Protein–RNA Interactions

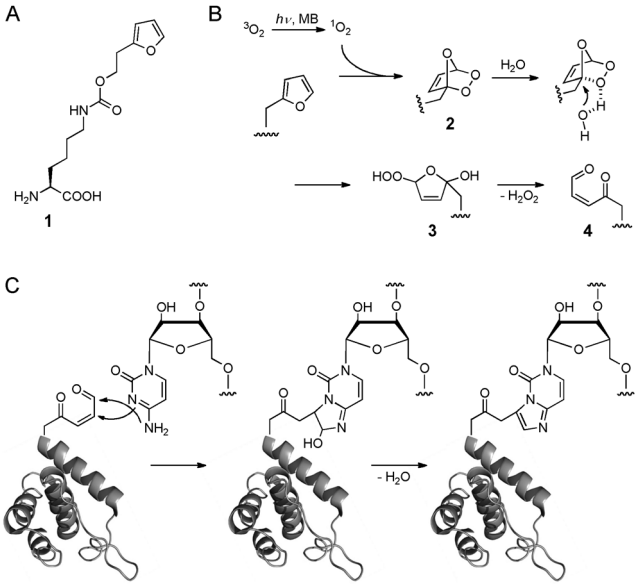
DOI: 10.1002/anie.201300754

Red-Light-Controlled Protein–RNA Crosslinking with a Genetically Encoded Furan**

*Moritz J. Schmidt and Daniel Summerer**

Protein–RNA interactions are extensively involved in the transcription, maturation, transport, translation, and decay of RNA and thus key to the regulation of gene expression. The ability to selectively induce the formation of a covalent crosslink between a peptide or protein of interest and its RNA binding partner(s) thus has therapeutic implications and enables the discovery and mapping of weak and/or transient protein–RNA interactions. An ideal crosslinking reagent for that purpose would be red light. This approach would offer high penetration depth in biological samples (e.g. required for photodynamic therapy)[1] and spatiotemporal reaction control. However, this approach also requires the site-specific installing of suited functional groups with unique, photocontrollable reactivity into the chosen peptide or protein. An especially useful strategy for that has been the genetic encoding of noncanonical amino acids (ncAA) by means of orthogonal pairs of tRNAs and aminoacyl-tRNA-synthetases (aaRS) as introduced by Schultz.[2] This method allows novel functional groups to be incorporated into a protein of interest directly in vivo, in response to a unique codon. Recently, photo-crosslinking ncAA based on arylazides,[3] diazirines,[4] and benzophenone[5] have been genetically encoded and employed to study biomolecular interactions with an emphasis on protein–protein interactions.[4d,6] However, these ncAA require irradiation with UV light, which restricts applications in complex cellular samples and tissues owing to its poor penetration depth[1] and can induce undesired nucleic acid damaging photoreactions.[7]

We aimed at the genetic encoding of RNA-crosslinking chemistry that can be controlled by irradiation with red light. We designed the ncAA *N*ε-[2-(furan-2-yl)ethoxy]carbonyl lysine which has a furan moiety as a precursor of the crosslinking group (**1**, Scheme 1 A). In contrast to short and rigid phenylalanine-based photo-crosslinking ncAAs,[4a,5] the long, flexible linker of **1** should enable wide-range reactivity.[4b–d]



Scheme 1. A) Structure of *N*ε-[2-(furan-2-yl)ethoxy]carbonyl lysine (**1**). B) Proposed mechanism of 1,4-enedione (**4**) formation from a furan moiety in an aqueous environment through oxidation by singlet oxygen. MB = methylene blue. C) Protein–RNA crosslinking by reaction of 1,4-enedione **4** with cytidine.

Furans can be oxidized to reactive intermediates by singlet oxygen ($^1O_2$), generated in situ by the light-induced excitation of photosensitizers.[8] After a [4+2] cycloaddition between $^1O_2$ and the furan (Scheme 1 B), the resulting ozonide **2** can be opened by hydroxy nucleophiles to yield hydroperoxides.

It has been proposed that in an aqueous environment, a water molecule serves as a nucleophile leading to the formation of **3** that affords the electrophilic 1,4-enedione **4** after elimination of a H_2O_2 molecule.[9] 1,4-enediones spontaneously form adducts with cytosine, adenine, and guanine nucleobases under physiological conditions through reaction of the exocyclic amino groups with the terminal aldehyde, 1,4-addition of the N3 or N1 atom (in cytosine and purines, respectively), and potential subsequent aromatization (Scheme 1 C, shown for cytidine).[10] Recently, this reaction sequence has been exploited for the formation of DNA interstrand crosslinks by incorporating monosubstituted furans into DNA oligonucleotides using solid-phase synthesis. High levels of crosslinking were observed when furans were incorporated opposite dC and dA nucleotides and oxidized with *N*-bromosuccinimide (NBS)[10d,11] or $^1O_2$ generated in situ by irradiation of methylene blue (MB) with red light.[12]

[*] M. Sc. M. J. Schmidt, Dr. D. Summerer
Department of Chemistry and Konstanz Research School Chemical Biology, University of Konstanz
Universitätsstrasse 10, 78457 Konstanz (Germany)
E-mail: daniel.summerer@uni-konstanz.de
Homepage: http://cms.uni-konstanz.de/summerer/

[**] We acknowledge support by the Zukunftskolleg of the University of Konstanz and the Konstanz Research School Chemical Biology. We thank Peter G. Schultz for plasmid pREP-CM3 + RBS. We thank Lei Wang for *E. coli* strain JX33. We thank Edward Lemke for plasmids pEVOL_PylRS_wt, _AF and pBAD_GFP-39TAG. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SPP1623).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201300754.

We analyzed the $tRNA^{Pyl}$/pyrrolysyl-RS (PylRS) pair of *Methanosarcina mazei* for its capability to direct the selective incorporation of **1** into proteins in *E. coli* in response to the amber codon, TAG.[13] However, *E. coli* expressing the wild-type (wt) pair (encoded by pEVOL-PylRS_wt) and mRNA encoding green fluorescent protein (GFP) with an amber codon at position Y39 and a C-terminal His_6-tag (encoded by pBAD_GFP-39TAG)[14] exhibited significant expression of full-length GFP protein only in the presence of the positive control ncAA *N*ε-Boc-L-lysine (Boc-Lys).[13e] In presence of **1**, no expression was detected by cellular fluorescence and SDS PAGE analysis (Figure 1 A).

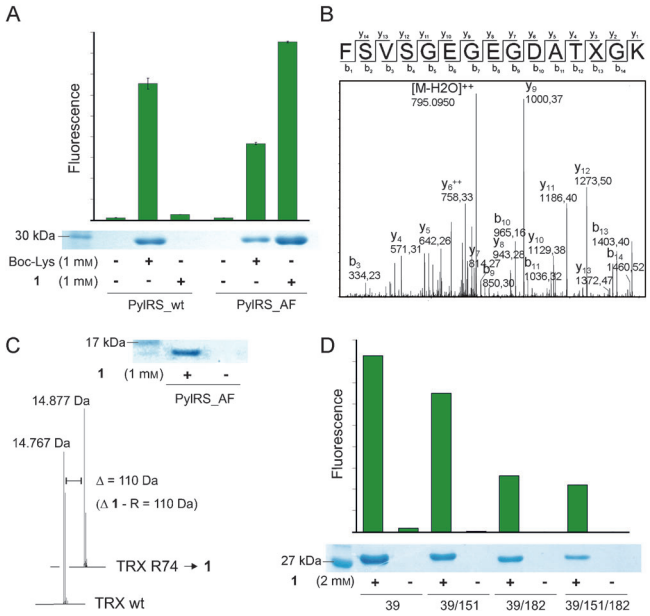


Figure 1. Efficiency and fidelity of incorporation of **1** into proteins in *E. coli*. A) Expression of GFP-Y39→**1** and GFP-Y39→Boc-Lys under co-expression of PylRS_wt or PylRS–AF. Top: Cellular GFP fluorescence. Bottom: SDS PAGE analysis of Ni-NTA purified GFP expressions. B) ESI MS/MS spectrum of GFP-Y39→**1** tryptic peptide fragment. **1** is marked as X. C) Top: SDS PAGE analysis of purified TRX-R74→**1** expression under co-expression of PylRS–AF. Bottom: ESI MS spectra of purified TRX-wt and TRX-R74→**1**. D) Incorporation of **1** into GFP at multiple positions as depicted, under co-expression of PylRS_AF in *E. coli* JX33 with analysis as in Figure 1 A.

To evolve a PylRS that accepts **1** as a substrate, we subjected three focused libraries with random mutations of positions in α-helix 6 of the pyrrolysine binding pocket[15] (Y306, Y306/L309 and Y306/A302) together with a fixed Y384F mutation[13e,14,16] to a selection as described elsewhere.[17] To maximize the response of cell growth to the aminoacylation activity of the PylRS mutant, selection was performed in the release factor 1 (RF1) knockout *E. coli* strain JX33 introduced by Wang.[18]

We identified two mutants, Y306A/L309M/Y384F and Y306A/Y384F (PylRS_AF) that promoted GFP-Y39→**1** expression with comparable efficiencies. For PylRS_AF, 17 mg L^{-1} of purified GFP-Y39→**1** were obtained, whereas no expression could be detected in the absence of **1**, indicating a high incorporation fidelity (Figure 1 A). Electrospray ionization tandem mass spectrometry (ESI MS/MS) of a trypsin-digested GFP-Y39→**1** sample confirmed the identity and site-selective incorporation of **1** (Figure 1 B, calculated: 803.8652 Da, found: 803.8650 Da).

We expressed thioredoxin (TRX) with an amber site at position R74 and a C-terminal His_6-tag (encoded by pBAD_TRX-His_6-R74TAG) in *E. coli* as above. SDS PAGE analysis again indicated high efficiency (yield: 14 mg L^{-1} purified TRX-R74→**1**) and fidelity of incorporation (Figure 1 C, top). ESI MS analysis confirmed the identity and high fidelity incorporation of **1** (Figure 1 C, bottom).

The ability to introduce **1** at multiple sites of a protein would allow the likelihood of crosslinking to increase and enable multiple-crosslinking in higher-order complexes. However, this multiple incorporation is usually restricted by the limited efficiency of amber suppression owing to competition between $tRNA_{CUA}$ and RF1.[18,19]

We expressed GFP bearing one to three amber codons (encoded by pBAD_GFP-39TAG, -39/151TAG, -39/182TAG, and -39/151/182TAG) in *E. coli* JX33 as above. In all cases, significant GFP expression was observed only in presence of **1**, indicating incorporation at all sites with high fidelity (Figure 1 D). Though efficiency was site-dependent (ca. 80 % for Y151 and ca. 32 % for Y182), over 4.5 mg L^{-1} of purified protein was obtained even for the triple incorporation experiment.

One of the most frequently occurring secondary structures of RNA is the hairpin fold. Selective recognition of these structures by regulatory proteins is critical for transcription, RNA transport, and translation. Consequently, they are promising targets for therapeutic approaches.[20] A central drug target and classical example for viral protein–RNA hairpin complexes has been the interaction between the HIV-1 trans-activator of transcription (TAT) protein and its cognate trans-activation response RNA (TAR, Figure 2 A), owing to its ubiquitous, essential role for transcription of the

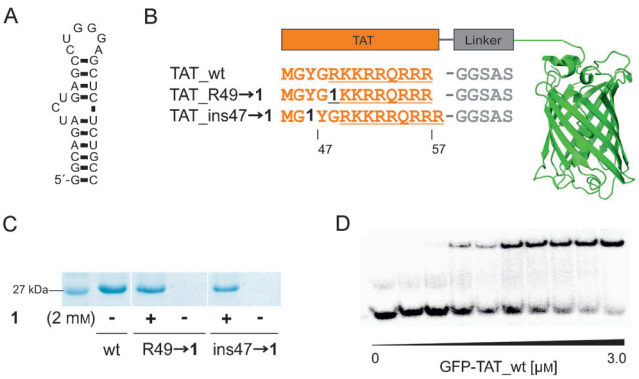


Figure 2. Incorporation of **1** into the arginine-rich motif (ARM) of HIV-1 TAT and binding to HIV-1 TAR RNA. A) Nucleotide sequence of employed HIV-1 TAR RNA. B) Overall design and amino acid sequences of employed TAT peptide constructs. TAT sequence is shown in orange with ARM underlined, linker in gray and GFP in green. C) SDS-PAGE analysis of purified TAT-wt, TAT-ins47→**1**, and TAT-R49→**1** expressions. D) Electrophoretic mobility shift assay of binding of GFP-TAT-wt to HIV-1 TAR RNA.

HIV genome. However, HIV-1 TAR exhibits high conformational flexibility, which complicates inhibitor design and has restricted structural insights. Nevertheless, cationic peptides, such as TAT-derived arginine-rich motifs (ARM), have recently emerged as promising class of TAR-inhibitors.[21] The ability to introduce **1** into ARM could thus lead to insights into TAR recognition that are otherwise difficult to access and provide chemical "warheads" for the design of peptide therapeutics.

We expressed two GFP fusion constructs bearing an N-terminal HIV-1 TAT ARM with or without **1** replacing R49, which is part of the ARM, but not critical for binding (Figure 2B, encoded by pBAD_GFP-TAT_wt or _R49TAG). SDS-PAGE analysis (Figure 2C) and ESI MS indicated incorporation of **1** with high efficiency and fidelity (calculated: 29544.9 and 29654.9 Da, found: 29542.0 and 29653.0 Da). Both constructs were able to bind TAR, as indicated by electromobility shift assays (Figure 2D; $K_D = (0.47 \pm 0.05)$ μM and (1.32 ± 0.17) μM for TAT_wt and _R49→**1**, respectively). We incubated 5′-^{32}P-labelled TAR with or without these constructs in the presence or absence of the photosensitizer MB and irradiated the mixtures with red light under aerobic conditions. Denaturing PAGE analysis revealed no crosslink formation in the absence of MB or light (Supporting Information). Similarly, reactions performed without the TAT construct or with TAT_wt did not show bands differing from the TAR band (Figure 3A). In contrast,

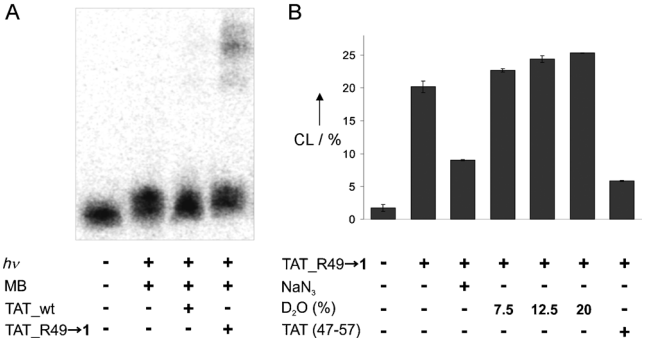


Figure 3. Red-light-controlled TAT–TAR crosslinking. A) Denaturing PAGE analysis of crosslinking reactions with TAT_wt and TAT_R49→**1**. TAR RNA was incubated with MB (8 equivalents of TAT construct) and with or without TAT constructs (2 μM) and irradiated for 2 min with red light. B) Yields of TAR/TAT_R49→**1** crosslink (CL) products observed in Figure 3A in absence or presence of 4 mM sodium azide (NaN_3), with different contents of D_2O or 6 μM of a synthetic TAT peptide as a competitor. Samples were treated as in Figure 3A.

for reactions performed with TAT_R49→**1**, a band with low electromobility was observed, indicating the formation of crosslink product (Figure 3A). We analyzed if the observed crosslink was the result of an $^1O_2$-dependent reaction. Addition of sodium azide, a quencher of $^1O_2$,[22] resulted in the decrease of the crosslink yield from 20 to 9%. Conversely, addition of the $^1O_2$-stabilizer D_2O[23] led to an increase of yields in a concentration-dependent manner, confirming that the observed crosslink formation was a result of the reaction of **1**, $^1O_2$, and TAR (Figure 3B).

The discovery of RNA binding partners and the mapping of protein–RNA interaction surfaces using our new approach requires the dependence of crosslinking on both complex formation and the spatial orientation of **1**. Crosslinking reactions with TAT_R49→**1** as above in the presence of a TAT competitor peptide (amino acids 47–57 from TAT) led to reduced yields (6%), indicating a dependence on complex formation. Additionally, we expressed a third TAT construct with **1** inserted at the N-terminus of the ARM (Figure 2B and C, TAT_ins47→**1**, ESI-MS: calculated: 29811.0, found: 29809.0 Da, $K_D = (1.34 \pm 0.41)$ μM). Crosslink yield with TAT_ins47→**1** was increased 1.5-fold compared to TAT_R49→**1**, indicating that **1** was in the proximity of TAR and that its crosslink reactivity was indeed altered by a difference in spatial orientation compared to TAT_R49→**1** (see the Supporting Information).

In conclusion, we present the first genetically encoded crosslinking chemistry that can be controlled with red light. The strategy offers high tissue penetration depths, is sensitive to protein–RNA complex formation and to the spatial orientation of **1** within a complex, which makes it applicable for the discovery and mapping of transient protein–RNA interactions with spatiotemporal resolution. Both the applied wavelengths and photosensitizer are used in photodynamic therapy and introduction of **1** is flexible in terms of the target peptide or protein and the position and number of incorporation sites. This represents a promising basis for the development of RNA-targeted biologics with photoactivatable units for photodynamic therapy.

Received: January 28, 2013
Published online: March 19, 2013

Keywords: crosslinking reactions · noncanonical amino acids · photodynamic therapy · pyrrolysyl tRNA synthetase · RNA recognition

[1] J. F. Lovell, T. W. B. Liu, J. Chen, G. Zheng, *Chem. Rev.* **2010**, *110*, 2839–2857.
[2] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.
[3] a) J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027; b) A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782–11783.
[4] a) E. M. Tippmann, W. Liu, D. Summerer, A. V. Mack, P. G. Schultz, *ChemBioChem* **2007**, *8*, 2210–2214; b) C. J. Chou, R. Uprety, L. Davis, J. W. Chin, A. Deiters, *Chem. Sci.* **2011**, *2*, 480–483; c) T. Yanagisawa, N. Hino, F. Iraha, T. Mukai, K. Sakamoto, S. Yokoyama, *Mol. Biosyst.* **2012**, *8*, 1131–1135; d) M. Zhang, S. Lin, X. Song, J. Liu, Y. Fu, X. Ge, X. Fu, Z. Chang, P. R. Chen, *Nat. Chem. Biol.* **2011**, *7*, 671–677.
[5] a) J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020–11024; b) J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science* **2003**, *301*, 964–967.
[6] a) C. Liu, A. L. Young, A. Starling-Windhof, A. Bracher, S. Saschenbrecker, B. V. Rao, K. V. Rao, O. Berninghausen, T. Mielke, F. U. Hartl, R. Beckmann, M. Hayer-Hartl, *Nature* **2010**, *463*, 197–202; b) S. Tagami, S. Sekine, T. Kumarevel, N. Hino, Y. Murayama, S. Kamegamori, M. Yamamoto, K. Sakamoto, S. Yokoyama, *Nature* **2010**, *468*, 978–982; c) P. Carvalho, A. M.

Stanley, T. A. Rapoport, *Cell* **2010**, *143*, 579 – 591; d) I. S. Farrell, R. Toroney, J. L. Hazen, R. A. Mehl, J. W. Chin, *Nat. Methods* **2005**, *2*, 377 – 384; e) H. S. Lee, R. D. Dimla, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5222 – 5224.

[7] R. P. Sinha, D. P. Hader, *Photochem. Photobiol. Sci.* **2002**, *1*, 225 – 236.

[8] a) B. L. Feringa, *Recl. Trav. Chim. Pays-Bas* **1987**, *106*, 469 – 488; b) K. Gollnick, A. Griesbeck, *Tetrahedron* **1985**, *41*, 2057 – 2068; c) T. Montagnon, M. Tofi, G. Vassilikogiannakis, *Acc. Chem. Res.* **2008**, *41*, 1001 – 1011.

[9] D. Noutsias, I. Alexopoulou, T. Montagnon, G. Vassilikogiannakis, *Green Chem.* **2012**, *14*, 601 – 604.

[10] a) L. A. Peterson, *Drug Metab. Rev.* **2006**, *38*, 615 – 626; b) M. C. Byrns, C. C. Vu, L. A. Peterson, *Chem. Res. Toxicol.* **2004**, *17*, 1607 – 1613; c) L. Gingipalli, P. C. Dedon, *J. Am. Chem. Soc.* **2001**, *123*, 2664 – 2665; d) M. Op de Beeck, A. Madder, *J. Am. Chem. Soc.* **2011**, *133*, 796 – 807.

[11] a) A. M. Jawalekar, M. Op de Beeck, F. L. van Delft, A. Madder, *Chem. Commun.* **2011**, *47*, 2796 – 2798; b) K. Stevens, A. Madder, *Nucleic Acids Res.* **2009**, *37*, 1555 – 1565.

[12] M. Op de Beeck, A. Madder, *J. Am. Chem. Soc.* **2012**, *134*, 10737 – 10740.

[13] a) W. Wan, Y. Huang, Z. Y. Wang, W. K. Russell, P. J. Pai, D. H. Russell, W. R. Liu, *Angew. Chem.* **2010**, *122*, 3279 – 3282; *Angew. Chem. Int. Ed.* **2010**, *49*, 3211 – 3214; b) A. Gautier, D. P. Nguyen, H. Lusic, W. A. An, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2010**, *132*, 4086 – 4088; c) T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem.* **2009**, *121*, 1661 – 1663; *Angew. Chem. Int. Ed.* **2009**, *48*, 1633 – 1635; d) P. R. Chen, D. Groff, J. T. Guo, W. J. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112 – 4115; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052 – 4055; e) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187 – 1197; f) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232 – 234; g) C. R. Polycarpo, S. Herring, A. Berube, J. L. Wood, D. Soll, A. Ambrogelly, *FEBS Lett.* **2006**, *580*, 6695 – 6700.

[14] T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew. Chem.* **2011**, *123*, 3964 – 3967; *Angew. Chem. Int. Ed.* **2011**, *50*, 3878 – 3881.

[15] a) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2008**, *378*, 634 – 652; b) J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Soll, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11268 – 11273.

[16] T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew. Chem.* **2012**, *124*, 4242 – 4246; *Angew. Chem. Int. Ed.* **2012**, *51*, 4166 – 4170.

[17] C. E. Melançon III, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3845 – 3847.

[18] a) D. B. Johnson, C. Wang, J. Xu, M. D. Schultz, R. J. Schmitz, J. R. Ecker, L. Wang, *ACS Chem. Biol.* **2012**, *7*, 1337 – 1344; b) D. B. F. Johnson, J. F. Xu, Z. X. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs, L. Wang, *Nat. Chem. Biol.* **2011**, *7*, 779 – 786.

[19] a) T. Mukai, A. Hayashi, F. Iraha, A. Sato, K. Ohtake, S. Yokoyama, K. Sakamoto, *Nucleic Acids Res.* **2010**, *38*, 8188 – 8195; b) Y. Huang, W. K. Russell, W. Wan, P. J. Pai, D. H. Russell, W. Liu, *Mol. Biosyst.* **2010**, *6*, 683 – 686.

[20] a) J. R. Thomas, P. J. Hergenrother, *Chem. Rev.* **2008**, *108*, 1171 – 1224; b) L. R. Guan, M. D. Disney, *ACS Chem. Biol.* **2012**, *7*, 73 – 86.

[21] a) A. Davidson, T. C. Leeper, Z. Athanassiou, K. Patora-Komisarska, J. Karn, J. A. Robinson, G. Varani, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11931 – 11936; b) S. J. Lee, S. Hyun, J. S. Kieft, J. Yu, *J. Am. Chem. Soc.* **2009**, *131*, 2224 – 2230; c) J. Pai, T. Yoon, N. D. Kim, I. S. Lee, J. Yu, I. Shin, *J. Am. Chem. Soc.* **2012**, *134*, 19287 – 19296.

[22] M. Y. Li, C. S. Cline, E. B. Koker, H. H. Carmichael, C. F. Chignell, P. Bilski, *Photochem. Photobiol.* **2001**, *74*, 760 – 764.

[23] C. Schweitzer, R. Schmidt, *Chem. Rev.* **2003**, *103*, 1685 – 1757.

