DOI: 10.1002/ejoc.201000204

Pyrrolysine Analogs for Translational Incorporation into Proteins

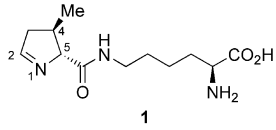
Tomasz Fekner,[a] Xin Li,[a] and Michael K. Chan*[a]

Keywords: Pyrrolysine / Amino acids / Amber codon / Protein modifications / Genetic encoding

The discovery that pyrrolysine is translationally incorporated into methylamine methyltransferases from certain methanogens in response to an in-frame amber codon prompted detailed studies of the biochemical machinery associated with its genetic encoding. Over the past few years, significant progress has also been made in identifying a set of structural and electronic features that enable a compound to act as an effective pyrrolysine mimic. With such empirical knowledge in hand, it is becoming increasingly feasible to rationally design and synthesize new analogs with useful chemical, physical, or biochemical properties and then site-specifically incorporate them into proteins. The present microreview highlights recent key developments in this rapidly expanding area of research.

Introduction

We and the Krzycki group first detected and identified pyrrolysine (Pyl, **1**, Figure 1), the 22nd genetically-encoded amino acid, from the crystal structure of the *Methanosarcina barkeri* monomethylamine methyltransferase (MtmB, Figure 2) and mass-spectrometry studies.[1–3] The name was coined to reflect its chemical identity as N^6-[(4*R*,5*R*)-4-methyl-1-*pyrro*line-5-carbonyl]-L-*lysine*.



1

Figure 1. Pyrrolysine (**1**).

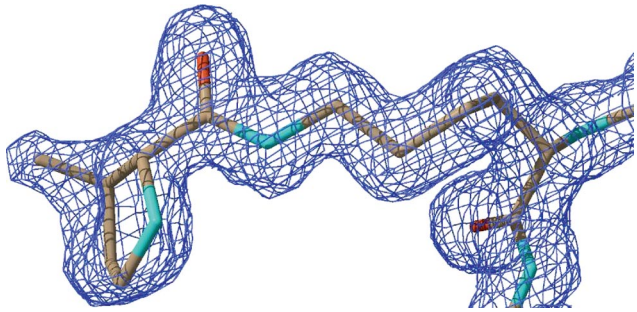


Figure 2. Electron density around the pyrrolysine residue in the X-ray crystal structure of *M. barkeri* MtmB.[1]

[a] Department of Chemistry, Department of Biochemistry, and Biophysics Graduate Program, The Ohio State University, 484 W 12th Ave., Columbus, OH 43210, USA
Fax: +1-614-292 6773
E-mail: chan@chemistry.ohio-state.edu

Subsequent biochemical studies unequivocally demonstrated that pyrrolysine is a new genetically-encoded amino acid whose incorporation is mediated by the otherwise nonsense UAG (amber) codon.[4,5] Specifically, **1** is directly recognized and charged by a pyrrolysyl-tRNA synthetase (PylRS) onto its cognate amber suppressor $tRNA^{Pyl}$ that, in turn, is used by the ribosome to incorporate **1** during protein synthesis.[6,7] An alternative pathway involving charging $tRNA^{Pyl}$ with lysine followed by its modification to pyrrolysine was shown not to be operational.[6] It was also established that synthetic **1** could be incorporated into recombinant MtmB in *Escherichia coli* by co-expression with *M. barkeri* PylRS and $tRNA^{Pyl}$ using the pETDuet vector (Novagen), thus demonstrating that the pyrrolysine incorporation system could be transferred to prokaryotic and potentially other biological systems.[6]

Certain unique features of the PylRS–$tRNA^{Pyl}$ pair should enable its widespread use for site-specific modification of proteins using structural analogs of pyrrolysine. There is a growing body of evidence that PylRS–$tRNA^{Pyl}$ pairs from *Methanosarcinaceae* are not only fully functional in *E. coli* but also orthogonal[8] to the endogenous aminoacyl-tRNA synthetase (aaRS)–tRNA sets.[6,9] The specificity and exclusivity of PylRS towards $tRNA^{Pyl}$ is also retained in mammalian cells.[9,10] Recently, the molecular basis of these properties has been elucidated from the X-ray single-crystal structure of the *Desulfitobacterium hafniense* PylRS–$tRNA^{Pyl}$ complex.[11] From the results of these studies, it was concluded that the highly specialized and intricate interaction surface between the two units, which is markedly distinct from those observed for other known aaRS–tRNA complexes, is the key feature responsible for the PylRS–$tRNA^{Pyl}$ orthogonality. Moreover, crystallographic analyses of the *M. mazei*[12,13] and *D. hafniense*[14] PylRS have revealed a remarkably large binding pocket for **1** that, in turn, suggests broad specificity for lysine analogs bearing a wide variety of N-6 side-chains.

Throughout the review, we refer to various compounds as pyrrolysine analogs not so much to emphasize their structural similarity with **1**, but rather to indicate that they are competent substrates for native or genetically-evolved PylRS–tRNA^{Pyl} pairs.

Exploratory Search for Pyrrolysine Analogs

For a chemical compound to serve as a competent analog of a genetically-encoded amino acid, such as pyrrolysine, it has to be a suitable surrogate in three critical events. Firstly, it has to be recognized by an appropriate aaRS which then catalyzes its reaction with adenosine 5′-triphosphate (ATP) to form an amino acid adenylate that primes the acyl group for the next key process – the aminoacylation of the cognate tRNA. Finally, the tRNA-charged amino acid has to be incorporated into the protein during translation. For each stage, appropriate assays have been developed that enable biochemists to evaluate the effectiveness of the potential analog. Of course, only detailed analysis of the final product makes it possible to ascertain the level of efficiency and fidelity of translational incorporation.

Initial efforts towards the synthesis of pyrrolysine and its analogs were born out of necessity stemming from the lack of readily available sources of the parent amino acid that severely hampered relevant biochemical studies. Although our own synthetic work in conjunction with the Krzycki group[2] provided us with enough material to demonstrate that **1** is directly charged onto tRNA^{Pyl} by PylRS and then inserted into a protein,[6] the approach was not practical for large scale preparations. Moreover, pyrrolysine turned out to be relatively unstable, which precluded its storage for extended periods of time. After being unable, in contrast to others,[12,15] to reproduce our synthetic protocol for the parent compound **1**, Ambrogelly et al.[16] reported on the synthesis of the pyrrolysine isomer **2** (Figure 3) but were also dissatisfied with the shortcomings of their approach. To alleviate these problems, a series of lysine derivatives **2**–**8** (Figure 3) was screened for their ability to serve as pyrrolysine surrogates using an aminoacylation assay based on acid–urea gel electrophoresis and Northern blotting. Some successful analogs identified in this fashion were then subjected to in vivo readthrough tests.

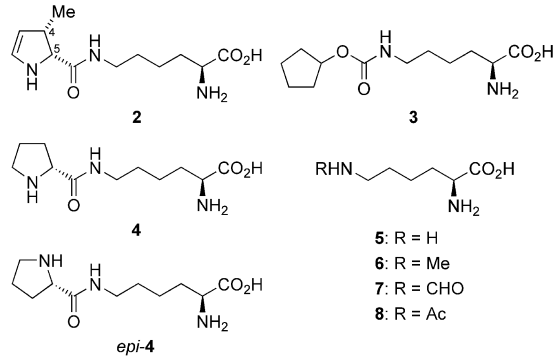
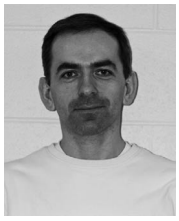
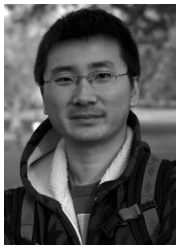


Figure 3. Potential pyrrolysine analogs **2**–**8** tested as substrates for *M. barkeri* PylRS–tRNA^{Pyl} by Ambrogelly et al.[16] **2** was screened as an equimolar mixture of (4*S**,5*R**)-diastereomers.

Of all the compounds screened, pyrrolysine isomer **2**, carbamate **3**, and proline derivative **4** were shown to be activated and subsequently ligated onto tRNA^{Pyl} by PylRS in vitro. The lack of any activation for *epi*-**4** in contrast to its diastereomer **4** led the authors to postulate the crucial



Tomasz Fekner obtained his M.S. degree in Chemistry from the Warsaw University of Technology and his D.Phil. degree in Organic Chemistry from the University of Oxford for his work on the synthesis of unnaturally configured penicillins and cephalosporins under the direction of Prof. Jack Baldwin. After post-doctoral stays at the University of Sheffield with Prof. Alan Spivey (homochiral DMAPs), The Ohio State University with Prof. Michael Chan (cyclic benzimidazole ligands), and University College Dublin with Prof. Pat Guiry (quinazoline–oxazoline ligands), he returned to Ohio State as a senior research associate. His primary scientific interests center on the application of organic synthesis to the study of biochemical problems.



Xin Li received his B.S. and M.S. in Biochemistry from Beijing Normal University, where he studied the regulation of protein phosphatase 2B with Prof. Qun Wei. Since 2005, he has been a Ph.D. student in Biophysics at The Ohio State University, working with Prof. Michael Chan. His current interests include the structural biology of carbon dioxide-associated enzymes and membrane proteins, and the applications of the emerging pyrrolysine-based non-natural amino acids into protein engineering and functional studies. He is a recipient of an American Heart Association Great Rivers Affiliate Predoctoral Fellowship and an OSU Presidential Fellowship.

Michael Chan received his B.S. in Chemistry from Harvey Mudd College, and his Ph.D. in Inorganic Chemistry from the University of California, Berkeley, working with William H. Armstrong on functional models of the Photosystem II Oxygen Evolving Complex. He then moved to protein crystallography where as an NIH postdoctoral fellow with Douglas Rees at Caltech, he focused on the structural biology of molybdenum and tungsten containing enzymes. Current interests include efforts to elucidate the fundamental enzymes associated with carbon dioxide and methane biochemistry, and the development of novel biotechnologies based on the pyrrolysine incorporation system. Michael has been a faculty member at The Ohio State University since 1995. He is a recipient of an NSF Career Award, an Alfred P. Sloan Foundation Fellowship, and most recently, an Ohio State Distinguished Scholar Award.

importance of the side-chain stereocenter for recognition by PylRS. They also suggested that the actual presence of the ring nitrogen atom is not particularly beneficial because **3** and **4** are, more or less, equally competent substrates. Taking this argument further, it was subsequently put forward that the improper positioning of the ring heteroatom, like in *epi*-**4**, is actually detrimental to the level of PylRS-catalysed activation. However, as carbamate **3** possesses an additional oxygen atom that could potentially contribute to PylRS binding, the analogy between it and **4** might not be so straightforward. Although lysine (**5**) was not expected to be activated by PylRS, it is noteworthy that its analogs **6**–**8** bearing simple acyclic substituents at N-6 were similarly ineffective, with the possible exception of the acetyl derivative **8** that showed very low, "slightly above background",[16] levels of PylRS-catalysed activation. Both **3** and **4** were also demonstrated to be competent amber suppressors during protein synthesis in *E. coli*.

In related exploratory studies, the Chan and Krzycki groups subjected analogs **9**–**13** (Figure 4) to the [^{32}P]PPi–ATP exchange (PPi = pyrophosphate), acid–urea gel electrophoresis, and Northern blotting assays as well as in vivo protein readthrough experiments.[17] These studies ultimately identified the THF derivative **10**, bearing an oxygen atom at the position corresponding to that of the imine nitrogen atom in **1**, to be an exceptionally effective pyrrolysine mimic. The catalytic efficiency of *M. barkeri* PylRS in the presence of **10** is ~10 and ~100 times higher than in the presence of Ambrogelly's analog **3** and the cyclopentyl derivative **9**, respectively. These results indicate that a properly placed heteroatom at the N-6 acyl substituent of lysine has a profound and beneficial effect on the level of activation. Moreover, the results for **3** and **9** demonstrate that the lack of a heteroatom in the five-membered ring can be somewhat offset by the presence of the carbamate oxygen atom. In agreement with the observations made by Ambrogelly et al.,[16] we also noted that misplacement of the heteroatom within the five-membered ring, as in the diastereomeric THF derivatives **11** and *epi*-**11**, renders these compounds unsuitable as pyrrolysine surrogates. This seems to indicate that the active site of PylRS is highly sensitive to the position of a heteroatom within the cyclic acyl substituents of its potential lysine-based substrates. It is also notable that the two analogs **12** and **13**, bearing aromatic substituents at lysine N-6, failed to be activated by PylRS. In addition to the activation studies, we were also able to demonstrate that **10** can be ligated to tRNA^{Pyl} and subsequently incorporated into a protein in vivo with the level of efficiency comparable to that observed for pyrrolysine itself.

In the previous studies, the structure of the wild-type PylRS was kept constant and it was the potential pyrrolysine analogs whose structure was modified until a suitable analog, acceptable by the synthetase, was found. Because N-6 acetylation of lysine residues is a vital post-translational modification of proteins regulating a wide variety of cellular processes,[18] Chin et al. attempted to genetically-encode N^6-acetyllysine (**8**) using *M. barkeri* PylRS–tRNA^{Pyl} pair.[19] However, as previously mentioned, Ambrogelly et al. found that **8** was very poorly activated by the native form of this synthetase.[16] To circumvent this obstacle, Chin et al. first confirmed by readthrough studies with the pyrrolysine analog **3**, that the PylRS–tRNA^{Pyl} pair was highly efficient and orthogonal in *E. coli*. Next, a library of ~10^8 PylRS mutants was generated, in which six active-site residues were randomized, based on the crystallographic studies of *M. mazei* PylRS bound to a substrate.[13] By applying the directed evolution principles[8] involving three cycles of selection, two most suitable mutants were then identified. The authors postulated that the hydrophobic cavity used to bind the pyrroline ring of **1** in the native synthetase was rearranged in the mutants to accommodate the much smaller acetyl group of **8**, with the difference in volume between the two substituents being compensated by an increased volume of the amino acid residues in the evolved synthetases. To demonstrate their efficiency and orthogonality, the newly-engineered enzymes were subsequently used to site-specifically incorporate **8** into a selected set of proteins (vide infra).

Yokoyama et al. utilized a similar evolution strategy to facilitate the incorporation of a range of suitable pyrrolysine analogs (Figure 5).[15] While carbamates **14** and **15** aminoacylate tRNA^{Pyl} in the presence of *M. mazei* PylRS as efficiently as pyrrolysine does, much bulkier analogs **16**–**18** require significantly higher concentrations in order to attain a similar level of ligaton. Moreover, after screening a series of other lysine analogs lacking the N-6 carbonyl group and finding no suitable substrate among them, it was postulated that the functional group was a necessary, albeit not sufficient, structural feature to be present in a successful pyrrolysine analog. Subsequent detailed crystallographic studies led the authors to conclude that PylRS most readily recog-

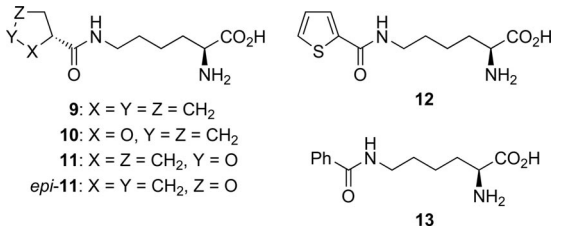


Figure 4. Potential pyrrolysine analogs **9**–**13** tested as substrates for *M. barkeri* PylRS–tRNA^{Pyl} by Chan, Krzycki et al.[17]

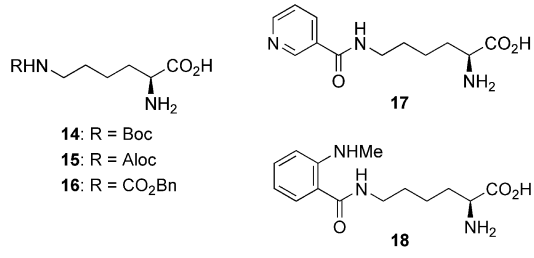
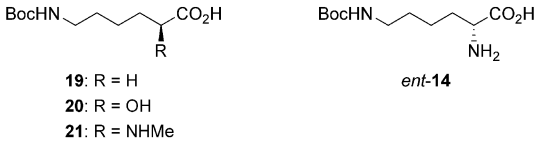


Figure 5. Potential pyrrolysine analogs **14**–**18** screened by Yokoyama et al.[15] Boc = *tert*-butoxycarbonyl, Aloc = allyloxycarbonyl, Bn = benzyl.

nizes substrates that possess some additional common structural characteristics, i.e., a properly sized and shaped hydrophobic substituent (not necessarily cyclic) at N-6 of lysine that can be accommodated in the hydrophobic pocket of the enzyme, a hydrogen-bond acceptor (N or O) in the vicinity of the N-6 carbonyl group, and a side-chain spacer of a suitable length. Accompanying mutational studies led to the identification of much more effective variants of PylRS that allowed for remarkably efficient in vivo incorporation of all the three carbamate-type analogs **14**–**16** selected for the initial screening.

While the length of the spacer, i.e., the distance between the N-6 hydrophobic substituent and the α-carbon bearing the CO_2H and NH_2 groups, in lysine-based pyrrolysine analogs is presumably important for PylRS recognition, the main-chain NH_2 group itself is a very insignificant recognition element.[12] Building on this observation, Yokoyama et al. identified a series of pyrrolysine analogs **19**–**21** and *ent*-**14** (Figure 6) that were shown to be efficiently ligated to $tRNA^{Pyl}$ by *M. mazei* PylRS.[20] Moreover, the hydroxy acid **20** was incorporated into a protein in vivo, thus enabling the site-specific introduction of an ester bond.

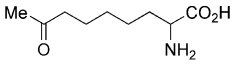


19: R = H
20: R = OH
21: R = NHMe

ent-**14**

Figure 6. Pyrrolysine analogs **19**–**21** and *ent*-**14** with a modified main-chain backbone as developed by Yokoyama et al.[20]

Guided by the results of crystallographic studies of Steitz, Söll et al. on *M. mazei* PylRS bound to either pyrrolysine (**1**) or its analog **3**,[13] Liu et al.[21] concluded that the carbonyl-bearing side-chain NH group in pyrrolysine mimics has little influence on their binding potential. To test this hypothesis, they prepared racemic ketone **22** (Figure 7) and successfully incorporated it into the green fluorescent protein (GFP) using the evolved PylRS–$tRNA^{Pyl}$ pair developed by Chin et al. for the translational incorporation of N^6-acetyllysine (**8**).[19]



22

Figure 7. Ketone-based pyrrolysine analog **22** reported by Liu et al.[21]

Practical Applications of Pyrrolysine Analogs

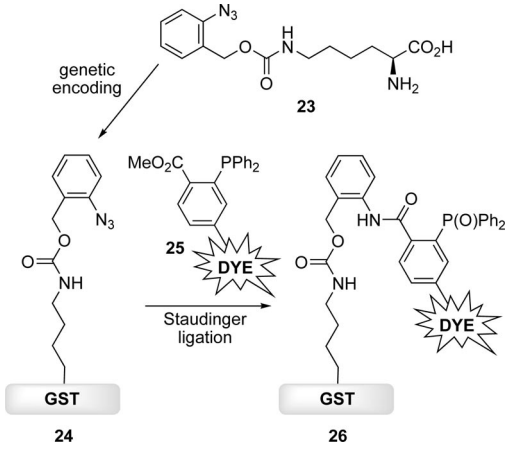
Considering the complexity and diversity of living organisms, it is remarkable that all of them utilize the same set of 20 proteinogenic amino acids (not including the rarely utilized selenocysteine and pyrrolysine) to biosynthesize a vast array of proteins they need. However, the ability to genetically encode additional amino acids makes it possible to enhance the desired properties of proteins or, by introducing useful functional groups in a site-specific manner, enable a detailed study of their function and structure. From a practical standpoint, it also increases the utility of the widely used *E. coli* heterogeneous expression system to proteins bearing post-translational modifications. In principle, the most convenient way to expand the genetic code is to engage any of the three termination codons. This approach makes less daunting the task of constructing a unique set comprising a non-natural amino acid, a novel tRNA, and its cognate aaRS. Of course, each of these components must comply with the stringent orthogonality rules that ensure that the novel amino acid is genetically incorporated into a protein with high efficiency and fidelity.[8]

Prior to work on the use of the PylRS–$tRNA^{Pyl}$ system to genetically encode pyrrolysine analogs, the Schultz group applied the elegant method of directed evolution to the construction of new orthogonal aaRS–tRNA pairs for translational incorporation of non-natural amino acids.[8] This method, which involves generating libraries of tRNA and aaRS mutants and subjecting them to cycles of negative and positive selection in order to improve their orthogonality in respect to the endogenous aaRS–tRNA sets, has resulted in nearly 70 non-natural amino acids being site-specifically incorporated into proteins. Because the Schultz group has been focused primarily on the evolution of *Methanocaldococcus jannaschii* TyrRS, most of the non-natural amino acids they have incorporated bear aromatic substituents. The PylRS–$tRNA^{Pyl}$ pair, with its origins in charging proteins with pyrrolysine (**1**), a non-aromatic amino acid, could be therefore complementary to the originally developed system. Moreover, because the PylRS–$tRNA^{Pyl}$ pair likely appeared some 3 billion years ago and has persisted in organisms that use methylamines as their energy source ever since,[13] it has evolved unique structural features in both partners that make it naturally orthogonal to other aaRS–tRNA pairs.[6,22,11] Over the past few years, a range of pyrrolysine analogs bearing useful functional groups or post-translational modifications have been site-specifically incorporated into proteins using PylRS–$tRNA^{Pyl}$ systems. In the remainder of this review, we will highlight the recent developments in this rapidly growing area of research.

A Pyrrolysine Analog for Staudinger Ligation

Following the development of the PylRS(Y306A·Y384F) mutant for site-specific incorporation of carbamate **16** into glutathione S-transferase (GST), Yokoyama et al.[15] analyzed a docking model of this substrate with a PylRS catalytic fragment. They concluded that if an additional *ortho* substituent at the aromatic ring of **16** was introduced then the resulting analog could still be accommodated into the large active-site pocket of the synthetase. To test their hypothesis and provide an example of a practically useful pyrrolysine analog, they successfully incorporated aromatic azide **23** into the GST protein **24** (Scheme 1). Subsequent non-traceless Staudinger ligation[23] between the side-chain azido group of **24** and the fluorescein-containing triaryl-

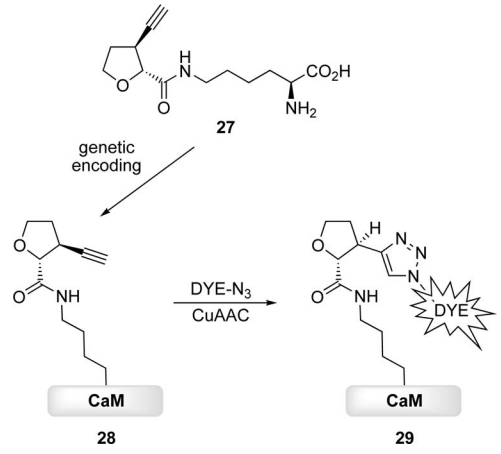
phosphane **25** gave the fluorescent protein **26**. Following an analogous treatment, the GST protein incorporating **16** in lieu of **23** was not fluorescent proving that the azide-containing residue was the only reactive site within protein **24** that was reactive towards dye **25**.



Scheme 1. Protein Staudinger ligation by Yokoyama et al.[15]

Pyrrolysine Analogs for Protein Click Chemistry

In our search for functionalized pyrrolysine derivatives amenable to further site-specific labeling, we decided to develop a suitable analog bearing a terminal alkyne group that could be used as a reactive handle for copper(I)-catalyzed azide–alkyne cycloaddition reaction (CuAAC or click reaction).[24] Because our THF-based derivative **10** proved to be an excellent substrate for the PylRS–tRNA^{Pyl} system,[17] we synthesised its closely related congener **27** (Scheme 2).[25] The new analog was subsequently demonstrated to be a competent substrate by translational incorporation into calmodulin (CaM), a 17-kDa protein that plays a pivotal role in calcium signaling in eukaryotes,[26] to produce the terminal alkyne-containing protein **28** [i.e., CaM mutant T34(**27**)]. It could then be tagged with an azide-bearing coumarin dye via a CuAAC process to give the fluorescent product **29**. We also prepared the CaM T34(**27**)·T100C double mutant and, thanks to the chemical orthogonality between the two reactive sites, were able to use them to install two distinct dyes. This allowed us to study conformational change of CaM in the presence of Ca^{II} and the M13 peptide by FRET measurements.



Scheme 2. Protein click chemistry protocol developed by Chan et al.[25]

The design of alkyne **27** was guided by our working hypothesis that a successful pyrrolysine mimic needed to be both structurally and electronically similar to the parent compound **1**. As a result, our initial synthetic targets were relatively complex which, as expected, required multi-step syntheses to prepare them. For instance, the synthesis of alkyne **27** involved 16 steps starting from ascorbic acid. It inevitably rendered this compound too expensive for widespread use. Inspired by the work of Yokoyama et al. on the incorporation of H-Lys(Boc)-OH (**14**) (vide supra),[15] we decided to look for acyclic analogs of alkyne **27**. Our second-generation studies culminated in the straightforward preparation of **30** (Figure 8),[27] which turned out to be a more effective pyrrolysine mimic than both **10** and **27**. As expected based on the previously discussed results,[16,17] diastereomer *epi*-**30** was demonstrated not to be a competent pyrrolysine mimic. Analogs bearing other functional groups in place of the side-chain NH_2 group (OMe, OH, H) were also found to be ineffective.

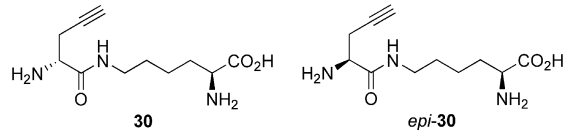


Figure 8. Second-generation pyrrolysine analog **30** for protein click chemistry and its ineffective diastereomer *epi*-**30** developed by Chan et al.[27]

Notably, while our second-generation studies were ongoing, Chin et al. reported on two simple pyrrolysine analogs **31** and **32** (Figure 9) for protein labelling via CuAAC.[28] The two analogs were successfully site-specifically incorporated into proteins. Additionally, myoglobin-His_6 with **31** incorporated at position 4 was reacted via a CuAAC process with either a biotin- or fluorophore-containing azide to provide suitably site-specifically tagged proteins.

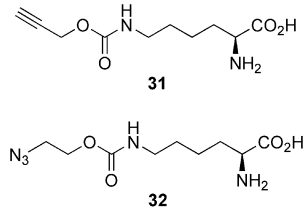


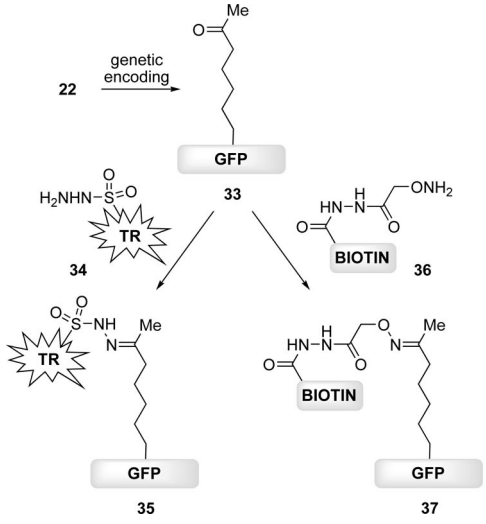
Figure 9. Pyrrolysine analogs **31** and **32** for protein click chemistry developed by Chin et al.[28]

In their study on the synthesis of glycosylated proteins, Carell et al. incorporated up to three molecules of alkyne **31** into the yellow fluorescent protein (YFP) and subsequently

functionalized it with azide-containing sugars by a click reaction.[29] As anticipated, the requirement to readthrough three UAG codons resulted in a very low yield of the desired modified protein. This notwithstanding, the demonstration that it is possible to incorporate site-specifically multiple non-natural amino acid residues into a protein is remarkable.

Protein Tagging with Hydrazides and Alkoxyamines

After successfully incorporating pyrrolysine analog **22** into GFP, Liu et al.[21] decided to take advantage of the known reactivity of the keto group towards highly nucleophilic species like hydrazides and alkoxyamines to tag the protein **33** (Scheme 3) with different probes in a site-specific manner. When treated with Texas Red (TR) hydrazide (**34**), protein **33** was converted into its fluorescent Schiff-base derivative **35**. In a control experiment, GFP incorporating N^6-acetyllysine (**8**) instead of **22** remained non-fluorescent after an analogous treatment, which proved that the reaction between protein **33** and dye **34** was limited to the site containing the keto group. Similarly, when reacted under physiological conditions with a fivefold excess of biotin-derived alkoxyamine **36**, protein **33** was converted into its biotin-tagged analog **37** with high efficiency and specificity.



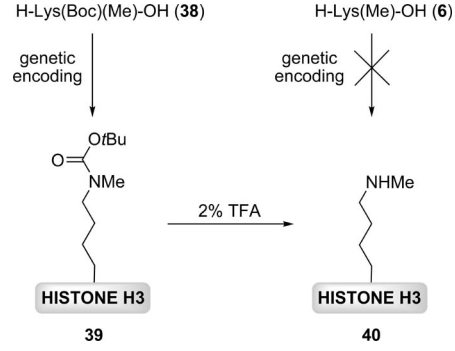
Scheme 3. Protein tagging by Liu et al.[21]

The pyrrolysine analog **22** can be used to introduce a wide range of biochemical and biophysical probes into proteins. Moreover, the authors propose that, because **22** is a non-hydrolysable structural mimic of **8**, it could potentially aid the study of the regulatory role of post-translational (reversible) acetylation in a variety of proteins, including histones.

Pyrrolysine Analogs for Synthesis of Methylated and Acetylated Proteins

The nucleophilic side-chain amino group makes protein-incorporated lysine residues favorite targets for numerous post-translational modifications, including methylation, acetylation, biotinylation, ubiquitination, sumoylation,[30] as well as the recently discovered propionylation and butyrylation.[31] As pyrrolysine is also a modified lysine, Chin et al. decided to use the PylRS–$tRNA^{Pyl}$ system to co-translationally incorporate suitably modified lysines in a site-specific manner.[32] This would facilitate the study of the effect of lysine post-translational modifications on the properties of proteins. As mentioned earlier, a mutated version of *M. barkeri* PylRS that accepts N^6-acetyllysine (**8**) was developed first, and these studies culminated in the synthesis of site-specifically acetylated myoglobin.[19] Their focus then shifted to histones due to the important role that reversible post-translational N-6 modifications of their lysine residues play in nucleosome function and gene expression.[32] By subjecting the previously developed mutant *M. barkeri* PylRS used to incorporate N^6-acetyllysine (**8**)[19] to yet another round of directed evolution, Chin et al. identified a new mutant with even higher efficiency for **8**. It was subsequently used to prepare a series of histones with different site-specifically acetylated lysines in the H2A, H2B, or H3 subunits. Notably, the authors successfully assembled the histones to form a nucleosome with H3 specifically acetylated at Lys56, a residue that was previously inaccessible for modification by other methods. Studies of this novel assembly seem to suggest that rather than affecting the compaction of nucleosomes in chromatin fibers, acetylation of Lys56 in H3 regulates nucleosome function by allowing for increased DNA breathing.

Chin et al. then focused their attention on site-specific methylation of lysine residues in histones. A deeper understanding of cellular phenomena controlled by this post-translational modification has been lacking due to the absence of straightforward methods for generating proteins altered in this manner.[33] Following their successful incorporation of **8**,[19] they decided to find a way to incorporate H-Lys(Me)-OH (**6**).[34] Unfortunately, as already established by Ambrogelly et al.[16] and Yokoyama et al.,[15] **6** is not accepted by wild-type PylRSs as a competent substrate. Moreover, a directed evolution approach, so effective during the studies on the incorporation of **8**,[19] failed to provide any viable synthetase mutant. As H-Lys(Boc)-OH (**14**)



Scheme 4. A two-step incorporation of H-Lys(Me)-OH (**6**) by Chin et al.[34]

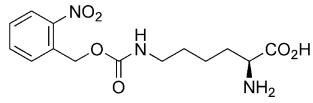
is known to be an excellent substrate for PylRS,[15] it was decided to check if H-Lys(Boc)(Me)-OH (**38**) was also acceptable (Scheme 4). Gratifyingly, **38** proved to be accommodated by *M. barkeri* PylRS and, as a result, it was possible to site-specifically insert it into full-length histone H3 in good yield and with high fidelity. The modified histone **39** was then deprotected with diluted trifluoroacetic acid (TFA) to produce the fully functional site-specifically *N*-methylated histone H3 (**40**).

Incorporation of Novel Amino Acids in Mammalian Cells

Although the directed evolution approach based on *M. jannaschii* TyrRS has been successfully implemented in *E. coli* to generate orthogonal aaRS–tRNA pairs for a range of non-natural amino acids,[8,35] these pairs cannot be typically transferred and used in mammalian cells due to the loss of orthogonality relative to the endogenous set of aaRS–tRNA pairs.[36] Recently, several groups have began to explore the application of the PylRS–tRNA^{Pyl} system to protein expression in mammalian cells.[9,10]

Yokoyama et al. were first to report a solution to this challenging problem.[10] Following a series of exploratory studies, they identified the human U6 promoter as being optimal for the expression of *M. mazei* tRNA^{Pyl} in Chinese hamster ovary (CHO) cells. In particular, it was demonstrated that when tRNA^{Pyl} under control of this promoter is co-expressed with PylRS in the presence of H-Lys(Boc)-OH (**14**), effective amber suppression can be achieved for specific reporter proteins. The efficiency of the full-length reporter-protein expression could be further improved by transfection of the host cell with a vector containing nine tandem-linked repeats of the U6–tRNA^{Pyl} gene cluster. When the PylRS–tRNA^{Pyl} pair was combined with the EBNA1-oriP transient expression system, overexpression of proteins incorporating **14** in human embryonic kidney (HEK) 293 cells was achieved. Suitable PylRS mutants evolved in the *E. coli* selection system were then transplanted into this mammalian expression system, so that both the bulky carbamate **16** and the small N^6-acetyllysine (**8**) could be site-specifically inserted into the target protein in vivo. The translational incorporation of **8** at a defined site of the target protein in mammalian cell provides a useful tool for the study of the role of lysine acetylation under physiological conditions.

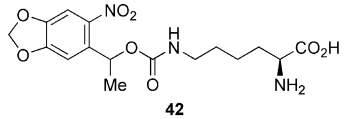
In separate studies on encoding non-natural amino acids in mammalian cells, Schultz, Geierstanger et al.[9] selected the photolabile derivative **41** (Figure 10), which is structurally closely related to Yokoyama's pyrrolysine analogs **16** and **23**, as their model substrate.[15] Carbamate **41** expands the set of genetically-encoded photocaged amino acids (Ser, Cys, and Tyr)[37] and could be used as a tool to aid the study of a wide range of biochemical processes pertinent to lysine residues such as ubiquitination, methylation, and acetylation. Using Ambrogelly's carbamate **3**,[16] they first reaffirmed findings of others that *M. mazei* PylRS–tRNA^{Pyl} pair is orthogonal both in *E. coli* and mammalian cells.[6,7,10] Next, in order to alter the specificity of the enzyme in favour of **41**, its five residues that, on the basis of the crystallographic studies of **1** bound to *M. mazei* PylRS,[13] surround the pyrroline ring, were randomized to generate a large ($>10^7$) library of mutants. Subsequent directed evolution performed in *E. coli* identified a mutant PylRS–tRNA^{Pyl} pair specific for **41** which was later proved to retain the orthogonality of its wild-type counterpart in both *E. coli* and, more importantly, mammalian cells.



41

Figure 10. Photolabile pyrrolysine analog (caged lysine) **41** developed by Schultz, Geierstanger et al.[9]

Inspired by the results of the previously discussed studies,[9,10] Chin, Deiters et al.[38] introduced another caged lysine derivative **42** (Figure 11). Its main advantages over **41** are a lower reactivity of its de-caging by-product (a benzophenone vs. a benzaldehyde for **41**) towards proteins and very rapid deprotection using nonphototoxic light (365 nm), although **41** can also be de-caged with near-visible light upon more prolonged treatment.[9] By following a well-established directed evolution protocol[9,19] involving a set of ~10^8 *M. barkeri* PylRS mutants, an orthogonal PylRS–tRNA^{Pyl} pair for translational incorporation of **42** in HEK 293 cells was developed. The practical applicability of **42** was illustrated by the study of the classical bipartite nuclear localization signal/sequence (NLS) of nucleoplasmin. First, a caged-NLS–GFP fusion, in which a key lysine residue participating in binding of importin α was replaced with **42**, was constructed. Next, it was demonstrated by fluorescence imaging that the one-point mutation (caging) blocks the critical NLS function leading to the partial transfer of the caged fusion into the cytoplasm. Finally, it was possible to study the kinetics of NLS nuclear import by real-time fluorescence microscopy following a de-caging event brought about by a short (1 s) pulse of 365-nm light. In a similar fashion, a caged tumor suppressor p53 was used to study a bipartite NLS-mediated nuclear import of its de-caged form from the cytosol.



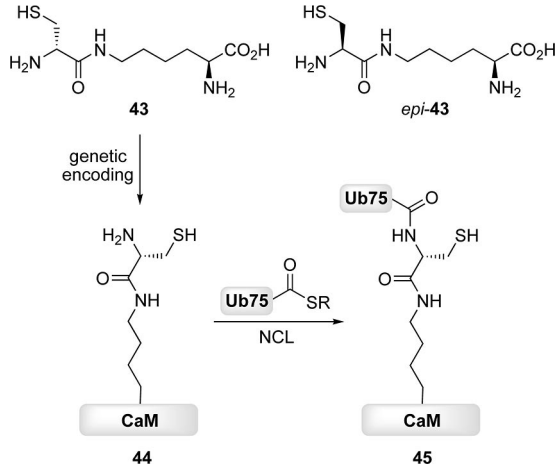
42

Figure 11. Highly photolabile pyrrolysine analog (caged lysine) **42** developed by Chin, Deiters et al.[38]

All in all, the use of these PylRS–tRNA^{Pyl} systems has the potential to provide a general strategy for expanding the set of genetically-encoded amino acids available to both prokaryotic and eukaryotic organisms.[39]

A Pyrrolysine Analog for Native Chemical Ligation

Native chemical ligation (NCL) is a well-established method of protein synthesis by joining large peptidic fragments.[40] It relies on the chemoselective reaction between a coupling partner possessing a C-terminal thioester with a peptide segment bearing an N-terminal cysteine residue. As ubiquitination, a special post-translational modification of lysine residues with a small protein (ubiquitin, Ub), plays an important role in a plethora of cellular processes,[41] we set about to develop a pyrrolysine analog that could be used to arm a peptide with the standard NCL handle. We therefore screened two diastereomeric dipeptides, **43** and *epi*-**43** (Scheme 5), for their ability to read through the UAG codon.[42] By using a modified mCherry florescence assay, we were able to determine that while the two dipeptides were competent substrates for *M. mazei* PylRS–tRNA^{Pyl}, **43** was superior to *epi*-**43** in terms of readthrough efficiency. It was therefore selected for our further studies.



Scheme 5. Incorporation of dipeptide **43** followed by NCL according to Chan et al.[42] R = $CH_2CH_2SO_3Na$.

After selecting CaM as our model system, we were then able to incorporate **43** into it and generate the recombinant protein **43**-CaM (**44**) as a suitable partner for NCL. Subsequent coupling with a thioester derived from a truncated ubiquitin (Ub75, i.e., containing residues 1–75) gave the ubiquitinated calmodulin **45**. As our product exhibited the same functional properties as enzymatically prepared CaM-Ub, it appears that the replacement of the Gly76 residue of Ub with D-Cys is inconsequential.

Conclusions

The discovery of pyrrolysine expanded the set of Nature's basic building blocks of life known to us. Realization that incorporation of the new amino acid is directed by a termination codon spurred vigorous research into the detailed mechanism of its translation. It was subsequently revealed that the PylRS–tRNA^{Pyl} system, apart from being orthogonal in a range of organisms, is relatively promiscuous in accepting a wide variety of substrates. These two features alone assure a bright future for pyrrolysine, which otherwise could have been known as the least favourite and most underutilized genetically-encoded amino acid.

As outlined in this microreview, following a period of exploratory studies in which random modified lysines were screened, some general rules have emerged which enable us to design suitable pyrrolysine analogs with a high level of confidence. For wild-type PylRS–tRNA^{Pyl} pairs, three main classes of pyrrolysine analogs **46a**–**c** (Figure 12) have been identified so far. The first and the most widely utilized class, **46a**, comprises N-6 carbamate-type lysine derivatives. There is a clear upper limit to the size of the R^1 group, as *t*Bu (**14**) is tolerable while Bn (**16**) is not. So far, the lower size limit for the R^1 group, if it at all exists, has not been established. The second class, **46b**, contains N-6 amide-type lysine derivatives with the acyl substituent bearing a heterocyclic (Y = O, NH), non-aromatic, five-membered ring. These compounds are the most direct steric and electronic analogs of pyrrolysine. The position of the heteroatom within the ring and the configuration of the stereogenic center adjacent to the amide carbonyl are of utmost importance. In general, it is not clear if the heterocyclic ring can be decorated with extra substituents, although small groups at C-4 (**27**) are acceptable. The third class, **46c**, includes N-6 amide-type lysine derivatives bearing acyclic acyl substituents. They require a properly positioned NH_2 group for effective pyrrolysine mimicry.

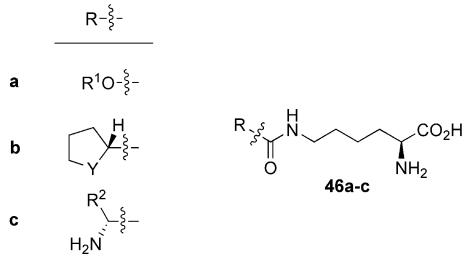


Figure 12. Three classes of pyrrolysine analogs (**46a**–**c**) for wild-type PylRS–tRNA^{Pyl}.

Mutated versions of PylRS that accept substrates that do not belong to any of the above categories (e.g., **8**, *ent*-**14**, **19**, and **22**) have also been developed. It seems that the carbonyl group separated from the carboxylate group by six atoms is the only common feature among all the successful analogs. Whether the length of the spacer between the two functional groups can be changed without any detriment to the readthrough efficiency has so far not been established.

Alongside the studies aimed at probing the specificity profile of the PylRS–tRNA^{Pyl} system, practical applications of pyrrolysine mimics are also beginning to materialize.

Acknowledgments

Financial support by the National Institute of General Medical Sciences, the National Institutes of Health (grant GM061796, supplemented by American Recovery and Reinvestment Act funds) to M. K. C. and by the American Heart Association via its Great Riv-

ers Affiliate Predoctoral Fellowship (0815449D) to X. L. is gratefully acknowledged.

[1] B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki, M. K. Chan, *Science* **2002**, *296*, 1462–1466.
[2] B. Hao, G. Zhao, P. T. Kang, J. A. Soares, T. K. Ferguson, J. Gallucci, J. A. Krzycki, M. K. Chan, *Chem. Biol.* **2004**, *11*, 1317–1324.
[3] J. A. Soares, L. Zhang, R. L. Pitsch, N. M. Kleinholz, R. B. Jones, J. J. Wolff, J. Amster, K. B. Green-Church, J. A. Krzycki, *J. Biol. Chem.* **2005**, *280*, 36962–36969.
[4] C. M. James, T. K. Ferguson, J. F. Leykam, J. A. Krzycki, *J. Biol. Chem.* **2001**, *276*, 34252–34258.
[5] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459–1462.
[6] 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.
[7] C. Polycarpo, A. Ambrogelly, A. Bérubé, S. M. Winbush, J. A. McCloskey, P. F. Crain, J. L. Wood, D. Söll, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12450–12454.
[8] J. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775–782.
[9] 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.
[10] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822.
[11] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Söll, O. Nureki, *Nature* **2009**, *457*, 1163–1168.
[12] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2008**, *378*, 634–652.
[13] J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Söll, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11268–11273.
[14] M. M. Lee, R. Jiang, R. Jain, R. C. Larue, J. Krzycki, M. K. Chan, *Biochem. Biophys. Res. Commun.* **2008**, *374*, 470–474.
[15] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187–1197.
[16] C. R. Polycarpo, S. Herring, A. Bérubé, J. L. Wood, D. Söll, A. Ambrogelly, *FEBS Lett.* **2006**, *580*, 6695–6700.
[17] W.-T. Li, A. Mahapatra, D. G. Longstaff, J. Bechtel, G. Zhao, P. T. Kang, M. K. Chan, J. A. Krzycki, *J. Mol. Biol.* **2009**, *385*, 1156–1164.
[18] S. C. Kim, R. Sprung, Y. Chen, Y. Xu, H. Ball, J. Pei, T. Cheng, Y. Kho, H. Xiao, L. Xiao, N. V. Grishin, M. White, X.-J. Yang, Y. Zhao, *Mol. Cell* **2006**, *23*, 607–618.
[19] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234.
[20] T. Kobayashi, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2009**, *385*, 1352–1360.
[21] Y. Huang, W. Wan, W. K. Russell, P.-J. Pai, Z. Wang, D. H. Russell, W. Liu, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 878–880.
[22] S. Herring, A. Ambrogelly, S. Gundllapalli, P. O'Donoghue, C. R. Polycarpo, D. Söll, *FEBS Lett.* **2007**, *581*, 3197–3203.
[23] a) E. Saxon, C. R. Bertozzi, *Science* **2000**, *287*, 2007–2010; b) K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19–24.
[24] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.
[25] T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 1633–1635.
[26] a) H. P. Jennissen, M. Laub, *Biol. Chem. Hoppe-Seyler* **1988**, *369*, 1325–1330; b) M. Majetschak, M. Laub, C. Klocke, J. A. Steppuhn, H. P. Jennissen, *Eur. J. Biochem.* **1998**, *255*, 492–500; c) M. Majetschak, M. Laub, H. E. Meyer, H. P. Jennissen, *Eur. J. Biochem.* **1998**, *255*, 482–491; d) M. Laub, J. A. Steppuhn, M. Bluggel, D. Immler, H. E. Meyer, H. P. Jennissen, *Eur. J. Biochem.* **1998**, *255*, 422–431; e) H. P. Jennissen, G. Botzet, M. Majetschak, M. Laub, R. Ziegenhagen, A. Demiroglou, *FEBS Lett.* **1992**, *296*, 51–56; f) M. Laub, H. P. Jennissen, *Biochim. Biophys. Acta* **1997**, *1357*, 173–191.
[27] X. Li, T. Fekner, M. K. Chan, *Chem. Asian J.*, DOI:10.1002/asia.201000205.
[28] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721.
[29] E. Kaya, K. Gutsmield, M. Vrabel, M. Müller, P. Thumbs, T. Carell, *ChemBioChem* **2009**, *10*, 2858–2861.
[30] C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto Jr., *Angew. Chem. Int. Ed.* **2005**, *44*, 7342–7372.
[31] Y. Chen, R. Sprung, Y. Tang, H. Ball, B. Sangras, S. C. Kim, J. R. Falck, J. Peng, W. Gu, Y. Zhao, *Mol. Cell. Proteomics* **2007**, *6*, 812–819.
[32] 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.
[33] a) C. Köhler, C. B. R. Villar, *Trends Cell Biol.* **2008**, *18*, 236–243; b) M. Spivakov, A. G. Fisher, *Nat. Rev. Genet.* **2007**, *8*, 263–271; c) C. Martin, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838–849.
[34] D. P. Nguyen, M. M. G. Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 14194–14195.
[35] L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.
[36] a) W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239–244; b) D. Kiga, K. Sakamoto, K. Kodama, T. Kigawa, T. Matsuda, T. Yabuki, M. Shirouzu, Y. Harada, H. Nakayama, K. Takio, Y. Hasegawa, Y. Endo, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9715–9720; c) 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 Acid. Res.* **2002**, *30*, 4692–4699.
[37] a) A. Deiters, D. Groff, Y. Ryu, J. Xie, P. G. Schultz, *Angew. Chem. Int. Ed.* **2006**, *45*, 2728–2731; b) E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, *Nat. Chem. Biol.* **2007**, *3*, 769–772; c) N. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 14306–14307.
[38] A. Gautier, D. P. Nguyen, H. Lusic, W. An, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2010**, *132*, 4086–4088.
[39] D. Schwarzer, *ChemBioChem* **2009**, *10*, 1602–1604.
[40] a) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776–779; b) S. B. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338–351.
[41] a) C. M. Pickart, M. J. Eddins, *Biochim. Biophys. Acta* **2004**, *1695*, 55–72; b) C. M. Pickart, D. Fushman, *Curr. Opin. Chem. Biol.* **2004**, *8*, 610–616; c) R. L. Welchman, C. Gordon, R. J. Mayer, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 599–609; d) D. Mukhopadhyay, H. Riezman, *Science* **2007**, *315*, 201–205; e) W. Li, Y. Ye, *Cell. Mol. Life Sci.* **2008**, *65*, 2397–2406.
[42] X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 9184–9187.

Received: February 16, 2010
Published Online: June 22, 2010

