

## *N*<sup>6</sup>-(2-(*R*)-Propargylglycyl)lysine as a Clickable Pyrrolysine Mimic

Xin Li, Tomasz Fekner, and Michael K. Chan\*<sup>[a]</sup>

The site-specific tagging of proteins with organic fluorophores<sup>[1]</sup> has been proven to be a powerful method in structural and functional studies of a wide range of biological systems.<sup>[2–4]</sup> The most challenging aspect of this approach is the regiospecific incorporation of a suitable fluorophore on or nearby a rationally selected amino acid residue within a protein chain. The demonstration that pyrrolysine (**1**, Figure 1),

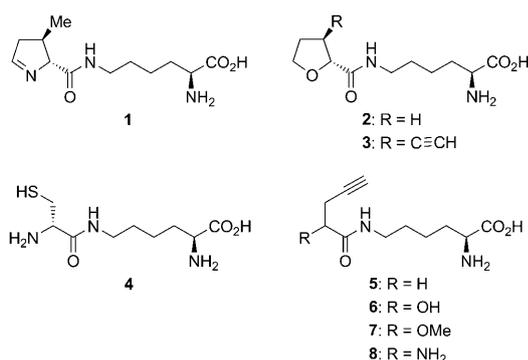


Figure 1. Pyrrolysine (**1**) and analogs **2–8**.

the 22nd genetically-encoded amino acid, can be incorporated into recombinant proteins in response to the UAG codon,<sup>[5,6]</sup> prompted us to search for similarly incorporable analogs of **1** harboring reactive functionalities suitable for anchoring small organic fluorophores. We previously reported the synthesis of the THF-containing lysine derivative **2** that reads through the UAG codon,<sup>[7]</sup> as does its close struc-

tural analog **3**.<sup>[8]</sup> The latter compound, thanks to the presence of the terminal alkyne functionality, enables site-specific post-translational modification of the resulting protein with azide-based fluorophores via the Cu<sup>I</sup>-catalyzed azide-alkyne cycloaddition reaction (CuAAC). Our recent work has gone beyond the field of protein click chemistry resulting in the synthesis and application of the cysteine derivative **4** for protein ubiquitination via native chemical ligation.<sup>[9]</sup>

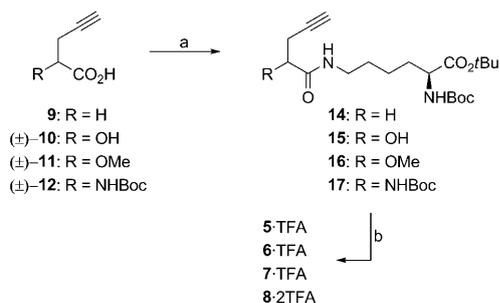
The main obstacle in the widespread use of **3** stems from its inaccessibility and high cost associated with the length of its synthesis (16 steps from ascorbic acid).<sup>[8]</sup> The guiding principle in the design of both **2** and **3** was the desire to maintain the steric and electronic similarity between **1** and its analogs, as it was believed to be a prerequisite for a successful pyrrolysine mimic. An earlier report by Polycarpo et al.<sup>[10]</sup> on simple pyrrolysine analogs had seemed to reinforce this notion. In particular, lysine derivatives bearing an acyclic acyl substituent at N-6 had been reported not to read through the UAG codon. However, a more recent work by Yanagisawa et al.<sup>[11]</sup> revealed that some acyclic analogs can also serve as viable pyrrolysine substitutes. Moreover, it was also reported that even *N*<sup>6</sup>-acetyllysine, when supplied at high concentration<sup>[12]</sup> or in the presence of a PylS (pyrrolysyl-tRNA synthetase) mutant,<sup>[13]</sup> can be charged onto PylT (tRNA<sup>Pyl</sup>) or even incorporated into recombinant proteins.

In light of these more recent findings, we decided to carry out systematic studies to establish which structural and electronic features associated with **1–3** are necessary to be present in an effective clickable pyrrolysine mimic. Our hope was to find an inexpensive and readily available alternative to the THF analog **3**. Notably, while our studies were ongoing, Nguyen et al.,<sup>[14]</sup> building on the structural studies by Polycarpo et al.,<sup>[10]</sup> have reported on a simple carbamate-based pyrrolysine analog for protein click chemistry. In our work, we focused on identifying alternative structural motifs beneficial for effective mimicking of **1**. With this in mind, we carried out readthrough experiments with a series of analogs **5–8**, that is, derivatives of lysine acylated at N-6 with various pent-4-ynoic acids **9–12** (Scheme 1). The requisite compounds **5–8** were prepared by coupling (BOP/NMM)

[a] X. Li,<sup>+</sup> Dr. T. Fekner,<sup>+</sup> Prof. M. K. Chan  
The Ohio State Biophysics Program, Departments of Chemistry and Biochemistry, The Ohio State University  
484 W 12th Ave., Columbus, OH 43210 (USA)  
Fax: (+1) 614-292-6773  
E-mail: chan@chemistry.ohio-state.edu  
Homepage: <http://www.chemistry.ohio-state.edu/~chan/>

[<sup>+</sup>] These authors contributed equally.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/asia.201000205>.



Scheme 1. Synthesis of pyrrolysine analogs **5–8**: a) Boc-Lys-OrBu (**13**), BOP, NMM, CH<sub>2</sub>Cl<sub>2</sub>, room temperature, 14–24 h, 91–95 %; b) TFA, room temperature, 1 h, ~100%. BOP: (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate, NMM: *N*-methylmorpholine, TFA: trifluoroacetic acid.

Boc-Lys-OrBu (**13**)<sup>[8]</sup> with the corresponding acids **9–12** to give amides **14–17** in excellent yield (>90%) in each case. Their subsequent treatment with neat TFA provided the desired pyrrolysine analogs **5–8** as TFA salts. Because the racemic acids **10–12** were used as substrates, the preliminary screening tests were performed with analogs **6–8** as ~1:1 diastereomeric mixtures.

The site-specific incorporation efficiency of analogs **5–8** in response to the UAG codon was then measured with our modified fluorescence protein assay. The use of the green fluorescent protein (GFP) as a reporter to monitor stop-codon readthrough of non-natural amino acids has been previously described.<sup>[15,16]</sup> We further improved on this methodology by substituting GFP with a red fluorescent protein, mCherry, to reduce background fluorescence associated with the growth medium. Moreover, being able to cultivate the host cells in 24-well plates facilitates handling of multiple samples and minimizes random disparities between them. Finally, measuring the fluorescence intensity of the cell culture directly in the plates avoids labor-intensive electrophoresis and western blotting.

In the current study, the Lys55 codon of mCherry was site-specifically mutated to UAG and the resulting mCherry coding sequence was inserted into pPylST (PylS-PylT plasmid) bearing PylS and PylT genes from *Methanosarcina mazei*.<sup>[8]</sup> *Escherichia coli* strain BL21(DE3) transformed with this plasmid was then grown in the Terrific Broth medium (pH 7.4) supplied with a suitable pyrrolysine analog (2 mM) to be examined. The results of these experiments revealed significant differences between compounds **5–8** in terms of readthrough efficiency (Figure 2). The most simplified analog **5**, bearing neither a ring nor an additional heteroatom at the lysine N-6 acyl substituent, failed to promote any detectable mCherry production, as did its hydroxyl congener **6**. A small level of UAG readthrough was observed for the methoxy analog **7**, although the compound is ~30 times less efficient than the original clickable analog **3**. In contrast, the NH<sub>2</sub>-bearing analog **8** turned out to be a much better substrate, with its readthrough efficiency exceeding that of **3** by ~70%. Similar results were observed when the

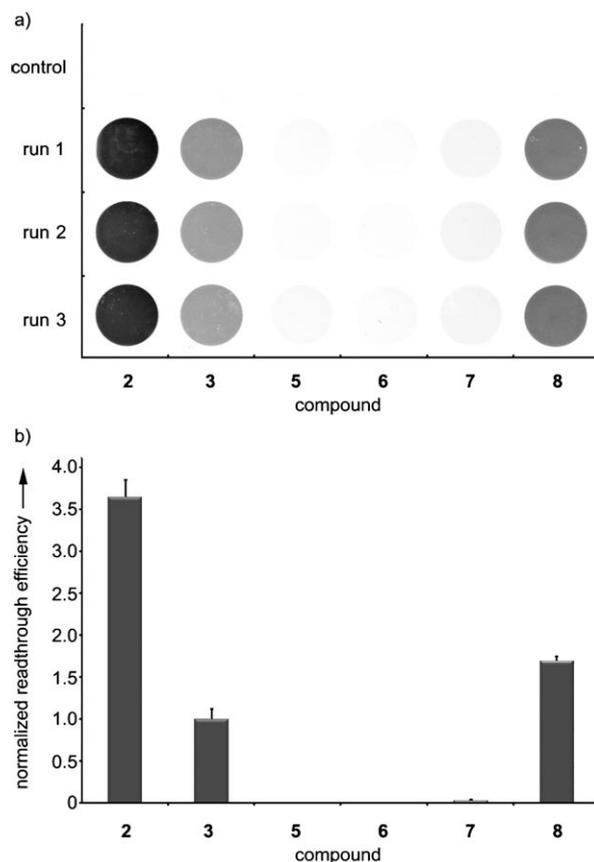


Figure 2. Readthrough efficiency of pyrrolysine analogs **2–3** and **5–8** at 2 mM (pH 7.4). a) A top-down view of the fluorescent image of the 24-well plate. The top row is the controls without supplying any pyrrolysine analogs. b) Normalized readthrough efficiency (relative to that of **3**) based on the fluorescence measurements. Error bars denote standard deviations calculated from three independent runs.

supplied analog concentration was varied from 0.5 mM to 20 mM (Figure S2 in the Supporting Information). The only notable difference was that, at higher concentrations, the level of incorporation of **8** matches even that of **2**. Importantly, the site-specificity of incorporation for **8** was confirmed by tandem mass spectrometry (Figure S3 and Tables S1, S2 in the Supporting Information).

These preliminary screening experiments indicated that dipeptide **8**, *N*<sup>6</sup>-(2-propargylglycyl)lysine (Pra-ε-Lys), is by far the best substrate in terms of the readthrough efficiency among the four new compounds tested. Because analogs **6–8** were prepared and screened as equimolar mixtures of two diastereomers, we carried out further experiments with isomerically pure samples of **8** prepared from commercially available homochiral propargylglycines (see the Supporting Information).<sup>[17]</sup> The readthrough efficiency of (*R*)-**8** (D-Pra-ε-Lys), (*S*)-**8** (L-Pra-ε-Lys), and their equimolar mixture **8** was measured at various concentrations and the results showed that (*R*)-**8** is incorporated much more efficiently than (*S*)-**8** (Figure 3). For instance, at 2 mM, (*R*)-**8** incorporates ~35 times more readily than its isomeric counterpart (*S*)-**8**. Interestingly, it is apparent from the data collected

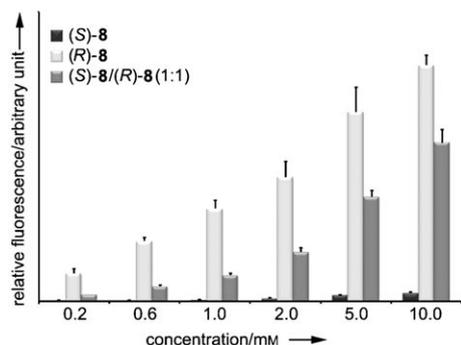


Figure 3. Concentration-dependent readthrough efficiency of pyrrolysine analogs (*R*)-**8**, (*S*)-**8**, and their equimolar mixture **8** at pH 8.5. Error bars indicate standard deviations calculated from three independent runs.

that (*R*)-**8** is markedly more efficient on its own than in the presence of (*S*)-**8**. One possible explanation is that the significantly less readily incorporable isomer (*S*)-**8** serves nonetheless as an effective competitive inhibitor of incorporation of (*R*)-**8**.

The observation that the *D*-Pra derivative (*R*)-**8** exhibits a much higher readthrough efficiency than its *L*-Pra counterpart (*S*)-**8** is notable. It was previously established by Polycarpo et al.<sup>[10]</sup> and Li et al.<sup>[7]</sup> that in amide-type analogs possessing cyclic acyl substituents at the lysine N-6, a heteroatom needs to be placed within the ring at a position strictly corresponding to that of the imine nitrogen in **1** (carbamate-type analogs do not require an additional heteroatom).<sup>[11]</sup> Therefore, while **2** and (*R*)-**18** (Figure 4) are competent pyrrolysine mimics, (*S*)-**18**, (*S*)-**19**, and (*R*)-**19** are not.<sup>[7,10]</sup>

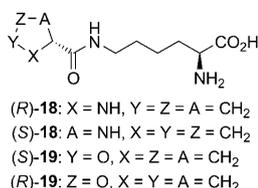


Figure 4. Potential pyrrolysine analogs studied by Polycarpo et al. (**18**)<sup>[10]</sup> and Li et al. (**19**).<sup>[7]</sup>

Having identified (*R*)-**8** as a strong outperformer of **3** in both the ease of synthesis and incorporation efficiency, we then focused on optimizing the expression conditions by adjusting the pH of the *E. coli* medium. Our working rationale for this study was based on the previous structural analysis of *M. mazei* PylS bound to adenylated **1** and pyrophosphate which suggested that a conserved tyrosine (Tyr384) on a flexible loop moves into the substrate-binding pocket to form, via its hydroxyl group, a hydrogen bond with the imine nitrogen of **1**.<sup>[18]</sup> This interaction might be important for pyrrolysine recognition and subsequent charging onto PylT. We speculated that, given its analogous position, the amine in (*R*)-**8** could participate in a similar hydrogen bond.<sup>[19]</sup> While the imine group of **1** would be fully depro-

tonated in standard *E. coli* media (pH 7.4), the amino group of (*R*)-**8** would be chiefly protonated which would hinder its ability to serve as a hydrogen-bond acceptor. However, at slightly higher pH, the amine should be deprotonated and, consequently, the readthrough efficiency of (*R*)-**8** would be expected to increase. Indeed, when 2 mM (*R*)-**8** was used, the production of full-length mCherry was higher at more basic pH of the growth medium (Figure 5). The total amount of the produced protein peaks at pH 8.5, with a yield five times greater than that at pH 6.5.

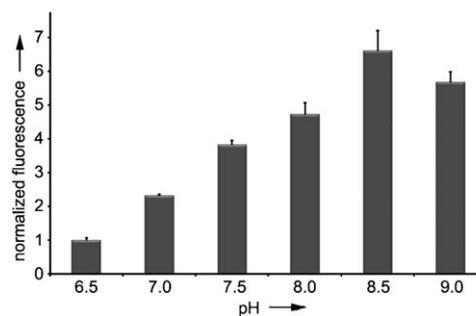


Figure 5. pH-dependent readthrough efficiency of 2 mM (*R*)-**8**. Fluorescence intensity of cell culture grown in media at various pH was normalized to that recorded at pH 6.5. Error bars indicate standard deviations calculated from three independent runs.

Although these data appeared to support our initial hypothesis, the results of other studies did not. Specifically, readthrough efficiency experiments with other pyrrolysine analogs, particularly **2** and **3**, yielded similar pH-dependent profiles (Figure S4 in the Supporting Information). These latter findings suggest that the hydrogen-bonding interaction which we invoked in our initial hypothesis (vide supra) is presumably not the major contributor to the observed pH effect. Instead, it is most likely that the  $\alpha$ -amino group of the lysine residue, which is an invariant structural feature in all the pyrrolysine analogs tested, is responsible for the observed pH effect. Another possibility is that the intracellular concentration of the pyrrolysine analogs is also pH dependent consistent with the fact that this parameter is also known to affect transport and metabolism of amino acids in *E. coli* resulting in changes to their cellular concentration.<sup>[20]</sup> Finally, the ability of PylS to recognize its substrates could be influenced by pH via changes in the protonation state of key residues such as Tyr384 (the hydrogen-bond donor in our original hypothesis). Ultimately, whatever the actual reason, higher pH levels appear to be favorable for readthrough regardless of the structural and electronic properties of our pyrrolysine analogs.

Following the demonstration that (*R*)-**8** can be efficiently incorporated into mCherry, we expanded our studies to determine whether the developed protocol could be applied to other proteins. Therefore, calmodulin (CaM),<sup>[21]</sup> a small 17-kDa model protein widely used in biochemical and structural studies, was chosen as the second target for the incorporation of (*R*)-**8**. *Rattus norvegicus* CaM cDNA bearing a UAG

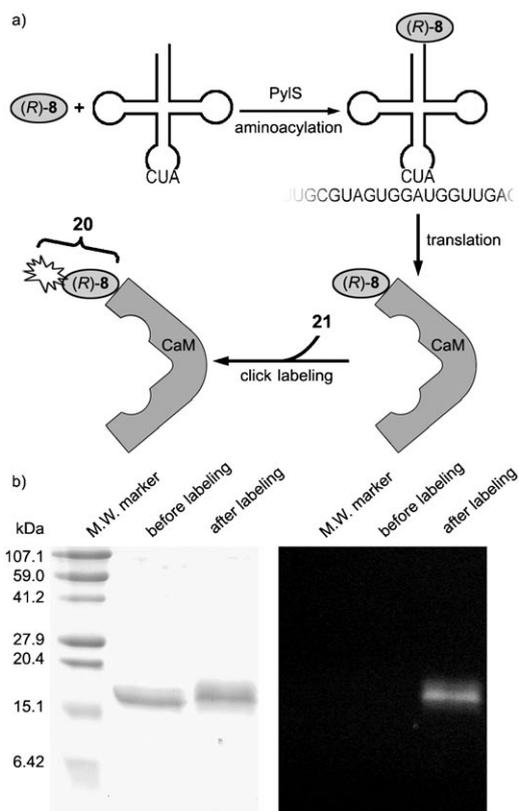
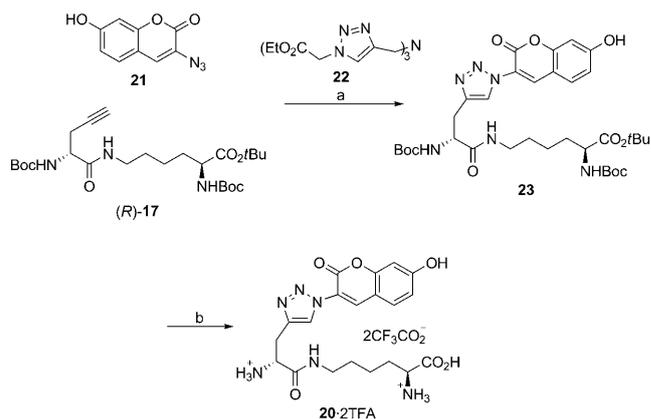


Figure 6. a) Incorporation of (*R*)-**8** into CaM and subsequent labeling with azidocoumarin **21** by CuAAC click reaction. b) Following the click reaction, the protein bands on the same SDS-PAGE gel were visualized first by UV excitation (right panel) and then by coomassie blue staining (left panel).

codon at position 34 was subcloned into pPylST and expressed in the host *E. coli* strain BL21(DE3) in the presence of 2 mM (*R*)-**8** (Figure 6). About 40 mg of (*R*)-**8**-CaM can be isolated from 1 L of the culture, which is a considerably higher yield than that reported for any other clickable pyrrolysine mimic.<sup>[11,14,22]</sup>

To facilitate our protein taggability studies, we first prepared **20** (Scheme 2), the click product between our most effective pyrrolysine mimic (*R*)-**8** and azidocoumarin **21**.<sup>[23]</sup> Thus, the CuAAC reaction between (*R*)-**17** and **21** in the presence of catalytic amounts of CuSO<sub>4</sub>, tris-triazole ligand **22**,<sup>[24,25]</sup> and sodium ascorbate gave the coupling product **23** in high yield (84%). Deprotection with TFA furnished **20** as its TFA salt. Subsequent spectroscopic studies revealed that when azidocoumarin **21** is CuAAC-coupled with either (*R*)-**8** or an (*R*)-**8**-containing protein, two major spectrometric changes are expected to be observed, that is, the red shift of absorbance from 350 nm to 396 nm and a concomitant significant increase in the fluorescence emission intensity at 470 nm (Figure S5 in the Supporting Information).<sup>[26]</sup> Armed with this knowledge, we set about to test the taggability of the purified (*R*)-**8**-CaM with azidocoumarin **21** under the standard CuAAC conditions. By continuously measuring either absorbance or fluorescence of the reaction mixture, we were able to determine that the coupling between **21**



Scheme 2. Synthesis of click product **20**: a) sodium ascorbate (40 mol %), **22** (10 mol %), CuSO<sub>4</sub> (10 mol %), *t*BuOH, H<sub>2</sub>O, room temperature, 20 h, 84%; b) TFA, room temperature, 2 h, ~100%.

and (*R*)-**8**-CaM was nearly completed within 1 h (Figure S5 in the Supporting Information). To confirm the successful labeling of the protein, (*R*)-**8**-CaM before and after labeling was characterized by SDS-PAGE gel electrophoresis and UV imaging (Figure 6).

Coumarin is a fluorophore whose spectral properties are strongly dependent on the environment it resides in.<sup>[27]</sup> It therefore serves as a useful dye for probing protein dynamics. CaM is known to undergo dramatic conformational changes in response to the concomitant presence of Ca<sup>II</sup> and a CaM-binding protein. With our long-term goal of developing a CaM-based Ca<sup>II</sup> sensor in mind, a highly efficient CaM-binding peptide, M13,<sup>[28]</sup> was genetically fused to the C-terminus of (*R*)-**8**-CaM. After labeling the resulting (*R*)-**8**-CaM-M13 with azidocoumarin **21**, we measured the fluorescence of this protein as a function of the Ca<sup>II</sup> concentration, and observed that with an increase from 0 to 1.0 × 10<sup>-4</sup> M, the fluorescence owing to the coumarin tag decreased by 45% (Figure 7). Thus, (*R*)-**8**-CaM-M13 can be

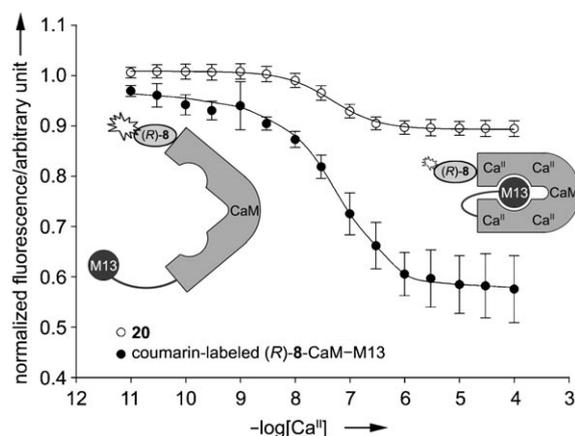


Figure 7. Fluorescence intensity of coumarin-labeled (*R*)-**8**-CaM-M13 and **20** as a function of the Ca<sup>II</sup> concentration. Data normalized to the fluorescence intensity in the absence of Ca<sup>II</sup>. Error bars indicate standard deviations calculated from three independent runs.

used as a tool to detect and quantitatively measure  $\text{Ca}^{II}$  within the concentration range of  $10^{-9}$ – $10^{-4}$  M.

In summary, through systematic studies, we have identified a structurally simple and easily accessible taggable pyrrolysine analog (*R*)-**8** (D-Pra-ε-Lys) and demonstrated its ability to be readily incorporated into two selected proteins in response to the UAG codon. The ability of both (*R*)-**8** and (*R*)-**4**<sup>[9]</sup> to be read through the UAG codon in the presence of PylST suggests the likely possibility that lysines acylated at N-6 with D-amino acids could constitute a general set of compounds amenable to translational incorporation by the pyrrolysine machinery. We have also developed a highly convenient UAG codon readthrough assay that should be indispensable in a search for new pyrrolysine analogs.

### Acknowledgements

This research was supported by an American Heart Association Great Rivers Affiliate Predoctoral Fellowship (0815449D) to X.L. and National Institutes of Health grant (GM061796) to M.K.C. supplemented by American Recovery and Reinvestment Act (ARRA) funds. We also thank the staff of the CCIC Mass Spectrometry and Proteomics Facility at OSU for assistance with protein analysis.

**Keywords:** amino acids • cycloaddition • fluorescent probes • protein modifications • pyrrolysine

- [1] M. S. Gonçalves, *Chem. Rev.* **2009**, *109*, 190–212.
- [2] B. N. Giepmans, S. R. Adams, M. H. Ellisman, R. Y. Tsien, *Science* **2006**, *312*, 217–224.
- [3] I. Chen, A. Y. Ting, *Curr. Opin. Biotechnol.* **2005**, *16*, 35–40.
- [4] T. L. Foley, M. D. Burkart, *Curr. Opin. Chem. Biol.* **2007**, *11*, 12–19.
- [5] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459–1462.
- [6] B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki, M. K. Chan, *Science* **2002**, *296*, 1462–1466.
- [7] 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.
- [8] T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem.* **2009**, *121*, 1661–1663; *Angew. Chem. Int. Ed.* **2009**, *48*, 1633–1635.
- [9] X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew. Chem.* **2009**, *121*, 9348–9351; *Angew. Chem. Int. Ed.* **2009**, *48*, 9184–9187.
- [10] C. R. Polycarpo, S. Herring, A. Berube, J. L. Wood, D. Soll, A. Ambrogelly, *FEBS Lett.* **2006**, *580*, 6695–6700.
- [11] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187–1197.
- [12] T. Kobayashi, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2009**, *385*, 1352–1360.
- [13] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234.
- [14] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721.
- [15] W. Liu, A. Brock, S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239–244.
- [16] W. Wang, J. K. Takimoto, G. V. Louie, T. J. Baiga, J. P. Noel, K. F. Lee, P. A. Slesinger, L. Wang, *Nat. Neurosci.* **2007**, *10*, 1063–1072.
- [17] H. K. Chenault, J. Dahmer, G. M. Whitesides, *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721.
- [18] J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Soll, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11268–11273.
- [19] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2008**, *378*, 634–652.
- [20] L. M. Stancik, D. M. Stancik, B. Schmidt, D. M. Barnhart, Y. N. Yoncheva, J. L. Slonczewski, *J. Bacteriol.* **2002**, *184*, 4246–4258.
- [21] A. S. Manalan, C. B. Klee, *Adv. Cyclic Nucleotide Protein Phosphorylation Res.* **1984**, *18*, 227–278.
- [22] A. Deiters, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1521–1524.
- [23] K. Sivakumar, F. Xie, B. M. Cash, S. Long, H. N. Barnhill, Q. Wang, *Org. Lett.* **2004**, *6*, 4603–4606.
- [24] S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis, *Nature* **2007**, *446*, 1105–1109.
- [25] H. B. Kramer, J. M. Chalker, B. G. Davis, *Nat. Protoc.* **2007**, DOI: 10.1038/nprot.2007.462.
- [26] P. Chittepudi, V. R. Sirivolu, F. Seela, *Bioorg. Med. Chem.* **2008**, *16*, 8427–8439.
- [27] G. Jones, W. R. Jackson, C. Y. Choi, W. R. Bergmark, *J. Phys. Chem.* **1985**, *89*, 294–300.
- [28] M. Ikura, G. M. Clore, A. M. Gronenborn, G. Zhu, C. B. Klee, A. Bax, *Science* **1992**, *256*, 632–638.

Received: March 15, 2010  
Published online: June 11, 2010