

## N-Terminal Protein Modification through a Biomimetic Transamination Reaction\*\*

Joshua M. Gilmore, Rebecca A. Scheck,  
Aaron P. Esser-Kahn, Neel S. Joshi, and  
Matthew B. Francis\*

An important challenge in the field of protein bioconjugation is the development of strategies that can modify a protein once at a single site. Although many bioconjugation reactions can functionalize specific amino acids in aqueous solution,<sup>[1]</sup> most proteins display multiple copies of the targeted residue on their surface. This commonly results in product mixtures that present the new functionality in multiple locations on the protein surface. As it is one of the rarest amino acids,<sup>[2]</sup> cysteine is the most commonly targeted residue when site-selective modification is required, but there remain many situations in which the modification of a unique copy of this residue is inconvenient or impossible. To address these limitations, several chemoselective techniques have been developed to target surface-accessible aromatic residues.<sup>[3]</sup> To complement these methods further, we report herein a biomimetic transamination reaction that can modify the N terminus of proteins and peptides under mild conditions. This technique introduces a uniquely reactive ketone or aldehyde group in a single location, thus allowing further modification through oxime or hydrazone formation. This simple strategy does not require the use of site-directed mutagenesis, and therefore has the potential to introduce virtually any functional group on a wide range of protein substrates.

The unique reactive properties of the N terminus have resulted in several strategies targeting this location to achieve site-selective protein modification. To a limited extent, the lower  $pK_a$  value of N-terminal amino groups (relative to lysine side chains) can be used to direct acylation reactions to

[\*] J. M. Gilmore,<sup>[†]</sup> R. A. Scheck,<sup>[†]</sup> A. P. Esser-Kahn, N. S. Joshi,  
Prof. M. B. Francis  
Department of Chemistry  
University of California  
Berkeley and Materials Sciences Division  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720 (USA)  
Fax: (+1) 510-643-3079  
E-mail: francis@cchem.berkeley.edu

[†] These authors contributed equally to this work.

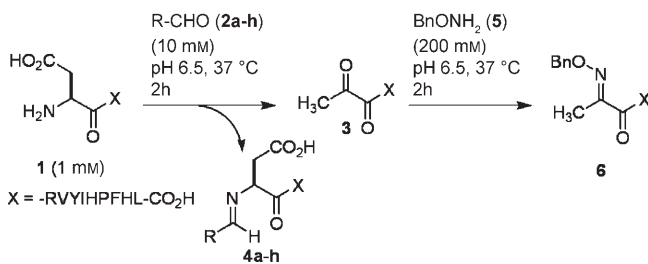
[\*\*] The authors thank the DOE Nanoscale Science and Engineering Technology (NSET) program for financial support, as well as the Department of Chemistry at the University of California, Berkeley, and the Berkeley Chemical Biology Graduate Program. N.S.J. gratefully acknowledges the Saegebarth family for generous fellowship support.



Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.





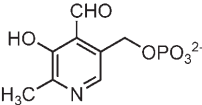

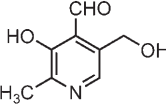

this site through careful control of the reaction pH.<sup>[4]</sup> However, the large number of competing lysine residues (10–40 on most proteins) limits the selectivity of this reaction in most cases. Alternative strategies have targeted the N terminus in combination with specific amino acid side chains. In particular, the reaction of N-terminal cysteine residues with thioesters has been used with great success, and affords fusion proteins through native chemical ligations.<sup>[5]</sup> N-terminal cysteine residues have also been modified through thiazolidine formation by using aldehyde reagents.<sup>[6]</sup> Reactive aldehydes can be formed through periodate oxidation of N-terminal serine and threonine residues,<sup>[7]</sup> and N-terminal tryptophan residues can be modified through Pictet–Spengler reactions.<sup>[8]</sup>

An alternative strategy that does not depend on the identity of the amino acid side chain can be envisioned through the oxidation of the N-terminal amino group to an imine, followed by hydrolysis to afford a ketone or an aldehyde. This reaction pathway has indeed been demonstrated for the chemospecific functionalization of N termini by using glyoxylic acid, copper(II) salts, 1M pyridine, and 1M HOAc.<sup>[9]</sup> However, these conditions are too harsh to maintain the folded structure of most proteins, and are therefore more appropriate for sequence-analysis applications. In contrast to these reports, however, initial studies by our research group found that the N-terminal aspartic acid residue of angiotensin I (**1**) unexpectedly decarboxylated to a small extent (ca. 10%) when the peptide was simply treated with glyoxylic acid (**2a**; Scheme 1). The product of the reaction was identified as pyruvamide **3** through subsequent reaction with *O*-benzylhydroxylamine (**5**) to form oxime **6**. Screening experiments indicated that the structure of the aldehyde substituent strongly influenced the reaction efficiency (Table 1) and that the decarboxylated product was often accompanied by various amounts of imines **4a–h**. Interestingly, pyridoxal-5-phosphate (PLP, **2c**) emerged as the most effective aldehyde, thus affording a 65% conversion of **1** into ketone **3** in 2 h at 37°C, with no observed imine formation.



**Scheme 1.** Formation of oxime **6** from the N-terminal aspartic acid residue of angiotensin I (**1**). Bn = benzyl.

**Table 1:** Modification of angiotensin I (**1**) using aldehyde reagents.<sup>[a]</sup>

Aldehyde		3 [%]	4a–h [%]	Aldehyde		3 [%]	4a–h [%]
2a		ca. 10	trace	2e		0	25
2b		14	17	2f		10	50
2c		65	0	2g		trace	> 90
2d		0	0	2h		20	40

[a] Conditions: 1 mM **1**, 10 mM aldehyde, 50 mM phosphate buffer (pH 6.5), 37°C, 2 h. Product distributions were determined by ESI-MS.

Subsequent optimization experiments determined that a 1 mM peptide solution could be modified by using 10 mM PLP in 50 mM phosphate buffer at pH 6.5 and 37°C (Table 2, entry 1). Consistent with other reactions involving carbonyl condensations,<sup>[10]</sup> reactions at pH 6.5 afforded the optimal

**Table 2:** Peptide modification using **2c**.<sup>[a]</sup>

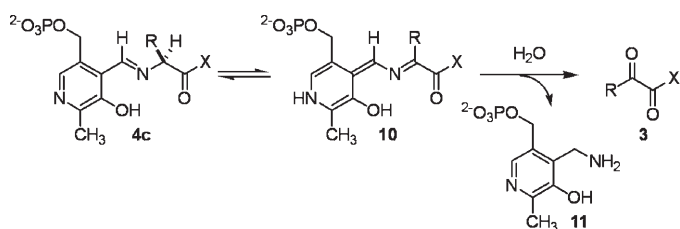
Entry	Peptide	Sequence	pH	T [°C]	t [h]	Conv [%]
1	<b>1</b>	DRYVIHPFHL	6.5	37	1	65
2	<b>1</b>	DRYVIHPFHL	5.5	37	1	29
3	<b>1</b>	DRYVIHPFHL	7.5	37	1	46
4	<b>1</b>	DRYVIHPFHL	8.5	37	1	43
5	<b>1</b>	DRYVIHPFHL	6.5	65	2	> 95
6	<b>1</b>	DRYVIHPFHL	6.5	22	2	55
7	<b>1</b>	DRYVIHPFHL	6.5	22	18	88
8	<b>7</b>	DNALRAFKYM	6.5	37	2	90
9	<b>8</b>	MYKFARLAND	6.5	37	2	71
10	<b>9</b>	GLEEEYWKWGDKNYDG	6.5	37	2	43

[a] Conditions for peptides: 1 mM peptide, 10 mM **2c**, 50 mM phosphate buffer. Product distributions were determined by ESI-MS after oxime formation with **5**.

amount of product, although slight variations were still tolerated (Table 2, entries 2–4). An increase in temperature to 65°C resulted in rapid conversion of **1** into **3** (> 95% in 2 h; Table 2, entry 5), although excellent conversion could also be obtained after 24 h at room temperature (Table 2, entry 7). The reaction was found to proceed readily in 2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid (HEPES), 3-(*N*-morpholine)propanesulfonic acid (MOPS), and phosphate buffers, and, contrary to previous transamination reports,<sup>[11]</sup> did not require the presence of divalent cations or denaturing organic cosolvents. A variety of Lewis acids were screened to confirm this, and little change in rate or conversion was observed.<sup>[12]</sup> Moreover, experiments performed in the presence of 10 mM ethylenediaminetetraacetate (EDTA) to sequester any metal ions in the reaction solution resulted in equivalent product yields.

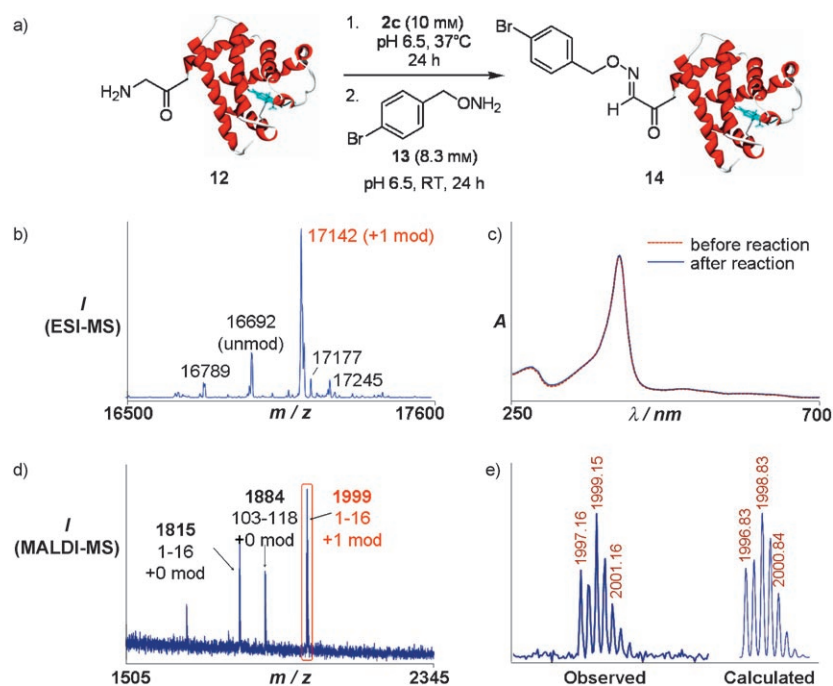
The most useful feature of this reaction is its selectivity for the N terminus of the peptide. Under the optimized conditions, peptide **7** afforded 90% conversion into the ketone product (Table 2, entry 8). These screening experiments also revealed that the N-terminal residue does not need to be aspartic acid, as peptides bearing N-terminal methionine (**8**) and glycine (**9**) residues still afforded good yields of the desired products. These peptides also provided the singly modified conjugate, despite the presence of internal lysine residues.

In a biological context, PLP is a cofactor that effects a variety of metabolic transformations under enzymatic control, including racemization, elimination, decarboxylation, and transamination.<sup>[13]</sup> Typically, PLP-mediated reactions first involve condensation of the aldehyde with amino groups on a peptide or protein, thus leading to the formation of imine **4c** (Scheme 2). In the absence of an enzyme, this process could occur with all the amino groups on a particular biomolecule. However, the imine formed with the N terminus has an  $\alpha$  proton with a much lower  $pK_a$  value, which allows tautomerization to occur uniquely at this site. Hydrolysis of the resulting glyoxyl imine **10** (accompanied by decarboxylation in the case of aspartic acid) yields an aldehyde or ketone specifically at the N terminus. Although in principle this reaction sequence could take place on the N termini of peptides and proteins in living cells, this pathway has not been reported to date. This is likely due to the low concentration of unbound PLP in the cytoplasm.



**Scheme 2.** Formation of pyruvamide **3**.

The development of the reaction for protein targets was initiated with horse heart myoglobin (**12**), which has an N-terminal glycine residue. Optimal conditions for myoglobin modification were found to be 50  $\mu$ M protein and 10 mM PLP for 20 h at 37°C in 25 mM phosphate buffer (pH 6.5; Figure 1a). Following treatment of the mixture with *p*-bromobenzoyloxyamine (**13**), it was determined by ESI-MS that a single product was obtained that corresponded to the expected oxime **14** (Figure 1b). The overall conversion of the reaction was determined to be greater than 75% by ESI-MS and 69% using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE analysis techniques, see below). Following removal of excess PLP, the absorbance of the heme component was found to be unchanged over the course of the reaction (Figure 1c). This observation provides good evi-



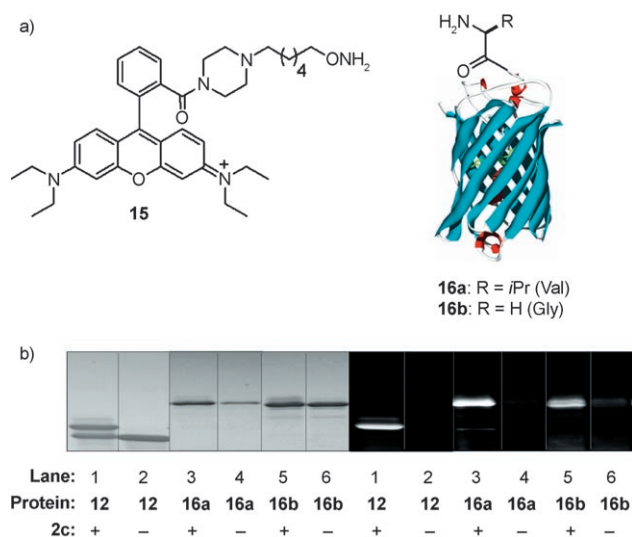
**Figure 1.** Site-selective protein modification with PLP. a) Myoglobin was treated with **2c** at 37°C for 24 h. After removal of small molecules by gel filtration, oxime **14** was formed after reaction with **13**. b) ESI-LCMS studies indicated that only one modification had occurred. c) UV/Vis spectra of **12** subjected to the reaction conditions in the absence (dotted red) and presence (dashed blue) of **2c**. d) Product **14** was digested with trypsin, and the resulting fragments were analyzed by MALDI-TOF MS. The only modified fragment observed contained the N-terminal residue. e) Predicted (left) and observed (right) isotope patterns for the modified myoglobin peptide fragment shown in (d). All MS values agree with those predicted within 0.01%.

dence that the tertiary structure of the protein was still intact following the modification procedure.

We were able to confirm the site-specificity of this modification after subjecting **14** to a proteolytic digest with trypsin. Analysis of the peptide fragments by MALDI-TOF-MS showed that the reaction had occurred on the N-terminal fragment of the protein (Figure 1d). The bromine atom of **13** provided a useful analysis tag, as the characteristic isotope pattern was specific to the modified fragment and could be observed clearly (Figure 1e). Despite the 17 lysine residues on the protein, no other modified peptides were observed, a result that is consistent with the observation that these residues do not participate in this reaction.

The oxime linkage is well suited for use as a probe or linker in biochemical systems, as it is selective for ketone and aldehyde functionalities, forms in high yield, and is stable under physiological conditions.<sup>[14]</sup> As an example, the N terminus of myoglobin was fluorescently labeled with chromophore **15**. SDS-PAGE analysis revealed selective labeling of the protein after activation with PLP (Figure 2b, lane 1), whereas no labeling was observed for unactivated myoglobin (lane 2).

This technique can also be used to label proteins when chemistry appropriate for cysteine groups is not available. For example, cysteine 70 in the green fluorescent protein<sup>[15]</sup> (EGFP, **16a**) is reactive toward maleimide reagents and this protein (in our hands) cannot be expressed in fluorescent

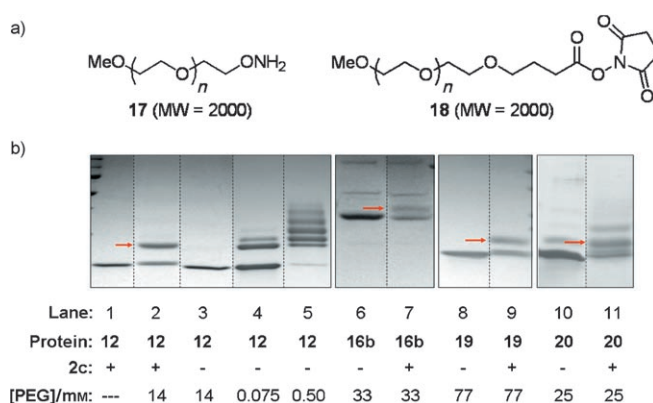


**Figure 2.** Site-specific chromophore labeling. Proteins possessing N-terminal carbonyl groups introduced by **2c** were labeled with **15**. Reaction success was confirmed by SDS-PAGE analysis using Coomassie staining (left) and fluorescence visualization (right).

form when this residue has been substituted. Thus, the labeling of other positions on this protein is difficult to achieve by standard mutagenesis techniques. We have found that the N-terminal modification method can overcome this limitation, and results in the attachment of **15** at the N-terminal valine residue of **16a** (Figure 2b, lanes 3 and 4) and at the N-terminal glycine residue of mutant **16b**. Subsequent experiments have confirmed that the internal chromophore retains its fluorescence after modification with PLP.<sup>[12]</sup>

The attachment of poly(ethylene glycol) (PEG) chains to proteins has been shown to improve circulation lifetime, lower immunogenicity, and confer proteolysis resistance.<sup>[16]</sup> These conjugates are typically prepared through lysine modification, which results in complex mixtures that exhibit varying levels of efficacy. To avoid this, we have used the two-step strategy outlined above to attach a single polymer chain to the N terminus of proteins.<sup>[17,18]</sup> Specifically, the aldehyde of PLP-activated myoglobin was treated with various equivalents of the previously reported<sup>[19]</sup> PEG alkoxyamine **17**. As seen in Figure 3b (lanes 1–3), a single bioconjugate was obtained even when using a large excess of **17**. Using similar conditions, this reaction strategy has also proven effective for EGFP (**16b**, lanes 6 and 7), RNase A (**19**, lanes 8 and 9), and thioredoxin (**20**, lanes 10 and 11). For comparison, PEG-NHS ester **18** yielded a ladder of polymer conjugates for the labeling of **12** (Figure 3b, lanes 4 and 5). Clean conversion to a singly modified protein was not observed under any conditions using this approach.

The shift on the gel caused by polymer attachment provides a useful method for the quantitation of reaction conversion by optical densitometry measurements after



**Figure 3.** Site-selective attachment of PEG polymers to proteins. a) PEG-alkoxyamine **17** was synthesized and coupled to N-terminal carbonyl groups introduced using **2c**. b) Analysis by SDS-PAGE with Coomassie staining confirmed single-site labeling of the proteins (lanes 1–3, 6–11). The red arrows indicate the shift in the position on the gel caused by the attached PEG polymer. PEG-NHS ester **18** was used to label lysine residues of **12** nonspecifically (lanes 4 and 5). A small portion of thioredoxin (**20**) reacts with the PEG-alkoxyamine without treatment with **2c** (lane 10). This background reaction gives rise to the small portion of doubly labeled product shown in lane 11. Protein **19** is RNase A.

Coomassie staining (Table 3). For myoglobin, 69% overall conversion was achieved at 37 °C, and for EGFP (**16b**) 53% conversion was obtained. High levels of modification could be obtained for **16a** and **16b** at 55 °C. RNase A (**19**) was cleanly modified in 50% conversion at 37 °C, as was thioredoxin (**20**) under identical conditions. The large number of disulfide and free thiol groups that proteins **19** and **20** possess render them difficult to modify in a single location by other methods. Finally, initial studies have found that protein G' (**21**) can be modified to 30% at 41 °C, thus indicating that proteins with N-terminal methionine residues are compatible with this technique. Studies to increase the conversion for this important class of substrates are in progress, as are studies to evaluate the compatibility of other N-terminal amino acids.

We anticipate that proteins possessing N-terminal serine, threonine, cysteine, and tryptophan residues will be incompatible with this technique because of known side reactions with aldehydes.<sup>[6,8]</sup> It is also presumed that N-terminal proline residues will be unreactive. Outside of these limitations, this reaction provides a generally applicable tool for the site-selective labeling of protein targets. Our current efforts use

**Table 3:** Protein modification using **2c**.<sup>[a]</sup>

Entry	Protein	N-terminal amino acid	Conc. [ $\mu$ M]	T [ $^{\circ}$ C]	Conv. [%]
1	myoglobin ( <b>12</b> )	glycine	50	37	69
2	GFP-1V ( <b>16a</b> )	valine	10	55	67
3	GFP-1G ( <b>16b</b> )	glycine	10	37	41
4	GFP-1G ( <b>16b</b> )	glycine	10	55	80
5	RNase A ( <b>19</b> )	lysine	50	37	50
6	thioredoxin ( <b>20</b> )	glycine	94	37	50
7 <sup>[b]</sup>	protein G' ( <b>21</b> )	methionine	33	41	30

[a] Conditions: 10 mM **2c**, 25 mM phosphate buffer, pH 6.5, 18–20 h. Product yields were determined by densitometry after derivitization with **17** and separation by SDS-PAGE (Coomassie stain, see Figure 3).

[b] Extent of protein modification was determined by ESI-MS after oxime formation with **5**.



this methodology for the site-specific placement of fluorophores for sensing applications and the dual modification of both protein termini for device integration.

Received: January 27, 2006

Revised: May 27, 2006

Published online: July 17, 2006

**Keywords:** bioconjugation · bioorganic chemistry · oximes · protein modifications

- [1] For an excellent review on protein bioconjugation, see: G. T. Hermanson, *Bioconjugate Techniques*, 1st ed., Academic Press, San Diego, **1996**.
- [2] R. F. Doolittle in *Prediction of Protein Structure and the Principles of Protein Conformation* (Ed.: G. D. Fasman), Plenum, New York, **1989**.
- [3] a) J. M. Antos, M. B. Francis, *J. Am. Chem. Soc.* **2004**, *126*, 10256; b) N. S. Joshi, L. R. Whitaker, M. B. Francis, *J. Am. Chem. Soc.* **2004**, *126*, 15924; c) J. M. Hooker, E. W. Kovacs, M. B. Francis, *J. Am. Chem. Soc.* **2004**, *126*, 3718–3719; d) S. D. Tilley, M. B. Francis, *J. Am. Chem. Soc.* **2006**, *128*, 1080–1081.
- [4] D. P. Baker, E. Y. Lin, K. Lin, M. Pellegrini, R. C. Petter, L. L. Chen, R. M. Arduini, M. Brickelmaier, D. Wen, D. M. Hess, L. Chen, D. Grant, A. Whitty, A. Gill, D. J. Lindner, R. B. Pepinsky, *Bioconj. Chem.* **2006**, *17*(1), 179–188.
- [5] a) P. E. Dawson, T. W. Muir, I. Clarklewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779; b) T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
- [6] J. P. Tam, Q. Yu, Z. Miao, *Biopolymers* **1999**, *51*, 311–332.
- [7] a) K. F. Geoghegan, J. G. Stroh, *Bioconjugate Chem.* **1992**, *3*, 138–146; b) J. K. Chen, W. S. Lane, A. W. Brauer, A. Tanaka, S. L. Schreiber, *J. Am. Chem. Soc.* **1993**, *115*, 12591–12592.
- [8] X. Li, L. Zhang, S. E. Hall, J. P. Tam, *Tetrahedron Lett.* **2000**, *41*, 4069–4073.
- [9] H. B. F. Dixon, *J. Protein Chem.* **1984**, *3*, 99–108.
- [10] W. P. Jencks, *J. Am. Chem. Soc.* **1959**, *81*, 475–481.
- [11] R. Fields, H. B. F. Dixon, *Methods Enzymol.* **1972**, *25*, 409–419.
- [12] See the Supporting Information for details.
- [13] For reviews of PLP-mediated reactions, see: a) M. D. Toney, *Arch. Biochem. Biophys.* **2005**, *433*, 279–287; b) A. C. Eliot, J. F. Kirsch, *Annu. Rev. Biochem.* **2004**, *73*, 383–415.
- [14] a) M. Calzadilla, T. Cordova, A. Malpica, *J. Phys. Org. Chem.* **1999**, *12*, 708–712; b) S. Rosenberg, S. M. Silver, J. M. Sayer, W. P. Jencks, *J. Am. Chem. Soc.* **1974**, *96*, 7987; see also: c) H. C. Hang, C. R. Bertozzi, *Acc. Chem. Res.* **2001**, *34*, 727–736; d) V. W. Cornish, K. M. Hahn, P. G. Schultz, *J. Am. Chem. Soc.* **1996**, *118*, 8150–8151.
- [15] R. Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509–544.
- [16] R. E. Offord, H. F. Gaertner, *Bioconjugate Chem.* **1996**, *7*, 38–44.
- [17] For an avidin/biotin-based approach to the site-selective attachment of polymers to proteins, see: D. Bontempo, H. D. Maynard, *J. Am. Chem. Soc.* **2005**, *127*, 6508–6509.
- [18] For an alternate strategy that employs solid-phase-assisted synthesis to yield an aminooxy-functionalized protein which has also been used to prepare PEG monoconjugates, see: H. Shao, M. M. Crnogorac, T. Kong, S.-Y. Chen, J. M. Williams, J. M. Tack, V. Gueriguian, E. N. Cagle, M. Carnevali, D. Tumelty, X. Paliard, L. P. Miranda, J. A. Bradburne, G. G. Kochendoerfer, *J. Am. Chem. Soc.* **2005**, *127*, 1350–1351; see also: G. G. Kochendoerfer, *Curr. Opin. Chem. Biol.* **2005**, *9*, 555–560.
- [19] T. L. Schlick, Z. Ding, E. W. Kovacs, M. B. Francis, *J. Am. Chem. Soc.* **2005**, *127*, 3718–3723.