a mixture of sodium borohydride and nickelous chloride, followed by protecting group modifications to afford the desired δ-thiol Lys (**39**).

Alternatively, thiolated Lys has been synthesized through the conversion of Lys to γ-chlorinated Lys **40** in concentrated hydrochloric acid under UV irradiation with continuous chlorine gas bubbling (Merkx et al., 2013; Van Der Heden Van Noort et al., 2017) (**Scheme 5C**). Despite low yield (11%), the reaction is highly stereospecific and regioselective. The α-amine and carboxylic acid of γ-chlorinated Lys (**40**) was subsequently masked by 9-BBN while the side-chain amine was protected with Boc anhydride. This was followed by the nucleophilic displacement of the chloride substituent by thioacetate (Merkx et al., 2013). Final protecting group manipulation afforded the desired γ-thiol Lys building block **41** for Fmoc-based SPPS.

As will be discussed in the later section, γ/δ-thiolated Lys has been frequently used in isopeptide ligation to facilitate site-specific incorporation of (poly) ubiquitin modifications into small proteins. However, this strategy poses a challenge in EPL, where a specialized expression strategy is required to incorporate protecting groups for cysteine residues native to the protein of interest. Therefore, the Metanis group sought to develop a γ-selenylated Lys that enables selenium-mediated isopeptide ligation followed by using site-specific deselenization chemistry to furnish the native Lys in the presence of unprotected Cys residues (**Scheme 6A**) (Dardashti et al., 2020). They reported the synthesis and application of γ-selenyl Lys **42** as a tool molecule in an isopeptide ligation-deselenization manifold as such, although the preparation of δ-selenyl Lys was not successful due to spontaneous selenium elimination. The synthetic route toward γ-selenyl Lys **42** started with preparing the suitably protected γ-chlorinated Lys **43** following the protocol reported by Merkx *et al.* (Merkx et al., 2013). Nucleophilic displacement by Na_2Se_2 was performed to generate **44** in moderate yield, in which the selenol moiety was subsequently masked by PMB group. Finally, unmasking the α-amine and α-carboxylate with ethylenediamine followed by Fmoc-protection gave the desired selenylated building block **42** for SPPS in 77% yield.

In parallel to this synthesis, the authors also set out to synthesize a γ-selenyl Lys analogue having an orthogonally protected ε-amine such that dual chemical ligation can be performed on the same residue, whereby the γ-selenol will first assist with constructing the main peptide chain *via* NCL followed by isopeptide ligation to append the ubiquitin protein to the ε-amine (**Scheme 6B**) (Dardashti et al., 2020). The authors were unable to incorporate an orthogonal protecting group on the ε-amine beginning with the same γ-chlorinated Lys building block; therefore, an alternative route was developed using aldehyde **45** as the starting material, which can be prepared from L-Asp. The aldehyde **45** was extended with a triphenylphosphorane reagent to afford the *trans*-ester **46** in 85% yield over two steps, which was subjected to conjugate addition reaction with PMB selenol generated *in situ* to provide compound **47** as a mixture of two stereoisomers. The side-chain ester was selectively reduced to a primary alcohol, which was subsequently activated and displaced by sodium azide to afford **48**. Staudinger reduction of azide **48** followed by protecting group manipulation provided the desired γ-Se-Lys **49** was afforded in 90% isolated yield over two steps.

γ-Thiol Derived Valine, Threonine and Isoleucine

A suitably protected Asp has been used in the synthesis of γ-thiol valine (Val) (**Scheme 7A**) (Chen et al., 2008). It began with introducing a [9-(9-phenylfluorenyl)] protecting group (PhFl) to the α-amine, which stabilized the α-chiral center and allowed for selective β-methylation leading to the amino acid **50** as a 1:1 mixture of *anti*- and *syn*-isomers. Selective reduction of the methyl ester produced γ-hydroxyl Val (**51**); and the *anti*- and *syn*-isomers were separated by chromatography. Finally,
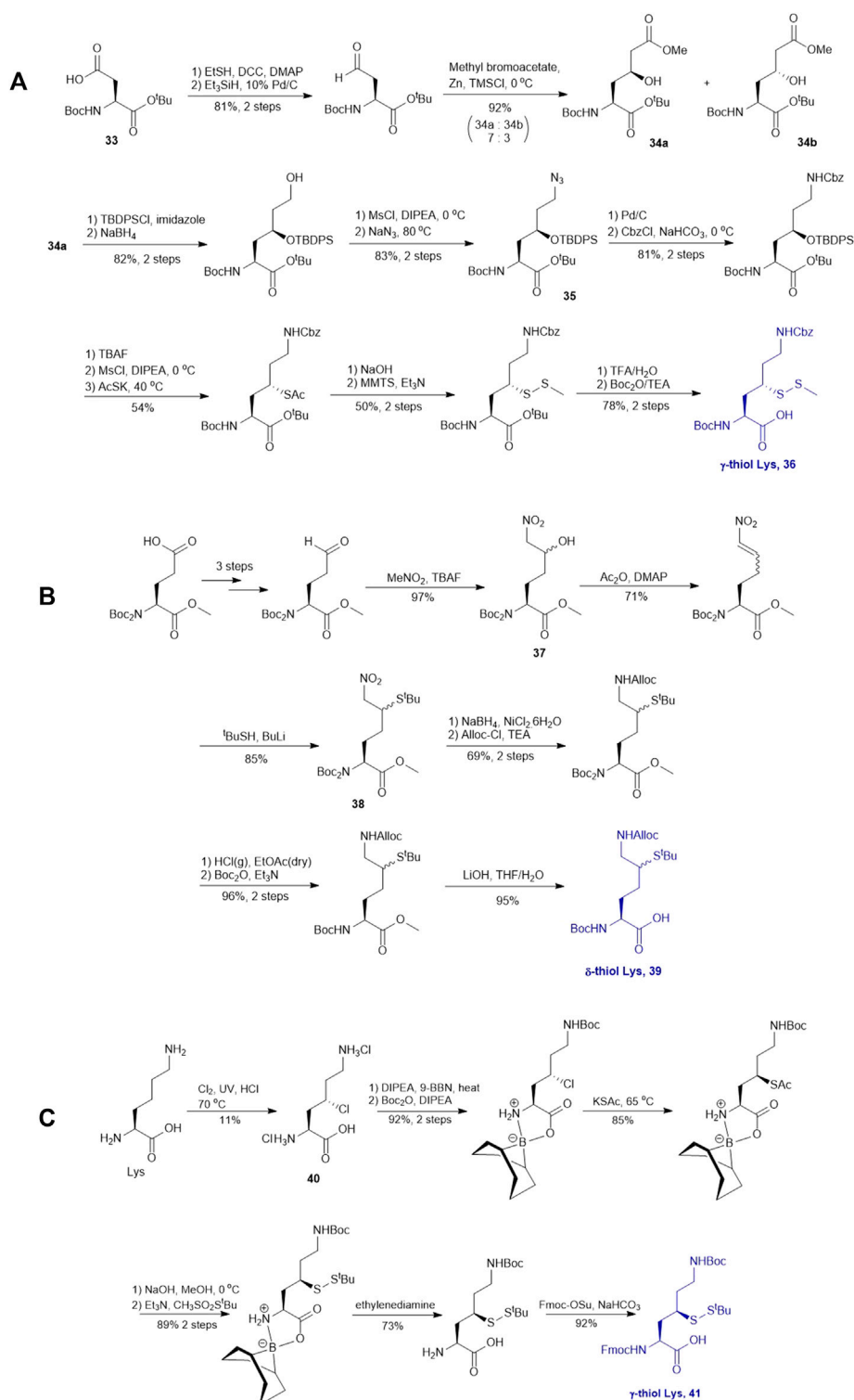
SCHEME 5 | (A) Synthesis of γ-thiolated Lys reported by Liu and co-workers; **(B)** Synthesis of δ-thiolated Lys via Brik's route; **(C)** Synthesis of γ-thiolated Lys reported by Merkx et al.

following the mesylation-nucleophilic displacement method, γ-thiol Val **52** was afforded in a good isolated yield (51% over five steps).

Vinylglycine (**53**) has been harnessed for the synthesis of γ-thiol threonine (Thr) through the formation of the *syn* epoxide intermediate **54** (Chen et al., 2010). Stereoinduction of the epoxidation reaction is

SCHEME 6 | Synthesis of two differentially functionalized γ-selenol Lys reported by Metanis and co-workers.
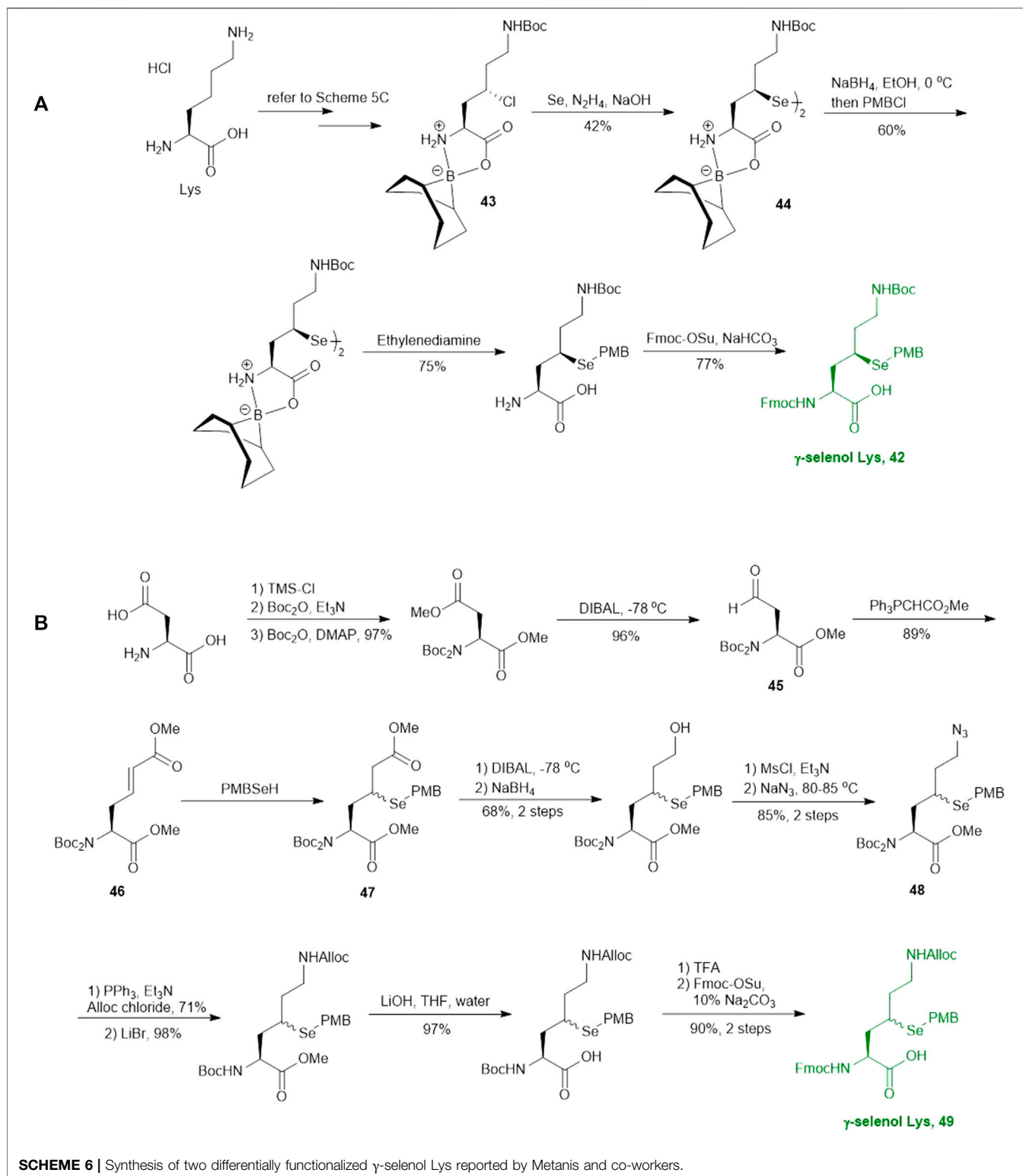
attributed to hydrogen-bond stabilization of the peroxide transition state (**55**) by the α-amine (**Scheme 7B**). The epoxide ring was subsequently opened by thioacetate to provide the Thr intermediate (**56**) in the native configuration bearing a γ-thioacetate moiety. The

final γ-thiol derived Thr **57** was afforded through several steps of protecting group manipulation. The reactivity of γ-thiol Thr has been demonstrated in model ligation studies using peptides bearing different C-terminal thioesters.

SCHEME 7 | (A) Synthesis of γ-thiol derived Val; **(B)** Synthesis of γ-thiol derived Thr; **(C)** Synthesis of γ-thiol derived Ile.
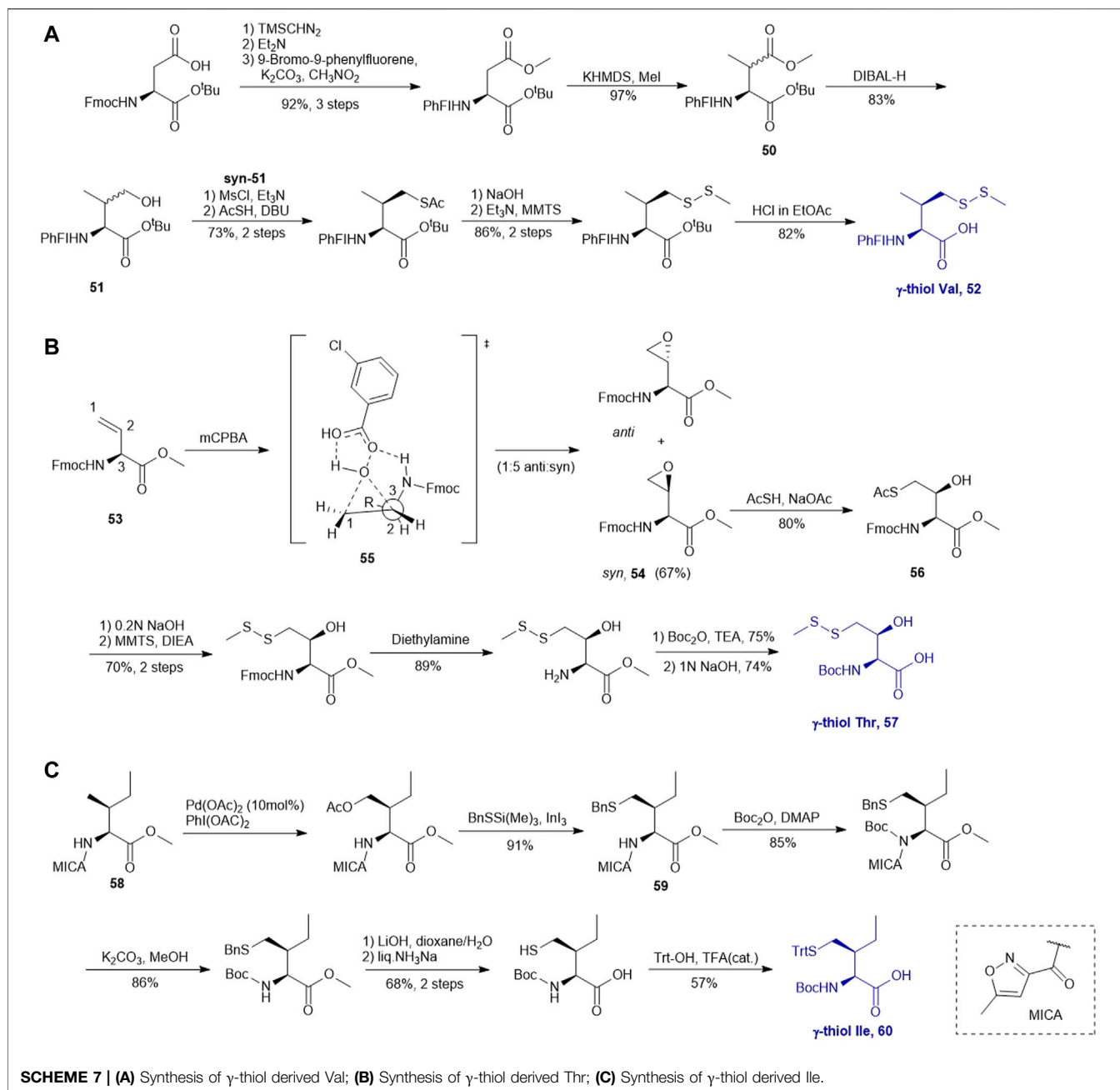
It is challenging to sulfenylate amino acids bearing hydrophobic or chemically unreactive side chains due to the lack of a directing group to facilitate the reaction. In 2015, Liu and coworkers had demonstrated a unique γ-C (sp$_3$)-H functionalization strategy using 5-methylisoxazole-3-carboxamide (MICA) as a directing group to activate γ-carbon (Pasunooti et al., 2015). This strategy had been successfully applied to synthesize γ-thiol isoleucine (Ile) for the first time (Pasunooti et al., 2016) (**Scheme 7C**). γ-thiol Val and γ-thiol Thr were also synthesized from their natural amino acid counterparts following a similar synthetic route. MICA was first introduced to the α-amine of a Ile amino acid to form **58**, which directed the

subsequent γ-acetoxylation. It was proposed that the MICA group first complexes with Pd^{II}, which is subsequently oxidized to Pd^{IV} by PhI (OAC)$_2$. The resulting complex proceeded through an octahedral transition state that exhibited an out-of-plane conformation favoring acetoxylation at the methyl branch of Ile (Buettner et al., 2019; Shabani et al., 2021). The resulting acetoxyl group was directly displaced by benzylthiol to form **59** in the presence of an indium triiodide catalyst (Nishimoto et al., 2012). The N-terminus of γ-thiol amino acid required Boc-protection before removing MICA using K$_2$CO$_3$ in MeOH. The final γ-thiol amino acid **60** was furnished after protecting group alternation. The utility of γ-thiol

SCHEME 8 | Synthesis of **(A)** β-thiolated Asp and **(B)** γ-thiolated Glu via electrophilic sulfenylation chemistry.
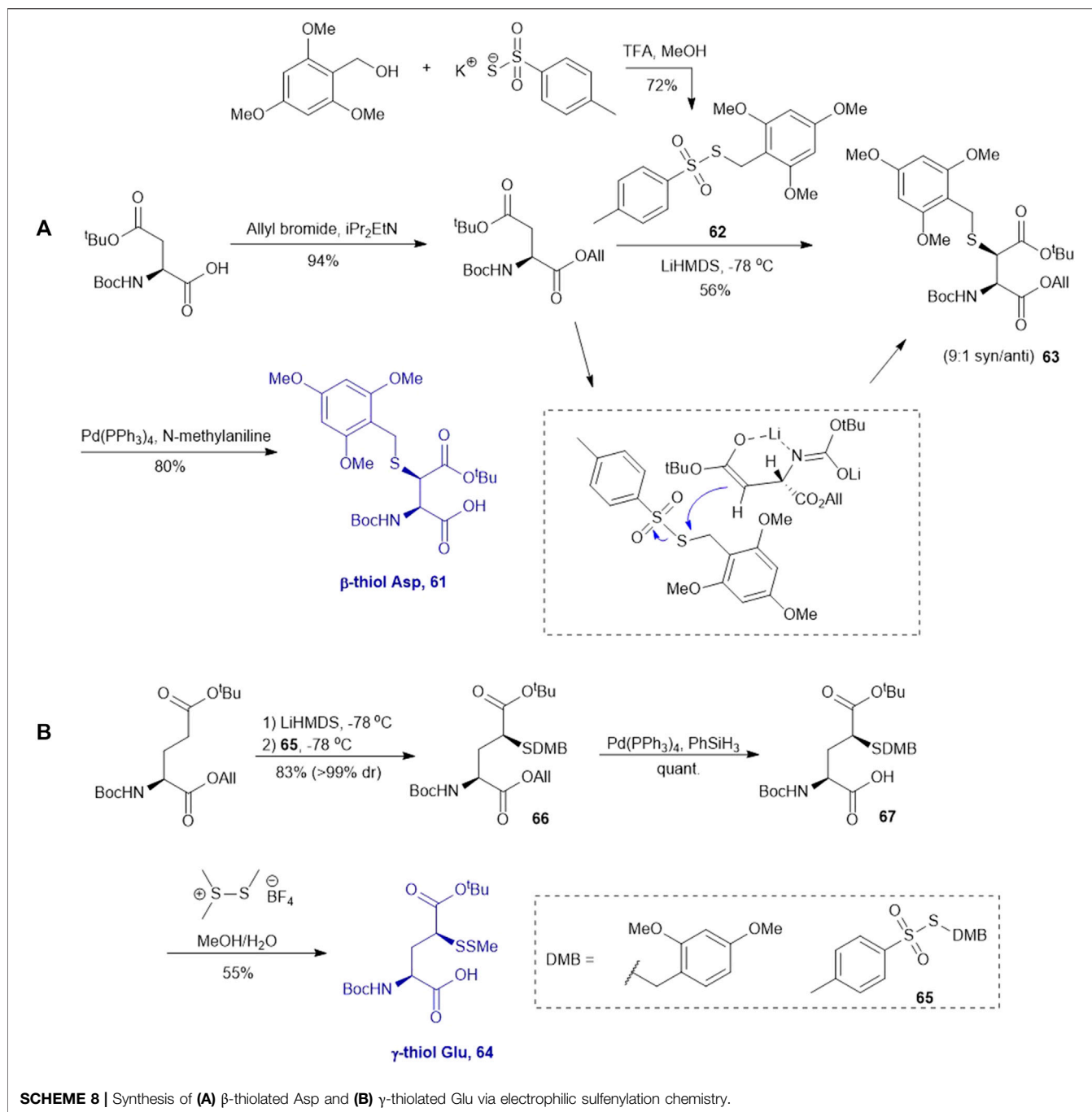
Ile (**60**) in peptide ligation has been demonstrated in the chemical synthesis of *Xenopus* H3.

Thiol/Selenol Derived Aspartate, Glutamate and Asparagine

Multiple thiolated amino acids have been synthesized *via* electrophilic sulfenylation chemistry due to its concise, direct approach to installing a thiol functional group on nucleophilic carbon in the side chain of amino acids. This approach has been proven effective in the synthesis of thiolated aspartic acid (Asp) and glutamic acid (Glu) with protected side chain esters that can be chemoselectively enolized for nucleophilic addition on an electrophilic sulfur center.

This approach was first conducted by Payne and coworkers in 2013 for the synthesis of β-thiol derived Asp **61**. (Thompson et al., 2013). They first prepared S-(2,4,6-trimethyoxybenzyl) toluenethiosulfonate (**62**), from 2,4,6-trimethylbenzyl alcohol and potassium toluene thiosulfonate (**Scheme 8A**). Using this sulfenylating reagent, the subsequent synthesis of Boc-Asp

(O^{t}Bu)-OH was conducted in three steps. Firstly, the free carboxylic acid was protected using allyl bromide (94%), then subjected to sulfenylation conditions, wherein the side-chain carboxylate ester was first enolized to form a lithium complex followed by nucleophilic attack at the electrophilic sulfur center of the sulfenylating reagent **62** at a low temperature leading to the formation of thiolated amino acid **63**. This reaction exploits the native α-chiral center to induce stereoselectivity. It proceeds through a 6-membered transition state where the Li coordinates with the enolate oxygen and protected nitrogen (**Scheme 8A** insert). This allows for preferential attack on the less hindered face, allowing for stereochemical control in favor of the *syn* configured diastereomer (9:1 *syn/anti*) (Shibata et al., 1996). The two diastereomers were separable by column chromatography to provide a stereochemically pure product in 56% yield. Finally, the allyl ester was deprotected using palladium tetrakis triphenylphosphine to obtain the desired thiolated amino acid **61** in an 80% yield (Thompson et al., 2013). The reactivity of β-thiol Asp was first evaluated in model ligation studies with various thioesters. These ligation reactions were conducted under standard NCL conditions and proceeded with high yields. They were also converted to the native peptides smoothly via Danishefsky's desulfurization approach. A one-pot ligation−desulfurization method was also developed with this amino acid. This approach also addressed the limitation of lacking chemoselective desulfurization chemistries in the presence of native, unprotected Cys residues. First the ligation was performed in identical denaturing conditions and the aryl thiol was subsequently removed by ether extraction upon ligation completion. TCEP and DTT were then added to initiate a desulfurization reaction occurring on the β-carbon at pH 3, 65°C for 20 h with reasonable yields ranging from 48–63%.

Electrophilic sulfenylation chemistry has also been applied to the synthesis of γ-thiolated Glu **64** through enolization of the side-chain carboxylate ester and sulfenylation *in situ* with **65** to afford **66** in a 83% isolated yield as a single diastereomer (Cergol et al., 2014). Product **66** was subjected to allyl deprotection conditions to afford the free acid **67** in quantitative yield (Cergol et al., 2014). This DMB protected thiolated amino acid **67** was incorporated into the N-terminal position of a model amino acid, however after acidolytic cleavage and deprotection, no desired product was isolated. The authors rationalized that acidolytic deprotection of the DMB group liberated the γ-thiol that could undergo nucleophilic attack at the amide bone leading to the deletion of this thiolated residue (Cergol et al., 2014). Therefore, the DMB protecting group was switched to an acid-stable but reductively labile methyl disulfide protecting group, which allowed for stable incorporation into peptide N-termini for NCL reactions (**Scheme 8B**). The utility of **64** in NCL reactions has been demonstrated in model ligation studies using a variety of peptide thioesters and in chemical synthesis of a known peptidic drug using a one-pot ligation−desulfurization method.

Similar to the synthetic methods mentioned above, Payne and coworkers also prepared β-thiol asparagine (Asn) using similar electrophilic sulfenylation chemistry (**Scheme 9A**) (Sayers et al., 2015). They began with the readily available Boc-Asp (OAll)-OH which was protected with phenylisopropyl trichloroacetimidate

to mask the α-carboxylate. The side-chain carboxylate was enolized by LiHMDS and reacted with the sulfenylating reagent **62** to form the thiolated derivate **68** as a mixture of two diastereomers (*syn:anti*, 9:2, **Scheme 9A**). The allyl ester was deprotected to liberate the free carboxylic acid, which was submitted to amination reaction containing Boc anhydride, ammonium bicarbonate, and pyridine to form the carboxamide side chain of **69**. The resulting *syn*- and *anti*-isomers were separable by column chromatography. From this, they attempted to directly couple this to a resin-bound peptide, however they found that lack of protection of the side chain led to quantitative succinimide formation. To suppress this side reaction, they used oxidative cyclization chemistry to form the final 2,4,6-trimethoxyphenyl-thiazolidine protected Asn **70** through generating a benzylic carbocation that underwent electrophilic addition onto the α-amine. This resulting building block was incorporated into peptides on resin smoothly, and was readily deprotected in 89:5:5:1 TFA:iPr$_3$SiH:H$_2$O:EDT to liberate the free thiol for NCL chemistry (Sayers et al., 2015).

Another synthetic route to β-thiol derived Asp was recently developed by Tan and coworkers harnessing the chemistry of direct thiylation of a protected Asp (**Scheme 9B**) (Guan et al., 2013). They first prepared the thiobenzamide ester (**71**) using 2-[(phenylthioxomethyl)thio]-acetic acid and triethylamine in pyridine (97%). They then formed the *trans*-thioazoline in a two-step synthesis involving side-chain enolization using LiHDMS followed by electrophilic iodination of the β-carbon, which was subsequently attacked by the sulfur to form a thiazolidine (**72**) in 38% yield. Through this process, no *cis*-isomer was observed and stereochemistry was maintained throughout the remaining synthetic steps (Harpaz et al., 2010). Next, the thiazoline was hydrolyzed in 2M HCl to afford the N-benzoyl derivative (**73**) in 54% yield. Finally, protection of the free thiol with a trityl group together with other protecting group transformation reactions furnished the desired building block **74**.

Analogous to the electrophilic sulfenylation chemistries utilized in the preparation of thiolated Asp (**61**) and Glu (**64**) building blocks, the synthesis of the suitably protected β-selenol derived Asp (**75**) began with enolization of the side-chain ester of Asp **76** followed by nucleophilic addition to the electrophilic selenium center within the PMB selenosulfonate **77** at a low temperature, which provided the selenylated amino acid **78** in an 81% yield as a mixture of two diastereomers (85:15 *syn:anti*, **Scheme 9C**). A final deallylation reaction afforded the desired building block **75** as a mixture of *syn* and *anti*-products in excellent yield. A similar synthetic route was also established for the suitably protected β-selenol derived Glu **79**. On this occasion, **79** was prepared as a single stereoisomer (2*S*, 4*S* as determined by NMR spectroscopy) (Mitchell et al., 2017).

Thiol Derived Tryptophan, Arginine and Glutamine

Payne and coworkers also reported an innovative approach to sulfenylate N-terminal tryptophane (Trp) residues in unprotected peptides (**Scheme 10A**). They first prepared a model peptide
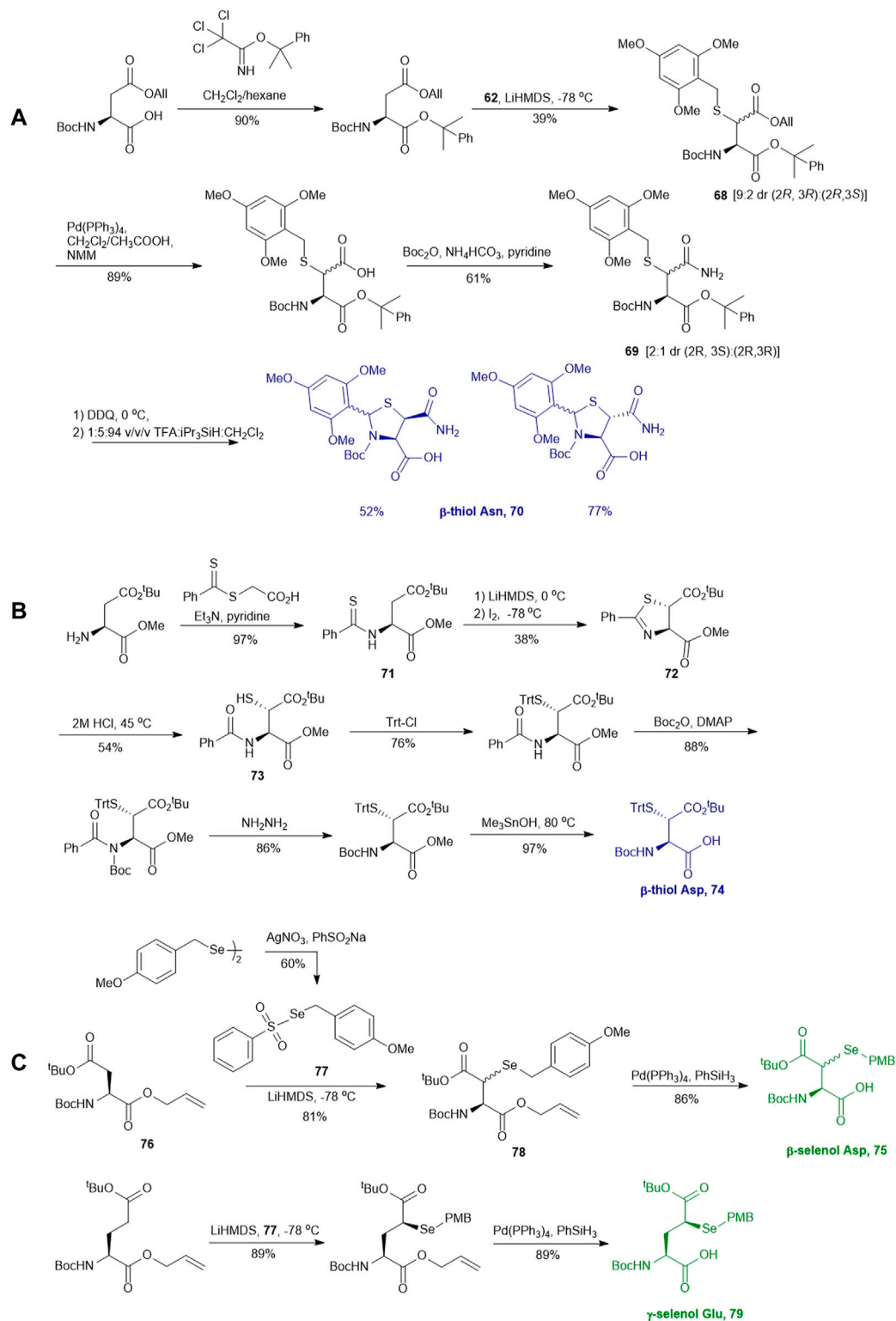
SCHEME 9 | (A) Synthesis of β-thiol derived Asn via electrophilic sulfenylation chemistry; **(B)** Tan's route to prepare β-thiol derived Asp; **(C)** Synthesis of β-selenol derived Asp and γ-selenol derived Glu via electrophilic selenylation chemistry.
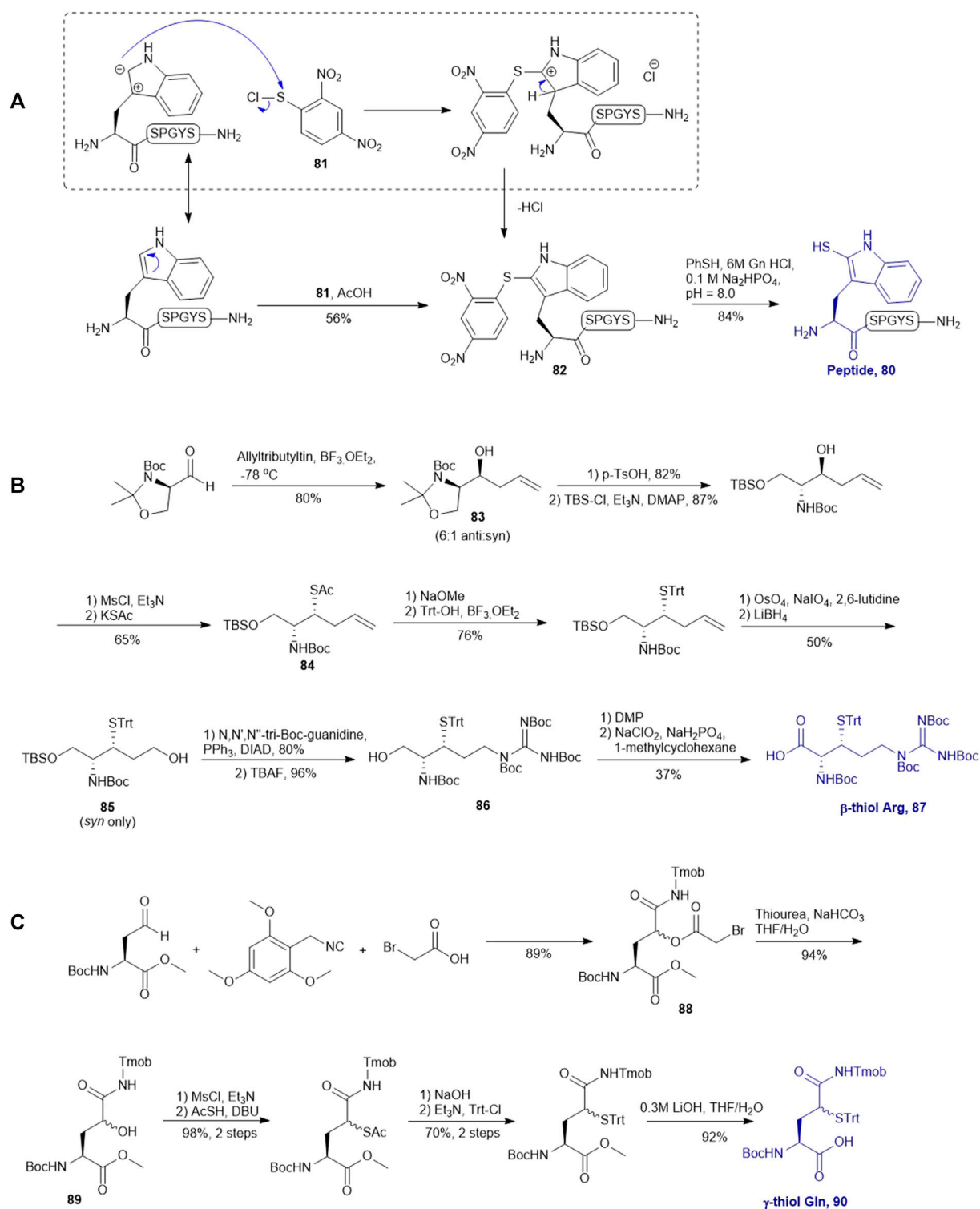
SCHEME 10 | (A) Electrophilic sulfeylation of peptidyl Trp; **(B)** Synthesis of β-thiol derived Arg based on Garner's aldehyde; **(C)** Preparation of γ-thiol derived Gln via a Passerini three-component reaction.

bearing an N-terminal Trp residue using Fmoc-based SPPS method. The peptide was then subjected to electrophilic aromatic substitution conditions containing 2,4-dinitrophenylsulfenyl chloride and acetic acid to furnish the Trp-sulfenylated peptide **80** in a 56% isolated yield. The reaction proceeded through electrophilic addition to the 2-position of the indole ring by the sulfenyl chloride **81** followed by

hydrogen elimination to re-form the aromatic ring in compound **82** (Malins et al., 2014). This modification can be effectively achieved both on and off resin.

Malins *et al.* described the chemical synthesis of β-thiol Arg based on Garner's aldehyde (Malins et al., 2013), which included eight high-yielding synthetic steps (**Scheme 10B**). They first generated an allyl alcohol via nucleophilic addition of allyltributylin to Garner's aldehyde (Garner, 1984). This reaction proceeded with reasonable stereochemical control, obtaining an inseparable mixture of 6:1 *anti* and *syn* diastereomers (**83**). The *anti*-isomer was generated through a Felkin-Ahn transition state which means that the nucleophile preferentially attacked on the *Re*-configured, less sterically hindered face. The thiazolidine was then deprotected using *p*-toluenesulfonic acid in 1,4-dioxane over 3 h to obtain the free alcohol in a good yield of 82%. After masking the primary alcohol with a TBS moiety and the secondary alcohol was transformed to thioacetate via a mesylation–nucleophilic displacement manifold to provide the *syn*-product **84** in a reasonable yield (62% over 2 steps). The thioacetate protecting group was then converted to an S-Trt protecting group while the allyl group was transformed into an alcohol (**85**) via oxidative cleavage with OsO$_4$ and NaIO$_4$ followed by reduction with LiBH$_4$. At this stage, it was noted that the minor *anti*-isomer formed earlier could be separated from the major product. Next, the *N,N,N*-tri-Boc-guanidine was installed onto **85** via a Mitsunobu reaction to obtain the protected guanidine side chain with a yield of 80%. Next the TBS protected alcohol was deprotected via treatment of tetrabutyl-ammonium fluoride (TBAF) to afford the free alcohol (**86**) in a great yield of 93%. Finally, a two-step sequential oxidation strategy was adopted to afford the desired β-thiol arginine (Arg) **87** in a 37% yield (Malins et al., 2013).

Brik reported the preparation of γ-thiol Gln based on Passerini three-component reaction (Siman et al., 2012) inspired by Kung's approach on the synthesis of γ-fluorinated Gln (**Scheme 10C**) (Qu et al., 2011). The side chain of the commercially available Asp was reduced to an aldehyde that was reacted with 2,4,6-trimethoxybenzyl isocyanide and bromoacetic acid to obtain a diastereomer mixture of the γ-acyloxy glutamine (Gln) derivative **88** (*syn/anti*, 1:1) in an 89% yield. (**Scheme 10C**) The bromoacetyl group was then removed using thiourea under basic conditions to give the γ-hydroxyl Gln (**89**), which was subjected to the mesylation-nucleophilic displacement manifold followed by protecting group manipulation to afford the desired γ-thiol derived Gln (**90**) as a mixture of two epimers. This isomeric mixture was used without chromatographic separation for SPPS. Peptide ligation using **90** has been demonstrated on model peptides.

Photoredox-Catalyzed Asymmetric Giese Reaction

Recently, Wang and coworkers have reported the synthesis of a library of enantiopure β-thiolated/selenylated amino acids via a photoredox-catalyzed asymmetric Giese reaction (Yin et al., 2020) (**Scheme 11**). This approach employed dehydrocysteine and dehydroselenocysteine amino acids protected as thiazoline

91 and selenazoline **92**, respectively. The side chain extension occurred at the β-carbon *via* radical-based 1,4-conjugate addition stimulated by visible light, leading to the formation of a β-thiolated/selenylated amino acid framework. This strategy enabled the preparation of a broad range of β-thiolated/selenylated amino acids on a multigram scale, which included natural and unnatural analogues.
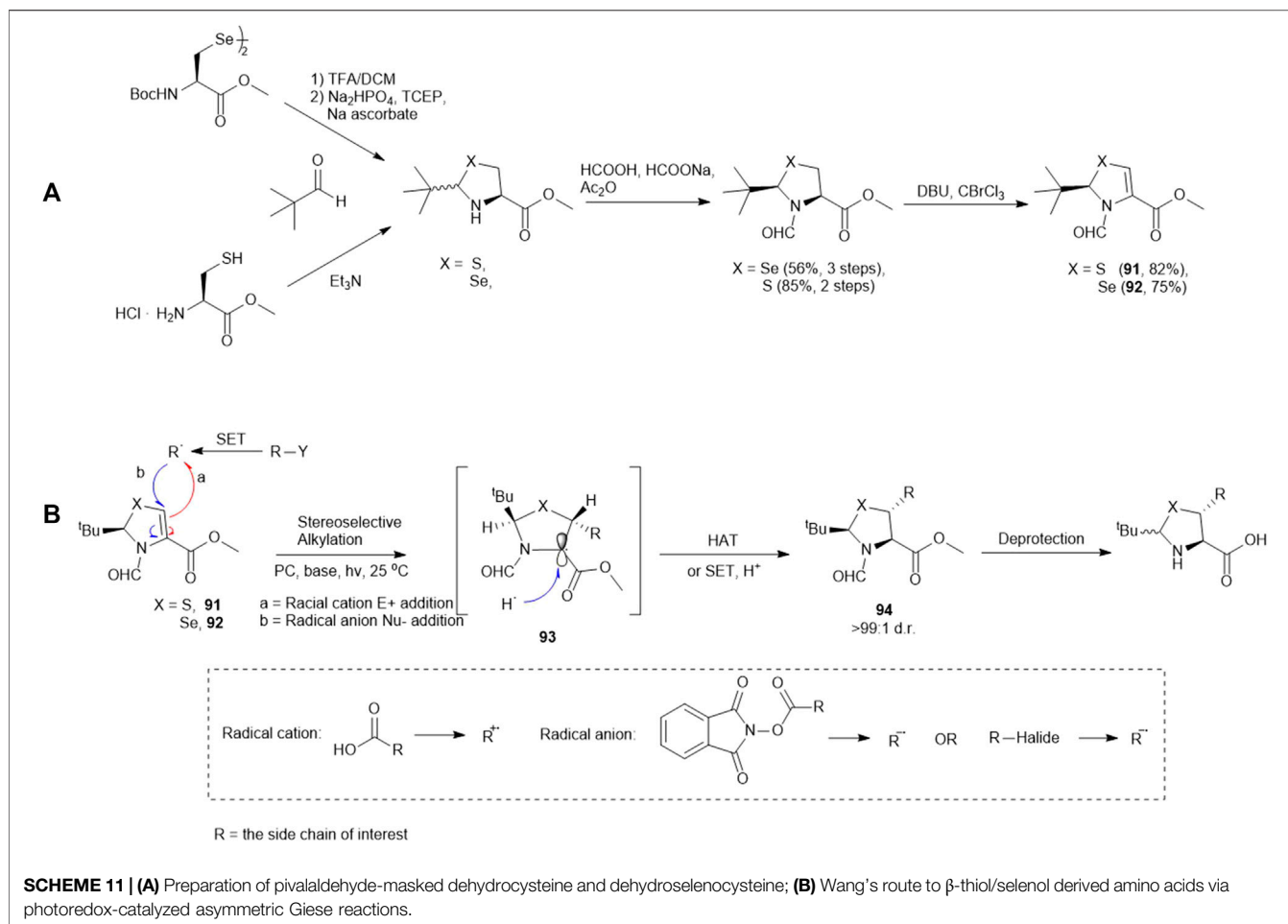
The thiazoline (**91**) and selenazoline (**92**) precursors were easily accessible from the protection of Boc-Sec-OMe and Boc-Cys-OMe with pivalaldehyde followed by hydride elimination to form **91** and **92**. (**Scheme 11**) The chirality of the auxiliary group determined the stereochemical outcome of the radical addition reaction, whereby the alkyl radical (R˙) preferred to approach the less hindered *Si*-face of the prochiral β-carbon center. Alkyl radicals can be generated from carboxylic acids (radical cations), redox-active esters (RAEs) or alkyl halides (radical anions) via single-electron transfer (SET) with a suitable photoredox catalyst and light source. The radical intermediate **93** generated in the radical addition was subsequently reduced via hydrogen atom abstraction (HAT) or SET to provide the product **94**. Subsequent removal of the formyl moiety and saponification furnished the desired β-thiolated/selenylated amino acids for SPPS. The *tert*-butylmethyl auxiliary group can be retained to facilitate one-pot, multi-component ligation methods, and removed under acidic conditions after peptide ligation.

Wang's method represents a generic approach to prepare β-thiolated/selenolated amino acids using Cys and Sec as universal donors. Apart from synthesizing the thiolated/selenylated variants of proteogenic amino acids, the modified forms of unnatural amino acids were also generated, including those with alkyl (primary, secondary and tertiary), pyridine, acetal, ketone and amide substituents on the side chains, suggesting the high structural and chemical tolerance of this method.

SELECTED EXAMPLES OF USING THIOLATED AND SELENYLATED AMINO ACIDS IN CHEMICAL SYNTHESIS OF POST-TRANSLATIONALLY MODIFIED PEPTIDES AND PROTEINS

β-Thiol Derived Aspartic Acid

Watson *et al.* reported the implementation of β-thiolated Asp **61** in the total synthesis of post-translationally sulfated anopheline proteins derived from *Anopheles* mosquitos (Watson et al., 2018). Selected sulfoproteins from *A. gambiae* and *A. albimanus* were first expressed in *Trichoplusia ni* insect cells and directed to the secretory pathway by a honeybee mellitin leader sequence. The secretory anopheline proteins in cell medium were submitted to tandem mass spectrometry analysis to reveal the tyrosine (Tyr) sulfation sites (Watson et al., 2018). They next sought to systematically characterize the anticoagulant activity of individual sulfation forms of each anopheline protein and identify the most potent sulfated form for therapeutic development. To this end, they initially attempted to integrate

SCHEME 11 | (A) Preparation of pivalaldehyde-masked dehydrocysteine and dehydroselenocysteine; **(B)** Wang's route to β-thiol/selenol derived amino acids via photoredox-catalyzed asymmetric Giese reactions.
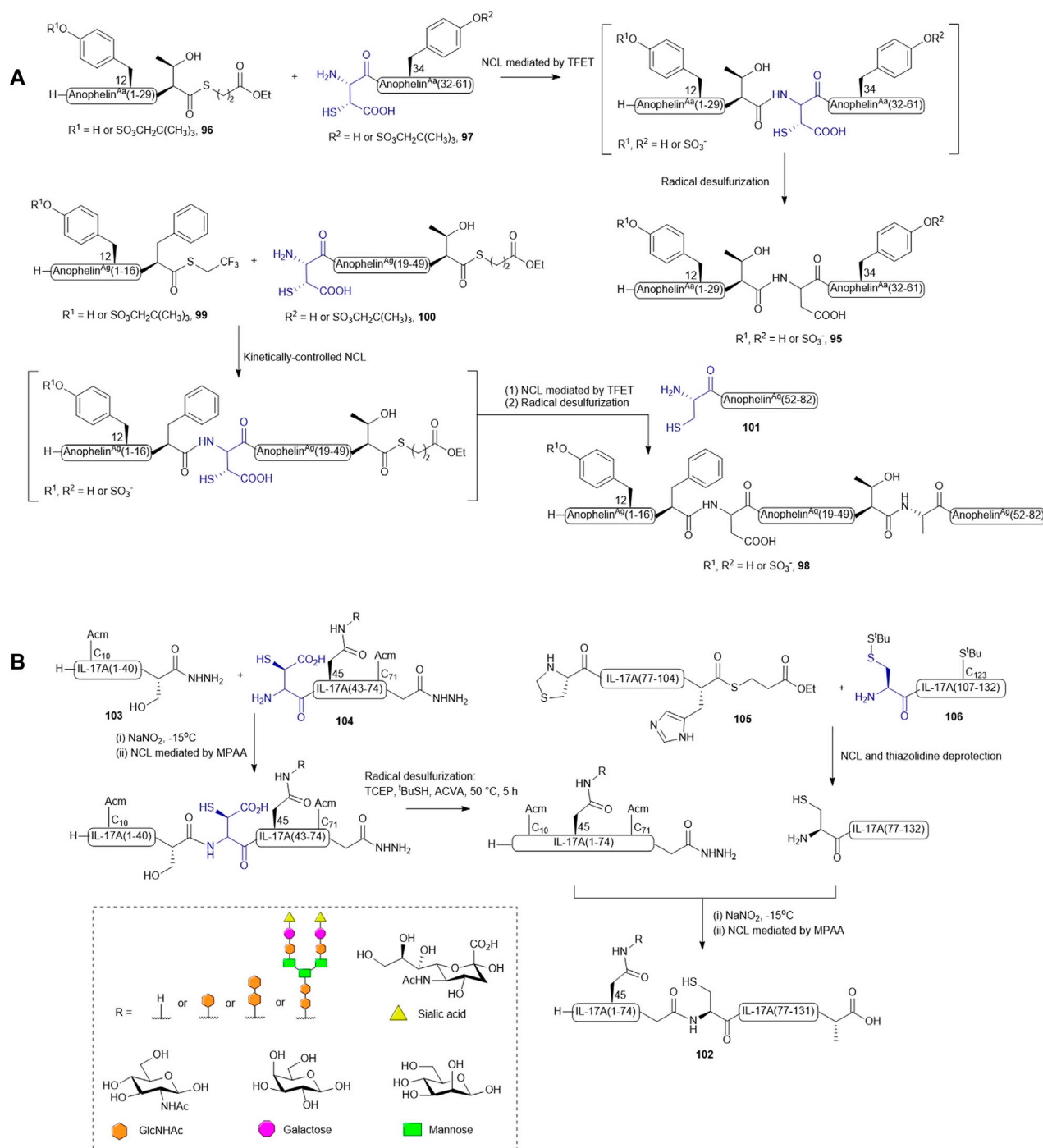
amber-stop codon suppression technology with a cell-free expression system to generate anopheline proteins with amber-codon incorporation of Tyr sulfates. However, this synthetic pipeline was unable to produce the doubly sulfated *A. albimanus* protein and cannot yield sulfoproteins in sufficient quantity for detailed *in vitro* and *in vivo* characterization. Therefore, it was decided to chemically synthesize the sulfoprotein libraries of anopheline proteins from *A. albuimanus and A. gambiae* under a ligation-desulfurization manifold (Watson et al., 2018).

As natural anopheline proteins do not possess a cysteine residue that enables peptide assembly through NCL, the retrosynthetic strategy was therefore designed based on ligation−desulfurization methodology (**Scheme 12A**). *A. albimanus* protein **95** was disconnected at the center between Thr30 and Asp31, resulting in two fragments: the N-terminal peptide thioester fragment (**96**) containing either sulfated or unsulfated Tyr12 and a C-terminal fragment (**97**) composed of a β-thiol Asp residue as ligation handle and a sulfated or unsulfated Tyr34 (Watson et al., 2018). These fragments were ligated in aqueous denaturing buffer aided by trifluoethanethiol (Thompson et al., 2014). The ligation reactions proceeded to completion within 16 h followed by radical desulfurization

chemistry mediated by VA-044, TCEP and reduced glutathione, which ultimately afforded the desired proteins in good yields. A sequential ligation−desulfurization methodology was implemented in the synthesis of sulfated *A. gambiae* protein **98** (Watson et al., 2018). The authors produced three suitably functionalized peptide fragments via Fmoc-based SPPS for convergent assembly of the desired protein: the N-terminal fragment **99** containing a C-terminal TFET thioester with the desired sulfated Tyr12 protected by a neopentyl group, the middle fragment (**100**) possessing an N-terminal β-thiol-derived Asp residue and a C-terminal deactivated alkyl thioester, and the last fragment (**101**) containing a cysteine residue as a ligation handle in place of the native Ala residue (Watson et al., 2018). The sulfoprotein was synthesized in a one-pot manner beginning with kinetically-controlled, sequential assembly of the three fragments from an N-to-C direction followed by radical desulfurization chemistry to eliminate the thiol appendages. All sulfated anopheline proteins prepared through this approach exhibited higher inhibitory activities against thrombin *in vitro* and *in vivo* compared to unmodified counterparts (Watson et al., 2018).

Recently, β-thiolated Asp has also been used for the design and synthesis of a homogenously N-glycosylated interleukin-17A (IL-17A) with the aim to elucidate the impacts of N-glycan on protein

SCHEME 12 | (A) Synthesis of sulfated anophelin variants via one-pot ligation-desulfurization chemistry; **(B)** Convergent synthesis of homogeneous glycoforms of IL-17A.

folding, thermal stability, and inflammatory cytokine inducing ability (Li et al., 2021b). In particular, the ligation-desulfurization methodology provided a powerful means to access individual glycosylated forms of IL-17A that are otherwise difficult to recombinantly derive. IL-17A (**102**) was disconnected into

four peptide segments prepared through Fmoc-based SPPS method, and methionine (Met) residues were mutated to norleucines to prevent unwanted oxidation. Both segments (1-41) **103** and (42-75) **104** contained a C-terminal peptide hydrazine which serves as a latent thioester, whilst segment

(76-105) **105** was modified with a C-terminal alkyl thioester (**Scheme 12B**) (Fang et al., 2011; Li et al., 2021b). A β-thiol-derived Asp was introduced to the N-terminus of segment (42-75) **104**, which permits a NCL reaction to form the first half of IL-17A (1–75), whereas two orthogonally protected cysteines were adopted in preparing the N-terminal residues of segments (76-105) **105** and (106–132) **106**. The peptide assembly of each glycosylated form of IL-17A began with performing NCL between segments (75-105) **105** and (106–132) **106** and peptide hydrazine ligation between segments (1-41) **103** and (42-75) **104** in parallel to furnish the first and second halves of IL-17A. Unmasking of the thiazolidine-protected Cys76 enabled the final convergent assembly of the two main fragments (1–75) and (76-105) to provide the full-length IL-17A **102** (Li et al., 2021b).

This convergent approach provided plasticity in the synthesis of discrete glycosylated forms of IL-17A since each segment could be modified and optimized independently to maximize the synthetic yield and PTM diversity. For peptide **104** derived with N-acetyl glucosamine (GlcNAc) or a disaccharide [GlcNAc(β1→4)GlcNAc], the respective N-glycosylated Asn amino acids were incorporated as Fmoc-protected building blocks in SPPS. However, to obtain the glycoform of **104** bearing a complex biantennary sialyloligosaccharide, the N-glycan chain was developed on the GlcNAc moiety within the target peptide using the Endo-M-catalyzed transglycosylation method. The N-terminal thio-aspartyl residue in the glycopeptide can still mediate the ligation reaction smoothly without an observable impact on the integrity of the glycan. However, an inseparable by-product containing a VA-044 adduct on the β-thiolated Asp residue was formed in the radical desulfurization reaction using Danishefsky's conditions. The authors postulated that the carboxylate side chain of the thioaspartyl residue forms a tight ion pair with the basic 4,5-dihydroimidazole radical generated from homolytic cleavage of VA-044, leading to competitive interception of the radical propagation chain mediated by sulfhydryl donors and TCEP. To their delight, using a different radical initiator ACVA possessing electronegative carboxylates that could repel the binding of Asp significantly diminished this side reaction and allowed completion of the total synthesis. The resulting homogeneous glycoproteins have enabled detailed structure-based evaluation of protein stability, proinflammatory activities, and receptor binding abilities (Li et al., 2021b).

β-Thiol Derived Arginine

Payne and coworkers highlighted the effectiveness of their β-thiol Arg amino acid in NCL chemistry by designing and synthesizing the glycopeptide (**107**) of the extracellular domain of mucin 1 (MUC1) (**Scheme 13**) (Malins et al., 2013). They employed a one-pot kinetically-controlled ligation method to assemble the target of interest, where fragment one (**108**) was a glycopeptide bearing an activated phenyl thioester, the middle glycopeptide fragment contained an N-terminal β-thiol Arg and a C-terminal alkyl thioester (**109**), and the last fragment also possessed an N-terminal β-thiol Arg as the ligation handle (**110**). Taking full advantage of the high reactivity of phenyl thioesters in

NCL, the intermolecular ligation reaction between **108** and **109** proceeded with much faster kinetics than intramolecular cyclization of **109**. Upon completion, the final fragment (**110**) was added along with 2 vol% thiophenol to activate the final ligation, which went to completion in 31 h to produce the glycopeptide (**107**) in 43% yield after purification. This peptide was then desulfurized to afford the native glycopeptide in 38% isolated yield (Malins et al., 2013).

δ-Thiol Derived Lysine

Liu and coworkers demonstrated the use of γ-thiol derived Lys in a dual peptide ligation methodology that enabled streamlined synthesis of ubiquitinated peptides and proteins (Yang et al., 2009). For example, a γ-thiol derived Lys having a Cbz-protected side chain was introduced to the N-terminus of a model peptide (**111**), allowing for the construction of the primary sequence of the target peptide (**111**) via NCL (**Scheme 14A**). The Cbz protecting group was then removed using a chilled cocktail of TFMSA/TFA/p-cresol/methyl phenyl sulfide (1:7:1:1, v/v/v/v). After HPLC purification, isopeptide ligation was performed by adding a ubiquitin mercaptoethanesulfonate (MES) thioester **112**. The reaction was complete within 45 min followed by desulfurization to furnish the ubiquitinated peptide **113** in excellent yield (90%).

δ-thiolated Lys has also been extensively utilized by other research groups for site-specific incorporation of a ubiquitin protein. Brik and coworkers published a synthetic route to all possible di-ubiquitin variants linked through disparate Lys-isopeptide bonds (**Scheme 14B**) (Kumar et al., 2010). The first ubiquitin unit within the dimer construct was disconnected at the Phe-Ala junction to afford two peptide segments (**114 and 115**): the N-terminal fragment **114** contained a phenylalaninyl MES thioester, whereas an unprotected Cys residue was integrated as the N-terminal residue of fragment **115**, and a δ-thiol Lys residue was introduced to either peptide segment according to the isopeptide linkage of preference. Each ligation was performed under standard NCL conditions, proceeding to completion after 7 h. This was followed by *in situ* deprotection of the side chain of the δ-thiol Lys residue to afford the desired ubiquitin unit (**116**) with a thiol-derived Lys residue at a designated position. The second ubiquitin unit was again divided at the Phe-Ala junction: the N-terminal segment was derived with a phenylalanyl thioester (**117**), whereas the C-terminal segment (**118**) was equipped with a ligation handle and a latent thioester (N-methylcysteine). After the ligation reaction, the crude product was incubated with 20% mercaptopropionic acid for 16 h at 40°C to afford the desired ubiquitin thioester **119** for final ligation. To assemble all the di-ubiquitin variants, seven peptide ligation reactions were carried out in parallel. Plausibly due to the large protein size, each reaction required more than 24 h to afford a moderate isolated yield (35–40%). The authors also noticed that not all the thiol auxiliaries could be removed under radical desulfurization conditions and thus, H$_2$/Raney nickel conditions in combination with TCEP treatment were adopted to desulfurize the di-ubiquitin variants.
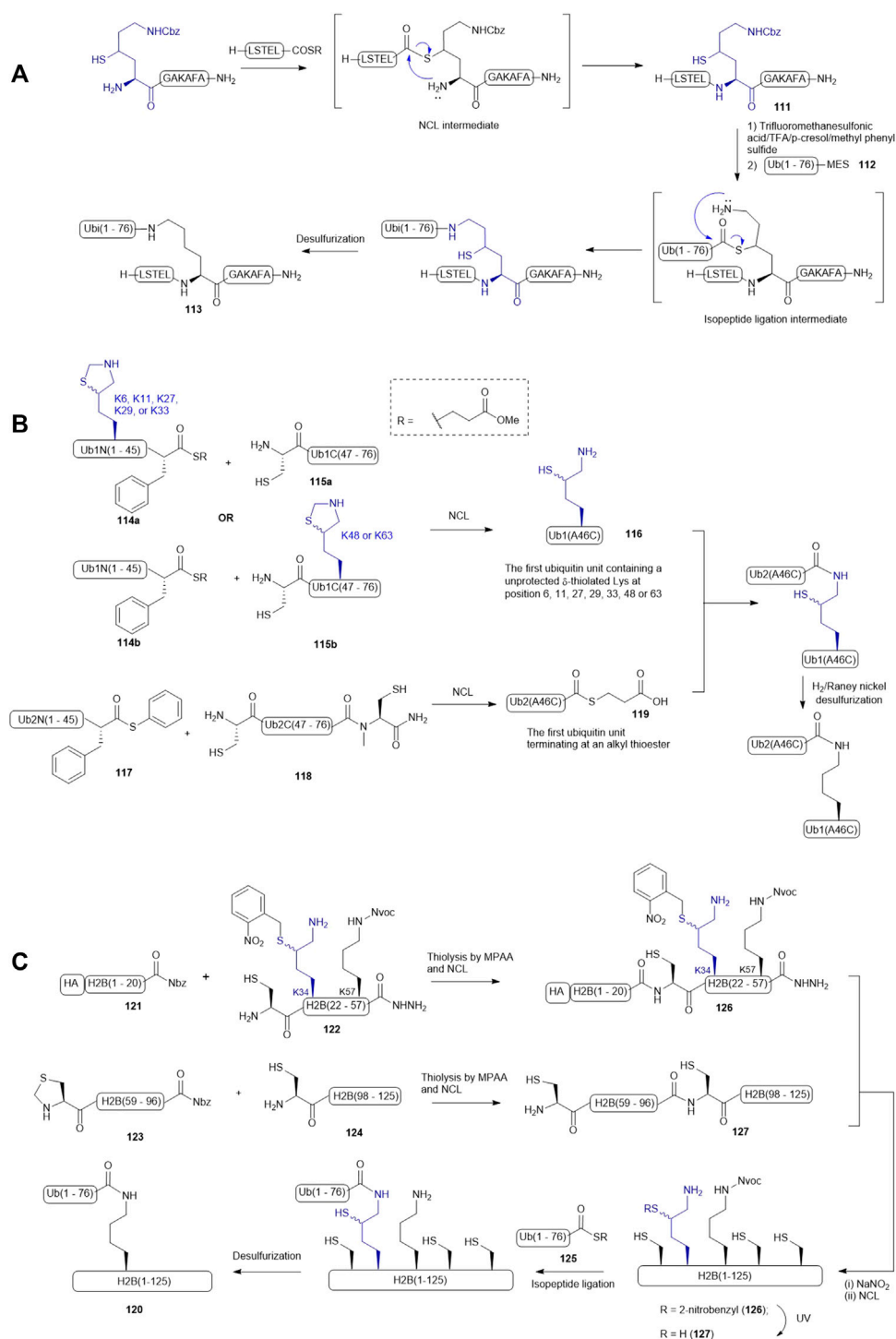
Chromatin packaging and function are tightly regulated by histone ubiquitination modifications, however, the biological
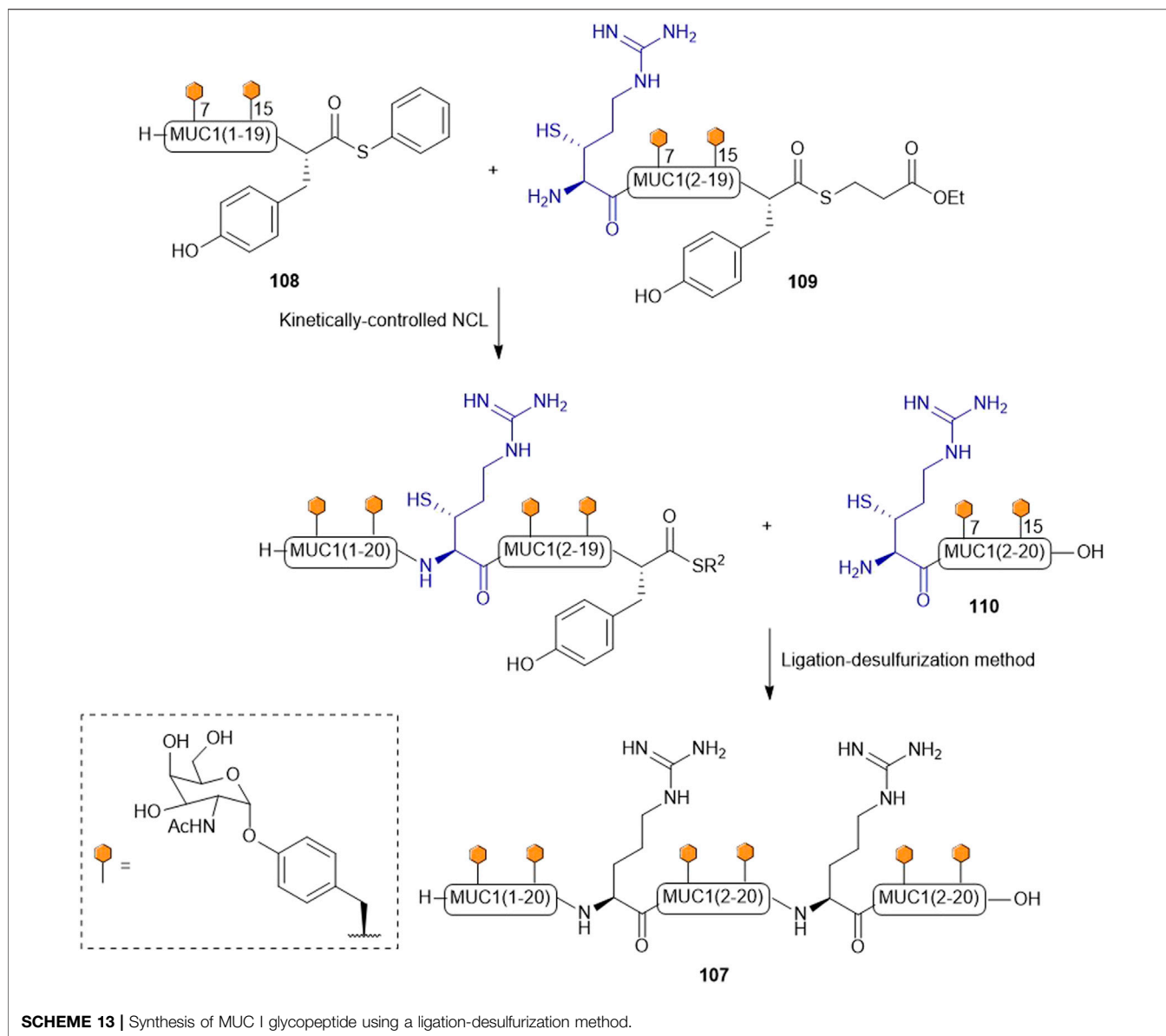
SCHEME 13 | Synthesis of MUC I glycopeptide using a ligation-desulfurization method.

roles of site-specific ubiquitin modifications of H2B remained unclear, prompting Siman *et al.* to chemically synthesize H2B homologues (**120**) with defined ubiquitination patterns (Siman et al., 2013). Since H2B lacks a native Cys residue, they adopted a ligation–desulfurization strategy wherein Ala21, Ala58, and Ala97 were mutated to Cys residues for NCL purposes, meaning the H2B protein would be divided into four segments, including H2B(1–20) **121** with an HA tag, H2B(21–57) **122**, H2B(58–96) **123**, and H2B(97–125) **124**. Initially, the synthesis started with ligating a ubiquitin thioester (**125**) to the side-chain of the δ-thiolated Lys at position 34 while the N-terminal Cys residue of H2B (21–57) was masked by an acetamidomethyl (Acm) group. The Acm group was then removed to allow for ligation with the H2B(1–20) thioester **121**. However, the resultant ligation product H2B(1–57) did not survive in the conditions employed to

activate acyl hydrazide. The synthetic strategy was therefore altered and focused on completing the synthesis of the full-length H2B protein before isopeptide ligation with the ubiquitin thioester **125**. Hence, the H2B(21–57) was re-synthesized by introducing a 2-nitrobenzyl (NB) group to mask the δ-thiol derived Lys34 (peptide **122** in **Scheme 14C**). The ligation of peptides **121** with **122** proceeded to completion in a good isolated yield; the following oxidation and thioesterification of the ligation product also proceeded in a straightforward manner to form **126**. It was subjected to NCL conditions containing the ligation product (**127**) of **123** and **124** to provide the full-length H2B (**128**) bearing a photo-caged thiolated Lys at position 34. After removing the NB group under UV irradiation, the unmasked thiolated H2B **129** was ligated to the ubiquitin thioester **125** and then subjected to desulfurization conditions to afford the desired Lys34-ubiquitinated H2B protein **120**. The

SCHEME 14 | (A) Liu's dual ligation strategy to access site-specific ubiquitination; **(B)** Brik's synthesis of the di-ubiquitin library **(C)** Siman's route to access Lys34-ubiquitinated H2B.

authors noticed a significant amount of Met oxidation products were formed in the photo-deprotection and isopeptide ligation conditions and suggested norleucine would overcome this issue in future studies.

β-Selenol Derived Phenylalanine

Recently, Payne and coworkers reported the preparation of four discrete glycosylated forms of human IFN-γ using a tandem DSL–deselenization strategy (Wang et al., 2020a). Specifically, the

human IFN-γ sequence was disconnected at Ser51–Phe52 and Lys108–Ala109 junctions, which generated three classes of differentially functionalized fragments to synthesize (Dawson et al., 1994): a C-terminal fragment, IFN-γ(109–138) **130**, where the N-terminal Ala was replaced with a Sec residue (Bang et al., 2006), a highly functionalized middle fragment [IFN-γ(52–108), **131**] having a PMB-protected β-selenol Phe at the N-terminus and a phenylselenoester at the C-terminus, with/without a β-GlcNAc moiety on the side chain of Asn97, and (Thompson et al., 2014) the N-terminal fragment **132** of IFN-γ(1–51) possessing an N-terminal pyroglutamate and a C-terminal phenylselenoester with either unmodified or β-GlcNAc-derivatized Asn25. The first DSL reaction was performed by simply mixing the diselenide dimer **130** of IFN-γ (109–138) bearing an N-terminal selenocystine (1.0 equiv.) with the middle fragment of IFN-γ(52–108) (**131**, 1.1 equiv.) in ligation buffer under additive-free conditions. After completion within 5 min, the desired product was composed of a mixture of the branched selenoester product and asymmetric diselenide products (termed peptide **133**). The resulting ligation mixture was subjected to deselenization conditions immediately, which involved removing diphenyl diselenide through hexane extraction followed by the addition of the deselenization buffer containing TCEP (50 equiv.) and DTT (50 equiv.). After 16 h, purification by HPLC afforded the PMB-protected IFN-γ (52–138) **134** in good isolated yields (62 and 57%, respectively). The β-selenol phenylalanine residues of **134** were subsequently unmasked under oxidative deprotection conditions comprising DMSO, ligation buffer (6 M Gdn HCl and 0.1 M Na$_2$HPO$_4$) and TFA in a ratio of 1:1:3 (v/v/v) to generate the desired peptide diselenide dimer **135** (81–90%). The deprotection conditions also caused concomitant oxidation of the Met residues. The final DSL reaction to assemble hIFN-γ (52–138) and (1–51) fragments was performed by dissolving **135** and hIFN-γ(1–51) selenoester **132** with/without β-GlcNAc-modified N25 residue in additive-free ligation buffer at a higher dilution (1–2 mM on the basis of the monomeric form of **132**). The ligation reaction proceeded smoothly within 30 min and was subjected to deselenization conditions to afford the four homogeneously glycoforms of hIFN-γ(1–138) in excellent yields (54–59% over two steps). Finally, the oxidized Met residues were reduced under Hackenberger's conditions to afford four glycosylated variants of IFN-γ (termed glycoprotein **136** in **Scheme 15A**) (Hackenberger, 2006).
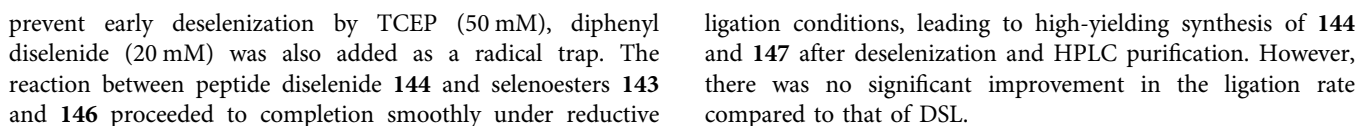
β-Selenol Derived Leucine

To expand the scope of one-pot DSL–deselenization methodologies, the utility of β-selenyl Leu was highlighted in the synthesis of a chemokine-binding protein from human cytomegalovirus (UL22A) (Wang et al., 2017) by Wang *et al.* UL22A has been shown to neutralize RANTES effectively *in vitro* and was demonstrated by the authors that the Tyr residues 65 and 69 were post-translationally sulfated, leading to further enhancement in chemokine binding. Taking advantage of the distinct reaction kinetics of DSL and NCL, a one-pot three-fragment ligation starting with DSL (using β-Se-Leu) followed by NCL (mediated by β-thiol-derived Asp) was planned (see
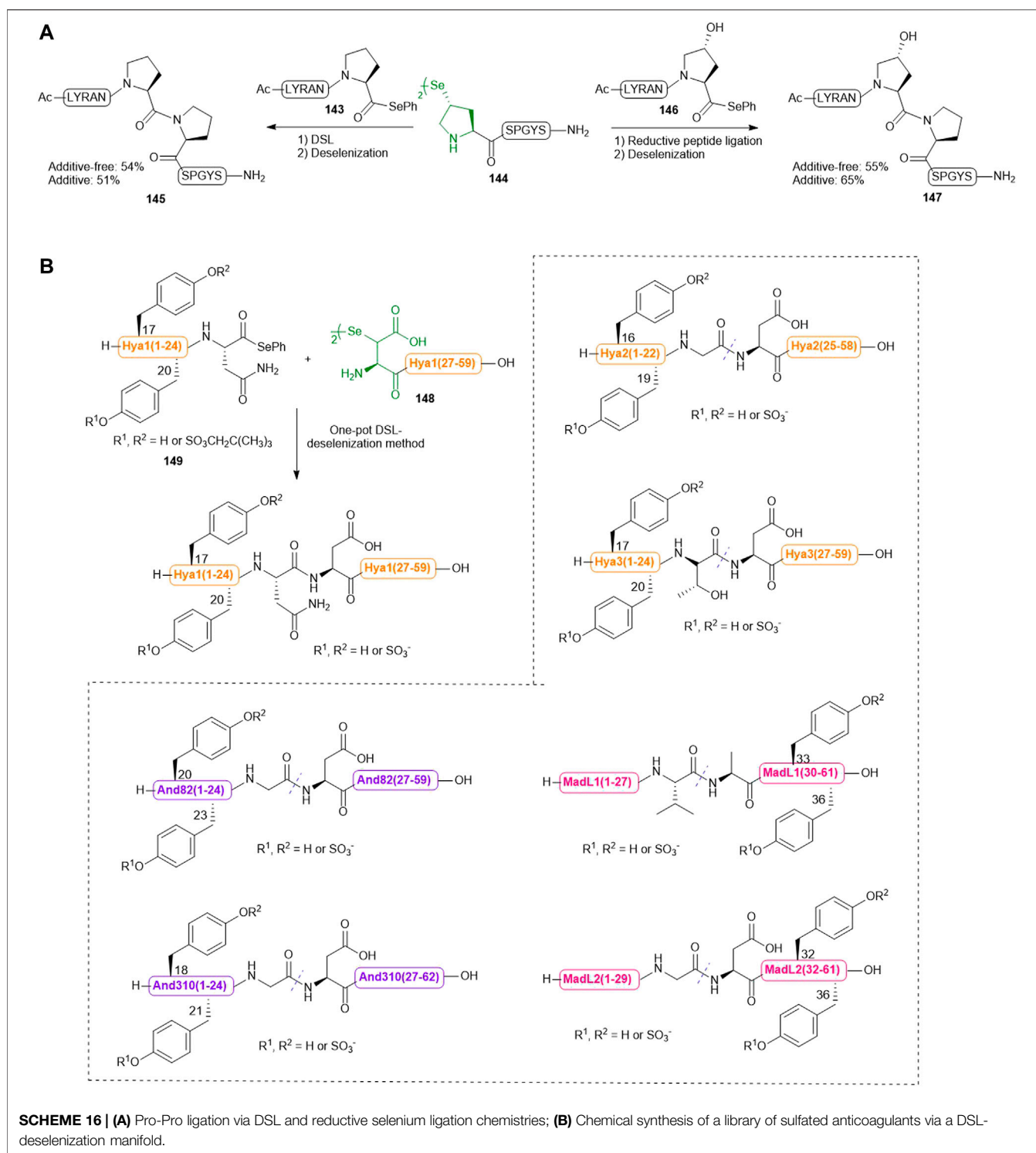
Scheme 15B) (Wang et al., 2020b). As such, the full-length protein **137** was divided into the three fragments, including an N-terminal segment UL22A 21–51 (**138**) bearing a phenyl selenoester at its C-terminus, a bifunctional middle fragment (UL22A 52–76, **139**) having a C-terminal alkyl thioester and an N-terminal β-selenyl Leu as diselenide dimer, and the last fragment UL22A (77-103) **140** possessing an N-terminal β-thiolated Asp. The phenyl selenoester **138** was first conjugated to the middle fragment dimer **139** following a standard DSL protocol, which proceeded to completion within 30 min. The fast ligation kinetics of DSL in the absence of a thiol additive meant that no intramolecular cyclization of **139** was possible. The C-terminal fragment UL22A (77–103) **140** was subsequently added together with TCEP (25 mM) and TFET (2 vol%) to effect the final ligation with concomitant deselenization. This was followed by radical desulfurization *in situ* to convert β-SH-Asp into native Asp, providing UL22A (21–103) **137** in 40% yield over 4 steps after HPLC purification.

In order to improve synthetic efficiency, a two-fragment ligation strategy was also employed to synthesize UL22A, where the protein was disconnected at the Val71-Leu72 junction to provide fragments **141** and **142** (**Scheme 15C**) (Wang et al., 2017). Interestingly, the DSL reaction did not proceed significantly, attributed to the steric congestion at this Val-Leu junction. The addition of diphenyl diselenide and TCEP enabled reduction of the diselenide dimer and unveiling the nucleophilicity of the selenolate allowed efficient convergent assembly within 1 h. The need for additives to facilitate ligation at the Val71-Leu72 junction but not in the DSL reaction at the Thr51-Leu52 junction highlighted the fact that DSL reactions are dependent on the specific nature of the fragments, not only on the steric hindrance around the selenoester center.

γ-Selenol Derived Proline

Recently, Sayers *et al.* harnessed the complementary reactivities of *trans*-γ-selenyl Pro and prolyl selenoesters to facilitate DSL reactions at the notorious proline–proline ligation junction for the first time (**Scheme 16A**) (Sayers et al., 2018a). The authors first demonstrated the capacity of γ-selenyl Pro in mediating DSL chemistry (within 5–45 min) using peptide selenoesters terminated at different C-terminal residues. After this initial assessment, the authors investigated whether this selenylated Pro warhead could effect DSL at intractable Pro–Pro junctions. Excitingly, the ligation reaction between a model peptide Ac-LYRANP-SePh (**143**) (2.0 equiv.) and the peptide diselenide **144** (1.0 equiv. with respect to the monomer) proceeded to completion to form **145** in 16 h in 6 M GnHCl/ 0.1 M phosphate buffer at pH 6.2. Upon completion, the ligation mixture was extracted with hexane to remove the diphenyl diselenide precipitate followed by *in situ* deselenization with TCEP and DTT to provide the desired peptide in a good yield. The authors also demonstrated the utility of this ligation chemistry in chemical synthesis of peptides bearing hydroxyPro-Pro junctions. They additionally examined whether the inclusion of TCEP could accelerate the Pro-Pro ligation reaction. To

SCHEME 15 | (A) Payne's route to access differentially glycosylated human IFN-γ; **(B)** Chemical synthesis of post-translationally sulfated UL22A via β-selenylated Leu-mediated DSL and **(C)** reductive diselenide-selenoester ligation.

prevent early deselenization by TCEP (50 mM), diphenyl diselenide (20 mM) was also added as a radical trap. The reaction between peptide diselenide **144** and selenoesters **143** and **146** proceeded to completion smoothly under reductive ligation conditions, leading to high-yielding synthesis of **144** and **147** after deselenization and HPLC purification. However, there was no significant improvement in the ligation rate compared to that of DSL.

SCHEME 16 | (A) Pro-Pro ligation via DSL and reductive selenium ligation chemistries; **(B)** Chemical synthesis of a library of sulfated anticoagulants via a DSL-deselenization manifold.

The enabling nature of such Pro-Pro ligation-deselenization methods was also demonstrated by the high-yielding chemical synthesis of the submaxillary gland androgen regulated protein 3B and lumbricin-1, which represented the first two successful examples of chemical protein synthesis through joining the sterically and electronically demanding Pro-Pro junctions.

β-Selenol Derived Aspartic Acid

Previously, Payne and coworkers demonstrated that hematophagous ticks and mosquitos produce a suite of Tyr-sulfated protein anticoagulants which hijack the host's central clotting machineries to facilitate blood ingestion (Thompson et al., 2017; Watson et al., 2018). Recently, they utilized a

bioinformatic approach to identify multiple putative thrombin inhibitors [hyalomin (Hya) 1, 2 and 3, andersonin (And) 82 and 310, and madanin-like protein (MDL) 1 and 2] and their two conserved Tyr sulfation sites. To confirm the biological activity of each sulfated Tyr residue and investigate the functional interplay between the PTM and the surrounding amino acids on anticoagulant mechanism, the sulfated variants of these anticoagulants were synthesized via a one-pot DSL−deselenization manifold (**Scheme 16B**) (Watson et al., 2019).

The requisite peptide selenoesters and their peptide diselenide companions were prepared by Fmoc-based SPPS method. In particular, the sulfated Tyr residues were masked by neopentyl groups and incorporated into the selenoester segments of Hya1, Hya2, Hya3, And82 and And310 proteins, or the peptide diselenide segments of MadL1 and MadL2. β-selenyl Asp amino acid was introduced to the N-terminus of each peptide diselenide segment (i.e., the Hya1 (27–59) fragment **148** shown in **Scheme 16B**) except that of MadL1, wherein selenocystine was incorporated instead. All proposed anticoagulants were assembled via DSL chemistry, whereby each peptide selenoester (i.e., the Hya1 (1–24) fragment **149** in Scheme 16B) and the ligation partner were first prepared as a 10 mM solution (with respect to the monomer) in ligation buffer (6 M Gn. HCl, 0.1 M Na_2HPO_4, pH 6.0–6.5) separately, followed by mixing of the two solutions in an equal volume to effect the ligation reaction. 10 vol% N,N-dimethylformamide was also added to the DSL mixture of Hyd and fragments in order to facilitate peptide dissolution. All reactions proceeded to completion within 20 min, which were subsequently subjected to deselenization conditions containing the final concentrations of 250 mM TCEP and 25 mM DTT. In contrast to the long reaction time (16 h) required for converting selenocystine to Ala residues, deselenization at the β-selenyl Asp residue was complete within 10 min owing to resonance stabilization of the β-carbon−centered radical. After deselenization, the crude mixtures were further incubated for 8–16 h to allow complete removal of neopentyl protecting groups before purification by reversed-phase HPLC, which afforded the target library of homogeneous sulfated variants of Hya, And and MDL in excellent yields.

Strategic Considerations in Chemical Protein Synthesis

NCL was developed to overcome the size limitation of synthetic polypeptides imposed by solid-phase peptide synthesis (SPPS), enabling one to convergently build larger polypeptides and proteins from smaller peptide fragments. However, the efficiency of the SPPS processes undertaken to prepare the requisite fragments is still a critical factor to consider, which may shape the retrosynthetic design. Typically, SPPS can afford polypeptides having 40–50 residues in length, but preparing shorter peptide segments may be required for difficult sequences and for those which do not involve turn-inducing elements (i.e., Pro and pseudoproline residues). In general, dividing the protein target into 30–40 residue fragments

provides a good chance to obtain the requisite ligation segments in reasonable yields and adequate purity (>95%). Another crucial factor to consider is the chemical reactivity of the amino acids within the ligation junctions and their potential influence on the processes of ensuing chemistries. Conventionally, retrosynthetic disconnections were decided based upon the location of native Cys residues within a target protein, but the development of thiol- and selenol-derived amino acids has massively expanded our choices. Careful consideration of the chemical orthogonality of these modified amino acids may allow one to develop streamlined synthesis of the target protein harnessing a one-pot, multicomponent ligation method or *in situ* desulfurization/deselenization method with a view to improve the overall synthetic efficiency and minimize laborious work of intermediary purification. For example, Thompson *et al.* reported a chemoselective desulfurization method to remove the β-thiol auxiliary at Asp residues in the presence of free Cys residues using TCEP and dithiothreitol (DTT) at pH 3, which makes β-mercapto Asp attractive in modern total synthesis of proteins possessing native Cys residues (Thompson et al., 2013). The influence of the thiolated/selenylated chiral centers on ligation efficiency should also be considered. For instance, Danishefsky and co-workers demonstrated that the ligation rate of the *erythro* isomer of β-thiol Leu is prohibitively slow while the *threo* form allows ligation with Gly, Ala, Phe and Val thioesters to complete within 8 h (Tan et al., 2010). Likewise, it has also been shown that the *trans* isomer of γ-thiol Pro can mediate fast NCL reactions comparable to other thiolated amino acids, whereas the *syn* isomer prevents S to N acyl shift during the process of NCL (Ding et al., 2011; Townsend et al., 2012).

The rate of NCL is also strongly influenced by the steric and electronic factors of the C-terminal amino acid residue that acts as an acyl donor. For example, thioester-derived β-branched amino acids (i.e., Val, Ile and Thr) react sluggishly with the ligation counterpart leading to the accumulation of by-products arising from the hydrolysis and other side reaction pathways (Hackeng et al., 1999; Siman et al., 2012). Prolyl thioesters were found to be poor acyl donors due to the n→π* electronic interaction between the carbonyl group on the Pro nitrogen to the thioester carbonyl carbon resulting in a reduction in electrophilicity (Hackeng et al., 1999; Pollock and Kent, 2011). Durek and Alewood provided a simple solution to these issues by introducing alkyl/aryl selenoesters to the C-terminal residues, which resulted in orders of magnitude enhancement of ligation rates, as exemplified by the model ligation reaction with a prolyl selenoester proceeding to completion within 2 h (Durek and Alewood, 2011). Finally, peptide thioesters terminating at a C-terminal Asp, Glu, Asn, Gln or Lys residue, are prone to cyclize *via* nucleophilic attack at the acyl center by the C-terminal side-chain functionality. The cyclization reaction at Lys, Asn and Gln residues may be suppressed by performing the ligation at pH 6–7; however, it is challenging to prevent the cyclization at Asp and Glu residues without masking their carboxylate side chains (Barnes et al., 2021).

It is also important to note the disparate reactivity profiles of thiolated amino acids in desulfurization chemistries, which may

allow one to design a more strategic approach to remove the thiol auxiliaries. Penicillamine as a precursor of Val was commonly used in developing ligation–desulfurization strategies due to its commercial availability (Chen et al., 2008; Haase et al., 2008; Reimann et al., 2015). Interestingly, Haase *et al.* reported that desulfurization yields were moderate to low when using metal-based methods whilst Danishefsky's radical approach could afford quantitative conversion to the Val residue (Haase et al., 2008). β-mercapto Asp and γ-mercapto Glu were commonly removed via the desulfurization chemistry initiated by VA-044 (Thompson et al., 2014); however, certain peptide sequences can promote a side reaction between the VA-044 radical and the sulfanyl radical generated via hydrogen abstraction from the sulfhydryl auxiliary (Cergol et al., 2014; Li et al., 2021b). An alternative radical initiator, 4,4′-azobis (4-cyanovaleric acid) reported by Li *et al.* was shown to effectively suppress this deleterious side reaction and facilitate conversion to the desired product (Li et al., 2021b). Brik and coworkers reported that ligation products containing γ-mercapto Gln were decomposed to unidentified compounds in metal-free desulfurization conditions due to the unstable nature of the γ-thiolated side chain (Siman et al., 2012), whereas clean conversion to the final product could be achieved in the nickel boride conditions. Likewise, Malins *et al.* reported that the sulfanyl moiety in 2-thiol Trp could not be eliminated in radical-based conditions plausibly due to the strength of the C–S bond and the contributions of the other tautomeric form, but a metal-based reductive cleavage protocol with hydrogen gas and Pd on Al_2O_3 was applied to facilitate the formation of the desulfurized product without detectable demethylthiolation on Met residues (Malins et al., 2014).
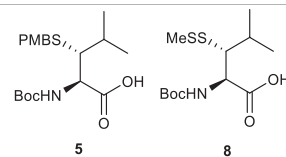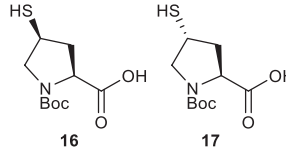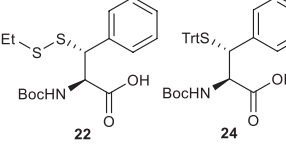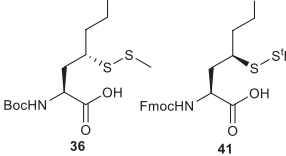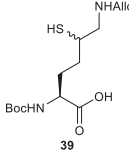
Despite being the chalcogenic relative of thiol, selenols manifest markedly different chemical reactivity to thiols. It is well documented that selenol auxiliaries possess a higher acidity and greater nucleophilicity than thiols and can be readily removed in the presence of other proteogenic functionalities (using TCEP and DTT at pH 6–7) (Gieselman et al., 2001; Metanis et al., 2010). The ease and practic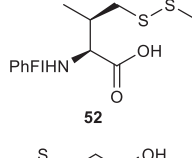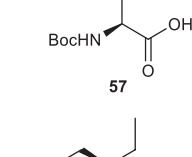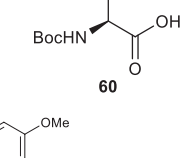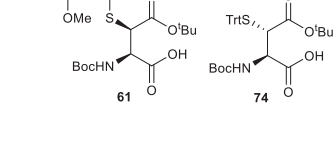ality of deselenization chemistry have attracted enormous interest in expanding the scope of selenol-derived amino acids in modern total synthesis of proteins. Due to the thermodynamically favourable reaction between TCEP and selenol leading to homolytic cleavage of the C–Se bond, aryl thiols were often used as alternative reductants in selenol-assisted NCL. However, the weak reducing power of aryl thiols only affords a low level of productive selenol during ligation; hence, the rate of ligation does not benefit from the greater reactivity of selenol. Metanis and coworkers demonstrated that the deleterious reaction between Sec and TCEP can be suppressed by the addition of ascorbate as a radical scavenger, and the resulting additive combination allows one to fully exploit the latent reactivity of selenol (Reddy et al., 2016) for peptide ligation. It is noteworthy that ascorbate is still incapable of preventing the homolytic cleavage of selenol at more activated positions (i.e. deselenization at β-seleno Asp and γ-seleno Glu residues) (Mitchell et al., 2017). Also harnessing the complementary reactivity between phosphine and selenol, Malins (Malins

et al., 2015b) and Dery et al. (Dery et al., 2015) have independently developed different oxidative deselenization protocols that effectively transform Sec to Ser at the ligation junction after ligation, which allow for retrosynthetic disconnection at X-Ser junctions. Recently, Metanis and coworkers have attempted to utilize the unique reactivity of γ-selenoLys to establish an isopeptide ligation–deselenization method for chemical synthesis of proteins modified by SUMO or ubiquitin (Dardashti et al., 2020). However, the resulting isopeptide ligation proceeded very slowly despite adding excess amounts of TCEP and sodium ascorbate. The authors proposed that the higher pKa of the ε-amine in Lys prevents transesterification to occur in neutral ligation buffer; however, the underlying mechanism remains to be investigated.

Payne and coworkers reported an additive-free ligation method (dubbed DSL) (Mitchell et al., 2015) that could potentially overcome the bottlenecks in selenium-mediated peptide ligation: first, the approach enables the reaction between a peptide diselenide and a peptide bearing C-terminal phenyl selenoester to complete within minutes without any reductive additives, which compares favorably to the rate of a respective NCL reaction. Second, this chemistry exhibits high tolerance to acidic pH and enables the ligation reaction to proceed at pH 3–7, which could effectively suppress hydrolysis of selenoesters. Third, the selenol auxiliary can be selectively removed *in situ* (using excess amounts of TCEP and DTT) after DSL without the need for intermediary purification. With these salient features and the demonstrated orthogonality between selenol and other proteogenic functionalities, it is anticipated that this methodology will find wide application in the research community of chemical protein synthesis. Recently, the harmony between NCL and DSL has been proven in the total synthesis of phosphorylated insulin-like growth factor binding protein 2 (Premdjee et al., 2021).
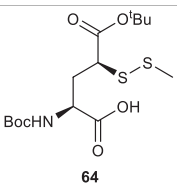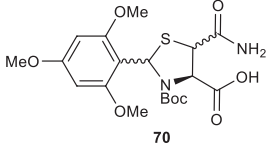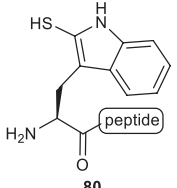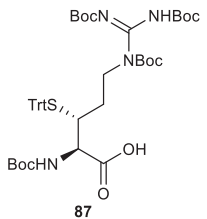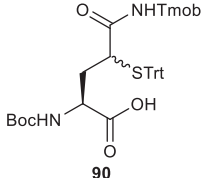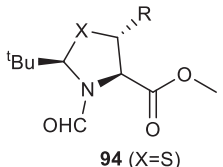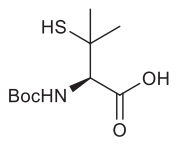
In the original forms, both NCL and DSL reactions require millimolar concentrations of ligation fragments to achieve productive kinetics and to afford high-yielding synthesis, however proceeding sluggishly at micromolar concentrations. Harnessing the superior reactivities of phenyl selenoester and Sec in reducing environment, Payne and co-workers reported a reductive diselenide-selenoester ligation (rDSL) method that enables efficient peptide ligation down to 50 nM (Chisholm et al., 2020). In addition, the authors demonstrated that an efficient photodeselenization process can be performed after ligation to rapidly afford the target polypeptides in a one-pot manner. The power of the rDSL-photodeselenization manifold has been showcased in the chemical synthesis of a lipidated peptide therapeutic (tesamorelin) and two palmitylated variants of an integral membrane protein (FXYD1) without the aid of solubility tags and hybridizing templates. With the global endeavor to develop more efficient strategies to access selenylated amino acids, it is expected that the scope of the DSL and rDSL methodologies will be continuously expanded and may allow one to exploit the scope beyond total synthesis of proteins (Sayers et al., 2018b).

TABLE 1 | Thiol derived amino acids and the relevant synthesis examples described in this article.

Thiolated amino acid	Synthetic strategy	Peptide and protein synthesis example	References(s)
5 **8**	1. Aziridine ring opening 2. Nucleophilic displacement	ATAD2 bromodomain region	Harpaz et al. (2010) Tan et al. (2010) Creech et al. (2014)
16 **17**	Nucleophilic displacement	1. Rat neuromedin U 2. hEPO (79–166) Glycopeptide	Ding et al. (2011) Townsend et al. (2012)
22 **24**	1. Nucleophilic displacement 2. Garner's aldehyde	1. Fragment of Augurin 2. Glycosylated human IFN-γ	Malins et al. (2015a) Wang et al. (2020a)
36 **41**	Nucleophilic displacement	1. Ubiquitinated peptide 2. Diubiquitin and tetraubiquitin	Pasunooti et al. (2009) Merkx et al. (2013)
39	Nucleophilic conjugate addition	1. Diubiquitin 2. Site-specific ubiquitinated H2B	Ajish Kumar et al. (2009) Siman et al. (2013)
52	Nucleophilic displacement	Model peptides	Chen et al. (2008)
57	Nucleophilic displacement	Model peptides	Chen et al. (2010)
60	C (sp³)-H activation	*Xenopus* H3	Pasunooti et al. (2016)
61 **74**	1. Electrophilic sulfenylation 2. Nucleophilic displacement	1. Extracellular N-terminal domain of CXCR4 2. Sulfated anopheline proteins 3. Glycosylated IL-17A	Watson et al. (2018) Thompson et al. (2013) Guan et al. (2013) Li et al. (2021b)

(Continued on following page)
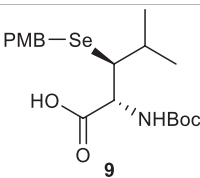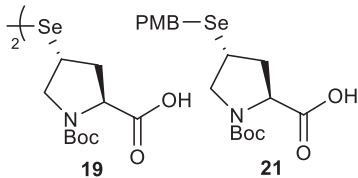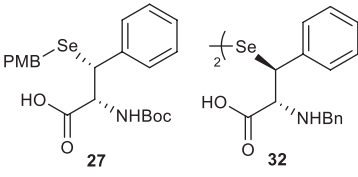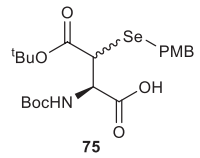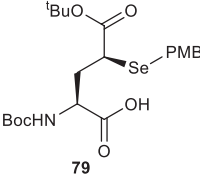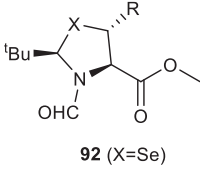
Thiolated amino acid	Synthetic strategy	Peptide and protein synthesis example	References(s)
64	Electrophilic sulfenylation	Teriparatide	Cergol et al. (2014)
70	Electrophilic sulfenylation	Enfuvirtide	Sayers et al. (2015)
80	Electrophilic sulfenylation	Extracellular N-terminal domain of CXCR1	Malins et al. (2014)
87	Garner's aldehyde	Glycosylated extracellular domain of MUC1	Malins et al. (2013)
90	Passerini three-component reaction	Model peptides	Siman et al. (2012)
94 (X=S)	Photoredox-catalyzed asymmetric Giese reaction	Model peptides	Yin et al. (2020)
	Commercially available	Tri-phosphorylated C-terminal Tau	Reimann et al. (2015) Haase et al. (2008)

CONCLUSION

Recent advances of peptide ligation chemistry have greatly enhanced our capacity to produce native proteins and even peptide therapeutics with defined, bespoke modifications. In particular, these new methods have expanded the scope of NCL beyond the need of adequately positioned Cys residues within a protein target, including the successes in preparing the thiol derived variants of Phe, Val, Lys, Thr, Ile, Leu, Pro, Arg, Asn, Gln, Asp, Glu, and Trp (**Table 1**), and the selenyl variants of Phe, Lys, Pro, Asp, and Glu (**Table 2**), and the recent approach to access a large variety of β-thiolated/selenylated amino acids based

TABLE 2 | Selenol derived amino acids and the relevant synthesis examples described in this article.

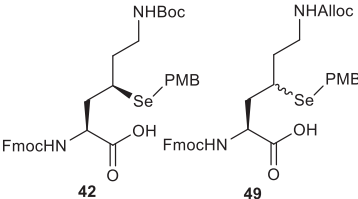| Selenolated amino acid	Synthetic strategy	Peptide and protein synthesis example	References(s)
9	Garner's aldehyde	Sulfated UL22A	Wang et al. (2017)
19 **21**	Nucleophilic displacement	1. Hydroxy-proline model peptide 2. SMR3B 3. Lumbricin-1	Townsend et al. (2012) Sayers et al. (2018a)
27 **32**	1. Nucleophilic displacement 2. Garner's aldehyde	Model peptides	Malins and Payne, (2012)
42 **49**	1. Nucleophilic displacement 2. Nucleophilic conjugate addition	SUMOylated GCK	Dardashti et al. (2020)
75	Electrophilic sulfenylation	1.Sulfated Hya1-3 2. Selenoprotein K 3. And82, And310 and MDL2 4. non-modified Hya2-4	Watson et al. (2019) Mitchell et al. (2017)
79	Electrophilic sulfenylation	Model peptides	Mitchell et al. (2017)
92 (X=Se)	Photoredox-catalyzed asymmetric Giese reaction	Cytpchrome c Oxidase subunit protein 7C Oxytocin analogues	Yin et al. (2020)

on photoredox-catalyzed asymmetric Giese reactions (**Table 1** and **2**). Integrating these important artificial amino acids into advanced ligation–desulfurization or ligation–deselenization methodologies has provided a viable alternative to access homogenously modified proteins in highly efficient manner and has advanced our knowledge on the roles of PTMs in protein function, activity, and structure. While such synthetic approaches cannot compete with the ultra-large protein libraries generated through phage display (Salmond and Fineran, 2015; Alfaleh et al., 2020), mRNA display (Murakami et al., 2006; Goto et al., 2011) and codon expansion and reprogramming technologies (Chin, 2014;

Wals and Ovaa, 2014), the latest ligation techniques offer an exciting opportunity to establish novel protein medicinal chemistry programs through constructing focused libraries, wherein distinct PTMs and bespoke modifications can be incorporated site-specifically allowing one to explore a unique and important chemical space in protein therapeutics. Furthermore, the fast kinetics and chemical specificity of the recently established selenium-based ligation technologies provide a potential avenue to engineer proteins at low concentrations in non-denaturing buffer (i.e., physiological buffer), which may offer an exciting possibility for other chemical biology applications, including cell-surface protein engineering, rapid antigen testing and antibody-drug conjugation.

AUTHOR CONTRIBUTIONS

IG and KW have contributed equally to the review conception, writing and figure design. IG was also involved in final figure editing and formatting. JL was actively involved in manuscript and reference editing and final revision. XL participated in conceptualization and writing, and actively revised this work.

FUNDING

REFERENCES

Agten, S. M., Watson, E. E., Ripoll-Rozada, J., Dowman, L. J., Wu, M. C. L., Alwis, I., et al. (2021). Potent Trivalent Inhibitors of Thrombin through Hybridization of Salivary Sulfopeptides from Hematophagous Arthropods. *Angew. Chem. Int. Ed.* 60 (10), 5348–5356. doi:10.1002/anie.202015127

Ajish Kumar, K. S., Haj-Yahya, M., Olschewski, D., Lashuel, H. A., and Brik, A. (2009). Highly Efficient and Chemoselective Peptide Ubiquitylation. *Angew. Chem. Int. Ed.* 48 (43), 8090–8094. doi:10.1002/anie.200902936

Alfaleh, M. A., Alsaab, H. O., Mahmoud, A. B., Alkayyal, A. A., Jones, M. L., Mahler, S. M., et al. (2020). Phage Display Derived Monoclonal Antibodies: From Bench to Bedside. *Front. Immunol.* 11, 1986. doi:10.3389/fimmu.2020.01986

Arnold, F. H. (2018). Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed.* 57 (16), 4143–4148. doi:10.1002/anie.201708408

Bang, D., Pentelute, B. L., and Kent, S. B. H. (2006). Kinetically Controlled Ligation for the Convergent Chemical Synthesis of Proteins. *Angew. Chem. Int. Ed.* 45 (24), 3985–3988. doi:10.1002/anie.200600702

Barnes, N. G., Nyandoro, K., Jin, H., and Macmillan, D. (2021). Rapid Access to Asp/Glu Sidechain Hydrazides as Thioester Precursors for Peptide Cyclization and Glycosylation. *Chem. Commun.* 57 (8), 1006–1009. doi:10.1039/d0cc07404g

Buettner, C. S., Willcox, D., Chappell, B. G. N., and Gaunt, M. J. (2019). Mechanistic Investigation into the C(sp3)-H Acetoxylation of Morpholinones. *Chem. Sci.* 10 (1), 83–89. doi:10.1039/c8sc03434f

Cergol, K. M., Thompson, R. E., Malins, L. R., Turner, P., and Payne, R. J. (2014). One-Pot Peptide Ligation-Desulfurization at Glutamate. *Org. Lett.* 16 (1), 290–293. doi:10.1021/ol403288n

Chen, J., Wan, Q., Yuan, Y., Zhu, J., and Danishefsky, S. J. (2008). Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angew. Chem. Int. Ed.* 47 (44), 8521–8524. doi:10.1002/anie.200803523

Chen, J., Wang, P., Zhu, J., Wan, Q., and Danishefsky, S. J. (2010). A Program for Ligation at Threonine Sites: Application to the Controlled Total Synthesis of Glycopeptides. *Tetrahedron* 66 (13), 2277–2283. doi:10.1016/j.tet.2010.01.067

Chin, J. W. (2014). Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* 83 (1), 379–408. doi:10.1146/annurev-biochem-060713-035737

Chisholm, T. S., Kulkarni, S. S., Hossain, K. R., Cornelius, F., Clarke, R. J., and Payne, R. J. (2020). Peptide Ligation at High Dilution via Reductive Diselenide-Selenoester Ligation. *J. Am. Chem. Soc.* 142 (2), 1090–1100. doi:10.1021/jacs.9b12558

Conibear, A. C., Watson, E. E., Payne, R. J., and Becker, C. F. W. (2018). Native Chemical Ligation in Protein Synthesis and Semi-synthesis. *Chem. Soc. Rev.* 47 (24), 9046–9068. doi:10.1039/c8cs00573g

Creech, G. S., Paresi, C., Li, Y.-M., and Danishefsky, S. J. (2014). Chemical Synthesis of the ATAD2 Bromodomain. *Proc. Natl. Acad. Sci.* 111 (8), 2891–2896. doi:10.1073/pnas.1400556111

Crich, D., and Banerjee, A. (2007). Native Chemical Ligation at Phenylalanine. *J. Am. Chem. Soc.* 129 (33), 10064–10065. doi:10.1021/ja072804l

Dardashti, R. N., Kumar, S., Sternisha, S. M., Reddy, P. S., Miller, B. G., and Metanis, N. (2020). Selenolysine: A New Tool for Traceless Isopeptide Bond Formation. *Chem. Eur. J.* 26 (22), 4952–4957. doi:10.1002/chem.202000310

Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. H. (1994). Synthesis of Proteins by Native Chemical Ligation. *Science* 266 (5186), 776–779. doi:10.1126/science.7973629

Dery, S., Reddy, P. S., Dery, L., Mousa, R., Dardashti, R. N., and Metanis, N. (2015). Insights into the Deselenization of Selenocysteine into Alanine and Serine. *Chem. Sci.* 6 (11), 6207–6212. doi:10.1039/c5sc02528a

Ding, H., Shigenaga, A., Sato, K., Morishita, K., and Otaka, A. (2011). Dual Kinetically Controlled Native Chemical Ligation Using a Combination of Sulfanylproline and Sulfanylethylanilide Peptide. *Org. Lett.* 13 (20), 5588–5591. doi:10.1021/ol202316v

Durek, T., and Alewood, P. F. (2011). Preformed Selenoesters Enable Rapid Native Chemical Ligation at Intractable Sites. *Angew. Chem. Int. Ed.* 50 (50), 12042–12045. doi:10.1002/anie.201105512

Easton, C. J., Hutton, C. A., Roselt, P. D., and Tiekink, E. R. T. (1994). Stereocontrolled Synthesis of β-hydroxyphenylalanine and β-hydroxytyrosine Derivatives. *Tetrahedron* 50 (24), 7327–7340. doi:10.1016/s0040-4020(01)85256-x

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 (33), 7645–7649. doi:10.1002/anie.201100996

Flavell, R. R., and Muir, T. W. (2009). Expressed Protein Ligation (EPL) in the Study of Signal Transduction, Ion Conduction, and Chromatin Biology. *Acc. Chem. Res.* 42 (1), 107–116. doi:10.1021/ar800129c

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 (36), 11634–11639. doi:10.1002/anie.201805191

Garner, P. (1984). Stereocontrolled Addition to a Penaldic Acid Equivalent: an Asymmetric -β-Hydroxy-L-Glutamic Acid. *Tetrahedron Lett.* 25 (51), 5855–5858. doi:10.1016/s0040-4039(01)81703-2

Gieselman, M. D., Xie, L., and Van Der Donk, W. A. (2001). Synthesis of a Selenocysteine-Containing Peptide by Native Chemical Ligation. *Org. Lett.* 3 (9), 1331–1334. doi:10.1021/ol015712o

Goto, Y., Katoh, T., and Suga, H. (2011). Flexizymes for Genetic Code Reprogramming. *Nat. Protoc.* 6 (6), 779–790. doi:10.1038/nprot.2011.331

Guan, X., Drake, M. R., and Tan, Z. (2013). Total Synthesis of Human Galanin-like Peptide through an Aspartic Acid Ligation. *Org. Lett.* 15 (24), 6128–6131. doi:10.1021/ol402984r

Haase, C., Rohde, H., and Seitz, O. (2008). Native Chemical Ligation at Valine. *Angew. Chem. Int. Ed.* 47 (36), 6807–6810. doi:10.1002/anie.200801590

Hackenberger, C. P. R. (2006). The Reduction of Oxidized Methionine Residues in Peptide Thioesters with NH4I-Me2S. *Org. Biomol. Chem.* 4 (11), 2291–2295. doi:10.1039/b603543d

Hackeng, T. M., Griffin, J. H., and Dawson, P. E. (1999). Protein Synthesis by Native Chemical Ligation: Expanded Scope by Using Straightforward

Methodology. *Proc. Natl. Acad. Sci.* 96 (18), 10068–10073. doi:10.1073/pnas.96.18.10068

Harpaz, Z., Siman, P., Kumar, K. S. A., and Brik, A. (2010). Protein Synthesis Assisted by Native Chemical Ligation at Leucine. *Chem. Eur. J. Chem. Bio.* 11 (9), 1232–1235. doi:10.1002/cbic.201000168

Hofmann, R. M., and Muir, T. W. (2002). Recent Advances in the Application of Expressed Protein Ligation to Protein Engineering. *Curr. Opin. Biotechnol.* 13 (4), 297–303. doi:10.1016/s0958-1669(02)00326-9

Hondal, R. J., Nilsson, B. L., and Raines, R. T. (2001). Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation. *J. Am. Chem. Soc.* 123 (21), 5140–5141. doi:10.1021/ja005885t

Hsieh, Y. S. Y., Wijeyewickrema, L. C., Wilkinson, B. L., Pike, R. N., and Payne, R. J. (2014). Total Synthesis of Homogeneous Variants of Hirudin P6: A Post-Translationally Modified Anti-thrombotic Leech-Derived Protein. *Angew. Chem. Int. Ed.* 53 (15), 3947–3951. doi:10.1002/anie.201310777

Jbara, M., Guttmann-Raviv, N., Maity, S. K., Ayoub, N., and Brik, A. (2017). Total Chemical Synthesis of Methylated Analogues of Histone 3 Revealed KDM4D as a Potential Regulator of H3K79me3. *Bioorg. Med. Chem.* 25 (18), 4966–4970. doi:10.1016/j.bmc.2017.04.015

Jbara, M., Maity, S. K., Morgan, M., Wolberger, C., and Brik, A. (2016). Chemical Synthesis of Phosphorylated Histone H2A at Tyr57 Reveals Insight into the Inhibition Mode of the SAGA Deubiquitinating Module. *Angew. Chem. Int. Ed.* 55 (16), 4972–4976. doi:10.1002/anie.201600638

Kilic, S., Boichenko, I., Lechner, C. C., and Fierz, B. (2018). A Bi-terminal Protein Ligation Strategy to Probe Chromatin Structure during DNA Damage. *Chem. Sci.* 9 (15), 3704–3709. doi:10.1039/c8sc00681d

Kumar, K. S. A., Bavikar, S. N., Spasser, L., Moyal, T., Ohayon, S., and Brik, A. (2011). Total Chemical Synthesis of a 304 Amino Acid K48-Linked Tetraubiquitin Protein. *Angew. Chem. Int. Ed.* 50 (27), 6137–6141. doi:10.1002/anie.201101920

Kumar, K. S. A., Spasser, L., Erlich, L. A., Bavikar, S. N., and Brik, A. (2010). Total Chemical Synthesis of Di-ubiquitin Chains. *Angew. Chem. Int. Edition* 49 (48), 9126–9131. doi:10.1002/anie.201003763

Li, H., Zhang, J., An, C., and Dong, S. (2021b). Probing N-Glycan Functions in Human Interleukin-17A Based on Chemically Synthesized Homogeneous Glycoforms. *J. Am. Chem. Soc.* 143 (7), 2846–2856. doi:10.1021/jacs.0c12448

Li, J.-B., Qi, Y.-K., He, Q.-Q., Ai, H.-S., Liu, S.-l., Wang, J.-X., et al. (2018). Chemically Synthesized Histone H2A Lys13 Di-ubiquitination Promotes Binding of 53BP1 to Nucleosomes. *Cell Res* 28 (2), 257–260. doi:10.1038/cr.2017.157

Li, Y., Heng, J., Sun, D., Zhang, B., Zhang, X., Zheng, Y., et al. (2021a). Chemical Synthesis of a Full-Length G-Protein-Coupled Receptor β2-Adrenergic Receptor with Defined Modification Patterns at the C-Terminus. *J. Am. Chem. Soc.* 143 (42), 17566–17576. doi:10.1021/jacs.1c07369

Malins, L. R., Cergol, K. M., and Payne, R. J. (2014). Chemoselective Sulfenylation and Peptide Ligation at Tryptophan. *Chem. Sci.* 5 (1), 260–266. doi:10.1039/c3sc51497h

Malins, L. R., Cergol, K. M., and Payne, R. J. (2013). Peptide Ligation-Desulfurization Chemistry at Arginine. *ChemBioChem* 14 (5), 559–563. doi:10.1002/cbic.201300049

Malins, L. R., Giltrap, A. M., Dowman, L. J., and Payne, R. J. (2015a). Synthesis of β-Thiol Phenylalanine for Applications in One-Pot Ligation-Desulfurization Chemistry. *Org. Lett.* 17 (9), 2070–2073. doi:10.1021/acs.orglett.5b00597

Malins, L. R., Mitchell, N. J., McGowan, S., and Payne, R. J. (2015b). Oxidative Deselenization of Selenocysteine: Applications for Programmed Ligation at Serine. *Angew. Chem. Int. Ed.* 54 (43), 12716–12721. doi:10.1002/anie.201504639

Malins, L. R., and Payne, R. J. (2012). Synthesis and Utility of β-Selenol-Phenylalanine for Native Chemical Ligation-Deselenization Chemistry. *Org. Lett.* 14 (12), 3142–3145. doi:10.1021/ol3012265

Marin, J., Didierjean, C., Aubry, A., Casimir, J.-R., Briand, J.-P., and Guichard, G. (2004). Synthesis of Enantiopure 4-Hydroxypipecolate and 4-Hydroxylysine Derivatives from a Common 4,6-Dioxopiperidinecarboxylate Precursor. *J. Org. Chem.* 69 (1), 130–141. doi:10.1021/jo0353886

Merkx, R., De Bruin, G., Kruithof, A., Van Den Bergh, T., Snip, E., Lutz, M., et al. (2013). Scalable Synthesis of γ-thiolysine Starting from Lysine and a Side by Side Comparison with δ-thiolysine in Non-enzymatic Ubiquitination. *Chem. Sci.* 4 (12), 4494. doi:10.1039/c3sc51599k

Metanis, N., Keinan, E., and Dawson, P. E. (2010). Traceless Ligation of Cysteine Peptides Using Selective Deselenization. *Angew. Chem. Int. Edition* 49 (39), 7049–7053. doi:10.1002/anie.201001900

Mitchell, N. J., Malins, L. R., Liu, X., Thompson, R. E., Chan, B., Radom, L., et al. (2015). Rapid Additive-free Selenocystine-Selenoester Peptide Ligation. *J. Am. Chem. Soc.* 137 (44), 14011–14014. doi:10.1021/jacs.5b07237

Mitchell, N. J., Sayers, J., Kulkarni, S. S., Clayton, D., Goldys, A. M., Ripoll-Rozada, J., et al. (2017). Accelerated Protein Synthesis via One-Pot Ligation-Deselenization Chemistry. *Chem* 2 (5), 703–715. doi:10.1016/j.chempr.2017.04.003

Moura, A., Savageau, M. A., and Alves, R. (2013). Relative Amino Acid Composition Signatures of Organisms and Environments. *PLoS ONE* 8 (10), e77319. doi:10.1371/journal.pone.0077319

Muir, T. W. (2003). Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem.* 72 (1), 249–289. doi:10.1146/annurev.biochem.72.121801.161900

Muir, T. W., Sondhi, D., and Cole, P. A. (1998). Expressed Protein Ligation: A General Method for Protein Engineering. *Proc. Natl. Acad. Sci.* 95 (12), 6705–6710. doi:10.1073/pnas.95.12.6705

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A Highly Flexible tRNA Acylation Method for Non-natural Polypeptide Synthesis. *Nat. Methods* 3 (5), 357–359. doi:10.1038/nmeth877

Murakami, M., Kiuchi, T., Nishihara, M., Tezuka, K., Okamoto, R., Izumi, M., et al. (2016). Chemical Synthesis of Erythropoietin Glycoforms for Insights into the Relationship between Glycosylation Pattern and Bioactivity. *Sci. Adv.* 2 (1), e1500678. doi:10.1126/sciadv.1500678

Nishimoto, Y., Okita, A., Yasuda, M., and Baba, A. (2012). Synthesis of a Wide Range of Thioethers by Indium Triiodide Catalyzed Direct Coupling between Alkyl Acetates and Thiosilanes. *Org. Lett.* 14 (7), 1846–1849. doi:10.1021/ol300450j

Nuijens, T., Toplak, A., Schmidt, M., Ricci, A., and Cabri, W. (2019). Natural Occurring and Engineered Enzymes for Peptide Ligation and Cyclization. *Front. Chem.* 7 (829), 829. doi:10.3389/fchem.2019.00829

Offer, J. (2010). Native Chemical Ligation with Nα Acyl Transfer Auxiliaries. *Biopolymers* 94 (4), 530–541. doi:10.1002/bip.21455

Pan, M., Zheng, Q., Gao, S., Qu, Q., Yu, Y., Wu, M., et al. (2019). Chemical Synthesis of Structurally Defined Phosphorylated Ubiquitins Suggests Impaired Parkin Activation by Phosphorylated Ubiquitins with a Non-phosphorylated Distal Unit. *CCS Chem.* 1 (5), 476–489. doi:10.31635/ccschem.019.20190001

Passiniemi, M., and Koskinen, A. M. (2013). Garner's Aldehyde as a Versatile Intermediate in the Synthesis of Enantiopure Natural Products. *Beilstein J. Org. Chem.* 9, 2641–2659. doi:10.3762/bjoc.9.300

Pasunooti, K. K., Banerjee, B., Yap, T., Jiang, Y., and Liu, C.-F. (2015). Auxiliary-Directed Pd-Catalyzed γ-C(sp3)-H Bond Activation of α-Aminobutanoic Acid Derivatives. *Org. Lett.* 17 (24), 6094–6097. doi:10.1021/acs.orglett.5b03118

Pasunooti, K. K., Yang, R., Banerjee, B., Yap, T., and Liu, C.-F. (2016). 5-Methylisoxazole-3-carboxamide-Directed Palladium-Catalyzed γ-C(sp3)-H Acetoxylation and Application to the Synthesis of γ-Mercapto Amino Acids for Native Chemical Ligation. *Org. Lett.* 18 (11), 2696–2699. doi:10.1021/acs.orglett.6b01160

Pasunooti, K. K., Yang, R., Vedachalam, S., Gorityala, B. K., Liu, C.-F., and Liu, X.-W. (2009). Synthesis of 4-Mercapto-L-Lysine Derivatives: Potential Building Blocks for Sequential Native Chemical Ligation. *Bioorg. Med. Chem. Lett.* 19 (22), 6268–6271. doi:10.1016/j.bmcl.2009.09.107

Pollock, S., and Kent, S. B. H. (2011). An Investigation into the Origin of the Dramatically Reduced Reactivity of Peptide-Prolyl-Thioesters in Native Chemical Ligation. *Chem. Commun.* 47 (8), 2342–2344. doi:10.1039/c0cc04120c

Premdjee, B., Andersen, A. S., Larance, M., Conde-Frieboes, K. W., and Payne, R. J. (2021). Chemical Synthesis of Phosphorylated Insulin-like Growth Factor Binding Protein 2. *J. Am. Chem. Soc.* 143 (14), 5336–5342. doi:10.1021/jacs.1c02280

Qi, Y.-K., Si, Y.-Y., Du, S.-S., Liang, J., Wang, K.-W., and Zheng, J.-S. (2019). Recent Advances in the Chemical Synthesis and Semi-synthesis of Poly-Ubiquitin-Based Proteins and Probes. *Sci. China Chem.* 62 (3), 299–312. doi:10.1007/s11426-018-9401-8

Qu, W., Zha, Z., Ploessl, K., Lieberman, B. P., Zhu, L., Wise, D. R., et al. (2011). Synthesis of Optically Pure 4-Fluoro-Glutamines as Potential Metabolic Imaging Agents for Tumors. *J. Am. Chem. Soc.* 133 (4), 1122–1133. doi:10.1021/ja109203d

Quaderer, R., Sewing, A., and Hilvert, D. (2001). Selenocysteine-Mediated Native Chemical Ligation. *Helvetica Chim. Acta* 84 (5), 1197–1206. doi:10.1002/1522-2675(20010516)84:5<1197:aid-hlca1197>3.0.co;2-#

Rashid Baig, N. B., Chandrakala, R. N., Sudhir, V. S., and Chandrasekaran, S. (2010). Synthesis of Unnatural Selenocystines and β-Aminodiselenides via Regioselective Ring-Opening of Sulfamidates Using a Sequential, One-Pot, Multistep Strategy. *J. Org. Chem.* 75 (9), 2910–2921. doi:10.1021/jo1001388

Reddy, P. S., Dery, S., and Metanis, N. (2016). Chemical Synthesis of Proteins with Non-strategically Placed Cysteines Using Selenazolidine and Selective Deselenization. *Angew. Chem.* 128 (3), 1004–1007. doi:10.1002/ange.201509378

Reimann, O., Glanz, M., and Hackenberger, C. P. R. (2015). Native Chemical Ligation between Asparagine and Valine: Application and Limitations for the Synthesis of Tri-phosphorylated C-Terminal Tau. *Bioorg. Med. Chem.* 23 (12), 2890–2894. doi:10.1016/j.bmc.2015.03.028

Salmond, G. P. C., and Fineran, P. C. (2015). A century of the Phage: Past, Present and Future. *Nat. Rev. Microbiol.* 13 (12), 777–786. doi:10.1038/nrmicro3564

Sayers, J., Karpati, P. M. T., Mitchell, N. J., Goldys, A. M., Kwong, S. M., Firth, N., et al. (2018a). Construction of Challenging Proline-Proline Junctions via Diselenide-Selenoester Ligation Chemistry. *J. Am. Chem. Soc.* 140 (41), 13327–13334. doi:10.1021/jacs.8b07877

Sayers, J., Payne, R. J., and Winssinger, N. (2018b). Peptide Nucleic Acid-Templated Selenocystine-Selenoester Ligation Enables Rapid miRNA Detection. *Chem. Sci.* 9 (4), 896–903. doi:10.1039/c7sc02736b

Sayers, J., Thompson, R. E., Perry, K. J., Malins, L. R., and Payne, R. J. (2015). Thiazolidine-Protected β-Thiol Asparagine: Applications in One-Pot Ligation-Desulfurization Chemistry. *Org. Lett.* 17 (19), 4902–4905. doi:10.1021/acs.orglett.5b02468

Schmidt, M., Toplak, A., Quaedflieg, P. J., and Nuijens, T. (2017). Enzyme-mediated Ligation Technologies for Peptides and Proteins. *Curr. Opin. Chem. Biol.* 38, 1–7. doi:10.1016/j.cbpa.2017.01.017

Seenaiah, M., Jbara, M., Mali, S. M., and Brik, A. (2015). Convergent versus Sequential Protein Synthesis: The Case of Ubiquitinated and Glycosylated H2B. *Angew. Chem. Int. Ed.* 54 (42), 12374–12378. doi:10.1002/anie.201503309

Shabani, S., Wu, Y., Ryan, H. G., and Hutton, C. A. (2021). Progress and Perspectives on Directing Group-Assisted Palladium-Catalysed C-H Functionalisation of Amino Acids and Peptides. *Chem. Soc. Rev.* 50 (16), 9278–9343. doi:10.1039/d0cs01441a

Shibata, N., Baldwin, J. E., Jacobs, A., and Wood, M. E. (1996). Electrophilic Sulfenylation in a Stereocontrolled Synthesis of Protected (2R,3R)-3-Mercaptoaspartic Acid from -aspartic Acid. *Tetrahedron* 52 (39), 12839–12852. doi:10.1016/0040-4020(96)00765-x

Shogren-Knaak, M. A., Fry, C. J., and Peterson, C. L. (2003). A Native Peptide Ligation Strategy for Deciphering Nucleosomal Histone Modifications. *J. Biol. Chem.* 278 (18), 15744–15748. doi:10.1074/jbc.m301445200

Siman, P., Karthikeyan, S. V., and Brik, A. (2012). Native Chemical Ligation at Glutamine. *Org. Lett.* 14 (6), 1520–1523. doi:10.1021/ol300254y

Siman, P., Karthikeyan, S. V., Nikolov, M., Fischle, W., and Brik, A. (2013). Convergent Chemical Synthesis of Histone H2B Protein for the Site-specific Ubiquitination at Lys34. *Angew. Chem. Int. Ed.* 52 (31), 8059–8063. doi:10.1002/anie.201303844

Tan, Z., Shang, S., and Danishefsky, S. J. (2010). Insights into the Finer Issues of Native Chemical Ligation: An Approach to Cascade Ligations. *Angew. Chem. Int. Ed.* 49 (49), 9500–9503. doi:10.1002/anie.201005513

Thompson, R. E., Chan, B., Radom, L., Jolliffe, K. A., and Payne, R. J. (2013). Chemoselective Peptide Ligation-Desulfurization at Aspartate. *Angew. Chem. Int. Ed.* 52 (37), 9723–9727. doi:10.1002/anie.201304793

Thompson, R. E., Liu, X., Alonso-García, N., Pereira, P. J. B., Jolliffe, K. A., and Payne, R. J. (2014). Trifluoroethanethiol: An Additive for Efficient One-Pot Peptide Ligation−Desulfurization Chemistry. *J. Am. Chem. Soc.* 136 (23), 8161–8164. doi:10.1021/ja502806r

Thompson, R. E., Liu, X., Ripoll-Rozada, J., Alonso-García, N., Parker, B. L., Pereira, P. J. B., et al. (2017). Tyrosine Sulfation Modulates Activity of Tick-Derived Thrombin Inhibitors. *Nat. Chem* 9 (9), 909–917. doi:10.1038/nchem.2744

Thompson, R. E., and Muir, T. W. (2020). Chemoenzymatic Semisynthesis of Proteins. *Chem. Rev.* 120 (6), 3051–3126. doi:10.1021/acs.chemrev.9b00450

Townsend, S. D., Tan, Z., Dong, S., Shang, S., Brailsford, J. A., and Danishefsky, S. J. (2012). Advances in Proline Ligation. *J. Am. Chem. Soc.* 134 (8), 3912–3916. doi:10.1021/ja212182q

Van Der Heden Van Noort, G. J., Kooij, R., Elliott, P. R., Komander, D., and Ovaa, H. (2017). Synthesis of Poly-Ubiquitin Chains Using a Bifunctional Ubiquitin Monomer. *Org. Lett.* 19 (24), 6490–6493. doi:10.1021/acs.orglett.7b03085

Wals, K., and Ovaa, H. (2014). Unnatural Amino Acid Incorporation in E. coli: Current and Future Applications in the Design of Therapeutic Proteins. *Front. Chem.* 2 (15), 15. doi:10.3389/fchem.2014.00015

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 (48), 9248–9252. doi:10.1002/anie.200704195

Wang, P., Dong, S., Shieh, J.-H., Peguero, E., Hendrickson, R., Moore, M. A. S., et al. (2013). Erythropoietin Derived by Chemical Synthesis. *Science* 342 (6164), 1357–1360. doi:10.1126/science.1245095

Wang, X., Ashhurst, A. S., Dowman, L. J., Watson, E. E., Li, H. Y., Fairbanks, A. J., et al. (2020). Total Synthesis of Glycosylated Human Interferon-γ. *Org. Lett.* 22 (17), 6863–6867. doi:10.1021/acs.orglett.0c02401

Wang, X., Corcilius, L., Premdjee, B., and Payne, R. J. (2020). Synthesis and Utility of β-Selenophenylalanine and β-Selenoleucine in Diselenide-Selenoester Ligation. *J. Org. Chem.* 85 (3), 1567–1578. doi:10.1021/acs.joc.9b02665

Wang, X., Sanchez, J., Stone, M. J., and Payne, R. J. (2017). Sulfation of the Human Cytomegalovirus Protein UL22A Enhances Binding to the Chemokine RANTES. *Angew. Chem. Int. Ed.* 56 (29), 8490–8494. doi:10.1002/anie.201703059

Wang, Z. A., and Cole, P. A. (2020). *Methods and Applications of Expressed Protein Ligation.* Springer US, 1–13. doi:10.1007/978-1-0716-0434-2_1

Watson, E. E., Liu, X., Thompson, R. E., Ripoll-Rozada, J., Wu, M., Alwis, I., et al. (2018). Mosquito-Derived Anophelin Sulfoproteins Are Potent Antithrombotics. *ACS Cent. Sci.* 4 (4), 468–476. doi:10.1021/acscentsci.7b00612

Watson, E. E., Ripoll-Rozada, J., Lee, A. C., Wu, M. C. L., Franck, C., Pasch, T., et al. (2019). Rapid Assembly and Profiling of an Anticoagulant Sulfoprotein Library. *Proc. Natl. Acad. Sci. USA* 116 (28), 13873–13878. doi:10.1073/pnas.1905177116

Weeks, A. M., and Wells, J. A. (2020). Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* 120 (6), 3127–3160. doi:10.1021/acs.chemrev.9b00372

Wilkinson, B. L., Stone, R. S., Capicciotti, C. J., Thaysen-Andersen, M., Matthews, J. M., Packer, N. H., et al. (2012). Total Synthesis of Homogeneous Antifreeze Glycopeptides and Glycoproteins. *Angew. Chem. Int. Ed.* 51 (15), 3606–3610. doi:10.1002/anie.201108682

Yan, L. Z., and Dawson, P. E. (2001). Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* 123 (4), 526–533. doi:10.1021/ja003265m

Yang, R., Pasunooti, K. K., Li, F., Liu, X.-W., and Liu, C.-F. (2009). Dual Native Chemical Ligation at Lysine. *J. Am. Chem. Soc.* 131 (38), 13592–13593. doi:10.1021/ja905491p

Yin, H., Zheng, M., Chen, H., Wang, S., Zhou, Q., Zhang, Q., et al. (2020). Stereoselective and Divergent Construction of β-Thiolated/Selenolated Amino Acids via Photoredox-Catalyzed Asymmetric Giese Reaction. *J. Am. Chem. Soc.* 142 (33), 14201–14209. doi:10.1021/jacs.0c04994

Directed Evolution Pipeline for the Improvement of Orthogonal Translation Machinery for Genetic Code Expansion at Sense Codons

Wil Biddle, David G. Schwark, Margaret A. Schmitt and John D. Fisk *

Department of Chemistry, University of Colorado Denver, Denver, CO, United States

The expansion of the genetic code beyond a single type of noncanonical amino acid (ncAA) is hindered by inefficient machinery for reassigning the meaning of sense codons. A major obstacle to using directed evolution to improve the efficiency of sense codon reassignment is that fractional sense codon reassignments lead to heterogeneous mixtures of full-length proteins with either a ncAA or a natural amino acid incorporated in response to the targeted codon. In stop codon suppression systems, missed incorporations lead to truncated proteins; improvements in activity may be inferred from increased protein yields or the production of downstream reporters. In sense codon reassignment, the heterogeneous proteins produced greatly complicate the development of screens for variants of the orthogonal machinery with improved activity. We describe the use of a previously-reported fluorescence-based screen for sense codon reassignment as the first step in a directed evolution workflow to improve the incorporation of a ncAA in response to the Arg AGG sense codon. We first screened a library with diversity introduced into both the orthogonal *Methanocaldococcus jannaschii* tyrosyl tRNA anticodon loop and the cognate aminoacyl tRNA synthetase (aaRS) anticodon binding domain for variants that improved incorporation of tyrosine in response to the AGG codon. The most efficient variants produced fluorescent proteins at levels indistinguishable from the *E. coli* translation machinery decoding tyrosine codons. Mutations to the *M. jannaschii* aaRS that were found to improve tyrosine incorporation were transplanted onto a *M. jannaschii* aaRS evolved for the incorporation of *para*-azidophenylalanine. Improved ncAA incorporation was evident using fluorescence- and mass-based reporters. The described workflow is generalizable and should enable the rapid tailoring of orthogonal machinery capable of activating diverse ncAAs to any sense codon target. We evaluated the selection based improvements of the orthogonal pair in a host genomically engineered for reduced target codon competition. Using this particular system for evaluation of arginine AGG codon reassignment, however, *E. coli* strains with genomes engineered to remove competing tRNAs did not outperform a standard laboratory *E. coli* strain in sense codon reassignment.

Keywords: genetic code expansion, sense codon reassignment, noncanonical amino acid, synthetic biology, directed evolution, protein engineering

INTRODUCTION

The genetically-encoded introduction of noncanonical amino acids (ncAAs) is a powerful tool for increasing the chemical diversity of proteins because it enables the precise placement of desired side chain functionality within a growing peptide chain. Two general approaches to expand the genetic code have been widely employed: nonsense suppression and residue specific reassignment (**Figure 1**) (Wiltschi and Budisa, 2007; Ngo and Tirrell, 2011). Advances in both technologies have increased the efficiency with which a single ncAA can be genetically encoded. Nonsense suppression technology has been improved through genome modifications, enabling the elimination of the release factor that typically competes to read the amber stop codon as a termination signal (Mukai et al., 2010; Johnson et al., 2011; Lajoie et al., 2013). Amino acid-specific reassignment has been updated by breaking the degeneracy of the genetic code to enable the reassignment of individual sense codons (Kwon et al., 2003;

Bohlke and Budisa, 2014; Zeng et al., 2014; Lee et al., 2015; Mukai et al., 2015; Ho et al., 2016; Kwon and Choi, 2016; Wang and Tsao, 2016). Improvements in both methods have focused primarily on genetic additions and deletions rather than on functional improvements of the evolved orthogonal translation components central to both technologies.

In vivo incorporation of ncAAs by either nonsense suppression or sense codon reassignment (SCR) requires the introduction of an orthogonal aminoacyl tRNA synthetase (aaRS) engineered to recognize and attach a ncAA onto its cognate, orthogonal tRNA (Furter, 1998; Wang et al., 2001). The orthogonal tRNA is not recognized by the set of endogenous aaRSs, and the orthogonal aaRS does not recognize the complement of endogenous tRNAs. The vast majority of the ribosomally incorporated ncAAs have been introduced using derivatives of only two orthogonal pairs: the tyrosine tRNA/aaRS pair from *Methanocaldococcus jannaschii* (*M. jannaschii*) and the pyrrolysine tRNA/aaRS pair from *Methanosarcina*
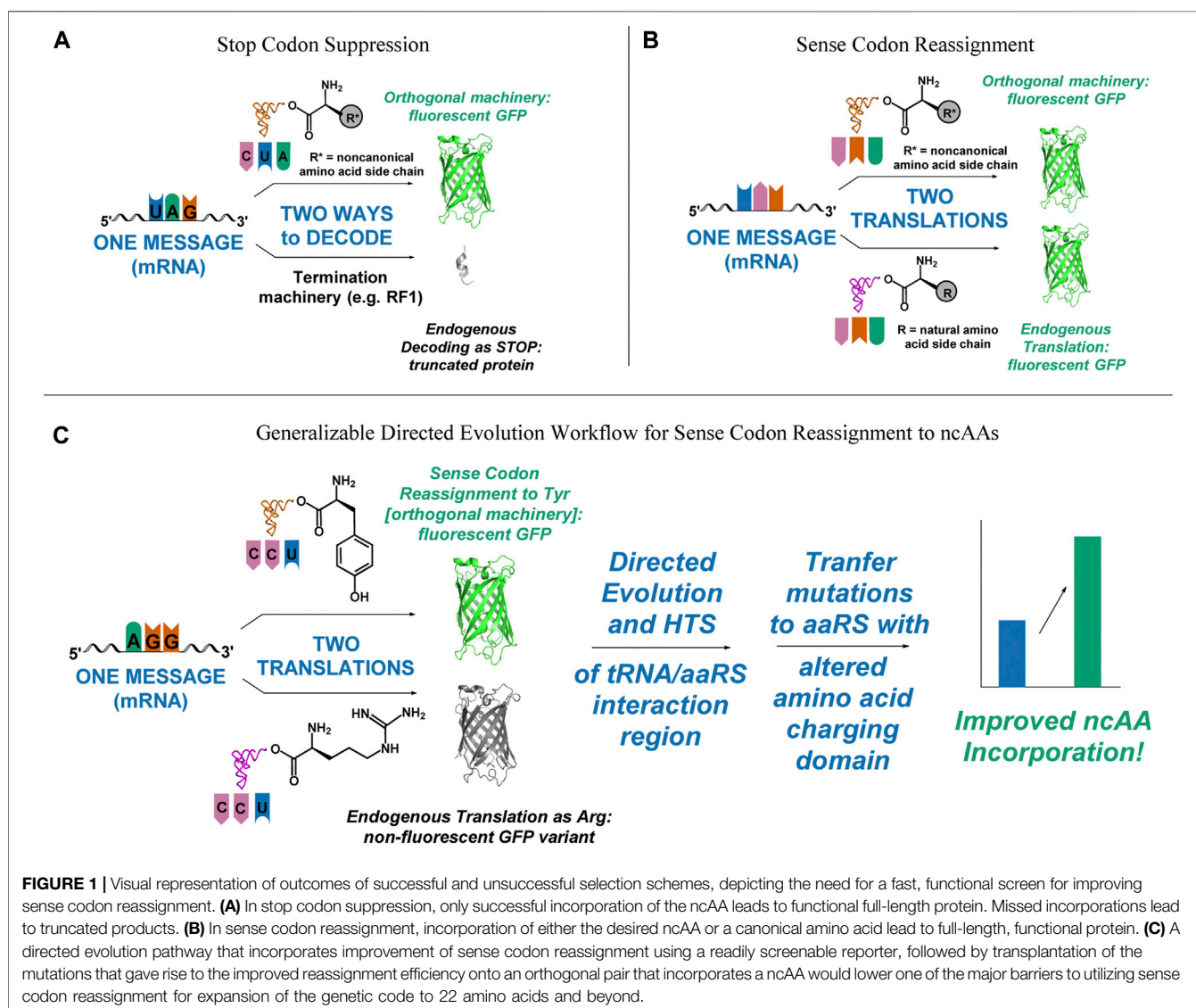
FIGURE 1 | Visual representation of outcomes of successful and unsuccessful selection schemes, depicting the need for a fast, functional screen for improving sense codon reassignment. **(A)** In stop codon suppression, only successful incorporation of the ncAA leads to functional full-length protein. Missed incorporations lead to truncated products. **(B)** In sense codon reassignment, incorporation of either the desired ncAA or a canonical amino acid lead to full-length, functional protein. **(C)** A directed evolution pathway that incorporates improvement of sense codon reassignment using a readily screenable reporter, followed by transplantation of the mutations that gave rise to the improved reassignment efficiency onto an orthogonal pair that incorporates a ncAA would lower one of the major barriers to utilizing sense codon reassignment for expansion of the genetic code to 22 amino acids and beyond.

species. (Wang et al., 2001; Polycarpo et al., 2006). Variants of these pairs that recognize and aminoacylate over 150 different ncAAs have been developed (Liu and Schultz, 2010; Wan et al., 2014; Dumas et al., 2015). Continuing efforts to find or engineer new orthogonal pairs have identified several additional potential orthogonal systems. These systems, however, have not yet found widespread use in genetic code expansion (Cervettini et al., 2020; Ding et al., 2020; Zhao et al., 2021).

Here, our previously-developed fluorescence-based screen is utilized as the first step in a directed evolution workflow to improve the incorporation of a ncAA in response to the Arg AGG sense codon (Kuhn et al., 2010; Biddle et al., 2015). The fluorescence-based screen takes advantage of the absolute requirement of tyrosine in the central position of a Thr-Tyr-Gly triad for mature GFP fluorophore formation. We first screened a library with diversity introduced into both the orthogonal *M. jannaschii* tyrosyl tRNA anticodon loop and the cognate aaRS anticodon binding domain for variants that improved incorporation of tyrosine in response to the AGG codon. The most efficient variants produced fluorescent protein at levels indistinguishable from the *Escherichia coli* (*E. coli*) translation machinery decoding Tyr codons, and reassignment efficiency improved from $56.9 \pm 2.4\%$ to $98.6 \pm 4.7\%$. Mutations to the *M. jannaschii* aaRS that were found to improve tyrosine incorporation were transplanted onto a *M. jannaschii* aaRS evolved for the incorporation of *para*-azidophenylalanine (pAzF) (Chin et al., 2002). The same suite of mutations improved incorporation of pAzF in response to the AGG codon from $29.5 \pm 0.1\%$ to $50.1 \pm 1.1\%$ as measured *via* fluorescence, an approximately 1.7 fold improvement for both amino acids. The improved *para*-azidophenylalanine incorporating tRNA/aaRS pair was evaluated in strains from which competition for the AGG codon was reduced by deletion of the gene for the primary endogenous tRNA that reads the AGG codon (Lee et al., 2015). Improvements in sense codon reassignment efficiency due to reduced competition did not appear to combine with improvements selected through directed evolution. We have previously quantified the reassignment efficiency of several sense codons by both the natural, tyrosine-incorporating *M. jannaschii* and an evolved tyrosine-incorporating *M. barkeri* orthogonal pair (Schmitt et al., 2018; Schwark et al., 2020; Schwark et al., 2021). This workflow should be generalizable for improving incorporation of other ncAAs in response to other sense codon targets.

Nonsense (stop codon) suppression has been the most widely used method for introduction of noncanonical amino acids into the genetic code. In *E. coli*, the amber stop codon is employed both because it is a termination signal and because it is the least frequently used codon. Nonsense suppression utilizes orthogonal aaRS variants to aminoacylate >150 different ncAAs onto essentially an identical cognate tRNA with a CUA anticodon. Variation across orthogonal systems that incorporate different ncAAs primarily resides within the amino acid recognition domain of the aaRS. tRNA/aaRS recognition is typically not altered across variants. The kinetic efficiencies of aaRSs evolved to recognize ncAAs are generally poor. *In vitro* kinetic

aminoacylation efficiencies for these enzymes are on the order of 1% that of natural enzymes (Guo et al., 2014; Amiram et al., 2015; Rauch et al., 2016). The measured efficiency of ncAA incorporation is dependent on systemic variables beyond the kinetics of aminoacylation (e.g. ncAA, tRNA and aaRS concentrations), and single site incorporation efficiencies of 10–30% are readily achievable. For systems in which a single amber stop codon is targeted for ncAA incorporation, this strategy and enzyme function level are sufficient to produce modified proteins in usable quantities. The expected protein yield drops as the nth power of the single site incorporation efficiency (where n is the number of incorporations attempted). Even for a highly efficient (e.g. 50%) ncAA-incorporating tRNA/aaRS system, the expected protein yield for a protein containing 3 amber stop codons would be 12.5% of the yield expected if the amber codon were a typical canonical amino acid (Schwark et al., 2018).

Advances in the technology that underpins both amber stop codon suppression and sense codon reassignment by breaking the degeneracy of the genetic code have focused primarily on extrinsic system improvements, such as increasing the expression level of the orthogonal pair, improving the interactions of the orthogonal tRNA with the endogenous translation components (e.g. EF-Tu), or reducing endogenous competition for the targeted codon. The orthogonal translation components are orthogonal because they are taken from a phylogenetically distant species, and the efficiency of incorporation could potentially be increased by improved assimilation of the orthogonal components into the host organism. A large amount of "orthogonal space" in the vicinity of each orthogonal system appears to exist, as mutually orthogonal variants are readily evolvable (Neumann et al., 2010; Willis and Chin, 2018; Cervettini et al., 2020). Orthogonal machinery levels have been adjusted by expressing both tRNAs and aaRSs from cassettes with different promoters in plasmids with different copy numbers (Ryu and Schultz, 2006; Young et al., 2010; Chatterjee et al., 2013). A handful of attempts to improve amber stop codon suppression *via* tRNA mutation have focused on changes to tRNA sequences that modulate interactions with elongation factor Tu (EF-Tu) (Guo et al., 2009; Schrader et al., 2011; Mittelstaet et al., 2013; Fan et al., 2015; Maranhao and Ellington, 2017). The extent to which tRNA mutations that improved incorporation of one ncAA in response to amber codons functioned when transferred between aaRSs specific for different ncAAs suggested that the extent of transfer is idiosyncratic, although in general, improved variants were identified (Guo et al., 2009). Reducing the competition from endogenous translation components has been the major mode of improving genetic code expansion systems. Three separate approaches involving different combinations of organismal genomic modifications enabled the deletion of the gene for the release factor that typically reads the amber stop codon as a termination signal (Mukai et al., 2010; Johnson et al., 2011; Lajoie et al., 2013). Additional genome rewriting projects seek to generate multiple "free" codons for ncAA incorporation (Ostrov et al., 2016; Wang et al., 2016; Fredens et al., 2019). Attempts to improve sense codon reassignment have involved

deletion of competing tRNA genes and antisense RNAs to reduce competition between orthogonal and host components reading the AGG arginine codon (Zeng et al., 2014; Lee et al., 2015).

An alternative strategy for improving the efficiency of orthogonal tRNA/aaRS pairs targets interactions intrinsic to the cognate pair itself. The anticodon is often an important identity element that allows a specific aaRS to recognize its appropriate tRNA (Giege et al., 1998). Changing the anticodon of the *M. jannaschii* tRNA, even to CUA for amber suppression, is known to strongly affect the efficiency of aminoacylation (Fechter et al., 2001). Increasing the recognition between the orthogonal tRNA and the anticodon binding domain of the aaRS is expected to lead to a higher concentration of aminoacylated tRNA and better kinetic competition against endogenous tRNAs capable of decoding the targeted codon. Improvements of the interactions between the components of the orthogonal translation machinery should facilitate improved incorporation efficiency with less dependence upon the system characteristics into which orthogonal translation system is placed. Several general selection strategies have been developed to improve the function of stop codon suppressing orthogonal pair systems, although the maturation of initially selected aaRSs utilized for ncAA incorporation is not commonly performed (Pott et al., 2014; Wang et al., 2015; Rauch et al., 2016; Maranhao and Ellington, 2017; Owens et al., 2017). Directed evolution of *M. jannaschii* tRNA/aaRS pair variants for improved reassignment of the amber stop codon as well as Lys AAG and His CAU sense codons have been described (Takimoto et al., 2009; Kuhn et al., 2010; Mittelstaet et al., 2013; Biddle et al., 2015; Wang et al., 2015; Biddle et al., 2016).

A major obstacle to using directed evolution to improve the efficiency of sense codon reassignment is that fractional sense codon reassignments lead to heterogeneous mixtures of full-length proteins with either a ncAA or a natural amino acid incorporated in response to the targeted codon. In stop codon suppression systems, missed incorporations lead to truncated proteins and improvements in activity can be inferred from increased protein yields or the production of downstream reporters (**Figure 1A**). In sense codon reassignment, heterogeneous mixtures of full-length proteins greatly complicate the development of screens for variants with improved activity (**Figure 1B**). Implementation of a general workflow that allows selection of orthogonal pair variants that reassign a particular sense codon first by evaluating variants with a fast, facile screen (e.g. fluorescence-activated cell sorting, FACS) followed by subsequent transfer of the mutations to a version of that orthogonal pair engineered to recognize a ncAA, for which screening would be more challenging, would lower the barrier to wider utilization of sense codon reassignment for genetic code expansion (**Figure 1C**).

METHODS AND MATERIALS

The Supplementary Materials file includes general reagents and materials, as well as detailed experimental protocols for site-directed mutagenesis, the fluorescence-based screen, and protein expression and isolation for reassignment efficiency analysis and mass spectrometry (Kunkel, 1985; Gao et al., 2003; Clackson and Lowman, 2004). This file also includes a graphical representation of the directed evolution workflow (**Supplementary Figure S1**), an image of a representative protein gel (**Supplementary Figure S2**), and a representative optical density *vs.* time plot for reassigning systems (**Supplementary Figure S3**). Finally, further experimental information, including a list of the oligonucleotide primers used for library creation, cell strain information, and preparation of electrocompetent cells are provided (Sambrook and Russell, 2001).

RESULTS AND DISCUSSION

Co-evolution of the *M. jannaschii* tRNA$_{CCU}$/TyrRS Improves Reassignment of AGG Codons to Tyrosine

We selected the arginine AGG codon for initial evaluation of this directed evolution workflow because AGG has been the most common target for sense codon reassignment using variants of both the orthogonal pyrrolysyl and *M. jannaschii* tyrosyl tRNA/aaRS pairs (Krishnakumar et al., 2013; Zeng et al., 2014; Lee et al., 2015; Mukai et al., 2015; Wang and Tsao, 2016). Zeng et al. reported approximately 90% efficient reassignment using a variant of the *Methanosarcina* pyrrolysyl tRNA/aaRS pair and media in which the concentration of arginine was controlled (Zeng et al., 2014). Mukai et al. mutated the 38 AGG codons found in essential *E. coli* genes to other arginine codons and subsequently removed the gene for endogenous tRNA$_{CCU}$, *argW*, from the genome (Mukai et al., 2015). These genetic transformations allowed incorporation of a close structural analogue of arginine at very high levels using a variant of the *Methanosarcina* pyrrolysyl tRNA/aaRS pair. Lee et al. described reassignment of the AGG codon using a variant of the *M. jannaschii* tyrosyl tRNA/aaRS pair (Lee et al., 2015). Their strategy also utilized an *argW* knockout strain, however the distribution of AGG codons in the genome was not adjusted. The fact that *argW* knockout strains are viable and exhibit only slightly reduced growth suggests that the remaining *E. coli* tRNA$_{UCU}$ is able to read AGG codons to some extent *via* an expected G/U wobble interaction. Double knockouts of both *E. coli* tRNA$_{UCU}$ and tRNA$_{CCU}$ are not viable (Mukai et al., 2015).

The efficiency of AGG reassignment was quantified in *E. coli* DH10B expressing the *M. jannaschii* tyrosyl aaRS, its cognate tRNA^{Opt} with a CCU anticodon (*M. jannaschii* tRNA$_{CCU}$/TyrRS), and a GFP reporter protein with an arginine AGG codon specifying the central fluorophore position (Young et al., 2010; Chatterjee et al., 2013; Biddle et al., 2015). The gene sequence of the GFP reporter uses a reduced codon set such that instances of the 20 sense codons decoded through wobble interactions and 10 additional least frequently used sense codons are eliminated or greatly reduced from the gene (Schmitt et al., 2018). No AGG codons are present in the aaRS gene. Changing the anticodon of the *M. jannaschii* tRNA^{Opt} to Watson—Crick base pair with the AGG codon results in 56.9 ± 2.4% reassignment efficiency. Significantly, this high efficiency is evident in rich
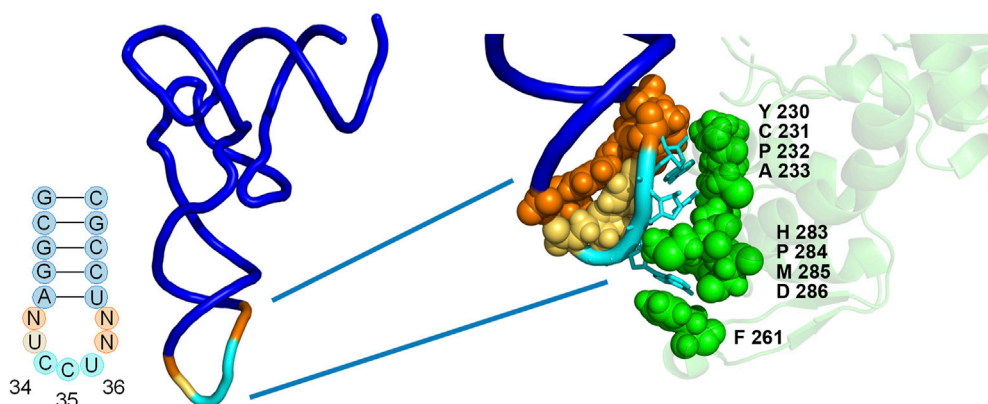
FIGURE 2 | Positions of library mutations mapped onto the co-crystal structure of the *M. jannaschii* tRNA/aaRS pair (PDB 1J1U). Positions 32, 37, and 38 within the tRNA anticodon loop were allowed to vary in conjunction with 9 amino acids in the anticodon binding domain of the aaRS. The diversity available at each position in the aaRS is described both in the text and in **Table 1**.

media without any modifications to the *E. coli* genome or alterations to the orthogonal tRNA and aaRS beyond the anticodon change. Reassignment efficiency is quantified by normalizing the observed fluorescence with a test codon at the fluorophore position to a "100% fluorescence" reference value determined by expressing sfGFP with a tyrosine UAC codon at position 66. A "0% fluorescence" reference is determined by expressing sfGFP with a non-tyrosine codon specifying the fluorophore. Both fluorescence reference systems include a plasmid expressing the orthogonal machinery to maintain a similar metabolic burden on each system.

A library of *M. jannaschii* tRNA anticodon loop and aaRS anticodon binding domain variants was constructed and screened for improved efficiency of reassignment of the AGG codon to Tyr (**Figure 2**). Co-evolution of the tRNA anticodon loop and the aaRS anticodon binding domain was hypothesized to increase the recognition and aminoacylation of the tRNA by the aaRS, which should, in turn, increase the effective concentration of the aminoacylated tRNA and lead to better competition against endogenous tRNAs for decoding the targeted codon. Furthermore, the positions within the *M. jannaschii* Tyr aaRS that contact the cognate tRNA anticodon are spatially distant from and within a separate domain relative to the amino acid binding pocket, suggesting that mutations which lead to improvement of incorporation for one amino acid may be transferable to a variant of the same enzyme that recognizes and aminoacylates a different amino acid (Kobayashi et al., 2003).

The library included diversity at 9 amino acid positions in the aaRS and 3 of 4 nucleotides in the tRNA anticodon loop outside of the anticodon (32, 37, and 38) (Biddle et al., 2015). The varied amino acid positions were chosen based on the proximity to the nucleotides of the anticodon in the co-crystal structure of the *M. jannaschii* tyrosyl tRNA/aaRS orthogonal pair (PDB 1J1U) (**Figure 2**) (Kobayashi et al., 2003). Amino acid residues with side chains within 5Å of anticodon nucleobases were varied. Amino acid diversity at each of the positions was limited in order to maintain a manageable library size, and degenerate nucleotides

were selected to include the parent *M. jannaschii* aaRS amino acid at each position (**Table 1**). Amino acids 230-233 were allowed to vary between four amino acids. Greater diversity was included at the remaining 5 positions; diversity at positions 261 and 283-286 varied between 9 and 15 different amino acids. Three of the four nucleotides flanking the anticodon in the tRNA were allowed to vary; the universally conserved U33 was preserved (**Figure 2**; **Table 1**). The library had a theoretical diversity of 4.1×10^{10} at the DNA level, 4.1×10^{9} at the expressed level.

The constructed library contained 3×10^{8} unique transformants, which were allowed to multiply 25-fold ($\sim 10^{9}$ cells). Of these, approximately 10^{7} cells were screened using fluorescence activated cell sorting (FACS), and the brightest 2% of cells were collected. Following amplification, the population was resorted, and the top 1% of fluorescent cells were collected. After amplification and a third round FACS, a portion of the collected cells were plated for single clone analysis (**Supplementary Figure S1** in Supplementary Material). 75 of 84 visibly green clones showed increased fluorescence relative to the starting *M. jannaschii* tRNA_{CCU}/TyrRS. The clones displayed a range of reassignment efficiencies, with the vast majority reassigning the AGG codon at greater than 75% efficiency. 10 of the 84 clones were selected for further analysis. Sequencing revealed eight unique sets of *M. jannaschii* tRNA_{CCU}/TyrRS variants (**Table 1**). Because false positives resulting from reporter fluorophore revertants or mutations increasing the expression of GFP have been observed in other aaRS selections, the orthogonal translation machinery plasmids were isolated and retransformed into cells with unselected GFP reporter plasmids. Analysis of the eight identified *M. jannaschii* tRNA_{CCU}/TyrRS variants suggested that the mutations within the tRNA and aaRS were responsible for the improved incorporation of tyrosine in response to the AGG codon. Sense codon reassignment efficiencies ranged from $83.2 \pm 0.9\%$ to $99.2 \pm 2.3\%$ (measurements from twelve colonies of each variant).

Strong consensus was apparent across the sequences of the selected *M. jannaschii* tRNA_{CCU}/TyrRS variants (**Table 1**). In the

TABLE 1 | Sequences and AGG reassignment efficiencies for *M. jannaschii* tyrosyl tRNA/aaRS variants selected from the library.

	Amino acid positions in aaRS anticodon binding domain									Nucleotide positions in tRNA					AGG % Eff[a]
	230	231	232	233	261	283	284	285	286	32	33	CCU	37	38	
M. jannaschii tRNA/aaRS	Y	C	P	A	F	H	P	M	D	C	U		A	A	56.9 ± 2.4%
—			amino acid residue diversity									nucleotide diversity			—
—	4	4	4	4	15	13	12	9	12	4	1		4	4	—
—			degenerate nucleotide triplet for each codon									degenerate nucleotide			—
—	KMT	TNT	MMG	KYC	NHH	VDS	BNT	WBB	VVW	N	U	—	N	N	—
C3[b]	Y	F	Q	A	F	L	R	S	G	A	U	—	A	G	98.6 ± 4.7%
F3[b]	S	F	T	A	F	L	R	S	A	A	U	—	A	G	86.6 ± 3.0%
F8	D	F	T	A	F	L	R	T	A	A	U	—	A	G	88.2 ± 1.5%
H1 (2)[c]	A	Y	T	A	F	L	R	S	A	A	U	—	A	G	99.2 ± 2.3%
D1	D	Y	T	A	Y	L	R	S	H	A	U	—	A	G	94.4 ± 2.6%
F6 (2)[c]	S	Y	K	A	Y	L	R	S	N	A	U	—	A	G	95.5 ± 1.1%
D6	A	Y	Q	A	F	Y	R	S	H	A	U	—	A	G	83.2 ± 0.9%
G4[b]	D	C	K	F	I	R	S	W	H	G	U	—	A	G	98.6 ± 2.1%

[a]*Error bars on reassignment efficiencies are the standard deviation of 12 biological replicates of each system, with the exception of* M. jannaschii *tRNA*CCU*/TyrRS, and tRNA*CCU*-C3/TyrRS-C3, which comprise evaluation of 24 biological replicates.*

[b]*Three clones (C3, F3, and G4) included mutations within the aaRS, outside of the varied positions. Clone C3 has E221K and R223G. Clones F3 and G4 have K228N.*

[c]*The number in parenthesis indicates the number of clones of the 10 characterized that had the listed sequence.*

case of the tRNA anticodon loop, the nucleotides identified at the three varied positions were identical for seven of the eight aaRS variants. The predominantly selected tRNA sequence was 5′-auCCUag-3′, where loop positions are written in lower case and anticodon nucleotides in upper case. Positions 32 and 38 were different than the starting *M. jannaschii* tRNACCU: 5′-cuCCUaa-3′. While the predominant tRNA variant included purine nucleotides at the 3 varied positions, the majority of *E. coli* tRNAs include a pyrimidine at position 32 and a purine at 38 as in the starting sequence. No *E. coli* tRNA has a purine in both positions 32 and 38. A purine/pyrimidine relationship also exists for tRNA positions 33/37, with the universally-conserved uridine at position 33. In *E. coli*, the choice of purine at position 37 is strongly dependent on the nucleotide at position 36. The majority of *E. coli* tRNAs feature adenosine (or a modified adenosine) at position 37; only a few of the tRNAs with G36 anticodons have a modified guanosine at position 37 (Machnicka et al., 2013).

The seven aaRS sequences (from nine characterized variants) that were selected along with the single tRNA sequence have strikingly similar amino acid sequences (**Table 1**). Alanine, the amino acid present at position 233 in the original aaRS, was present in each of the seven sequences. Despite having the greatest opportunity to vary, position 261 maintained a similar amino acid to the wild type aaRS, with all seven sequences either remaining Phe or mutating to Tyr. The changes selected in the remaining seven varied positions modified the amino acid size and polarity relative to the amino acid present in the starting aaRS. In all seven selected sequences, position 284 mutated from Pro to Arg. Six of seven sequences included H283L and M285S. For both Leu 283 and Ser 285, the multiple codons for each amino acid available as a result of degenerate codon choice were present at the DNA level. All seven sequences included an aromatic residue at position 231. Less of a consensus was apparent at positions 232 and 286. Position 232 universally changed from the proline found in the starting aaRS with a weak consensus for Thr

(4/7 sequences). Position 286 showed a weak consensus for alanine (3/7 sequences) or histidine (2/7 sequences). The amino acids identified at position 230 were evenly distributed between Tyr, Ala, Asp, and Ser, the 4 amino acids available at that position. The single selected tRNA with a different anticodon loop sequence (5′-guCCUag-3′) was co-selected with an aaRS with a unique suite of mutations (clone G4 tRNA/aaRS pair). The amino acids selected at six of nine varied positions in clone G4 were not identified in any of the other variants.

Three of the ten clones, including the highly efficient co-evolved tRNA/aaRS pair C3, had additional, spontaneous mutations to the aaRS sequence. Clone C3 had two mutations: E221K and R223G. Clones F3 and G4 had a K228N mutation. All three of these mutations occur within an *α*-helix near the anticodon binding domain. The E221K and R223G mutations in C3 occur near the end of the *α*-helix furthest removed from the tRNA. The K228N mutation was also observed in three of ten characterized clones identified from a similar tRNA anticodon loop/aaRS anticodon binding domain library targeting improvement of Lys AAG codon reassignment (Biddle et al., 2015). The anticodons required to Watson—Crick base pair with Lys AAG (CUU) and Arg AGG (CCU) differ only at position 35. That the same mutations arise in selections for improving the function of two very similar anticodon loops is not surprising.

Importantly, the orthogonality of the *M. jannaschii* tRNACCU-C3/TyrRS-C3 and *M. jannaschii* tRNACCU-G4/TyrRS-G4 variants was maintained. Incorporation of tyrosine in response to an AGG codon using orthogonal translation machinery vectors from which the tRNA was removed was below the limit of detection of the fluorescence-based screen (0.2%, 2 of every 1000 incorporation events in response to the AGG codon in the reporter). The anticodon is often an important identity element for aaRSs, and the *M. jannaschii* TyrRS is known to utilize the anticodon during the process of aminoacylation. The most likely way in which orthogonality of the pair might be

broken after mutation of the aaRS anticodon binding domain to better recognize the $tRNA_{CCU}$ would be aminoacylation of Tyr onto the *E. coli* $tRNA_{CCU}$. The absence of detectable fluorescence in vectors expressing either TyrRS-C3 or TyrRS-G4 without their cognate tRNA suggests that is not a frequent occurrence. These two aaRS variants were selected for the verification of orthogonality because they represent the 2 broad sequence spaces identified in the characterized clones.

As with our previous reports of sense codon reassignment using the fluorescence-based screen, the fluorescence measurements from the in cell assay were compared to fluorescence measurements from isolated protein. GFP protein production across the different AGG reassigning systems and non-reassigning controls is remarkably similar (**Supplementary Figure S2**). In this report, clarified lysate fluorescence measurements were normalized to protein amounts determined from image quantification of SDS-PAGE gels. Reassignment efficiency of AGG to tyrosine was identical within error using both types of measurement. Analysis of clarified lysates showed that *M. jannaschii* $tRNA_{CCU}$/TyrRS was 60.7 ± 2.4% efficient and that *M. jannaschii* $tRNA_{CCU}$-C3/TyrRS-C3 was 96.0 ± 2.0% efficient. The protein reassignment efficiencies from clarified lysates comprise at least 4 biological replicates of each system.

Reassignment efficiencies are calculated using the optical density-corrected fluorescence over a 4 h period of cell growth in order to eliminate biases from differences in the growth profiles of different codon reassigning systems. Bacterial cells are largely tolerant of induced missense mutations. We have previously discussed the average impact of high level codon reassignment on cellular fitness (Schmitt et al., 2018; Schwark et al., 2020). Using these orthogonal translation machinery and reporter vectors, the average fitness reduction for 20 different sense codon reassigning systems is about 2.5x the fitness reduction imposed by using 2 antibiotics to maintain 2 different protein expressing plasmids in the cells. The cell strain used for sense codon reassignment evaluation in the majority of our prior work was SB3930. Our recent reports have utilized the more commonly-available laboratory strain DH10B (Schwark et al., 2021). Codon reassignments have similar impacts on cell health in both strains.

Reassignment of Arg AGG using *M. jannaschii* $tRNA_{CCU}$/TyrRS results in reduced carrying capacity and instantaneous doubling times in DH10B relative to an "empty" translation machinery vector from which both the tRNA and aaRS gene cassettes have been deleted. *E. coli* DH10B expressing *M. jannaschii* no-tRNA/no-aaRS and the GFP reporter vector with a Tyr codon in the fluorophore have an instantaneous doubling time of 40.0 ± 3.0 min, while those reassigning the AGG codon to tyrosine have an instantaneous doubling time of 45.8 ± 2.3 min. The relative system fitness of the AGG reassigning system is ~87% that of the "empty" translation machinery control. Interestingly, expression of *M. jannaschii* $tRNA_{CCU}$-C3/TyrRS-C3 restores cell fitness back to that of the "empty" translational machinery vector control. *E. coli* DH10B reassigning the AGG codon with *M. jannaschii* $tRNA_{CCU}$-C3/TyrRS-C3 double every 36.7 ± 2.2 min in the exponential phase, for a

relative system fitness of ~109%. The tolerance toward high level reassignment of Arg AGG codons is expected to partly be the result of the low usage of the AGG codon throughout the *E. coli* genome. A representative optical density *vs.* time cell growth graph is provided as **Supplementary Figure S3**.

Mutations That Improve Incorporation of Tyr in Response to AGG Also Improve Incorporation of a ncAA

As part of a general process for directed evolution of sense codon reassigning systems, the fluorescence-based screen with the tyrosine-incorporating aaRSs would be utilized as a stand in for aaRS variants that activate ncAAs. A group of clones with different sequences (C3, H1, and G4) exhibited very high, near quantitative reassignment of AGG to Tyr using the fluorescence-based screen. The reassignment efficiencies for each of these systems were all within error of each other, based on evaluation of 12 or 24 biological replicates of each system. In order to evaluate the transferability of improvements in tRNA/aaRS interactions between aaRS variants that activate different amino acids, the tRNA anticodon loop and aaRS anticodon binding domain mutations present in clone C3 were transferred to an aaRS variant that activates *para*-azidophenylalanine (pAzFRS) and evaluated both using both the fluorescence-based screen and mass spectrometry (Chin et al., 2002; Kirshenbaum et al., 2002). The C3 mutations improved AGG reassignment to Tyr from 56.9 ± 2.4% to 98.6 ± 4.7% (**Figure 3A**).

The *M. jannaschii* pAzF incorporating variant has been employed to direct incorporation of pAzF in response to both amber stop and arginine AGG codons (Chin et al., 2002; Lee et al., 2015). pAzF has been widely used for protein crosslinking *via* photolysis and bioorthogonal derivatization *via* copper-catalyzed Huisgen cyclization with alkynes. pAzF is one of several ncAAs that produce fluorescent proteins when introduced at the fluorophore tyrosine position in GFP (Wang et al., 2003a; Kajihara et al., 2005). Introduction of pAzF into GFP results in a protein with less intense fluorescence and blue shifted maximum relative to sfGFP. The apparent brightness (quantum yield*extinction coefficient) of sfGFP with pAzF in the fluorophore is approximately 10% of wild type sfGFP (Wang et al., 2012; Reddington et al., 2013). The actual fluorescent species in GFP with pAzF substitution is the reduced form of pAzF, *para*-aminophenylalanine. Reports differ on the extent to which spontaneous reduction of the azide to an amine occurs in GFP (Wang et al., 2003a; Morris et al., 2013; Reddington et al., 2013). In biological contexts, azides are often subject to rapid reduction by common dithiols (Bayley et al., 1978; Staros et al., 1978).

The fluorescence of cells expressing both the AGG reassignment reporter and either the *M. jannaschii* $tRNA_{CCU}$/pAzFRS or *M. jannaschii* $tRNA_{CCU}$-C3/pAzFRS-C3 was monitored in both the presence and absence of pAzF. Meaningful *in vivo* fluorescence was not observed for these cultures, possibly due to the lower inherent brightness of *para*-aminophenylalanine-containing GFP, the absence of spontaneous reduction of the azide, or obfuscation of protein
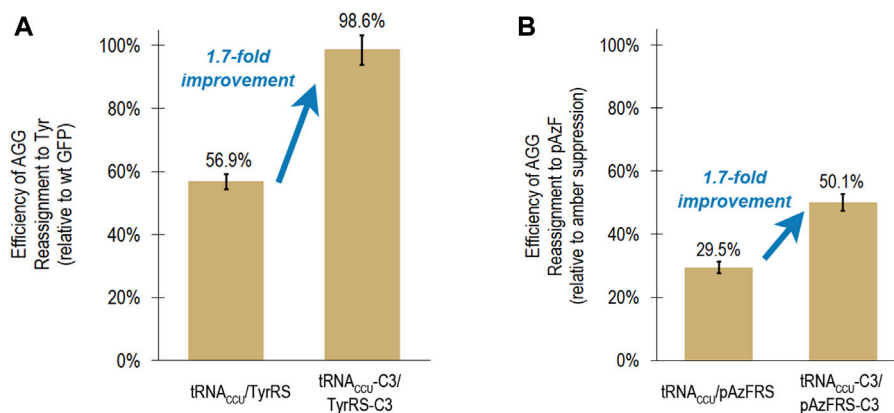
FIGURE 3 | Transferability of mutations that led to increased reassignment of the AGG codon to tyrosine to an aaRS capable of incorporating a ncAA (*para*-azidophenylalanine). In both cases, the same suite of mutations to the tRNA anticodon loop and aaRS anticodon binding domain resulted in a 1.7-fold improvement in reassignment efficiency. **(A)** Efficiency of AGG sense codon reassignment to tyrosine by two orthogonal pairs: *M. jannaschii* tRNA$_{CCU}$/TyrRS and tRNA$_{CCU}$-C3/TyrRS-C3. Reported efficiencies are the mean and standard deviation for 24 biological replicates of each system evaluated in several independent iterations of the fluorescence-based screen. Reassignment efficiencies are corroborated by fluorescence analysis of the GFP protein mixtures isolated from these cells. **(B)** Efficiency of AGG sense codon reassignment to a ncAA, pAzF, by two orthogonal pairs: *M. jannaschii* tRNA$_{CCU}$/pAzFRS and tRNA$_{CCU}$-C3/pAzFRS-C3. Reported efficiencies are the mean and standard deviation of at least 4 biological replicates of each system quantified using the fluorescence per protein of GFP mixtures isolated from these cells. The "100% pAzF incorporation" control in each experiment is based on the fluorescence per protein analysis of full length GFP proteins produced after suppression of an amber stop codon in the fluorophore position 66.

fluorescence by fluorescent components in the rich medium. A slight increase in fluorescence was observed after irradiation of the cell cultures with UV light; however, controlled irradiation of several cell cultures was challenging. Reduction of the azide side chain following protein isolation proved to be more controllable.
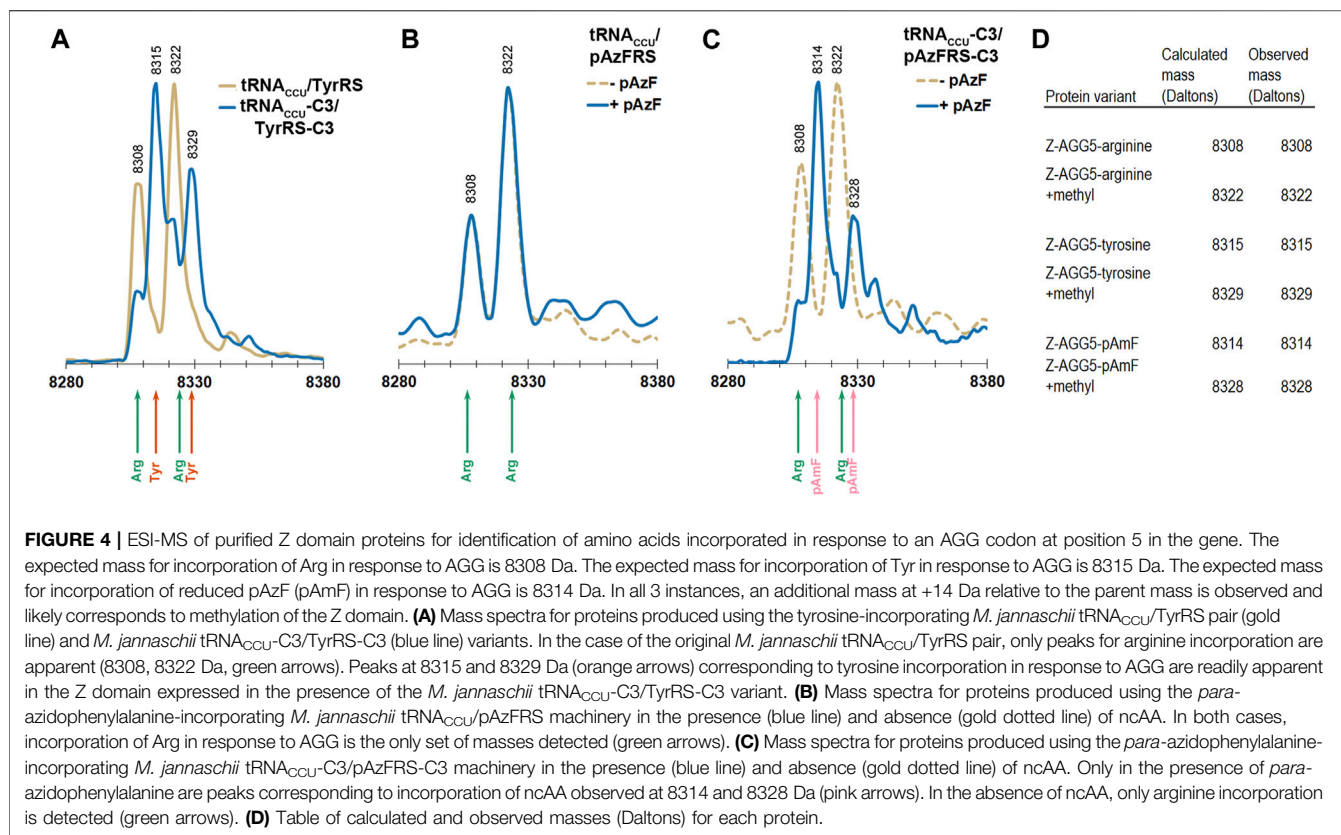
In order to use fluorescence to quantify ncAA incorporation efficiency, the differences in the photophysical properties of GFP with pAmF (reduced pAzF) in the fluorophore need to be included in the measurements. The 100% fluorescence control for reassignment to pAzF was the per protein fluorescence of pAzF incorporated in response to an amber stop codon in the fluorophore of GFP. Although the efficiencies of introduction of ncAAs in response to the amber stop codon using *M. jannaschii* tRNA/aaRS pair variants are typically in the 10–30% range relative to sense codons (Young et al., 2010), no protein is produced in the absence of ncAA. The protein isolated and subsequently quantified contains only the ncAA. The amount of protein was determined from SDS-PAGE gels using GFP standards to calibrate the visualization. The observed fluorescence was normalized to the protein amounts to derive a value for fluorescence per protein.

Based on fluorescence analysis of isolated proteins, *M. jannaschii* tRNA$_{CCU}$/pAzFRS incorporated pAzF in response to the AGG codon in the fluorophore of GFP with 29.5 ± 0.1% efficiency in rich media. The *M. jannaschii* tRNA$_{CCU}$-C3/pAzFRS-C3 orthogonal pair incorporated pAzF in response to the AGG codon with 50.1 ± 1.1% efficiency, a 1.7-fold improvement over the original *M. jannaschii* tRNA$_{CCU}$/pAzFRS (**Figure 3B**). This same relative improvement was observed between *M. jannaschii* tRNA$_{CCU}$/TyrRS (56.9 ± 2.4%) and *M. jannaschii* tRNA$_{CCU}$-C3/TyrRS-C3 (98.6 ± 4.7%) (**Figure 3A**). The improved reassignment efficiency

observed for the C3 variant is expected to be the result of increased effective concentration of aminoacylated tRNA due to improved recognition between the anticodon of the tRNA and the aaRS. In this instance, improvements made to the *M. jannaschii* tRNA/aaRS pair for incorporation one amino acid, Tyr, also increased the efficiency of incorporating another amino acid, pAzF, in response to AGG codons.

Electrospray ionization mass spectrometry (ESI-MS) was used to evaluate the levels of incorporation of arginine, tyrosine, and pAzF in response to a single AGG codon in the Z domain of protein A (Braisted and Wells, 1996). The Z domain is a small, soluble, 8.3 kDa, three-helix protein that has been employed as a reporter for stop codon suppression (Xie et al., 2007). The phenylalanine codon at position 5 in the native Z domain, a known permissive site, was mutated to AGG. The AGG codon containing Z domain variant was expressed in DH10B cells also expressing the *M. jannaschii* machinery to reassign AGG codons to tyrosine or pAzF (tRNA$_{CCU}$/TyrRS, tRNA$_{CCU}$-C3/TyrRS-C3, tRNA$_{CCU}$/pAzFRS, or tRNA$_{CCU}$-C3/pAzFRS-C3). Z domain proteins were isolated and analyzed using ESI-MS of the intact protein followed by deconvolution of the mass spectra using the Maximum Entropy algorithm (MassHunter Software, Agilent Technologies).

The calculated mass of the Z domain protein with arginine at the single AGG codon is 8308 Da, and the +14 Da peak observed for all parent peaks is likely due to a methylation of the Z domain (Stock et al., 1987; Apostol et al., 1995). The Z domain protein has been reported to be variously modified (Wang et al., 2003b; Zhang et al., 2003). Incorporation of tyrosine rather than arginine in response to the AGG codon results in a +7 Da shift. Incorporation of pAzF rather than arginine in response to the AGG codon results in a either a +6

FIGURE 4 | ESI-MS of purified Z domain proteins for identification of amino acids incorporated in response to an AGG codon at position 5 in the gene. The expected mass for incorporation of Arg in response to AGG is 8308 Da. The expected mass for incorporation of Tyr in response to AGG is 8315 Da. The expected mass for incorporation of reduced pAzF (pAmF) in response to AGG is 8314 Da. In all 3 instances, an additional mass at +14 Da relative to the parent mass is observed and likely corresponds to methylation of the Z domain. **(A)** Mass spectra for proteins produced using the tyrosine-incorporating *M. jannaschii* tRNA$_{CCU}$/TyrRS pair (gold line) and *M. jannaschii* tRNA$_{CCU}$-C3/TyrRS-C3 (blue line) variants. In the case of the original *M. jannaschii* tRNA$_{CCU}$/TyrRS pair, only peaks for arginine incorporation are apparent (8308, 8322 Da, green arrows). Peaks at 8315 and 8329 Da (orange arrows) corresponding to tyrosine incorporation in response to AGG are readily apparent in the Z domain expressed in the presence of the *M. jannaschii* tRNA$_{CCU}$-C3/TyrRS-C3 variant. **(B)** Mass spectra for proteins produced using the *para*-azidophenylalanine-incorporating *M. jannaschii* tRNA$_{CCU}$/pAzFRS machinery in the presence (blue line) and absence (gold dotted line) of ncAA. In both cases, incorporation of Arg in response to AGG is the only set of masses detected (green arrows). **(C)** Mass spectra for proteins produced using the *para*-azidophenylalanine-incorporating *M. jannaschii* tRNA$_{CCU}$-C3/pAzFRS-C3 machinery in the presence (blue line) and absence (gold dotted line) of ncAA. Only in the presence of *para*-azidophenylalanine are peaks corresponding to incorporation of ncAA observed at 8314 and 8328 Da (pink arrows). In the absence of ncAA, only arginine incorporation is detected (green arrows). **(D)** Table of calculated and observed masses (Daltons) for each protein.

or +32 Da shift (for the reduced amine or azide form of the amino acid, respectively).

Reassignment of the AGG codon to either tyrosine or pAzF was not detectable using the original forms of the *M. jannaschii* translation machinery where only the anticodon of the orthogonal tRNA was modified (tRNA$_{CCU}$/TyrRS or tRNA$_{CCU}$/pAzFRS). Masses corresponding to arginine incorporation were observed (**Figures 4A,B**). AGG reassignment is apparent in the deconvoluted mass spectra of cells expressing either tRNA$_{CCU}$-C3/TyrRS-C3 or tRNA$_{CCU}$-C3/pAzFRS-C3 (**Figures 4A,C**). In the case of protein isolated from cells expressing tRNA$_{CCU}$-C3/TyrRS-C3, the peaks corresponding to arginine incorporation (8308 and 8322 Da) shrink, and peaks corresponding to tyrosine incorporation (8315 and 8329 Da) become evident. A similar trend is observed for proteins isolated from cells expressing tRNA$_{CCU}$-C3/pAzFRS-C3 in the presence of pAzF.

The apparent difference in reassignment efficiencies in the Z domain protein as measured by ESI-MS and the GFP fluorescence-based screen is consistent with large observed variations in the efficiency of incorporation measured at different positions within the same protein and between various reporter proteins (Young et al., 2010). The efficiency of stop codon reassignment is thought to be sequence dependent based largely on differential interactions of release factors and suppressor tRNAs with regions outside the codon, codon context effects, additionally modulated by mRNA structure (Bossi and Roth 1980; Chevance et al., 2014; Gamble et al., 2016; Schwark

et al., 2018). The extent to which codon context affects the rate and efficiency of sense codon reading and reassignment has not been widely investigated. While the absolute reassignment efficiencies are decidedly different in the two reporter protein systems, the trends in both suggest that tRNA/aaRS variants selected from a library that improve reassignment of AGG codons to tyrosine are transferable and increase the reassignment of AGG codons to pAzF.

Increased Efficiencies Observed After Directed Evolution are Not Additive With Genomic Modification

The increase in sense codon reassignment efficiency that results from modifying interactions between the orthogonal tRNA/aaRS pair is expected to be independent from and potentially combinable with improvements that result from removing competition from endogenous tRNAs through genomic engineering. As near-quantitative reassignment of the AGG codon to tyrosine was observed using the *M. jannaschii* tRNA$_{CCU}$-C3/TyrRS-C3, genomic engineering was not expected to markedly improve this system. Our evaluation of the effect of host strain genomic engineering on sense codon reassignment began instead with *M. jannaschii* tRNA$_{CCU}$/TyrRS, which reassigns AGG to tyrosine with 56.9 ± 2.4% efficiency in *E. coli* DH10B.

One of the simplest genome modifications that may affect decoding the AGG codon is removal of the gene which encodes

the *E. coli* rare arginine tRNA$_{CCU}$, *argW*. The *argW* knockout of *E. coli* DH10B was prepared *via* the lambda red recombinase method (DS-dArgW) (Datsenko and Wanner, 2000). The strain mimics the previously-reported BS01 strain (DH10B Δ*arg*W) evaluated for reassignment of the AGG codon (Lee et al., 2015). An *argW/argA* double knockout named DS-dArgWArgA was also evaluated. The secondary *argA* knockout renders the cells auxotrophic for arginine. Sanger sequencing of the products of PCR amplifications of relevant segments of the chromosome and tRNA amplification tests indicated that the genomic knockouts were successful.

Eliminating competition for the AGG codon by removing the endogenous tRNA with a CCU anticodon resulted in a 1.6-fold *decrease* in measured reassignment efficiency. The *M. jannaschii* tRNA$_{CCU}$/TyrRS reassigned AGG to tyrosine with 37.3 ± 0.3% efficiency in *E. coli* DS-dArgWArgA (12 biological replicates), compared to 56.9 ± 2.4% in *E. coli* DH10B. The observation that knocking out the competing tRNA was detrimental, as opposed to neutral or even beneficial, to sense codon reassignment efficiency was surprising. Cells from the original report (BS02, DH10B Δ*arg*W Δ*arg*A) were obtained for comparison to the DS-dArgWArgA strain and yielded similar results with our orthogonal machinery (Lee et al., 2016).

The finding that expression of the minor *E. coli* tRNA$_{CCU}$ is not a major factor in allowing high level sense codon reassignment at the AGG codon is consistent with several other reports. Zeng et al. attempted to reduce the effective concentration of *E. coli* tRNA$_{CCU}$ by introducing an antisense RNA and found that it had a no effect on sense codon reassignment efficiency (Zeng et al., 2014). Mukai et al. reported cells that *E. coli* cells were viable after knocking out *argW*, but were unable to survive after knocking out the genes for both the *E. coli* tRNA$_{CCU}$ and *E. coli* tRNA$_{UCU}$, which is able to read both the Arg AGA and AGG codons (Mukai et al., 2015). Supplemental information in Lee et al. suggests that very high levels of AGG codon reassignment are present in *argW*[+] parent cells consistent with the AGG reassignment efficiencies measured by the GFP screen (Lee et al., 2016). Direct comparisons between these systems are difficult as the vectors employed to produce both the orthogonal machinery and reporter proteins differ significantly across the reports. Preliminary evaluation of combinations of the variously evolved versions of orthogonal machinery expressed from different plasmid vectors (e.g. origin of replication, antibiotic resistance marker, and the promoters driving expression of the aaRS and tRNA) in the wildtype and knockout cell lines utilizing both fluorescence and mass reporters produced confounding results. The general conclusion that can be drawn from these experiments is that the systemic context in which a set of improvements in orthogonal system performance is selected is important. Combinations of improvements selected under different conditions do not necessarily add in predictable ways.

CONCLUSION

We have described a general directed evolution pipeline for tailoring the tRNA/aaRS interactions of already-evolved ncAA-activating aaRSs for improved sense codon reassignment. Our pipeline exploits an easily screenable, tyrosine-activating aaRS as a stand in for ncAA-activating orthogonal aaRSs without readily screenable side chain chemistries. In the case of the ~100 ncAA activating *M. jannaschii* orthogonal aaRS variants, the stand in aaRS is the natural, parent, tyrosine-activating *M. jannaschii* aaRS. This workflow should be applicable to ncAA-activating variants of the *M. barkeri* pyrrolysine aaRS as well, as we recently reported the evolution of a highly efficient tyrosine-incorporating variant of that pair (Schwark et al., 2021). We demonstrate the utility of the pipeline in improving the efficiency of *para*-azidophenylalanine incorporation in response to the arginine AGG codon. We selected variants that could more efficiently translate the AGG codon placed at the fluorophore tyrosine position in sfGFP from a library of *M. jannaschii* Tyr tRNA and aaRSs with focused mutations in the tRNA anticodon loop and aaRS tRNA anticodon binding domain. Translation of the codon at position 66 in a GFP reporter as either tyrosine or the amino acid typically specified by the target codon results in fluorescence signals proportional to the extent of successful reassignment. Screening large combinatorial libraries for improved variants is accomplished quickly *via* FACS. We showed that mutations that led to improved AGG codon reassignment identified using a tyrosine-activating aaRS also improved the efficiency of pAzF incorporation in response to the arginine AGG codon.

We expect that this directed evolution workflow will be readily applicable to improving incorporation of ncAAs in response to other targeted sense codons exploiting orthogonal tRNA/aaRS pairs already-evolved to recognize ncAAs and target suppression of an amber stop codon. Although decreasing competition from the endogenous translation machinery *via* genomic alterations proved effective for improving reassignment of the Arg AGG codon in other reports, the genomic alteration strategy is not necessarily readily transferrable to other attractive sense codon targets. Given the idiosyncrasies of *in vivo* protein translation and of the orthogonal pairs transplanted into it, improvements for one amino acid may not result in the same magnitude of improvement for another amino acid each time. However, we expect this approach to serve as a reasonable starting point for rapid identification of positional mutants with the likelihood of improving tRNA/aaRS recognition across aaRSs evolved for many amino acids.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conceptualization: WB and JF; Funding acquisition: MS and JF; Investigation, WB, DS, and MS; Methodology, WB, DS, and MS; Supervision: JF; Visualization, WB and MS; Writing–original draft, WB; Writing–review and editing, MS and JF.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.815788/full#supplementary-material

REFERENCES

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

Apostol, I., Aitken, J., Levine, J., Lippincott, J., Davidson, J. S., and Abbott-Brown, D. (1995). Recombinant Protein Sequences Can Trigger Methylation of N-Terminal Amino Acids inEscherichia Coli. *Protein Sci.* 4 (12), 2616–2618. doi:10.1002/pro.5560041219

Bayley, H., Standring, D. N., and Knowles, J. R. (1978). Propane-1,3-dithiol: A Selective Reagent for the Efficient Reduction of Alkyl and Aryl Azides to Amines. *Tetrahedron Lett.* 19, 3633–3634. doi:10.1016/s0040-4039(01)95015-4

Biddle, W., Schmitt, M. A., and Fisk, J. D. (2015). Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen. *Biochemistry* 54 (50), 7355–7364. doi:10.1021/acs.biochem.5b00870

Biddle, W., Schmitt, M. A., and Fisk, J. D. (2016). Modification of Orthogonal tRNAs: Unexpected Consequences for Sense Codon Reassignment. *Nucleic Acids Res.* 44 (21), 10042–10050. doi:10.1093/nar/gkw948

Bohlke, N., and Budisa, N. (2014). Sense Codon Emancipation for Proteome-wide Incorporation of Noncanonical Amino Acids: Rare Isoleucine Codon AUA as a Target for Genetic Code Expansion. *FEMS Microbiol. Lett.* 351 (2), 133–144. doi:10.1111/1574-6968.12371

Bossi, L., and Roth, J. R. (1980). The Influence of Codon Context on Genetic Code Translation. *Nature* 286 (5769), 123–127. doi:10.1038/286123a0

Braisted, A. C., and Wells, J. A. (1996). Minimizing a Binding Domain from Protein A. *Proc. Natl. Acad. Sci.* 93 (12), 5688–5692. doi:10.1073/pnas.93.12.5688

Cervettini, D., Tang, S., Fried, S. D., Willis, J. C. W., Funke, L. F. H., Colwell, L. J., et al. (2020). Rapid Discovery and Evolution of Orthogonal Aminoacyl-tRNA Synthetase-tRNA Pairs. *Nat. Biotechnol.* 38 (8), 989–999. doi:10.1038/s41587-020-0479-2

Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013). A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia C. Biochemistry* 52 (10), 1828–1837. doi:10.1021/bi4000244

Chevance, F. F. V., Le Guyon, S., and Hughes, K. T. (2014). The Effects of Codon Context on *In Vivo* Translation Speed. *Plos Genet.* 10 (6), e1004392. doi:10.1371/journal.pgen.1004392

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 C. J. Am. Chem. Soc.* 124 (31), 9026–9027. doi:10.1021/ja027007w

Clackson, T., and Lowman, H. B. (2004). "Phage Display a Practical Approach," in *Practical Approach Series*. Editors T. Clackson and H. B. Lowman (Oxford: Oxford University Press).

Datsenko, K. A., and Wanner, B. L. (2000). One-Step Inactivation of Chromosomal Genes in *Escherichia C* K-12 Using PCR Products. *Proc. Natl. Acad. Sci.* 97 (12), 6640–6645. doi:10.1073/pnas.120163297

Ding, W., Zhao, H., Chen, Y., Zhang, B., Yang, Y., Zang, J., et al. (2020). Chimeric Design of Pyrrolysyl-tRNA Synthetase/tRNA Pairs and Canonical Synthetase/tRNA Pairs for Genetic Code Expansion. *Nat. Commun.* 11 (1), 3154. doi:10.1038/s41467-020-16898-y

Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015). Designing Logical Codon Reassignment - Expanding the Chemistry in Biology. *Chem. Sci.* 6 (1), 50–69. doi:10.1039/c4sc01534g

Fan, C., Xiong, H., Reynolds, N. M., and Söll, D. (2015). Rationally Evolving tRNA(Pyl) for Efficient Incorporation of Noncanonical Amino Acids. *Nucleic Acids Res.* 43 (22), e156. doi:10.1093/nar/gkv800

Fechter, P., Rudinger-Thirion, J., Tukalo, M., and Giege, R. (2001). Major Tyrosine Identity Determinants in Methanococcus Jannaschii and *Saccharomyces C* tRNA(Tyr) Conserved but Expressed Differently. *Eur. J. Biochem.* 268 (3), 761–767. doi:10.1046/j.1432-1327.2001.01931.x

Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., et al. (2019). Total Synthesis of *Escherichia coli* with a Recoded Genome. *Nature* 569 (7757), 514–518. doi:10.1038/s41586-019-1192-5

Furter, R. (1998). Expansion of the Genetic Code: Site-Directed P-Fluoro-Phenylalanine Incorporation in *Escherichia coli*. *Protein Sci.* 7 (2), 419–426. doi:10.1002/pro.5560070223

Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S., and Grayhack, E. J. (2016). Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell* 166 (3), 679–690. doi:10.1016/j.cell.2016.05.070

Gao, X., Yo, P., Keith, A., Ragan, T. J., and Harris, T. K. (2003). Thermodynamically Balanced Inside-Out (TBIO) PCR-Based Gene Synthesis: A Novel Method of Primer Design for High-Fidelity Assembly of Longer Gene Sequences. *Nucleic Acids Res.* 31 (22), e143. doi:10.1093/nar/gng143

Giege, R., Sissler, M., and Florentz, C. (1998). Universal Rules and Idiosyncratic Features in tRNA Identity. *Nucleic Acids Res.* 26 (22), 5017–5035. doi:10.1093/nar/26.22.5017

Guo, J., Melançon, C. E., Lee, H. S., Groff, D., and Schultz, P. G. (2009). Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angew. Chem. Int. Edition* 48 (48), 9148–9151. doi:10.1002/anie.200904035

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

Ho, J. M., Reynolds, N. M., Rivera, K., Connolly, M., Guo, L.-T., Ling, J., et al. (2016). Efficient Reassignment of a Frequent Serine Codon in Wild-Type *Escherichia coli*. *ACS Synth. Biol.* 5 (2), 163–171. doi:10.1021/acssynbio.5b00197

Johnson, D. B. F., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., et al. (2011). RF1 Knockout Allows Ribosomal Incorporation of Unnatural Amino Acids at Multiple Sites. *Nat. Chem. Biol.* 7 (11), 779–786. doi:10.1038/nchembio.657

Kajihara, D., Hohsaka, T., and Sisido, M. (2005). Synthesis and Sequence Optimization of GFP Mutants Containing Aromatic Non-natural Amino

Acids at the Tyr66 Position. *Protein Eng. Des. Selection* 18 (6), 273–278. doi:10.1093/protein/gzi033

Kirshenbaum, K., Carrico, I. S., and Tirrell, D. A. (2002). Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues. *ChemBioChem* 3 (2-3), 235–237. doi:10.1002/1439-7633(20020301)3:2/3<235:AID-CBIC235>3.0.CO;2-7

Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., et al. (2003). Structural Basis for Orthogonal tRNA Specificities of Tyrosyl-tRNA Synthetases for Genetic Code Expansion. *Nat. Struct. Mol. Biol.* 10 (6), 425–432. doi:10.1038/nsb934

Krishnakumar, R., Prat, L., Aerni, H.-R., Ling, J., Merryman, C., Glass, J. I., et al. (2013). Transfer RNA Misidentification Scrambles Sense Codon Recoding. *ChemBioChem* 14 (15), 1967–1972. doi:10.1002/cbic.201300444

Kuhn, S. M., Rubini, M., Fuhrmann, M., Theobald, I., and Skerra, A. (2010). Engineering of an Orthogonal Aminoacyl-tRNA Synthetase for Efficient Incorporation of the Non-natural Amino Acid O-Methyl-L-Tyrosine Using Fluorescence-Based Bacterial Cell Sorting. *J. Mol. Biol.* 404 (1), 70–87. doi:10.1016/j.jmb.2010.09.001

Kunkel, T. A. (1985). Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Proc. Natl. Acad. Sci.* 82 (2), 488–492. doi:10.1073/pnas.82.2.488

Kwon, I., and Choi, E. S. (2016). Forced Ambiguity of the Leucine Codons for Multiple-Site-Specific Incorporation of a Noncanonical Amino Acid. *PLoS One* 11 (3), e0152826. doi:10.1371/journal.pone.0152826

Kwon, I., Kirshenbaum, K., and Tirrell, D. A. (2003). Breaking the Degeneracy of the Genetic Code. *J. Am. Chem. Soc.* 125 (25), 7512–7513. doi:10.1021/ja0350076

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., et al. (2013). Genomically Recoded Organisms Expand Biological Functions. *Science* 342 (6156), 357–360. doi:10.1126/science.1241459

Lee, B. S., Shin, S., Jeon, J. Y., Jang, K.-S., Lee, B. Y., Choi, S., et al. (2015). Incorporation of Unnatural Amino Acids in Response to the AGG Codon. *ACS Chem. Biol.* 10 (7), 1648–1653. doi:10.1021/acschembio.5b00230

Lee, K. B., Hou, C. Y., Kim, C.-E., Kim, D.-M., Suga, H., and Kang, T. J. (2016). Genetic Code Expansion by Degeneracy Reprogramming of Arginyl Codons. *ChemBioChem* 17 (13), 1198–1201. doi:10.1002/cbic.201600111

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

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., et al. (2013). MODOMICS: A Database of RNA Modification Pathways-2013 Update. *Nucleic Acids Res.* 41 (D1), D262–D267. doi:10.1093/nar/gks1007

Maranhao, A. C., and Ellington, A. D. (2017). Evolving Orthogonal Suppressor tRNAs To Incorporate Modified Amino Acids. *ACS Synth. Biol.* 6 (1), 108–119. doi:10.1021/acssynbio.660014510.1021/acssynbio.6b00145

Mittelstaet, J., Konevega, A. L., and Rodnina, M. V. (2013). A Kinetic Safety Gate Controlling the Delivery of Unnatural Amino Acids to the Ribosome. *J. Am. Chem. Soc.* 135 (45), 17031–17038. doi:10.1021/ja407511q

Morris, J. L., Reddington, S. C., Murphy, D. M., Jones, D. D., Platts, J. A., and Tippmann, E. M. (2013). Aryl Azide Photochemistry in Defined Protein Environments. *Org. Lett.* 15 (4), 728–731. doi:10.1021/ol3028779

Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., et al. (2010). Codon Reassignment in the *Escherichia coli* Genetic Code. *Nucleic Acids Res.* 38 (22), 8188–8195. doi:10.1093/nar/gkq707

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

Neumann, H., Slusarczyk, A. L., and Chin, J. W. (2010). De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs. *J. Am. Chem. Soc.* 132 (7), 2142–2144. doi:10.1021/ja9068722

Ngo, J. T., and Tirrell, D. A. (2011). Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis. *Acc. Chem. Res.* 44 (9), 677–685. doi:10.1021/ar200144y

Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., et al. (2016). Design, Synthesis, and Testing toward a 57-Codon Genome. *Science* 353 (6301), 819–822. doi:10.1126/science.aaf3639

Owens, A. E., Grasso, K. T., Ziegler, C. A., and Fasan, R. (2017). Two-Tier Screening Platform for Directed Evolution of Aminoacyl-tRNA Synthetases with Enhanced Stop Codon Suppression Efficiency. *ChemBioChem* 18 (12), 1109–1116. doi:10.1002/cbic.201700039

Polycarpo, C. R., Herring, S., Bérubé, A., Wood, J. L., Söll, D., and Ambrogelly, A. (2006). Pyrrolysine Analogues as Substrates for Pyrrolysyl-tRNA Synthetase. *FEBS Lett.* 580 (28-29), 6695–6700. doi:10.1016/j.febslet.2006.11.028

Pott, M., Schmidt, M. J., and Summerer, D. (2014). Evolved Sequence Contexts for Highly Efficient Amber Suppression with Noncanonical Amino Acids. *ACS Chem. Biol.* 9 (12), 2815–2822. doi:10.1021/cb5006273

Rauch, B. J., Porter, J. J., Mehl, R. A., and Perona, J. J. (2016). Improved Incorporation of Noncanonical Amino Acids by an Engineered tRNA(Tyr) Suppressor. *Biochemistry* 55 (3), 618–628. doi:10.1021/acs.biochem.5b01185

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 (23), 5974–5977. doi:10.1002/anie.201301490

Ryu, Y., and Schultz, P. G. (2006). Efficient Incorporation of Unnatural Amino Acids into Proteins in *Escherichia coli. Nat. Methods* 3 (4), 263–265. doi:10.1038/nmeth864

Sambrook, J., and Russell, D. W. (2001). "Molecular Cloning : a Laboratory Manual," in *Cold Spring Harbor* (N.Y., Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

Schmitt, M. A., Biddle, W., and Fisk, J. D. (2018). Mapping the Plasticity of the *Escherichia coli* Genetic Code with Orthogonal Pair-Directed Sense Codon Reassignment. *Biochemistry* 57, 2762–2774. doi:10.1021/acs.biochem.8b00177

Schrader, J. M., Chapman, S. J., and Uhlenbeck, O. C. (2011). Tuning the Affinity of Aminoacyl-tRNA to Elongation Factor Tu for Optimal Decoding. *Proc. Natl. Acad. Sci.* 108 (13), 5215–5220. doi:10.1073/pnas.1102128108

Schwark, D., Schmitt, M., and Fisk, J. (2018). Dissecting the Contribution of Release Factor Interactions to Amber Stop Codon Reassignment Efficiencies of the Methanocaldococcus Jannaschii Orthogonal Pair. *Genes* 9 (11), 546. doi:10.3390/genes9110546

Schwark, D. G., Schmitt, M. A., Biddle, W., and Fisk, J. D. (2020). The Influence of Competing tRNA Abundance on Translation: Quantifying the Efficiency of Sense Codon Reassignment at Rarely Used Codons. *ChemBioChem* 21, 2274–2286. doi:10.1002/cbic.202000052

Schwark, D. G., Schmitt, M. A., and Fisk, J. D. (2021). Directed Evolution of the Methanosarcina Barkeri Pyrrolysyl tRNA/aminoacyl tRNA Synthetase Pair for Rapid Evaluation of Sense Codon Reassignment Potential. *Int. J. Mol. Sci.* 22 (2), 895. doi:10.3390/ijms22020895

Staros, J. V., Bayley, H., Standring, D. N., and Knowles, J. R. (1978). Reduction of Aryl Azides by Thiols: Implications for the Use of Photoaffinity Reagents. *Biochem. Biophysical Res. Commun.* 80 (3), 568–572. doi:10.1016/0006-291x(78)91606-6

Stock, A., Schaeffer, E., Koshland, D. E., and Stock, J. (1987). A Second Type of Protein Methylation Reaction in Bacterial Chemotaxis. *J. Biol. Chem.* 262 (17), 8011–8014. doi:10.1016/s0021-9258(18)47518-7

Takimoto, J. K., Adams, K. L., Xiang, Z., and Wang, L. (2009). Improving Orthogonal tRNA-Synthetase Recognition for Efficient Unnatural Amino Acid Incorporation and Application in Mammalian Cells. *Mol. Biosyst.* 5 (9), 931–934. doi:10.1039/b904228h

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 (Bba) - Proteins Proteomics* 1844 (6), 1059–1070. doi:10.1016/j.bbapap.2014.03.002

Wang, Y., and Tsao, M.-L. (2016). Reassigning Sense Codon AGA to Encode Noncanonical Amino Acids inEscherichia Coli. *ChemBioChem* 17 (23), 2234–2239. doi:10.1002/cbic.201600448

Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001). Expanding the Genetic Code of *Escherichia coli. Science* 292 (5516), 498–500. doi:10.1126/science.1060077

Wang, L., Xie, J., Deniz, A. A., and Schultz, P. G. (2003a). Unnatural Amino Acid Mutagenesis of Green Fluorescent Protein. *J. Org. Chem.* 68 (1), 174–176. doi:10.1021/jo026570u

Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003b). Addition of the Keto Functional Group to the Genetic Code of *Escherichia coli. Proc. Natl. Acad. Sci.* 100 (1), 56–61. doi:10.1073/pnas.0234824100

Wang, F., Niu, W., Guo, J., and Schultz, P. G. (2012). Unnatural Amino Acid Mutagenesis of Fluorescent Proteins. *Angew. Chem. Int. Ed.* 51 (40), 10132–10135. doi:10.1002/anie.201204668

Wang, N., Ju, T., Niu, W., and Guo, J. (2015). Fine-Tuning Interaction between Aminoacyl-tRNA Synthetase and tRNA for Efficient Synthesis of Proteins Containing Unnatural Amino Acids. *ACS Synth. Biol.* 4 (3), 207–212. doi:10.1021/sb500195w

Wang, K., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T., and Chin, J. W. (2016). Defining Synonymous Codon Compression Schemes by Genome Recoding. *Nature* 539 (7627), 59–64. doi:10.1038/nature20124

Willis, J. C. W., and Chin, J. W. (2018). Mutually Orthogonal Pyrrolysyl-tRNA Synthetase/tRNA Pairs. *Nat. Chem* 10 (8), 831–837. doi:10.1038/s41557-018-0052-5

Wiltschi, B., and Budisa, N. (2007). Natural History and Experimental Evolution of the Genetic Code. *Appl. Microbiol. Biotechnol.* 74 (4), 739–753. doi:10.1007/s00253-006-0823-6

Xie, J., Supekova, L., and Schultz, P. G. (2007). A Genetically Encoded Metabolically Stable Analogue of Phosphotyrosine in *Escherichia coli*. *ACS Chem. Biol.* 2 (7), 474–478. doi:10.1021/cb700083w

Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010). An Enhanced System for Unnatural Amino Acid Mutagenesis in *E. coli*. *J. Mol. Biol.* 395 (2), 361–374. doi:10.1016/j.jmb.2009.10.030

Zeng, Y., Wang, W., and Liu, W. R. (2014). Towards Reassigning the Rare AGG Codon inEscherichia Coli. *ChemBioChem* 15 (12), 1750–1754. doi:10.1002/cbic.201400075

Zhang, Z., Smith, B. A. C., Wang, L., Brock, A., Cho, C., and Schultz, P. G. (2003). A New Strategy for the Site-Specific Modification of Proteins *In Vivo*. *Biochemistry* 42 (22), 6735–6746. doi:10.1021/bi0300231

Zhao, H., Ding, W., Zang, J., Yang, Y., Liu, C., Hu, L., et al. (2021). Directed-Evolution of Translation System for Efficient Unnatural Amino Acids Incorporation and Generalizable Synthetic Auxotroph Construction. *Nat. Commun.* 12 (1), 7039. doi:10.1038/s41467-021-27399-x

Uncover New Reactivity of Genetically Encoded Alkyl Bromide Non-Canonical Amino Acids

Xin Shu[1,2†], Sana Asghar[1,2†], Fan Yang[3], Shang-Tong Li[4]*, Haifan Wu[5]* and Bing Yang[1,2]*

[1]Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China, [2]Cancer Center, Zhejiang University, Hangzhou, China, [3]Department of Biophysics, Kidney Disease Center of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, [4]Glbizzia Biosciences Co., Ltd, Beijing, China, [5]Department of Chemistry and Biochemistry, Wichita State University, Wichita, KS, United States

Genetically encoded non-canonical amino acids (ncAAs) with electrophilic moieties are excellent tools to investigate protein-protein interactions (PPIs) both *in vitro* and *in vivo*. These ncAAs, including a series of alkyl bromide-based ncAAs, mainly target cysteine residues to form protein-protein cross-links. Although some reactivities towards lysine and tyrosine residues have been reported, a comprehensive understanding of their reactivity towards a broad range of nucleophilic amino acids is lacking. Here we used a recently developed OpenUaa search engine to perform an in-depth analysis of mass spec data generated for Thioredoxin and its direct binding proteins cross-linked with an alkyl bromide-based ncAA, BprY. The analysis showed that, besides cysteine residues, BprY also targeted a broad range of nucleophilic amino acids. We validated this broad reactivity of BprY with Affibody/Z protein complex. We then successfully applied BprY to map a binding interface between SUMO2 and SUMO-interacting motifs (SIMs). BprY was further applied to probe SUMO2 interaction partners. We identified 264 SUMO2 binders, including several validated SUMO2 binders and many new binders. Our data demonstrated that BprY can be effectively used to probe protein-protein interaction interfaces even without cysteine residues, which will greatly expand the power of BprY in studying PPIs.

Keywords: protein-protein interactions, genetic code expansion, non-canonical amino acid, chemical cross-linking, and SUMO interactome

INTRODUCTION

Protein-protein interactions (PPIs) are essential for virtually all cellular processes in all living organisms. Thus, there is a significant effort in mapping protein-protein interaction networks to understand relevant biological processes in detail, and many techniques have been developed for this purpose (Low et al., 2021). Affinity purification mass spectrometry (AP-MS) have been successfully applied to map protein-protein interactomes in many organisms (Ho et al., 2002; Butland et al., 2005; Krogan et al., 2006; Gordon et al., 2020; Richards et al., 2021), although these methods cannot distinguish between direct binders and indirect binders of a protein of interest (POI) (Morris et al., 2014). Moreover, weak and transient interactions are typically not comprehensively detected under the conditions of AP-MS (Richards et al., 2021). Proximity labeling (PL) by introducing covalent labels to proteins proximal to a POI allows large scale analysis of protein-protein interactions with potential spatial and temporal resolutions in cells (Han et al., 2018; Roux et al., 2018). However, limited information is available to map the interaction interfaces to gain further structural understanding of these interactions.
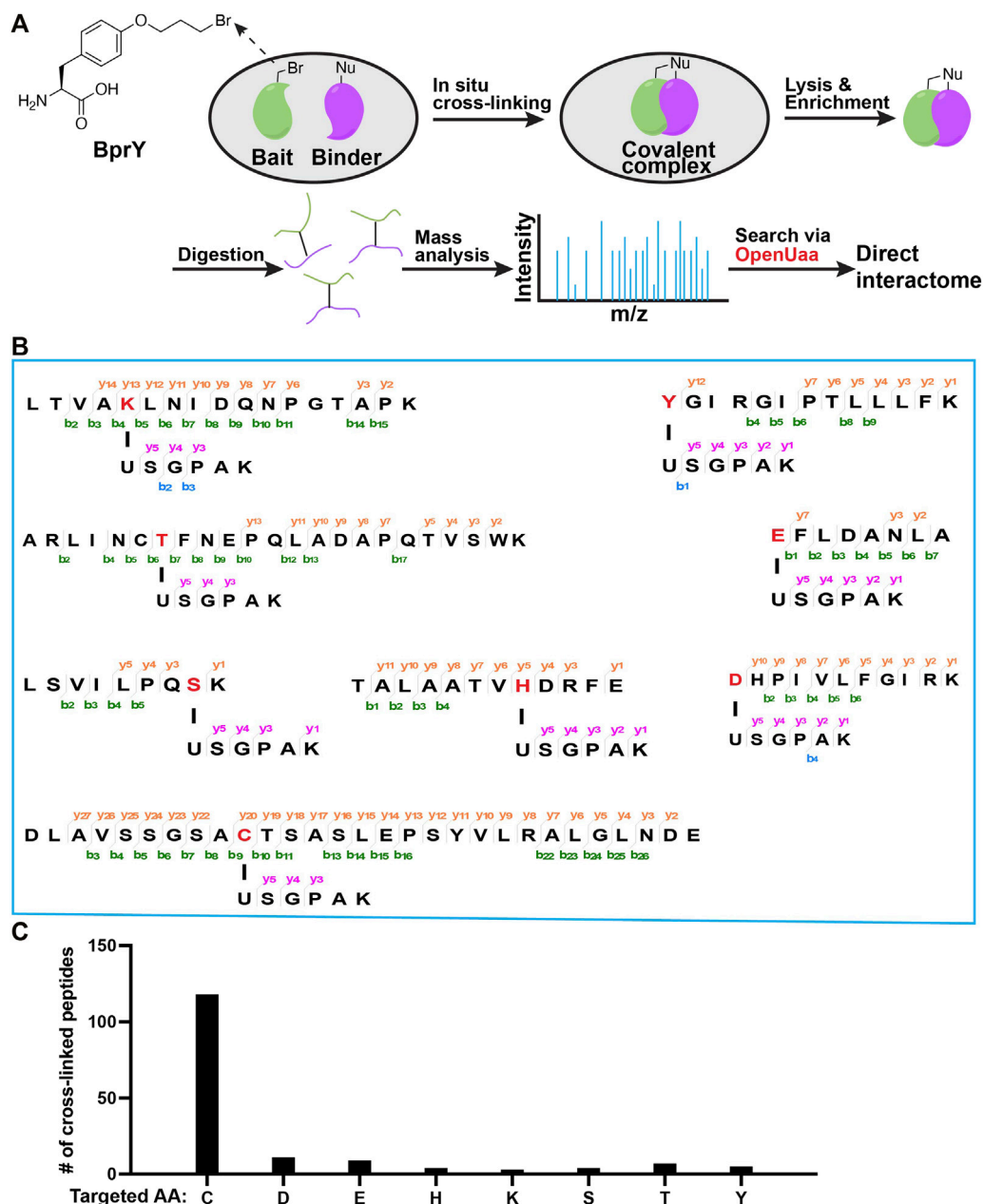
FIGURE 1 | Broad reactivity of BprY to nucleophilic amino acids identified by data mining. **(A)** Scheme of *in-situ* BprY cross-linking and in-depth data analysis by OpenUaa. **(B)** MS/MS fragmentation patterns of cross-linked peptides with different nucleophilic AAs targeted by BprY. **(C)** Number of cross-linked peptides identified for different targeted AAs.

A complementary method to AP-MS and PL for analyzing PPIs involves covalent cross-linking (Liu et al., 2015; Wang, 2017; Nguyen et al., 2018; Yu and Huang, 2018). Genetic code expansion by amber codon suppression has enabled site-specific incorporation of non-canonical amino acids (ncAAs) into proteins (Wang et al., 2001; Wang and Schultz, 2004; Wang et al., 2006; de la Torre and Chin, 2021; Shandell et al., 2021). Many ncAAs, including photo-activated ncAAs (Chin et al., 2002a; Chin et al., 2002b; Zhang et al., 2011; Lin et al., 2014;

Yang et al., 2016) and those with fine-tuned bio-reactivity to capture protein binders (Xiang et al., 2013; Chen X.-H. et al., 2014; Furman et al., 2014; Xiang et al., 2014; Xuan et al., 2016; Cigler et al., 2017; Wang, 2017; Nguyen et al., 2018; Shang et al., 2018; Wang et al., 2018; Yang et al., 2019; Liu et al., 2020; Liu J. et al., 2021), has been successfully incorporated into proteins. One example is a chemical cross-linking ncAA BprY with an electrophilic alkyl bromide group (**Figure 1A**), which is typically unreactive towards biomolecules in cells after incorporation into

a POI unless there is a proximal nucleophilic cysteine residue from the binder of this POI through the proximity-enabled reactivity. Given its ability to cross-link proximal cysteine residues, BprY has been successfully applied to capture proteome-wide protein-protein interactions in live cells (Yang et al., 2017), allowing the development of GECX-MS (Genetically Encoded Chemical Cross-linking of proteins coupled with Mass Spectrometry) as a powerful tool to identify direct binding partners of target proteins (**Figure 1A**). In GECX-MS, a cross-linkable ncAA, such as BprY, is genetically incorporated into a bait protein to covalently capture binder proteins including those weak and transient PPIs *in situ*. Analysis of cross-linked peptides by mass spectroscopy identifies not only binder proteins but also their corresponding cross-linking sites, which could be used to map binding interfaces or binding motifs.

One limitation of BprY and related alkyl halide-based ncAAs is that the cross-linking reaction requires proximal cysteine residues in binder proteins. Because cysteine residues are in relatively low abundance and they often form disulfide bonds, potentially binder proteins without proximal cysteine residues may not be captured. To maximize the potential of BprY to probe protein-protein interactions, its amino acid reactivity beyond cysteine is critical. Although there are some reports of extended reactivity of alkyl bromide based ncAAs towards glutamate, lysine and histidine (Chen X.-H. et al., 2014; Xiang et al., 2014; Cigler et al., 2017), a comprehensive analysis of BprY reactivity towards different nucleophilic amino acids under physiological conditions is not available. Here we show that BprY can target a broad range of nucleophilic amino acid residues, including Cys, Asp, Glu, Ser, Thr, His, and Tyr, among which the reactivity to Cys is still the highest. BprY can also be used to probe PPIs without cysteine at the binding interface. Finally, we successfully applied BprY to identify SUMO2-interacting proteins at a whole proteome level.

MATERIALS AND METHODS

Plasmid Construction

pBad-Affibody and mutants. The gene encoding Affibody was PCR amplified with Affibody-NdeI-F and Affibody-HindIII-R primers, and the PCR product was cloned into a commercial pBad vector pre-treated with NdeI and HindIII enzymes. Affibody-D2N mutant was generated by PCR amplification of pBad-Affibody with Affibody-D2N-F and Affibody-HindIII-R primers. Affibody mutants pBad-Afb*(K7X) were generated by PCR amplification of pBad-Affibody with Affibody-Mutant-F and Affibody-Mutant-R primers.

Affibody-NdeI-F:
GGAGATATACATATGGTAGACAACGCCTTCAAC
Affibody-HindIII-R:
AAAACAGCCAAGCTTTTAGTGATGGTGATGGTG
ATGA
Affibody-D2N-F:
GGAGATATACATATGGTAAACAACGCCTTCAACAAG
Affibody-Mutant-F:
AACAACGCCTTCAACCxxxCAACTATCAGTCGCC

Affibody-Mutant-R:
GGCGACTGATAGTTGxxxGTTGAAGGCGTTGTT
pBad-MBP-Z-24TAG. The gene encoding MBP-Z fusion protein was PCR amplified with the following primers, and the PCR product was cloned into the pBad vector pre-digested with NdeI and HindIII enzymes. pBad-MBP-Z-24TAG was generated by PCR amplification of pBad-MBP-Z with MBP-Z-24TAG-E25Q-Mutant primers.

MBP-Z-NdeI-F:
GGAGATATACATATGATGAAAATCGAAGAAGGT
AAACTG
MBP-Z-HindIII-R:
CAAAACAGCCAAGCTTTTAATGATGATGATGATGAT
GCTTAGG
MBP-Z-24TAG-E25Q-Mutant-F:
TTACCTAACCTGAATTAGCAGCAGCGTAATGCCTTC
MBP-Z-24TAG-E25Q -Mutant-R:
GAAGGCATTACGCTGCTGCTAATTCAGGTTAGGTAA
pBad-SUMO2 and mutants. The gene encoding *Homo sapiens* SUMO-2 (NCBI Reference Sequence: NM_006937.3) was PCR amplified with the following primers, and the PCR product was cloned into the pBad vector pre-treated with NdeI and HindIII enzymes. Sites of hSUMO2-E49 and R50 were mutated to a TAG codon respectively.

hSUMO2-NdeI-F:
GGAGATATACATATGATGGCCGACGAAAAGC
hSUMO2-HindIII-R:
AACAGCCAAGCTTTCAGTGATGGTGATGGTGATGGT
AGACACCTCCCGTCT
pET28a-MBP-RNF111$_{293-391}$. The gene encoding *Homo sapiens* RNF111$_{293-391}$ (NCBI Reference Sequence: NM_001270530.1) was PCR amplified with the following primers, and the PCR product was cloned into a pET28a-MBP vector.

pET28a-MBP-RNF111-F:
AAGTTCTGTTCCAGGGGCCCCATATGATGTCAGGAA
GTATTGATGAAGATGTTG
pET28a-MBP-RNF111-R:
CAGTGGTGGTGGTGGTGGTGCTCGAGTTCATCTTCA
TCAACGGTAAGGTC

Protein Expression

Affibody, SUMO2 and MBP-RNF111$_{293-391}$. The corresponding plasmids were individually transformed into DH10B cells, which were plated on LB agar plates supplemented with 100 μg/mL ampicillin. Colonies were picked from the plate of each plasmid and individually inoculated to 100 mL LB (5 g/L NaCl, 10 g/L Tryptone, 10 g/L Yeast extract). Cells were grown at 37°C and 200 rpm to an OD of 0.6 with good aeration and the relevant antibiotic selection. For induction of MBP-RNF111$_{293-391}$, 400 μM IPTG was added. 0.2% L-arabinose was added to induce the expression of Affibody and SUMO2. The expression was carried out at 30°C, 200 rpm for 5h. Cells were harvested by centrifugation at 6,000 *g*, 4°C for 10 min. The cell pellet was washed with cold PBS buffer and centrifuged again at 6,000 g, 4°C for 10 min. Cell pellets were then frozen in liquid nitrogen and stored at −80°C.

ncAA constructs: MBP-Z(E24BprY), SUMO2*(E49BprY), and SUMO2*(R50BprY). The corresponding plasmids were individually transformed into DH10B cells together with pEvol-BprY plasmid. Cells were plated on LB agar plates supplemented with 100 μg/mL ampicillin and 30 μg/mL chloramphenicol. Colonies were picked from these plates and individually inoculated to 25 mL LB (5 g/L NaCl, 10 g/L Tryptone, 10 g/L Yeast extract). Cells were grown at 37°C and 200 rpm to an OD of 0.4 with good aeration and the relevant antibiotic selection. Then the medium was added 1 mM ncAA BprY and 0.2% L-arabinose. The expression was carried out at 18°C and 200 rpm shaking for 18 h. Cells were harvested by centrifugation at 3,260 *g* and 4°C for 20 min. Cell pellets were washed with cold PBS buffer, centrifuged again at 3,260 g and 4°C for 20 min. Cell pellets were then frozen in liquid nitrogen and stored at −80°C.

His-Tag Protein Purification

Frozen cells were rapidly thawed and resuspended in lysis buffer (50mM Tris, pH8.0, 500 mM NaCl, 0.1% Tween-20). EDTA S3 free protease inhibitor cocktail was added followed by vortexing for 2 min. Cells were lysed by sonication after which the cell lysate was clarified by centrifugation at 13,000 rpm and 4°C for 30 min. The supernatant was collected and incubated with 200 μL Ni-NTA Affinity resin at 4°C for 1 h. The resin was washed with an equal volume of wash buffer (50 mM Tris pH8.0, 500 mM NaCl, 20 mM imidazole) for 2 times at 4°C. Elution was done with 200 μl elution buffer (50 mM Tris pH8.0, 100 mM NaCl, 250 mM imidazole) for five times. The fractions containing the target protein were determined by SDS-PAGE analysis with 10% Tricine gel.

Proteins Cross-Linking *in vitro*

A 20 μL reaction mixture containing MBP-Z(E24BprY) (0.6 mg/mL) and Affibody (1.2 mg/mL) in HEPES buffer (pH 7.5) was incubated at 37°C for 8 h. A 20 μL reaction mixture containing SUMO2 (10 μM) and RNF111 (293-391) (10 μM) in HEPES buffer (pH 7.5) was incubated at 37°C for 8 h.

Protein Digestion

Protein samples were precipitated by the addition of six volumes of cold acetone (−20°C) and incubated at −20°C for 30 min. Precipitated proteins were air dried and resuspended in 8 M urea, 100 mM Tris, pH 8.5. After reduction with 5 mM TCEP for 20 min and alkylation with 10 mM iodoacetamide for 15 min in the dark, samples were diluted to 2 M urea with 100 mM Tris, pH 8.5, and digested with trypsin (at 50:1 protein: enzyme ratio) at 37°C for 16 h. Digestion was terminated by adding formic acid to 5% final concentration, and digested peptides were desalted with StageTips.

Cell Lysate Cross-Linking and Two-step His-Tag Purification to Enrich Cross-Linked Peptides

Cell pellets expressing His-tagged SUMO2*(E49BprY) or SUMO2*(R50BprY) were resuspended in 4 ml lysis buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 0.1% Tween-20), separately. Cells were lysed by sonication after which the cell lysate was clarified by centrifugation at 13,000 rpm and 4°C for 30 min. The supernatants were collected and incubated with 200 μl Ni-NTA Affinity resin at 4°C for 2 h. The resin was washed with 4 ml of wash buffer (50 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole) for 3 times at 4°C, followed by a second rinsed with an equal volume of wash buffer 2 (50 mM Tris pH8.0, 500 mM NaCl, 40 mM imidazole) for 3 times at 4°C. Then the resin was equilibrated with lysis buffer and incubated with 293T cell lysates from one 10 cm plate at RT for overnight. The next day, the resin was washed with 4 mL wash buffer (50 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole) for 3 times at 4°C, and then rinsed with an equal volume of wash buffer 2 (50 mM Tris pH8.0, 500 mM NaCl, 40 mM imidazole) for 3 times at 4°C. Elution was done with 200 μl elution buffer (50 mM Tris pH 8.0, 100 mM NaCl, 250 mM imidazole) for five times. The eluates were concentrated, and buffer exchanged into 100 μl of protein storage buffer (50 mM HEPES, pH 7.5, and 100 mM NaCl) using 10 k Amicon Ultra columns. Purified proteins were digested with Lys-C at 37°C for overnight, and digested peptides were incubated with pre-equilibrated Ni-NTA Agarose resin (50 μL) at 4°C for 2 h to further enrich cross-linked peptides (all contain C-terminal His tag after Lys-C digestion). Resin was rinsed with wash buffer 2 for three times and 50 mM NH$_4$HCO$_3$ twice. Bound peptides were digested on-bead with Trypsin at 37°C for 8 h, and digested peptides were desalted with StageTips before MS analysis.

Tandem MS Analysis

MS experiments were performed on a Q Exactive HF-X instrument (ThermoFisher) coupled with an Easy-nLC 1200 system. Mobile phase A and B were water and 80% acetonitrile, respectively, with 0.1% formic acid. Digested peptides were loaded directly onto analytical column (75 μm × 20 cm, 1.9 μm C18, 5 μm tip) at a flow rate of 300 nL/min. All peptide samples were separated using a linear gradient of 6–22% B for 38 min, 22–35% B for 17 min, 35–90% B in 2 min, 90% B for 1 min, 100% B for 2 min. Survey scans of peptide precursors were performed from 350 to 1500 *m/z* at 60,000 FWHM resolution with a $1 × 10^6$ ion count target and a maximum injection time of 20 ms. The instrument was set to run in top-speed mode with 1-s cycles for the survey and the MS/MS scans. After a survey scan, tandem MS was then performed on the most abundant precursors exhibiting a charge state from 3 to 7 of greater than $1 × 10^5$ intensity by isolating them in the quadrupole at 1.6 m/z. Higher energy collisional dissociation (HCD) fragmentation was applied with 27% collision energy and resulting fragments detected in the Orbitrap detector at a resolution of 15,000. The maximum injection time limited was 30 ms, and dynamic exclusion was set to 30 s with a 10 ppm mass tolerance around the precursor.

MS Data Analysis

MS/MS spectra were extracted by parsing from RAW file. Datasets of model proteins were searched against the corresponding proteins by OpenUaa. OpenUaa was also used

to search data of two-step purified Trx sample and SUMO interaction protein sample against *E. coli* proteome and human proteome downloaded from the UniProt database and the reversed decoy proteins, separately. OpenUAA search parameters: 5% false discovery rate (FDR) at the peptide-spectrum match (PSM) level, 10 ppm precursor mass tolerance, 20 ppm fragment mass tolerance, variable modification Cys 57.02146, and three maximum number of missed cleavage sites.

Molecular Docking Study

For molecular docking, we first relaxed the structures of SUMO2 and RNF4 (PDB ID: 6JXX and 4PPE, respectively) in the Rosetta software suite (Leaver-Fay et al., 2011). 1000 models were generated for each protein and the model with the lowest total score was chosen for docking. In the protein-protein docking using Rosetta suite, RNF4 was first randomly oriented relative to SUMO2, and then 10,000 docking models were generated. The final docking model was chosen as the model with the largest binding energy (Rosetta energy term: I_sc). The docking model was rendered in the UCSF Chimera (Pettersen et al., 2004).

Bioinformatic Analysis

For the analysis of sequence motif around the cross-linked sites, MEME software was used to scan a ±15 amino acids sequence window around the cross-linked site to generate a consensus motif. Gene ontology (GO) term and KEGG pathway enrichment for functional analysis were performed using the clusterProfiler package under the R software (Yu et al., 2012). The human proteome was used as the background for enrichment analysis. The significance of the enrichment analysis was defined using a hypergeometric test, and the resulted *p* values were corrected for multiple hypothesis testing using the BH method (Hendriks and Vertegaal, 2016). The final reported enriched terms and pathways were filtered according to the adjusted *p* values <0.05. STRING network analysis was performed using the STRING database, which integrate multiple information sources for protein interaction speculation (Szklarczyk et al., 2019). Using all identified interacting proteins of SUMO2 as input, protein interactions between SUMO2 interacting proteins were filtered at a STRING interaction confidence of >0.4. Statistical analysis of the interaction network was performed based on the interactions in the networks compared to randomly expected frequency of interactions. Network visualization was performed using Cytoscape software version 3.7.2. The MCODE plugin of Cytoscape was used to extract the 8 interconnected modules under default settings for other parameters (Bader and Hogue, 2003).

RESULTS

Deep mining of interactome data identified a broad reactivity of BprY. We chose to investigate those nucleophilic amino acids known to carry out alkylation reactions (deGruyter et al., 2017). Nucleophilic substitution reactions with Asn/Gln/Arg sidechains are rare. Thus, we think they are unlikely to react with BprY. For Met, alkylation with benzyl bromides has been reported using model peptides under acidic conditions (Kramer and Deming, 2013). It is also known that iodoacetamide, a common alkylation reagent used in proteomics, can alkylate Met (Kruger et al., 2005). Therefore, we think it is possible that Met might also react with BprY. One concern is that the alkylation product sulfonium can undergo a dealkylation reaction (Kramer and Deming, 2013) and can dissociate upon collision to cause neutral mass loss (Kruger et al., 2005), which can further complicate the MS analysis. Due to these reasons, we decided to focus on Cys, Asp, Glu, His, Lys, Ser, Thr and Tyr in the current study. To systematically evaluate the reactivity of BprY to different types of amino acid residues in proteins, we re-analyzed previously published direct interactome data (Yang et al., 2017) of thioredoxin (Trx) probed by BprY (**Figure 1A**). Using a recently developed searching algorism OpenUaa (Liu C. et al., 2021), which allowed deeper mining of the interactome data, we observed a broad reactivity of BprY to multiple nucleophilic amino acids—Cys, Asp, Glu, His, Lys, Ser, Thr and Tyr, supported by high-quality MS/MS spectra (Figures 1B; **Supplementary Figure S1**). As expected, the reactivity is dominated by Cys with 118 Cys-targeted peptides identified. However, we also observed 43 non-Cys cross-linked peptides which still made a significant contribution to determining the direct interactome (**Figure 1C**).

Validate the reactivity using a model interacting protein pair. To validate this broad reactivity of BprY, we employed a model interacting protein pair—affibody (Afb) and protein Z fused to maltose-binding protein (MBP-Z) (**Figure 2A**). Efficient cross-linking was observed when a Cys residue and BprY were introduced to replace residue K7 of Afb and residue E24 of MBP-Z, respectively (**Supplementary Figure S2A**), consistent with a previous study (Yang et al., 2017). Several other AAs were also introduced at residue 7 of Afb, and the extent of cross-linking between Afb(K7X, X: variable AA) and MBP-Z(E24BprY) was evaluated. To our surprise, even some degree of cross-linking was observed between the "unreactive" control Afb(K7A) and MBP-Z(E24BprY) (**Supplementary Figure S2A**), suggesting nucleophilic amino acid residues adjacent to residue 7 in Afb can also facilitate cross-linking. We found that a triple mutation D2N, K4A, and E8Q of Afb completely eliminated cross-linking of the "unreactive" control (**Figure 2B**). Therefore, this triple mutation of Afb, denoted as Afb* thereafter, was further investigated. As shown in **Figure 2B**, Afb*(K7C) was efficiently cross-linked to MBP-Z(E24BprY), suggesting that the triple mutation didn't affect the binding of Afb to MBP-Z. Further cross-linking of MBP-Z with Afb*(K7X) showed that, compared to Afb*(K7C), several nucleophilic amino acids displayed reactivity to BprY, although the cross-linking efficiency as evaluated from the intensity of the cross-linking band is lower than that of Cys.

Apply BprY to study PPI without Cys at the interface in vitro. The study with model interacting protein pair further supported that BprY can target multiple nucleophilic amino acid residues beyond Cys, suggesting that BprY can be used to probe protein-protein interactions even without Cys residues at the interaction interface. To test this idea, we used BprY to capture the interaction between small ubiquitin-like modifiers (SUMOs)
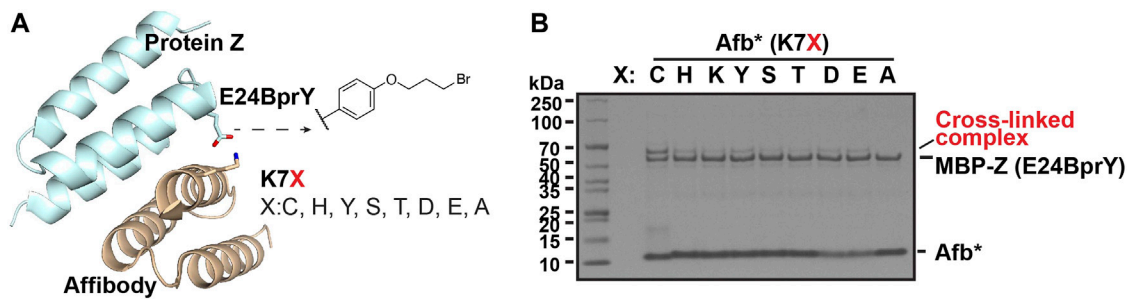
FIGURE 2 | Broad reactivity of BprY to nucleophilic AAs demonstrated using a model interacting protein pair. **(A)** The interaction between protein Z and affibody (PDB ID 1LP1) shows proximity between residue E24 in protein Z and residue K7 in affibody. **(B)** Cross-linking between MBP-Z(E24BprY) to Afb*(K7X). * denotes Afb with triple mutants D2N, K4A, and E8Q.
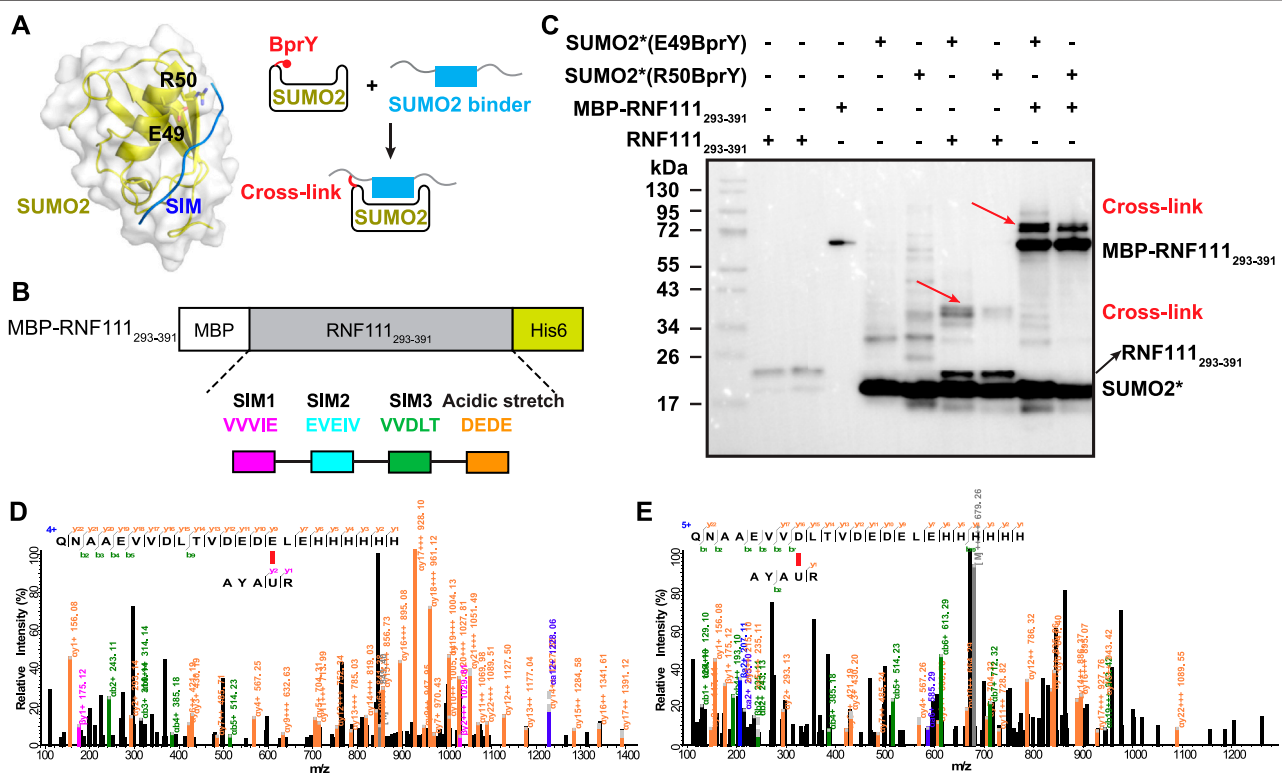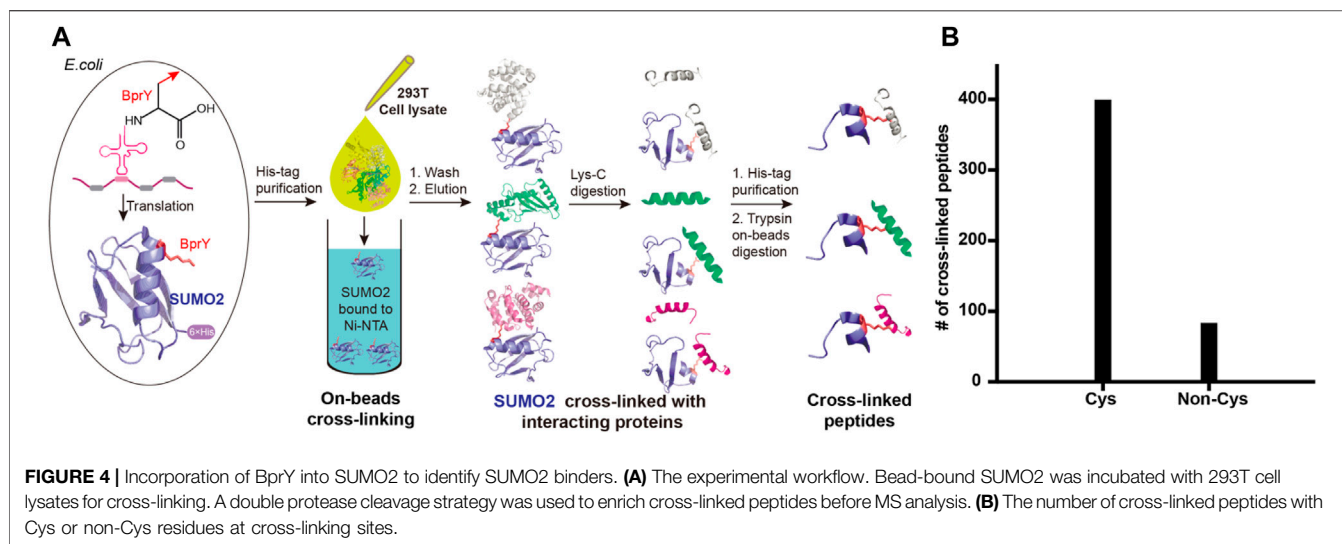

FIGURE 3 | BprY successfully probed a PPI interface without Cys. **(A)** The interaction between SUMO2 and SIM (PDB ID 2MP2) shows residues E49 and R50 next to the SIM-binding groove. A scheme illustrating the incorporation of BprY at residue 49 or 50 of SUMO2 to covalently capture SUMO2 binding proteins. **(B)** The construct of MBP-RNF111₂₉₃₋₃₉₁. **(C)** SDS-PAGE gel shows cross-linking of MBP-RNF111₂₉₃₋₃₉₁ and RNF111₂₉₃₋₃₉₁ with SUMO2* (E49BprY) or SUMO2* (R50BprY). Gel was stained by Coomassie brilliant blue. * denotes C48A mutation. Red arrows indicate cross-linking bands. **(D,E)** Representative MS/MS spectra of cross-linked peptides showing BprY in SUMO2 cross-linked to E391 and D384 of RNF111.

and one of their binding partners—RNF111. SUMOs can be reversibly conjugated to lysine side-chains of target proteins by an enzymatic cascade involving E1-E2-E3 enzymes, and this modification plays a key role in genome stability and transcription. The interaction between SUMOs and their binding partners is mediated primarily by SUMO-interacting motifs (SIMs) containing 3-4 aliphatic amino acid residues (**Figure 3A**) (Flotho and Melchior, 2013). RNF111 is a

SUMO-targeted ubiquitin ligase with three SIMs and an acidic stretch adjacent to SIM3 (Poulsen et al., 2013; Sriramachandran et al., 2019).

To probe the binding between SUMO2 and RNF111, we generated two SUMO2 constructs with BprY individually incorporated into SUMO2 at residue E49 or R50, adjacent to the SIM-binding groove (**Figure 3A**). In both constructs, residue C48 was mutated to Ala to prevent intra-protein cross-linking.

FIGURE 4 | Incorporation of BprY into SUMO2 to identify SUMO2 binders. **(A)** The experimental workflow. Bead-bound SUMO2 was incubated with 293T cell lysates for cross-linking. A double protease cleavage strategy was used to enrich cross-linked peptides before MS analysis. **(B)** The number of cross-linked peptides with Cys or non-Cys residues at cross-linking sites.
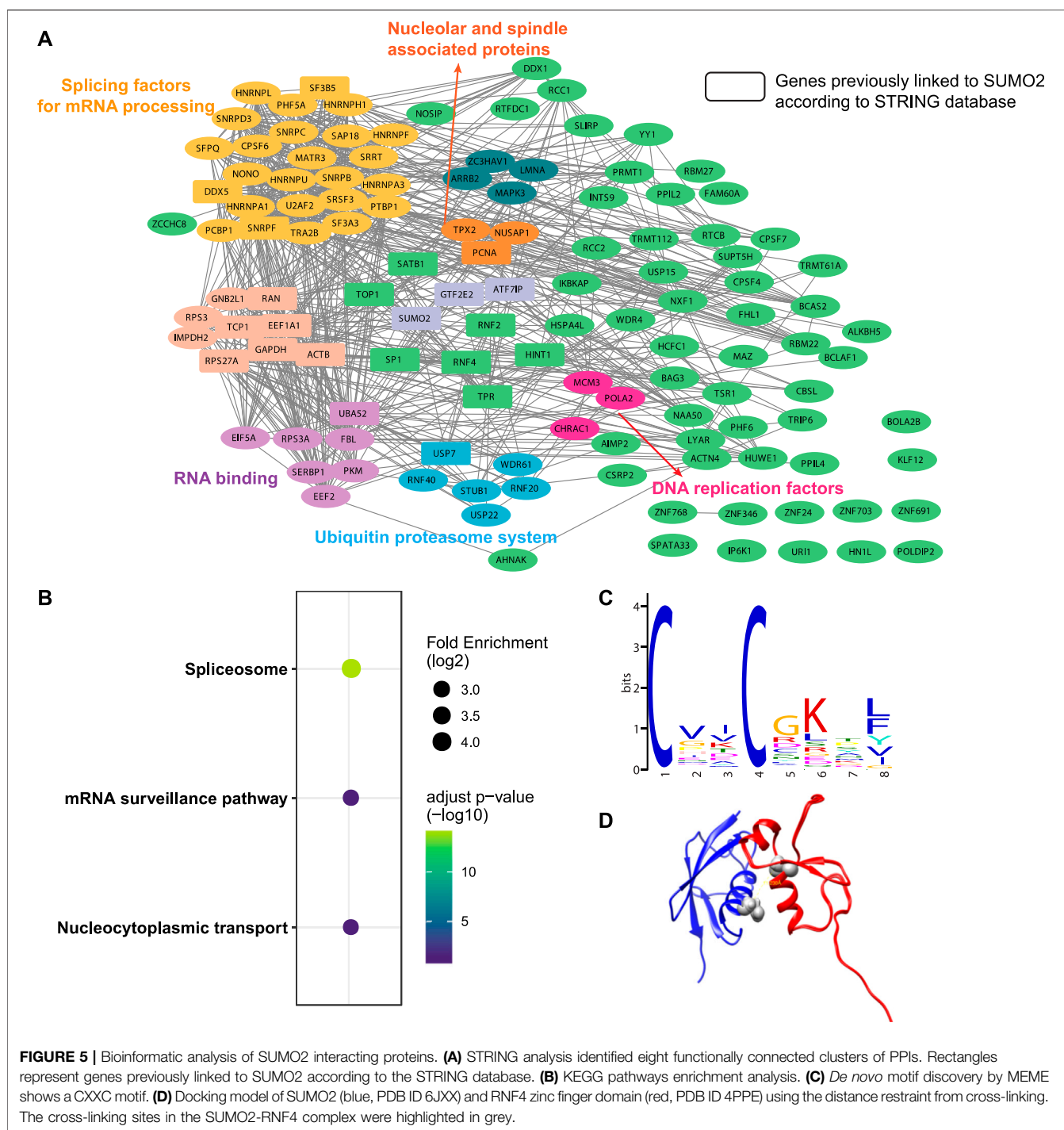
Mutation of these residues has been known to have a small effect on SUMO2 binding (Bouchenna et al., 2019; Bruninghoff et al., 2020). These two constructs SUMO2*(E49BprY) and SUMO2*(R50BprY) were then individually incubated with MBP-RNF111$_{293-391}$ (**Figure 3B**). As shown in **Figure 3C**, both SUMO2*(E49BprY) and SUMO2*(R50BprY) formed inter-protein cross-links with MBP-RNF111$_{293-391}$. The same result was also observed with RNF111$_{293-391}$, suggesting the cross-links are specific to RNF111. Incorporation of BprY at residue 49 of SUMO2* appeared to give more efficient cross-linking, suggesting that the side-chain of residue 49 might be better positioned for the cross-linking reaction. Another possibility is that mutation of R50 caused weaker binding to SIMs. Further mass analysis identified cross-links between BprY in SUMO2* to residues D384 in SIM3 and E391 in the acidic stretch of RNF111 (**Figures 3D,E**), supporting that this acidic stretch contributes to the binding of SIM3 in RNF111 to SUMO2.

Proteome-wide identification of SUMO2 interacting proteins by BprY. We next applied BprY to covalently capture SUMO2 interaction partners at the whole proteome level using 293T cell lysates (**Figure 4A**). After the expression of C-terminal His-tagged SUMO2*(E49BprY) and SUMO2*(R50BprY) in *E. coli* DH10B cells, both constructs were purified by Ni-NTA resin, and the resin-bound SUMO2 was incubated with 293T cell lysates to capture SUMO2 binders by forming cross-links. The resin was then stringently washed followed by elution to give SUMO2 and SUMO2 cross-linked to its interacting proteins. Because there is no lysine residue after K44 in SUMO2, Lys-C digestion followed by a second-step Ni-NTA purification can further enrich cross-linked peptides (**Figure 4A**). Final on-bead trypsin digestion was done before mass analysis.

A total of 482 cross-linked peptides were detected (399 Cys and 83 non-Cys, **Figure 4B**), corresponding to 264 SUMO2 binders among which 139 are proteins with nucleus localization. Cysteine residues only account for about 2% of the total residues in proteins and likely have low frequency to appear at binding interfaces compared with all the other amino acids investigated here. However, our SUMO interactome data showed that the number of identified cysteine-targeted cross-links is four times higher than that of non-cysteine cross-links. This is mainly due to higher intrinsic reactivity of cysteine sidechain to BprY. We believe that, for non-cysteine cross-links to occur efficiently, the interaction needs to be relatively strong, and the non-cysteine sidechain needs to be in an optimal geometry to react with BprY. Because SUMO2 is mainly localized in the nucleus, we first attempted to analyze these nuclear protein binders. Among them, we observed a previously reported SUMO2 binder RNF4 as well as many known SUMOylation substrates, including PCNA, SATB1, *etc* (**Figure 5A**). An analysis using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database showed that most of the SUMO2 binders were situated in a large network in which proteins are functionally/physically connected. At medium STRING confidence, 126 of the 139 proteins are in a single interaction network, with a PPI enrichment *p*-value < 1e-16, suggesting the identified SUMO2 binders are functional connected. Furthermore, we performed MCODE analysis on the SUMO2 binder network and identified 8 highly interconnected clusters within the core network (**Figure 5A**), including the spliceosome, ubiquitin proteasome system, DNA replication factors, and RNA processing factors. Interestingly, members (MCM3/CHRAC1/POLA2) of a cluster are all newly identified SUMO2 binders (Li et al., 2018; Dang and Morales, 2020). Since SUMOylation plays critical roles in DNA damage response (DDR) and numerous SUMO conjugates have been identified, the interaction between SUMO2 and MCM3/CHRAC1/POLA2 may establish a new link between SUMOylation mediated double strand breaks repair and maintenance of genome stability.

The KEGG pathway enrichment analysis highlighted three major pathways associated with these binders—spliceosome, mRNA surveillance pathway, and nucleocytoplasmic transport, consistent with some of the major functions of SUMOylation (**Figure 5B**). Interestingly, *de novo* motif discovery using MEME

FIGURE 5 | Bioinformatic analysis of SUMO2 interacting proteins. **(A)** STRING analysis identified eight functionally connected clusters of PPIs. Rectangles represent genes previously linked to SUMO2 according to the STRING database. **(B)** KEGG pathways enrichment analysis. **(C)** *De novo* motif discovery by MEME shows a CXXC motif. **(D)** Docking model of SUMO2 (blue, PDB ID 6JXX) and RNF4 zinc finger domain (red, PDB ID 4PPE) using the distance restraint from cross-linking. The cross-linking sites in the SUMO2-RNF4 complex were highlighted in grey.

based on sequences flanking the cross-linking site of SUMO2 binders revealed a CXXC motif (**Figure 5C**), which is commonly found in zinc finger proteins. Indeed, many identified proteins have zinc finger domains (ZNF24/346/691/703/768 and ZC3HAV1). This finding is consistent with previous reports suggesting that other than SIMs, zinc fingers can also bind SUMOs (Danielsen et al., 2012; Guzzo et al., 2014; Diehl et al., 2016). Besides, the enrichment of lysine in the motif suggests a close distance between cross-linked sites and SUMOylation sites.

One advantage of GECX-MS is the ability to identify cross-linking sites, which can be used to generate distance restraints for structural modeling. We focused on RNF4, a known binder of SUMO2, and attempted to reveal the potential conformation of the SUMO2-RNF4 complex with molecular docking. RNF4 has four SIMs and a zinc finger domain. Previous studies have identified SIM2 and SIM3 as the major contributor to SUMO binding (Kung et al., 2014; Xu et al., 2014). Interestingly, in this study, we found a cross-link between SUMO2 to the zinc finger domain of RNF4,

suggesting that this domain may also be involved in SUMO binding. By applying the distance restraint from cross-linking, we performed molecular docking of SUMO2 and RNF4 zinc finger domain using the Rosetta software suite (Leaver-Fay et al., 2011). The docking model (**Figure 5D**) revealed a binding interface on SUMO2 close to the SIM-binding groove. One thing to be noted is that because the BprY mediated cross-linking in this study was done in cell lysates with mixed cellular compartments, we also identified many non-nuclear proteins as SUMO2 binders. Gene ontology (GO) analysis suggested a potential link between SUMO2 to protein translation and localization (**Supplementary Figure S3**). Although there have been studies showing this link (Xu et al., 2010a; Xu et al., 2010b; Chen L.-z. et al., 2014), more detailed investigation will be needed in future studies.

CONCLUSION

In conclusion, we have demonstrated that the alkyl bromide containing ncAA BprY can react with not only cysteine residues but also a broad range of nucleophilic amino acids. Therefore, the application of BprY will not be limited to PPIs containing cysteine residues at the binding interface. Indeed, this aspect has been successfully demonstrated by our *in vitro* study of SUMO2/ RNF111 interaction in which there are no cysteine residues at the binding site. With this broad reactivity, we applied BprY to covalently capture and identify 264 SUMO2 interacting proteins at a whole proteome level. This study further demonstrated that BprY and the relevant alkyl halide ncAAs are excellent tools to study protein-protein interactions.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD031159 and 10.6019/PXD031159.

AUTHOR CONTRIBUTIONS

XS conducted experiments. SA and S-TL performed bioinformatic analysis. FY performed protein modeling. BY and HW conceived and directed the project. S-TL, BY, and HW wrote the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.815991/full#supplementary-material

REFERENCES

Bader, G. D., and Hogue, C. W. (2003). An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks. *BMC Bioinformatics* 4, 2. doi:10.1186/1471-2105-4-2

Bouchenna, J., Sénéchal, M., Drobecq, H., Stankovic-Valentin, N., Vicogne, J., and Melnyk, O. (2019). The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis. *Bioconjug. Chem.* 30 (10), 2684–2696. doi:10.1021/acs.bioconjchem.9b00598

Brüninghoff, K., Aust, A., Taupitz, K. F., Wulff, S., Dörner, W., and Mootz, H. D. (2020). Identification of SUMO Binding Proteins Enriched after Covalent Photo-Cross-Linking. *ACS Chem. Biol.* 15 (9), 2406–2414. doi:10.1021/acschembio.0c00609

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., et al. (2005). Interaction Network Containing Conserved and Essential Protein Complexes in Escherichia coli. *Nature* 433 (7025), 531–537. doi:10.1038/nature03239

Chen, L.-z., Li, X.-y., Huang, H., Xing, W., Guo, W., He, J., et al. (2014a). SUMO-2 Promotes mRNA Translation by Enhancing Interaction between eIF4E and eIF4G. *PLoS One* 9 (6), e100457. doi:10.1371/journal.pone.0100457

Chen, X.-H., Xiang, Z., Hu, Y. S., Lacey, V. K., Cang, H., and Wang, L. (2014b). Genetically Encoding an Electrophilic Amino Acid for Protein Stapling and Covalent Binding to Native Receptors. *ACS Chem. Biol.* 9 (9), 1956–1961. doi:10.1021/cb500453a

Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002a). Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia coli. *Proc. Natl. Acad. Sci.* 99 (17), 11020–11024. doi:10.1073/pnas.172226299

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002b). 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

Dang, T. T., and Morales, J. C. (2020). Involvement of POLA2 in Double Strand Break Repair and Genotoxic Stress. *Ijms* 21 (12), 4245. doi:10.3390/ijms21124245

Danielsen, J. R., Povlsen, L. K., Villumsen, B. H., Streicher, W., Nilsson, J., Wikström, M., et al. (2012). DNA Damage-Inducible SUMOylation of HERC2 Promotes RNF8 Binding via a Novel SUMO-Binding Zinc finger. *J. Cel Biol* 197 (2), 179–187. doi:10.1083/jcb.201106152

de la Torre, D., and Chin, J. W. (2021). Reprogramming the Genetic Code. *Nat. Rev. Genet.* 22 (3), 169–184. doi:10.1038/s41576-020-00307-7

deGruyter, J. N., Malins, L. R., and Baran, P. S. (2017). Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* 56 (30), 3863–3873. doi:10.1021/acs.biochem.7b00536

Diehl, C., Akke, M., Bekker-Jensen, S., Mailand, N., Streicher, W., and Wikström, M. (2016). Structural Analysis of a Complex between Small Ubiquitin-like Modifier 1 (SUMO1) and the ZZ Domain of CREB-Binding Protein (CBP/p300) Reveals a New Interaction Surface on SUMO. *J. Biol. Chem.* 291 (24), 12658–12672. doi:10.1074/jbc.M115.711325

Flotho, A., and Melchior, F. (2013). Sumoylation: a Regulatory Protein Modification in Health and Disease. *Annu. Rev. Biochem.* 82, 357–385. doi:10.1146/annurev-biochem-061909-093311

Furman, J. L., Kang, M., Choi, S., Cao, Y., Wold, E. D., Sun, S. B., et al. (2014). A Genetically Encoded Aza-Michael Acceptor for Covalent Cross-Linking of Protein-Receptor Complexes. *J. Am. Chem. Soc.* 136 (23), 8411–8417. doi:10.1021/ja502851h

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* 583 (7816), 459–468. doi:10.1038/s41586-020-2286-9

Guzzo, C. M., Ringel, A., Cox, E., Uzoma, I., Zhu, H., Blackshaw, S., et al. (2014). Characterization of the SUMO-Binding Activity of the Myeloproliferative and Mental Retardation (MYM)-type Zinc Fingers in ZNF261 and ZNF198. *PLoS One* 9 (8), e105271. doi:10.1371/journal.pone.0105271

Han, S., Li, J., and Ting, A. Y. (2018). Proximity Labeling: Spatially Resolved Proteomic Mapping for Neurobiology. *Curr. Opin. Neurobiol.* 50, 17–23. doi:10.1016/j.conb.2017.10.015

Hendriks, I. A., and Vertegaal, A. C. O. (2016). A High-Yield Double-Purification Proteomics Strategy for the Identification of SUMO Sites. *Nat. Protoc.* 11 (9), 1630–1649. doi:10.1038/nprot.2016.082

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S.-L., et al. (2002). Systematic Identification of Protein Complexes in Saccharomyces cerevisiae by Mass Spectrometry. *Nature* 415 (6868), 180–183. doi:10.1038/415180a

Kramer, J. R., and Deming, T. J. (2013). Reversible Chemoselective Tagging and Functionalization of Methionine Containing Peptides. *Chem. Commun.* 49 (45), 5144–5146. doi:10.1039/c3cc42214c

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., et al. (2006). Global Landscape of Protein Complexes in the Yeast Saccharomyces cerevisiae. *Nature* 440 (7084), 637–643. doi:10.1038/nature04670

Krüger, R., Hung, C.-W., Edelson-Averbukh, M., and Lehmann, W. D. (2005). Iodoacetamide-alkylated Methionine Can Mimic Neutral Loss of Phosphoric Acid from Phosphopeptides as Exemplified by Nano-Electrospray Ionization Quadrupole Time-Of-Flight Parent Ion Scanning. *Rapid Commun. Mass. Spectrom.* 19 (12), 1709–1716. doi:10.1002/rcm.1976

Kung, C. C.-H., Naik, M. T., Wang, S.-H., Shih, H.-M., Chang, C.-C., Lin, L.-Y., et al. (2014). Structural Analysis of Poly-SUMO Chain Recognition by the RNF4-SIMs Domain. *Biochem. J.* 462 (1), 53–65. doi:10.1042/BJ20140521

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

Li, M., Xu, X., Chang, C.-W., Zheng, L., Shen, B., and Liu, Y. (2018). SUMO2 Conjugation of PCNA Facilitates Chromatin Remodeling to Resolve Transcription-Replication Conflicts. *Nat. Commun.* 9 (1), 2706. doi:10.1038/s41467-018-05236-y

Lin, S., He, D., Long, T., Zhang, S., Meng, R., and Chen, P. R. (2014). Genetically Encoded Cleavable Protein Photo-Cross-Linker. *J. Am. Chem. Soc.* 136 (34), 11860–11863. doi:10.1021/ja504371w

Liu, C., Wu, T., Shu, X., Li, S. T., Wang, D. R., Wang, N., et al. (2021a). Identification of Protein Direct Interactome with Genetic Code Expansion and Search Engine OpenUaa. *Adv. Biol.* 5 (3), 2000308. doi:10.1002/adbi.202000308

Liu, F., Rijkers, D. T. S., Post, H., and Heck, A. J. R. (2015). Proteome-wide Profiling of Protein Assemblies by Cross-Linking Mass Spectrometry. *Nat. Methods* 12 (12), 1179–1184. doi:10.1038/nmeth.3603

Liu, J., Cao, L., Klauser, P. C., Cheng, R., Berdan, V. Y., Sun, W., et al. (2021b). A Genetically Encoded Fluorosulfonyloxybenzoyl-L-Lysine for Expansive

Covalent Bonding of Proteins via SuFEx Chemistry. *J. Am. Chem. Soc.* 143 (27), 10341–10351. doi:10.1021/jacs.1c04259

Liu, J., Cheng, R., Van Eps, N., Wang, N., Morizumi, T., Ou, W.-L., et al. (2020). Genetically Encoded Quinone Methides Enabling Rapid, Site-specific, and Photocontrolled Protein Modification with Amine Reagents. *J. Am. Chem. Soc.* 142 (40), 17057–17068. doi:10.1021/jacs.0c06820

Low, T. Y., Syafruddin, S. E., Mohtar, M. A., Vellaichamy, A., A Rahman, N. S., Pung, Y.-F., et al. (2021). Recent Progress in Mass Spectrometry-Based Strategies for Elucidating Protein-Protein Interactions. *Cell. Mol. Life Sci.* 78 (13), 5325–5339. doi:10.1007/s00018-021-03856-0

Morris, J. H., Knudsen, G. M., Verschueren, E., Johnson, J. R., Cimermancic, P., Greninger, A. L., et al. (2014). Affinity Purification-Mass Spectrometry and Network Analysis to Understand Protein-Protein Interactions. *Nat. Protoc.* 9 (11), 2539–2554. doi:10.1038/nprot.2014.164

Nguyen, T.-A., Cigler, M., and Lang, K. (2018). Expanding the Genetic Code to Study Protein-Protein Interactions. *Angew. Chem. Int. Ed.* 57 (44), 14350–14361. doi:10.1002/anie.201805869

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., et al. (2022). The PRIDE Database Resources in 2022: a Hub for Mass Spectrometry-Based Proteomics Evidences. *Nucleic Acids Res.* 50 (D1), D543–D552. doi:10.1093/nar/gkab1038

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25 (13), 1605–1612. doi:10.1002/jcc.20084

Poulsen, S. L., Hansen, R. K., Wagner, S. A., van Cuijk, L., van Belle, G. J., Streicher, W., et al. (2013). RNF111/Arkadia Is a SUMO-Targeted Ubiquitin Ligase that Facilitates the DNA Damage Response. *J. Cel Biol* 201 (6), 797–807. doi:10.1083/jcb.201212075

Richards, A. L., Eckhardt, M., and Krogan, N. J. (2021). Mass Spectrometry-based Protein-Protein Interaction Networks for the Study of Human Diseases. *Mol. Syst. Biol.* 17 (1), e8792. doi:10.15252/msb.20188792

Roux, K. J., Kim, D. I., Burke, B., and May, D. G. (2018). BioID: A Screen for Protein-Protein Interactions. *Curr. Protoc. Protein Sci.* 91, 19 23 11–19 23 15. doi:10.1002/cpps.51

Shandell, M. A., Tan, Z., and Cornish, V. W. (2021). Genetic Code Expansion: A Brief History and Perspective. *Biochemistry* 60, 3455–3469. doi:10.1021/acs.biochem.1c00286

Shang, X., Chen, Y., Wang, N., Niu, W., and Guo, J. (2018). Oxidation-induced Generation of a Mild Electrophile for Proximity-Enhanced Protein-Protein Crosslinking. *Chem. Commun.* 54 (33), 4172–4175. doi:10.1039/c8cc01639a

Sriramachandran, A. M., Meyer-Teschendorf, K., Pabst, S., Ulrich, H. D., Gehring, N. H., Hofmann, K., et al. (2019). Arkadia/RNF111 Is a SUMO-Targeted Ubiquitin Ligase with Preference for Substrates Marked with SUMO1-Capped SUMO2/3 Chain. *Nat. Commun.* 10 (1), 3678. doi:10.1038/s41467-019-11549-3

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi:10.1093/nar/gky1131

Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001). Expanding the Genetic Code of Escherichia coli. *Science* 292 (5516), 498–500. doi:10.1126/science.1060077

Wang, L. (2017). Genetically Encoding New Bioreactivity. *New Biotechnol.* 38 (Pt A), 16–25. doi:10.1016/j.nbt.2016.10.003

Wang, L., and Schultz, P. G. (2005). Expanding the Genetic Code. *Angew. Chem. Int. Ed.* 44 (1), 34–66. doi:10.1002/anie.200460627

Wang, L., Xie, J., and Schultz, P. G. (2006). Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* 35, 225–249. doi:10.1146/annurev.biophys.35.101105.121507

Wang, N., Yang, B., Fu, C., Zhu, H., Zheng, F., Kobayashi, T., et al. (2018). Genetically Encoding Fluorosulfate-L-Tyrosine to React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins In Vivo. *J. Am. Chem. Soc.* 140 (15), 4995–4999. doi:10.1021/jacs.8b01087

Xiang, Z., Lacey, V. K., Ren, H., Xu, J., Burban, D. J., Jennings, P. A., et al. (2014). Proximity-enabled Protein Crosslinking through Genetically Encoding Haloalkane Unnatural Amino Acids. *Angew. Chem. Int. Ed.* 53 (8), 2190–2193. doi:10.1002/anie.201308794

Xiang, Z., Ren, H., Hu, Y. S., Coin, I., Wei, J., Cang, H., et al. (2013). Adding an Unnatural Covalent Bond to Proteins through Proximity-Enhanced Bioreactivity. *Nat. Methods* 10 (9), 885–888. doi:10.1038/nmeth.2595

Xu, X., Vatsyayan, J., Gao, C., Bakkenist, C. J., and Hu, J. (2010a). HDAC2 Promotes eIF4E Sumoylation and Activates mRNA Translation Gene Specifically. *J. Biol. Chem.* 285 (24), 18139–18143. doi:10.1074/jbc.C110.131599

Xu, X., Vatsyayan, J., Gao, C., Bakkenist, C. J., and Hu, J. (2010b). Sumoylation of eIF4E Activates mRNA Translation. *EMBO Rep.* 11 (4), 299–304. doi:10.1038/embor.2010.18

Xu, Y., Plechanovová, A., Simpson, P., Marchant, J., Leidecker, O., Kraatz, S., et al. (2014). Structural Insight into SUMO Chain Recognition and Manipulation by the Ubiquitin Ligase RNF4. *Nat. Commun.* 5, 4217. doi:10.1038/ncomms5217

Xuan, W., Li, J., Luo, X., and Schultz, P. G. (2016). Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins. *Angew. Chem. Int. Ed.* 55 (34), 10065–10068. doi:10.1002/anie.201604891

Yang, B., Tang, S., Ma, C., Li, S.-T., Shao, G.-C., Dang, B., et al. (2017). Spontaneous and Specific Chemical Cross-Linking in Live Cells to Capture and Identify Protein Interactions. *Nat. Commun.* 8 (1), 2240. doi:10.1038/s41467-017-02409-z

Yang, B., Wang, N., Schnier, P. D., Zheng, F., Zhu, H., Polizzi, N. F., et al. (2019). Genetically Introducing Biochemically Reactive Amino Acids Dehydroalanine and Dehydrobutyrine in Proteins. *J. Am. Chem. Soc.* 141 (19), 7698–7703. doi:10.1021/jacs.9b02611

Yang, Y., Song, H., He, D., Zhang, S., Dai, S., Lin, S., et al. (2016). Genetically Encoded Protein Photocrosslinker with a Transferable Mass Spectrometry-Identifiable Label. *Nat. Commun.* 7, 12299. doi:10.1038/ncomms12299

Yu, C., and Huang, L. (2018). Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* 90 (1), 144–165. doi:10.1021/acs.analchem.7b04431

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., et al. (2011). A Genetically Incorporated Crosslinker Reveals Chaperone Cooperation in Acid Resistance. *Nat. Chem. Biol.* 7 (10), 671–677. doi:10.1038/nchembio.644

Creating a Selective Nanobody Against 3-Nitrotyrosine Containing Proteins

Elise M. Van Fossen[1], Sonia Grutzius[1], Carl E. Ruby[2], Dan V. Mourich[2], Chris Cebra[2], Shay Bracha[3], P. Andrew Karplus[1], Richard B. Cooley[1] and Ryan A. Mehl[1]*

[1]Oregon State University, Department of Biochemistry and Biophysics, Agricultural and Life Sciences, Corvallis, OR, United States, [2]Oregon State University, Department of Clinical Sciences, College of Veterinary Medicine, Corvallis, OR, United States, [3]Department of Small Animal Clinical Sciences (VSCS), Texas A&M College of Veterinary Medicine and Biomedical Sciences, College Station, TX, United States

A critical step in developing therapeutics for oxidative stress-related pathologies is the ability to determine which specific modified protein species are innocuous by-products of pathology and which are causative agents. To achieve this goal, technologies are needed that can identify, characterize and quantify oxidative post translational modifications (oxPTMs). Nanobodies (Nbs) represent exquisite tools for intracellular tracking of molecules due to their small size, stability and engineerability. Here, we demonstrate that it is possible to develop a selective Nb against an oxPTM protein, with the key advance being the use of genetic code expansion (GCE) to provide an efficient source of the large quantities of high-quality, homogenous and site-specific oxPTM-containing protein needed for the Nb selection process. In this proof-of-concept study, we produce a Nb selective for a 3-nitrotyrosine (nitroTyr) modified form of the 14-3-3 signaling protein with a lesser recognition of nitroTyr in other protein contexts. This advance opens the door to the GCE-facilitated development of other anti-PTM Nbs.

Keywords: genetic code expansion, nanobody, oxidative post translational modification, nitrotyrosine, single domain antibody

INTRODUCTION

Oxidative post-translational modifications (oxPTMs) are formed by small molecule oxidants reacting with proteins under both normal and oxidative stress conditions and their heterogeneity in terms of the types, locations and the extents of the modification make it challenging to pin-point the effects they exert. Regardless, the study of oxPTMs remains critical as these modifications are often identified in disease pathology and show promise as disease prediction tools (Tomin et al., 2019).

The oxPTM 3-nitrotyrosine (nitroTyr, nY) is produced by the formation of peroxynitrite-derived radicals and their subsequent reaction with tyrosine side chains (Souza et al., 2008). Its accumulation in over 100 distinct proteins is associated with numerous disease pathologies (Pacher et al., 2007). NitroTyr-modified proteins have been shown to cause functional changes that can contribute to disease (Franco et al., 2013; Bartesaghi and Radi 2018; Ferrer-Sueta et al., 2018; Radi, 2018; Porter et al., 2020), but this has been possible in only a few cases, as it is challenging to define which specific nitroTyr protein species are innocuous by-products of pathology and which are causative agents. As such knowledge is required for the development of effective therapeutics for oxidative stress-related pathologies, there is a critical need for tools to identify, characterize and quantify nitroTyr-modified proteins.

One such tool has been anti-nitroTyr antibodies (nitroTyr-Abs), which have enabled immune-localization of nitroTyr proteins (Moller et al., 2019) and nitroTyr-specific proteomic profiling (Herce-Pagliai et al., 1998). Most commonly employed is a nitroTyr-Ab specific to the PTM itself (Beckmann et al., 1994), though protein and site-specific nitroTyr-Abs have also been developed (Khan et al., 2017). Despite the utility of nitroTyr-Abs for the identification and enrichment of nitrated proteins, their large size has prevented their use for tracking proteins in live cells, which would allow nitroTyr-induced changes in client binding and subcellular location to be monitored.

Recently, nanobodies (Nbs)—single-domain antibody fragments derived from the heavy-chain of immunoglobulins of camelids and much smaller than Abs (~15 kDa vs.–150 kDa)—have been shown to be effective for the intracellular tracking of proteins (Moeglin et al., 2021). Additionally, Nbs are well suited for binding epitopes that are inaccessible to traditional Abs due to their protruding, convex paratopes (Muyldermans 2013; Pardon et al., 2014). The ability to generate nitroTyr-protein specific Nbs would allow the full suite of Nb capabilities to be harnessed for both *in vitro* and in cell work, including for instance, the modulation of protein activity by targeting nitroTyr proteins with a covalent nanobody (or "GlueBody") for degradation in order to observe the downstream effects on redox signaling (Bery et al., 2019; Cheloha et al., 2020; de Beer and Giepmans 2020; Zhang et al., 2021).

Nb selections require high-quality, homogenous target protein (Pardon et al., 2014), and the inability to make such protein has been a general barrier to creating Nbs against oxPTMs. Traditionally, oxPTMs are chemically introduced into proteins but this process generates multiple types of chemical modifications with minimal ability to control site-specificity (Ischiropoulos et al., 1992; Neumann et al., 2008; Gerding et al., 2019). Indeed, to our knowledge, no Nb has yet been developed against any oxPTM and only a single example has been published for any type of PTM (Moeglin et al., 2021). An avenue to bypass this barrier for PTMs is genetic code expansion (GCE), which can install a variety of PTMs into proteins to generate large quantities of homogenous, site-specific PTM-containing proteins (Neumann et al., 2008; Franco et al., 2013; Cooley et al., 2014a; Cooley et al., 2014b; DiDonato et al., 2014; Franco et al., 2015; Porter and Mehl 2018; Porter et al., 2019; Randall et al., 2019; Beyer et al., 2020; Jang et al., 2020; Porter et al., 2020). In GCE, non-canonical amino acids (ncAAs) are site-specifically incorporated into stop codons *via* orthogonal amino-acyl tRNA synthetase (aaRS)/tRNA$_{UAG}$ pairs. As GCE tools already exist to incorporate nitroTyr, we sought here to use GCE-produced nitroTyr-modified proteins as proof of concept to show how Nbs can be developed against oxPTM targets.

As trial nitrated targets, we selected two hub proteins involved in cell signaling pathways: 14-3-3 (Pennington et al., 2018) and calmodulin [CaM; (Smallwood et al., 2003)]. Both govern critical processes in biology and contain multiple biologically relevant sites of tyrosine nitration (Ghesquiere et al., 2009; Nuriel et al., 2015; Zhao et al., 2017; Porter et al., 2020). After generating a nitroTyr-Nb library from an immunized camelid, we performed selections with the most promising candidate protein target and

obtained a single Nb with a reasonable level of selectivity for certain nitrated protein targets, even in highly proteinaceous solutions. Furthermore, as a step toward developing covalent Nb binders, we incorporate a photocrosslinking ncAA at multiple positions in this anti-nitroTyr Nb. Though these positions did not result in high yielding crosslinking, we showcase the versatility of GCE-technology in Nb development, from Nb-selections to Nb engineering.

MATERIALS AND METHODS

Non-Canonical Amino Acids

3-nitro-tyrosine (nitroTyr, nY) and 4-azido-phenylalanine (azidoPhe, pAzF) were purchased from Alfa Aesar (Cat. no. A11018) and Chem impex (Cat. no. 06162). NcAA solutions were prepared fresh before each use by suspending the amino acids in water and solubilizing with 1–2 M equivalents of NaOH for a final concentration of 100 mM ncAA.

Molecular Cloning

Molecular cloning of plasmids used in this study (**Supplementary Table S1**) were constructed in the following manner. Previously published sequences containing at least 25 bp of homology at their flanking ends to either the vector backbone or other fragments, were optimized for expression in *E. coli* and synthesized by Integrated DNA Technologies (Coralville, IA, USA, **Supplementary Table S2**). Gene fragments, as well as fragments containing amber stop codons were amplified with primers listed in **Supplementary Table S3** using touchdown PCR (Green and Sambrook 2018) and the resulting products were separated on 0.8–1.2% (w/v) agarose gels and purified using Gene Jet Gel Extraction Kit (ThermoFisher Scientific, Waltham, MA, United States) according to the manufacturer's instructions. Vector backbones were prepared through restriction digestion and purified through gel extraction. Fragments and vector backbones were then ligated using the SLiCE cloning protocol (Zhang et al., 2012) transformed into chemically-competent DH10B *E. coli* cells and selected on LB agar plates containing the appropriate antibiotic. Colonies were selected and propagated prior to purification. Genetic sequences of each plasmid were confirmed using Sanger sequencing. The primers and template used for PCR amplification, as well as the vector backbone and restriction enzymes used for linearization are summarized for each plasmid in **Supplementary Table S1**. Plasmids not prepared in this study were prepared previously (see references in **Supplementary Table S1**).

Protein Expression

For the expression of nitroTyr and azidoPhe proteins, each TAG site containing expression plasmid was co-transformed with a GCE-machinery plasmid (see **Supplementary Table S1**) into BL21ai cells. All pDule2 machinery plasmids contain a p15a origin of replication and constitutively express the indicated amber suppressing *Methanocaldococcus jannaschii* (*Mj*) aminoacyl tRNA synthetase (aaRS)/tRNA$_{CUA}$ pair. Cultures (5 ml) were inoculated by scraping a swath of cells containing

the appropriate plasmids from a fresh LB agar plate and were grown overnight in 2XYT (no more than 16 h) in the presence of appropriate antibiotics. Overnight cultures were then used to inoculate expression cultures of ZY-AIM media [0.1–1.0 L, (Studier 2005)], supplemented with appropriate antibiotics. For the expression of nitroTyr and azidoPhe proteins, the media was also supplemented with nitroTyr and azidoPhe (at final concentration of 1 mM, diluted from a freshly prepared 100 mM solution as described above). Cultures were grown with constant shaking at 275 rpm in baffled flasks at 37°C until they reached an OD_{600} of 1.5, wherein they were induced with a final concentration of 0.1% arabinose. Upon induction, the temperature was decreased to 25°C and cultures expressed for an additional 30 h prior to cell harvesting.

Protein Purification

Cell pellets were resuspended in a lysis/wash buffer (50 mM Tris, 500 mM NaCl, 5 mM imidazole, pH 7.5) and lysed using a M-110P microfluidizer system (Microfluidics Corp, Newton, MA, United States) set at 18,000 psi. Cell debris was pelleted at 20,000 rcf for 25 min at 4°C and clarified cell lysate was recovered. To bind His$_6$-tagged protein, clarified cell lysate was then incubated with TALON resin (Takara Bio, Japan) at 4°C for 1 h with rocking. Resin was collected and extensively washed with 50 resin bed volumes (bv) of lysis buffer. At this point, for proteins containing an N-terminal His-*bd*NEDD8 tag (Frey and Gorlich 2014), the protein-bound resin was resuspended in 3 bvs of storage buffer (50 mM Tris, 500 mM NaCl, 10% glycerol). His-tag free *bd*NEDP1 (a kind gift from Dirk Gorlich, Addgene #104129, 100 nM) was added to the resuspension and the solution was incubated for 1.5 h at room temperature with rocking. The resin was retained in a column and cleaved nanobody was collected in the flow-through.

For all other constructs, bound protein was eluted from the TALON resin by incubation with five bv's of Elution Buffer (lysis buffer with 200 mM imidazole) and desalted into storage buffer using a PD-10 desalting column (GE Health Sciences). When cleavage of a His$_6$-SUMO tag was required, Ubiquitin-Like Protease 1 (ULP1) was added to the desalted solution and was incubated overnight at 4°C. Following incubation, cleaved protein was re-flowed over TALON resin to bind the cleaved tag, and the flow-through containing the Nb was collected.

Nb-G5 was further purified by size exclusion chromatography using a Superdex 200 10/300 column in storage buffer. When necessary, the protein solution was concentrated by using a 3 kDa MWCO Vivaspin spin-concentration filter (GE Health Sciences). Protein concentration was determined by absorbance at 280 nm using primary sequence calculated extinction coefficients, flash-frozen in liquid nitrogen and stored at −80°C.

Mass Spectrometry of Nitrated Proteins

Purified protein was exchanged into LC-MS grade water or 50 mM tri-ethyl-ammonium bicarbonate with PD-10 columns (Cytiva) diluted to 50 μM and analyzed with the Waters Synapt G2 Mass Spectrometer at the Mass Spectrometry Facility at Oregon State University. The deconvoluted masses were obtained by using Waters MassLynx MaxEnt1 software.

Expression and Purification of Biotinylated Proteins for Biolayer Interferometry

Biotinylated 14-3-3 proteins were expressed in BL21ai by co-transforming pBAD-AVI-14-3-3-His plasmids (**Supplementary Table S1**) with the GCE machinery plasmid pDule2-3-nitroTyr-A7 (Addgene #174079) and the pEVF-GST-BirA plasmid. This plasmid expresses the fusion protein GST-BirA in a vector compatible with GCE machinery plasmids and most common expression plasmids (RSF origin, CmR antibiotic resistance). pBAD-AVI-eGFP-His was also co-transformed with pEVF-GST-BirA to generate biotinylated eGFP as a control for Biolayer Interferometry (BLI) experiments.

Expression of biotinylated 14-3-3ß proceeded as described above except with the addition of 50 μM biotin and 25 μg/μL chloramphenicol to expression cultures, and at an OD of 1.5, 0.1 mM IPTG was added to induce the expression of GST-BirA. The purification of biotinylated 14-3-3ß was identical to non-biotinylated 14-3-3ß and biotinylation of Nb targets was validated by streptavidin motility-shift assay (Fairhead and Howarth 2015).

Peptide Design for Library Generation

Peptides were designed to encompass nitrated portions of calmodulin (nY99 and nY138) and 14-3-3 (nY133). The peptide for 14-3-3- nY133 BMHI shared a sequence identity of 61.9% with 14-3-3 nitroTyr human isoform b. A peptide for general nitration was also designed, containing multiple nitroTyr flanked by β-sheets. Peptides were designed by following the described parameters for Nb generation against a folded protein (Trier et al., 2012) and were submitted to B-epitope predictors (http://crdd.osdd.net/raghava/abcpred/) to determine epitope effectiveness. The peptides contained a N-terminal cysteine to facilitate keyhole limpet hemocyanin (KLH) conjugation. The peptides were synthesized by Genscript, resuspended in TBS and conjugated to KLH with the Thermo Scientific Imject Maleimide-Activated mcKLH Spin Kit by following the manufacturer's instructions.

Nitration of KLH

KLH was nitrated by following a previously described method (Ischiropoulos et al., 1992; Beckmann et al., 1994). Briefly, KLH (100 μg) was diluted into 100 mM potassium phosphate pH 7.4 to a final concentration of 0.160 μg/μL. Peroxynitrite (160 mM stock concentration) was added to the solution to achieve a final concentration of 2 mM. The solution was incubated at room temperature for 14 h, after which nitrated KLH was exchanged into PBS by overnight dialysis at 4°C (MWCO 3,000 Da). Nitration was confirmed by Western Blotting using Millipore's Polyclonal anti-nitroTyr Ab (Cat # 06-284, see below section "Western Blots").

Construction of Nb Display Libraries
Immunization of Animal

A healthy Oregon State University owned alpaca was the host for Nb production. The animal was immunized over a series of six injections with a total of 700 μg of a mixture of the nitrated peptides conjugated to KLH in 1X PBS mixed with equal volume

(1 ml:1 ml) Sigma adjuvant System. The peptides were injected subcutaneously over the right caudal neck. Pre- and post-immune serum samples were collected. Immunizations were followed by a production bleed not exceeding 200 ml and whole blood collected into heparinized tubes. All animal protocols were approved by the Oregon State University Institutional Animal Care and Use Committee (IACUC).

Serum Analysis

Pre- and post-immune serum samples were collected and screened for the presence of Nbs binding nitrated targets. 14-3-3 (WT and nY133), CaM (WT, nY99 and nY138) (5 μg), and HCT116 cell lysate as a positive control were separated on 15% SDS-PAGE gels, transferred to PVDF membrane, blocked with 5% (w/v) nonfat milk in TBST and probed with either pre- or post-immunization serum, rocking for 16 h at room temperature. After rinsing three times with TBST, the membranes were than incubated with anti-llama horseradish peroxidase (HRP) conjugated secondary antibody (goat anti-llama IgG HRP, ab112786, Abcam, Cambridge, United Kingdom) diluted 1:20,000 in 5% nonfat milk/TBST for 1 h at room temperature. The membranes were washed three times with TBST and visualized with Clarity Western ECL Substrate.

Library Generation

After the immunization procedure, blood samples were collected (100 ml), peripheral blood lymphocytes (PBLs) were enriched by gradient centrifugation from blood and total RNA was isolated from the PBLs to prepare cDNA (RNAeasy, Qiagen, San Diego, CA, United States). The open reading frames encoding all immunoglobulin heavy-chains were amplified by RT-PCR with primers. Nb open reading frames were amplified through a nested PCR using primers to generate flanking sequences amenable to homologous recombination into pSEX81 (PR3005, Progen Biotechnik GmbH, Heidelberg, Baden- Wuerttemberg, Germany). PCR products (10 μg) were mixed with NcoI/BamHI linearized pSEX81 vector (Progen, 10 μg) and ligated by with T4 DNA ligase and electro-transformed into competent *E. coli* TG1 cells. Transformants were grown in 2TY medium containing 2% glucose and 100 μg/ml ampicillin at 37°C overnight (Sabir et al., 2014).

Library Selections

The transformed TG1 cells were incubated with hyperphage (PRHYPE, Progen Biotechnik GmbH, Heidelberg, Baden-Wuerttemberg, Germany). The phage particles presenting the VHH library on their tips were collected. Phages containing the Nb fragments were enriched with solid phase panning. As a negative selection, Enzyme-Linked Immunosorbent Assay (ELISA) wells were coated with 5 μg of KLH at 4°C and phage particles were added to the wells and incubated at room temperature for 1.5 h. The unbound phage was moved to another round of negative selection for a total of three rounds. Afterwards, unbound phage were moved to positive selection with nitrated KLH. Bound phages were eluted with 0.1 M triethylamine and used for reinfection of TG1 cells, which were then used for one round of negative

selection with WT 14-3-3ß bound by the (6x) His to Pierce® Nickel Coated Plates (ThermoScientific). The unbound phage were used for two subsequent panning rounds of positive sections with nitroTyr 14-3-3ß immobilized to the Ni-coated plates resulting two full length Nb sequences (Nb-G5 and Nb-F110).
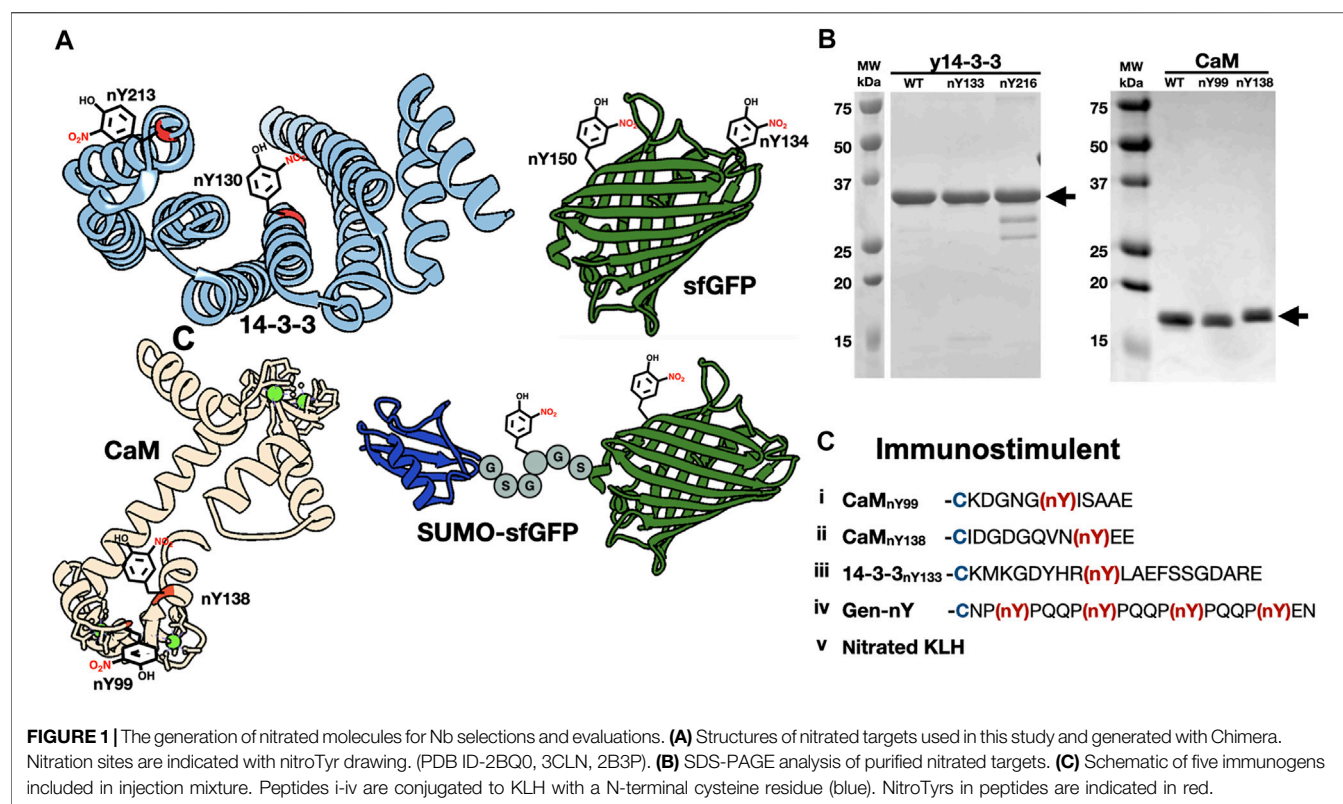
Nb Validation With Dot Blots

Immobilon-FL PVDF Membrane was activated by a brief incubation in methanol then equilibrated in MilliQ water followed by equilibration in Towbin's buffer [25 mM Tris, 192 mM glycine, 20% (v/v) methanol (pH 8.3)]. Three microlitres of 14-3-3 targets were blotted onto the PVDF membrane obtaining final masses ranging from 0.5 to 5 μg. The membrane was incubated briefly in Ponceau S to visualize the protein load. The membrane was then rinsed with TBST and blocked with 5% milk for 1 hour then rinsed 3 × 5 min with TBST. Nbs were prepared in TBS to achieve a 1:3,000 desired dilution. Nb was incubated with the PVDF membrane at RT at times ranging from 1 h to overnight. After incubation, membranes were rinsed 3 × 5 min with TBST. The secondary antibody, horseradish peroxidase (HRP) conjugated anti-llama, was diluted 1:20,000 in 5% milk and incubated with the PVDF membrane overnight. Bound Nbs were incubated for 1 h at room temperature with an anti-llama HRP conjugated secondary antibody diluted 1:20,000 in 5% nonfat milk/TBST. The membranes were washed three times with TBST and visualized with Clarity Western ECL Substrate.

Nb Validation With Enzyme-Linked Immunosorbent Assay

Nbs were prepared in TBS to achieve a 1:3,000 desired dilution and were incubated with immobilized WT 14-3-3 (5 μg) and nY 130/133 and nY 213/216 14-3-3 (0.1–5 μg) in Ni-coated ELISA wells. Nbs were incubated at room temperature for 1 hour and was visualized after washing 10x with TBST. Bound Nbs were incubated for 1 h at room temperature with an anti-llama HRP conjugated secondary antibody diluted 1:1,000 in 5% nonfat milk/TBST. The wells were washed ten times with TBST and visualized with Clarity Western ECL Substrate.

Nb Validation With Biolayer Interferometry

All BLI measurements were made on a fortéBIO (Menlo Park, CA, United States) Octet Red96 system using streptavidin (SA) or nickel (Ni) sensors. Assays were performed in 96-well microplates at 37°C. All sample volumes were 200 μL. SA sensors were loaded with 14-3-3ß homogeneously biotinylated at the N-terminus (0.020 μg/μL). Preliminary assays were done to determine an appropriate amount of biotinylated 14-3-3 to be loaded on the SA sensors, defined as the lowest amount of 14-3-3 that would provide acceptable signal above background for the lowest concentration of Nb-G5 as well as reasonable signal with minimal distortion at near-saturating concentrations of Nb-G5. After loading biotinylated 14-3-3ß onto SA sensors, a baseline was established in buffer composed of 50 mM Tris pH 7.5,

FIGURE 1 | The generation of nitrated molecules for Nb selections and evaluations. **(A)** Structures of nitrated targets used in this study and generated with Chimera. Nitration sites are indicated with nitroTyr drawing. (PDB ID-2BQ0, 3CLN, 2B3P). **(B)** SDS-PAGE analysis of purified nitrated targets. **(C)** Schematic of five immunogens included in injection mixture. Peptides i-iv are conjugated to KLH with a N-terminal cysteine residue (blue). NitroTyrs in peptides are indicated in red.

150 mM NaCl and BSA (0.1–1 mg/ml) or Tween (0.05%). Association with the Nb was then carried out in the same buffer for 60–90 s at Nb concentrations that ranged from 0.0125–1.5 μM. Dissociation was subsequently measured the same buffer for 120–300 s. BSA was used in all buffers to reduce non-specific binding. These experiments were repeated a minimum of three times, with at least three different preparations of Nb-G5.

Statistical Analysis of Biolayer Interferometry Fits

Data were reference-subtracted and aligned with each other in the Octet Data Analysis software (FortéBIO). Sensograms were fit with a 1:1 binding model to obtain kinetic binding constants. Equilibrium dissociation constant (K_D) values were calculated from the ratio of K_{off} to K_{on}. Global fits with an R^2 higher than 0.98 were considered acceptable.

Crosslinking Reactions

Crosslinking reactions were conducted in the following manner. azidoPhe-incorporated protein and target were combined in a 1:6 molar ratio in 50 mM Tris, pH 7.5, 150 mM NaCl for a total reaction volume of 50 μL. This solution was allowed to incubate for 30 min on ice, and then exposed to UV light (254 nM) *via* a UVP Inc. Model UVGL-25 Multiband UV-254/366 nm Mineralight Lamp for 5 minutes in a 96 well UV-transparent plate also on ice. After exposure to UV light, the solutions were removed from the wells and mixed with SDS sample loading buffer and analyzed with SDS-PAGE.

Western Blots

Western blot samples were separated on 4–22% gradient SDS-PAGE gels, transferred to PVDF membrane (30 V, overnight) blocked with Licor blocking buffer in TBST, and probed with anti-His (Takara, 1:1,000), or anti-V5 (Invitrogen, 1:500) primary antibodies rocking for 16 h at 4°C. After rinsing three times with TBST, the membranes were than incubated with Li-Cor IRDye 800CW Goat anti-rabbit or anti-mouse IgG (1:10,000) secondary antibody, rocking for 1 h at room temperature, and washed three times for 5 min in TBST. The membrane was then scanned using a Li-Cor Odyssey 9,120 Imaging System. If antibody overlays were desired, after imaging, the additional antibody was added to the membrane and incubated for 1 h at room temperature. After incubation, an appropriate secondary antibody was added, and the blot visualized as described above.

RESULTS

Strategy and nitroTyr Protein Generation
Strategy

With the goal of generating a Nb that is either selective for a nitration site in a specific protein or one that binds nitroTyr regardless of protein context, we chose to immunize an animal with peptides that represented multiple nitrated targets (general and protein-specific) and use the results of a preliminary serum screen to decide the best target(s) to pursue. Per usual protocols

TABLE 1 | NitroTyr proteins produced for Nb evaluation.

Protein	Nitration sites	Biological significance	Structural context	Ref
y14-3-3	Y133	Nitrated *in vivo*	Key component of pSer/pThr recognition triad	Zhao et al. (2017), Pennington et al. (2018)
y14-3-3	Y216	Nitrated *in vivo*	Solvent exposed	Nuriel et al. (2015), Pennington et al. (2018)
h14-3-3	Y130	Nitrated *in vivo*	Key component of pSer/pThr recognition triad	Zhao et al. (2017), Pennington et al. (2018)
h14-3-3	Y213	Nitrated *in vivo*	Solvent exposed	Nuriel et al. (2015), Pennington et al. (2018)
CaM	Y99	Nitration site is a target for denitrase	Found in globular Ca^{2+} binding domain	Corti et al. (1999), Smallwood et al. (2007)
CaM	Y138	Nitration at site modulates signaling behavior	Found in a conformationally dynamic region	Mukherjea et al. (1996), Porter et al. (2020)
GFP	Y134	None	In protein loop	N/A
GFP	N150	None	On side of GFP barrel	N/A
SUMO-GFP	Link, N150 (2x nitroTyr)	None	Dual nitration sites; one on a disordered linker	N/A

(Vincke et al., 2012), from the blood of a seropositive immunized animal, we generated a phage-display library from peripheral blood lymphocyte RNA and carried out phage display selections with select nitrated target proteins. The selected nitroTyr-specific Nbs were then characterized for their affinity and selectivity using both semi-quantitative (ELISA, dot-blot) and quantitative (biolayer interferometry) methods.

NitroTyr-Protein Targets for Selections and for Immunizations
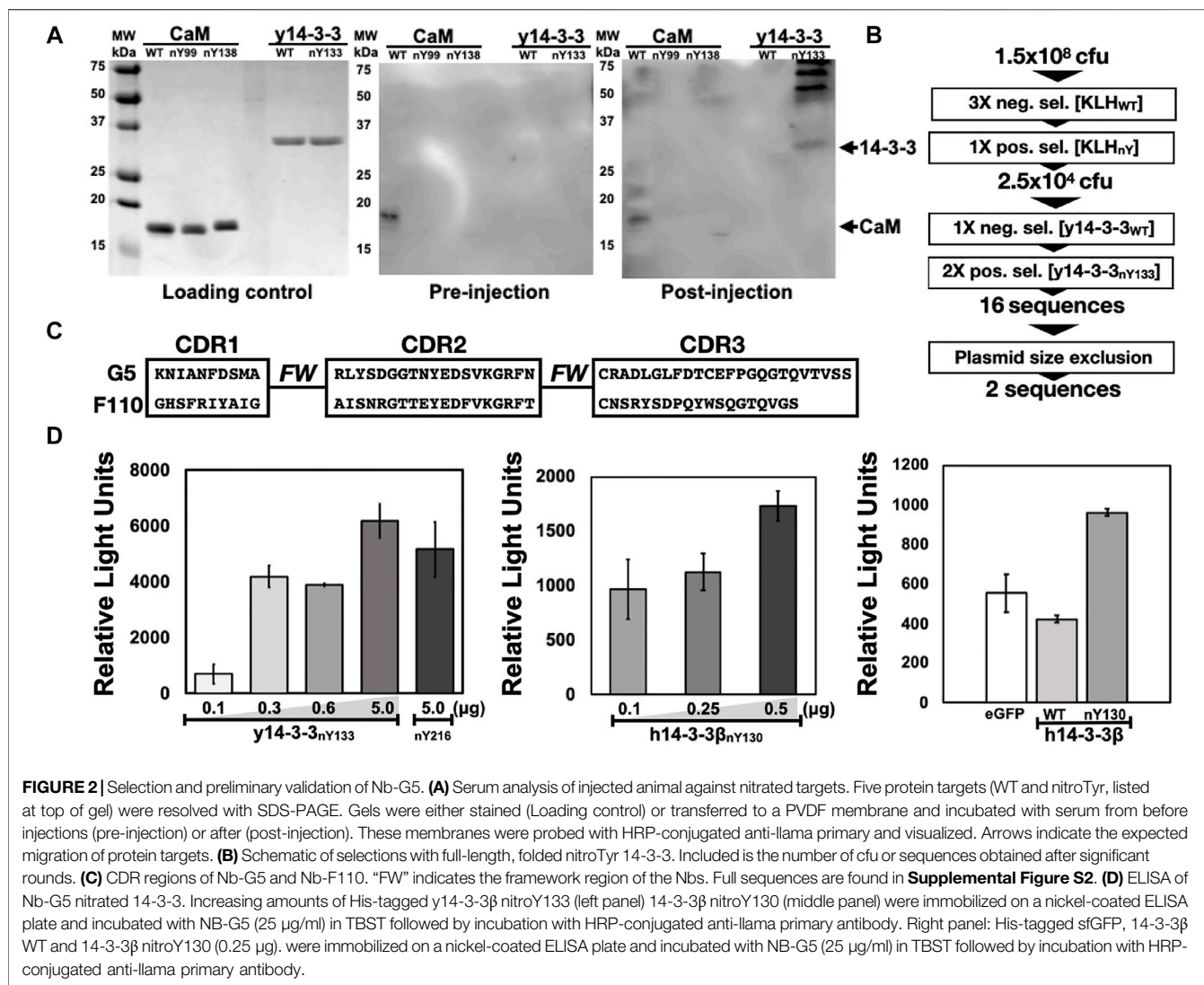
We selected signaling hub proteins 14-3-3 and CaM as nitroTyr-protein targets (**Figure 1A**). The yeast ortholog of 14-3-3 (commonly known as BMH1 and referred to here as "y14-3-3" for clarity) was used initially as it is a functional model for human 14-3-3 [(Clapp et al., 2012), ~55% identity, h14-3-3] and we had established expression and purification systems for it. While proteomics studies have identified multiple sites of *in vivo* nitration of y14-3-3 (Nuriel et al., 2015; Pennington et al., 2018), we initially focused on Y133, which is a key part of pSer/pThr binding site and fully conserved in all eukaryotic 14-3-3 proteins. For CaM, we focused on Y99 which when nitrated serves as a biomarker of oxidative stress (Smallwood et al., 2003) and Y138 which both enhances CaM binding to eNOS and has been shown to be transient in activated macrophages (Porter et al., 2020). These two sites also provide contrasting protein contexts, as Y99 is found in the less solvent-exposed globular Ca^{2+}-binding domain and Y138 is contained in a conformationally dynamic region (Smallwood et al., 2007; Piazza et al., 2012; Porter et al., 2020) (**Figure 1A**). For our immunization series, we designed peptides encompassing all three of these nitration sites along with an additional peptide containing several nitroTyrs to enhance the chance of obtaining Nbs that were generally specific for nitroTyr. With these peptides, we also included in the immunization series chemically nitrated keyhole limpet hemocyanin (KLH) (**Figure 1C**) which emulated the successful strategy employed for the original nitroTyr-Ab generation (Beckmann et al., 1994).

NitroTyr-Proteins for Further Binding Characterization

In addition to nitrated immunogens and the specific target proteins to be used in selections, we generated some nitroTyr-proteins to further characterize the specificities of the generated Nbs (**Figure 1A** and **Table 1**). To that end, we identified y14-3-3 Y216 whose nitration is also physiologically relevant and is more solvent-exposed than Y133. Additionally, the human variants of 14-3-3β (nY130 and nY213) were used because their nitration is more relevant to human disease. We further made superfolder green fluorescent proteins (sfGFPs) with nitroTyr inserted in the place of N150 or D134 to study the ability of the Nb to bind a nitroTyr that juts out from a smooth protein surface (site 150) as compared to a nitroTyr located in the loops of the barrel (site 134). And finally, to evaluate the effects of having multiple sites of nitration contained in a single protein, we created a SUMO-linker-sfGFP containing two sites of nitration. Using GCE, we produced all full-length, homogenously nitrated proteins, purified to >95% purity with yields ranging from 30–45 mg/L of culture (**Figure 1B**). The fidelity of site-specific nitroTyr incorporation was verified by mass spectrometry, with correct masses obtained for both unmodified and nitrated targets (**Supplementary Figure S1**). In total, nine specific nitroTyr proteins were produced for generating and testing Nbs (**Table 1**).

Identification of Two Promising Nbs Selected to Recognize 14-3-3$_{nY133}$
Immune-Serum Generation and Screening

All five nitrated immunogens were simultaneously injected into an alpaca following a standard immunization procedure (*Materials and Methods*). To monitor the production of Camelid antibodies that target nitroTyr-proteins, pre- and post-injection sera (collected before and after full immunization) were incubated with membrane-immobilized y14-3-3 and CaM targets. Development of the membranes revealed bands in the lane containing y14-3-3$_{nY133}$ and a faint

FIGURE 2 | Selection and preliminary validation of Nb-G5. (A) Serum analysis of injected animal against nitrated targets. Five protein targets (WT and nitroTyr, listed at top of gel) were resolved with SDS-PAGE. Gels were either stained (Loading control) or transferred to a PVDF membrane and incubated with serum from before injections (pre-injection) or after (post-injection). These membranes were probed with HRP-conjugated anti-llama primary and visualized. Arrows indicate the expected migration of protein targets. (B) Schematic of selections with full-length, folded nitroTyr 14-3-3. Included is the number of cfu or sequences obtained after significant rounds. (C) CDR regions of Nb-G5 and Nb-F110. "FW" indicates the framework region of the Nbs. Full sequences are found in **Supplemental Figure S2**. (D) ELISA of Nb-G5 nitrated 14-3-3. Increasing amounts of His-tagged y14-3-3β nitroY133 (left panel) 14-3-3β nitroY130 (middle panel) were immobilized on a nickel-coated ELISA plate and incubated with NB-G5 (25 µg/ml) in TBST followed by incubation with HRP-conjugated anti-llama primary antibody. Right panel: His-tagged sfGFP, 14-3-3β WT and 14-3-3β nitroY130 (0.25 µg). were immobilized on a nickel-coated ELISA plate and incubated with NB-G5 (25 µg/ml) in TBST followed by incubation with HRP-conjugated anti-llama primary antibody.

band in a lane containing CaM_{nY138}, none of which appear on the pre-serum membrane providing evidence that antibodies specific to nitrated targets had been produced (**Figure 2A**). Additional prominent higher molecular weight bands were also visible in the y14-3-3$_{nY133}$ lane; these may represent camelid antibody affinity for dimerized forms of 14-3-3. Due to the intensity of the y14-3-3$_{nY133}$ bands compared to the other targets (CaM_{nY99} and CaM_{nY138}), we sought to develop a protein and nitration site-specific Nb and pursued selections against y14-3-3$_{nY133}$.

Library Generation and Phage Display Selections Using y14-3-3$_{nY133}$

To generate a phage display library, the total RNA from peripheral blood lymphocytes (PBLs) after the immunization procedure were used to make Nb-encoding cDNA. The cDNAs were amplified and subcloned into the phagemid vector pSEX81, generating a phage display library. With this library, we performed seven rounds of phage display selections

(**Figure 2B**). The first three rounds were negative selections using KLH as the target, designed to exclude anti-KLH library members. Then, mimicking the process that yielded the initial nitroTyr-Abs (Beckmann et al., 1994), we carried out a positive selection with nitroTyr-KLH to ensure that remaining Nbs had some direct recognition of nitroTyr. These four rounds of selection diminished the library size from 1.5×10^8 cfu (colony forming units) to 2.5×10^4 cfu (**Figure 2B**).

We then panned for library members that were selective for y14-3-3$_{nY133}$ by using one round of negative selection with wild type y14-3-3, followed by two rounds of positive selections with y14-3-3$_{nY133}$ (**Figure 2B**). After a plasmid size-exclusion step to eliminate truncated Nbs, two full-length Nbs were obtained: Nb-G5 and Nb-F110 (**Figure 2C**, **Supplementary Figure S2**). Expression and purification of Nb-G5 yielded protein sufficient for downstream characterization (**Supplementary Figure S3**) but we were unable to identify suitable *E. coli* expression strategies for

Nb-F110. As a result, we moved forward with evaluating only Nb-G5.

Characterization of the Binding Properties of Nb-G5
Semi-Quantitative Survey of Nb-G5 Binding to 14-3-3 Protein Forms

Binding of Nb-G5 was qualitatively evaluated with both dot-blot assays and enzyme-linked immunosorbent assays (ELISA). With the dot-blot assay, we evaluated Nb-G5 selectivity and Nb binding sensitivity. Nitrated targets y14-3-3$_{nY133}$ and y14-3-3$_{nY216}$ along with y14-3-3ß$_{WT}$ were dotted in varying amounts onto a membrane and probed with Nb-G5; binding was observed for both nitroTyr proteins but not y14-3-3ß$_{WT,}$ indicating a selectivity for nitrated protein (**Supplementary Figure S4**). y14-3-3$_{nY216}$ was included to see if Nb-G5 possessed any sensitivity for nitration independent of the protein context of site nY133. Having at this point successfully expressed the more biomedically relevant nitrated human 14-3-3ß (*h*14-3-3ß) isoforms, we also probed these and found they were recognized with a similar level of sensitivity (**Supplementary Figure S4**).

The range of selectivity and sensitivity of Nb-G5 to immobilized, folded protein in solution was assayed by ELISA. For both the yeast and human 14-3-3 variants, significant signal above background was observed, down to 0.1 μg of immobilized nitrated 14-3-3. The selectivity of Nb-G5 for nitroTyr was evaluated by assaying immobilized 14-3-3ß$_{WT}$ and sfGFP (as a negative control) alongside h14-3-3ß$_{nY130}$. Significant signal for h14-3-3ß$_{nY130}$ over sfGFP and h14-3-3ß $_{WT,}$ which both exhibited a similar level of background was also observed by ELISA (**Figure 2D**). Based on this, we decided to carry out quantitative binding studies with the h14-3-3ß forms rather than with the y14-3-3 forms.

Quantitative Analyses Nb-G5 Affinity and Selectivity

Encouraged by our preliminary evaluation of Nb binding, we employed biolayer interferometry (BLI) to quantitatively assess the sensitivity and selectivity of Nb-G5. Avi-tagged h14-3-3ß variants (WT, nY130 and nY213) were immobilized onto BLI tips and incubated with Nb-G5 to measure the association and dissociation kinetics of binding. Global fits of the BLI sensograms in 0.25% bovine serum albumin (BSA) gave a K$_D$ value for Nb-G5 against h14-3-3ß$_{nY130}$ of ~14 nM, a value that is an order of magnitude tighter than that obtained for Nb-G5 against h14-3-3ß$_{WT}$ under the same assay conditions, and 2–5 times tighter than the K$_D$ for h14-3-3ß$_{nY213}$ (**Figure 3A** and **Table 2**). The main contributor to the observed differences in K$_D$ is the dissociation rate, consistent with the nitroTyr protein forms making specific interactions with Nb-G5 that slow its dissociation (**Table 2**). To minimize the contribution of non-specific binding, the assays were also performed in 1% BSA. In these experiments, binding to h14-3-3ß$_{WT}$ was almost completely ablated and kinetic constants could not be determined, although binding to h14-3-3ß$_{nY130}$ was still observed (K$_D$ ~40 nM, **Figure 3B**). This is consistent with

binding to WT protein being non-specific, whereas binding to both nitrated protein forms involving some level of specific recognition.

To further assess the relative contributions of the nitroTyr group itself *versus* the protein context to the binding of Nb-G5 for nitrated targets, we tested Nb-G5 binding with proteins containing nitroTyr in different protein locations (**Figure 1A**). Proteins included three CaM forms (WT, nY99, nY138 in 0.05% Tween) and four sfGFP forms [WT, nY134, nY150, SUMO-link-sfGFP (2x nitroTyr) in 0.2% BSA, **Table 1**]. Although none of these proteins bound as well as either of the nitrated forms of h14-3-3ß, local fits revealed some binding to both CaM$_{nY138}$ (**Figure 3C** panel A), sfGFP$_{nY134}$ and SUMO-link-sfGFP (2x nitroTyr) (**Figure 3C** panel B) for these targets at high Nb-G5 concentrations (>200 nM). These interactions generated K$_D$ values >100 nM, and no notable binding to the wild type proteins or to the nitrated targets GFP$_{nY150}$ or CaM$_{nY99}$ was observed (**Supplementary Figure S5**).
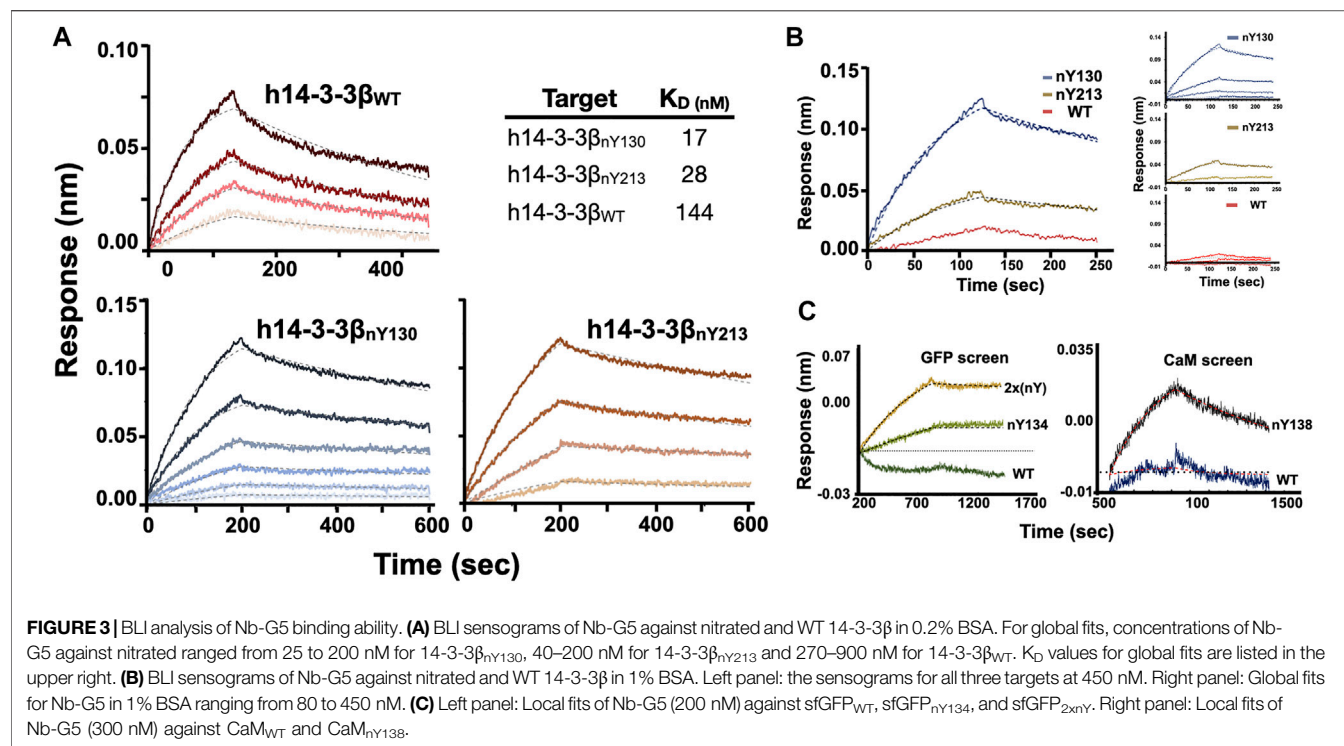
Nb-G5 Nanobody as Basis for Creating nitroTyr Specific Cross-Linking Tools
Identification of Potential Sites and Construct Design for Crosslinking Experiments

As noted in the introduction, generating covalent Nb binders (i.e., "GlueBodies") through incorporating ncAAs with crosslinking abilities (ncAA-CL), is a means to target proteins for degradation (Zhang et al., 2021). Nbs that target oxPTMs for degradation would be powerful tools for discovering possible physiological changes triggered by specific oxidized proteins. A variety of ncAA-CL can be incorporated by GCE, including chemically reactive, proximity-induced and photo-reactive ncAA-CLs (**Figure 4A**). Because the GCE machinery for photocrosslinking ncAA-CL 4-azido phenylalanine (azidoPhe, pAzF) is robust, the ncAA is commercially available and the crosslinking assays are simple to perform, we decided to investigate the suitability of Nb-G5 as a covalent crosslinker by the installation of azidoPhe into Nb-G5 with GCE.

We first sought to identify locations on Nb-G5 near the edge of the paratope, so that upon binding they would be close to the target, yet not interfere with Nb binding itself. We noticed the CDR3 loop of Nb-G5 contained an unpaired cysteine residue which is unusual in Nbs (Bannas et al., 2017) and so we speculated this Cys was part of the antigen binding site. To evaluate its contribution, we generated a C115S mutant of Nb-G5 and observed by BLI that the binding to h14-3-3ß$_{nY130}$ decreased by six-fold, and that the selectivity for h14-3-3ß$_{nY130}$ over h14-3-3$_{WT}$ was lost (**Figure 4B**).

We created two different strategies to visualize the crosslinked material (**Figure 4C**). The first was appending a C-terminal GFP fusion to the Nb-G5, which would give a sensitive fluorescent signature to the Nb and allow CL-dependent mass shifts to be easily observed with in-gel fluorescence. Concerned that a bulky GFP tag could impact the binding of Nb-G5, we also designed a construct with a smaller C-terminal V5 tag, which could be visualized by Western blotting (**Figure 4C**, "ii"). Also, since we expected the Nb-G5 incorporating azidoPhe would express

FIGURE 3 | BLI analysis of Nb-G5 binding ability. **(A)** BLI sensograms of Nb-G5 against nitrated and WT 14-3-3β in 0.2% BSA. For global fits, concentrations of Nb-G5 against nitrated ranged from 25 to 200 nM for 14-3-3β_{nY130}, 40–200 nM for 14-3-3β_{nY213} and 270–900 nM for 14-3-3β_{WT}. K_D values for global fits are listed in the upper right. **(B)** BLI sensograms of Nb-G5 against nitrated and WT 14-3-3β in 1% BSA. Left panel: the sensograms for all three targets at 450 nM. Right panel: Global fits for Nb-G5 in 1% BSA ranging from 80 to 450 nM. **(C)** Left panel: Local fits of Nb-G5 (200 nM) against sfGFP$_{WT}$, sfGFP$_{nY134}$, and sfGFP$_{2xnY}$. Right panel: Local fits of Nb-G5 (300 nM) against CaM$_{WT}$ and CaM$_{nY138}$.

TABLE 2 | Kinetic constants for Nb-G5:14-3-3 binding.

TARGET (in 0.2% BSA)	K_D (nM)	k_{on} (10^4 M^{-1} s^{-1})	k_{off} (10^3 s^{-1})	BSA
14-3-3β_{nY130}	17 ± 3	2.2 ± 0.2	0.33 ± 0.02	0.2%
14-3-3β_{nY213}	28 ± 2	2.6 ± 0.1	0.73 ± 0.02	0.2%
14-3-3β_{WT}	144 ± 8	2.1 ± 0.6	2.3 ± 0.5	0.2%

TARGET (1.0% BSA)	K_D (nM)	k_{on} (10^4 M^{-1} s^{-1})	k_{off} (10^3 s^{-1})	BSA
14-3-3β_{nY130}	36 ± 3	3.2 ± 0.1	1.19 ± 0.07	1.0%
14-3-3β_{nY213}	110 ± 3	2.05 ± 0.02	2.26 ± 0.05	1.0%
14-3-3β_{WT}	Undeterminable	Undeterminable	Undeterminable	1.0%

less well, we sought to enhance the expression through adding a cleavable N-terminal bdNEDD8 solubility tag (Pleiner et al., 2018). With the systems designed, we selected six potential sites flanking C115 for azidoPhe incorporation, and all six sites for both designs (GFP and V5) yielded sufficient cleaved, azidoPhe-containing Nb-G5 for crosslinking experiments (**Figure 4D**, **Supplementary Figure S6**).

Crosslinking Attempts With h1413-3ß$_{nY133}$ and CaM$_{nY138}$

As nitroTyr-protein targets, we selected the two that were bound best by Nb-G5: 14-3-3$_{nY133}$ and CaM$_{nY138}$. Both targets possess C-terminal His tags, allowing them to be detected with Western Blot. Screening of the six azidoPhe sites with both nitrated targets for their ability to yield nitroTyr-dependent crosslinked products yielded disappointing results (**Supplementary Figures S7, S8**). Some bands appear at the correct molecular weight range for pAzF113 with the CaM target and pAzF108 and pAzF110 for the

14-3-3 target, but even for these the amount of CL product is minimal and masked by the presence of non-specific crosslinked products. In addition, the photoactivation appears to dimerize the nitroTyr-containing target protein, further complicating analysis and making the approach with this Nb-G5 and azidoPhe photocrosslinker challenging to implement in a more complex environment (live cell).

DISCUSSION

With the aid of GCE technology, we show it is possible to generate a Nb selective for an oxidatively-modified protein target, opening the doors to dynamic studies of oxPTM-impacted proteins in living cells. This technology is not limited to oxPTMs, as GCE can provide site-specifically modified protein for many biologically important PTMs, including phosphoserine (Rogerson et al., 2015; Zhu et al., 2019) and acetyl lysine (Neumann-Staubitz et al.,

FIGURE 4 | Generating an anti-nitroTyr cross-linking tool. **(A).** Examples of three common crosslinking strategies: (i) Chemical crosslinkers, such as the carbamate-containing ncAA FPheK, display a high reactivity to proximal nucleophilic residues (e.g., Lys, Cys, and Tyr) under weakly basic conditions (Xuan et al., 2017); (ii) Proximity-based crosslinkers, such as the haloalkane ncAA shown, react with Cys residues in the target protein facilitated by an increased local effective concentration (Xiang et al., 2014); and (iii) Photoreactive crosslinkers, such as benzophenones, diazirines, and azides (as in the azidoPhe ncAA shown), contain UV-induced cross-linking moieties that react with diverse proximal groups (Hoffmann 2020). **(B).** K_Ds of Nb-G5 and Nb-G5 (C115S) against nitrated and WT 14-3-3 are shown. **(C).** Constructs of Nb-G5-GFP (i) and Nedd8-Nb-G5-V5 (ii). **(D).** A model of Nb-G5, generated by ABodyBuilder: automated antibody structure predictor (Leem et al., 2016). The CDR3 loop unpaired cysteine (C115) is indicated in red and the adjacent resides individually substituted with azidoPhe are colored teal and labeled. An anti-V5 western blot image (top panel) and an in-gel fluorescence assay (bottom panel) show both kinds of azidoPhe-encoded Nb-G5 constructs expressed well.

2021). This allows, for the first time, a rigorous characterization of the site and protein specificity of Nbs developed against these targets contributing to both the engineering of nanobodies for specific purposes as well our general knowledge of Nb-target recognition.

Starting from the serum of an animal injected with a mixture of five nitrated immunogens, we developed an anti-nitroTyr nanobody library and performed phage-display selections for the nitrated target 14-3-3$_{nY130}$, obtaining the sequence for a single soluble nanobody, Nb-G5. BLI analysis of Nb-G5 with 14-3-3$_{nY130}$ revealed a 10-fold selectivity for 14-3-3$_{nY130}$ over 14-3-3$_{WT}$ in 0.2% BSA solutions, exhibiting a modest K_D of ~14 nM. In 1% BSA solutions, Nb-G5 exhibited a higher K_D but complete selectivity for 14-3-3$_{nY130}$ over 14-3-3$_{WT}$. Nb-G5 also showed some weak yet selective binding to other non-14-3-3 nitrated targets, which may indicate that the Nb is interacting with the nitroTyr molecule itself. This weak binding was site-dependent and appeared to be strongest when the nitroTyr molecule was exposed, but still surrounded by ample protein context.

After obtaining some information about a cysteine that is critical to Nb selectivity, we were able to site-specifically install the ncAA azidoPhe in the CDR3 region of Nb-G5. Using robust GCE machinery for azidoPhe, all azidoPhe sites were successfully installed at reasonable yields allowing us to easily screen the library against two nitrated targets for crosslinking. However, the amount of crosslinked product produced was small, and the presence of nitroTyr in proteins appeared to increase the

propensity of the nitrated proteins to crosslink under UV exposure, limiting the suitability of this specific crosslinking strategy for nitroTyr targets. Nitrotyrosine increases the UV absorbance of the target proteins (Crow and Beckman 1995) and appears to be competing with the desired photocrosslinking pathway. With GCE, we are not limited to one type of crosslinking chemistry and could utilize crosslinkers that instead are proximity-based and operate independently of UV irradiation, including ncAA with bromoalkyl moieties (Cigler et al., 2017). These would be well-suited for Nbs targeting chemically labile modifications which may be more sensitive to UV irradiation such as those installed under conditions of oxidative stress. Regardless, we were encouraged by the ability to form ncAA-CL encoded-Nbs, showing that GCE-technology is well suited to be implemented in every step of Nb development, from selections to additional functionalities.

The mid-nanomolar K_D obtained for NB-G5 with no additional rounds of affinity maturation is promising, however, Nbs that are designed to be employed as intracellular tools require low to sub-nanomolar affinities for their targets (Mahajan et al., 2018). Originally, when we were performing the selections, we included a positive selection step with nitrated KLH to emulate the strategy that produced the first anti-nitroTyr antibodies. It may be that more selective binders for 14-3-3nY were excluded from the set in that KLH$_{nY}$ positive selection step, although Nb-G5 has slight general specificity. The pre-KLH positive step library still exists, and it represents a valuable

resource in itself that could be used to select for Nbs against other nitroTyr proteins, or a tighter 14-3-3nY133 binder. With respect to the selections themselves, Nb-selection technology is rapidly improving. With GCE-produced proteins and multiple rounds off affinity maturation, the development of anti-oxPTM Nbs with sub-nanomolar affinities, appropriate of intracellular interrogation, are within reach. Such Nbs will be powerful tools that will open up a window into the complex intracellular dynamics of oxidatively modified proteins and their impacts.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The animal study was reviewed and approved by Oregon State University Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

Conceptualization: EV, CR, DM, CC, RC, and RM. Methodology: EV, SG, DM, CC, SB, RC, and RM. Investigation: EV and SG. Supervision: RM and SB. Writing—original draft: EV and SG. Writing—review and editing: EV, CR, DM, CC, RC, and RM.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.835229/full#supplementary-material

REFERENCES

Bannas, P., Hambach, J., and Koch-Nolte, F. (2017). Nanobodies and Nanobody-Based Human Heavy Chain Antibodies as Antitumor Therapeutics. *Front. Immunol.* 8, 1603. doi:10.3389/fimmu.2017.01603

Bartesaghi, S., and Radi, R. (2018). Fundamentals on the Biochemistry of Peroxynitrite and Protein Tyrosine Nitration. *Redox Biol.* 14, 618–625. doi:10.1016/j.redox.2017.09.009

Beckmann, J. S., Ye, Y. Z., Anderson, P. G., Chen, J., Accavitti, M. A., Tarpey, M. M., et al. (1994). Extensive Nitration of Protein Tyrosines in Human Atherosclerosis Detected by Immunohistochemistry. *Biol. Chem. Hoppe-Seyler* 375 (2), 81–88. doi:10.1515/bchm3.1994.375.2.81

Bery, N., Keller, L., Soulié, M., Gence, R., Iscache, A.-L., Cherier, J., et al. (2019). A Targeted Protein Degradation Cell-Based Screening for Nanobodies Selective toward the Cellular RHOB GTP-Bound Conformation. *Cell Chem. Biol.* 26 (11), 1544–1558. doi:10.1016/j.chembiol.2019.08.009

Beyer, J. N., Hosseinzadeh, P., Gottfried-Lee, I., Van Fossen, E. M., Zhu, P., Bednar, R. M., et al. (2020). Overcoming Near-Cognate Suppression in a Release Factor 1-Deficient Host with an Improved Nitro-Tyrosine tRNA Synthetase. *J. Mol. Biol.* 432 (16), 4690–4704. doi:10.1016/j.jmb.2020.06.014

Cheloha, R. W., Harmand, T. J., Wijne, C., Schwartz, T. U., and Ploegh, H. L. (2020). Exploring Cellular Biochemistry with Nanobodies. *J. Biol. Chem.* 295 (45), 15307–15327. doi:10.1074/jbc.rev120.012960

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

Clapp, C., Portt, L., Khoury, C., Sheibani, S., Norman, G., Ebner, P., et al. (2012). 14-3-3 Protects against Stress-Induced Apoptosis. *Cell Death Dis* 3, e348. doi:10.1038/cddis.2012.90

Cooley, R. B., Feldman, J. L., Driggers, C. M., Bundy, T. A., Stokes, A. L., Karplus, P. A., et al. (2014a). Structural Basis of Improved Second-Generation 3-nitro-tyrosine tRNA Synthetases. *Biochemistry* 53 (12), 1916–1924. doi:10.1021/bi5001239

Cooley, R. B., Karplus, P. A., and Mehl, R. A. (2014b). Gleaning Unexpected Fruits from Hard-Won Syntheses: Probing Principles of Permissivity in Non-canonical Amino Acid-tRNA Synthetases. *Chembiochem* 15 (12), 1810–1819. doi:10.1002/cbic.201402180

Corti, C., Leclerc L'Hostis, E., Quadroni, M., Schmid, H., Durussel, I., Cox, J., et al. (1999). Tyrosine Phosphorylation Modulates the Interaction of Calmodulin with its Target Proteins. *Eur. J. Biochem.* 262 (3), 790–802. doi:10.1046/j.1432-1327.1999.00441.x

Crow, J. P., and Beckman, J. S. (1995). Quantitation of Protein Tyrosine, 3-Nitrotyrosine, and 3-Aminotyrosine Utilizing HPLC and Intrinsic Ultrviolet Absorbance. *Methods* 7 (1), 116–120. doi:10.1006/meth.1995.1017

de Beer, M. A., and Giepmans, B. N. G. (2020). Nanobody-Based Probes for Subcellular Protein Identification and Visualization. *Front. Cell. Neurosci.* 14, 573278. doi:10.3389/fncel.2020.573278

DiDonato, J. A., Aulak, K., Huang, Y., Wagner, M., Gerstenecker, G., Topbas, C., et al. (2014). Site-specific Nitration of Apolipoprotein A-I at Tyrosine 166 Is Both Abundant within Human Atherosclerotic Plaque and Dysfunctional. *J. Biol. Chem.* 289 (15), 10276–10292. doi:10.1074/jbc.m114.556506

Fairhead, M., and Howarth, M. (2015). Site-specific Biotinylation of Purified Proteins Using BirA. *Methods Mol. Biol.* 1266, 171–184. doi:10.1007/978-1-4939-2272-7_12

Ferrer-Sueta, G., Campolo, N., Trujillo, M., Bartesaghi, S., Carballal, S., Romero, N., et al. (2018). Biochemistry of Peroxynitrite and Protein Tyrosine Nitration. *Chem. Rev.* 118 (3), 1338–1408. doi:10.1021/acs.chemrev.7b00568

Franco, M. C., Ricart, K. C., Gonzalez, A. S., Dennys, C. N., Nelson, P. A., Janes, M. S., et al. (2015). Nitration of Hsp90 on Tyrosine 33 Regulates Mitochondrial

Metabolism. *J. Biol. Chem.* 290 (31), 19055–19066. doi:10.1074/jbc.m115.663278

Franco, M. C., Ye, Y., Refakis, C. A., Feldman, J. L., Stokes, A. L., Basso, M., et al. (2013). Nitration of Hsp90 Induces Cell Death. *Proc. Natl. Acad. Sci.* 110 (12), E1102–E1111. doi:10.1073/pnas.1215177110

Frey, S., and Görlich, D. (2014). A New Set of Highly Efficient, Tag-Cleaving Proteases for Purifying Recombinant Proteins. *J. Chromatogr. A* 1337, 95–105. doi:10.1016/j.chroma.2014.02.029

Gerding, H. R., Karreman, C., Daiber, A., Delp, J., Hammler, D., Mex, M., et al. (2019). Reductive Modification of Genetically Encoded 3-nitrotyrosine Sites in Alpha Synuclein Expressed in E.Coli. *Redox Biol.* 26, 101251. doi:10.1016/j.redox.2019.101251

Ghesquiere, B., Colaert, N., Helsens, K., Dejager, L., Vanhaute, C., Verleysen, K., et al. (2009). *In Vitro* and *In Vivo* Protein-Bound Tyrosine Nitration Characterized by diagonal Chromatography. *Mol. Cell Proteomics* 8 (12), 2642–2652. doi:10.1074/mcp.m900259-mcp200

Green, M. R., and Sambrook, J. (2018). Touchdown Polymerase Chain Reaction (PCR). *Cold Spring Harb Protoc.* 2018 (5). doi:10.1101/pdb.prot095133

Herce-Pagliai, C., Kotecha, S., and Shuker, D. E. G. (1998). Analytical Methods for 3-nitrotyrosine as a Marker of Exposure to Reactive Nitrogen Species: a Review. *Nitric Oxide* 2 (5), 324–336. doi:10.1006/niox.1998.0192

Hoffmann, J. E. (2020). Bifunctional Non-canonical Amino Acids: Combining Photo-Crosslinking with Click Chemistry. *Biomolecules* 10 (4), 578. doi:10.3390/biom10040578

Ischiropoulos, H., Zhu, L., Chen, J., Tsai, M., Martin, J. C., Smith, C. D., et al. (1992). Peroxynitrite-mediated Tyrosine Nitration Catalyzed by Superoxide Dismutase. *Arch. Biochem. Biophys.* 298 (2), 431–437. doi:10.1016/0003-9861(92)90431-u

Jang, H. S., Gu, X., Cooley, R. B., Porter, J. J., Henson, R. L., Willi, T., et al. (2020). Efficient Site-specific Prokaryotic and Eukaryotic Incorporation of Halotyrosine Amino Acids into Proteins. *ACS Chem. Biol.* 15 (2), 562–574. doi:10.1021/acschembio.9b01026

Khan, M. A., Alam, K., Hassan, S. M., and Rizvi, M. M. A. (2017). Nitration of H2B Histone Elicits an Immune Response in Experimental Animals. *Autoimmunity* 50 (4), 232–240. doi:10.1080/08916934.2017.1347643

Leem, J., Dunbar, J., Georges, G., Shi, J., and Deane, C. M. (2016). ABodyBuilder: Automated Antibody Structure Prediction with Data-Driven Accuracy Estimation. *MAbs* 8 (7), 1259–1268. doi:10.1080/19420862.2016.1205773

Mahajan, S. P., Meksiriporn, B., Waraho-Zhmayev, D., Weyant, K. B., Kocer, I., Butler, D. C., et al. (2018). Computational Affinity Maturation of Camelid Single-Domain Intrabodies against the Nonamyloid Component of Alpha-Synuclein. *Sci. Rep.* 8 (1), 17611. doi:10.1038/s41598-018-35464-7

Moeglin, E., Desplancq, D., Stoessel, A., Massute, C., Ranniger, J., McEwen, A. G., et al. (2021). A Novel Nanobody Precisely Visualizes Phosphorylated Histone H2AX in Living Cancer Cells under Drug-Induced Replication Stress. *Cancers (Basel)* 13 (13), 3317. doi:10.3390/cancers13133317

Möller, M. N., Rios, N., Trujillo, M., Radi, R., Denicola, A., and Alvarez, B. (2019). Detection and Quantification of Nitric Oxide-Derived Oxidants in Biological Systems. *J. Biol. Chem.* 294 (40), 14776–14802. doi:10.1074/jbc.rev119.006136

Mukherjea, P., Maune, J. F., and Beckingham, K. (1996). Interlobe Communication in Multiple Calcium-Binding Site Mutants ofDrosophilacalmodulin. *Protein Sci.* 5 (3), 468–477. doi:10.1002/pro.5560050308

Muyldermans, S. (2013). Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* 82, 775–797. doi:10.1146/annurev-biochem-063011-092449

Neumann, H., Hazen, J. L., Weinstein, J., Mehl, R. A., and Chin, J. W. (2008). Genetically Encoding Protein Oxidative Damage. *J. Am. Chem. Soc.* 130 (12), 4028–4033. doi:10.1021/ja710100d

Neumann-Staubitz, P., Lammers, M., and Neumann, H. (2021). Genetic Code Expansion Tools to Study Lysine Acylation. *Adv. Biol. (Weinh)* 5 (12), e2100926. doi:10.1002/adbi.202100926

Nuriel, T., Whitehouse, J., Ma, Y., Mercer, E. J., Brown, N., and Gross, S. S. (2015). ANSID: A Solid-phase Proteomic Approach for Identification and Relative Quantification of Aromatic Nitration Sites. *Front. Chem.* 3, 70. doi:10.3389/fchem.2015.00070

Pacher, P., Beckman, J. S., and Liaudet, L. (2007). Nitric Oxide and Peroxynitrite in Health and Disease. *Physiol. Rev.* 87 (1), 315–424. doi:10.1152/physrev.00029.2006

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S. G. F., Wohlkönig, A., Ruf, A., et al. (2014). A General Protocol for the Generation of Nanobodies for Structural Biology. *Nat. Protoc.* 9 (3), 674–693. doi:10.1038/nprot.2014.039

Pennington, K., Chan, T., Torres, M., and Andersen, J. (2018). The Dynamic and Stress-Adaptive Signaling Hub of 14-3-3: Emerging Mechanisms of Regulation and Context-dependent Protein-Protein Interactions. *Oncogene* 37 (42), 5587–5604. doi:10.1038/s41388-018-0348-3

Piazza, M., Futrega, K., Spratt, D. E., Dieckmann, T., and Guillemette, J. G. (2012). Structure and of Calmodulin (CaM) Bound to Nitric Oxide Synthase Peptides: Effects of a Phosphomimetic CaM Mutation. *Biochemistry* 51 (17), 3651–3661. doi:10.1021/bi300327z

Pleiner, T., Bates, M., and Görlich, D. (2018). A Toolbox of Anti-mouse and Anti-rabbit IgG Secondary Nanobodies. *J. Cell Biol* 217 (3), 1143–1154. doi:10.1083/jcb.201709115

Porter, J. J., and Mehl, R. A. (20182018). Genetic Code Expansion: A Powerful Tool for Understanding the Physiological Consequences of Oxidative Stress Protein Modifications. *Oxid Med. Cell Longev* 2018, 7607463. doi:10.1155/2018/7607463

Porter, J. J., Jang, H. S., Haque, M. M., Stuehr, D. J., and Mehl, R. A. (2020). Tyrosine Nitration on Calmodulin Enhances Calcium-dependent Association and Activation of Nitric-Oxide Synthase. *J. Biol. Chem.* 295 (8), 2203–2211. doi:10.1074/jbc.ra119.010999

Porter, J. J., Jang, H. S., Van Fossen, E. M., Nguyen, D. P., Willi, T. S., Cooley, R. B., et al. (2019). Genetically Encoded Protein Tyrosine Nitration in Mammalian Cells. *ACS Chem. Biol.* 14 (6), 1328–1336. doi:10.1021/acschembio.9b00371

Radi, R. (2018). Oxygen Radicals, Nitric Oxide, and Peroxynitrite: Redox Pathways in Molecular Medicine. *Proc. Natl. Acad. Sci. USA* 115 (23), 5839–5848. doi:10.1073/pnas.1804932115

Randall, L. M., Dalla Rizza, J., Parsonage, D., Santos, J., Mehl, R. A., Lowther, W. T., et al. (2019). Unraveling the Effects of Peroxiredoxin 2 Nitration; Role of C-Terminal Tyrosine 193. *Free Radic. Biol. Med.* 141, 492–501. doi:10.1016/j.freeradbiomed.2019.07.016

Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S. M., et al. (2015). Efficient Genetic Encoding of Phosphoserine and its Nonhydrolyzable Analog. *Nat. Chem. Biol.* 11 (7), 496–503. doi:10.1038/nchembio.1823

Sabir, J. S. M., Atef, A., El-Domyati, F. M., Edris, S., Hajrah, N., Alzohairy, A. M., et al. (2014). Construction of Naïve Camelids VHH Repertoire in Phage Display-Based Library. *Comptes Rendus Biologies* 337 (4), 244–249. doi:10.1016/j.crvi.2014.02.004

Smallwood, H. S., Galeva, N. A., Bartlett, R. K., Urbauer, R. J. B., Williams, T. D., Urbauer, J. L., et al. (2003). Selective Nitration of Tyr99 in Calmodulin as a Marker of Cellular Conditions of Oxidative Stress. *Chem. Res. Toxicol.* 16 (1), 95–102. doi:10.1021/tx025566a

Smallwood, H. S., Lourette, N. M., Boschek, C. B., Bigelow, D. J., Smith, R. D., Paša-Tolić, L., et al. (2007). Identification of a Denitrase Activity against Calmodulin in Activated Macrophages Using High-Field Liquid Chromatography–FTICR Mass Spectrometry. *Biochemistry* 46 (37), 10498–10505. doi:10.1021/bi7009713

Souza, J. M., Peluffo, G., and Radi, R. (2008). Protein Tyrosine Nitration-Functional Alteration or Just a Biomarker? *Free Radic. Biol. Med.* 45 (4), 357–366. doi:10.1016/j.freeradbiomed.2008.04.010

Studier, F. W. (2005). Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expr. Purif.* 41 (1), 207–234. doi:10.1016/j.pep.2005.01.016

Tomin, T., Schittmayer, M., Honeder, S., Heininger, C., and Birner-Gruenberger, R. (2019). Irreversible Oxidative post-translational Modifications in Heart Disease. *Expert Rev. Proteomics* 16 (8), 681–693. doi:10.1080/14789450.2019.1645602

Trier, N. H., Hansen, P. R., and Houen, G. (2012). Production and Characterization of Peptide Antibodies. *Methods* 56 (2), 136–144. doi:10.1016/j.ymeth.2011.12.001

Vincke, C., Gutiérrez, C., Wernery, U., Devoogdt, N., Hassanzadeh-Ghassabeh, G., and Muyldermans, S. (2012). Generation of Single Domain Antibody Fragments Derived from Camelids and Generation of Manifold Constructs. *Methods Mol. Biol.* 907, 145–176. doi:10.1007/978-1-61779-974-7_8

Xiang, Z., Lacey, V. K., Ren, H., Xu, J., Burban, D. J., Jennings, P. A., et al. (2014). Proximity-enabled Protein Crosslinking through Genetically Encoding

Haloalkane Unnatural Amino Acids. *Angew. Chem. Int. Ed.* 53 (8), 2190–2193. doi:10.1002/anie.201308794

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

Zhang, H., Han, Y., Yang, Y., Lin, F., Li, K., Kong, L., et al. (2021). Covalently Engineered Nanobody Chimeras for Targeted Membrane Protein Degradation. *J. Am. Chem. Soc.* 143 (40), 16377–16382. doi:10.1021/jacs.1c08521

Zhang, Y., Werling, U., and Edelmann, W. (2012). SLiCE: a Novel Bacterial Cell Extract-Based DNA Cloning Method. *Nucleic Acids Res.* 40 (8), e55. doi:10.1093/nar/gkr1288

Zhao, Y., Zhang, Y., Sun, H., Maroto, R., and Brasier, A. R. (2017). Selective Affinity Enrichment of Nitrotyrosine-Containing Peptides for Quantitative Analysis in Complex Samples. *J. Proteome Res.* 16 (8), 2983–2992. doi:10.1021/acs.jproteome.7b00275

Zhu, P., Gafken, P. R., Mehl, R. A., and Cooley, R. B. (2019). A Highly Versatile Expression System for the Production of Multiply Phosphorylated Proteins. *ACS Chem. Biol.* 14 (7), 1564–1572. doi:10.1021/acschembio.9b00307

Check for
updates

Non-Canonical Amino Acid-Based Engineering of (R)-Amine Transaminase

Amol D. Pagar[1], Hyunwoo Jeon[1], Taresh P. Khobragade[1], Sharad Sarak[1], Pritam Giri[1], Seonga Lim[1], Tae Hyeon Yoo[2], Byoung Joon Ko[3] and Hyungdon Yun[1]*

[1]Department of Systems Biotechnology, Konkuk University, Seoul, South Korea, [2]Department of Molecular Science and Technology, Ajou University, Suwon, South Korea, [3]School of Biopharmaceutical and Medical Sciences, Sungshin Women's University, Seoul, South Korea

Non-canonical amino acids (ncAAs) have been utilized as an invaluable tool for modulating the active site of the enzymes, probing the complex enzyme mechanisms, improving catalytic activity, and designing new to nature enzymes. Here, we report site-specific incorporation of *p*-benzoyl phenylalanine (*p*BpA) to engineer (R)-amine transaminase previously created from D-amino acid aminotransferase scaffold. Replacement of the single Phe88 residue at the active site with *p*BpA exhibits a significant 15-fold and 8-fold enhancement in activity for 1-phenylpropan-1-amine and benzaldehyde, respectively. Reshaping of the enzyme's active site afforded an another variant F86A/F88*p*BpA, with 30% higher thermostability at 55°C without affecting parent enzyme activity. Moreover, various racemic amines were successfully resolved by transaminase variants into (S)-amines with excellent conversions (~50%) and enantiomeric excess (>99%) using pyruvate as an amino acceptor. Additionally, kinetic resolution of the 1-phenylpropan-1-amine was performed using benzaldehyde as an amino acceptor, which is cheaper than pyruvate. Our results highlight the utility of ncAAs for designing enzymes with enhanced functionality beyond the limit of 20 canonical amino acids.

Keywords: (R)-amine transaminase, genetic code expansion, enzyme engineering, non-canonical amino acid, *p*-benzoyl-L-phenylalanine

INTRODUCTION

Chiral amines are valuable and versatile building blocks for the pharmaceutical, agricultural, and fine chemical industries (Mathew and Yun, 2012; Kelly et al., 2018; Patil et al., 2018). Due to the increasing environmental awareness, the development of new biocatalysts to produce optically pure amines is of primary interest (Fuchs et al., 2015; Slabu et al., 2017). Transaminases (TAs), a class of pyridoxal 5′-phosphate (PLP)-dependent enzymes represent an important biocatalysts that provides a green and cost-effective alternative to transition metal catalysts (Guo and Berglund, 2017). Transaminases (TAs) can be classified into two groups: α-transaminases (α-TAs) and ω-transaminases (ω-TAs) depending on the type of substrate that is converted (Steffen-Munsberg et al., 2015). α-TAs requires the presence of a carboxyl group at the α-position with respect to the carbonyl functionality (Rocha et al., 2019). On the other hand, ω-TAs can accept aliphatic ketones and amines as substrates (that is, not only α-keto acids and amino acids). ω-TAs can be further divided into two subgroups, β-TAs, and amine transaminases (ATAs), the latter being commonly used as a synonym for all ω-TAs (Rocha et al., 2019). High enantiospecificity, broad substrate scope,

and no need for external cofactor are the beneficial properties of TAs in an industrial context (Steffen-Munsberg et al., 2015). Being PLP-dependent enzymes, TAs can also be classified as (S)-selective ATAs belonging to fold-type I (aspartate transaminase superfamily) and (R)-selective ATAs belonging to fold-type IV (D-amino acid aminotransferase superfamily) (Slabu et al., 2017). The (S)-selective ATAs have been known and widely investigated for industrial use due to their abundance in nature (Łyskowski et al., 2014). In contrast, (R)-ATAs have recently been discovered and are less studied than their (S)-selective counterparts (Telzerow et al., 2019). The past decade has seen a surge in the interest in (R)-ATAs after successfully utilizing engineered (R)-ATA-117-Rd11 from Arthrobacter sp. KNK168 for the asymmetric synthesis of the antidiabetic drug (R)-sitagliptin (Iwasaki et al., 2006; Savile et al., 2010). (R)-ATAs catalyze the transfer of an amino group from (R)-aromatic or (R)-primary aliphatic amines to pyruvate producing ketones or aldehydes and D-Ala (Schätzle et al., 2011; Iwasaki et al., 2012; Sayer et al., 2014; Iglesias et al., 2017; Lakó et al., 2020). Thermodynamically, deamination of amines is preferable; nonetheless, amination of ketones is possible if a sufficient strategy to shift the equilibrium to amine synthesis is utilized (Schätzle et al., 2011; Iglesias et al., 2017; Telzerow et al., 2019, 2021). Several research groups contributed to this field by identification and characterization of new members and employing protein engineering tools to expand the substrate scope and enhance the thermostability (Iwasaki et al., 2006, 2012; Łyskowski et al., 2014; Guan et al., 2015; Hou et al., 2016; Pavkov-Keller et al., 2016; Bezsudnova et al., 2019; Zeifman et al., 2019; Cheng et al., 2020a; Telzerow et al., 2021). For instance, (R)-ATA from Exophiala xenobiotica has been identified to synthesize biaryl amines, which are considered privileged scaffolds for pharmaceuticals (Telzerow et al., 2019). In parallel, Bornscheuer and colleagues showed the one-pot synthesis of biaryl amine by combining enzymatic and chemical reaction utilizing engineered (R)-ATA from Aspergillus fumigatus (Dawood et al., 2018). Moreover, by employing an evolutionary approach, substrate scope and thermostability of (R)-ATAs from Mycobacterium vanbaalenii and Aspergillus terreus were enhanced, respectively (Cheng et al., 2020b; Liu et al., 2021). Also, a rationally designed variant V37A of newly characterized (R)-ATA from Luminiphilus syltensis showed broader substrate scope towards bulkier substrates (Konia et al., 2021). Therefore, it is of interest to identify, characterize and engineer more (R)-ATAs to provide broad diversity of applications in the chiral amine synthesis.

(R)-ATAs are not the only fold type IV PLP dependent enzymes. (Telzerow et al., 2021). D-amino acid aminotransferases (DATAs), L-branched chain aminotransferases (BCATs), and 4-Amino-4-deoxychorismate lyases (ADCLs) hold an overall structural similarity, therefore, belong to the same family (Telzerow et al., 2021). Still, their amino acid sequence, activity, stereo preference, and substate scope considerably differ (Bezsudnova et al., 2020). ADCLs catalyze the production of p-aminobenzoate from 4-amino-4-deoxychorismate as part of folate biosynthesis (Dai et al., 2013). On the other hand, the specificity of DATAs, BCATs is concerned mainly about α-amino and α-keto acids while (R)-ATAs concerns (R)-primary amines (Bezsudnova et al., 2020). Unlike (S)-ATAs,

class IV TAs are inert against polyamines and β-, γ-, and ω-amino acids (Bezsudnova et al., 2020). Besides their function, these different subfamilies can also be distinguished by the sequence-based motifs assigned by Bornscheuer and co-workers (Höhne et al., 2010). These motifs have advanced the identification and characterization of several new class IV TAs, thereby providing valuable insights about this fold type (Kelly et al., 2020). Among them (R)-ATAs are of particular interest due to their ability to synthesize enantiomerically pure (R)-amines from respective ketones (Ferrandi and Monti, 2018). Thus, it is essential to extend the knowledge of this important enzyme class by identifying and characterizing new members and structure-guided rational design approaches, thereby supporting protein engineering efforts to expand the toolbox of these industrially valuable enzymes (Bezsudnova et al., 2020)., To this end, Voss et al. demonstrated an evolutionary approach for the (R)-ATA from DATA using bioinformatic analysis combined with a computational redesign. A sextuple variant (Y31F/H86F/Y88F/H100L/S180A/T242I) was obtained with a specific activity of 326 mU mg^{-1} in the conversion of (R)-phenylethylamine (**1a**) (Voss et al., 2020). This study paves an alternate route to expand the toolbox and deepens our understanding of the class IV transaminases.

Although, protein engineering techniques such as rational design and directed evolution have found widespread applications in improving or altering the intrinsic activities of numerous enzymes (Arnold, 2018; Chen and Arnold, 2020). These strategies are primarily based on nature's alphabet of twenty canonical amino acids (cAAs) (Qu et al., 2020; Wang et al., 2021). However, less than half of cAAs have side chains with functional groups involved in enzymes' catalytic mechanisms (Green et al., 2016; Yu et al., 2021). Because of the limited chemical and physical repertoire of cAAs, not surprisingly, enzymes recruit reactive cofactors and post-transnationally modify existing amino acids in the active site (Yu et al., 2018). The advancements in the genetic code expansion (GCE) allow enhanced protein properties by introducing unique functional groups beyond nature's limited building blocks (Wang and Schultz, 2004; Young and Schultz, 2018). To this end, a series of orthogonal amino-acyl transfer RNA (tRNA) synthetase (aaRS)/tRNA pairs have been developed to encode distinct non-canonical amino acids (ncAAs) in vivo (Chin, 2014, 2017; Young and Schultz, 2018; Tseng et al., 2020; Koch et al., 2021; Lee et al., 2021). Over the last 2 decades, more than 200 ncAAs have been genetically encoded in prokaryotes and eukaryotes (Pagar et al., 2021). Apart from peptide modification, antibody development for pharmaceutical use, ncAAs has been widely applied in enzyme engineering research to illustrate the enzyme mechanisms, enhance enzyme activity, and even generate new catalytic mechanisms into protein scaffolds (Agostini et al., 2017; Won et al., 2019b; Drienovská and Roelfes, 2020; Giri et al., 2021; Pagar et al., 2021). The examples include ncAAs with metal chelating, photo-crosslinking, extended disulfide forming, orthogonal reactive, and unique pi-pi interaction properties (Pagar et al., 2021). Although progress in this field is fast, only a handful of examples have been reported for remodelling enzyme active sites by introducing single ncAA resulting in improved activity and substrate scope compared to cAAs

(Pagar et al., 2021). The major bottlenecks for the practical utilization of ncAAs in enzyme engineering are lower expression yields, high cost or unavailability of the ncAAs, background incorporation of cAAs, and limited structural diversity of ncAAs utilizing only orthogonal tyrosyl/pyrrolysyl-tRNA synthetases (Agostini et al., 2017; Drienovská and Roelfes, 2020). Nonetheless, continued efforts to overcome the above limitations open new doors in enzyme engineering and serve as a toolkit for evolution and designing enzymes with desired functionality (Pagar et al., 2021). In this work, we report ncAA-based engineering of (*R*)-ATA previously created from DATA scaffold. The rational incorporation of ncAA has significantly improved enzyme functionality. These results uncover the great potential of engineering enzymes with ncAAs for the efficient synthesis of chiral amines.

MATERIALS AND METHODS

Materials

The plasmids pDule-tfmF A65V S158A (ID: 85484) (Miyake-Stoner et al., 2010) and PylRS-AS (ID:137908) (Lee et al., 2016) were purchased from Addgene (Watertown, MA, United States). The gene for (*R*)-ATA and primers were synthesised and sequenced by BIONICS Co., Ltd. (Seoul, South Korea). The 2,3,4-triflourophenylalanine (F3F) was purchased from ChemImpex Inc (Wood Dale, IL, United States), *p*BpA and *p*-methylphenylalanine (*p*MeF) were purchased from BACHEM (Bubendorf, Switzerland), and *p*-trifluoromethylphenylalanine (*p*tFMF) from Alfa Aesar. Ni–NTA affinity columns were purchased from Qiagen (Valencia, CA, United States). All the other chemicals like pyruvate, acetophenone, pyridoxal $5'$-phosphate and amino donors were purchased from Sigma-Aldrich, Korea.

Site Directed Mutagenesis

The plasmid pET24ma harboring a gene for (*R*)-ATA was utilized as a template for inserting TAG codon as well as Ala mutations. The list of primers used is given in **Supplementary Table S1**. The thermal cycler was programmed as 1) initial denaturation at 95°C for 2 min, 2) 18 cycles of denaturation at 95°C for 30 s, annealing at 58°C for 30 s (depending on the T_m of the primers) and extension at 72°C for 5 min; 3) a final extension at 72°C for 5 min (Sarak et al., 2021). The resulting PCR product was digested with DpnI for 2 h at 37°C and transformed into chemically competent *E. coli* DH5-alpha cells. A single colony was picked from LB plates supplemented with Kanamycin and cultured overnight in 5 ml of LB media. The plasmids were isolated, and the gene was sequenced to confirm desired mutation.

ncAA Incorporation and Purification of the Mutant Enzymes

The plasmid pET24ma harboring the gene encoding for (*R*)-ATA and the TAG codon variants were individually co-transformed with PylRS-AS or pDule-tfmF A65V S158A or pEVOL-*p*BpARS (Park et al., 2018) into chemically competent *E. coli.* (BL-21) cells.

A single colony of host cells was incubated overnight into 5 ml LB-media supplemented with respective antibiotics. The selected ncAAs were incorporated into target enzyme by the reported protocols for each ncAAs (Chin et al., 2002; Jackson et al., 2007; Miyake-Stoner et al., 2010; Lee et al., 2016). The expression of *p*BpARS and F₃FPylRS was induced by 0.2% arabinose and expression of target enzyme was induced by 0.1 mM IPTG. The overexpressed cells were harvested by centrifugation at 5,000 rpm, 4°C for 10 min. The pellet was suspended in an appropriate volume of the lysis buffer (NaH₂PO₄ (50 mM), NaCl (300 mM), imidazole (5 mM, pH 8.0)). The cell suspension was then subjected to ultrasonic disruption by a horn-type sonicator (Sonics& Material Inc, United States). During sonication, the sample tube containing the cell mass suspension was held in an ice bath. The sonication was carried for the total duration of 30 min, with a duty cycle of 37.5%. The cell lysate was centrifuged at 17,000 rpm for 30 min and supernatant was loaded on Ni-NTA agarose resin column [BabyBio] with the flow rate of 1 ml/min. The non-target proteins were washed with 100 ml of washing buffer (50 mM NaH₂PO₄, 300 mM NaCl and 20 mM imidazole, pH 8.0) with flow rate of 2 ml/min. Next, the desired protein containing C-terminal hexa-His-tag was eluted using elution buffer 1 (50 mM NaH₂PO₄, 300 mM NaCl and 250 mM imidazole, pH 8.0). The eluted solution containing the purified protein was dialyzed against 20 mM Tris-HCl buffer (pH 8.0) and concentrated using an Amicon 30K ultrafiltration unit.

Determination of Enzyme Activity

All enzyme assays were carried out at 37°C and 100 mM Tris-HCl buffer (pH-9.0). Standard substrate conditions for activity assay were 10 mM (*R*)-1a, 10 mM pyruvate and 0.1 mM PLP. The typical reaction volume was 500 μL, and the enzyme reaction was stopped after 30 min by adding 500 μL 10% perchloric acid. Acetophenone produced was analysed by HPLC. Here, different concentrations of enzymes were used to measure the exact specific activity.

Determination of Substrate Specificity

Amino acceptor specificity was measured by using 10 mM (*R*)-1a as an amino donor and 10 mM amino acceptor at 37°C and 100 mM Tris Hcl buffer (pH-9.0) for 30 min. Acetophenone produced was analysed by HPLC. Amine substrate specificity was measured using 10 mM pyruvate and 10 mM respective amine compound (1a-1e) solubilized using 10% DMSO. Respective ketone product formed in each reaction after 30 min was analyzed by HPLC.

Analytical Conditions

Acetophenone and other ketone products were analyzed using a C₁₈ Symmetry column (Agilent) with an elution mixture of 50% methanol (0.1% TFA) and 50% water (0.1% TFA) at a flow rate of 1.0 ml/min at 244 nm detection wavelength. The quantitative analysis of amines was measured using HPLC with a Crownpak CR (Daicel Co., Japan) column at 210 nm with an elution of pH 1.5 perchloric acid solution (0.6 ml min⁻¹).
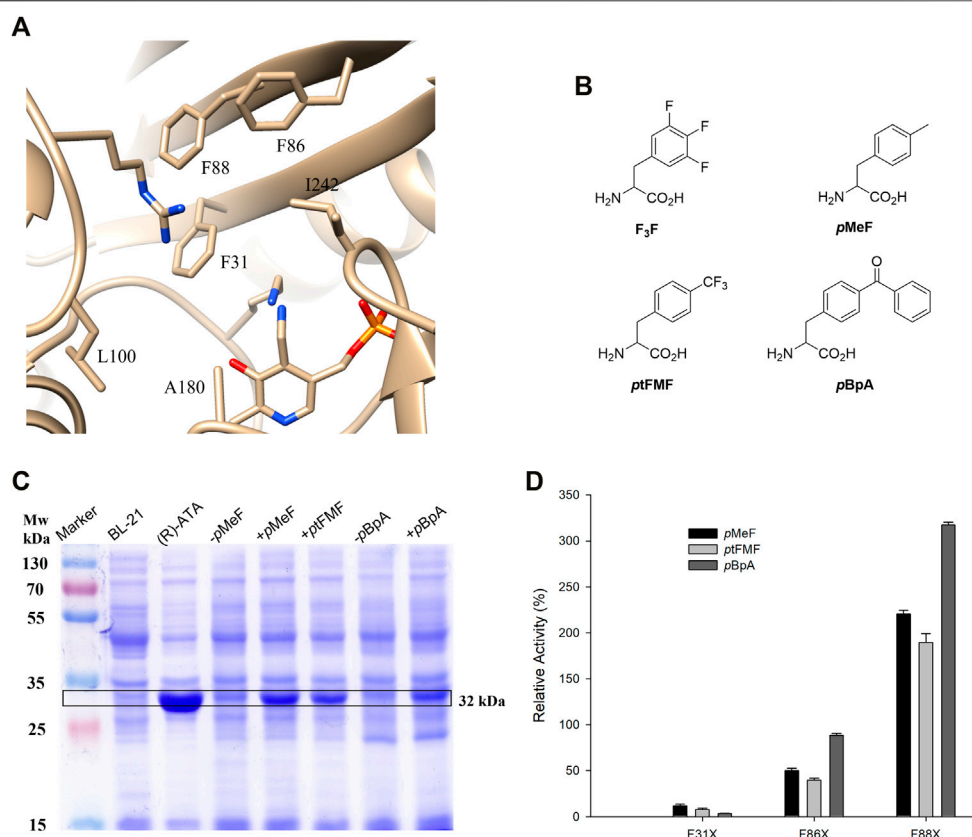
FIGURE 1 | (A) Active site residues of rationally designed (*R*)-ATA from DATA scaffold. **(B)** Structures of ncAAs selected for incorporation into active site of (*R*)-ATA **(C)** Coomassie-stained SDS-PAGE showing expression of proteins in presence and absence of ncAAs; **(D)** Relative activities of variants containing three selected ncAAs at 31st, 86th and 88th position. The specific activity of parent (*R*)-ATA (324 mU mg^{-1}) was taken as 100%.

RESULTS AND DISCUSSION

Incorporation of ncAAs Into (*R*)-ATA

Class IV TAs are promising catalysts for synthesizing enantiomerically pure amines and D-amino acids (Bezsudnova et al., 2020). The development of (*R*)-ATA for manufacturing sitagliptin demonstrated the practical viability of (*R*)-ATAs for asymmetric synthesis on an industrial scale (Savile et al., 2010). It sparked new research in the field of (*R*)-ATAs in synthetic applications. Yet, narrow ketone substrate range, unfavourable reaction equilibria, and strong ketone inhibition limit the applicability of these enzymes (Steffen-Munsberg et al., 2015; Guo and Berglund, 2017). In conjunction with identifying new TAs, structure-guided rational design approaches could help overcome these bottlenecks (Steffen-Munsberg et al., 2015). To this end, Voss et al. studied two members from fold type IV, BCAT and DATA, and their respective variants to accept the benchmark amine substrate 1a. By employing bioinformatic analysis combined with the computational redesign, a DATA variant (M2-6) harbouring six mutations (Y31F/H86F/Y88F/ H100L/S180A/T242I) (**Figure 1A**), was generated yielding a total activity of 326 mU mg^{-1} toward 1a. A significant increase

in active site volume by non-polar side chain mutations contributed to the favourable binding of bulkier and hydrophobic 1a over D-alanine. Despite bearing moderate (*R*)-ATA activity, the rational engineering of DATA deepens our knowledge of how substrate specificity in α-ATA is affected and can be altered toward accepting arylamines. Notably, three among six mutations introduced were Y31F/H86F/Y88F, vastly contributing to the active site's hydrophobicity. The rational design approach based on cAAs limits to create a more hydrophobic environment into the active site as Phe is highly hydrophobic among all cAAs. The GCE strategies allow the incorporation of more than 200 ncAAs with different functional properties into proteins (Won et al., 2019b; Pagar et al., 2021). Therefore, we speculate that substitution of Phe by a ncAA with a more hydrophobic side chain could enhance the activity of this rationally designed enzyme. Initially, we selected four Phe analogs, namely, F$_3$F, *p*tFMF, *p*-methylphenylalanine *p*MeF, and *p*BpA, based on commercial availability and hydrophobicity of the side chains (**Figure 1B**). An amber stop codon was substituted for three individual Phe residues (F31, F86, and F88). Initially, the incorporation efficiencies of selected ncAAs were determined by incorporating each ncAA
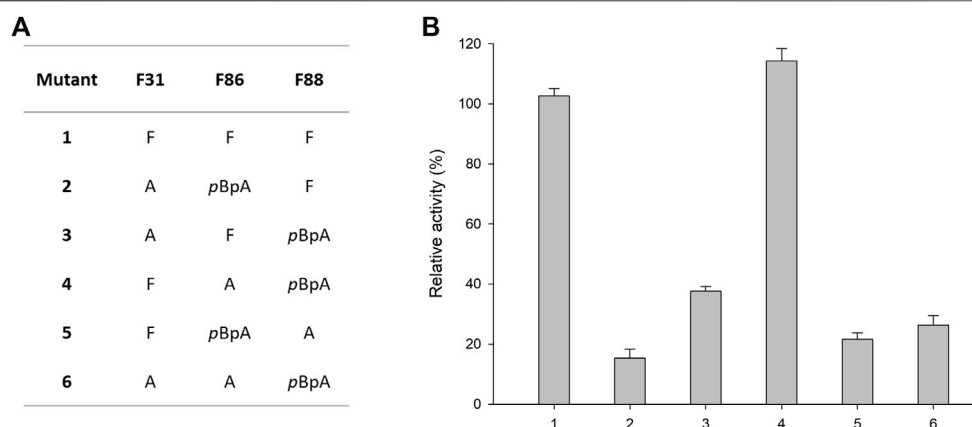
FIGURE 2 | Alanine scanning of the active site of (R)-ATA **(A)** Mutations introduced into active site of (R)-ATA having one pBpA and another Phe mutated to Ala. **(B)** The relative activity of the pBpA/Ala mutants towards 1a. The specific activity of the parent (R)-ATA (324 mU mg^{-1}) was taken as 100%.

at the 31st position using available orthogonal tRNA synthetase/tRNA pairs evolved for each ncAA (**Supplementary Table S2**). The incorporation efficiency of F$_3$F using evolved *Methanosarcina mazei* pyrrolysyl-tRNA synthetase/tRNA pair was comparatively poor (**Supplementary Figure S1**), and media shift, high concentration of ncAA (5 mM) longer expression time, is quite unpractical for obtaining higher concentrations of the mutant enzyme. The tRNA synthetase/tRNA pair evolved from *Methanococcus jannaschii* tyrosyl tRNA synthetase for pMeF, ptFMF and pBpA incorporation showed better incorporation efficiency and fidelity (**Figure 1C**). Therefore, total nine mutants containing each of three ncAAs at 31st, 86th and 88th position were purified along with their parent (R)-ATA and subjected to an activity assay, which was conducted using 10 mM 1a and 10 mM pyruvate in 100 mm Tris/HCl buffer (pH 9.0). One unit was defined as the amount of enzyme that catalysed the formation of 1 µmol of acetophenone per minute. Incorporation of ncAAs at 31st position showed deleterious effects on the TA activity, implying F31 is highly conserved. Remarkably, replacing F88 with all three ncAAs displayed higher activities than the parent enzyme with highest 3-fold activity by F88pBpA (**Figure 1C**). This enhanced activity was in close agreement with the previous finding where replacement of the Y88 with Phe showed a beneficial effect on (R)-PEA acceptance (Voss et al., 2020). On the other hand, replacing F86 with pMeF and ptFMF showed ~50 and 61% reduction in activity, respectively, while pBpA hardly affected the enzyme activity.

Reshaping of the Active Site

Encouraged by the above results, we further aimed to reshape the active site of (R)-ATA with ncAA. Three Phe residues provide enough hydrophobicity to the active site but significantly compromise the substrate-binding pocket's size and thus narrow the substrate scope. Therefore, we envisioned that the hydrophobicity provided by single pBpA introduced into the

active site might compensate for Phe residue's replacement. Ala and Phe have hydrophobic side chains, but the earlier is much smaller and more flexible than the latter. Therefore, an Ala scan was performed on three selected Phe residues while retaining one residue as pBpA. Mutant combinations with F31pBpA were omitted owing to the detrimental effect of pBpA substitution at 31st position (**Figure 1C**). Total five mutants were designed (**Figure 2A**), and activities of the resulting variants were measured against 1a. Significant reductions in enzyme activity toward 1a were observed, apparently due to the impairment of the hydrophobic environment required to recognize the phenyl group of 1a (**Figure 2B**). In contrast, variant F86A/F88pBpA showed activity comparable to the parent enzyme, suggesting F31 and F88 residues play an essential role in recognizing the phenyl ring of 1a. Additionally, the loss of hydrophobicity by replacing F86 with Ala was compensated by pBpA, thereby retaining the
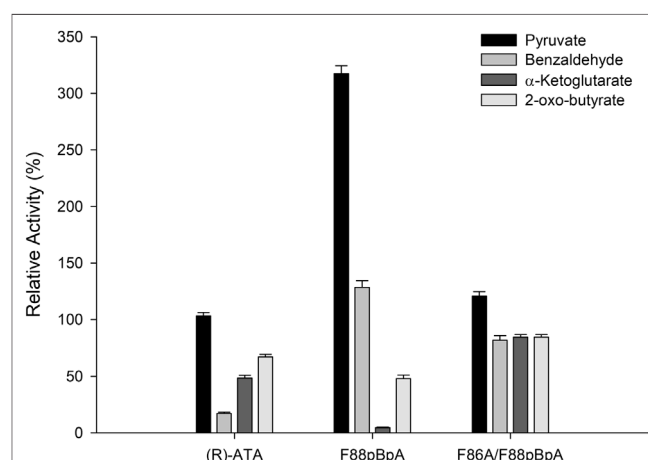

FIGURE 3 | Substrate specificities of the (R)-ATA and its variants towards various amino acceptors. (R)-ATA activity towards pyruvate (324 mU mg^{-1}) was taken as 100%.
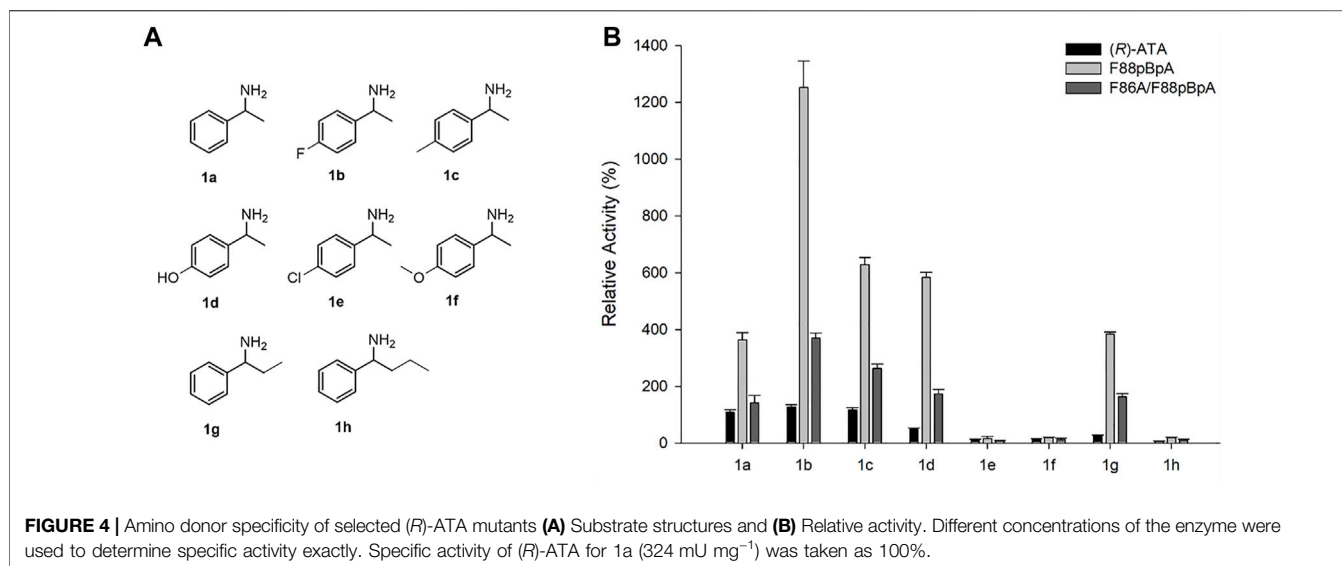
FIGURE 4 | Amino donor specificity of selected (*R*)-ATA mutants **(A)** Substrate structures and **(B)** Relative activity. Different concentrations of the enzyme were used to determine specific activity exactly. Specific activity of (*R*)-ATA for 1a (324 mU mg^{-1}) was taken as 100%.

overall enzyme activity. The variant F86A/F88*p*BpA was taken for further characterization.

Substrate Specificity of the Designed Variants

Next, the substrate specificities of parent (*R*)-ATA, F88*p*BpA, and F86A/F88*p*BpA were investigated using a series of commonly used amino acceptors like pyruvate, benzaldehyde, α-ketoglutarate, and 2-oxobutyrate, using 1a as an amino donor. Pyruvate served as the best amino acceptor for all the variants, and F88*p*BpA showed a specific activity of 1.03 U mg^{-1}. Interestingly, F88*p*BpA and F86A/F88*p*BpA showed ~8- and 5-fold higher activities respectively, for benzaldehyde than the parent enzyme (**Figure 3**). This increased activity could be attributed to the enhanced acceptance of aromatic substrates by *p*BpA incorporation. As a cheaper amino acceptor than pyruvate, benzaldehyde can be suitable for TA reactions. In contrast, the activity of F88*p*BpA variant towards α-ketoglutarate has been drastically reduced while F86A/F88*p*BpA retained its activity towards all amino acceptors.

Moreover, the substrate specificities of variants were examined for various substituted arylamines (1a-h) as shown in **Figure 4A** using pyruvate as an amino acceptor. The respective ketone product formed in 10 mM reaction was quantified by HPLC. While parent enzyme showed mediocre activities towards 1a-d and 1g, F88*p*BpA and F86A/F88*p*BpA showed significantly enhanced activities (**Figure 4B**). Notably, 1g carrying a 26% parental reactivity relative to 1a showed the highest ~15-fold enhancement in activity upon F88*p*BpA substitution. Notably, parent (*R*)-ATA and its variants were able to accept ethyl side chain into P-pocket (small-binding pocket) which is uncommon among most of the naturally occurring (*R*)-ATAs (Cheng et al., 2020b; Konia et al., 2021). In addition, ~12-fold enhanced activity for 1d which is carrying *p*-hydroxy substitution, suggests that even a polar substituted aryl amine's acceptance has been

improved by F88*p*BpA substitution. Unfortunately, none of the mutant showed activity for 1e, 1f and 1h. F86A/F88*p*BpA was expected to have a broader substrate scope due to enlarged active site by F86A mutation. Nevertheless, more active site remodelling efforts are still necessary to improve the substrate scope of this rationally designed enzyme.

Effect of pBpA Incorporation on Stability

It has been observed that improved activity or substrate specificity of enzymes leads to a trade-off between stability in many instances (Siddiqui, 2017; Yu and Dalby, 2018). However, it is worth mentioning that the (*R*)-ATA utilized herein was previously created from the DATA from *Bacillus subtilis* without affecting overall stability (Voss et al., 2020). Incorporation of halogenated ncAAs by selective pressure incorporation or site-specific incorporation in several instances has enhanced the stability of enzymes (Hoesl et al., 2011; Deepankumar et al., 2014; Carlsson et al., 2018; Ohtake et al., 2018; Won et al., 2019a). Nevertheless, incorporation of single *p*BpA has not yet been reported for improved stability. Therefore, we were curious whether incorporating *p*BpA into (*R*)-ATA affected the stability. To this end, we examined the residual activity of the variants by incubating enzyme samples at various temperatures (37–60°C) in the presence of 100 mM Tris HCl (pH-9.0) for 30 min (**Figure 5A**). Unfortunately, the most active variant F88*p*BpA and the parent enzyme lost almost 45% of their intrinsic activity after incubating at 55°C, while F86A/F88*p*BpA retained 85% of its original activity. These results demonstrate that enhanced activity does not necessarily lead to a trade-off with stability as best variant F88*p*BpA showed a similar stability profile with the parent enzyme. On the other hand, F86A/F88*p*BpA showed greater thermostability than the parent enzyme, indicating that it is indeed possible to tune the activity and stability profiles of the enzymes using ncAAs. Moreover, the thermal stability of enzyme variants was examined under different conditions because 1a diminishes stability due to
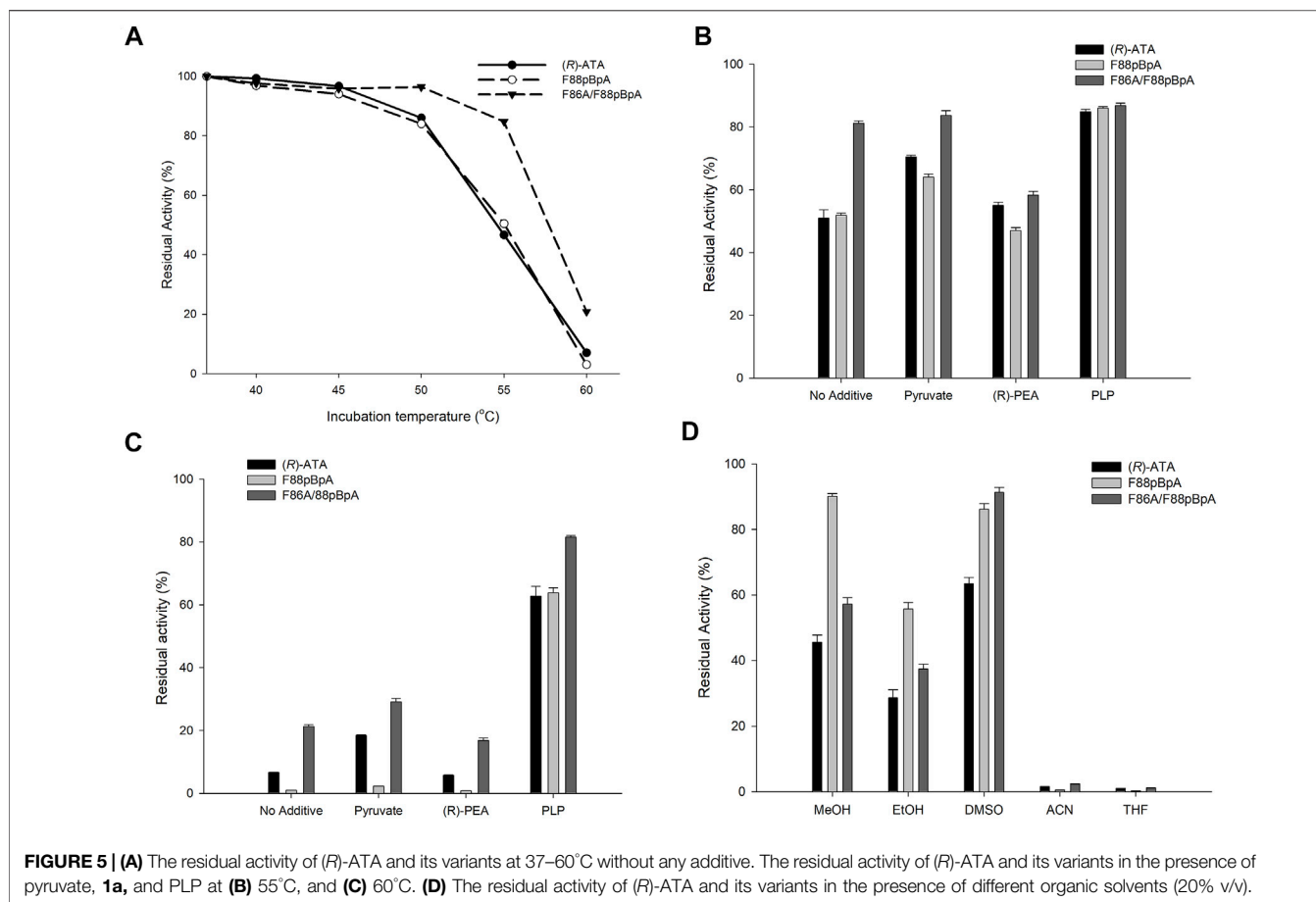
FIGURE 5 | (A) The residual activity of (R)-ATA and its variants at 37–60°C without any additive. The residual activity of (R)-ATA and its variants in the presence of pyruvate, **1a,** and PLP at **(B)** 55°C, and **(C)** 60°C. **(D)** The residual activity of (R)-ATA and its variants in the presence of different organic solvents (20% v/v).

formation of E-PMP (Deepankumar et al., 2014; Mathew et al., 2016). In contrast, pyruvate and PLP have a beneficial effect on stability. The residual activities of enzymes were further examined after 30 min of incubation in 100 mM Tris/HCl buffer (pH-9.0) at 55 and 60°C and in the presence of 10 Mm 1a, 10 mM pyruvate, or 0.1 mM PLP. Interestingly parent (R)-ATA and F88pBpA regain their stability at 55°C in presence of PLP displaying 83% overall activity similar to F86A/F88pBpA (**Figure 5B**). Next, the enzyme activities were significantly decreased at 60°C in presence of 1a and without additive, whereas PLP showed stabilizing effect by forming E-PLP complex. Variant F86A/F88pBpA retained highest 81% of activity where native enzyme and F88pBpA lost ~35% of their original activity at 60°C (**Figure 5C**). Owing to the poor solubility of keto or amine substrates in aqueous buffer solutions, organic solvents are generally used as a co-solvent. Therefore, the stability of enzyme variants was also examined in presence of various organic solvents (20% v/v) like methanol, ethanol, DMSO, acetonitrile and tetrahydrofuran. All the variants showed better stability after incubation in presence of DMSO and methanol whereas, complete loss of activity was observed in presence of acetonitrile and tetrahydrofuran (**Figure 5D**). In agreement with enhanced stability, F86A/F88pBpA also showed better organic solvent tolerance in presence of 20% DMSO than other variants.

Kinetic Resolution of Amines

Incorporation of ncAAs into enzymes enables a dramatic expansion of their catalytic features. However, the biocatalytic use of enzymes containing ncAAs lacks behind the use of ncAAs as tools for therapy and research (Pagar et al., 2021). Low expression levels, high cost or unavailability of ncAAs, unfolded enzymes, incompletely translated enzymes and background incorporation of the cAAs are the major bottlenecks for the GCE method from being a standard method for enzyme engineering (Drienovská et al., 2020). Expression yield of the protein containing ncAA also reflects the degree by which mutation is tolerated (Kolev et al., 2014). For instance, the expression yields of F88pBpA and F86A/F88pBpA were 32–35 and 68–70 mg/L respectively whereas parent (R)-ATA co-expressed with pBpARS yielded ~134–140 mg/L of protein. Also, affinity purification is rather unpractical for screening of large set of mutants or large scale biocatalytic process. Therefore, we performed the kinetic resolution of racemic amines using whole cell despite the fact that expression levels of soluble enzymes harboring pBpA will be less than parent (R)-ATA (**Figure 1C**). Since, pyruvate was a good amino acceptor for all variants, the kinetic resolution of 20 mM 1a-1d and 1g was carried out with equimolar concentration of pyruvate (**Supplementary Figure S2–6**). As shown in **Table 1** after 24 h reaction, 50% conversion of 1g with >99% ee was

TABLE 1 | Kinetic resolution of various amine compounds using parent (R)-ATA and its variants.

Substrate	Parent (R)-ATA		F88pBpA		F86A/F88pBpA	
	Conversion (%)	ee (%)	Conversion (%)	ee (%)	Conversion (%)	ee (%)
1a	22	28	17	21	27	38
1b	50	>99	44	78	50	>99
1c	38	60	36	57	43	76
1d	34	52	41	70	40	68
1 g	36	56	45	83	50	>99
1 g*	50	>99	50	>99	50	>99

Reaction conditions: Reaction Vol. 1 ml. 20 mM rac-(1a-1d and 1g), 20 mM pyruvate, 1.98 mg$_{CDW}$/mL E. coli, 0.1 mM PLP, 100 mM Tris-HCl buffer (pH-9.0) at 37°C. *kinetic resolution of 10 mM rac-1g using 20 mM benzaldehyde as an amino acceptor instead of pyruvate.

achieved by F86A88pBpA whereas, F88pBpA and parent (R)-ATA showed only 45 and 36% conversion, respectively. Mutant F86A/F88pBpA performed well than F88pBpA for all the substrates perhaps due to its higher stability and expression levels. For the kinetic resolution of amines, it is necessary to add stoichiometric quantities of pyruvate which is used as an amino acceptor. However, pyruvate itself is an expensive compound and thus increases the overall cost of the biocatalytic production of amines (Shin et al., 2015). Interestingly, F88pBpA showed considerable reactivity towards benzaldehyde (**Figure 3**) as 463 mU mg^{-1} suggesting that very cheap benzaldehyde can be used as a good amino acceptor. Therefore, kinetic resolution of 10 mM 1g was performed using 20 mM benzaldehyde. Highly enantiomerically pure (S)-1g with 50% conversion and >99% ee was obtained by parent (R)-ATA and its variants (**Supplementary Figure S7**). This result clearly demonstrates the applicability of mutant TAs using benzaldehyde as a cheap amino acceptor in a kinetic resolution reaction system for producing optically pure amines.

The kinetic resolution is not an efficient way to utilize (R)-ATAs for synthesizing (S)-amines. Several (S)-selective ATAs can be employed to produce (S)-enantiomers with a theoretical yield up to 100% by asymmetric synthesis (Mathew and Yun, 2012; Slabu et al., 2017; Kelly et al., 2018; Patil et al., 2018). Therefore, we tried to perform (R)-ATA catalyzed asymmetric synthesis of (R)-1g from propiophenone using benzylamine as an amino donor. Though enzymes utilized in this study were reactive for deamination of (R)-1g, they could not perform the reverse reaction, i.e., amination of the propiophenone. This indicates that, unlike most of the ω-TAs, enzyme variants utilized herein do not follow the "good donor-acceptor pair relationship" (Han et al., 2017). However, we believe that F88pBpA or F86A/F88pBpA may serve as a good template for protein engineering to obtain efficient biocatalysts to produce bulky (R)-amines.

DISCUSSION

ATAs are the PLP dependent enzymes which catalyze the transfer of an amino group from an amino donor to a carbonyl moiety and vice versa (Steffen-Munsberg et al., 2015). Transaminase reactions gain popularity for the production of chiral amines because of their outstanding optical purity, high yield, broad substrate specificity, and environmental friendliness (Patil et al., 2018). On account of

their importance in asymmetric biocatalysis, the identification and characterization of new members and engineering efforts to improve their functional properties are of utmost importance (Steffen-Munsberg et al., 2015).

Transaminase classification has recently been revised (Gao et al., 2017). To date, only a few (R)-ATA are identified, and their crystal structures were made available, allowing researchers to get insight into the enantioselectivity and explore the catalytic mechanism compared to their (S)-selective counterparts (Konia et al., 2021). Furthermore, researchers used molecular modeling and the structure-activity relationship to create the enzymes (Voss et al., 2020). However, problems like an unfavorable equilibrium constant and the inability to accept bulky substrates still need to be addressed (Bezsudnova et al., 2020). Although engineering (R)-ATA from Arthrobacter sp. was a noteworthy, the rational quest for more relevant (R)-ATAs is still ongoing because it will lay the groundwork for these enzymes' future applications (Savile et al., 2010).

The chemical modification and ncAA incorporation methods have emerged as an important alternative to the traditional enzyme engineering approaches like directed evolution and rational design (Giri et al., 2021; Pagar et al., 2021). These methods used independently or together have dramatically expanded the chemical diversity for proteins, which has provided protein engineers with powerful tools for enzymes engineering (Pagar et al., 2021). Over the past few decades, numerous studies have reported increased enzyme activity and stability by incorporating the ncAAs (Agostini et al., 2017; Drienovská and Roelfes, 2020). In a pioneering example, Jackson and others demonstrated that site-specific incorporation of ncAAs can be used to redesign the enzyme's active site for diverse substrates, resulting in a ~30-fold increased activity that cannot be reproduced by substituting any other cAA (Jackson et al., 2006).

In this study, we demonstrated the ncAA-based engineering approach to enhance the functionality of rationally designed (R)-ATA. Based on the preliminary knowledge that hydrophobicity is a major contributing factor for the creation of (R)-ATA activity, we explored it further using ncAA incorporation. The designed variant containing pBpA at 88th position has significantly enhanced the activity towards several arylamine substrates. Moreover, another engineered variant F86A/F88pBpA showed enhanced thermostability and organic solvent tolerance. These results clearly demonstrate that a wise selection of ncAAs for rational engineering of the enzymes can render beneficial results

for functional enhancement. However, further engineering efforts are still needed to improve the substrate scope of this enzyme. To this goal, simultaneous incorporation of ncAAs using mutually orthogonal aminoacyl tRNA synthetases/tRNA pairs for the incorporation of distinct ncAAs into one protein could be beneficial (Neumann et al., 2010; Italia et al., 2019). Moreover, directed evolution with ncAAs can also be utilized to enlarge the chemical and sequence space of proteins and, in turn, increase the probability of evolving the desired mutant (Pagar et al., 2021; Pan et al., 2021).

The main obstacle in the ncAA-based enzyme engineering and directed evolution approach is exogenous supplementation of often expensive ncAA into the growth medium (Pagar et al., 2021). This increases the overall production cost of the ncAA containing enzymes. Autonomous biosynthesis of ncAAs and their concurrent incorporation into enzyme of interest *in vivo* could significantly reduce the production cost and permeability issues (Pagar et al., 2021). Using enzymatic and metabolic pathways, some ncAAs like *p*-amino-phenylalanine, 5-hydroxytryptophan, L-phosphothreonine, L-dihydroxyphenylalanine, fluorotyrosine, and *S*-allylcysteine were biosynthesized in *E. coli* and concurrently incorporated into target proteins (Ma et al., 2014; Exner et al., 2017; Kim et al., 2018; Won et al., 2019a; Nojoumi et al., 2019; Schipp et al., 2020). The production of an increasing number of ncAAs in engineered cells will be aided by advances in biochemistry, molecular biology, and synthetic biology. In addition, advancements in ncAA mutagenesis procedures may drive ncAAs more valuable in protein engineering and enzyme evolution (Pagar 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

AP: Experiment, methodology, data collection and analysis, writing original draft preparation. HJ, TK, SS, PG, LS: Methodology and experiment. TY and BK: review and editing, HY: Conceptualization, review and editing, project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.839636/full#supplementary-material

REFERENCES

Agostini, F., Völler, J. S., Koksch, B., Acevedo-Rocha, C. G., Kubyshkin, V., and Budisa, N. (2017). Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed.* 56, 9680–9703. doi:10.1002/anie.201610129

Arnold, F. H. (2018). Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed.* 57, 4143–4148. doi:10.1002/anie.201708408

Bezsudnova, E. Y., Boyko, K. M., Nikolaeva, A. Y., Zeifman, Y. S., Rakitina, T. V., Suplatov, D. A., et al. (2019). Biochemical and Structural Insights into PLP Fold Type IV Transaminase from Thermobaculum Terrenum. *Biochimie* 158, 130–138. doi:10.1016/j.biochi.2018.12.017

Bezsudnova, E. Y., PopovPopov, V. O. A., and Boyko, K. M. (2020). Structural Insight into the Substrate Specificity of PLP Fold Type IV Transaminases. *Appl. Microbiol. Biotechnol.* 104, 2343–2357. doi:10.1007/s00253-020-10369-6

Carlsson, A.-C. C., Scholfield, M. R., Rowe, R. K., Ford, M. C., Alexander, A. T., Mehl, R. A., et al. (2018). Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds. *Biochemistry* 57, 4135–4147. doi:10.1021/acs.biochem.8b00603

Chen, K., and Arnold, F. H. (2020). Engineering New Catalytic Activities in Enzymes. *Nat. Catal.* 3, 203–213. doi:10.1038/s41929-019-0385-5

Cheng, F., Chen, X.-L., Li, M.-Y., Zhang, X.-J., Jia, D.-X., Wang, Y.-J., et al. (2020a). Creation of a Robust and R-Selective ω-amine Transaminase for the Asymmetric Synthesis of Sitagliptin Intermediate on a Kilogram Scale. *Enzyme Microb. Tech.* 141, 109655. doi:10.1016/j.enzmictec.2020.109655

Cheng, F., Chen, X.-L., Xiang, C., Liu, Z.-Q., Wang, Y.-J., and Zheng, Y.-G. (2020b). Fluorescence-based High-Throughput Screening System for R-ω-Transaminase Engineering and its Substrate Scope Extension. *Appl. Microbiol. Biotechnol.* 104, 2999–3009. doi:10.1007/s00253-020-10444-y

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

Dai, Y.-N., Chi, C.-B., Zhou, K., Cheng, W., Jiang, Y.-L., Ren, Y.-M., et al. (2013). Structure and Catalytic Mechanism of Yeast 4-Amino-4-Deoxychorismate Lyase. *J. Biol. Chem.* 288, 22985–22992. doi:10.1074/jbc.M113.480335

Dawood, A. W. H., Bassut, J., de Souza, R. O. M. A., and Bornscheuer, U. T. (2018). Combination of the Suzuki-Miyaura Cross-Coupling Reaction with Engineered Transaminases. *Chem. Eur. J.* 24, 16009–16013. doi:10.1002/chem.201804366

Deepankumar, K., Shon, M., Nadarajan, S. P., Shin, G., Mathew, S., Ayyadurai, N., et al. (2014). Enhancing Thermostability and Organic Solvent Tolerance of ω-Transaminase through Global Incorporation of Fluorotyrosine. *Adv. Synth. Catal.* 356, 993–998. doi:10.1002/adsc.201300706

D. Patil, M., Grogan, G., Bommarius, A., and Yun, H. (2018). Recent Advances in ω-Transaminase-Mediated Biocatalysis for the Enantioselective Synthesis of Chiral Amines. *Catalysts* 8, 254. doi:10.3390/catal8070254

Drienovská, I., Gajdoš, M., Kindler, A., Takhtehchian, M., Darnhofer, B., Birner-Gruenberger, R., et al. (2020). Folding Assessment of Incorporation of Noncanonical Amino Acids Facilitates Expansion of Functional-Group Diversity for Enzyme Engineering. *Chem. Eur. J.* 26, 12338–12342. doi:10.1002/chem.202002077

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

Exner, M. P., Kuenzl, T., To, T. M. T., Ouyang, Z., Schwagerus, S., Hoesl, M. G., et al. (2017). Design ofS-Allylcysteine *In Situ* Production and Incorporation Based on a Novel Pyrrolysyl-tRNA Synthetase Variant. *ChemBioChem* 18, 85–90. doi:10.1002/cbic.201600537

Ferrandi, E. E., and Monti, D. (2018). Amine Transaminases in Chiral Amines Synthesis: Recent Advances and Challenges. *World J. Microbiol. Biotechnol.* 34, 1–10. doi:10.1007/s11274-017-2395-2

Fuchs, M., Farnberger, J. E., and Kroutil, W. (20152015). The Industrial Age of Biocatalytic Transamination. *Eur. J. Org. Chem.* 2015, 6965–6982. doi:10.1002/ejoc.201500852

Gao, S., Su, Y., Zhao, L., Li, G., and Zheng, G. (2017). Characterization of a (R)-selective Amine Transaminase from Fusarium Oxysporum. *Process Biochem.* 63, 130–136. doi:10.1016/j.procbio.2017.08.012

Giri, P., Pagar, A. D., Patil, M. D., and Yun, H. (2021). Chemical Modification of Enzymes to Improve Biocatalytic Performance. *Biotechnol. Adv.* 53, 107868. doi:10.1016/j.biotechadv.2021.107868

Green, A. P., Hayashi, T., Mittl, P. R. E., and Hilvert, D. (2016). A Chemically Programmed Proximal Ligand Enhances the Catalytic Properties of a Heme Enzyme. *J. Am. Chem. Soc.* 138, 11344–11352. doi:10.1021/jacs.6b07029

Guan, L.-J., Ohtsuka, J., Okai, M., Miyakawa, T., Mase, T., Zhi, Y., et al. (2015). A New Target Region for Changing the Substrate Specificity of Amine Transaminases. *Sci. Rep.* 5, 1–8. doi:10.1038/srep10753

Guo, F., and Berglund, P. (2017). Transaminase Biocatalysis: Optimization and Application. *Green. Chem.* 19, 333–360. doi:10.1039/c6gc02328b

Han, S.-W., Kim, J., Cho, H.-S., and Shin, J.-S. (2017). Active Site Engineering of ω-Transaminase Guided by Docking Orientation Analysis and Virtual Activity Screening. *ACS Catal.* 7, 3752–3762. doi:10.1021/acscatal.6b03242

Hoesl, M. G., Acevedo-Rocha, C. G., Nehring, S., Royter, M., Wolschner, C., Wiltschi, B., et al. (2011). Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem* 3, 213–221. doi:10.1002/cctc.201000253

Höhne, M., Schätzle, S., Jochens, H., Robins, K., and Bornscheuer, U. T. (2010). Rational Assignment of Key Motifs for Function Guides In Silico Enzyme Identification. *Nat. Chem. Biol.* 6, 807–813. doi:10.1038/nchembio.447

Hou, A., Deng, Z., Ma, H., and Liu, T. (2016). Substrate Screening of Amino Transaminase for the Synthesis of a Sitagliptin Intermediate. *Tetrahedron* 72, 4660–4664. doi:10.1016/j.tet.2016.06.039

Iglesias, C., Panizza, P., and Rodriguez Giordano, S. (2017). Identification, Expression and Characterization of an R-ω-Transaminase from Capronia Semiimmersa. *Appl. Microbiol. Biotechnol.* 101, 5677–5687. doi:10.1007/s00253-017-8309-2

Italia, J. S., Addy, P. S., Erickson, S. B., Peeler, J. C., Weerapana, E., and Chatterjee, A. (2019). Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J. Am. Chem. Soc.* 141, 6204–6212. doi:10.1021/jacs.8b12954

Iwasaki, A., Matsumoto, K., Hasegawa, J., and Yasohara, Y. (2012). A Novel Transaminase, (R)-amine:pyruvate Aminotransferase, from Arthrobacter Sp. KNK168 (FERM BP-5228): Purification, Characterization, and Gene Cloning. *Appl. Microbiol. Biotechnol.* 93, 1563–1573. doi:10.1007/s00253-011-3580-0

Iwasaki, A., Yamada, Y., Kizaki, N., Ikenaka, Y., and Hasegawa, J. (2006). Microbial Synthesis of Chiral Amines by (R)-specific Transamination with Arthrobacter Sp. KNK168. *Appl. Microbiol. Biotechnol.* 69, 499–505. doi:10.1007/s00253-005-0002-1

Jackson, J. C., Duffy, S. P., Hess, K. R., and Mehl, R. A. (2006). Improving Nature's Enzyme Active Site with Genetically Encoded Unnatural Amino Acids. *J. Am. Chem. Soc.* 128, 11124–11127. doi:10.1021/ja061099y

Jackson, J. C., Hammill, J. T., and Mehl, R. A. (2007). Site-specific Incorporation of a 19F-Amino Acid into Proteins as an NMR Probe for Characterizing Protein Structure and Reactivity. *J. Am. Chem. Soc.* 129, 1160–1166. doi:10.1021/ja064661t

Kelly, S. A., Mix, S., Moody, T. S., and Gilmore, B. F. (2020). Transaminases for Industrial Biocatalysis: Novel Enzyme Discovery. *Appl. Microbiol. Biotechnol.* 104, 4781–4794. doi:10.1007/s00253-020-10585-0

Kelly, S. A., Pohle, S., Wharry, S., Mix, S., Allen, C. C. R., Moody, T. S., et al. (2018). Application of ω-Transaminases in the Pharmaceutical Industry. *Chem. Rev.* 118, 349–367. doi:10.1021/acs.chemrev.7b00437

Kim, S., Sung, B. H., Kim, S. C., and Lee, H. S. (2018). Genetic Incorporation Ofl-Dihydroxyphenylalanine (DOPA) Biosynthesized by a Tyrosine Phenol-Lyase. *Chem. Commun.* 54, 3002–3005. doi:10.1039/c8cc00281a

Koch, N. G., Goettig, P., Rappsilber, J., and Budisa, N. (2021). Engineering Pyrrolysyl-Trna Synthetase for the Incorporation of Non-canonical Amino Acids with Smaller Side Chains. *Ijms* 22, 11194. doi:10.3390/ijms222011194

Kolev, J. N., Zaengle, J. M., Ravikumar, R., and Fasan, R. (2014). Enhancing the Efficiency and Regioselectivity of P450 Oxidation Catalysts by Unnatural Amino Acid Mutagenesis. *ChemBioChem* 15, 1001–1010.

Konia, E., Chatzicharalampous, K., Drakonaki, A., Muenke, C., Ermler, U., Tsiotis, G., et al. (2021). Rational Engineering of Luminiphilus Syltensis (R)-selective Amine Transaminase for the Acceptance of Bulky Substrates. *Chem. Commun.* 57, 12948–12951. doi:10.1039/d1cc04664k

Lakó, Á., Molnár, Z., Mendonça, R., and Poppe, L. (2020). Transaminase-mediated Synthesis of Enantiopure Drug-like 1-(3′,4′-disubstituted Phenyl)propan-2-Amines. *RSC Adv.* 10, 40894–40903. doi:10.1039/d0ra08134e

Lee, B. S., Choi, W. J., Lee, S. W., Ko, B. J., and Yoo, T. H. (2021). Towards Engineering an Orthogonal Protein Translation Initiation System. *Front. Chem.* 9, 1–9. doi:10.3389/fchem.2021.772648

Lee, Y.-J., Schmidt, M. J., Tharp, J. M., Weber, A., Koenig, A. L., Zheng, H., et al. (2016). Genetically Encoded Fluorophenylalanines Enable Insights into the Recognition of Lysine Trimethylation by an Epigenetic Reader. *Chem. Commun.* 52, 12606–12609. doi:10.1039/c6cc05959g

Liu, C.-Y., Cecilia Severin, L., Lyu, C.-J., Zhu, W.-L., Wang, H.-P., Jiang, C.-J., et al. (2021). Improving Thermostability of (R)-selective Amine Transaminase from Aspergillus terreus by Evolutionary Coupling Saturation Mutagenesis. *Biochem. Eng. J.* 167, 107926. doi:10.1016/j.bej.2021.107926

Łyskowski, A., Gruber, C., Steinkellner, G., Schürmann, M., Schwab, H., Gruber, K., et al. (2014). Crystal Structure of an (R)-Selective ω-Transaminase from Aspergillus terreus. *PLoS One* 9, e87350. doi:10.1371/journal.pone.0087350

Ma, Y., Biava, H., Contestabile, R., Budisa, N., and Di Salvo, M. (2014). Coupling Bioorthogonal Chemistries with Artificial Metabolism: Intracellular Biosynthesis of Azidohomoalanine and its Incorporation into Recombinant Proteins. *Molecules* 19, 1004–1022. doi:10.3390/molecules19011004

Mathew, S., Deepankumar, K., Shin, G., Hong, E. Y., Kim, B.-G., Chung, T., et al. (2016). Identification of Novel Thermostable ω-transaminase and its Application for Enzymatic Synthesis of Chiral Amines at High Temperature. *RSC Adv.* 6, 69257–69260. doi:10.1039/c6ra15110h

Mathew, S., and Yun, H. (2012). ω-Transaminases for the Production of Optically Pure Amines and Unnatural Amino Acids. *ACS Catal.* 2, 993–1001. doi:10.1021/cs300116n

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

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010). Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* 464, 441–444. doi:10.1038/nature08817

Nojoumi, S., Ma, Y., Schwagerus, S., Hackenberger, C. P. R., and Budisa, N. (2019). In-cell Synthesis of Bioorthogonal Alkene Tag S-Allyl-Homocysteine and its Coupling with Reprogrammed Translation. *Ijms* 20, 2299. doi:10.3390/ijms20092299

Ohtake, K., Mukai, T., Iraha, F., Takahashi, M., Haruna, K.-i., Date, M., et al. (2018). Engineering an Automaturing Transglutaminase with Enhanced Thermostability by Genetic Code Expansion with Two Codon Reassignments. *ACS Synth. Biol.* 7, 2170–2176. doi:10.1021/acssynbio.8b00157

Pagar, A. D., Patil, M. D., Flood, D. T., Yoo, T. H., Dawson, P. E., and Yun, H. (2021). Recent Advances in Biocatalysis with Chemical Modification and Expanded Amino Acid Alphabet. *Chem. Rev.* 121, 6173–6245. doi:10.1021/acs.chemrev.0c01201

Pan, X., Yang, J., Xie, P., Zhang, J., Ke, F., Guo, X., et al. (2021). Enhancement of Activity and Thermostability of Keratinase from *Pseudomonas aeruginosa* CCTCC AB2013184 by Directed Evolution with Noncanonical Amino Acids. *Front. Bioeng. Biotechnol.* 9, 1–15. doi:10.3389/fbioe.2021.770907

Park, J., Lee, Y., Ko, B. J., and Yoo, T. H. (2018). Peptide-Directed Photo-Cross-Linking for Site-specific Conjugation of IgG. *Bioconjug. Chem.* 29, 3240–3244. doi:10.1021/acs.bioconjchem.8b00515

Pavkov-Keller, T., Strohmeier, G. A., Diepold, M., Peeters, W., Smeets, N., Schürmann, M., et al. (2016). Discovery and Structural Characterisation of New Fold Type IV-Transaminases Exemplify the Diversity of This Enzyme Fold. *Sci. Rep.* 6, 1–12. doi:10.1038/srep38183

Qu, G., Li, A., Acevedo-Rocha, C. G., Sun, Z., and Reetz, M. T. (2020). The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes. *Angew. Chem. Int. Ed.* 59, 13204–13231. doi:10.1002/anie.201901491

Rocha, J. F., Pina, A. F., Sousa, S. F., and Cerqueira, N. M. F. S. A. (2019). PLP-dependent Enzymes as Important Biocatalysts for the Pharmaceutical, Chemical and Food Industries: A Structural and Mechanistic Perspective. *Catal. Sci. Technol.* 9, 4864–4876. doi:10.1039/c9cy01210a

Sarak, S., Sung, S., Jeon, H., Patil, M. D., Khobragade, T. P., Pagar, A. D., et al. (2021). An Integrated Cofactor/Co-Product Recycling Cascade for the Biosynthesis of Nylon Monomers from Cycloalkylamines. *Angew. Chem. Int. Ed.* 60, 3481–3486. doi:10.1002/anie.202012658

Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., et al. (2010). Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* 329, 305–309. doi:10.1126/science.1188934

Sayer, C., Martinez-Torres, R. J., Richter, N., Isupov, M. N., Hailes, H. C., Littlechild, J. A., et al. (2014). The Substrate Specificity, Enantioselectivity and Structure of the (*R*)-selective Amine : Pyruvate Transaminase fromNectria Haematococca. *FEBS J.* 281, 2240–2253. doi:10.1111/febs.12778

Schätzle, S., Steffen-Munsberg, F., Thontowi, A., Höhne, M., Robins, K., and Bornscheuer, U. T. (2011). Enzymatic Asymmetric Synthesis of Enantiomerically Pure Aliphatic, Aromatic and Arylaliphatic Amines with (*R*)-selective Amine Transaminases. *Adv. Synth. Catal.* 353, 2439–2445. doi:10.1002/adsc.201100435

Schipp, C. J., Ma, Y., Al-Shameri, A., D'Alessio, F., Neubauer, P., Contestabile, R., et al. (2020). An Engineered *Escherichia coli* Strain with Synthetic Metabolism for in-Cell Production of Translationally Active Methionine Derivatives. *ChemBioChem* 21, 3525–3538. doi:10.1002/cbic.202000257

Shin, G., Mathew, S., and Yun, H. (2015). Kinetic Resolution of Amines by (*R*)-selective omega-transaminase from Mycobacterium Vanbaalenii. *J. Ind. Eng. Chem.* 23, 128–133. doi:10.1016/j.jiec.2014.08.003

Siddiqui, K. S. (2017). Defying the Activity-Stability Trade-Off in Enzymes: Taking Advantage of Entropy to Enhance Activity and Thermostability. *Crit. Rev. Biotechnol.* 37, 309–322. doi:10.3109/07388551.2016.1144045

Slabu, I., GalmanLloyd, J. L, Richard, C. A., Lloyd, R. C., and Turner, N. J. (2017). Discovery, Engineering, and Synthetic Application of Transaminase Biocatalysts. *ACS Catal.* 7, 8263–8284. doi:10.1021/acscatal.7b02686

Steffen-Munsberg, F., Vickers, C., Kohls, H., Land, H., Mallin, H., Nobili, A., et al. (2015). Bioinformatic Analysis of a PLP-dependent Enzyme Superfamily Suitable for Biocatalytic Applications. *Biotechnol. Adv.* 33, 566–604. doi:10.1016/j.biotechadv.2014.12.012

Telzerow, A., Paris, J., Håkansson, M., González-Sabín, J., Ríos-Lombardía, N., Gröger, H., et al. (2021). Expanding the Toolbox of R -Selective Amine Transaminases by Identification and Characterization of New Members. *ChemBioChem* 22, 1232–1242. doi:10.1002/cbic.202000692

Telzerow, A., Paris, J., Håkansson, M., González-Sabín, J., Ríos-Lombardía, N., Schürmann, M., et al. (2019). Amine Transaminase from Exophiala Xenobiotica-Crystal Structure and Engineering of a Fold IV Transaminase that Naturally Converts Biaryl Ketones. *ACS Catal.* 9, 1140–1148. doi:10.1021/acscatal.8b04524

Tseng, H.-W., Baumann, T., Sun, H., Wang, Y.-S., Ignatova, Z., and Budisa, N. (2020). Expanding the Scope of Orthogonal Translation with Pyrrolysyl-tRNA Synthetases Dedicated to Aromatic Amino Acids. *Molecules* 25, 4418–18. doi:10.3390/molecules25194418

Voss, M., Xiang, C., Esque, J., Nobili, A., Menke, M. J., André, I., et al. (2020). Creation of (R)-Amine Transaminase Activity within an α-Amino Acid Transaminase Scaffold. *ACS Chem. Biol.* 15, 416–424. doi:10.1021/acschembio.9b00888

Wang, L., and Schultz, P. G. (2005). Expanding the Genetic Code. *Angew. Chem. Int. Ed.* 44, 34–66. doi:10.1002/anie.200460627

Wang, Y., Xue, P., Cao, M., Yu, T., Lane, S. T., and Zhao, H. (2021). Directed Evolution: Methodologies and Applications. *Chem. Rev.* 121, 12384–12444. doi:10.1021/acs.chemrev.1c00260

Won, Y., Jeon, H., Pagar, A. D., Patil, M. D., Nadarajan, S. P., Flood, D. T., et al. (2019a). *In Vivo* biosynthesis of Tyrosine Analogs and Their Concurrent Incorporation into a Residue-specific Manner for Enzyme Engineering. *Chem. Commun.* 55, 15133–15136. doi:10.1039/c9cc08503c

Won, Y., Pagar, A. D., Patil, M. D., Dawson, P. E., and Yun, H. (2019b). Recent Advances in Enzyme Engineering through Incorporation of Unnatural Amino Acids. *Biotechnol. Bioproc. E* 24, 592–604. doi:10.1007/s12257-019-0163-x

Young, D. D., and Schultz, P. G. (2018). Playing with the Molecules of Life. *ACS Chem. Biol.* 13, 854–870. doi:10.1021/acschembio.7b00974

Yu, H., and Dalby, P. A. (2018). Exploiting Correlated Molecular-Dynamics Networks to Counteract Enzyme Activity-Stability Trade-Off. *Proc. Natl. Acad. Sci. USA* 115, E12192–E12200. doi:10.1073/pnas.1812204115

Yu, Y., Hu, C., Xia, L., and Wang, J. (2018). Artificial Metalloenzyme Design with Unnatural Amino Acids and Non-native Cofactors. *ACS Catal.* 8, 1851–1863. doi:10.1021/acscatal.7b03754

Yu, Z., Yu, H., Tang, H., Wang, Z., Wu, J., Yang, L., et al. (2021). Site-specifically Incorporated Non-Canonical Amino Acids into *Pseudomonas alcaligenes* Lipase to Hydrolyze L -menthol Propionate Among the Eight Isomers. *ChemCatChem* 13, 2691–2701. doi:10.1002/cctc.202100358

Zeifman, Y. S., Boyko, K. M., Nikolaeva, A. Y., Timofeev, V. I., Rakitina, T. V., Popov, V. O., et al. (2019). Functional Characterization of PLP Fold Type IV Transaminase with a Mixed Type of Activity from Haliangium Ochraceum. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1867, 575–585. doi:10.1016/j.bbapap.2019.03.005

Studying Acetylation of Aconitase Isozymes by Genetic Code Expansion

Jessica Araujo[1], Sara Ottinger[2], Sumana Venkat[3], Qinglei Gan[2] and Chenguang Fan[1,2]*

[1]Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR, United States, [2]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR, United States, [3]Children's Research Institute, The University of Texas Southwestern Medical Center, Dallas, TX, United States

Aconitase catalyzes the second reaction of the tricarboxylic acid cycle, the reversible conversion of citrate and isocitrate. *Escherichia coli* has two isoforms of aconitase, AcnA and AcnB. Acetylomic studies have identified acetylation at multiple lysine sites of both *E. coli* aconitase isozymes, but the impacts of acetylation on aconitases are unknown. In this study, we applied the genetic code expansion approach to produce 14 site-specifically acetylated aconitase variants. Enzyme assays and kinetic analyses showed that acetylation of AcnA K684 decreased the enzyme activity, while acetylation of AcnB K567 increased the enzyme activity. Further *in vitro* acetylation and deacetylation assays were performed, which indicated that both aconitase isozymes could be acetylated by acetyl-phosphate chemically, and be deacetylated by the CobB deacetylase at most lysine sites. Through this study, we have demonstrated practical applications of genetic code expansion in acetylation studies.

Keywords: lysine acetylation, aconitase, TCA cycle, genetic code expansion, deacetylase

INTRODUCTION

Aconitase catalyzes the reversible conversion of citrate and isocitrate in the tricarboxylic acid (**TCA**) and glyoxylate cycles (**Scheme 1**). It is an iron-sulfur (**Fe-S**) enzyme (Beinert et al., 1996). Depending on the state of the Fe-S cluster, aconitase has three forms: the active (4Fe-4S) form, the inactive (3Fe-4S) form, and the apo-enzyme form. The (4Fe-4S) cluster is sensitive to reactive oxygen species (**ROS**) and iron depletion which impair aconitase activities, so aconitases are widely used as biomarkers for oxidative stress and intracellular sensors of iron and redox states (Gardner, 2002; Castro et al., 2019). In *E. coli* cells, there are two aconitase isozymes, AcnA and AcnB (Gruer and Guest, 1994). Enzymological and regulatory analyses indicated that AcnB is the major TCA enzyme expressed during the exponential phase while AcnA is synthesized during the stationary phase or under stress conditions (Cunningham et al., 1997). Inactivated by ROS or iron depletion, both AcnA and AcnB apo-enzymes can bind to 3'-untranslated region of *acnA* and *acnB* mRNAs to stabilize them and increase their own expression, mediating a post-transcriptional positive autoregulation (Tang and Guest, 1999). Furthermore, aconitases are also regulated by post-translational modifications (**PTM**s), mostly oxidation, nitrosylation, and thiolation of cysteine residues around the Fe-S cluster (Lushchak et al., 2014). Recently, a number of acetylated lysine residues have been identified in aconitase isozymes of both mammals and bacteria (Xu et al., 2017). Two studies on mammalian mitochondrial aconitase (**mAcn**) indicated that acetylation can increase its enzyme activity. By acetylating mitochondria *in vitro* with acetic anhydride or acetyl-CoA and increasing the acetylation level in mice with high-fat diet, it was found that mAcn with higher acetylation levels had increased activities. Sequence alignment and structural modeling suggested the acetylation of the conserved K144 residue could induce such enhancement (Fernandes et al., 2015). In another study, K258 of mAcn was identified to be acetylated in human prostate cancer cells by mass spectrometry

analysis. While replacing K258 with arginine, such acetylation-deficient variant had decrease enzyme activities. The deacetylase SIRT3 deacetylates mAcn and decreases its enzyme activity (Sawant Dessai et al., 2021). However, both *E. coli* AcnA and AcnB have low sequence identities with mammalian mAcn, so the impacts of acetylation on aconitase isozymes in *E. coli* remain unknown.

The classic approach to study lysine acetylation is to use glutamine (KQ mutation) as a mimic of acetyllysine. However, this method undermines the structural difference between glutamine and acetyllysine. The side chain of glutamine residue is ~4 Å shorter than acetyllysine, so it may not reflect the real impacts of lysine acetylation. Indeed, our previous study on lysine acetylation of isocitrate dehydrogenase compared its activity with KQ mutations and real acetylated lysine residues, showing that at some acetylation sites the KQ mutation method derived different or even opposite conclusions (Venkat et al., 2018). To overcome this problem, the genetic code expansion technique has been used to generate site-specifically acetylated enzyme variants. This technique introduces an aminoacyl-tRNA synthetase which has been engineered to recognize acetyllysine and a tRNA which can decode a stop codon (UAG) as acetyllysine to produce site-specifically and purely acetylated proteins (Neumann et al., 2009). In this work, we used this approach to study lysine acetylation of aconitase isozymes, demonstrating a practical application of genetic code expansion in protein PTM studies.

MATERIALS AND METHODS

General Molecular Biology and Protein Analyses

Chemicals were purchased from VWR International (Radnor, PA, United States) or Chem-Impex International (Wood Dale, IL, United States). Plasmid were constructed by the NEBuilder HiFi DNA Assembly Kit (New England Biolabs, Ipswich, MA, United States). Point mutations were generated by the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). For western blotting, purified aconitase isozymes and their variants were separated on SDS PAGE gels and transferred to the PVDF membranes. The horseradish peroxidase (HRP)-conjugated acetyllysine antibody (Cell Signaling Technology, Danvers, MA, United States) was used as the primary antibody, and chemiluminescence for detection was generated by Pierce ECL Western Blotting substrates (Thermo Scientific, Waltham, MA, United States).

Expression and Purification of Aconitases and Acetylated Variants

The gene of *acnA* or *acnB* or their mutants was cloned into the *pCDF-1b* plasmid (EMD Millipore, Burlington, MA, United States) with a C-terminal His₆-tag, individually. Then it was transformed into BL21 (DE3) cells together with the acetyllysine incorporation system routinely used in our group (Venkat et al., 2017a). Cells was grown in 400 ml of LB medium with 100 μg/ml streptomycin, 50 μg/ml chloramphenicol, 10 mM acetyllysine, and 20 mM nicotinamine (NAM, the deacetylase

inhibitor) at 37°C to OD 600 nm of 0.6–0.8, then 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression. Cells were then incubated at 16°C for an additional 12 h and harvested by centrifugation at 4,000 × g for 20 min at 4°C. Cell pellets were suspended in 12 ml of 50 mM Tris (pH 7.8), 300 mM NaCl, 20 mM imidazole, 20 mM NAM, and 5 mM β-mercaptoethanol with cocktail protease inhibitors (Roche, Basel, Switzerland), and then broken by sonication. The crude extract was centrifuged at 20,000 × g for 30 min at 4°C. The soluble fraction was filtered through a 0.45-μm membrane and loaded onto a column containing 2 ml of Ni-NTA resin (Qiagen, Hilden, Germany). The column was then washed with 25 ml of 50 mM Tris (pH 7.8), 300 mM NaCl, 1 mM DTT, and 50 mM imidazole, and eluted with 2 ml of 50 mM Tris (pH 7.8), 300 mM NaCl, 1 mM DTT, and 200 mM imidazole. SDS-PAGE electrophoresis was performed to check the purity of aconitases and their variants. Western blotting and mass spectrometry were performed to confirm the incorporation of acetyllysine at correct sites.

The Aconitase Activity Assay and Kinetic Analyses

Before enzyme activity assays, purified aconitases and their acetylated variants were reactivated by incubating with 1 mM $(NH_4)_2Fe(SO_4)_2$ and 5 mM DTT in 50 mM Tris (pH 8) for 30 min following previous protocols (Bradbury et al., 1996). Enzyme assays were performed with the commercial aconitase assay kit from BioAssay System (Hayward, CA). Briefly, it measures the isocitrate generated as a product of the aconitase reaction. The isocitrate is then oxidized producing NADPH and the oxidation product. The NADPH converts the dye to an intense violet color with an absorption maximum at 565 nm. The increase in absorbance at 565 nm is directly proportional to aconitase activity. To determine steady-state kinetic parameters, the concentration of the substrate citrate was varied from 0.1 to 50 mM. Kinetic parameters were calculated by non-liner regression with software Grafit (Erithacus Software).

Mass Spectrometry Analyses

The LC-MS/MS analyses were performed by the Yale Keck Proteomics facility and followed the previous protocol (Venkat et al., 2018). Briefly, aconitases and their variants were separated by SDS-PAGE electrophoresis. Protein bands were cut and digested in gel by trypsin, and analyzed by LC-MS/MS on an LTQ Orbitrap XL equipped with a nanoACQUITY UPLC system. The Mascot search algorithm was used to search for the substitution of the lysine residue with acetyllysine. All the MS data have been deposited as ProteomeXchange dataset PXD031431.

The *In Vitro* Acetylation Assay

The acetylation reaction was performed in the buffer of 50 mM Tris (pH 8.0), 0.1 mM EDTA, 1 mM DTT and 10 mM sodium butyrate, initiated by mixing 10 μg enzyme and 3 mM AcP in a total volume of 100 μL, and then incubated at 37°C for 1 h.

The *In Vitro* Deacetylation Assay

The deacetylation reaction was performed in the buffer of 50 mM HEPES (pH 7.0), 5 mM $MgCl_2$, 1.0 mM NAD^+, and 1 mM DTT, initiated by mixing 10 μg enzyme and 10 μg purified CobB in a total volume of 100 μL, and then incubated at 37°C for 1 h.

RESULTS

Generation of Site-Specifically Acetylated Aconitase Variants

Several quantitative acetylomic studies of *E. coli* cells have demonstrated that both aconitase isoenzymes have higher acetylation stoichiometry than many other *E. coli* proteins (Baeza et al., 2014; Meyer et al., 2016; Weinert et al., 2017). Thus, we aimed to identify the effects of acetylation on aconitase isoenzymes site-specifically. Although a series of acetylomic studies have been performed for *E. coli* cells, the sets of acetylation sites identified in aconitase isoenzymes do not overlap well because of differences in strains, growth media, and MS detection and resolutions (Yu et al., 2008; Zhang et al., 2009; Colak et al., 2013; Weinert et al., 2013; Zhang et al., 2013; Castano-Cerezo et al., 2014; Kuhn et al., 2014; AbouElfetouh et al., 2015; Schilling et al., 2015). To be feasible and avoid biased selection, we chose all the lysine residues identified to be acetylated by more than three independent acetylomic studies, which were K164, K342, K482, K684 of AcnA and K77, K373, K396, K407, K539, K559, K567, K728, K759, K835 of AcnB.

In this study, we utilized our optimized acetyllysine incorporation system to produce site-specifically acetylated aconitase variants at selected sites listed above individually (Venkat et al., 2017a). To minimize the non-specific acetylation of other lysine residues in aconitases, we used the BL21 (DE3) strain as the host cell line, which has a low level of acetylation globally (Weinert et al., 2013). Our previous studies on acetylation of malate dehydrogenase, isocitrate dehydrogenase, and citrate synthase have shown that those wild-type enzymes purified from BL21 (DE3) cells have low levels of non-specific acetylation (Venkat et al., 2017b; Venkat et al., 2018; Venkat et al., 2019). As expected, wild-type AcnA and AcnB overexpressed in BL21 (DE3) cells had no or very weak level of acetylation (**Supplementary Figure S1**). We fused the His_6-tag to the C-terminus of aconitase variants for easy purification and to remove truncated proteins terminated at inserted UAG codons. All the purified acetylated aconitase variants had clear single bands in SDS-PAGE gels and were detected by the acetyllysine antibody in western blots (**Supplementary Figure S1**). The positions of acetyllysine incorporation were confirmed by LC-MS/MS analyses (**Supplementary Figures S2–S15**).

The Site-Specific Effects of Lysine Acetylation on Aconitase Activities

First, we measured the enzyme activities of purified AcnA and AcnB as well as their site-specifically acetylated variants with the commercial kit individually (**Figure 1**). Acetylation of most lysine sites had no significant effects on aconitase activities ($p = 0.726$, 0.110, and 0.804 for AcnA variants at K164, K342 and K482, individually; $p = 0.186$, 0.198, 0.649, 0.580, 0.094, 0.257, 0.625, 0.257, 0.674 for AcnB variants at K77, K373, K396, K407, K539, K559, K728, K759, and K835, individually). There were only two variants which significantly affect aconitase activities ($p < 0.001$ for both AcnA K684 and AcnB K567). Interestingly, the impacts of acetylation on enzyme activities were different in aconitase isozymes. Acetylation of AcnA K684 decreased the activity by ~3-fold while acetylation of AcnB K567 increased the activity by ~2-fold.

To obtain insights into the impacts of acetylation on aconitase activities, we performed steady-state kinetic analyses of AcnA and AcnB as well as those two variants (AcnA-684AcK and AcnB-567AcK) which significantly affected activities (**Figure 1**). Acetylation of AcnA K684 impairs both the substrate binding and the turnover number while acetylation of AcnB K567 only enhances the turnover number.

The Acetyl-Phosphate-Dependent Acetylation of Aconitase Isozymes

It is known that lysine acetylation in *E. coli* is mostly generated non-enzymatically with acetyl-phosphate (**AcP**) as the acetyl-donor while acetyl-CoA-dependent enzymatic acetylation only applies for a small portion of proteins (Weinert et al., 2013; Hentchel and Escalante-Semerena, 2015). Our previous studies also showed that AcP itself can acetylate several TCA cycle enzymes chemically (Venkat et al., 2017b; Venkat et al., 2018; Venkat et al., 2019). In this study, WT AcnA and AcnB expressed in BL21 (DE3) cells were purified and treated with 3 mM AcP *in vitro* which is the physiological concentration of AcP inside *E. coli* cells (Klein et al., 2007). Western blots showed that AcP acetylated both AcnA and AcnB in a time-dependent manner (**Figure 2A**). The activities of AcnA and AcnB before and after AcP-treatment were measured. Consistent with site-specific results above, acetylation of AcnA impaired its activity while acetylation of AcnB enhanced its activity ($p < 0.001$ for AcnA and $p = 0.004$ for AcnB) (**Figure 2B**). The impacts of acetylation by AcP-treatment were not as significant as the site-specific acetylation above, probably because AcP cannot acetylate lysine residues completely while purely acetylated variants were tested in above site-specific experiments.

Then we performed LC-MS/MS analyses to identify acetylation sites in AcnA and AcnB by AcP-treatment. Those 14 sites selected for site-specific tests above were all acetylated by *in vitro* AcP-treatment. Besides them, we also identified 12 additional acetylation sites in AcnA (K10, K16, K116, K257, K283, K391, K406, K453, K578, K758, K770, and, K823) and 11 additional acetylation sites in AcnB (K20, K64, K73, K135, K137, K144, K356, K387, K571, K613, and K722). Among them, K116, K257, K406, K453, K578, K823 of AcnA and K64, K137 of AcnB have not been identified to be acetylated in *E. coli* cells ever before, probably because acetylation of these sites has low stoichiometry and can be deacetylated easily in living cells. On the other hand, K18, K30, K71, K161, K460, K585, K832 of AcnA
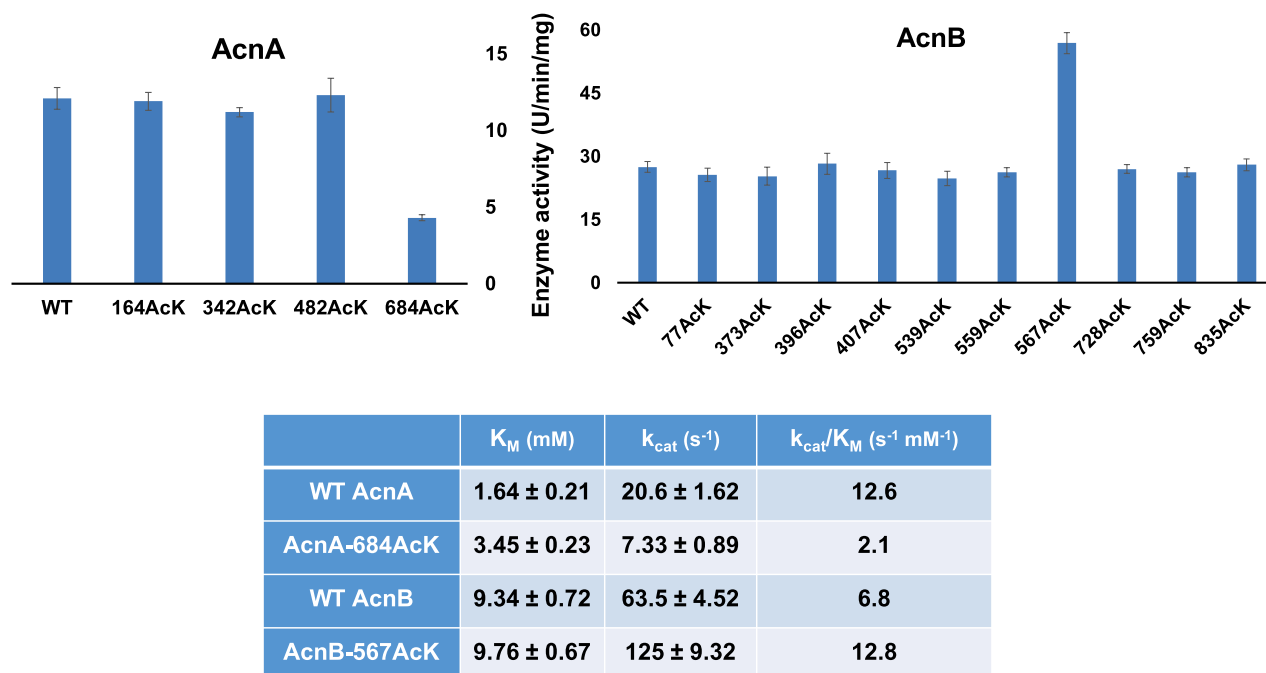
FIGURE 1 | The enzyme activities and kinetic analyses of AcnA, AcnB and their acetylated variants. The **(A)** is enzyme activities measured by the commercial kit with 50 mM citrate as the substrate concentration. The **(B)** is the steady-state kinetic parameters. Kinetic parameters were calculated by non-liner regression with software Grafit. Mean values and standard deviations were calculated based on three replicates. The Student's t test was used to analyze the significance of differences between activities of each acetylated variant and the corresponding WT enzyme.
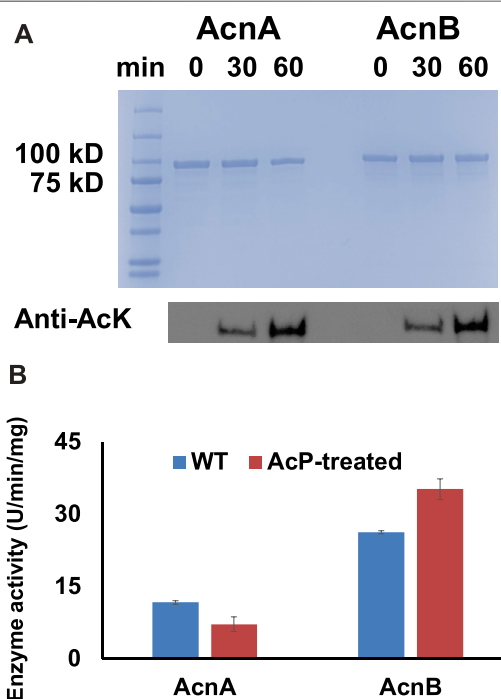
FIGURE 2 | AcP-dependent acetylation of AcnA and AcnB. **(A)** SDS-PAGE and wester blots of purified WT AcnA and WT AcnB treated with AcP *in vitro*. 2 μg of proteins were loaded for each lane. The full image of western blots is in **Supplementary Figure S16**. **(B)** Enzyme activities of purified WT AcnA and WT AcnB before and after AcP-treatment. Mean values and standard deviations were calculated based on three replicates. The Student's t test was used to analyze the significance of differences between activities of enzymes before and after AcP-treatment.
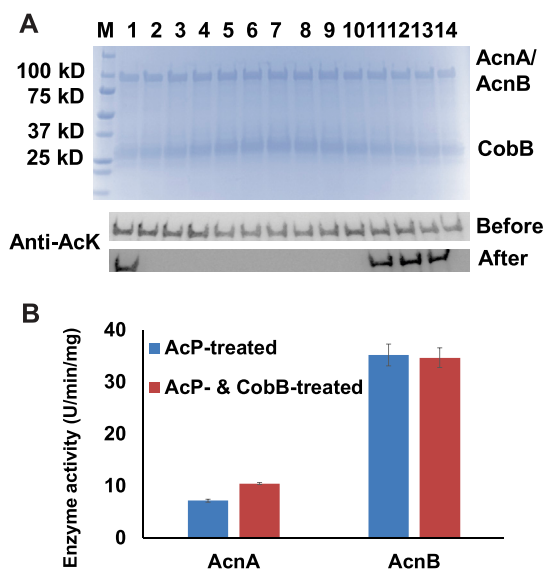
FIGURE 3 | CobB-catalyzed deacetylation of AcnA and AcnB. **(A)** SDS-PAGE and wester blots of site-specifically acetylated AcnA and AcnB variants incubated with CobB *in vitro*. 2 µg of AcnA/AcnB and CobB were loaded for each lane. The full image of western blots is in **Supplementary Figure S17**. Samples from lane 1 to 4 are AcnA-164AcK, -342AcK, -482AcK, and -684AcK. Samples from lane 5 to 14 are AcnB-77AcK, 373-AcK, -396AcK, -407AcK, -539AcK, -559AcK, -567AcK, -728AcK, -759AcK, and -835AcK. **(B)** Enzyme activities of AcP-treated AcnA and AcnB before and after CobB-incubation. Mean values and standard deviations were calculated based on three replicates. The Student's t test was used to analyze the significance of differences between activities of enzymes before and after CobB-treatment.
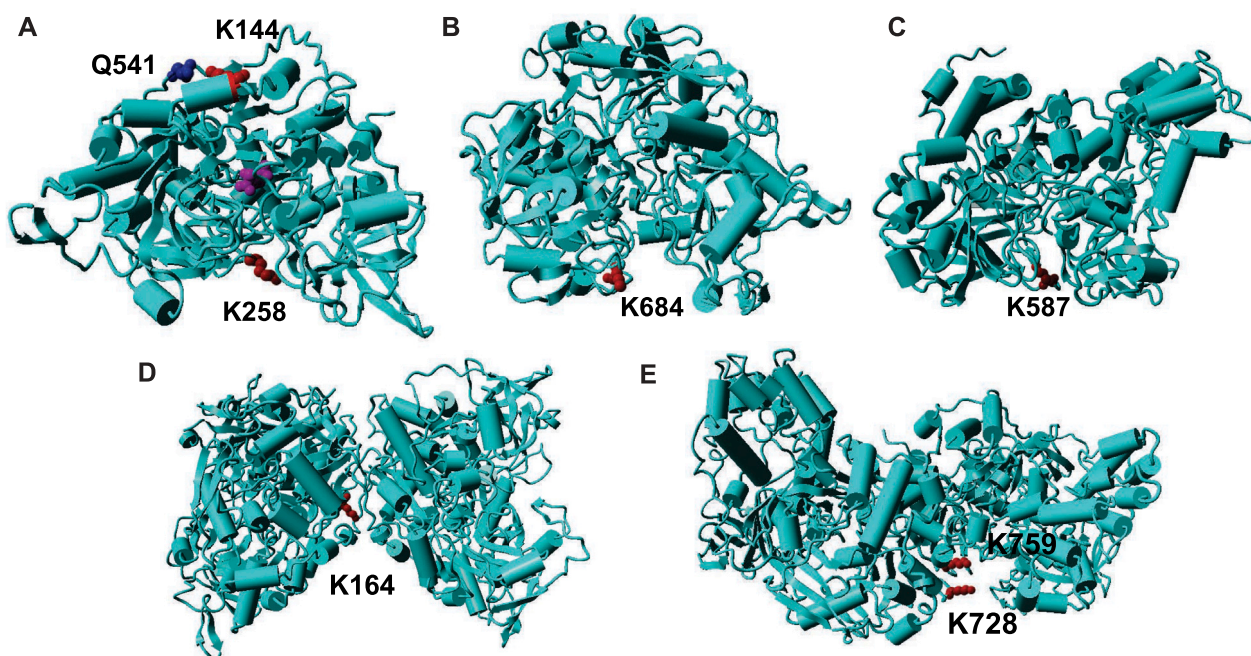
FIGURE 4 | Mapping of acetylated lysine residues on the structures of AcnA and AcnB. **(A)** The crystal structure of human cAcn (PDB ID: 2B3Y). **(B)** The homology model of AcnA structure based on human cAcn. **(C)** The crystal structure of AcnB (PDB ID: 1L5J). **(D)** The homology model of AcnA dimer based on human cAcn. **(E)** The structure of AcnB dimer. The backbone of the protein is in cyan; lysine residues are in red; other key residues are in blue; the substrate is in purple.
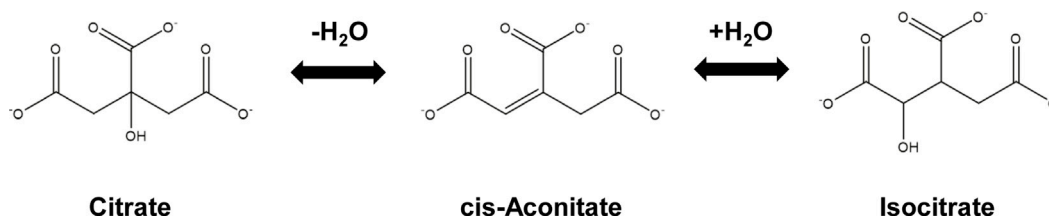
SCHEME 1 | The reaction catalyzed by aconitase.

and K85, K110, K117, K221, K229, K267, K537, K588 of AcnB listed in the *E. coli* acetylation database (Xu et al., 2017) were not identified in our *in vitro* AcP-acetylation tests, implying that specific acetyltransferases or cofactors could be necessary for acetylation of these lysine sites in cells.

The CobB-Dependent Deacetylation of Aconitase Isozymes

Acetylation of lysine residues is reversible, and the deacetylation of acetylated lysine residues is catalyzed by protein lysine deacetylases (**KDAC**). To date, CobB is still the only well-known KDAC in *E. coli* (Starai et al., 2002). Our previous studies showed that CobB can deacetylate acetylated lysine residues in several TCA cycle enzymes, but not for all the acetylation sites (Venkat et al., 2017b; Venkat et al., 2018; Venkat et al., 2019). In this study, we incubated those 14 site-specifically acetylated AcnA and AcnB variants with CobB, and used western blotting to determine the site specificity of CobB for AcnA and AcnB (**Figure 3A**). Most of acetylation sites were sensitive to CobB, while K164 of AcnA and K567, K728, K759 of AcnB were resistant to CobB.

Then we incubated AcP-treated AcnA and AcnB with CobB *in vitro*. After that, we measured the enzyme activities (**Figure 3B**). CobB-dependent acetylation restored AcnA activity ($p < 0.001$), but did not affect AcnB activity significantly ($p = 0.775$). Acetylation of K684 in AcnA decreases AcnA activity and K684 is sensitive to CobB, so deacetylation of K684 could restore its enzyme activity. On the other hand, acetylation of K567 in AcnB increases AcnB activity but K567 is resistant to CobB, so CobB-dependent deacetylation could not restore AcnB activity. These results also indicated that K684 of AcnA and K567 of AcnB are the two lysine residues whose acetylation affects their enzyme activities the most, which is consistent to our site-specific results above.

DISCUSSION

The Effects of Acetylation on Aconitase Activities

Before our work, only two papers have reported acetylation of aconitase isoenzymes, which indicated that acetylation of K144 and K258 in mammalian mAcn enhances enzyme activities (Fernandes et al., 2015; Sawant Dessai et al., 2021). To find the structure-function relationships of aconitase acetylation,

we mapped those two lysine residues onto the crystal structures of mAcn. Because there is no reported human mAcn structures, we used bovine mAcn as the model which has 96% sequence identify with human mAcn (**Figure 4A**). The active site is at the center of the enzyme (the purple molecule is the intermediate analog methyl-isocitrate). K144 is at the back of the active site. The previous study proposed that acetylation of K144 strengthens its interaction with the nearby Q541, hence inducing conformational changes to facilitate substrate binding (Fernandes et al., 2015). K258 is at the entrance of the active site, so its acetylation could also cause conformational changes to enhance enzyme activities.

The structure of human cytosolic Acn (**cAcn**) which has a 53% sequence identify with *E. coli* AcnA has been solved (Dupuy et al., 2006). The structure of *E. coli* AcnB has also been solved (Williams et al., 2002). Although both *E. coli* AcnA and AcnB have low sequence identifies with human mAcn (28 and 24%, respectively), the overall structures of different aconitases are very similar (Williams et al., 2002). Hence, we mapped K684 of AcnA and K567 of AcnB onto their structures. Both K684 of AcnA and K567 of AcnB are located at the entrance of the active sites of aconitase isozymes, but the orientations of two lysine residues are different (**Figures 4B,C**). K684 of AcnA faces outside the active site, and the addition of an acetyl group could limit the accessibility of the active sites, thus decreasing substrate binding and catalytic efficiency. On the other hand, K567 of AcnB faces inside the active site, and its acetylation may not affect the substrate binding. The additional acetyl group could fill the space of the active site to provide extra interactions to stabilize the intermediate, hence increasing the turnover of the reaction. Further structural studies such as crystalizing those site-specifically acetylated variants are necessary to identify the mechanisms of impacts from acetylation.

The Sensitivities of Acetylated Lysine Residues in Aconitases Towards CobB Deacetylase

Our deacetylation tests showed that CobB can remove most of acetylated lysine resides in both AcnA and AcnB (**Figure 3**). Then we mapped those residues which are sensitive to CobB (**Supplementary Figure S18**). CobB-sensitive sites are all located at protein surface for easy CobB access, consistent with our previous studies on other TCA cycle enzymes (Venkat et al., 2017b; Venkat et al., 2018; Venkat et al., 2019). On the other hand, we also mapped those residues which are resistant to CobB. Both AcnA and AcnB

form dimers in solutions (Jordan et al., 1999). We used human cAcn as the template to model the AcnA structure. K164 of AcnA is located at the interface of two subunits (**Figure 4D**). K728 and K759 of AcnB are also at the subunit interface (**Figure 4E**). Such steric hindrance limits the access of CobB for deacetylation. Although K567 is located at the entrance of the active site, the primary amine group points to interior of the active site (**Figure 4C**), and this orientation also limits the access of CobB for deacetylation. However, acetylation of K567 increases AcnB activity, so the acetylation of K567 should be regulated. Because we used purified CobB in *in vitro* experiments, some cofactors that are necessary for deacetylation of K567 may be missed in *in vitro* experiments. Furthermore, there could be other deacetylases in *E. coli* which can deacetylate those CobB-resistant positions in Acn. Indeed, besides CobB which belongs to class III lysine deacetylases, putative members of class I, II and IV lysine deacetylases have also been identified in bacteria (Hildmann et al., 2007).

In summary, we applied the genetic code expansion approach to generate site-specifically acetylated aconitase variants. With these purely acetylated variants on hand, we were able to identify key lysine residues, whose acetylation has significant impacts on enzyme functions. We could also test acetylation and deacetylation specificities by this approach. Thus, we have demonstrated practical applications of genetic code expansion in protein acetylation studies. Further studies will be implemented to investigate aconitase acetylation regulations in living cells. Together with genetic studies such as knocking out proteins involved in acetylation and deacetylation processes, the results from this work will provide solid evidence.

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: MassIVE (Accession: MSV000088779), ProteomeXchange (Accession: PXD031431).

AUTHOR CONTRIBUTIONS

JA, SO, SV, and QG performed the experiments. JA and CF analyzed the data and wrote the manuscript, while all authors edited the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.862483/full#supplementary-material

REFERENCES

AbouElfetouh, A., Kuhn, M. L., Hu, L. I., Scholle, M. D., Sorensen, D. J., Sahu, A. K., et al. (2015). The *E. coli* Sirtuin CobB Shows No Preference for Enzymatic and Nonenzymatic Lysine Acetylation Substrate Sites. *Microbiologyopen* 4 (1), 66–83. doi:10.1002/mbo3.223

Baeza, J., Dowell, J. A., Smallegan, M. J., Fan, J., Amador-Noguez, D., Khan, Z., et al. (2014). Stoichiometry of Site-specific Lysine Acetylation in an Entire Proteome. *J. Biol. Chem.* 289 (31), 21326–21338. doi:10.1074/jbc.M114.581843

Beinert, H., Kennedy, M. C., and Stout, C. D. (1996). Aconitase as Iron–Sulfur Protein, Enzyme, and Iron-Regulatory Protein. *Chem. Rev.* 96 (7), 2335–2374. doi:10.1021/cr950040z

Bradbury, A. J., Gruer, M. J., Rudd, K. E., and Guest, J. R. (1996). The Second Aconitase (AcnB) of *Escherichia coli. Microbiology (Reading)* 142 (Pt 2), 389–400. doi:10.1099/13500872-142-2-389

Castaño-Cerezo, S., Bernal, V., Post, H., Fuhrer, T., Cappadona, S., Sánchez-Díaz, N. C., et al. (2014). Protein Acetylation Affects Acetate Metabolism, Motility and Acid Stress Response in *Escherichia coli. Mol. Syst. Biol.* 10, 762. doi:10.15252/msb.20145227

Castro, L., Tórtora, V., Mansilla, S., and Radi, R. (2019). Aconitases: Non-redox Iron-Sulfur Proteins Sensitive to Reactive Species. *Acc. Chem. Res.* 52 (9), 2609–2619. doi:10.1021/acs.accounts.9b00150

Colak, G., Xie, Z., Zhu, A. Y., Dai, L., Lu, Z., Zhang, Y., et al. (2013). Identification of Lysine Succinylation Substrates and the Succinylation Regulatory Enzyme CobB in *Escherichia coli. Mol. Cell Proteomics* 12 (12), 3509–3520. doi:10.1074/mcp.M113.031567

Cunningham, L., Gruer, M. J., and Guest, J. R. (1997). Transcriptional Regulation of the Aconitase Genes (acnA and acnB) of *Escherichia coli. Microbiology* 143 (Pt 12), 3795–3805. doi:10.1099/00221287-143-12-3795

Dupuy, J., Volbeda, A., Carpentier, P., Darnault, C., Moulis, J.-M., and Fontecilla-Camps, J. C. (2006). Crystal Structure of Human Iron Regulatory Protein 1 as Cytosolic Aconitase. *Structure* 14 (1), 129–139. doi:10.1016/j.str.2005.09.009

Fernandes, J., Weddle, A., Kinter, C. S., Humphries, K. M., Mather, T., Szweda, L. I., et al. (2015). Lysine Acetylation Activates Mitochondrial Aconitase in the Heart. *Biochemistry* 54 (25), 4008–4018. doi:10.1021/acs.biochem.5b00375

Gardner, P. R. (2002). Aconitase: Sensitive Target and Measure of Superoxide. *Methods Enzymol.* 349, 9–23. doi:10.1016/s0076-6879(02)49317-2

Gruer, M. J., and Guest, J. R. (1994). Two Genetically-Distinct and Differentially-Regulated Aconitases (AcnA and AcnB) in *Escherichia coli. Microbiology* 140 (Pt 10), 2531–2541. doi:10.1099/00221287-140-10-2531

Hentchel, K. L., and Escalante-Semerena, J. C. (2015). Acylation of Biomolecules in Prokaryotes: a Widespread Strategy for the Control of Biological Function and Metabolic Stress. *Microbiol. Mol. Biol. Rev.* 79 (3), 321–346. doi:10.1128/MMBR.00020-15

Hildmann, C., Riester, D., and Schwienhorst, A. (2007). Histone Deacetylases-An Important Class of Cellular Regulators with a Variety of Functions. *Appl. Microbiol. Biotechnol.* 75 (3), 487–497. doi:10.1007/s00253-007-0911-2

Jordan, P. A., Tang, Y., Bradbury, A. J., Thomson, A. J., and Guest, J. R. (1999). Biochemical and Spectroscopic Characterization of *Escherichia coli* Aconitases (AcnA and AcnB). *Biochem. J.* 344, 739–746. doi:10.1042/bj3440739

Klein, A. H., Shulla, A., Reimann, S. A., Keating, D. H., and Wolfe, A. J. (2007). The Intracellular Concentration of Acetyl Phosphate in *Escherichia coli* Is Sufficient for Direct Phosphorylation of Two-Component Response Regulators. *J. Bacteriol.* 189 (15), 5574–5581. doi:10.1128/JB.00564-07

Kuhn, M. L., Zemaitaitis, B., Hu, L. I., Sahu, A., Sorensen, D., Minasov, G., et al. (2014). Structural, Kinetic and Proteomic Characterization of Acetyl Phosphate-dependent Bacterial Protein Acetylation. *PLoS One* 9 (4), e94816. doi:10.1371/journal.pone.0094816

Lushchak, O. V., Piroddi, M., Galli, F., and Lushchak, V. I. (2014). Aconitase post-translational Modification as a Key in Linkage between Krebs Cycle, Iron Homeostasis, Redox Signaling, and Metabolism of Reactive Oxygen Species. *Redox Rep.* 19 (1), 8–15. doi:10.1179/1351000213Y.0000000073

Meyer, J. G., D'Souza, A. K., Sorensen, D. J., Rardin, M. J., Wolfe, A. J., Gibson, B. W., et al. (2016). Quantification of Lysine Acetylation and Succinylation Stoichiometry in Proteins Using Mass Spectrometric Data-independent Acquisitions (SWATH). *J. Am. Soc. Mass. Spectrom.* 27 (11), 1758–1771. doi:10.1007/s13361-016-1476-z

Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., et al. (2009). A Method for Genetically Installing Site-specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Mol. Cel.* 36 (1), 153–163. doi:10.1016/j.molcel.2009.07.027

Sawant Dessai, A., Palestino Dominguez, M., Chen, U.-I., Hasper, J., Prechtl, C., Yu, C., et al. (2021). Transcriptional Repression of SIRT3 Potentiates Mitochondrial Aconitase Activation to Drive Aggressive Prostate Cancer to the Bone. *Cancer Res.* 81 (1), 50–63. doi:10.1158/0008-5472.CAN-20-1708

Schilling, B., Christensen, D., Davis, R., Sahu, A. K., Hu, L. I., Walker-Peddakotla, A., et al. (2015). Protein Acetylation Dynamics in Response to Carbon Overflow inEscherichia Coli. *Mol. Microbiol.* 98 (5), 847–863. doi:10.1111/mmi.13161

Starai, V. J., Celic, I., Cole, R. N., Boeke, J. D., and Escalante-Semerena, J. C. (2002). Sir2-dependent Activation of Acetyl-CoA Synthetase by Deacetylation of Active Lysine. *Science* 298 (5602), 2390–2392. doi:10.1126/science.1077650

Tang, Y., and Guest, J. R. (1999). Direct Evidence for mRNA Binding and post-transcriptional Regulation by *Escherichia coli* Aconitases. *Microbiology (Reading)* 145 (Pt 11), 3069–3079. doi:10.1099/00221287-145-11-3069

Venkat, S., Gregory, C., Meng, K., Gan, Q., and Fan, C. (2017a). A Facile Protocol to Generate Site-Specifically Acetylated Proteins in Escherichia Coli. *JoVE* 130, e57061. doi:10.3791/57061

Venkat, S., Gregory, C., Sturges, J., Gan, Q., and Fan, C. (2017b). Studying the Lysine Acetylation of Malate Dehydrogenase. *J. Mol. Biol.* 429 (9), 1396–1405. doi:10.1016/j.jmb.2017.03.027

Venkat, S., Chen, H., Stahman, A., Hudson, D., McGuire, P., Gan, Q., et al. (2018). Characterizing Lysine Acetylation of Isocitrate Dehydrogenase in *Escherichia coli*. *J. Mol. Biol.* 430 (13), 1901–1911. doi:10.1016/j.jmb.2018.04.031

Venkat, S., Chen, H., McGuire, P., Stahman, A., Gan, Q., and Fan, C. (2019). Characterizing Lysine Acetylation of *Escherichia coli* Type II Citrate Synthase. *FEBS J.* 286 (14), 2799–2808. doi:10.1111/febs.14845

Weinert, B. T., Iesmantavicius, V., Wagner, S. A., Schölz, C., Gummesson, B., Beli, P., et al. (2013). Acetyl-phosphate Is a Critical Determinant of Lysine Acetylation in *E. coli*. *Mol. Cel.* 51 (2), 265–272. doi:10.1016/j.molcel.2013.06.003

Weinert, B. T., Satpathy, S., Hansen, B. K., Lyon, D., Jensen, L. J., and Choudhary, C. (2017). Accurate Quantification of Site-specific Acetylation Stoichiometry Reveals the Impact of Sirtuin Deacetylase CobB on the *E. coli* Acetylome. *Mol. Cell Proteomics* 16 (5), 759–769. doi:10.1074/mcp.M117.067587

Williams, C. H., Stillman, T. J., Barynin, V. V., Sedelnikova, S. E., Tang, Y., Green, J., et al. (2002). *E. coli* Aconitase B Structure Reveals a HEAT-like Domain with Implications for Protein-Protein Recognition. *Nat. Struct. Biol.* 9 (6), 447–452. doi:10.1038/nsb801

Xu, H., Zhou, J., Lin, S., Deng, W., Zhang, Y., and Xue, Y. (2017). PLMD: An Updated Data Resource of Protein Lysine Modifications. *J. Genet. Genomics* 44 (5), 243–250. doi:10.1016/j.jgg.2017.03.007

Yu, B. J., Kim, J. A., Moon, J. H., Ryu, S. E., and Pan, J. G. (2008). The Diversity of Lysine-Acetylated Proteins in *Escherichia coli*. *J. Microbiol. Biotechnol.* 18 (9), 1529–1536.

Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., et al. (2009). Lysine Acetylation Is a Highly Abundant and Evolutionarily Conserved Modification in *Escherichia coli*. *Mol. Cell Proteomics* 8 (2), 215–225. doi:10.1074/mcp.M800187-MCP200

Zhang, K., Zheng, S., Yang, J. S., Chen, Y., and Cheng, Z. (2013). Comprehensive Profiling of Protein Lysine Acetylation in *Escherichia coli*. *J. Proteome Res.* 12 (2), 844–851. doi:10.1021/pr300912q

Unnatural Amino Acid: 4-Aminopyrazolonyl Amino Acid Comprising *Tri*-Peptides Forms Organogel With Co-Solvent (EtOAc: Hexane)

Amarnath Bollu [1,2], Prajnanandan Giri [1,2], Nihar Ranjan Dalabehera [1,2], Asmita Rani Asmi [1,2] and Nagendra K. Sharma [1,2]*

[1]National Institute of Science Education and Research (NISER), Bhubaneswar, India, [2]Homi Bhabha National Institute (HBNI), Mumbai, India

Ampyrone is an amino-functionalized heterocyclic pyrazolone derivative that possesses therapeutic values such as analgesic, anti-inflammatory, and antipyretics. The chemical structure of ampyrone exhibits excellent hydrogen bonding sites and is considered as the potential scaffold of supramolecular self-assembly. Recently, this molecule has been derived into unnatural amino acids such as aminopyrazolone amino acid and its peptides. This report describes that one of its amino acids, *O*-alkylated ampyrone, containing hybrid (α/β) peptides forms organogel after sonication at 50–55°C with 0.7–0.9% (w/v) in ethyl acetate: hexane (1:3). The formation/morphology of such organogels is studied by nuclear magnetic resonance Fourier-transform infrared (FT-IR), circular dichroism (CD), scanning electron microscope (SEM), transmission electron microscopy (TEM), powder X-ray diffraction (Powder-XRD), and thermogravimetric analysis (TGA). Energy-minimized conformation of APA-peptides reveals the possibility of intermolecular hydrogen bonding. Hence, APA-peptides are promising peptidomimetics for the organogel-peptides.

Keywords: ampyrone, aminopyrazolonyl amino acid, hybrid-β-peptides, organogelation, unnatural amino acid

INTRODUCTION

Peptides form self-assembly structures through non-covalent interactions, such as hydrogen bonding, van der Waals interactions, and π–π stacking (Zweep and Van Esch, 2013). The amide bonds and side chains of amino acid residues play a significant role in stabilizing the non-covalent interactions in peptides, which impart in the self-assembly of supramolecular structures including hydrogels and organogels (Hanabusa et al., 1993; Aggeli et al., 1997; Shaikh et al., 2018). Oligopeptides and small peptides are widely applied for the formation of versatile supramolecular organogels through these non-covalent interactions (Tomasini and Castellucci, 2013; Biswas et al., 2016). Sono-gels are a class of gels that are formed under ultrasound sonication and are widely applied for peptide-based gels (Li et al., 2007; Cravotto and Cintas, 2009). The process

---

Abbreviations: APA 4, aminopyrazolone acid.
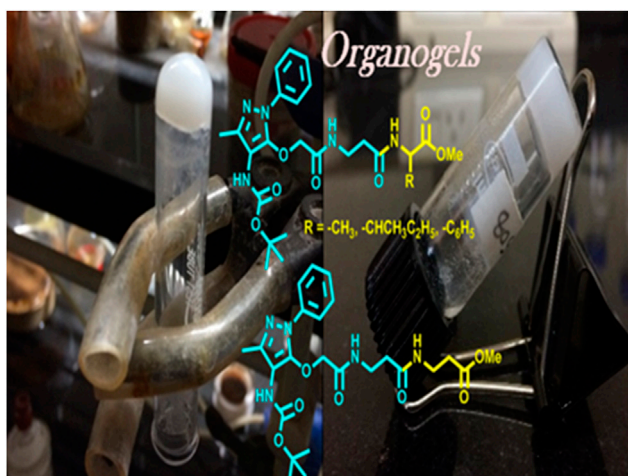
GRAPHICAL ABSTRACT |

of gelation by ultrasound could involve breaking the larger aggregates or disordered aggregates to induce the formation of well-defined larger uniform aggregates which may lead to the formation of gels (Chatterjee and Maitra, 2017). The peptide-based organogels are biocompatible materials and considered promising biomaterials for various applications such as drug delivery (Couffin-Hoarau et al., 2004; Baral et al., 2014; Rouse et al., 2017), oil recovery in the petroleum industry (Chetia et al., 2020), and removal of toxic dyes (Li et al., 2020). Recently, the sequence-specific small peptides are explored to prepare thermally stable reversible/irreversible organogel biomaterials from natural/unnatural/hybrid peptides (Chakraborty et al., 2002; Banerjee et al., 2008; Maity et al., 2015; Wang and Yan, 2018). The insertion of aromatic structural unit/aromatic amino acid at the *N*-terminal of *di-tri*-peptides leads to the formation of stable organogel materials (Babu et al., 2014). 4-Aminopyrazole containing aromatic unnatural amino acids/dipeptides have abilities to interact with several bio-macromolecules such as

interaction with specific β-sheet-rich targets in Aβ-protein and serine proteases via non-covalent interactions (Schrader and Kirsten, 1996; Kirsten and Schrader, 1997; Gilfillan et al., 2015; Hellmert et al., 2015). Previously, we have explored the structural and conformational studies of 4-aminopyrazolone amino acids/di-tri-peptide scaffold for non-covalent interactions, which is one of the important criteria for gelation (Bollu and Sharma, 2019b; a). We report the synthesis of 4-aminopyrazolone acid (APA) containing hybrid peptides with α-/β-amino acids and preparation of their *organogels* (**Figure 1**). These supramolecular self-assemblies are studied by NMR, FT-IR, CD, SEM, TEM, Powder-XRD, and TGA.

RESULTS AND DISCUSSION

We began the synthesis of the unnatural amino acid (**1**), aminopyrazolonyl acid (APA) by following the previously reported procedure (**Scheme 1**) (Bollu and Sharma, 2019b). In the literature, bipolar organic molecules have a higher propensity for organogelation (Bardelang et al., 2008; Loic, 2017). Thus, we planned to prepare bipolar molecules by introducing an APA unit at *N*-terminal of *di*-peptides containing hydrophobic side chain residues. We therefore synthesized APA *tri*-peptides (**2a-2e**) from α-β-hybrid peptide derivatives (NH$_2$-AA-OMe) and APA (**1**). The hybrid peptides (NH$_2$-AA-OMe) were prepared from β-alanine and α-amino acid (Gly/Ala/Ile/Phe). APA-β-Ala-α-Gly-OMe (**2a**) was prepared from dipeptide β-Ala-α-Gly-OMe, **2b** from β-Ala-α-Ala-OMe, **2c** from β-Ala-α-Ile, **2d** from β-Ala-α-Phe-OMe, and **2e** from β-Ala-β-Ala-OMe. These APA-peptides are well characterized by ^{1}H-/^{13}C-NMR/ESI-HRMS. Their respective spectra are provided in the **Supplementary Material**.

In the literature, the sequence-specific aromatic *tri*-peptides reportedly form organogel in the co-solvent systems (hexane: ethylacetate) after sonication (Maity et al., 2011; Maity et al., 2015). We also attempted the organogelation of unnatural

FIGURE 1 | APA-appended peptides.
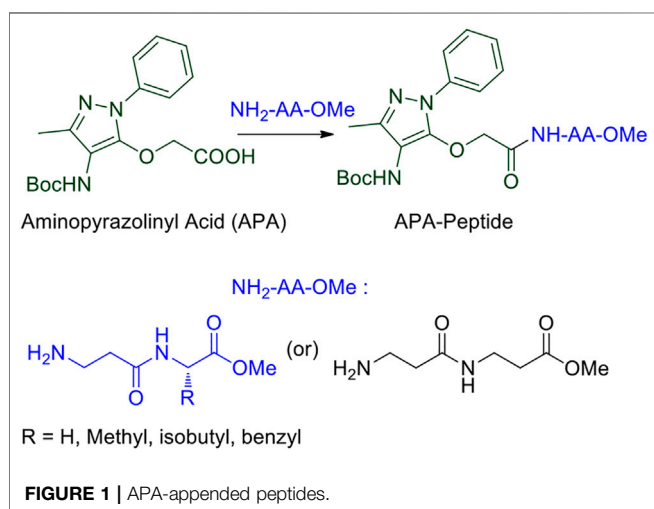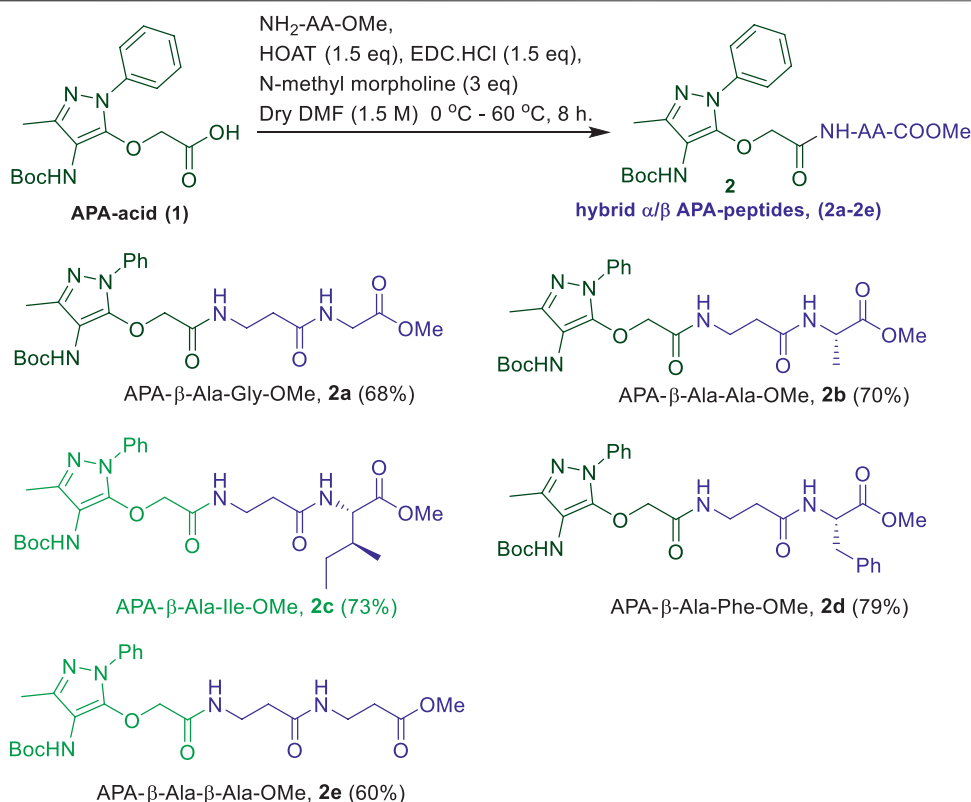
SCHEME 1 | Synthesis of α-/β-hybrid APA-peptides.

aromatic amino acid, aminopyrazolonyl amino acid (APA), containing peptides (**2a-2e**) in the same co-solvent systems (hexane: ethylacetate) by sonication. The synthesized APA-peptides (**2a-2e**) 0.7–0.9% (wt/vol) were dissolved in solvent systems EtOAc:Hexane (1:3, v/v) and sonicated for 2 minutes above the room temperature (~50°C) and then allowed to cool at room temperature. We noticed that the homogenous solutions of peptides (**2b-2e**) were transformed into colorless organogel within 10 min. However, the organogel formation was not noticed with hybrid APA-peptide **2a**. In the case of APA-peptide (**2d**), precipitation occurred at room temperature, however, upon heating converted into a homogenous solution. The hot homogenous solution was sonicated to form organogel within 2 minutes by allowing to cool at room temperature. In the literature, precipitates can also help in the formation of larger aggregates which can transform into gels (Li et al., 2007; Cravotto and Cintas, 2009). Importantly, the physical appearances of organogels of APA-peptides are different, such as transparent or opaque. We repeated a similar experiment with other solvents such as hexane, ethylacetate, benzene, chloroform, acetonitrile, and methanol but could not observe the gel formation. Mostly these peptides are sparingly soluble/or appeared as precipitates in those solvents. Moreover, for NMR studies, we attempted the organogelation of peptide **2e** in the deuterated solvent toluene-d_8, and the formation of organogel was noticed after 2–3 days. The APA-peptide organogels are stable up to 50–55°C. At higher temperatures (above ~55 °C), these gels are melted and eventually result in clear solutions. The formation of organogels was validated by the widely accepted inverted test tube method (Wang et al., 2003; Nagahama et al., 2008; Yoshida et al., 2014). Importantly, gelation is not observed when the β-Ala residue in APA-peptide is replaced with α-amino acid residues such as Gly, Ilu, and Ala (Bollu and Sharma, 2019b), indicating that the presence of β-Ala at that position is crucial for the formation of gels. Possibly, the presence of the β-Ala (extra methylene) group increases the chain length (extra–CH$_2$-), which affects the intramolecular H-bonding interactions and flexibility in sol-state that can reorganize easily during the formation of the rigid gel networking aggregates, appearing as physical gels (Dado and Gellman, 1994; Roy et al., 2004; Chatterjee et al., 2007). However under similar conditions, compound **2a** (Gly residue at the C-terminal) did not form a gel.

Morphology
We studied the surface morphology of APA-peptide organogels by TEM (Transmission Electron Microscope) and SEM (Scanning Electron Microscope) imaging techniques. Their TEM images are depicted in **Figure 2**, while SEM images are provided in the SM (**Supplementary Figure S20**). We also inverted the sample vials containing APA-peptide organogel to confirm the formation of organogels (**Figure 2**). The TEM image of organogel of APA-peptide (**2b**) shows the formation of
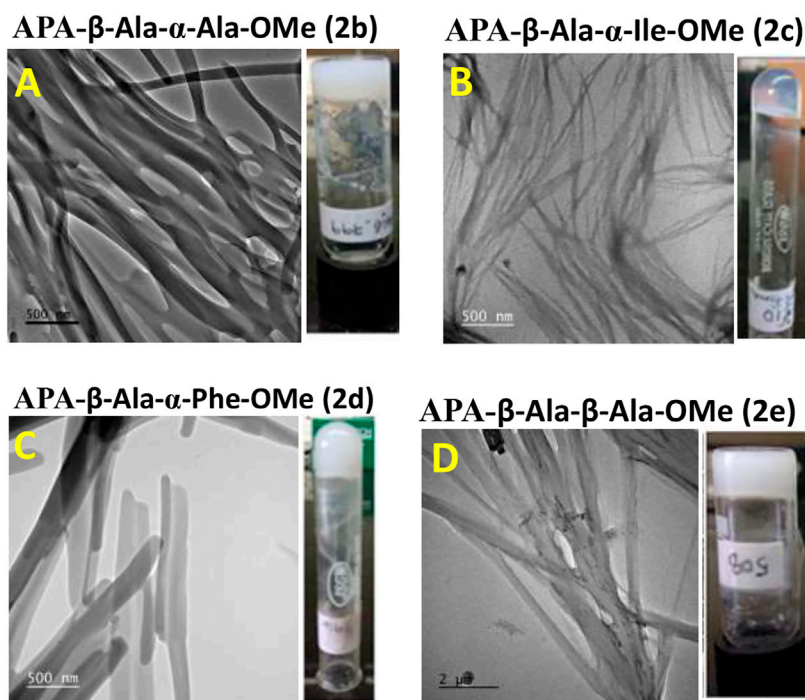
FIGURE 2 | TEM image of APA-peptide organogels. Peptide APA-β-Ala-α-Ala-OMe, **2b (A)**; peptide APA-β-Ala-α-Ile-OMe, **2c (B)**; peptide APA-β-Ala-α-Phe-OMe, **2d (C)**; and peptide APA-β-Ala-β-Ala-OMe, **2e (D)**.

supramolecular self-assembly structure as a group of thick long linear fiber-forming complex structure (**Figure 2A**). The organogel of APA-peptide (**2c**) forms a supramolecular self-assembly structure as a thin short linear fiber structure (**Figure 2B**). The organogel of APA-peptide (**2d**) forms a supramolecular self-assembly structure as a small strip-type structure (**Figure 2C**). The organogel of APA-peptide (**2e**) forms a supramolecular self-assembly structure as a long rod-type structure (**Figure 2D**).
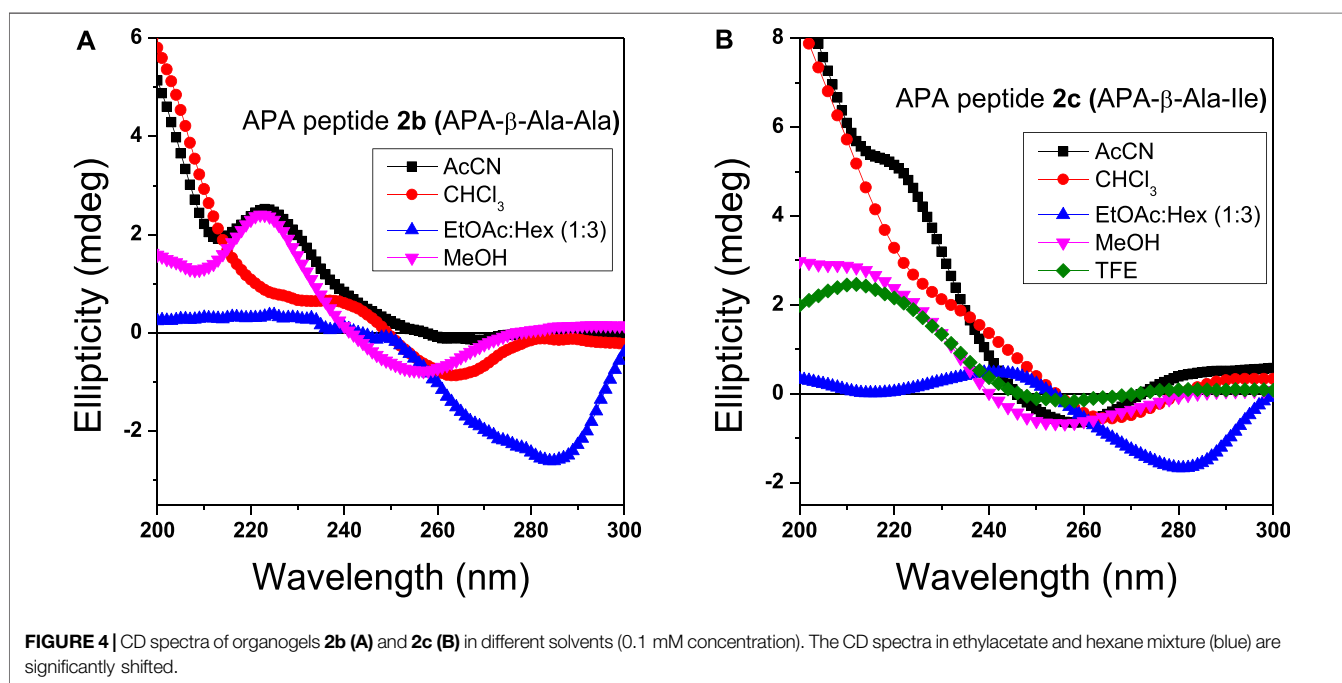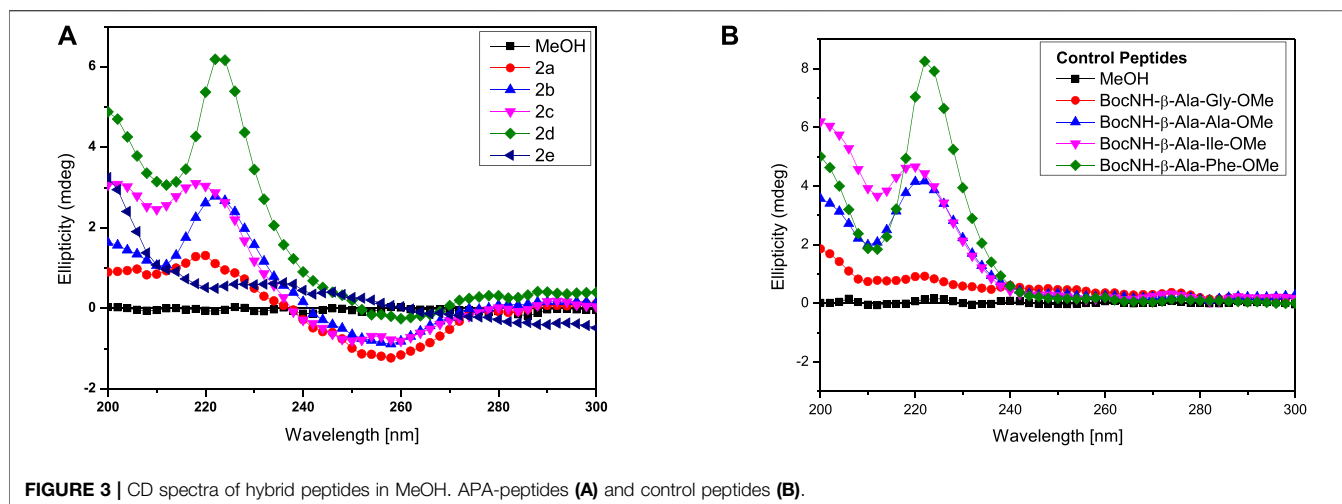
Thermogravimetric Analysis

TGA of APA-peptides **2b-2e** in xerogel (dried organogel) and powder forms is measured with increasing temperature (with 5°C/min). (Haines, 1995). From TGA plots, we also extracted differential thermogravimetric (DTG) (Thürmer et al., 2014) plots (first order derivative plots), and all these plots are provided in the SM (**Supplementary Figure S25**). In xerogel and powder forms, weight loss from trapped solvent evaporation is observed below 100°C. In xerogel and powder form of peptides **2b-2d**, significant weight loss transitions are observed with two peaks between 200 and 300°C, whereas in APA-peptides **2e**, these weight loss transition peaks are observed at 170–225°C. Importantly, all APA-peptides in the xerogel form exhibit higher weight loss temperatures than the respective powder forms. Presumably, these weight loss peaks are either due to the loss of the sensitive Boc-protecting group or decompositions. These TGA and DTG plots demonstrate slightly enhanced stability of xerogels than their respective powder forms.

UV Studies

We attempted to record the UV-vis spectra of APA-peptides (**2a-2e**) in polar/non-polar solvents. The UV-vis spectra of peptides **2a/2e** in MeOH exhibit an absorption peak at λ_{245nm} owing to the pyrazolone ring (**Supplementary Material**, **Supplementary Figure S26**). However, we were unable to record the UV spectra of peptides in ethylacetate and hexane owing to the poor solubility/precipitation.

Circular Dichroism Studies

Circular Dichroism (CD) studies reveal the configuration and chirality of molecules including the nature of regular secondary structure (α-helix and β-strand) in protein, peptides, hydro-/organo-gels, and other chiral self-assembly materials (López Deber et al., 2014). However, the structure and conformation of peptides are sensitive to the nature of the solvent environment, which plays a significant role in peptides' secondary structure formation (Cerpa et al., 1996; Awasthi et al., 2001). Previously, we have reported that the APA residue is involved in conformational changes of APA-peptides. We recorded the CD spectra of APA-peptides (**2a–2e**) of 0.1 mM concentrations in different solvent systems such as AcCN, MeOH, CHCl₃, and TFE. Their CD spectra in polar solvent MeOH are provided in **Figure 3**, while their CD spectra in other solvents are provided in the SM (**Supplementary Figures S12–S14**). For control studies, we also recorded the CD spectra of control peptides, without containing the APA-residue (**Figure 3B**, **Supplementary Figures S15–S18**). In MeOH solvent (polar protic), the CD spectra of APA-

FIGURE 3 | CD spectra of hybrid peptides in MeOH. APA-peptides **(A)** and control peptides **(B)**.

FIGURE 4 | CD spectra of organogels **2b (A)** and **2c (B)** in different solvents (0.1 mM concentration). The CD spectra in ethylacetate and hexane mixture (blue) are significantly shifted.

peptides (**2a–2e**) exhibit maxima at wavelength (λ) 220 nm (λ_{220nm}) and minima at λ_{260nm}. In contrast, the CD spectra of control peptides (in MeOH) exhibit only maxima at ~λ_{220nm}. The CD spectra of APA-peptides (**2a-2e**) exhibit almost similar CD signal maxima (λ_{220nm}) and minima (λ_{250nm}) in aprotic polar solvent acetonitrile (AcCN). However, the CD signals of APA-peptides (2a–2e) exhibit poorly resolved maxima and minima in solvent chloroform (CHCl$_3$), and only maxima (λ_{200nm} and λ_{220nm}) are observed in solvent trifluoroethanol (TFE). In the literature, TFE is well known to induce intramolecular hydrogen bonding which stabilizes possible helical structures, and such kind of CD structure is not observed with APA-peptides (**2a-2e**) (Sonnichsen et al., 1992). The CD signals of these peptides are possibly due to

electronic transitions of the amide carbonyl group (π–π^*/n–π^*) at ~λ_{220nm} and pyrazolonyl/phenyl aromatic rings (π–π^*) at λ_{250nm}. From the CD spectra of APA-peptides, overlapping of aromatic chromophoric (pyrazolonyl/phenyl) absorption (220–280 nm) with the finger print region of peptide secondary structure (190–240 nm) is observed. This made the interpretation of the secondary structures difficult. However, the maxima at ~λ_{220nm} in APA-peptides (**2a-2e**) are presumed from the characteristics β-type of secondary structures (Maity et al., 2015).

We also studied the CD spectra of organogels of representative APA-peptides (**2b/2c**) in the co-solvent system EtOAc:Hexane (1:3, v/v) and other different polarity solvents such as AcCN, CHCl$_3$, MeOH, and TFE (**Figure 4**).
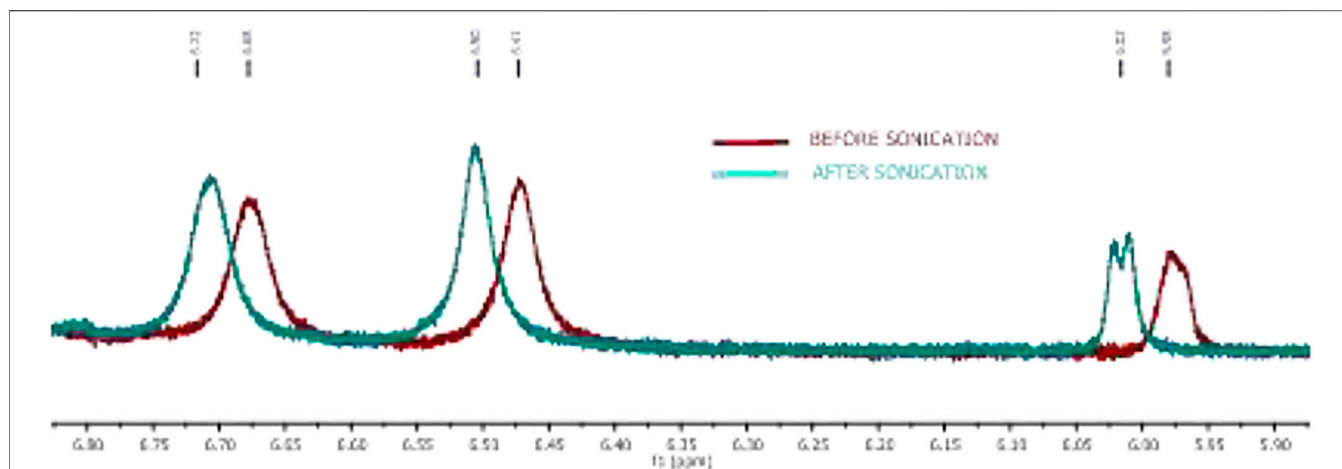
FIGURE 5 | Expanded NMR of the peptide **2c** amide N-H region of organogel in toluene-d_8 before (red) and after sonication (turquoise).

The CD spectra of APA-peptide organogels (**2b/2c**) in the co-solvent system EtOAc:Hexane (1:3, v/v) exhibit only minima at ~λ_{290nm}, and remarkable red-shift from λ_{260nm} strongly supports the existence of strong π–π interactions, possibly between two aromatic moieties (phenylpyrazolonyl unit) in organogel (**Figures 4A,B**). However the CD spectra of APA-peptides (**2b/2c**) in MeOH/TFE exhibit maxima (λ_{260nm}) and minima (λ_{260nm}). The CD spectra of that peptide organogel in other solvents are relatively non-characteristic. The solvent polarity and interaction of these solvents with APA-peptides resulted in diverse CD structures. The CD structures of APA-peptides in other solvents are presumed due to intermolecular H-bonding; this is further supported by our NMR and X-ray studies.

NMR-Studies

In the literature, the formation of peptide organogels is also studied by ^{1}H-NMR in the deuterated solvent (toluene-d_8) which exhibits a significant downfield chemical shift of amide N-H (Maity et al., 2015). We performed similar NMR studies with representative organogel-forming APA-peptide **2c** in the NMR solvent, toluene-d_8 (**Figure 5**). The NMR spectra of amide N-H proton of peptide **2c** before/after organogelation are depicted in **Figure 5** that exhibit the notable chemical shift in those amide N-H protons after sonication. This indicates that amide N-H is involved in hydrogen bonding after sonication that provides a relatively stronger hydrogen bonding environment. Similar NMR experiments were attempted with other organogel-forming peptides (**2b/2d/2e**) but were unable to record ^{1}H-NMR in the same solvent, toluene-d_8, because of instant solubility/precipitation.

APA-peptides (**2a-2e**) have three amide bonds which can involve in the hydrogen bonding network. We recorded 2D-NMR (^{1}H-COSY) spectra for representative APA-peptide (**2c**) in CDCl$_3$ and assigned their NH protons chemical shifts (δ) as Boc-NH (δ6.24), Ile-NH (δ6.41), and β-Ala-NH (δ7.31) (**Supplementary Figure S22A**). Notably, the β-Ala-NH is overlapped with aromatic protons; the cross-peaks in the

^{1}H-COSY experiment are used to assign its chemical shift value. To study the amide bonds of APA-peptide (**2c**) involved in the hydrogen bonding network, we performed the ^{1}H-NMR DMSO-d_6 titration experiment in CDCl$_3$ (**Supplementary Figure S23**)·(Malik et al., 2002; Balachandra and Sharma, 2014; Bollu and Sharma, 2019b) Since the amide bond (β-Ala-NH) appeared in the aromatic region, after DMSO-d_6 titration, we again recorded 1H-COSY to confirm the respective amide NH (**Supplementary Figure S22B**). From these titration 1H-NMR spectra, with increasing concentration of DMSO-d_6 (up to 19 μL), Boc-NH and Ile-NH exhibit a significant downfield shift; however, β-Ala-NH (appended at APA moiety) shows a marginal shift (**Supplementary Figure S24**). It appears that Boc-NH and Ile-NH are involved in intermolecular hydrogen bonding, and β-Ala-NH is involved in intramolecular hydrogen bonding for the formation of the secondary structure.

FT-IR Studies of Organogels

FT-IR spectral analyses also support the formation of organogel in the sequence-specific peptides (Malik et al., 2002; Maity et al., 2015). It is reported that the IR frequency of free N-H stretching (Amide-A band) appears at ~3400 cm^{-1}, while hydrogen-bonded N-H appears at a lower frequency of ~3300 cm^{-1}s (Vass et al., 2003; Adochitei and Drochioiu, 2011). Also, the frequency of free amide-1 band (C=O stretching vibration) appears at 1680 cm^{-1}, while hydrogen-bonded C=O vibration appears at lower frequency ~1650 cm^{-1} in organogels/xerogels (Bardelang et al., 2008; Maity et al., 2015). Importantly, IR peaks in organogel/xerogel are more structured than those in synthesized peptides. To prevent the self-aggregations through intermolecular hydrogen bonding, we planned to record the IR spectra of APA-peptides (**2a-2e**) organogel in hexafluoroisopropanol (HFIP) solvent. Thus, we recorded the FT-IR spectra of clear xerogels of APA-peptides (**2a-2e**) and compared with IR spectra of organogel in HFIP solvent. Their carbonyl and amide region spectra are depicted in **Figure 6**, while their whole spectra are provided in the SM (**Supplementary Figure S19**). The FT-IR spectra of clear
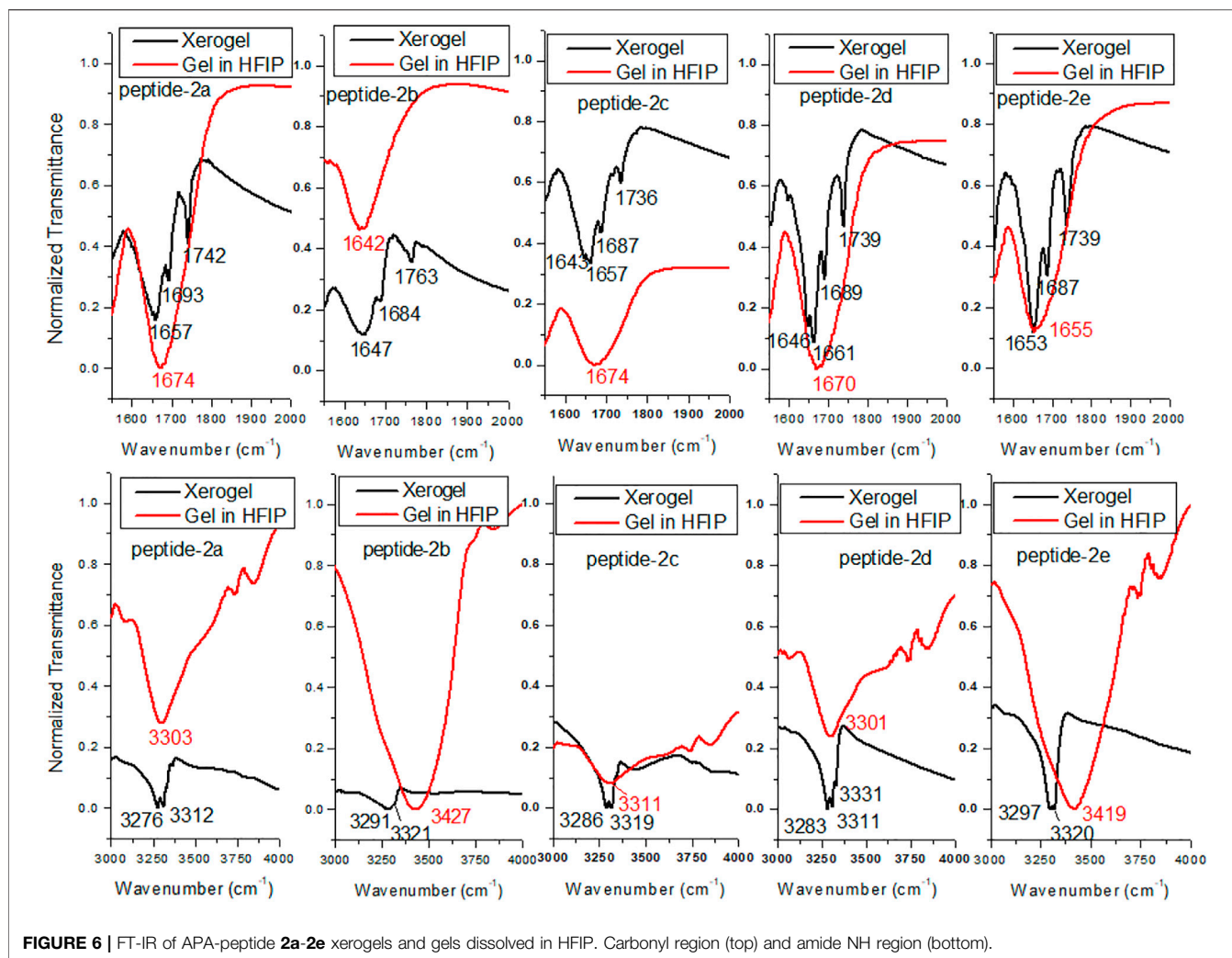
FIGURE 6 | FT-IR of APA-peptide **2a**-**2e** xerogels and gels dissolved in HFIP. Carbonyl region (top) and amide NH region (bottom).
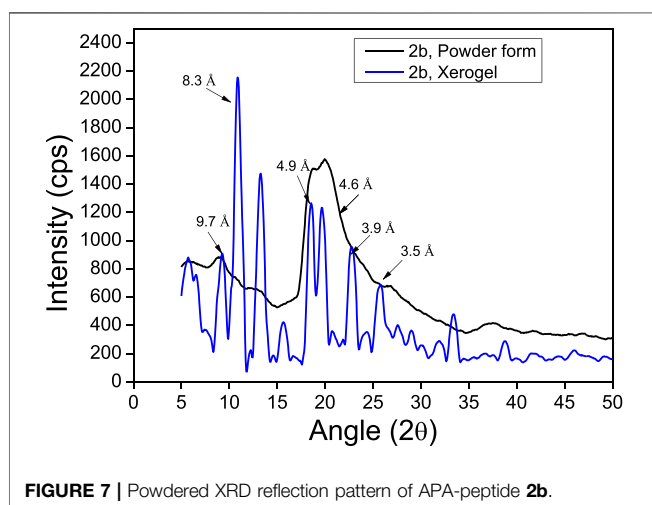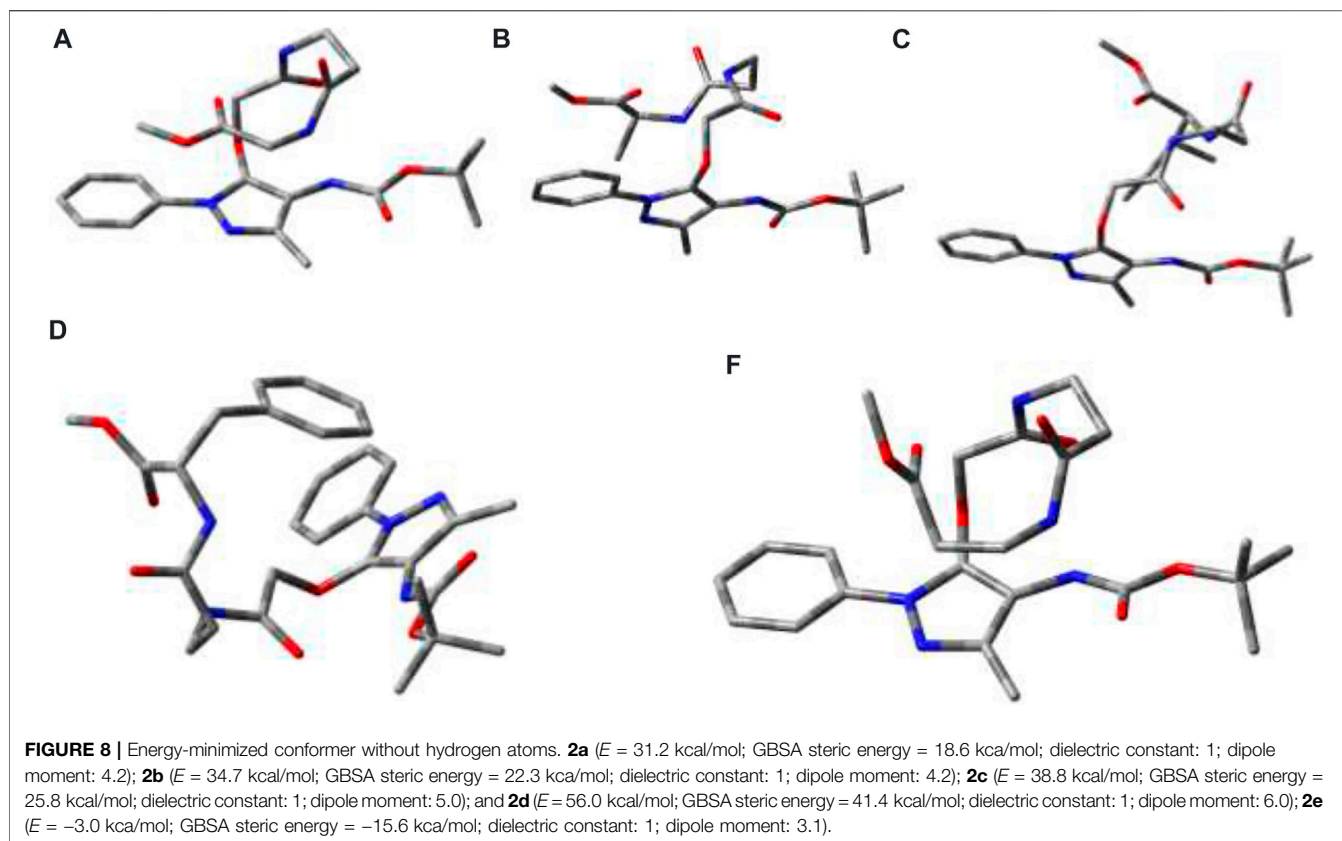

FIGURE 7 | Powdered XRD reflection pattern of APA-peptide **2b**.

xerogels of APA-peptides (**2a**–**2e**) exhibit resolved peaks at ~1,645–1,657 cm^{-1}, ~1,684–1,693 cm^{-1}, and ~1736–1763 cm^{-1} which belong to the stretching frequency of the amide carbonyl,

carbamate carbonyl, and ester carbonyl, respectively. However, the FT-IR spectra of those organogels in HFIP solvent exhibit a non-resolved broad peak at 1674 cm^{-1}. We also found that the N-H (Amide-A) stretching vibrations appear at ~3,276–3,312 cm^{-1}, which is lower than free N-H stretching frequency (~3400 cm^{-1}). In the literature, the β-sheet-forming peptides show amide-1 (amide carbonyl) stretching frequency at ~1625–1650 cm^{-1}, while α-helix-forming peptides at ~1,650–1,660 cm^{-1} that is lower than free amide carbonyl stretching frequency (~1680 cm^{-1} (Vass et al., 2003; Yuran et al., 2012). The FT-IR spectra of other xerogel peptides/organogel in HFIP are almost same. Thus, our FT-IR spectral analyses support the formation of secondary structure as α-helix/β-sheet types in xerogel (**2a**-**2e**).

X-Ray Diffraction Analysis

The powder X-ray diffraction studies are used to confirm the supramolecular self-assembly structure in xerogels including peptide-based xerogels (dry organogels) (Marchesan et al., 2012; Marchesan et al., 2014). A typical peptide xerogel exhibit sharp reflection peaks at 5–35° 2θ (reflection angle) range, while

FIGURE 8 | Energy-minimized conformer without hydrogen atoms. **2a** (E = 31.2 kcal/mol; GBSA steric energy = 18.6 kca/mol; dielectric constant: 1; dipole moment: 4.2); **2b** (E = 34.7 kcal/mol; GBSA steric energy = 22.3 kca/mol; dielectric constant: 1; dipole moment: 4.2); **2c** (E = 38.8 kcal/mol; GBSA steric energy = 25.8 kca/mol; dielectric constant: 1; dipole moment: 5.0); and **2d** (E = 56.0 kcal/mol; GBSA steric energy = 41.4 kca/mol; dielectric constant: 1; dipole moment: 6.0); **2e** (E = −3.0 kca/mol; GBSA steric energy = −15.6 kca/mol; dielectric constant: 1; dipole moment: 3.1).

non-xerogel peptides (synthetic) exhibit broad reflection peaks at 20° 2θ range. We also performed a powder X-ray diffraction experiment with organogel-forming APA-peptides (**2b-2c**). We recorded the X-ray diffraction (XRD) spectra of peptides **2b-2e** in powder form (before organogelation) and their respective xerogel. The XRD spectra of the APA-peptide (**2b**) solid powder (before/after gelation) are depicted in **Figure 7**, while for other APA-peptides (**2c-2e**) are provided in the SM (**Supplementary Figure S21**). The XRD spectra of peptide (**2b**) show that its xerogel powders are structurally organized than the powder before gelation. We calculated d-spacing values in angstroms (Å) from their experimental 2θ reflection peaks by applying Bragg's equation (nλ = 2dSinθ) (Bardelang et al., 2008; Marchesan et al., 2012). Their d-spacing values are provided above the reflection peaks. In xerogel spectra, the reflection peaks at 4.5–5.1Å are characteristics for hydrogen bonding between β-strands, while peaks at 9.7–10.8Å are associated with the distance between anti-parallel strands (i.e. every other strand) or to inter-sheet distances. The peaks at 3.8–4.2Å are attributed to π-π stacking possible from aromatic N-phenyl pyrazole rings (Marchesan et al., 2012; Marchesan et al., 2014). Thus, APA-peptide xerogels have the β-sheet type of structure in their supramolecular self-assembly structure.

Conformational Studies

Global-MMX (GMMX) is a steric energy minimization program that uses the supported force field (MMX, MM3, or MMFF94)

and operates in batch mode to search conformational space and to list the lowest energy unique conformers. The generalized born/surface area (GB/SA) solvation model gives free energies of aqueous solvation (Cheng et al., 2000). GMMX and GBSA solvation calculation models are being frequently applied to find the energy-minimized conformation of peptides in the gas phase and water medium (Lee et al., 2001; Biswas et al., 2013). We performed the theoretical calculation to find the energy-minimized conformation into the gas phase and solution with GMMX and GBSA solvation methods with the MMFF94 force field. The details are proved in the Supplementary Material. The structurally minimized conformers of APA-peptides (**2a-2e**) without hydrogen atoms are provided in **Figure 8** while with hydrogen atoms in **Supplementary Figure S27**. The stabilization energy of APA-peptide (**2a-2e**) solution phase (dielectric constant, equivalent to water) is lower than that of the gas phase by 12~kcal/mol without affecting the significant changes in structural conformation. Importantly, we could not find *intramolecular* hydrogen bonding in the energy-minimized conformers of APA-peptides. Their phenyl-aminopyrazolone residues are planar, and the polar groups are exposed in solvents which could participate in the intermolecular hydrogen bonding with other molecules. Generally, intramolecular hydrogen bonding prevents the formation of organogels. Presumably, these APA-peptides form *intermolecular* hydrogen bonding in the organic solvent system (EtOAc:Hexane) after sonication and produce

organogels. Hence, APA-peptides have the ability to form organogel.

CONCLUSION

Aminopyrazolonyl amino acid (APA) containing α-/β-hybrid peptides are explored further for supramolecular self-assembly structure by the formation of organogel in the organic solvent system. Most of them form organogels, but their physical appearances are different such as opaque and translucent. These organogels are characterized as β-sheet types of the structure by NMR, IR, CD, powder-XRD, TGA, SEM, and TEM techniques. Theoretically, the energy-minimized structure suggests that there is no *intra*molecular hydrogen bonding in the polar solvent. There could be possibility of the formation of intermolecular hydrogen bonding after sonication in the organic solvent which leads to the formation of organogel in the EtOAc:Hexane solvent system. Hence, the APA acid could be employed at the *N*-terminal of target di-/tri-peptides for organogelation in the organic co-solvent (EtOAc:Hexane) system.

EXPERIMENTAL DETAILS

Materials

All required materials were obtained from commercial suppliers and used without any further purification. Dimethylformamide was distilled with calcium hydride. Reactions were monitored by TLC (thin layer chromatography) and visualized by UV and ninhydrin. Column chromatography was performed in a 230–400 mesh silica gel. Mass spectra and HRMS were obtained using the Bruker micrOTOF-Q II spectrometer. ^{1}H NMR and ^{13}C NMR were recorded on Bruker AV-400 or 700 MHz at 298 K. ^{1}H and ^{13}C NMR chemical shifts were recorded in ppm downfield from tetramethylsilane or residual solvent peak. Splitting patterns are abbreviated as follows: s, singlet; d, doublet; dd, doublet of doublet; t, triplet; q, quartet; dq, doublet of the quartet; and m, multiplet. Powder X-ray diffraction data were collected on a Bruker D8 Advance with DA VINCI design fitted with an HTK 16 temperature chamber X-ray powder diffractometer using CuKα radiation (λ = 1.5418 Å). Transmission electron microscopy (TEM) data were recorded using JEOL 2100F.

General Experimental Procedure for Compounds (2a-2e)

The experimental procedures for the synthesis of control dipeptides and APA-peptides (**2a-2e**) were followed from the literature. (Bollu and Sharma, 2019b).

APA-β-Ala-Gly-OMe (**2a**). R_f 0.18 (0.4:9.6 MeOH/CH$_2$Cl$_2$); yield 68%; ^{1}H NMR (400 MHz, DMSO) δ 8.37 (s, 1H), 8.17 (s, 1H), 8.05 (s, 1H), 7.71 (d, *J* = 7.5 Hz, 2H), 7.46 (t, *J* = 7.6 Hz, 2H), 7.30 (t, *J* = 7.1 Hz, 1H), 4.60 (s, 2H), 3.83 (d, *J* = 5.5 Hz, 2H), 3.61 (s, 3H), 3.32 (s, 2H), 2.36 (t, *J* = 6.9 Hz, 2H), 2.03 (s, 3H), and 1.44 (s, 9H); ^{13}C NMR (176 MHz, DMSO) δ 171.56, 170.95, 167.16, 162.16, 155.30, 147.65, 146.78, 138.90, 129.55, 126.76, 122.28, 103.74, 79.39, 70.54, 52.22, 45.65, 41.05, 35.63, 35.22, 28.61, and

12.40. HRMS (ESI-TOF) m/z [M + H]$^+$ Calcd for C$_{23}$H$_{31}$N$_5$O$_7$ 490.2296; found 490.2295.

APA-β-Ala-Ala-OMe (**2b**). R_f 0.33 (0.4:9.6 MeOH/CH$_2$Cl$_2$); yield 70%; ^{1}H NMR (400 MHz, DMSO) δ 8.35 (d, *J* = 6.9 Hz, 1H), 8.17 (s, 1H), 8.04 (s, 1H), 7.70 (d, *J* = 7.8 Hz, 2H), 7.46 (t, *J* = 7.9 Hz, 2H), 7.30 (t, *J* = 7.3 Hz, 1H), 4.60 (s, 2H), 4.26 (p, *J* = 7.2 Hz, 1H), 3.61 (d, *J* = 5.4 Hz, 3H), 3.40–3.21 (m, 3H), 2.39–2.27 (m, 2H), 2.02 (s, 3H), 1.43 (s, 9H), 1.25 (d, *J* = 7.3 Hz, 4H). ^{13}C NMR (176 MHz, DMSO) δ 173.75, 170.98, 167.17, 155.32, 147.66, 146.80, 138.93, 129.58, 126.79, 122.30, 103.77, 79.41, 70.55, 52.40, 48.06, 35.63, 35.20, 28.64, 17.49, and 12.43. HRMS (ESI-TOF) m/z [M + Na]$^+$ Calcd for C$_{24}$H$_{33}$N$_5$O$_7$Na 526.2272; found 526.2272.

APA-β-Ala-Ile-OMe (**2c**). R_f 0.35 (0.3:9.7 MeOH/CH$_2$Cl$_2$); yield 73%; ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.61 (d, *J* = 7.9 Hz, 2H), 7.43 (t, *J* = 7.7 Hz, 2H), 7.37–7.24 (m, 1H), 6.34 (d, *J* = 72.2 Hz, 1H), 4.57 (m, *J* = 41.5, 26.2 Hz, 3H), 3.67 (d, *J* = 27.7 Hz, 4H), 3.51 (s, 1H), 2.48 (d, *J* = 4.8 Hz, 2H), 2.17 (s, 3H), 1.81 (s, 1H), 1.48 (s, 9H), 0.89 (d, *J* = 6.0 Hz, 6H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 172.81, 172.04, 167.72, 155.29, 147.54, 146.92, 138.24, 129.04, 126.72, 122.38, 102.45, 80.59, 70.89, 56.91, 52.08, 37.06, 35.57, 28.25, 25.25, 15.50, 11.87, and 11.40. HRMS (ESI-TOF) m/z [M + H]$^+$ Calcd for C$_{27}$H$_{40}$N$_5$O$_7$ 546.2933; found 546.2812.

APA-β-Ala-Phe-OMe (**2d**). R_f 0.36 (0.4:9.6 MeOH/CH$_2$Cl$_2$); yield 79%; ^{1}H NMR (400 MHz, DMSO) δ 8.42 (d, *J* = 7.7 Hz, 1H), 8.16 (s, 1H), 7.98 (s, 1H), 7.69 (d, *J* = 7.7 Hz, 2H), 7.45 (t, *J* = 7.7 Hz, 2H), 7.33–7.16 (m, 7H), 4.58 (s, 2H), 4.47 (dd, *J* = 13.9, 8.2 Hz, 1H), 3.60 (s, 3 Hz), 3.22 (dd, *J* = 20.2, 10.2 Hz, 2H), 3.01 (dd, *J* = 13.6, 5.3 Hz, 1H), 2.88 (dd, *J* = 13.4, 9.5 Hz, 1H), 2.30 (m, *J* = 14.6, 7.4 Hz, 2H), 2.02 (s, 3H), 1.43 (s, 9H). ^{13}C NMR (176 MHz, DMSO) δ 172.66, 172.57, 171.15, 170.40, 167.16, 155.33, 147.67, 146.81, 138.90, 137.77, 129.59, 128.80, 127.13, 126.81, 122.32, 103.72, 79.43, 70.53, 54.12, 52.41, 45.69, 37.24, 36.33, 35.58, 35.21, 34.29, and 28.63, 12.42. HRMS (ESI-TOF) m/z [M + H]$^+$ Calcd for C$_{30}$H$_{38}$N$_5$O$_7$ 580.2766; found 580.2760.

APA-β-Ala-β-Ala-OMe (**2e**). R_f 0.28 (0.3:9.7 MeOH/CH$_2$Cl$_2$); yield 60%; ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.56 (d, *J* = 7.6 Hz, 2H), 7.41 (t, *J* = 7.7 Hz, 2H), 7.36–7.14 (m, 2H), 6.45 (s, 1H), 6.35 (s, 1H), 4.58 (s, 2H), 3.66 (s, 3H), 3.60–3.43 (m, 2H), 2.44 (d, *J* = 40.0 Hz, 2H), 2.18 (s, 3H), 1.47 (s, 9H). HRMS (ESI-TOF) m/z [M + H]$^+$ Calcd for C$_{24}$H$_{34}$N$_5$O$_7$ 504.2453; found 504.2453.

Organogelation

A measure of 10 mg of APA-peptides (**2a-2e**) was dissolved in 1 ml of hexane–ethylacetate (3:1) solvent mixture and sonicated at 50°C for 2 min and then allowed to cool at room temperature. Under these conditions, APA-peptides (**2b-2e**) formed organogels.

Field Emission Scanning Electron Microscopy

A measure of 10 mg of APA-peptides (**2b-2e**) was dissolved in 1 ml of hexane–ethylacetate (3:1) mixture and sonicated at 50°C for 2 min. Then, the gel was casted on the silicon wafer and dried under high vacuum, and SEM images were obtained at 3.00 kV.

Field Emission Transmission Electron Microscopy

A measure of 10 mg of APA-peptides (**2b-2e**) was dissolved in 1 ml of hexane–ethylacetate (3:1) mixture and sonicated at 50°C

for 2 min. Then, the gel was diluted 3–4 times and casted on a copper grid and dried under high vacuum; TEM images were obtained.

Circular Dichroism Spectroscopy

CD spectra were recorded in degassed CH_3OH, AcCN, $CHCl_3$, CF_3CH_2OH, and hexane–ethylacetate (3:1) at 20°C from 300–200 nm with peptide concentrations of 0.1 mM. CD data are collected with following parameters: data pitch at 2 nm, DIT for 2 s, bandwidth at 2 nm, and scanning speed at 100 nm/min.

Fourier-Transform Infrared Spectroscopy

Peptide gels were drop-casted on the KBr window and dried under high vacuum. For HFIP, xerogels were dissolved in HFIP and drop-casted on the KBr window and then dried under high vacuum. The spectra are the average of 250 scans.

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**.

REFERENCES

Adochitei, A., and Drochioiu, G. (2011). Rapid Characterization of Peptide Secondary Structure by FT-IR Spectroscopy. *Rev. Roum. Chim.* 56, 783–791.

Aggeli, A., Bell, M., Boden, N., Keen, J. N., Knowles, P. F., Mcleish, T. C. B., et al. (1997). Responsive Gels Formed by the Spontaneous Self-Assembly of Peptides into Polymeric β-sheet tapes. *Nature* 386, 259–262. doi:10.1038/386259a0

Awasthi, S. K., Shankaramma, S. C., Raghothama, S., and Balaram, P. (2001). Solvent-induced β-hairpin to helix Conformational Transition in a Designed Peptide. *Biopolymers* 58, 465–476. doi:10.1002/1097-0282(20010415)58:5<465::aid-bip1022>3.0.co;2-t

Babu, S. S., Praveen, V. K., and Ajayaghosh, A. (2014). Functional π-Gelators and Their Applications. *Chem. Rev.* 114, 1973–2129. doi:10.1021/cr400195e

Balachandra, C., and Sharma, N. K. (2014). Synthesis and Conformational Analysis of New Troponyl Aromatic Amino Acid. *Tetrahedron* 70, 7464–7469. doi:10.1016/j.tet.2014.08.014

Banerjee, A., Palui, G., and Banerjee, A. (2008). Pentapeptide Based Organogels: the Role of Adjacently Located Phenylalanine Residues in Gel Formation. *Soft Matter* 4, 1430–1437. doi:10.1039/b802205b

Baral, A., Roy, S., Dehsorkhi, A., Hamley, I. W., Mohapatra, S., Ghosh, S., et al. (2014). Assembly of an Injectable Noncytotoxic Peptide-Based Hydrogelator for Sustained Release of Drugs. *Langmuir* 30, 929–936. doi:10.1021/la4043638

Bardelang, D., Camerel, F., Margeson, J. C., Leek, D. M., Schmutz, M., Zaman, M. B., et al. (2008). Unusual Sculpting of Dipeptide Particles by Ultrasound Induces Gelation. *J. Am. Chem. Soc.* 130, 3313–3315. doi:10.1021/ja711342y

Biswas, G., Moon, H. J., Boratyński, P., Jeong, B., and Kwon, Y.-U. (2016). Structural Sensitivity of Peptoid-Based Low Molecular Mass Organogelator. *Mater. Des.* 108, 659–665. doi:10.1016/j.matdes.2016.07.059

Biswas, S., Abo-Dya, N. E., Oliferenko, A., Khiabani, A., Steel, P. J., Alamry, K. A., et al. (2013). Oxyazapeptides: Synthesis, Structure Determination, and Conformational Analysis. *J. Org. Chem.* 78, 8502–8509. doi:10.1021/jo401234g

Bollu, A., and Sharma, N. K. (2019a). Cleavable Amide Bond: Mechanistic Insight into Cleavable 4-Aminopyrazolyloxy Acetamide at Low pH. *J. Org. Chem.* 84, 5596–5602. doi:10.1021/acs.joc.9b00535

Bollu, A., and Sharma, N. K. (2019b). Synthesis and Conformational Analysis of Aminopyrazolonyl Amino Acid (APA)/Peptides. *Eur. J. Org. Chem.* 2019, 1286–1292. doi:10.1002/ejoc.201801640

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.821971/full#supplementary-material

Cerpa, R., Cohen, F. E., and Kuntz, I. D. (1996). Conformational Switching in Designed Peptides: the helix/sheet Transition. *Folding Des.* 1, 91–101. doi:10.1016/s1359-0278(96)00018-1

Chakraborty, T. K., Jayaprakash, S., Srinivasu, P., Madhavendra, S. S., Ravi Sankar, A., and Kunwar, A. C. (2002). Furanoid Sugar Amino Acid Based Peptidomimetics: Well-Defined Solution Conformations to Gel-like Structures. *Tetrahedron* 58, 2853–2859. doi:10.1016/s0040-4020(02)00158-8

Chatterjee, S., and Maitra, U. (2017). Hierarchical Self-Assembly of Photoluminescent CdS Nanoparticles into a Bile Acid Derived Organogel: Morphological and Photophysical Properties. *Phys. Chem. Chem. Phys.* 19, 17726–17734. doi:10.1039/c7cp02519j

Chatterjee, S., Roy, R. S., and Balaram, P. (2007). Expanding the Polypeptide Backbone: Hydrogen-Bonded Conformations in Hybrid Polypeptides Containing the Higher Homologues of α-amino Acids. *J. R. Soc. Interf.* 4, 587–606. doi:10.1098/rsif.2006.0203

Cheng, A., Best, S. A., Merz, K. M., Jr, and Reynolds, C. H. (2000). GB/SA Water Model for the Merck Molecular Force Field (MMFF). *J. Mol. Graphics Model.* 18, 273–282. doi:10.1016/s1093-3263(00)00038-3

Chetia, M., Debnath, S., Chowdhury, S., and Chatterjee, S. (2020). Self-assembly and Multifunctionality of Peptide Organogels: Oil Spill Recovery, Dye Absorption and Synthesis of Conducting Biomaterials. *RSC Adv.* 10, 5220–5233. doi:10.1039/c9ra10395c

Couffin-Hoarau, A.-C., Motulsky, A., Delmas, P., and Leroux, J.-C. (2004). In Situ-forming Pharmaceutical Organogels Based on the Self-Assembly of L-Alanine Derivatives. *Pharm. Res.* 21, 454–457. doi:10.1023/b:pham.0000019299.01265.05

Cravotto, G., and Cintas, P. (2009). Molecular Self-Assembly and Patterning Induced by Sound Waves. The Case of Gelation. *Chem. Soc. Rev.* 38, 2684–2697. doi:10.1039/b901840a

Dado, G. P., and Gellman, S. H. (1994). Intramolecular Hydrogen Bonding in Derivatives of .beta.-Alanine and .gamma.-Amino Butyric Acid; Model Studies for the Folding of Unnatural Polypeptide Backbones. *J. Am. Chem. Soc.* 116, 1054–1062. doi:10.1021/ja00082a029

Gilfillan, L., Artschwager, R., Harkiss, A. H., Liskamp, R. M. J., and Sutherland, A. (2015). Synthesis of Pyrazole Containing α-amino Acids via a Highly Regioselective Condensation/aza-Michael Reaction of β-aryl α,β-unsaturated Ketones. *Org. Biomol. Chem.* 13, 4514–4523. doi:10.1039/c5ob00364d

Haines, P. (1995). *Thermal Methods of Analysis*. London: Blackie Academic Professional.

Hanabusa, K., Tange, J., Taguchi, Y., Koyama, T., and Shirai, H. (1993). Small Molecular Gelling Agents to Harden Organic Liquids: Alkylamide of N-Benzyloxycarbonyl-L-Valyl-L-Valine. *J. Chem. Soc. Chem. Commun.*, 390–392. doi:10.1039/c39930000390

Hellmert, M., Müller-Schiffmann, A., Peters, M. S., Korth, C., and Schrader, T. (2015). Hybridization of an Aβ-specific Antibody Fragment with Aminopyrazole-Based β-sheet Ligands Displays Striking Enhancement of Target Affinity. *Org. Biomol. Chem.* 13, 2974–2979. doi:10.1039/c4ob02411g

Kirsten, C. N., and Schrader, T. H. (1997). Intermolecular β-Sheet Stabilization with Aminopyrazoles. *J. Am. Chem. Soc.* 119, 12061–12068. doi:10.1021/ja972158y

Lee, H.-J., Choi, K.-H., Ahn, I.-A., Ro, S., Jang, H. G., Choi, Y.-S., et al. (2001). The β-turn Preferential Solution Conformation of a Tetrapeptide Containing an Azaamino Acid Residue. *J. Mol. Struct.* 569, 43–54. doi:10.1016/s0022-2860(00)00861-9

Li, Y., Wang, T., and Liu, M. (2007). Ultrasound Induced Formation of Organogel from a Glutamic dendron. *Tetrahedron* 63, 7468–7473. doi:10.1016/j.tet.2007.02.070

Li, Z., Luo, Z., Zhou, J., Ye, Z., Ou, G.-C., Huo, Y., et al. (2020). Monopeptide-based Powder Gelators for Instant Phase-Selective Gelation of Aprotic Aromatics and for Toxic Dye Removal. *Langmuir* 36, 9090–9098. doi:10.1021/acs.langmuir.0c01101

Loic, S. (2017). "Amino Acids Modification to Improve and fine-tune Peptide-Based Hydrogels," in *Amino Acid-New Insights Roles Plant Animal* (London: IntechOpen), 31–73. doi:10.5772/intechopen.68705

López Deber, M. P., Hickman, D. T., Nand, D., Baldus, M., Pfeifer, A., and Muhs, A. (2014). Engineering Amyloid-like Assemblies from Unstructured Peptides via Site-specific Lipid Conjugation. *PLoS One* 9, e105641. doi:10.1371/journal.pone.0105641

Maity, S., Das, P., and Reches, M. (2015). Inversion of Supramolecular Chirality by Sonication-Induced Organogelation. *Sci. Rep.* 5, 16365. doi:10.1038/srep16365

Maity, S., Kumar, P., and Haldar, D. (2011). Sonication-induced Instant Amyloid-like Fibril Formation and Organogelation by a Tripeptide. *Soft Matter* 7, 5239–5245. doi:10.1039/c1sm05277b

Malik, S., Maji, S. K., Banerjee, A., and Nandi, A. K. (2002). A Synthetic Tripeptide as Organogelator: Elucidation of Gelation mechanismElectronic Supplementary Information (ESI) Available: the 500 MHz 1-D 1H NMR Spectrum, the 500 MHz 1H-1H DQF COSY Spectrum of the Tripeptide in CDCl3 and the MALDI-MS Spectrum of the Tripeptide. See http://www.rsc.Org/suppdata/p2/b1/b111598g/. *J. Chem. Soc. Perkin Trans.* 2, 1177–1186. doi:10.1039/b111598g

Marchesan, S., Easton, C. D., Styan, K. E., Waddington, L. J., Kushkaki, F., Goodall, L., et al. (2014). Chirality Effects at Each Amino Acid Position on Tripeptide Self-Assembly into Hydrogel Biomaterials. *Nanoscale* 6, 5172–5180. doi:10.1039/c3nr06752a

Marchesan, S., Waddington, L., Easton, C. D., Winkler, D. A., Goodall, L., Forsythe, J., et al. (2012). Unzipping the Role of Chirality in Nanoscale Self-Assembly of Tripeptide Hydrogels. *Nanoscale* 4, 6752–6760. doi:10.1039/c2nr32006a

Nagahama, K., Ouchi, T., and Ohya, Y. (2008). Temperature-Induced Hydrogels through Self-Assembly of Cholesterol-Substituted Star PEG-B-PLLA Copolymers: An Injectable Scaffold for Tissue Engineering. *Adv. Funct. Mater.* 18, 1220–1231. doi:10.1002/adfm.200700587

Rouse, C. K., Martin, A. D., Easton, C. J., and Thordarson, P. (2017). A Peptide Amphiphile Organogelator of Polar Organic Solvents. *Sci. Rep.* 7, 43668. doi:10.1038/srep43668

Roy, R. S., Karle, I. L., Raghothama, S., and Balaram, P. (2004). α,β Hybrid Peptides: A Polypeptide helix with a central Segment Containing Two Consecutive β-amino Acid Residues. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16478–16482. doi:10.1073/pnas.0407557101

Schrader, T., and Kirsten, C. (1996). Intermolecular Stabilisation of the β-sheet Conformation in Dipeptides. *Chem. Commun.*, 2089–2090. doi:10.1039/cc9960002089

Shaikh, H., Rho, J. Y., Macdougall, L. J., Gurnani, P., Lunn, A. M., Yang, J., et al. (2018). Hydrogel and Organogel Formation by Hierarchical Self-Assembly of Cyclic Peptides Nanotubes. *Chem. A Eur. J* 24, 19066–19074. doi:10.1002/chem.201804576

Sonnichsen, F. D., Van Eyk, J. E., Hodges, R. S., and Sykes, B. D. (1992). Effect of Trifluoroethanol on Protein Secondary Structure: an NMR and CD Study Using a Synthetic Actin Peptide. *Biochemistry* 31, 8790–8798. doi:10.1021/bi00152a015

Thürmer, M. B., Diehl, C. E., Brum, F. J. B., and Santos, L. a. D. (2014). Preparation and Characterization of Hydrogels with Potential for Use as Biomaterials. *Mater. Res.* 17, 109–113. doi:10.1590/1516-1439.223613

Tomasini, C., and Castellucci, N. (2013). Peptides and Peptidomimetics that Behave as Low Molecular Weight Gelators. *Chem. Soc. Rev.* 42, 156–172. doi:10.1039/c2cs35284b

Vass, E., Hollósi, M., Besson, F., and Buchet, R. (2003). Vibrational Spectroscopic Detection of Beta- and Gamma-Turns in Synthetic and Natural Peptides and Proteins. *Chem. Rev.* 103, 1917–1954. doi:10.1021/cr000100n

Wang, J., and Yan, X. (2018). "Peptide-Based Hydrogels/Organogels: Assembly and Application," in *Nano/Micro-Structured Materials for Energy and Biomedical Applications* (Springer), 205–226. doi:10.1007/978-981-10-7787-6_6

Wang, X.-Z., Li, X.-Q., Shao, X.-B., Zhao, X., Deng, P., Jiang, X.-K., et al. (2003). Selective Rearrangements of Quadruply Hydrogen-Bonded Dimer Driven by Donor-Acceptor Interaction. *Chem. Eur. J.* 9, 2904–2913. doi:10.1002/chem.200204513

Yoshida, Y., Takahashi, A., Kuzuya, A., and Ohya, Y. (2014). Instant Preparation of a Biodegradable Injectable Polymer Formulation Exhibiting a Temperature-Responsive Sol-Gel Transition. *Polym. J.* 46, 632–635. doi:10.1038/pj.2014.30

Yuran, S., Razvag, Y., and Reches, M. (2012). Coassembly of Aromatic Dipeptides into Biomolecular Necklaces. *Acs Nano* 6, 9559–9566. doi:10.1021/nn302983e

Zweep, N., and Van Esch, J. H. (2013). "CHAPTER 1. The Design of Molecular Gelators," in *Functional Molecular Gels*. Editors B. Escuder and J. F. Miravet (The Royal Society of Chemistry), 1–29. doi:10.1039/9781849737371-00001

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