

ChemComm

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

COMMUNICATION

Site-specific fluorescence labelling of RNA using bio-orthogonal reaction of *trans*-cyclooctene and tetrazine.

Cite this: DOI: 10.1039/x0xx00000x

P. N. Asare-Okai,[‡] E. Agustin,[‡] D. Fabris and M. Royzen^a

Received 00th January 2012,

Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

This communication describes a general approach for site-specific fluorescence labelling of RNA using a cytidine triphosphate (CTP) analogue derivatized with a *trans*-cyclooctene group. The analogue was efficiently incorporated into a model RNA strand using *in vitro* transcription. Bio-orthogonal reaction with fluorescein-labelled tetrazine was utilized to fluorescently tag the synthetic RNA strand.

There is a strong interest in *in vivo* fluorescent imaging of RNA as a tool to observe and uncover the details of its cellular regulation and its dynamic behaviour inside the cell. Recent literature reports accentuate the significance of visualizing RNA's movement in real time.¹ Singer and co-authors described a GFP-fusion protein approach to visualize neuronal β -actin messenger RNA that allowed the investigation of molecular mechanisms underlying memory formation. These discoveries foster enthusiasm to tackle many yet unanswered questions regarding RNA's biology, such as its role in cell differentiation during embryogenesis as well as its regulation of gene expression.²

GFP-based imaging approaches inevitably have to prove that drastic changes in size and structure did not significantly alter RNA's properties, and the studies described above certainly achieved that by genetically engineering a healthy animal model. Yet, a small molecule fluorescence imaging approach would be advantageous due to minimal modification of the size and the native structure of RNA. An RNA nucleobase can be synthetically modified and thereby converted into a chromophore.³ Alternatively, fluorescence labelling of RNA can be split into two steps using bio-orthogonal chemistry. By this approach, nucleosides are modified with a small bio-orthogonal group that minimally perturbs the native structure and preserves enzymatic recognition. Upon incorporation into a strand, the RNA is labelled using fluorescently modified bio-orthogonal partner.

A number of bio-orthogonal reactions have been developed to date, yet only copper-catalysed [3+2] alkyne azide "click" cycloaddition reaction (CuAAC) has been utilized for fluorescent labelling of RNA. Uridine, as well as cytidine analogues, derivatized

with a terminal alkyne group, have been incorporated into RNA strands and labelled using fluorescent azides.⁴ Ethynyl uridine allows imaging of total cellular RNA in fixed cells and is sold as part of a kit.⁵ *In vivo* applications of CuAAC, however, are limited due to cytotoxicity of copper.

In addition to fluorescence labelling, a wide range of CuAAC-based synthetic tools have been exploited for covalent modification of oligonucleotides. CuAAC chemistry has been utilized to form DNA crosslinks that result in unnatural three-dimensional structures.⁶ Several drug delivery platforms for siRNA were developed using 'click' chemistry.⁷ Also, there are several commercially available reagents for internal, as well as terminal modification of DNA and RNA strands assembled via solid phase synthesis.⁸

Most recently, copper-free, strain-promoted azide-alkyne [3+2] cycloaddition reaction (SPAAC) has been described for covalent modification of oligonucleotides. SPAAC chemistry has been successfully applied for imaging of glycans in live cells and organisms.⁹ By the same token, SPAAC can be used for fluorescence imaging of RNA *in vivo*. So far it has mainly been applied for covalent modification of oligonucleotide strands formed by solid phase synthesis.¹⁰

Inverse Electron Demand Diels-Alder reaction (IEDDA) between *trans*-cyclooctenes (TCOs) and tetrazines does not require a metal catalyst, making it significantly less cytotoxic than CuAAC.¹¹ IEDDA has been utilized for live cell imaging of proteins,¹² glycans¹³ and fatty acids¹⁴. Presently, to the best of our knowledge, there is only a single report describing solid phase synthesis of DNA, fluorescently labelled using IEDDA reaction.¹⁵ IEDDA chemistry has not been applied to RNA. Herein we attempt to fill this void by describing a methodology for incorporating the bio-orthogonal *trans*-cyclooctene group into an RNA strand by *in vitro* transcription and subsequent fluorescent labelling.

The bio-orthogonal pair described in this work is shown in Fig. 1. The cytidine triphosphate (CTP) analogue, **1**, was derivatized with a TCO group at the 5-position of the nucleobase in a way to minimally alter the electrostatics of cytosine, which is important to preserve enzymatic recognition.¹⁶ Previous reports have shown that T7 RNA polymerase can tolerate substitutions at this position.¹⁷ The

IEDDA partner is a fluorescein-labelled tetrazine, **2**. Based on the previously characterized library of IEDDA reagents, we decided to focus on 4-methyl-substituted tetrazine that shows higher stability under physiological conditions.¹⁸

The synthesis of **1** is described in Fig 2A. It commenced with a commercially available 5-iodocytidine. The N-trifluoroacetyl-propargyl amine was coupled using Sonogashira conditions. The alkyne group was reduced using palladium – catalysed hydrogenation procedure. Compound **4** was afforded upon deprotection of trifluoroacetyl group using aqueous ammonia, and was subsequently coupled to TCO-NHS ester, **5**. Compound **6** was converted into a 5'-triphosphate by stepwise phosphorylation with POCl₃ and proton sponge, followed by tributylammonium pyrophosphate. Analytically pure **1** was obtained upon reverse-phase HPLC purification.

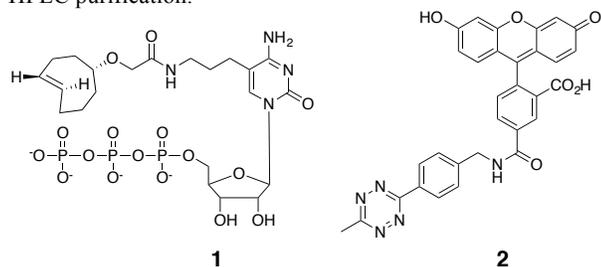


Fig. 1 The bio-orthogonal pair utilized herein for site-specific fluorescence labelling of RNA.

Synthesis of the fluorescein-labelled tetrazine is shown in Fig. 2B. The tetrazine **7** was prepared using previously described procedure.¹⁹ The Boc-group was subsequently deprotected using trifluoroacetic acid (TFA) and compound **8** was isolated as a TFA salt. Compound **2** was afforded by coupling **8** to the fluorescein-NHS ester. Spectroscopically pure **2** was achieved after purification by preparative thin layer chromatography.

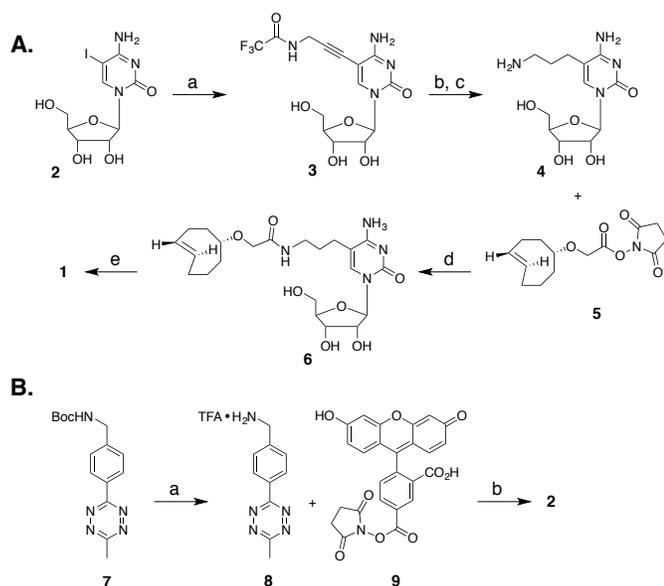


Fig. 2 (A) Synthesis of the CTP analogue, **1**. (a) TFA-propargyl amine, Pd(PPh₃)₄, CuI, Et₃N, DMF; (b) Pd(OH)₂, Et₃SiH, MeOH; (c) NH₄OH, H₂O; (d) Et₃N, DMF; (e) proton sponge, POCl₃, then Bu₃N, tributylammonium pyrophosphate, DMF. (B) Synthesis of the fluorescein-labelled tetrazine, **2**. (a) TFA, CH₂Cl₂; (b) Et₃N, DMF.

In vitro transcription (IVT) experiments were carried out to investigate the ability of T7 RNA polymerase to recognize the CTP

analogue and site-specifically incorporate **1** into a sequence of interest. These experiments were performed using *mirVana*TM miRNA Probe Construction Kit (Life Technologies) using DNA template that contains T7 polymerase promoter sequence linked to the sequence of miR-122.²⁰ The DNA template used for the transcription using the aforementioned kit has the sequence: 5'-ACAAACACCATTGTCACACTCCACCTGTCTC-3' and generates the RNA product: 5'-GGGAGACAGGUGGAGUGUGACAAUGGUGUUUGU-3' in which **1** is incorporated at the two highlighted positions.

The *in vitro* transcription was carried out using variable ratios of CTP and **1** and was correlated to the standard sequences, prepared by solid phase synthesis. Following transcription, the samples were treated with **2** (250 μM) to fluorescently label the RNA strands containing **1**. Subsequently, the samples were processed with 2 units of DNaseI to degrade the DNA template, denatured in 7M urea in the presence of 25 mM EDTA and resolved on a 20% denaturing polyacrylamide gel electrophoresis (PAGE). The PAGE results are shown in Fig. 3. Lane 1 represents a synthetic 26 nucleotide-long RNA sequence, modified with a fluorescein-based (FAM) dye at the 5' end, while lane 2 is a synthetic 33 nucleotide-long RNA sequence which does not carry any fluorescent label. Lanes 3 through 5 correspond to the IVT experiments carried out using 0, 50% and 100% **1**, respectively.

Fluorescence imaging of the gel, shown in Fig. 3A, illustrates the results of the post-transcriptional fluorescence labelling of the RNA strands. In addition to the synthetic FAM-labelled RNA in lane 1, fluorescent bands are observed in lanes 4 and 5. These bands correspond to the IVT products which were fluorescently labelled via bio-orthogonal IEDDA chemistry. Meanwhile, lane 2 (negative control), as well as lane 3, corresponding to the RNA lacking TCO, are fluorescently silent.

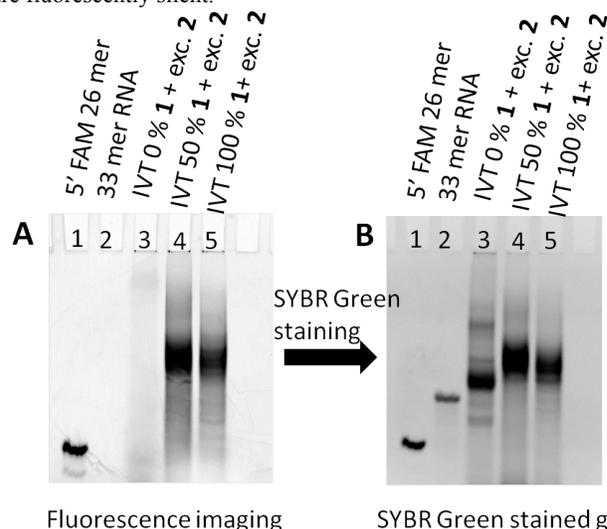


Fig. 3 Denaturing gel (20%) analysis of the *in vitro* transcription experiments. (A) fluorescence image of the unstained gel; (B) image of the same gel after SYBR Green staining. Lane 1 shows the 26 nt. long RNA sequence, carrying FAM dye (positive control), while lane 2 is a 33 nt. long RNA sequence prepared by solid phase synthesis (negative control); lanes 3, 4 and 5 show *in vitro* transcripts obtained using variable amounts of CTP and **1**. Lane 3: 100% CTP; lane 4: 50% CTP, 50% **1**; lane 5 100% **1**.

SYBR Green staining of the same gel to visualize all RNA sequences is shown in Fig. 3B. The observed bands in lanes 4 and 5, corresponding to fluorescein labelled RNA, have slower mobility than the unmodified strand in lane 3. This is consistent with the additional mass and negative charge introduced by the TCO-

chromophore conjugate. Lane 4 has multiple bands showing strong SYBR Green staining, which is due to the heterogeneity of the RNA transcripts that could contain either zero, one, or two fluorescein groups. Additional weakly staining bands, observed in Fig. 3B are likely due to aborted sequences and untemplated nucleotide additions at the 5' end by the T7 polymerase.²¹ Based on the relative pixel intensity, the transcription efficiency of the CTP analogue was calculated to be $60 \pm 5\%$ (Fig. S1).

In addition to PAGE, the IVT products were analysed by electrospray ionization mass spectrometry (ESI-MS). To facilitate the ESI-MS analysis, the RNA transcripts were subjected to RNase T1 treatment that cleaves 3' ends of guanine residues thereby forming two unique fragments: ACAGp and ACAAUGp. Fig. 4A shows the mass spectrum of the digested strand containing 100% CTP. The two unmodified unique fragments ACAGp and ACAAUGp were observed in the 2- charge state having a mass of 1308.18 Da and 1943.26 Da (1308.19 Da and 1943.27 Da calculated from sequence). The mass spectrum of the IVT product obtained using 100% **1** and treated with RNase T1, is shown in Fig. 4B. The spectrum is indicative of the incorporation of the CTP analogue into the RNA strand due to a mass shift of 225.173 Da. The sequence identity of the two fragments was confirmed by tandem mass spectrometry experiments. (Fig. S2 and S3, Table S1).

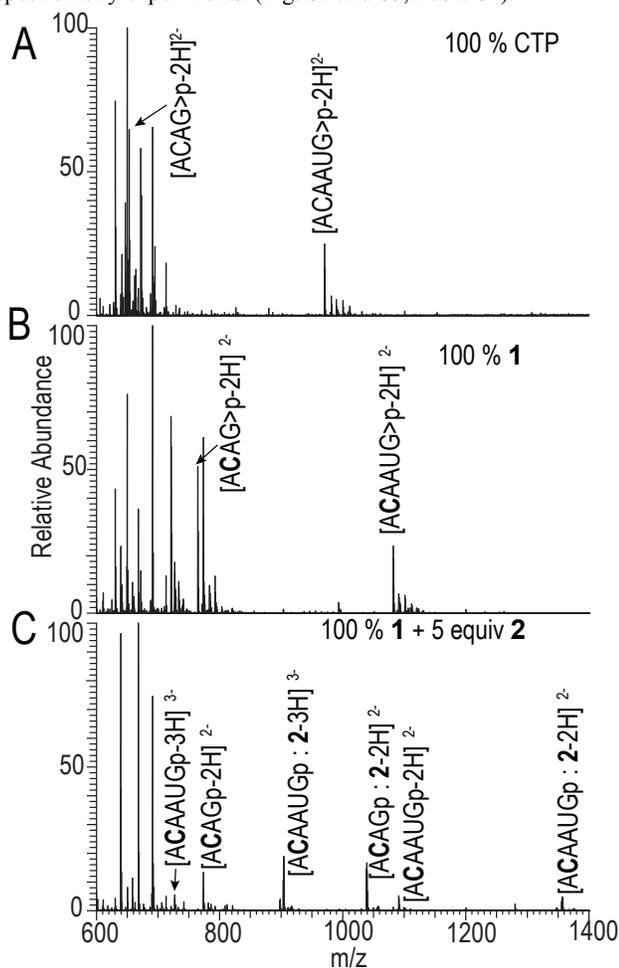


Fig. 4 Mass spectrometric analysis of RNase T1 digest of: A) IVT product using 100% CTP; B) IVT product using 100% **1**; and C) reaction between the RNA strand containing two TCO groups and 5 equiv **2**.

Post-transcriptional labelling of the RNA strands was also analysed by fluorescence spectroscopy. Fig. 5 illustrates the fluorescence spectra of: (a) TCO-modified RNA which is fluorescently silent, (b) 100 nM **2**, (c) 100 nM **2** incubated with 1 equiv of control RNA strand lacking TCO and (d) 100 nM **2** incubated with 1 equiv of IVT-synthesized RNA strand containing two TCO groups. The observed 3-fold enhancement of fluorescence is consistent with the previously characterized spectroscopic behaviour of fluorescently-labelled tetrazines.²² The observed fluorescence enhancement was however lower than the one observed between **1** + **2** (Fig. S4). This could be due to homo fluorescence resonance energy transfer (homo-FRET) between the two fluorescein chromophores attached to the RNA strand.²³

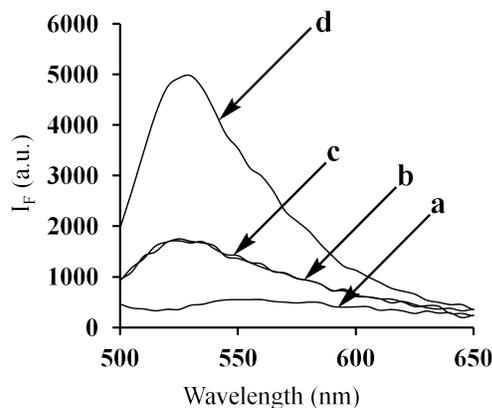


Fig. 5 Fluorescent spectra of: (a) 100 nM of IVT TCO-modified RNA by itself (b) 100 nM **2**, (c) 100 nM **2** treated with 1 equiv. of unmodified 33 mer RNA strand and (d) 100 nM **2** treated with 1 equiv. of IVT TCO-modified RNA. Spectra were obtained in 1X PBS buffer (pH 7.4), $\lambda_{\text{ex}} = 490$ nm.

Fluorescently labelled RNA strands were also characterized by mass spectrometry. The IEDDA reaction between 5 μM of the RNA and 5 equiv of **2** in 150 mM ammonium acetate (pH 7.5) at 37 °C was carried out in the dark for 30 min. The cycloaddition product was subjected to RNase T1 treatment and the product was confirmed by mass spectrometry analysis (Fig. 4C). The observed spectrum contains peaks corresponding to unreacted RNA, as well as the IEDDA product. In this type of analysis, the fact that the negative charges are carried by phosphate groups that are not affected by chemical modification enables one to estimate the relative abundances of species in solution directly from the respective signal intensities. Based on the relative peak intensities, the cycloaddition reaction achieved 60% conversion. The unreacted ACAAUGp fragment was observed at charge states 2- and 3- with a mass of 2184.42 Da (2184.43 Da calculated). The corresponding cycloaddition product with a mass shift of 531.14 Da was also observed at charge states 2- and 3-. The ACAGp fragment with mass 1549.34 Da (1549.35 Da calculated) and its corresponding cycloaddition product with mass shift of 531.14 Da were both observed in the -2 charge state. We believe that the reaction was incomplete due to partial isomerization of the TCO group to a *cis*-cyclooctene. The latter is significantly less reactive with tetrazines.²⁴ The isomerization is thought to have occurred during one of the steps in preparation to PAGE, which entailed heating of the RNA sample to 90 °C in the presence of 7M urea and 25 mM EDTA. The precursors to that step, **1** or IVT products containing **1** showed full conversion upon reacting with **2**.

To test the scope of the IEDDA chemistry for fluorescent labelling of RNA, transcription was performed using a 435 nucleotide-long DNA template, which contains over 100 cytidine

residues and encodes a ribozyme at the 3' end. After transcription, the ribozyme undergoes self-cleavage resulting in two RNA components. IVT was carried out using varying amounts of **1** and cytidine. The resulting IVT products were treated with **2** (250 μ M) and analysed on a 20% denaturing PAGE. Fig. S5 shows that **1** can be successfully incorporated at multiple sites of a much longer RNA sequence. Bio-orthogonal modification of cytidine did not disrupt the ability of the ribozyme sequence to self-cleave.

Conclusions

This communication describes a general approach for fluorescent labelling of RNA strands using a pair of bio-orthogonal reagents, a cytidine analogue derivatized with a TCO group and fluorescein-modified tetrazine. Attachment of the TCO group at the 5-position of the nucleobase of cytidine triphosphate does not disrupt the recognition of the nucleotide by T7 RNA polymerase that efficiently incorporates the analogue into a strand. In principle, this approach could be used for fluorescent modification of any RNA strand. The fluorescent labelling step could in principle be done *in vivo*, thus allowing to image the RNA of interest in real time.

M.R. would like to thank the Research Foundation of the State University of New York at Albany for financial support of this project, as well as Prof. Paul Agris for his advice. In addition, we would like to thank Qishan Lin for high resolution mass spectrometry analysis of the compounds described in this work.

Notes and references

^a University at Albany, SUNY, Department of Chemistry, 1400 Washington Avenue, Albany, NY 12222, United States. E-mail: mroyzen@albany.edu; Tel: 518-437-4463.

‡ E.A. synthesized the compounds described in this report; P.N.A. carried out all the bio-analytical work.

Electronic Supplementary Information (ESI) available: Detailed synthetic and analytical procedures, as well as compound characterization. See DOI: 10.1039/c000000x/

- (a) H. Y. Park, H. Lim, Y. J. Yoon, A. Follenzi, C. Nwokafor, M. Lopez-Jones, X. Meng and R. H. Singer, *Science*, 2014, **343**, 422-424; (b) A. R. Buxbaum, B. Wu and R. H. Singer, *Science*, 2014, **343**, 419-422.
- (a) S. Tyagi, *Nat Methods*, 2009, **6**, 331-338; (b) H. G. Garcia, M. Tikhonov, A. Lin and T. Gregor, *Curr. Biol.*, 2013, **23**, 2140-2145.
- (a) S. G. Srivatsan and Y. Tor, *Nat. Protoc.*, 2007, **2**, 1547-1555; (b) D. Shin, R. W. Sinkeldam and Y. Tor, *J. Am. Chem. Soc.*, 2011, **133**, 14912-14915.
- (a) C. Y. Jao and A. Salic, *Proc. Natl. Acad. Sci. U S A*, 2008, **105**, 15779-15784; (b) D. Qu, L. Zhou, W. Wang, Z. Wang, G. Wang, W. Chi and B. Zhang, *Anal. Biochem.*, 2013, **434**, 128-135.
- Click-iT® detection kit, sold by Life Technologies™.
- (a) R. Kumar, A. El-Sagheer, J. Tumpene, P. Lincoln, L. M. Wilhelmsson, and T. Brown, *J. Am. Chem. Soc.* 2007, **129**, 6859-6864. (b) H. Xiong, P. Leonard, and F. Seela, *Bioconjug. Chem.* 2012, **23**, 856-870. (c) R. Bonnet, P. Murat, N. Spinelli, and E. Defrancq, *Chem. Commun.* 2012, **48**, 5992-5994.
- (a) S. Seidu-Larry, B. Krieg, M. Hirsch, M. Helm, and O. Domingo, *Chem. Commun.* 2012, **48**, 11014-11016. (b) S. E. Averick, E. Paredes, S. K. Dey, K. M. Snyder, N. Tapinos, K. Matyjaszewski, and S. R. Das, *J. Am. Chem. Soc.* 2013, **135**, 12508-12511. (c) J. Willibald, J. Harder, K. Sparrer, K. K. Conzelmann, and T. Carell, *J. Am. Chem. Soc.* 2012, **134**, 12330-12333.
- (a) T. Ishizuka, M. Kimoto, A. Sato, and I. Hirao, *Chem. Commun.* 2012, **48**, 10835-10837. (b) V. R. Sirivolu, P. Chittepudi, and F. Seela, *Chembiochem*, 2008, **9**, 2305-2316. (c) S. Berndt, N. Herzig, P. Kele, D. Lachmann, X. Li, O. S. Wolfbeis, and H. A. Wagenknecht, *Bioconjug. Chem.* 2009, **20**, 558-564.
- N. J. Agard and C. R. Bertozzi, *Acc. Chem. Res.* 2009, **42**, 788-797.
- (a) M. Shelbourne, T. Brown, A. H. El-Sagheer, and T. Brown, *Chem. Commun.* 2012, **48**, 11184-11186. (b) I. Singh and F. Heaney, *Chem. Commun.* 2011, **47**, 2706-2708. (c) I. S. Marks, J. S. Kang, B. T. Jones, K. J. Landmark, A. J. Cleland, and T. A. Taton, *Bioconjug. Chem.* 2011, **22**, 1259-1263.
- (a) M. L. Blackman, M. Royzen and J. M. Fox, *J. Am. Chem. Soc.*, 2008, **130**, 13518-13519. (b) N. K. Devaraj, R. Upadhyay, J. B. Haun, S. Hilderbrand and R. Weissleder, *Angew. Chem. Int. Ed.*, 2009, **48**, 7013-7016.
- D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox, and A. Y. Ting, *J. Am. Chem. Soc.* 2012, **134**, 792-795.
- C. M. Cole, J. Yang, J. Seckute, and N. K. Devaraj, *Chembiochem*. 2013, **14**, 205-208.
- J. Yang, J. Seckute, C. M. Cole and N. K. Devaraj, *Angew. Chem. Int. Ed.* 2012, **51**, 7476-7479.
- J. Schoch, M. Wiessler, and A. Jaschke, *J. Am. Chem. Soc.* 2010, **132**, 8846-8847.
- S. K. Jarchow-Choy, E. Sjuvarsson, H. O. Sintim, S. Eriksson, and E. T. Kool, *J. Am. Chem. Soc.* 2009, **131**, 5488-5494.
- S. G. Srivatsan and Y. Tor, *Chem. Asian J.* 2009, **4**, 419-427.
- M. R. Karver, R. Weissleder, and S. A. Hilderbrand, *Bioconjug. Chem.* 2011, **22**, 2263-2270.
- J. Yang, M. R. Karver, W. Li, S. Sahu and N. K. Devaraj, *Angew. Chem. Int. Ed.* 2012, **51**, 5222-5225.
- miR-122 was chosen as a model strand because of our interest in developing a fluorescent analogue that could be studied by microscopy in live cells. miR-122 is abundantly expressed in healthy liver cells but is down regulated in hepatocellular carcinomas. Fluorescent analogue of miR-122 will be used to study factors causing its downregulation.
- (a) J. A. Pleiss, M. L. Derrick and O. C. Uhlenbeck, *RNA*, 1998, **4**, 1313-1317. (b) S. E. Moroney and J. A. Piccirilli, *Biochem.* 1991, **30**, 10343-10349. (c) E. T. Schenborn and R. C. Mierendorf, Jr. *Nucleic Acids Res.* 1985, **13** 6223-6236.
- N. K. Devaraj, S. Hilderbrand, R. Upadhyay, R. Mazitschek and R. Weissleder, *Angew. Chem. Int. Ed.*, 2010, **49**, 2869-2872.
- L. W. Runnels and S. F. Scarlata, *Biophys. J.* 1995, **69**, 1569-1583.
- F. Thalhammer, U. Wallfaher and J. Sauer, *Tetrahedron Lett.* 1990, **31**, 6851-6854.