JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Genetic Encoding of Bicyclononynes and *trans*-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions

Kathrin Lang,[†] Lloyd Davis,[†] Stephen Wallace,[†] Mohan Mahesh,[†] Daniel J. Cox,[†] Melissa L. Blackman,[‡] Joseph M. Fox,[‡] and Jason W. Chin*[,†]

[†]Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, United Kingdom
[‡]Brown Laboratory, Department of Chemistry & Biochemistry, University of Delaware, Newark, Delaware 19716, United States

Supporting Information

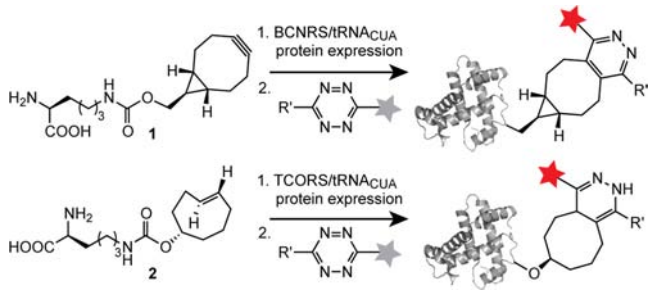
ABSTRACT: Rapid, site-specific labeling of proteins with diverse probes remains an outstanding challenge for chemical biologists. Enzyme-mediated labeling approaches may be rapid but use protein or peptide fusions that introduce perturbations into the protein under study and may limit the sites that can be labeled, while many "bioorthogonal" reactions for which a component can be genetically encoded are too slow to effect quantitative site-specific labeling of proteins on a time scale that is useful for studying many biological processes. We report a fluorogenic reaction between bicyclo[6.1.0]non-4-yn-9-ylmethanol (BCN) and tetrazines that is 3−7 orders of magnitude faster than many bioorthogonal reactions. Unlike the reactions of strained alkenes, including *trans*-cyclooctenes and norbornenes, with tetrazines, the BCN–tetrazine reaction gives a single product of defined stereochemistry. We have discovered aminoacyl-tRNA synthetase/tRNA pairs for the efficient site-specific incorporation of a BCN-containing amino acid, **1**, and a *trans*-cyclooctene-containing amino acid **2** (which also reacts extremely rapidly with tetrazines) into proteins expressed in *Escherichia coli* and mammalian cells. We demonstrate the rapid fluorogenic labeling of proteins containing **1** and **2** in vitro, in *E. coli*, and in live mammalian cells. These approaches may be extended to site-specific protein labeling in animals, and we anticipate that they will have a broad impact on labeling and imaging studies.

Inverse electron-demand Diels–Alder reactions between strained alkenes (including norbornenes and *trans*-cyclooctenes) and tetrazines have emerged as an important class of rapid bioorthogonal reactions.[1−4] The rates reported for some of these reactions are incredibly high.[3,4] Very recently, three approaches for specific labeling of proteins using these reactions have been reported. A lipoic acid ligase variant that accepts a *trans*-cyclooctene substrate was used to label proteins bearing a 13 amino acid lipoic acid ligase tag in a two-step procedure.[5] A tetrazine was introduced at a specific site in a protein expressed in *Escherichia coli* via genetic code expansion and then derivatized with a strained *trans*-cyclooctene–diacetylfluorescein.[6] We demonstrated the incorporation of a strained alkene (a norbornene-containing amino acid) via genetic code expansion and site-specific fluorogenic labeling with tetrazine fluorophores in vitro, in *E. coli*, and in live mammalian cells.[7] Related work was recently reported by others,[8,9] and while the present work was under review, the low-efficiency incorporation of one of the amino acids reported here (**2**) and the detection of some fluorescent labeling in fixed cells was described.[9]

Recent work in organic solvents suggests that the reaction between tetrazines and BCN, which was first described in strain-promoted reactions with azides,[10] may proceed very rapidly.[11] However, this reaction, unlike the much slower reaction of simple cyclooctynes with azides, nitrones,[12−16] and tetrazines,[9,17] has not been explored in aqueous media or as a chemoselective route to label macromolecules.

Here we present a rapid and fluorogenic reaction between tetrazines and BCN and demonstrate the genetic encoding of BCN- and *trans*-cyclooctene-containing amino acids **1** and **2**, respectively, in *E. coli* and mammalian cells. We show the specific, rapid, and fluorogenic labeling of proteins in *E. coli* and live mammalian cells with tetrazine probes, and explicitly demonstrate the advantages of this approach (Scheme 1).

Scheme 1. Genetic Encoding and Fluorogenic Labeling of Unnatural Amino Acids 1 and 2



The rate constants for the reactions of various dienophiles [BCN, *trans*-cyclooctene-4-ol (TCO), and bicyclo[6.1.0]non-4-ene-9-ylmethanol (sTCO)] with tetrazines have been determined.[3−5,9,11] However, in many cases, different tetrazines,

Received: March 23, 2012
Published: June 13, 2012

   dx.doi.org/10.1021/ja302832g | J. Am. Chem. Soc. 2012, 134, 10317−10320

solvent systems, or measurement methods were used, making it challenging to compare quantitatively the reactivities of the dienophiles with tetrazines of interest. Our initial experiments confirmed that the reactions of the dienophiles **1**–**3** with tetrazine **6** (Figure 1) were too fast to study by manual mixing

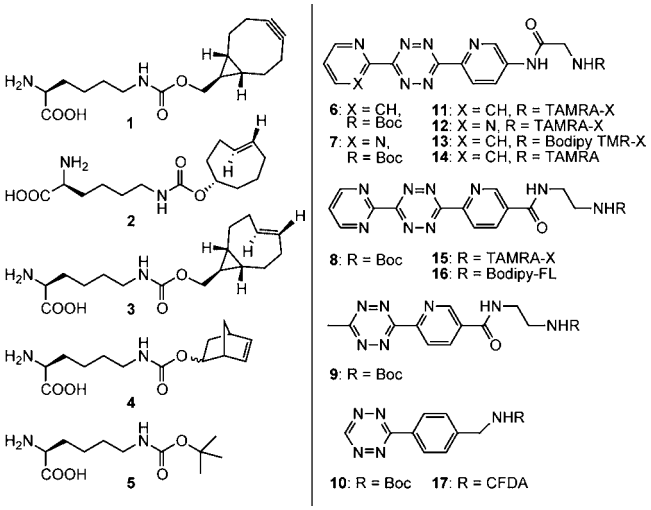


Figure 1. Structures of unnatural amino acids **1**–**5** and tetrazine derivatives **6**–**17** used in this study. For the structures of TAMRA-X, Bodipy TMR-X, Bodipy-FL, and CFDA, see Figure S4.

under pseudo-first-order conditions. We therefore turned to stopped-flow techniques to determine the pseudo-first-order rate constants for these reactions. By following the exponential decay of the tetrazine absorbance at 320 nm upon reaction with a 10–100-fold excess of BCN in 55:45 methanol/water, we determined the rate constants for the reactions of BCN with **6** and **7** as 437 ± 13 and 1245 ± 45 M^{-1} s^{-1}, respectively [Figure 2a,b and Figure S2a,b in the Supporting Information (SI)].

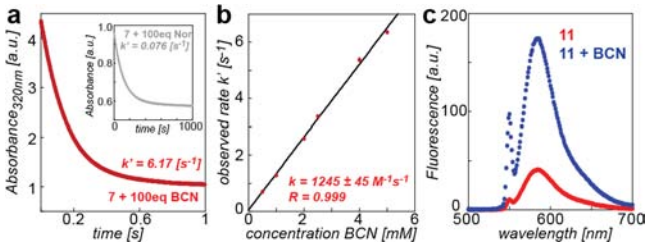


Figure 2. Characterization of the reaction of BCN with **7**. (a) Stopped-flow kinetics of the reaction. The inset shows the conjugation of **7** to 5-norbornene-2-ol (Nor); the different time scales should be noted. Conditions: c_7 = 0.05 mM and c_{BCN} = c_{Nor} = 5 mM in 55:45 MeOH/H_2O at 25 °C. (b) Determination of the second-order rate constant k for the reaction of **7** and BCN. (c) Fluorogenic reaction of **11** with BCN.

LC–MS and NMR analysis confirmed the products (see the SI and Figure S1). Under the same conditions, the rate constants for the reactions of TCO with **6** and **7** were 5235 ± 258 and 17248 ± 3132 M^{-1} s^{-1}, respectively (Figure S3). The reaction between BCN and **6** is ∼1000 times faster than the reaction between 5-norbornene-2-ol and **6**,[7] while the TCO reaction is 10–15 times faster than that with BCN. The sTCO reaction was too fast for accurate measurements by stopped-flow techniques, and we estimate that it is at least 50 times faster than the TCO reaction. Similar rate accelerations were observed for the reactions of BCN with **7**, **8**, and **9** (SI Table 1 and Figure S2).

Several tetrazine–fluorophore conjugates, including **11**, **13**, **14**, and **16** (Figure 1 and Figure S4) are substantially quenched with respect to the free fluorophores.[7,18] We found that the reaction of BCN with **11**, **13**, **14**, and **16** leads to a 5–10-fold increase in fluorescence, suggesting that the formation of the pyridazine product efficiently relieves the fluorophore quenching (Figure 2c and Figure S5). These fluorogenic reactions with BCN, like those between strained alkenes and tetrazines,[7,18] are advantageous for imaging experiments since they maximize the labeling signal while minimizing the fluorescence arising from the free tetrazine fluorophore.

Next, we aimed to design, synthesize, and genetically encode amino acids bearing BCN, TCO, and sTCO for site-specific protein labeling with a diverse range of probes both in vitro and in cells. The pyrrolysyl-tRNA synthetase (PylRS)/$tRNA_{CUA}$ pairs from *Methanosarcina* species, including *M. barkeri* (*Mb*) and *M. mazei* (*Mm*), and their evolved derivatives have been used to direct the site-specific incorporation of a growing list of structurally diverse unnatural amino acids in response to the amber codon[19−26] in a range of hosts, allowing synthetases evolved in *E. coli* to be used for genetic code expansion in a growing list of cells and organisms, including *E. coli*, *Salmonella typhimurium*, yeast, human cells, and *Caenorhabditis elegans*.[7,27−31] We designed the unnatural amino acids **1**, **2**, and **3** (Figure 1) with the goal of incorporating them into proteins using the PylRS/$tRNA_{CUA}$ pair or an evolved derivative. The amino acids were synthesized as described in the SI.

We screened the *Mb*PylRS/$tRNA_{CUA}$ pair along with a panel of *Mb*PylRS mutants for their ability to direct the incorporation of **1**–**3** in response to an amber codon introduced at position 150 in a C-terminally His_6-tagged superfolder green fluorescent protein (sfGFP). Cells containing a mutant of *Mb*PylRS with the three amino acid substitutions Y271M, L274G, and C313A[32] in the enzyme active site [which we named BCN-tRNA synthetase (BCNRS)] and a plasmid that encodes *Mb*$tRNA_{CUA}$ and sfGFP-His_6 with an amber codon at position 150 (*psfGFP150TAGPylT*-His_6) led to the amino acid-dependent synthesis of full-length sfGFP-His_6 (Figure 3a and Figure S6). We found an additional synthetase mutant bearing the mutations Y271A, L274 M, and C313A,[32] which we named TCO-tRNA synthetase (TCORS). The TCORS/$tRNA_{CUA}$ pair led to the amino acid-dependent synthesis of sfGFP from *psfGFP150TAGPylT*-His_6 in the presence of **2**. Finally, we found that both the BCNRS/$tRNA_{CUA}$ pair and the TCORS/$tRNA_{CUA}$ pair lead to the amino acid-dependent synthesis of sfGFP from *psfGFP150TAGPylT*-His_6 in the presence of **3**. For each amino acid, sfGFP was isolated in good yield after His-tag and gel-filtration purification (6–12 mg/L of culture; Figure 3b). This is comparable to the yields obtained for other well-incorporated unnatural amino acids, including **5**. Electrospray ionization mass spectrometry (ESI-MS) data for sfGFP produced from *psfGFP150TAGPylT*-His_6 in the presence of each unnatural amino acid were consistent with their site-specific incorporation (Figure 3c–e).

To demonstrate that the tetrazine–dye probes react efficiently and specifically with recombinant proteins bearing site-specifically incorporated **1**, we labeled purified sfGFP-**1**-His_6 with 10 equiv of tetrazine–fluorophore conjugate **11** for 1 h at room temperature. SDS-PAGE and ESI-MS analyses confirmed the quantitative labeling of sfGFP-**1** (Figure 4a,b). Control experiments demonstrated that sfGFP-**4** was labeled

dx.doi.org/10.1021/ja302832g | J. Am. Chem. Soc. 2012, 134, 10317−10320

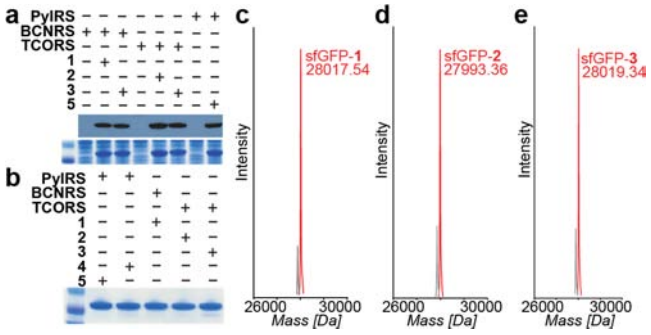


Figure 3. Efficient genetically encoded incorporation of unnatural amino acids in *E. coli*. (a) Amino acid-dependent overexpression of sfGFP-His_6 bearing an amber codon at position 150. The expressed protein was detected in lysates using an anti-His_6 antibody and Coomassie staining. (b) Coomassie-stained gel showing purified proteins. (c–e) ESI-MS data for amino acid incorporation. For sfGFP-**1**-His_6: found, 28017.54 Da; calcd, 28017.62 Da. For sfGFP-**2**-His_6: found, 27993.36 Da; calcd, 27992.82 Da. For sfGFP-His_6 produced with **3** as described: found, 28019.34 Da; calcd, 28019.63 Da. The minor peaks in the mass spectra correspond to loss of the N-terminal methionine.

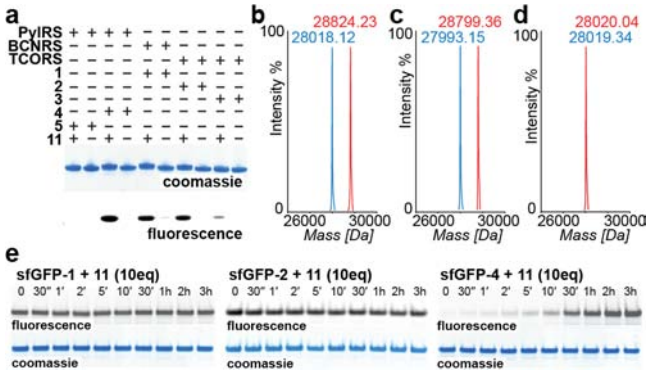


Figure 4. Rapid and specific labeling of recombinant proteins with tetrazine–fluorophore conjugates. (a) Specific labeling of sfGFP bearing **1**, **2**, or **4** with **11** (10 equiv) demonstrated by SDS-PAGE and in-gel fluorescence. (b) Quantitative labeling of sfGFP-**1** with **11** demonstrated by ESI-MS. Before bioconjugation (blue): found, 28018.1 Da; calcd, 28017.6 Da. After bioconjugation (red): found, 28824.2 Da; calcd, 28823.2 Da. (c) Quantitative labeling of sfGFP-**2** with **11** demonstrated by ESI-MS. Before bioconjugation (blue): found, 27993.2 Da; calcd, 27992.8 Da. After bioconjugation (red): found, 28799.4 Da; calcd, 28799.1 Da. (d) No labeling with **11** of sfGFP-His_6 expressed in the presence of **3** could be detected by MS. (e) Very rapid labeling of proteins containing **1** or **2**.

under the same conditions and that no nonspecific labeling occurred with sfGFP-**5**. ESI-MS showed that sfGFP-**1** could be efficiently and specifically derivatized with **6**–**9** (Figure S7) and with **12**–**14** and **16** (Figure S8). We also demonstrated that purified sfGFP-**2** could be quantitatively labeled with **11** (Figure 4a,c). Interestingly, we observed only very weak labeling of sfGFP-His_6 purified from cells expressing TCORS/$tRNA_{CUA}$ and *psfGFP150TAGPylT*-His_6 and grown in the presence of **3** (Figure 4a,d) and substoichiometric labeling of this protein prior to purification (Figure S9). Since the sfGFP expressed in the presence of **3** has a mass corresponding to the incorporation of **3**, these observations are consistent with the in vivo conversion of a fraction of the *trans*-alkene in **3** to its unreactive *cis* isomer. This isomerization is known to occur in the presence of thiols.[4]

To demonstrate that the reaction between BCN and tetrazines is highly selective within a cellular context, we performed the reaction on *E. coli* expressing sfGFP-**1** (Figure S10). In-gel fluorescence imaging of this proteome labeling demonstrated that while many proteins in the lysates were present at a comparable abundance to sfGFP-**1**, there was very little background labeling. This suggests that the reaction is specific with respect to the *E. coli* proteome.

To investigate whether the rates of the BCN– and TCO–tetrazine cycloadditions observed for small molecules translate into exceptionally rapid protein labeling, we compared the labeling of purified sfGFP bearing **1**, **2**, or **4** with 10 equiv of **11**. In-gel fluorescence imaging of the labeling reaction as a function of time (Figure 4e) indicated that the reaction of sfGFP-**4** reached completion in ~1 h. In contrast, the labeling of sfGFP-**1** and sfGFP-**2** was complete within the few seconds it took to measure the first time point, demonstrating that the rate acceleration of the BCN– and TCO–tetrazine reactions translates into much more rapid protein labeling.

To demonstrate the incorporation of amino acids **1** and **2** into mammalian cells, we transplanted the mutations allowing the incorporation of **1** or **2** into a mammalian-optimized *Mb*PylRS. Western blots showed that both **1** and **2** can be genetically encoded with high efficiency into proteins in mammalian cells using the BCNRS/$tRNA_{CUA}$ and TCORS/$tRNA_{CUA}$ pairs, respectively (Figure 5a).

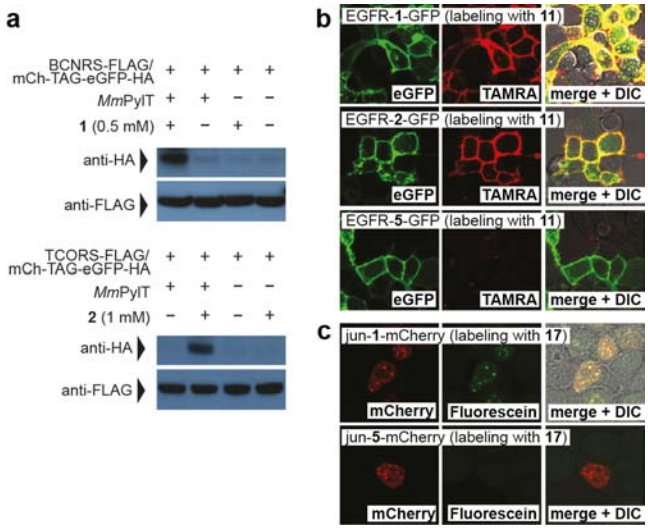


Figure 5. Site-specific incorporation of **1** and **2** into proteins in mammalian cells and their rapid and specific labeling with tetrazine fluorophores. (a) Western blots demonstrate that the expression of full-length mCherry(TAG)eGFP-HA is dependent on the presence of **1** or **2** and $tRNA_{CUA}$. BCNRS and TCORS were FLAG-tagged. (b) Specific and ultrarapid labeling of a cell-surface protein in live mammalian cells. Left: EGFR-GFP bearing **1**, **2**, or **5** at position 128 is visible as green fluorescence at the membrane of transfected cells. Middle: treatment of cells with **11** (400 nM) selectively labels EGFR containing **1** or **2**. Right: merged green and red fluorescence images with differential interference contrast (DIC). Cells were imaged 2 min after the addition of **11**. (c) Specific and rapid labeling of a nuclear protein in live mammalian cells. Left: jun-**1**-mCherry and jun-**5**-mCherry are visible as red fluorescence in the nuclei of transfected cells. Middle: Selective labeling of jun-**1**-mCherry with **17** (200 nM). Right: merged red and green fluorescence with DIC. No labeling was observed for cells bearing jun-**5**-mCherry.

dx.doi.org/10.1021/ja302832g | *J. Am. Chem. Soc.* 2012, 134, 10317–10320

To investigate whether the rapid BCN−tetrazine ligation would provide advantages for site-specific labeling of proteins on mammalian cells, we expressed an epidermal growth factor receptor (EGFR)−GFP fusion bearing an amber codon at position 128 (*EGFR(128TAG)GFP*) in HEK-293 cells containing the BCNRS/$tRNA_{CUA}$ pair cultured in the presence of **1** (0.5 mM). Full-length EGFR-**1**-GFP was produced in the presence of **1**, resulting in bright green fluorescence at the cell membrane. To label **1** with tetrazine−fluorophore conjugates, we incubated cells with **11** (400 nM), changed the medium, and imaged the red fluorescence arising from TAMRA labeling. The TAMRA fluorescence colocalized nicely with the cell-surface EGFR-GFP fluorescence. Clear labeling of cells bearing EGFR-**1**-GFP was observed within 2 min, the first time point we could measure; additional time points demonstrated that the labeling was saturated within 2 min (Figure 5b and Figures S11−S14); similar results were obtained with **12**. Incorporation of **2** into the EGFR−GFP fusion led to similarly rapid and efficient labeling with **11** (Figure 5b and Figures S15 and S16). In contrast, it took 2 h before we observed any specific labeling of cells bearing EGFR-**4**-GFP under identical conditions (Figure S14).[7] In control experiments we observed neither labeling of cells bearing EGFR-**5**-GFP nor nonspecific labeling of cells that did not express EGFR-GFP. We observed weak but measurable labeling of EGFR-GFP expressed in HEK-293 cells from *EGFR(128TAG)GFP* in the presence of the BCNRS/$tRNA_{CUA}$ pair and **3** (Figure S17). These observations are consistent with the isomerization of a fraction of **3** in mammalian cells and with our observations in *E. coli*.

To demonstrate the rapid labeling of an intracellular protein in mammalian cells, we expressed a transcription factor, jun, with a C-terminal mCherry fusion from a gene bearing an amber codon in the linker between JunB (jun) and mCherry. In the presence of amino acid **1** and the BCNRS/$tRNA_{CUA}$ pair, the jun-**1**-mCherry protein was produced in HEK cells and, as expected, localized in the nuclei of cells (Figure 5c and Figure S18). Labeling with cell-permeable conjugate **17** (200 nM) resulted in green fluorescence that colocalized nicely with the mCherry signal at the first time point analyzed (after 15 min of labeling and 90 min of washing). No specific labeling was observed in nontransfected cells in the same sample or in control cells expressing jun-**5**-mCherry, further confirming the specificity of intracellular labeling.

In conclusion, we have reported the exceptionally rapid fluorogenic reaction of BCN with a range of tetrazines under aqueous conditions at room temperature. The reaction between a strained alkene (e.g., a norbornene or *trans*-cyclooctene) and a tetrazine may lead to a mixture of diastereomers and regioisomers as well as isomers from dihydropyridazine isomerization.[3,4] In contrast, the BCN−tetrazine reaction leads to the formation of a single product. This will be an advantage in applications where homogeneity in the orientation of probe attachment may be important, including single-molecule spectroscopy, super-resolution microscopy, and FRET approaches.

We have demonstrated the efficient site-specific incorporation of **1** and **2** into proteins in *E. coli* and mammalian cells and their efficient, specific, fluorogenic, and rapid labeling with tetrazine fluorophore conjugates. While we have demonstrated the advantages of this approach in vitro, in *E. coli*, and in live mammalian cells, the ability to incorporate unnatural amino acids in *C. elegans* using the PylRS/$tRNA_{CUA}$ pair[29] suggests that it may be possible to extend the labeling approach described here to site-specific protein labeling in animals.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Complete references and supplementary results and methods. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

chin@mrc-lmb.cam.ac.uk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the MRC (Grants U105181009 and UD99999908) and the NIH (Grant P20 RR017716 from the NCRR COBRE Program). We thank Han Zhang, Fiona Townsley, and Trevor Rutherford for experimental assistance.

■ REFERENCES

(1) Devaraj, N. K.; et al. *Bioconjugate Chem.* **2008**, *19*, 2297.
(2) Devaraj, N. K.; Weissleder, R. *Acc. Chem. Res.* **2011**, *44*, 816.
(3) Blackman, M. L.; et al. *J. Am. Chem. Soc.* **2008**, *130*, 13518.
(4) Taylor, M. T.; et al. *J. Am. Chem. Soc.* **2011**, *133*, 9646.
(5) Liu, D. S.; et al. *J. Am. Chem. Soc.* **2012**, *134*, 792.
(6) Seitchik, J. L.; et al. *J. Am. Chem. Soc.* **2012**, *134*, 2898.
(7) Lang, K.; et al. *Nat. Chem.* **2012**, *4*, 298.
(8) Kaya, E.; et al. *Angew. Chem., Int. Ed.* **2012**, *51*, 4466.
(9) Plass, T.; et al. *Angew. Chem., Int. Ed.* **2012**, *51*, 4166.
(10) Dommerholt, J.; et al. *Angew. Chem., Int. Ed.* **2010**, *49*, 9422.
(11) Chen, W. X.; et al. *Chem. Commun.* **2012**, *48*, 1736.
(12) McKay, C. S.; et al. *Chem. Commun.* **2011**, *47*, 10040.
(13) McKay, C. S.; et al. *Org. Biomol. Chem.* **2012**, *10*, 3066.
(14) Ning, X.; et al. *Angew. Chem., Int. Ed.* **2010**, *49*, 3065.
(15) Agard, N. J.; et al. *J. Am. Chem. Soc.* **2004**, *126*, 15046.
(16) Sletten, E. M.; Bertozzi, C. R. *Acc. Chem. Res.* **2011**, *44*, 666.
(17) Karver, M. R.; et al. *Angew. Chem., Int. Ed.* **2012**, *51*, 920.
(18) Devaraj, N. K.; et al. *Angew. Chem., Int. Ed.* **2010**, *49*, 2869.
(19) Fekner, T.; et al. *Angew. Chem., Int. Ed.* **2009**, *48*, 1633.
(20) Nguyen, D. P.; et al. *J. Am. Chem. Soc.* **2009**, *131*, 14194.
(21) Nguyen, D. P.; et al. *J. Am. Chem. Soc.* **2009**, *131*, 8720.
(22) Nguyen, D. P.; et al. *J. Am. Chem. Soc.* **2011**, *133*, 11418.
(23) Neumann, H.; et al. *Nat. Chem. Biol.* **2008**, *4*, 232.
(24) Polycarpo, C. R.; et al. *FEBS Lett.* **2006**, *580*, 6695.
(25) Li, X.; et al. *Angew. Chem., Int. Ed.* **2009**, *48*, 9184.
(26) Wang, Y. S.; et al. *J. Am. Chem. Soc.* **2012**, *134*, 2950.
(27) Mukai, T.; et al. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818.
(28) Hancock, S. M.; et al. *J. Am. Chem. Soc.* **2010**, *132*, 14819.
(29) Greiss, S.; Chin, J. W. *J. Am. Chem. Soc.* **2011**, *133*, 14196.
(30) Lin, S. X.; et al. *J. Am. Chem. Soc.* **2011**, *133*, 20581.
(31) Gautier, A.; et al. *J. Am. Chem. Soc.* **2010**, *132*, 4086.
(32) Virdee, S.; et al. *J. Am. Chem. Soc.* **2011**, *133*, 10708.