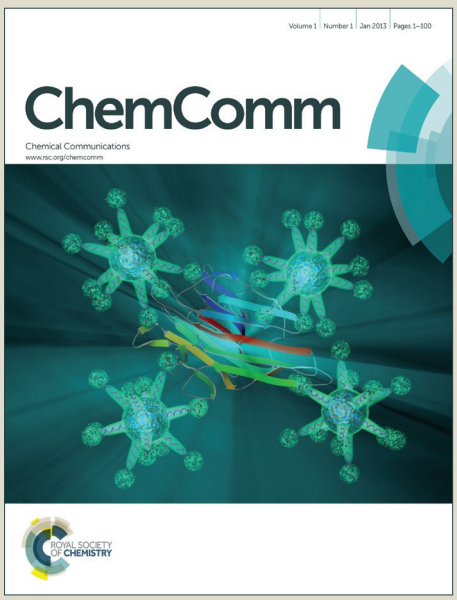
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

COMMUNICATION

Sequence-Selective DNA Recognition and Enhanced Cellular Uptake by Peptide-Steroid Conjugates

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Yara Ruiz García[a], Abhishek Iyer[a], Dorien Van Lysebetten[a], Y. Vladimir Pabon[c], Benoit Louage[b], Malgorzata Honcharenko[d], Bruno G. De Geest[b], C. I. Edvard Smith[c], Roger Strömberg[d] and Annemieke Madder*,[a]

Several GCN4 bZIP TF models have previously been designed and synthesized. However, the synthetic routes towards these constructs are typically tedious and difficult. We here describe the substitution of the Leucine zipper domain of the protein by a deoxycholic acid derivative appending the two GCN4 binding region peptides through an optimized double azide-alkyn cycloaddition click reaction. In addition to achieving sequence specific dsDNA binding, we have investigated the potential of these compounds to enter cells. Confocal microscopy and flow cytometry show the beneficial influence of the steroid on cell uptake. This unique synthetic model of the bZIP TF thus combines sequence specific dsDNA binding properties with enhanced cell-uptake. Given the unique properties of deoxycholic acid and the convergent nature of the synthesis, we believe this work represents a key achievement in the field of TF mimicry.

Gene expression at the transcriptional level is mainly regulated by proteins that bind DNA in a sequence-specific manner. These proteins, known as transcription factors (TFs), are responsible for controlling the transfer of genetic information from DNA to mRNA. More specifically, many oncoproteins are transcription factors responsible for cell-growth proliferation and tumor formation[1]. As a consequence of the specificity of these oncoproteins in the DNA sequence recognition during transcription, several approaches have been explored to develop inhibitors or modifiers of gene expression that can prevent specific genes from being transcribed[2].

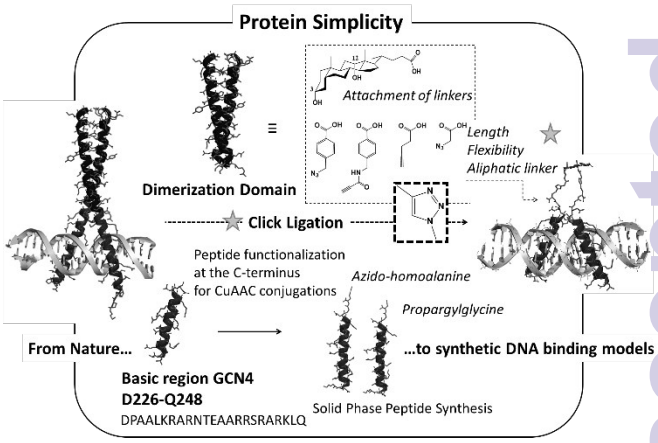


Figure 1. Artificial DNA binder design

In addition, the lack of a general recognition code for the interaction between amino acid sequences within a protein and its specific DNA-binding site has promoted the study of th structure of TFs and their interaction with the DNA[3]. We her present our efforts towards mimicking the GCN4 bZIP T (General Control Protein Leucine Zipper Transcription Factor) by a simplified synthetic construct and towards understanding hov these protein mimics behave in a cellular context[4]. The GCN bZIP TF, which controls the activation of several genes in response to amino acid starvation in yeast, has been chosen a model system in order to allow comparison with previously published TF miniaturisation attempts. Many eukary ic transcription factors feature homologous protein sequen as forming a family generally referred to as the basic-region-leucine-zipper or bZIP motif of which the mammalian ATF/CRE family of transcription factors represent a large group. Ou target has been to develop a new type of site-specific DNA binders which can recognize dsDNA by specific binding in th major groove and additionally show enhanced cellular uptake by exploiting the unique properties of the steroid moiety. W have been inspired by the bZIP (leucine zipper) and the b-H...

[a] Organic and Biomimetic Chemistry Research Group. Department of Organic and Macromolecular Chemistry, Ghent University. Krijgslaan 281 (S4), B-9000 Ghent (Belgium). E-mail: annemieke.madder@ugent.be.
[b] Department of Pharmaceutics. Ghent University. Harelbekestraat 72, 9000 Ghent. Belgium. E-mail: br.degeest@ugent.be.
[c] Clinical Research Center, Department of Laboratory Medicine, Karolinska University Hospital Huddinge, SE-141 86, Stockholm, Sweden. E-mail: edvard.smith@ki.se.
[d] Department of Biosciences and Nutrition (BioNut), H2. Karolinska Institutet. Novum 141 83 Huddinge, Stockholm, Sweden. E-mail: Roger.Stromberg@ki.se.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

ZIP(helix-loop-helix leucine zipper) motifs, in which the basic DNA recognition region binds to the major groove as a dimer, inserting two α-helices held in the correct position by a dimerization domain[5]. The main residues of the protein involved in the DNA recognition comprise the amino acids 226-248 of the basic region of the GCN4 protein. Previously, models of such transcription factors have been synthesized by different research groups employing a series of different dimerization domain substitutes. Pioneering work was carried out by the group of Kim[6], developing an analogue in which the complete dimerization domain was substituted by a simple disulfide bond. Building on this successful idea of miniaturisation, Morii, Schepartz, Mascareñas and our group have enforced the proof-of-concept by using a variety of small dimerizing moieties[7–12]. Subsequently, we have shown that the attachment of the basic region peptides to a rigid scaffold, a derivative of deoxycholic acid in this case, also allows selective recognition of DNA. Indeed, our previous work on cMyc-Max b-HLH-ZIP and GCN4-bZIP proteins showed that this type of steroid-based constructs show potential for binding DNA.[10,13] The specific choice of the steroid scaffold was inspired by the fact that it is inexpensive, commercially available, versatile and easy to modify synthetically. In addition, its known ability to enhance proteolytic stability of attached peptides[14–16], amphiphilicity[17] , the conformational properties ensuring correct positioning of the two appended chains[18] and the potential to increase cellular uptake and bioavailability[19] render it an excellent candidate for the attachment of the two DNA recognizing arms of the zipper motif of GCN4. The incorporation of a spacer between the peptides and the steroid scaffold was shown important in order to provide the final conjugate with enough flexibility to adopt an optimal conformation for specific interaction with the major groove of the target DNA sequence.[13] However, the concomitant further lengthening of the 23-mer peptide chains caused increased tendency of peptide aggregation necessitating the use of microwave assistance for successful synthesis.[13] In order to synthesize these peptidosteroids in a more effective and convenient way, we here explored a convergent approach involving the CuAAC reaction (Copper-catalyzed Alkyne-Azide Cycloaddition) to conjugate the recognition domains to a functionalized steroid scaffold. The geometric, steric and electronic properties of the 1,2,3-triazole resemble a trans-amide bond, while also affording resistance to enzymatic degradation[20–23], hydrolysis and oxidation. In addition, successful replacement of two amino acids in α-helical peptides by a triazole unit has been shown to not significantly influence the secondary peptide structure[20]. To date the triazole linkage has scarcely been used for peptide-steroid conjugation and only to assemble short apolar tripeptides onto bile acid scaffolds[24], despite the increasing interest of these type of conjugates in diverse applications such as HIV inhibitors[25,26] and immunogens for vaccine development[27].

In the current work we have designed and synthesized four different scaffolds for peptide dimerization of the GCN4 basic region, which was made possible by optimizing conditions for the CuAAC mediated conjugation of the long, unprotected and functionalized zipper peptides. A small series of deoxycholic acid derivatives was conceived, differing with respect to spacer between the peptide and the steroid skeleton, which have different lengths, rigidities and functionalities. Hereto commercially available deoxycholic acid was modified at the alcohol positions by attachment of different linkers. The linkers chosen for the study encompass pentynoic acid, azido glycine, 4-azidomethyl-benzoic acid and (N-Propynoylamino)-p-toluic acid (PATA). The PATA linker has been specifically developed for bioconjugation purposes as an active alkyne for preparation of peptide-oligonucleotide conjugates via CuAAC[28]. Functionalization of the steroid nucleus was performed by Steglich esterification, affording final scaffolds (**1**, **2** & **3**). In case of PATA as linker (**4**), the diamino derivative of deoxycholic acid proved necessary for the coupling of the linker, as esterification gave rise to byproducts due to the high reactivity of the alkyne.

The GCN4 basic DNA binding region consists of 23 amino acids that specifically recognize the ATF/CREB binding site (5'-ATGA C/G TCAT-3'), which is the functional target of GCN4 *in vivo* and involved in inducing amino acid biosynthesis in yeast.

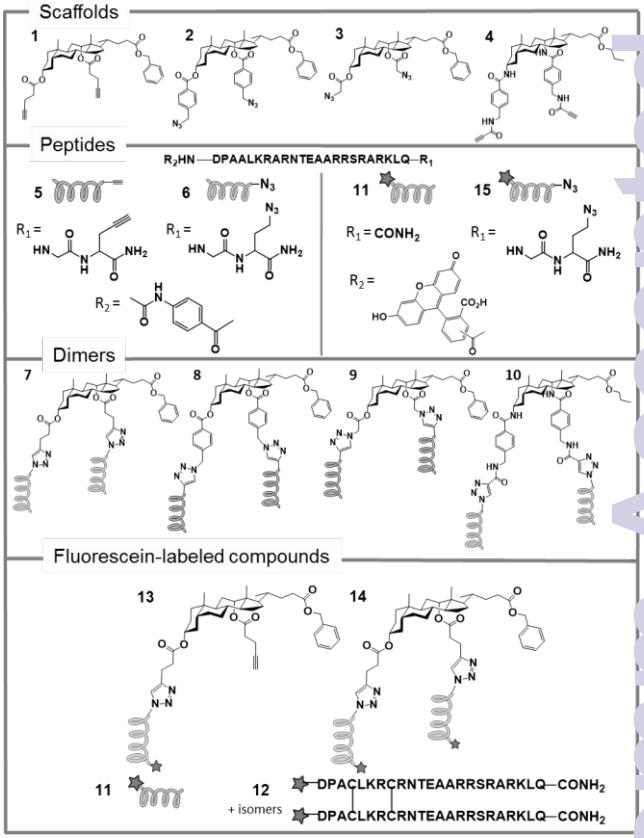


Figure 2. Deoxycholic acid derivatives for the substitution of the dimerization doma with the corresponding peptides.

In order to append the peptides to the bile acid scaffolds, peptides were modified at the C-terminus with unnatural amino acids bearing an alkyne or an azide. Peptides **5** and **6** were synthesized in an automated fashion using Fmoc/tBu based solid phase peptide synthesis (SPPS). The alkyne and azide functionalized GCN4 basic region peptides, **5** and **6** were then attached to the central steroid core, affording four different transcription factor models. With this hydrophilic deprotected peptide and the hydrophobic scaffold, the DMSO/H_2O combination was found to be optimal for the CuAAC reaction.

As a catalyst, $Cu(CH_3CN)_4PF_6$ gave the best results. A high excess of catalyst was needed possibly due to complexation of copper with the nitrogen containing side chains of the peptide. However, reaction proceeds well with excess of copper ion complex and the copper ions can be readily removed after reaction by use of EDTA.[28] An excess of scaffold was also required for complete reaction of the peptide, the dipodal construct being favoured over the monopodal one under these conditions. The reaction was complete after 3 hours at room temperature and compatible with the presence of all unprotected amino acids in the sequence. Purification of final constructs after completion of the reaction was possible via RP-HPLC, affording compounds **7**, **8**, **9** and **10** in high purity for DNA binding studies. DNA binding affinity of mimics **7**, **8**, **9** and **10** was evaluated using an Electrophoretic Mobility Shift Assay (EMSA). The study was based on titration of a duplex DNA sequence containing the ATF/CREB recognition site with increasing concentrations of compounds **7**, **8**, **9** and **10**.

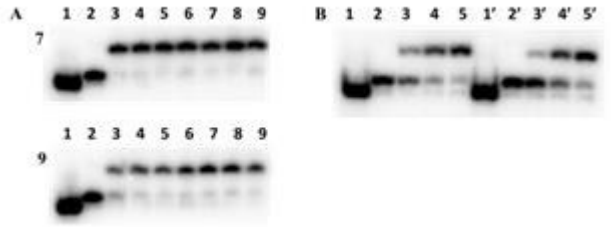


Figure 3. A) EMSA titration of the dipodal peptidosteroid conjugates **7** and **9** to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') and its complementary strand (5' - GAA AAA AAA TGA CGT CAT CCG – 3') at 5nM: First lane in all the gels: pyrimidine strand (5'-labeled ^{32}P-CRE sequence). Lanes 2-9 contain peptide concentrations of 0, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.1125 and 0.125 µM for **7** and **9**. B) EMSA titration of the dipodal peptidosteroid conjugates **7** and **9** to the 5'-labeled ^{32}P-CRE sequence (5 – CGG ATG ACG TCA TTT TTT TTC – 3') at 5nM in the presence of competitor DNA sequence (5'-AGCAGAGGGCGTGGGGGAAAAGAAAAAAGATCCACCGGTCGCCAC-3') at 500 nM: First lane in all the gels: pyrimidine strand. Lanes 2 and 2' to 5 and 5' contain peptide concentrations of 0, 0.05, 0.125 and 0.312 µM for **7** and **9** respectively.

No clear specific binding to DNA could be observed for mimics **8** and **10** (gel not shown). On the other hand, EMSA for compounds **7** and **9** in the absence of any competitor DNA shows that both compounds have a high affinity for the CRE DNA sequence (figure 3A). A closer visual inspection of the gels reveals that **7** has slightly higher DNA binding affinity as compared to **9**. Due to the more or less complete up-shifting of bands even at the lowest concentration, accurate determination of the dissociation constant (KD) was not possible under these conditions. However, by a simple visual inspection we can conclude that almost all the DNA is completely bound at 50 nM in case of compound **7** and at 62.5 nM in case of **9**. Decreasing the peptide concentration to a value < 50 nM results in large standard deviations in KD calculations. Also, decreasing the ratio of DNA:peptide further without lowering the DNA concentration is not feasible due to problems in detection of the isotope. Therefore we conducted a second series of EMSA studies which involved a competitor DNA which allows for both checking of nonspecific interactions and calculation of the KD. EMSA for compounds **7** and **9** in the presence of 500 nM competitor DNA reveal that nonspecific interactions do exist, however the specificity of the compounds towards the CRE sequence is still present to a large extent. This can be seen from the gradual gel shift in contrast to the previous gel (figure 3B). In support of this is that KD calculations from competition experiments with 0.5-2 µM competitor DNA give values of 4-8*10 nm for **7** and 1-3*10^2 nm for **9** (see ESI). We can also state here more conclusively that **7** is clearly not only the better DNA binding construct but also shows greater specificity as compared to **9**.

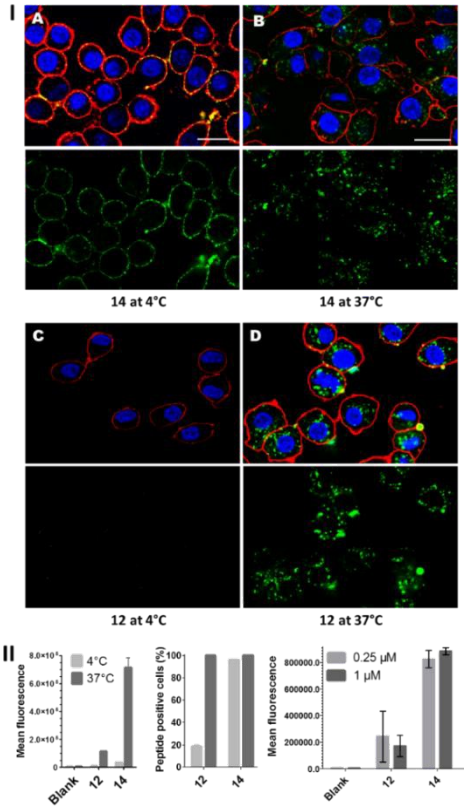


Figure 4. **I**. Confocal microscopy of RAW264.7 cells incubated with (A) **14** at 4 °C, (B) **14** at 37°C, (C) **12** at 4 °C, and (D) **12** at 37 °C at 0.25 µM (green fluorescence signal). Cell nuclei were labeled with Hoechst (blue) and cell membranes with AlexaFluor-conjugated cholera toxin subunit B (red). The lower panels only show the green fluorescence channel. (Scale bar = 20 µm). **II.** Flow cytometry analysis of mean cell fluorescence (left) and peptide positive cells (middle) of the synthesized compounds 12 and 14 incubated with RAW264.7 cells at 4 °C and 37°C at 0.25 µM. Mean fluorescence of compounds 12 and 14 at 0.25 µM and 1 µM.

A series of experiments were set up to investigate the cell uptake capacity of the various synthetic constructs. For this purpose the best DNA binding cholic acid based mimic **7** of the GCN4 protein was resynthesized incorporating a fluorescein tag to give **14**. The properties of this construct were compared to the labelled but otherwise unmodified GCN4 peptide dimer **12** and monomer **11** as well as the monomeric steroid conjugate **13**. The toxicity (MTT assay), quantification of uptake (flow cytometry) and intracellular localization of these constructs were studied on RAW264.7 mouse macrophages, which have been already used for cell penetration studies with a similar GCN4 peptide[30]. Compounds **12** and **14** showed a cell viability at 0,25 µM of 88 % and 86 % respectively (SI) and both are taken up at 37 °C but to a different degree. There is clearly an enhanced uptake when deoxycholic acid is used as a scaffold, as seen from the mean fluorescence values obtained by flow cytometry analysis, as they are more than four times higher for

14 than for **12** (figure 4.II). The monopodal cholic acid derivative **13**, although non-DNA binding and more hydrophobic, also exhibits enhanced uptake. However, a decreased uptake is seen when the concentration is increased to 1 μM (SI). This could be attributed to the denaturant like properties of the cholic acid. The most interesting results were obtained when comparing the localisation of the DNA binding cholic acid dimer **14** at 4°C and 37°C (figure 4.IA & 4.IB). At 4°C, where endocytosis is blocked, only binding to the cell membrane was observed. This resulted in a high percentage of peptide positive cells (figure 4.II left and middle). However, at 37°C, whereby both passive and active uptake is possible, the uptake is considerably higher as evidenced by the higher mean cell fluorescence. This indicates that the deoxycholic acid coupled peptides are mainly internalized via active transport at 37°C. The most likely explanation is that the peptides follow an endocytotic pathway.

In conclusion, we here have illustrated a strategy to conjugate relatively long, unprotected peptides to bile acid scaffolds in a convergent manner. From the four models of the GCN4 bZIP transcription factor presented, the one with the most flexible linker (**7**) proved to be the best synthetic DNA binder. This can be attributed to the fact that the linker allows the construct to grip the major groove of the DNA like a pair of tweezers which is less optimal in the case of the other linkers, which may be too long, too short, or too inflexible. The binding affinity of **7** is comparable with that of earlier described models of GCN4 TFs with KDs in the nM range, while the synthetic route is considerably less complicated. In addition, compound **7** is more readily taken up by cells than non-steroid constructs as evidenced by the uptake in RAW264.7 mouse macrophages. This illustrates the particular properties of the peptide combined with the steroid nucleus, which allows uptake at low concentrations. Using four models of the GCN4 bZIP TF we were able to identify some of the parameters that affect dsDNA recognition in synthetic constructs. Additionally we also for the first time discuss and reveal the interesting cell-uptake properties of this type of peptidosteroid based TF mimics.

Acknowledgement

Tim Courtin is acknowledged for the determination of the concentration of the final compound by ERETIC-NMR and Jan Goeman for the LC-MS analysis. Yara Ruiz García and Abhishek Iyer are indebted to the Marie Curie Early Stage Research Training Fellowship of the European Community's Seventh Framework Programme under contract number PITN-GA-2010-238679. The FWO-Flanders, BOF (UGent) and the Swedish Research Council are also acknowledged for their financial support. Y. Vladimir Pabon is indebted to the Departamento Administrativo de Ciencia, Tecnología e Innovación COLCIENCIAS, grant 02007/24122010.

Notes and references

1. D. Hanahan and R. A. Weinberg, *Cell*, 2011, **144**, 646–74.
2. F. Fontaine, J. Overman, and M. François, *Cell Regen. (London, England)*, 2015, **4**, 2.
3. T. Suganuma, M. Kawabata, T. Ohshima, and M.-A. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 13073–8.
4. K. Arndt and G. R. Fink, *Proc. Natl. Acad. Sci. U. S. A.*, 198[illegible] **83**, 8516–20.
5. T. E. Ellenberger, C. J. Brandl, K. Struhl, and S. C. Harrison, *Cell (Cambridge, Mass.)*, 1992, **71**, 1223–1237.
6. R. V Talanian, C. J. McKnight, and P. S. Kim, *Sci. (Washington, D. C., 1883-)*, 1990, **249**, 769–771.
7. A. M. Caamano, M. E. Vazquez, J. Martinez-Costas, L. Castedo, and J. L. Mascarenas, *Angew. Chem., Int. Ed.*, 2000, **39**, 3104–3107.
8. M. Ueno, A. Murakami, K. Makino, and T. Morii, *J. Am. Chem. Soc.*, 1993, **115**, 12575–12576.
9. J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vazquez, and J. L. Mascarenas, *Nat. Commun.*, 2013, **4**, 1874–1878.
10. L. L. G. Carrette, T. Morii, and A. Madder, *European J. Org. Chem.*, 2014, **2014**, 2883–2891.
11. B. Cuenoud and A. Schepartz, *Sci. (Washington, D. C., 1883-)*, 1993, **259**, 510–513.
12. Y. Ruiz García, J. Zelenka, V. Pabon, A. Iyer, M. Buděšínský, T. Kraus, C. I. E. Smith, A. Madder, *Org. Biomol. Chem.*, 2015.
13. D. Verzele, A. Madder, *Eur. J. Org. Chem.*, 2013, 673–68[illegible]
14. W. Kramer, G. Wess, A. Enhsen, E. Falk, A. Hoffmann, G. Neckermann, G. Schubert, and M. Urmann, *J. Control. Release*, 1997, **46**, 17–30.
15. C. a Bodé, T. Bechet, E. Prodhomme, K. Gheysen, P. Gregoir, J. C. Martins, C. P. Muller, and A. Madder, *Org. Biomol. Chem.*, 2009, **7**, 3391–9.
16. D. B. Salunke, B. G. Hazra, and V. S. Pore, *Curr. Med. Che[illegible]* 2006, **13**, 813–47.
17. C. A. Bode, C. P. Muller, and A. Madder, *J. Pept. Sci.*, 2007, **13**, 702–708.
18. A. Madder, L. Li, M. H. De, N. Farcy, H. D. Van, F. Fant, G. Vanhoenacker, P. Sandra, A. P. Davis, and C. P. J. De, *J. Comb. Chem.*, 2002, **4**, 552–562.
19. C. A. Bode, T. Bechet, E. Prodhomme, K. Gheysen, P. Gregoir, J. C. Martins, C. P. Muller, and A. Madder, *Org. Biomol. Chem.*, 2009, **7**, 3391–3399.
20. W. S. Horne, M. K. Yadav, C. D. Stout, and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2004, **126**, 15366–15367.
21. H. C. Kolb and K. B. Sharpless, *Drug Discov. Today*, 2003, **8**, 1128–1137.
22. C. W. Tornoe, C. Christensen, and M. Meldal, *J. Org. Chem* 2002, **67**, 3057–3064.
23. I. E. Valverde, A. Bauman, C. A. Kluba, S. Vomstein, M. A. Walter, and T. L. Mindt, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 8957–60.
24. R. Echemendía, O. Concepción, F. E. Morales, M. W. Paixão, and D. G. Rivera, *Tetrahedron*, 2014, **70**, 3297.
25. A. Hollmann, P. M. Matos, M. T. Augusto, M. A. R. B. Castanho, and N. C. Santos, *PLoS One*, 2013, **8**, e60302.
26. P. Ingallinella, E. Bianchi, N. A. Ladwa, Y.-J. Wang, R. Hrin, M. Veneziano, F. Bonelli, T. J. Ketas, J. P. Moore, M. D. Miller, and A. Pessi, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5801–5806.
27. J. P. Tam, *J. Immunol. Methods*, 1996, **196**, 17–32.
28. M. Wenska, M. Alvira, P. Steunenberg, A. Stenberg, M. Murtola, and R. Stroemberg, *Nucleic Acids Res.*, 2011, **39**, 9047–9059.
29. T. Hai and M. G. Hartman, *Gene*, 2001, **273**, 1–11.
30. S. Futaki, T. Suzuki, W. Ohashi, T. Yagami, S. Tanaka, K. Ueda, and Y. Sugiura, *J. Biol. Chem.*, 2001, **276**, 5836–40.