

This is a repository copy of *Pyrrolysine Amber Stop Codon Suppression:Development and Applications*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/119828/>

Version: Accepted Version

Article:

Brabham, Robin Louis and Fascione, Martin Anthony orcid.org/0000-0002-0066-4419
(2017) *Pyrrolysine Amber Stop Codon Suppression:Development and Applications*.
Chembiochem. ISSN: 1439-7633

<https://doi.org/10.1002/cbic.201700148>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Pyrrolysine Amber Stop Codon Suppression: Development and Applications

Robin Brabham,^[a] and Martin A. Fascione*^[a]

^[a]York Structural Biology Lab, Department of Chemistry, University of York, Heslington Road, York, YO10 5DD, UK; martin.fascione@york.ac.uk



Robin Brabham was born in Southampton (UK) in 1993, and was awarded a MChem degree from the Department of Chemistry at the University of York, UK (July 2015). In October 2015 he commenced Ph.D. studies in the Fascione group, focusing on the development of new methods for protein bioconjugation using unnatural amino acid mutagenesis.



Martin Fascione received his Ph.D. from the University of Leeds in 2009, working under the tutelage of W. Bruce Turnbull on the stereoselective synthesis of 1,2-cis-glycosides. Following a postdoctoral period in Leeds, he was then awarded a Marie Curie International Outgoing Fellowship to study the mechanisms of carbohydrate-processing enzymes with Professor Steve Withers, FRS, at the University of British Columbia in Vancouver, Canada (2012-2013) and Professor Gideon Davies, FRS, FMedSci, at the University of York, UK (2013-2014). In August 2014 he took up a lectureship in the York Structural Biology Laboratory, within the Department of Chemistry.

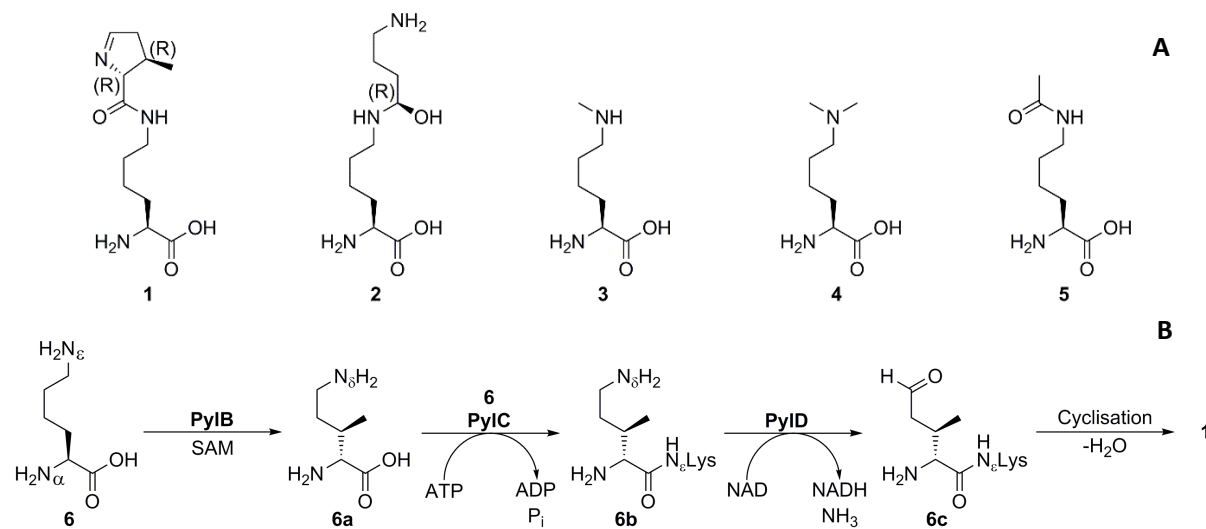
Abstract

The pyrrolysine tRNA synthetase-tRNA pair is likely one of the most promiscuous tRNA-synthetase pairs found in nature, capable of genetically encoding a plethora of non-canonical amino acids through stop codon reassignment. Proteins containing reactive handles, post-translational modification mimics or both can be produced in practical quantities, allowing *inter alia* the probing of biological pathways, the generation of antibody-drug conjugates, and enhancing protein function. This Minireview summarises the development of pyrrolysine amber stop codon suppression, presents some of the considerations required to utilise this technique to its greatest potential, and showcases the creative ways in which this technique has led to a better understanding of biological systems.

1. Pyrrolysine: The 22nd Canonical Amino Acid

The “22nd canonical amino acid” pyrrolysine (Pyl, Scheme 1a) **1** is a rare amino acid utilised by only a handful of organisms, primarily archaeal methanogens of the *Methanosarcinacea* family.^[1] In contrast to other lysine derivatives such as hypusine **2** and alkylated/acylated lysines **3-5** which are post-translational modifications (PTMs) of lysine, pyrrolysine is genetically encoded,^[2] with biosynthesis occurring prior to translation (Scheme 1b) starting from lysine **6** *via* intermediates **6a-c**^[3] utilising biosynthetic enzymes **PylB**,^[4] **PylC**^[5] and **PylD**.^[6] The usage of this amino acid is largely confined to the active site of methyltransferases, wherein the electrophilic imine group is used to capture methylamine and mediate the transfer of CH₃⁺ to cobalt(I) in the active site.^[7] A proposed reason for the evolution of this additional canonical amino acid system is the adaptation of various

Archaea to mono/di/trimethylamine-rich environments such as cattle rumen;^[8] indeed, the bacterium *Acetohalobium arabaticum* has been found to expand its genetic code to include pyrrolysine only in the presence of trimethylamine.^[9]



Scheme 1: **A** Native lysine derivatives found in proteins; **B** biosynthetic pathway of pyrrolysine.

Pyrrolysine is encoded by the amber stop codon, TAG in DNA and CUA anticodon on tRNA; hence incorporation into proteins has necessitated overriding the stop function of this codon. This has occurred through the natural evolution of an orthogonal tRNA-tRNA synthetase (RS) pair in pyrrolysine-utilising organisms. Pyrrolysine is delivered to the ribosome during translation in the form of Pyl-tRNA^{Pyl}, a complex formed through the charging of tRNA^{Pyl} with pyrrolysine by PylRS, which has the corresponding CUA anticodon. PylRS is a Class II tRNA synthetase, a homodimer in the active form, with an amino acid binding pocket which extends deep into the protein and exhibits significant hydrophobic character (Figure 1).^[10]

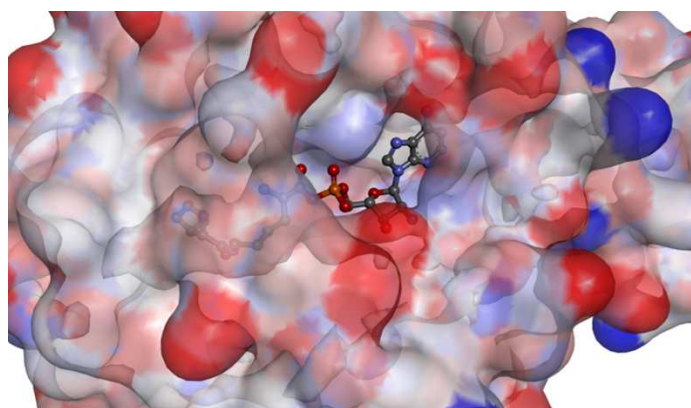


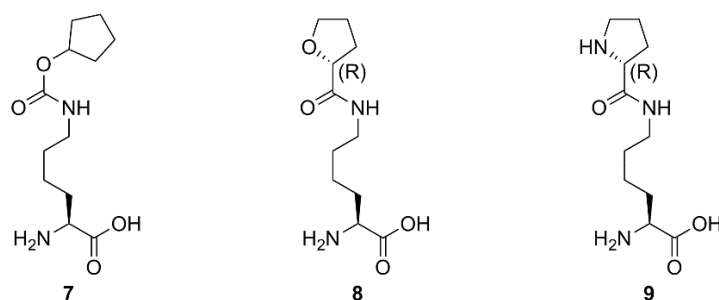
Figure 1: the hydrophobic amino acid binding pocket extends far inside PylRS (*Methanosarcina mazei*). PDB: 2Q7H.^[10]

As is expected of a tRNA synthetase, specificity for the substrate amino acid versus other native amino acids is very high: no other canonical amino acids are recognised by pylRS. Attempts to hijack the amber stop codon for protein expression in *E. coli* cells using the tRNA^{Pyl}-PylRS pair in the absence of pyrrolysine failed, with only truncated protein being produced even at abnormally high

concentrations of canonical amino acid,^[11] unequivocally demonstrating the orthogonality of this pair to canonical amino acids.

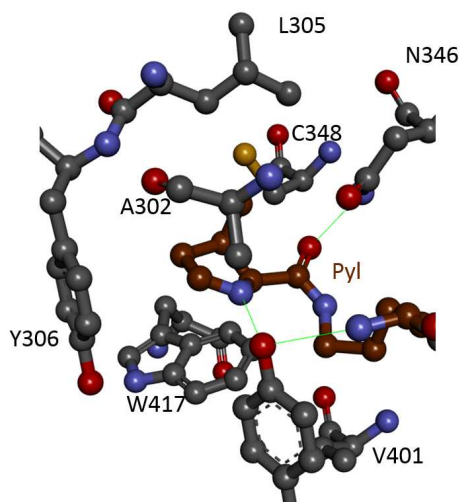
2. Hijacking the Pyrrolysine System

As specific as pylRS may be in the pool of natural amino acids, this enzyme has been shown to be highly promiscuous with unnatural amino acids. PylRS could be co-crystallised with not only pyrrolysine-AMP but also pyrrolysine analogue **7** and ATP (Scheme 2).^[10] Whilst the addition of substrate analogues is hardly uncommon in protein crystallography, it was quickly realised that the proteinogenic role of PylRS could be highly exploitable if structural analogues of pyrrolysine could be recognised by pylRS and hence incorporated into proteins,^[12] analogously to other orthogonal tRNA-RS pairs^[13] such as the *Methanocaldococcus janaschii* pair capable of introducing various tyrosine analogues into proteins.^[14] Hence simple pyrrolysine analogues **8** and **9** were successfully incorporated into β -galactosidase as a test protein in *E. coli*.^[15]



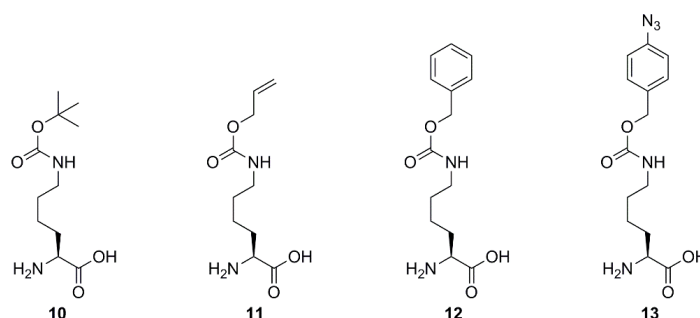
Scheme 2: Simple pyrrolysine analogues incorporated into proteins.

The limits on the design of pyrrolysine analogues chiefly arise from the specificity of PylRS: any analogue must have some affinity for the hydrophobic binding pocket. Key interactions involve Tyr384, Asn346, Trp417, Cys348 and Val401 *inter alia* in PylRS from *M. mazei* (Figure 2). The former two are important in mediating hydrogen bonding interactions between the α -amino group, the imine nitrogen and the pyrroline carbonyl, whilst the latter three are the main residues defining the hydrophobic pocket.^[16] Given this reasoning, most pyrrolysine analogues involve a different amide (or carbamate) on the ϵ -amino group, with an additional hydrogen bond acceptor replacing the imine and a moderate-size hydrophobic group replacing the pyrroline ring. This is illustrated by analogues **7-9**, with a common set of analogues being dipeptides such as **9**. Notably, the α -amino group plays only a minor role in substrate recognition, likely acting only as a hydrogen bond



donor/acceptor, to the end that pyrrolysine-like α -hydroxy acids are suitable substrates for PylRS.^[17]

Whilst the promiscuity of the wild-type PylRS variants from *Methanosarcina barkeri* and *M. mazei* (between which the active site residues are largely conserved) is already of significant utility, rational enzyme engineering has further expanded the domain of pyrrolysine analogues. Whilst Boc-Lys **10** and Aloc-Lys **11** are suitable substrates for the wild type PylRS,^[18] other lysine derivatives such as Ac-Lys **5**,^[19] Z-Lys **12**^[20] and AzZ-Lys **13**^[18] with traditional “protecting groups” or variants thereof required mutations to the active site (Scheme 3). Incorporating the small derivative **5** required shrinking the hydrophobic pocket through a C313F mutation, amongst other mutations, to the *M. barkeri* PylRS, whilst the more hydrophobic **12** and **13** necessitated the key mutations of C348V and Y384F respectively in PylRS from *M. mazei*.



Scheme 3: Protected pyrrolysine analogues.

Once a suitable pyrrolysine analogue has been selected, protein production is governed by addition of the unnatural amino acid, tRNA^{Pyl} and PylRS. This is typically achieved in a one-pot fashion by co-transforming genes for pylRS and tRNA^{Pyl} alongside the gene of interest into the cell line and inducing both genes during overexpression in the presence of the unnatural amino acid (Figure 3). A number of constructs have combined the genes for tRNA^{Pyl}, PylRS and other expression factors into single vectors for greater yields and expression control in a variety of cell lines,^[21] as well as making use of evolved pyrrolysine tRNAs for further improved yields, even with multiple reassigned stop codons.^[22] A common problem experienced is protein truncation arising from translation termination at the reassigned codon, lowering protein yields; this has been circumvented through the use of release factor one (RF1) knockout strains,^[23] increasing protein yields with minimal negative effects on the host cells.^[24]

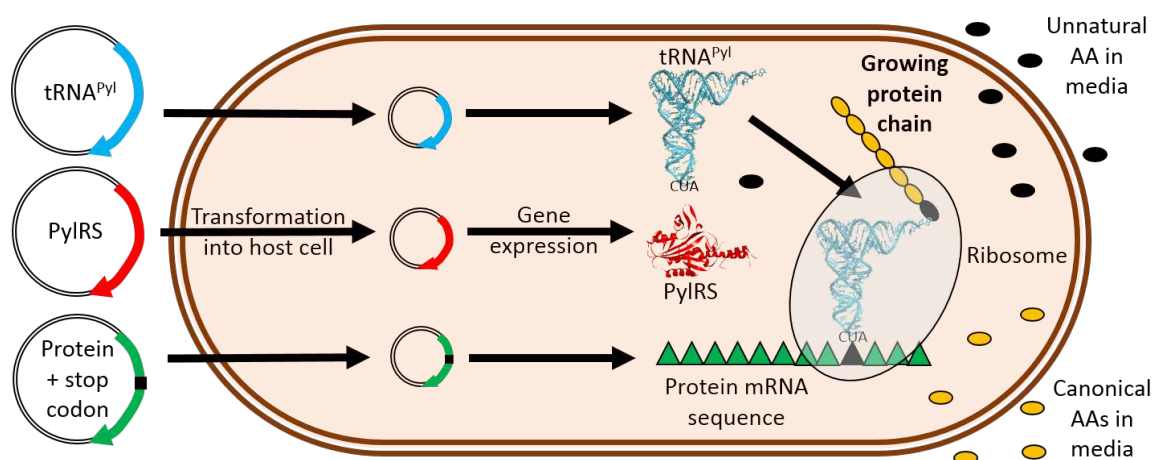
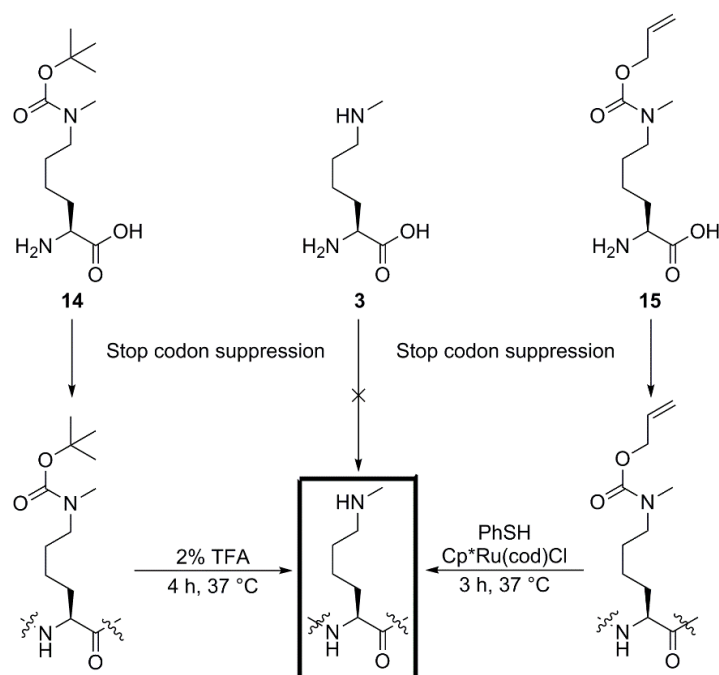


Figure 3: Amber stop codon suppression in recombinant cells using unnatural amino acids and the tRNA^{Pyl}-PylRS pair.

3. PTM Mimicry and Protecting Groups

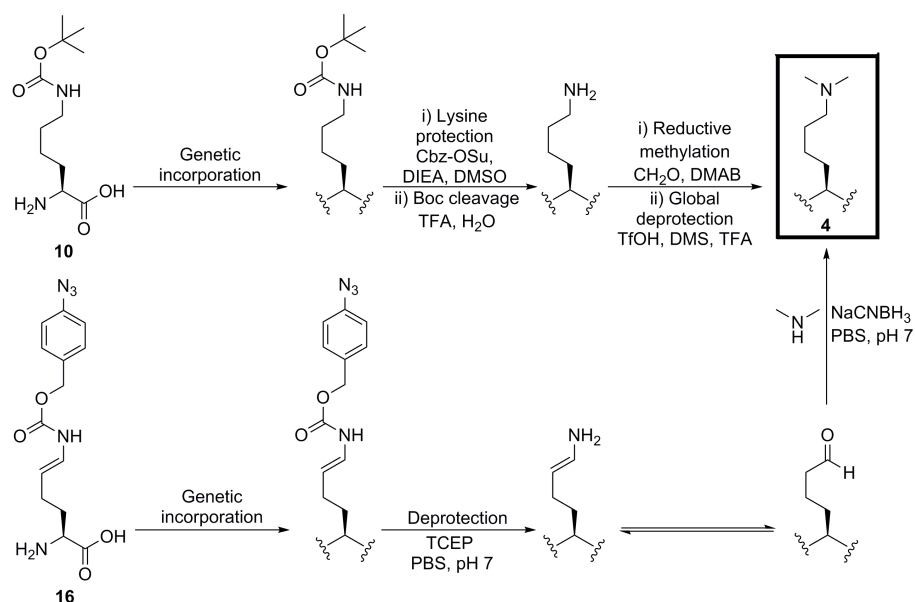
The expansion of the genetic code has been used extensively in the study of histones, in which lysine residues frequently undergo PTMs which have implications in disease states. As discussed, evolution of a suitable PylRS permitted incorporation of acetylated lysine **5** into green fluorescent protein (GFP), myoglobin^[19] and chloramphenicol acetyltransferase,^[25] with further work applied to the biological target of histones. Histone H3 was prepared with and without K56 acetylation using this method,^[26] allowing a full evaluation of acetylation in various processes such as chromatin remodelling and DNA breathing using both variants. Notably, this residue has been resistant to modification through other methods. Differentially acetylated variants of H2A and H2B subunits were also prepared. Optimisation of this method later allowed production of full-length H3 containing four acetylated lysine residues in *E. coli* cells,^[27] and further work allowed concomitant production of six differentially acetylated H3 variants directly in multiple mammalian cell lines, demonstrating the influence of histone acetylation on the expression of particular genes.^[28]

Attempts to incorporate methylated lysine **3** into H3 using directed evolution of PylRS were unable to replicate the success achieved with **5**. Ingeniously, a strategy was developed to utilise a “masked” derivative of **3** to achieve incorporation: given that Boc-Lys **10** can be recognised by the wild-type PylRS^[20] and deprotected with dilute TFA to yield the free amine, it was hypothesised that the lysine surrogate **14** should be compatible with an analogous strategy to yield methylated lysine (Scheme 4). Indeed, incorporation and post-translational deprotection yielded modified myoglobin and histone H3K9me1,^[29] with K9 methylation hence shown to be essential for the binding of H3 to heterochromatin protein 1. Further work demonstrated how this route could be used to install multiple monomethylated lysine residues in histone H3 through the use of RF1 knockout *E. coli* strain RFO.^[30] A later strategy achieved the same overall incorporation of **3** into a protein, but using an alternative protecting group. As Alloc-Lys **11** is a suitable substrate for PylRS, the N-methylated derivative **15** could serve as an analogous “caged” precursor to **3** if the Alloc group could be cleaved in a bioorthogonal manner akin to a Boc group. Histone H2B was produced containing **15** at position 27 and the post-translational deprotection succeeded with a ruthenium catalyst and thiophenol to yield H2BK27me1.^[31] The cleavage conditions for Alloc are evidently preferable as the biologically incompatible acidic conditions required for Boc cleavage cause protein denaturation and require refolding for further work, which is not always feasible in the case of large or complex proteins.



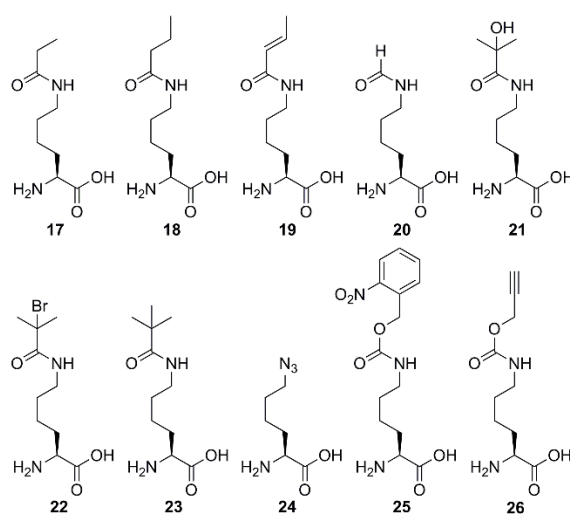
Scheme 4: Installation of methylated lysine using amber stop codon suppression, either methylated Boc-Lys **14** or Alloc-Lys **15**, and the respective deprotection conditions.

The installation of dimethylated lysine **4** in proteins has been a far more elusive goal. Efforts to evolve a pylRS recognising **4** failed, and deprotection strategies are not possible due to the limited valence of the nitrogen atom and the instability of an analogous quaternary ammonium salt. An early strategy towards dimethyllysine-containing histones made use of lysine protecting groups and reductive amination, more reminiscent of a traditional synthetic strategy (Scheme 5).^[32] The target lysine was installed as Boc-Lys **10**, and all 12 other lysine amines protected chemically. Deprotection and alkylation afforded the dimethylated lysine residue, and all other amines could be deprotected to afford the desired histone. Whilst impressive, this method is somewhat cumbersome, and hence a modified strategy was designed to install **4** in a more facile manner. The UAA **16**, an unsaturated derivative of **13**, was incorporated into GFP with an engineered *M. mazei* PylRS and the azido-Cbz group cleaved in a self-immolative reductive manner using TCEP, after which the free imine is hydrolysed to leave an aldehyde which can undergo reductive amination (Scheme 5).^[33] These two strategies reverse the location of the aldehyde and amine, but the second route offered a reduced reliance on protecting groups and milder reaction conditions.



Scheme 5: Dimethyllysine installed using noncanonical lysine derivatives, making extensive use of protection/deprotection strategies.

In addition to methylation and acetylation, histone lysine residues have been found modified by other carbonyl groups, and further work has naturally involved using stop codon suppression to produce such modified histones. Small fatty acid derivatives of lysine **17-19** (Scheme 6) were conveniently found to be substrates for wild-type PylRS, leading to various K9-acylated derivatives of H3.^[34] This work was later refined using directed evolution of PylRS, leading to the synthesis of octamers of propionylated, butyrate and crotonylated H4 proteins.^[35] Two additional lysine PTMs found in native histones could also be introduced using lysine derivatives **20**^[36] and **21**,^[37] in addition to novel analogue **22** and protected Piv-Lys **23**, with directed evolution of PylRS again proving essential. A creative strategy to incorporate acylated lysine derivatives made use of azidohomoleucine **24** in combination with on-protein traceless Staudinger ligations to afford acetylated and photocaged succinylated lysine residues, with the key property of being able to acylate independently of genetic code expansion through modifying the phosphinothioester reagent.^[38]



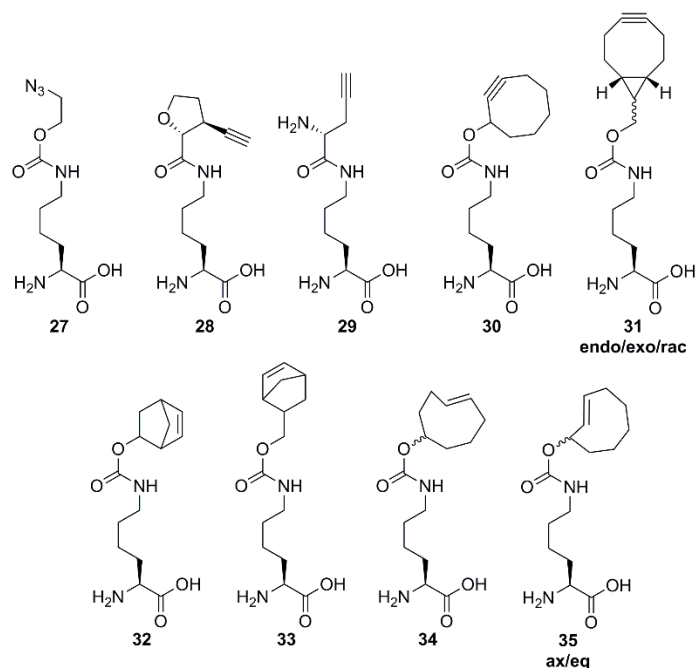
Scheme 6: Acylated lysine derivatives or precursors compatible with PylRS, either as PTM mimics or of novel use.

The ability to incorporate traditional organic chemistry protecting groups into proteins, and subsequently deprotect using bioorthogonal reagents, has seen use beyond histones. Following on from the use of Boc and Alloc, given that Z-Lys **12**^[20] and oNZ-Lys **25**^[39] could be genetically encoded, protected derivatives of methyllysine **3** could also be incorporated using the *N*-methylated derivatives of **12** and **25** which are unmasked by palladium^[40] and long-wave UV irradiation^[41] respectively. These two deprotection steps are generally considered to be bioorthogonal, although the use of UV radiation is not desirable in live cell work. Palladium deprotection was further used with Alloc-Lys **15** and Proc-Lys **26** to demonstrate site-specific caging of individual lysine residues in proteins and bioorthogonally decaging *in vitro* and *in vivo*, with the application of palladium-activated toxicity through release of a key lysine residue in a bacterial toxin.^[42]

4. Pyrrolysine Analogues in Click Chemistry and Diels-Alder Cycloadditions

Click chemistry has rapidly become a staple of bioorthogonal chemistry: its broad substrate range, biocompatibility and stability of coupling partners and fast kinetics have facilitated its widespread adoption in chemical biology. Use for protein modification was initially documented using a viral coat protein non-selectively labelled with an azido-acyl linker which subsequently underwent ligation with alkynyl fluorescein and an alkyne-containing dansyl-BSA conjugate.^[43] This first foray into copper-catalysed click chemistry (CuAAC) in protein modification was an unequivocal demonstration of the potential of the technique, albeit suffering from the drawback of a lack of site specificity. Stop codon suppression is a near-perfect solution to this problem, with the highly modifiable nature of both techniques being pleasingly synergistic. Azides and alkynes **13** and **27-31** (Scheme 7) have been encoded using pyrrolysine analogues and used in CuAAC to create bioconjugates ranging from fluorophore-tagged proteins,^[44] biotinylated myoglobin,^[45] reversibly SUMOylated proteins,^[46] and even functional antibody-drug conjugates (ADCs). The development of strain-promoted azide-alkyne cycloadditions (SPAAC) as a “copper-free click chemistry” was further reflected in additional pyrrolysine analogues developed. Strained cyclooctyne (SCO) **30**^[47] and diastereomers of bicyclononyne (BCN) **31**^[48] opened up even wider possibilities for click chemistry, including SPAAC-induced amine group unmasking,^[49] imaging of HeLa cells without the additional concern of cytotoxic copper catalysts and producing stable ADCs with anti-cancer effects in mice.^[50]

The rise of strain-promoted inverse electron-demand Diels Alder cycloaddition (SPIEDAC), in which strained dienophiles such as methylcyclopropenes^[51] or **31-35** are coupled to tetrazines, prompted a similar introduction of SPIEDAC-capable pyrrolysine analogues, with SPIEDAC offering the advantage of faster kinetics than CuAAC and SPAAC^[52] whilst also retaining fluorogenicity. An assortment of proteins containing norbornenes and *trans*-cyclooctenes (TCOs) **32-35** could be produced using engineered variants of the *M. mazei* PylRS. Applications have included fluorescently tagged proteins *in vitro*,^[53] rapid labelling of proteins *in vivo*,^[54] live cell labelling,^[55] synthesis of glycoconjugates using glycosyl azides,^[56] protein cross-linking,^[57] and SPIEDAC-initiated rapid decaging *via* a novel rearrangement mechanism.^[58] Stop codon suppression and bioorthogonal chemistry are highly complementary techniques in the field of protein modification and have facilitated innovative developments in the field of chemical biology.

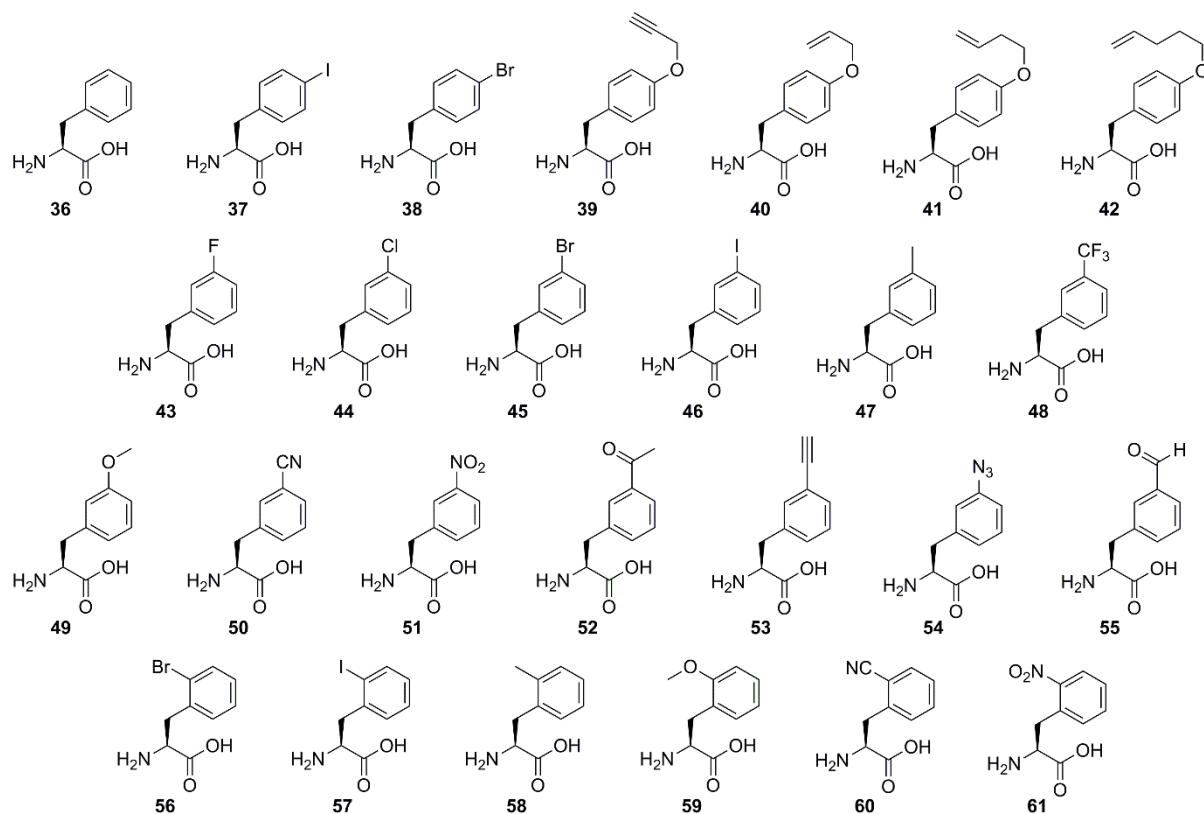


Scheme 7: Pyrrolysine analogues used in click chemistry, SPAAC and SPIEDAC.

5. Recognising Phenylalanine and Cysteine Analogues using Pyrrolysine tRNA Synthetase

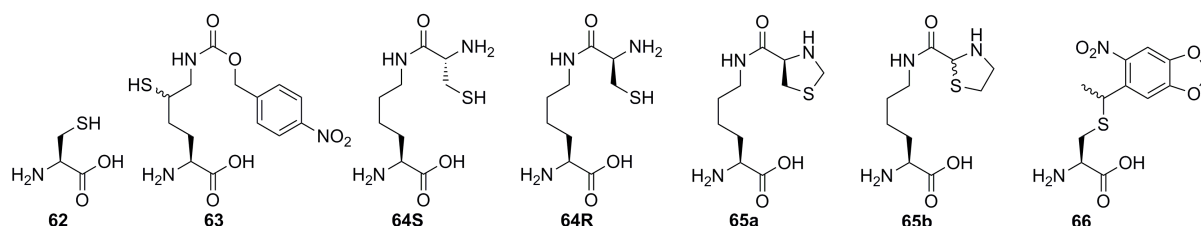
Whilst lysine derivatives are clearly highly useful targets for incorporation into proteins using stop codon suppression, greater utility could be afforded through the recognition of an even wider range of non-canonical amino acids by the *M. mazei* PylRS. Six key residues in the active site- L305, Y306, L309, N346, C348 and W417- were mutated randomly and screened against phenylalanine (Phe) **36** and *p*-iodo-Phe **37** (Scheme 8). Whilst **36** required only N346 and C348 mutations for successful incorporation, the most successful PylRS with **37** required mutations to five of the six selected active site residues. GFP containing **36** and **37** could be produced, and brominated analogue **38** was also shown to be a substrate for the **37**-recognising PylRS.^[59] The synthetic utility of this method was starkly shown through the ability to perform a site-specific Suzuki-Miyaura coupling between a Z-domain protein containing **37** and a dansylated boronic acid. The promiscuity of the engineered *M. mazei* PylRS was further examined and, following N346A and C348A mutations (PylRS^{AA}), additional Phe analogues **39-42** proved to be suitable substrates, with **39** possessing the additional function of CuAAC reactivity.^[60] The same mutant could also recognise an array of *meta*-substituted Phe derivatives **43-54**, with a range of functional groups including halides, ketones, alkynes and azides.^[61] Trifluoromethylated UAA **48** found instant use as a sensitive ¹⁹F handle, allowing monitoring of protein folding states through NMR alongside other fluorinated unnatural amino acids.^[62] A key discovery was the ability to genetically encode aldehyde **55** using PylRS^{AA},^[63] producing proteins which can undergo key bioorthogonal reactions such as oxime ligation. Other methods to install aldehydes into proteins such as using formylglycine-generating enzyme (FGE) and *N*-terminal oxidation are less flexible in where the aldehyde can be installed, as well as furnishing electronically different aldehydes.^[64] This finding circumvented a requirement to utilise the pyrrolysine biosynthetic pathway to install aldehydes in proteins, with previous work making use of the imine in pyrrolysine **1** itself, and its demethylated analogue, as aldehyde equivalents for protein bioconjugation.^[65] Further work confirmed that the promiscuity of the PylRS^{AA} mutant was such that several *ortho*-substituted phenylalanine analogues **56-61** could be incorporated into proteins, even in live cells.^[66] Orthogonality was generally retained with PylRS^{AA}, although some canonical **36** was incorporated in competition with the desired Phe analogue. It is worth noting that other Phe

analogues can be incorporated using the *M. janaschii* TyrT/TyrRS pair,^[14] including **37**^[67] *inter alia*, although the *M. mazei* and *M. barkeri* PylT/PylRS pairs exhibit an overall greater level of promiscuity.



Scheme 8: Phenylalanine and derivatives compatible with a mutated PylRS.

Whilst other tRNA/RS pairs have permitted the genetic encoding of protected cysteine derivatives,^[68] the ability to incorporate derivatives of unprotected cysteine **62** using genetic code expansion has also been a highly sought-after goal, given the broad range of reactions in which the thiol group can participate. Protected lysine thiol **63** (Scheme 9) was shown to be a suitable substrate for a mutant PyIRS from *M. barkeri*, with the deprotected protein ready for native chemical ligation (NCL) with a ubiquitin thioester.^[69] The wild type *M. barkeri* PyIRS also recognised the unprotected Cys-Lys dipeptides **64S** and **64R**, again allowing ubiquitination using NCL.^[70] **64S**, with stereochemical configuration matching that of pyrrolysine, was also shown to be a better substrate for this PyIRS variant than diastereomer **64R**. Further work demonstrated the importance

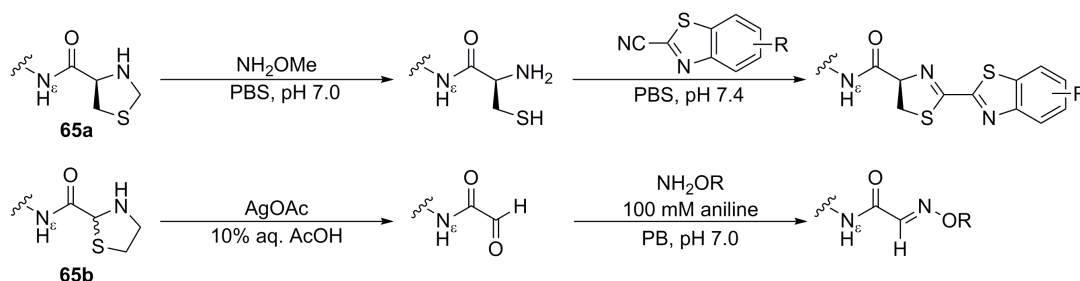


of configuration and PyIRS variant as the wild type *M. mazei* PyIRS could recognise **64S** but not **64R**, exhibiting greater sensitivity to the configuration of the Cys chiral centre.^[71] Directed evolution of the *M. barkeri* PyIRS led to improved recognition of **64S** and **64R**, as well as compatibility with the reactive amino acid **65a**, a 4-thiazolidine derivative of **64R** (to circumvent pyruvate adduct formation) which, once unmasked with methoxyamine, could undergo bioorthogonal cyanobenzothiazole condensation with fluorophores or aniline-catalysed thiazolidine formation with aldehyde-tagged ubiquitin (Scheme 10).^[72] The analogous 2-thiazolidine **65b**, used with wild-type *M. barkeri* pyIRS as a racemate, was also found to be a useful cage for reactivity. Silver-mediated cleavage exposed a glyoxyl aldehyde capable of undergoing oxime ligation, although notably the thiazolidine cleavage step required harsh protein-denaturing conditions.^[73] Photocaged cysteine **66** could be installed using a mutant *M. barkeri* PyIRS without the need for a lysine backbone,^[74] allowing the controlled photoactivation of the catalytic cysteine residue in tobacco etch virus protease.^[75]

Scheme 9: Cysteine and surrogates used in protein modification.

6. Further Applications of Pyrrolysine Analogues

In addition to the previously discussed applications of PTMs, ADCs and live cell labelling, the tRNA^{Pyl}-PyIRS pair has seen use in wider studies such as photochemistry, in ventures beyond the initial photocaging strategies used with pyrrolysine analogues **25**^[76] and **67**^[77] and the fluorescent analogue

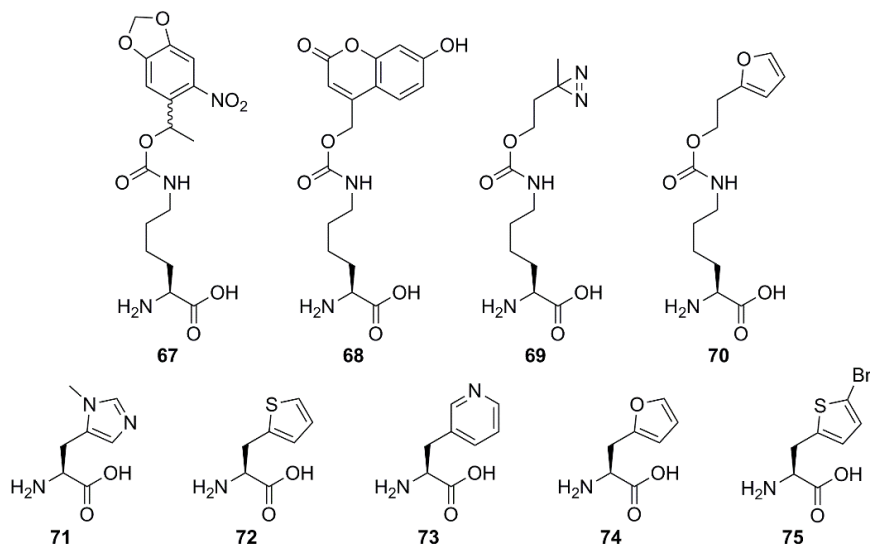


Scheme 10: Bioorthogonal of thiazolidines incorporated using genetic expansion techniques.

68 (Scheme 11).^[78] The introduction of photo-crosslinking amino acid **69** into the genetic code was used to produce dimers of mutated glutathione S-transferase in live cells, with the short (10-20 min)

reaction times permitting the use of long-wave UV irradiation in live cells with minimal interference.^[79] Crosslinking was also possible in live cells between kinase Cdk5 containing **69** and its substrate kinase Pak1, demonstrating the phosphorylation cascade occurring between these two kinases, albeit with the unwanted side-reaction of homodimerisation.^[80] Milder photochemical activation could be achieved using **70** in the presence of biocompatible photosensitiser methylene blue and oxygen, forming a reactive enedione species *in situ*.^[81] **70** was installed in an arginine-rich motif of HIV-1 trans-activator of transcription (TAT) protein, the motif responsible for binding to trans-activation response RNA (TAR), and upon irradiation with red light in the presence of methylene blue and oxygen the covalent TAT-TAR complex was formed *in vitro*.

Another field making intelligent use of amber stop codon suppression is enzyme engineering. The activity of an enzyme is naturally dependent on active site residues, and hence mutations in such residues can impede or improve enzyme performance. Stop codon suppression allows such mutations to include non-canonical amino acids to further allow *e.g.* fine-tuning of hydrogen bonding, the pK_a values of residues or metal chelation beyond that of the canon. Methylated histidine (His) **71**, along with other His analogues **72-75**, have been found to be taken up by a PylRS mutant engineered to contain an amino acid binding site much smaller than the wild type, and incorporation into blue fluorescent protein resulted in altered protein UV-visible spectral characteristics,^[82] as has been observed with canonical amino acid mutagenesis of blue fluorescent protein. This was found to be useful in the engineering of ascorbate peroxidase APX2, where introduction of **71** in the active site had the effect of greatly increasing the enzyme turnover in the oxidation of guaiacol and a slight overall increase in catalytic efficiency.^[83] The exact reason for this has not been definitively elucidated, but disruption of the hydrogen bonding in the active site and additional stabilisation of radical intermediates through hyperconjugation are suspected to explain the greater stability of the APX2 mutant. Thus, non-canonical amino acid mutagenesis has clear potential in the fields of enzyme engineering and mechanistic studies, offering far more flexibility in retuning an enzyme active site.



Scheme 11: Non-canonical amino acids used in photochemistry or as histidine analogues.

7. Reflection and Outlook

Chemical biology has benefited greatly from the pyrrolysine amber stop codon suppression system. In tandem with bioorthogonal reactions such as SPAAC, SPIEDAC and palladium-catalysed cross-coupling reactions, a wealth of new site-specific yet broadly applicable protein modification

strategies have been developed *in vitro*. Even protein modification and labelling *in vivo*, prokaryotic and eukaryotic, has become a feasible reality. The pyrrolysine tRNA_{CUA}-RS pair is able to genetically encode a far broader range of amino acid analogues than other stop codon suppression pairs. The *M. janaschii* pair, whilst generally offering higher protein yields, suffers some drawbacks, including smaller substrate scope and incompatibility with eukaryotic cell lines,^[84] although engineered eukaryotic cell lines have been used in the successful generation of antibody-drug conjugates.^[85] For the task of installing bioorthogonal handles, pyrrolysine analogues present an attractive route as the only sequence constraint is the stop codon requirement, whilst other enzymatic methods such as FGE-catalysed elimination frequently require recognition sequences. Previous chemical methods to install azides by protein acylation have been made entirely redundant due to the lack of site specificity compared to that inherent in the pyrrolysine amber stop codon suppression system. Expansion of the genetic code from 22 to over 70 amino acids has also permitted exploration of numerous biological systems impervious to previous human intervention, and will continue to be a source of inspiration and innovation.

Acknowledgements

The authors thank The University of York, and the European Commission for the award of a Marie Curie Fellowship to MAF (FP7-PEOPLE-2011-IOF-302246).

Keywords: pyrrolysine • unnatural amino acid • post translational modifications

References

- [1] B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki, M. K. Chan, *Science* **2002**, 296, 1462-1466.
- [2] C. Polycarpo, A. Ambrogelly, A. Berube, S. M. Winbush, J. A. McCloskey, P. F. Crain, J. L. Wood, D. Soll, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 12450-12454.
- [3] M. A. Gaston, L. Zhang, K. B. Green-Church, J. A. Krzycki, *Nature* **2011**, 471, 647-650.
- [4] F. Quitterer, A. List, W. Eisenreich, A. Bacher, M. Groll, *Angew. Chem. Int. Ed.* **2012**, 51, 1339-1342.
- [5] F. Quitterer, A. List, P. Beck, A. Bacher, M. Groll, *J. Mol. Biol.* **2012**, 424, 270-282.
- [6] aS. E. Cellitti, W. J. Ou, H. P. Chiu, J. Grunewald, D. H. Jones, X. S. Hao, Q. Fan, L. L. Quinn, K. Ng, A. T. Anfora, S. A. Lesley, T. Uno, A. Brock, B. H. Geierstanger, *Nat. Chem. Biol.* **2011**, 7, 528-530; bF. Quitterer, P. Beck, A. Bacher, M. Groll, *Angew. Chem. Int. Ed.* **2013**, 52, 7033-7037; cF. Quitterer, P. Beck, A. Bacher, M. Groll, *Angew. Chem. Int. Ed.* **2014**, 53, 8150-8153.
- [7] J. A. Krzycki, *Curr Opin Chem Biol* **2004**, 8, 484-491.
- [8] G. Borrel, N. Gaci, P. Peyret, P. W. O'Toole, S. Gribaldo, J. F. Brugere, *Archaea* **2014**, 2014, 374146.
- [9] L. Prat, I. U. Heinemann, H. R. Aerni, J. Rinehart, P. O'Donoghue, D. Soll, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109, 21070-21075.
- [10] J. M. Kavan, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Soll, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, 104, 11268-11273.
- [11] S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Green-Church, M. K. Chan, J. A. Krzycki, *Nature* **2004**, 431, 333-335.
- [12] B. Geierstanger, W. Ou, S. E. Cellitti, T. Uno, T. Crossgrove, H. P. Chiu, J. Grunewald, X. Hao, Google Patents, **2010**.
- [13] K. Sakamoto, A. Hayashi, A. Sakamoto, D. Kiga, H. Nakayama, A. Soma, T. Kobayashi, M. Kitabatake, K. Takio, K. Saito, M. Shirouzu, I. Hirao, S. Yokoyama, *Nucleic Acids Res.* **2002**, 30, 4692-4699.
- [14] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, 292, 498-500.
- [15] C. R. Polycarpo, S. Herring, A. Berube, J. L. Wood, D. Soll, A. Ambrogelly, *FEBS Lett.* **2006**, 580, 6695-6700.
- [16] J. K. Takimoto, N. Dellas, J. P. Noel, L. Wang, *ACS Chem. Biol.* **2011**, 6, 733-743.
- [17] T. Kobayashi, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2009**, 385, 1352-1360.
- [18] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, 15, 1187-1197.
- [19] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, 4, 232-234.
- [20] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, 371, 818-822.
- [21] aH. Xiao, A. Chatterjee, S. H. Choi, K. M. Bajjuri, S. C. Sinha, P. G. Schultz, *Angew. Chem. Int. Ed.* **2013**, 52, 14080-14083; bS. Cohen, E. Arbely, *ChemBioChem* **2016**, 17, 1008-1011; cT. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, 395, 361-374.
- [22] aA. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao, P. G. Schultz, *Biochemistry* **2013**, 52, 1828-1837; bC. Fan, H. Xiong, N. M. Reynolds, D. Soll, *Nucleic Acids Res.* **2015**, 43, e156.
- [23] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H. R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, *Science* **2013**, 342, 357-360.
- [24] D. B. Johnson, C. Wang, J. Xu, M. D. Schultz, R. J. Schmitz, J. R. Ecker, L. Wang, *ACS Chem. Biol.* **2012**, 7, 1337-1344.
- [25] T. Umehara, J. Kim, S. Lee, L. T. Guo, D. Soll, H. S. Park, *FEBS Lett.* **2012**, 586, 729-733.
- [26] H. Neumann, S. M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes, J. W. Chin, *Mol. Cell* **2009**, 36, 153-163.
- [27] I. A. Young, C. Mittal, M. A. Shogren-Knaak, *Protein Expr. Purif.* **2016**, 118, 92-97.

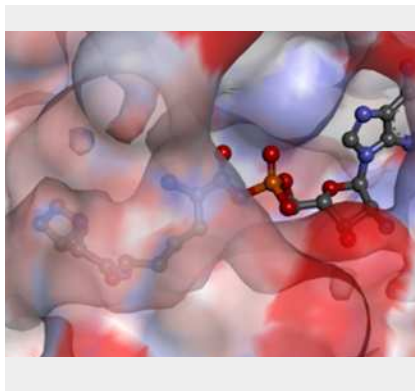
- [28] S. J. Elsasser, R. J. Ernst, O. S. Walker, J. W. Chin, *Nat. Methods* **2016**, *13*, 158-164.
- [29] D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 14194-14195.
- [30] T. Yanagisawa, M. Takahashi, T. Mukai, S. Sato, M. Wakamori, M. Shirouzu, K. Sakamoto, T. Umehara, S. Yokoyama, *ChemBioChem* **2014**, *15*, 1830-1838.
- [31] H. W. Ai, J. W. Lee, P. G. Schultz, *Chem. Commun.* **2010**, *46*, 5506-5508.
- [32] D. P. Nguyen, M. M. Garcia Alai, S. Virdee, J. W. Chin, *Chem. Biol.* **2010**, *17*, 1072-1076.
- [33] Z. A. Wang, Y. Zeng, Y. Kurra, X. Wang, J. M. Tharp, E. C. Vatansever, W. W. Hsu, S. Dai, X. Fang, W. R. Liu, *Angew. Chem. Int. Ed.* **2017**, *56*, 212-216.
- [34] M. J. Gattner, M. Vrabel, T. Carell, *Chem. Commun.* **2013**, *49*, 379-381.
- [35] B. J. Wilkins, L. E. Hahn, S. Heitmuller, H. Frauendorf, O. Valerius, G. H. Braus, H. Neumann, *ACS Chem. Biol.* **2015**, *10*, 939-944.
- [36] T. Wang, Q. Zhou, F. Li, Y. Yu, X. Yin, J. Wang, *ChemBioChem* **2015**, *16*, 1440-1442.
- [37] H. Xiao, W. Xuan, S. Shao, T. Liu, P. G. Schultz, *ACS Chem. Biol.* **2015**, *10*, 1599-1603.
- [38] Z. A. Wang, Y. Kurra, X. Wang, Y. Zeng, Y. J. Lee, V. Sharma, H. Lin, S. Y. Dai, W. R. Liu, *Angew. Chem. Int. Ed.* **2017**, *56*, 1643-1647.
- [39] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem. Int. Ed.* **2009**, *48*, 4052-4055.
- [40] Y. S. Wang, B. Wu, Z. Wang, Y. Huang, W. Wan, W. K. Russell, P. J. Pai, Y. N. Moe, D. H. Russell, W. R. Liu, *Mol. Biosyst.* **2010**, *6*, 1557-1560.
- [41] D. Groff, P. R. Chen, F. B. Peters, P. G. Schultz, *ChemBioChem* **2010**, *11*, 1066-1068.
- [42] J. Li, J. Yu, J. Zhao, J. Wang, S. Zheng, S. Lin, L. Chen, M. Yang, S. Jia, X. Zhang, P. R. Chen, *Nat. Chem.* **2014**, *6*, 352-361.
- [43] V. Hong, S. I. Presolski, C. Ma, M. G. Finn, *Angew. Chem. Int. Ed.* **2009**, *48*, 9879-9883.
- [44] aT. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 1633-1635; bX. Li, T. Fekner, M. K. Chan, *Chem. Asian J.* **2010**, *5*, 1765-1769.
- [45] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720-8721.
- [46] M. M. Lee, T. Fekner, T. H. Tang, L. Wang, A. H. Chan, P. H. Hsu, S. W. Au, M. K. Chan, *ChemBioChem* **2013**, *14*, 805-808.
- [47] T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew. Chem. Int. Ed.* **2011**, *50*, 3878-3881.
- [48] A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. Verkade, M. Wiessler, C. Schultz, J. C. van Hest, F. L. van Delft, E. A. Lemke, *ChemBioChem* **2012**, *13*, 2094-2099.
- [49] Y. Ge, X. Fan, P. R. Chen, *Chem. Sci.* **2016**, *7*, 7055-7060.
- [50] M. P. VanBrunt, K. Shanebeck, Z. Caldwell, J. Johnson, P. Thompson, T. Martin, H. Dong, G. Li, H. Xu, F. D'Hooge, L. Masterson, P. Bariola, A. Tiberghien, E. Ezeadi, D. G. Williams, J. A. Hartley, P. W. Howard, K. H. Grabstein, M. A. Bowen, M. Marelli, *Bioconjugate Chem.* **2015**, *26*, 2249-2260.
- [51] T. S. Elliott, F. M. Townsley, A. Bianco, R. J. Ernst, A. Sachdeva, S. J. Elsasser, L. Davis, K. Lang, R. Pisa, S. Greiss, K. S. Lilley, J. W. Chin, *Nat. Biotechnol.* **2014**, *32*, 465-U186.
- [52] M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, *130*, 13518-13519.
- [53] aE. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 4466-4469; bK. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox, J. W. Chin, *J. Am. Chem. Soc.* **2012**, *134*, 10317-10320; cA. Sachdeva, K. Wang, T. Elliott, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 7785-7788; dK. Kipper, E. G. Lundius, V. Curic, I. Nikic, M. Wiessler, E. A. Lemke, J. Elf, *ACS Synth. Biol.* **2017**, *6*, 233-255.
- [54] aK. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, J. W. Chin, *Nat. Chem.* **2012**, *4*, 298-304; bT. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew. Chem. Int. Ed.* **2012**, *51*, 4166-4170; cJ. E. Hoffmann, T. Plass, I. Nikic, I. V.

- Aramburu, C. Koehler, H. Gillandt, E. A. Lemke, C. Schultz, *Chemistry* **2015**, *21*, 12266-12270; dY. Yang, S. Lin, W. Lin, P. R. Chen, *ChemBioChem* **2014**, *15*, 1738-1743.
- [55] T. Peng, H. C. Hang, *J. Am. Chem. Soc.* **2016**, *138*, 14423-14433.
- [56] T. Machida, K. Lang, L. Xue, J. W. Chin, N. Winssinger, *Bioconjugate Chem.* **2015**, *26*, 802-806.
- [57] A. Rutkowska, T. Plass, J. E. Hoffmann, D. A. Yushchenko, S. Feng, C. Schultz, *ChemBioChem* **2014**, *15*, 1765-1768.
- [58] aJ. Li, S. Jia, P. R. Chen, *Nat. Chem. Biol.* **2014**, *10*, 1003-1005; bX. Fan, Y. Ge, F. Lin, Y. Yang, G. Zhang, W. S. Ngai, Z. Lin, S. Zheng, J. Wang, J. Zhao, J. Li, P. R. Chen, *Angew. Chem. Int. Ed.* **2016**, *55*, 14046-14050.
- [59] Y. S. Wang, W. K. Russell, Z. Wang, W. Wan, L. E. Dodd, P. J. Pai, D. H. Russell, W. R. Liu, *Mol. Biosyst.* **2011**, *7*, 714-717.
- [60] Y. S. Wang, X. Fang, A. L. Wallace, B. Wu, W. R. Liu, *J. Am. Chem. Soc.* **2012**, *134*, 2950-2953.
- [61] aY. S. Wang, X. Fang, H. Y. Chen, B. Wu, Z. U. Wang, C. Hilty, W. R. Liu, *ACS Chem. Biol.* **2013**, *8*, 405-415; bA. Tuley, Y. S. Wang, X. Fang, Y. Kurra, Y. H. Rezenom, W. R. Liu, *Chem. Commun.* **2014**, *50*, 2673-2675.
- [62] aJ. C. Jackson, J. T. Hammill, R. A. Mehl, *J. Am. Chem. Soc.* **2007**, *129*, 1160-1166; bS. E. Cellitti, D. H. Jones, L. Lagpacan, X. Hao, Q. Zhang, H. Hu, S. M. Brittain, A. Brinker, J. Caldwell, B. Bursulaya, G. Spraggon, A. Brock, Y. Ryu, T. Uno, P. G. Schultz, B. H. Geierstanger, *J. Am. Chem. Soc.* **2008**, *130*, 9268-9281.
- [63] A. Tuley, Y. J. Lee, B. Wu, Z. U. Wang, W. R. Liu, *Chem. Commun.* **2014**, *50*, 7424-7426.
- [64] R. J. Spears, M. A. Fascione, *Org. Biomol. Chem.* **2016**, *14*, 7622-7638.
- [65] W. Ou, T. Uno, H. P. Chiu, J. Grunewald, S. E. Cellitti, T. Crossgrove, X. Hao, Q. Fan, L. L. Quinn, P. Patterson, L. Okach, D. H. Jones, S. A. Lesley, A. Brock, B. H. Geierstanger, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10437-10442.
- [66] J. M. Tharp, Y. S. Wang, Y. J. Lee, Y. Yang, W. R. Liu, *ACS Chem. Biol.* **2014**, *9*, 884-890.
- [67] J. M. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon, P. G. Schultz, *Nat. Biotechnol.* **2004**, *22*, 1297-1301.
- [68] aN. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 14306-14307; bD. H. Jones, S. E. Cellitti, X. Hao, Q. Zhang, M. Jahnz, D. Summerer, P. G. Schultz, T. Uno, B. H. Geierstanger, *J. Biomol. NMR* **2010**, *46*, 89-100.
- [69] S. Virdee, P. B. Kapadnis, T. Elliott, K. Lang, J. Madrzak, D. P. Nguyen, L. Riechmann, J. W. Chin, *J. Am. Chem. Soc.* **2011**, *133*, 10708-10711.
- [70] X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 9184-9187.
- [71] D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir, J. W. Chin, *J. Am. Chem. Soc.* **2011**, *133*, 11418-11421.
- [72] X. Bi, K. K. Pasunooti, A. H. Tareq, J. Takyi-Williams, C. F. Liu, *Org. Biomol. Chem.* **2016**, *14*, 5282-5285.
- [73] X. Bi, K. K. Pasunooti, J. Lescar, C. F. Liu, *Bioconjugate Chem.* **2017**, *28*, 325-329.
- [74] R. Uprety, J. Luo, J. Liu, Y. Naro, S. Samanta, A. Deiters, *ChemBioChem* **2014**, *15*, 1793-1799.
- [75] D. P. Nguyen, M. Mahesh, S. J. Elsasser, S. M. Hancock, C. Uttamapinant, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 2240-2243.
- [76] A. Gautier, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2011**, *133*, 2124-2127.
- [77] J. Luo, E. Arbely, J. Zhang, C. Chou, R. Uprety, J. W. Chin, A. Deiters, *Chem. Commun.* **2016**, *52*, 8529-8532.
- [78] J. Luo, R. Uprety, Y. Naro, C. Chou, D. P. Nguyen, J. W. Chin, A. Deiters, *J. Am. Chem. Soc.* **2014**, *136*, 15551-15558.
- [79] C. J. Chou, R. Uprety, L. Davis, J. W. Chin, A. Deiters, *Chem. Sci.* **2011**, *2*, 480-483.
- [80] H. W. Ai, W. Shen, A. Sagi, P. R. Chen, P. G. Schultz, *ChemBioChem* **2011**, *12*, 1854-1857.
- [81] M. J. Schmidt, D. Summerer, *Angew. Chem. Int. Ed.* **2013**, *52*, 4690-4693.
- [82] H. Xiao, F. B. Peters, P. Y. Yang, S. Reed, J. R. Chittuluru, P. G. Schultz, *ACS Chem. Biol.* **2014**, *9*, 1092-1096.

- [83] A. P. Green, T. Hayashi, P. R. Mittl, D. Hilvert, *J. Am. Chem. Soc.* **2016**, *138*, 11344-11352.
- [84] M. Tsunoda, Y. Kusakabe, N. Tanaka, S. Ohno, M. Nakamura, T. Senda, T. Moriguchi, N. Asai, M. Sekine, T. Yokogawa, K. Nishikawa, K. T. Nakamura, *Nucleic Acids Res.* **2007**, *35*, 4289-4300.
- [85] aJ. Y. Axup, K. M. Bajjuri, M. Ritland, B. M. Hutchins, C. H. Kim, S. A. Kazane, R. Halder, J. S. Forsyth, A. F. Santidrian, K. Stafin, Y. Lu, H. Tran, A. J. Seller, S. L. Biroc, A. Szydlík, J. K. Pinkstaff, F. Tian, S. C. Sinha, B. Felding-Habermann, V. V. Smider, P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 16101-16106; bF. Tian, Y. Lu, A. Manibusan, A. Sellers, H. Tran, Y. Sun, T. Phuong, R. Barnett, B. Hehli, F. Song, M. J. DeGuzman, S. Ensari, J. K. Pinkstaff, L. M. Sullivan, S. L. Biroc, H. Cho, P. G. Schultz, J. DiJoseph, M. Dougher, D. Ma, R. Dushin, M. Leal, L. Tchistiakova, E. Feyfant, H. P. Gerber, P. Sapra, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 1766-1771.

MINIREVIEW

The pyrrolysine tRNA synthetase-tRNA pair is capable of genetically encoding a plethora of non-canonical amino acids through amber stop codon suppression. This Minireview summarises the development of pyrrolysine amber stop codon suppression, presents some of the considerations required to utilise this technique to its greatest potential, and showcases the creative ways in which this technique has led to a better understanding of biological systems.



Robin Brabham, Martin A. Fascione*

Page No. – Page No.

Pyrrolysine Amber Stop Codon Suppression: Development and Applications