ChemComm

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

View Article Online
DOI: 10.1039/C7CC01705G

ChemComm

RSCPublishing

COMMUNICATION

Specific dsDNA Recognition by a mimic of the DNA binding domain of the c-Myc/Max transcription factor

Cite this: DOI: 10.1039/xoxxoooooox

Yara Ruiz García [a], Y. Vladimir Pabon [b], C. I. Edvard Smith[b] and Annemieke Madder*[a]

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

We here report on the synthesis of the first mimic of the DNA binding domain of the c-Myc/Max-bHLH-ZIP transcription factor able to selectively recognize its cognate E-box sequence 5'-CACGTG-3' through the major groove of the double-stranded DNA. The designed peptidosteroid conjugate was shown to be effective as DNA binder in the presence of excess competitor DNA.

In the last decades, developments in the areas of biotechnology and chemical biology have opened up new prospects in anti-cancer research. Currently, the inhibition of oncogenes and/or the reactivation of tumor suppressor genes by transcription factors (TFs) are of particular interest[1,2]. In this context, protein-DNA interactions have become a principal target of study through several strategies, as they are a keystone in the regulation of gene expression[3,4]. Amongst other strategies, the design of synthetic TF models has provided insights into the latter process[5].

The Myc/Max/Mad network is a family of TFs containing a basic Helix-Loop-Helix Zipper motif (bHLH-ZIP TF), which is responsible for gene specific transcriptional regulation[6]. These oncoproteins are heterodimers composed of proteins from the Myc, Max and Mad families[7–9]. The c-Myc-Max heterodimer binds the double-stranded DNA (dsDNA) in a sequence-specific manner and its function consists in the transcription initiation from promoters containing the hexameric sequence (E-box) 5'-**CACGTG**-3'[10]. The dysregulation of the c-*MYC* gene causes its overexpression, which is related to the progression of certain types of cancer[11–13].

The crystal structure of cMyc-Max bHLH-ZIP TF has been studied in detail. The structural motif of bHLH-ZIP proteins is similar to the basic Leucine Zipper domain (bZIP) one. Both are characterized by two domains: (i) a dimerization domain, which is a prerequisite for DNA binding, and (ii) a DNA recognition domain known as the basic region that contains the α-helical peptides responsible for the specific DNA binding. However, the main difference is situated in the interface between the two domains. In the case of bZIP proteins the α helices are continuous, while in the case of bHLH-ZIP proteins they are interrupted by a loop, which does not contribute to DNA binding, rendering the design and synthesis of the mimics of the bHLH-ZIP protein considerably more complicated (figure **1**) [14,15].

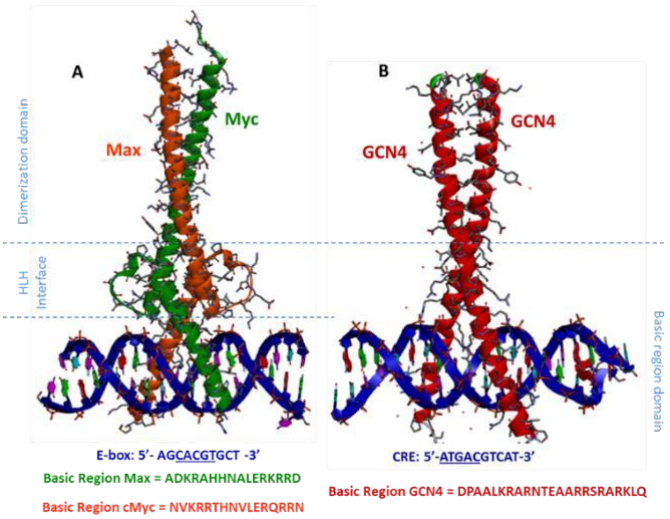

Figure **1**. Representation of the experimentally-determined structures of cMyc/Max bHLH-ZIP, PBD -ID: 1NKP (left) and GCN4 bZIP, PBD -ID: 1YSA (right) Transcription Factors.

Due to the relevant role of the c-Myc-Max heterodimer oncoprotein in cancer development, Kent et al[14] developed the first approach directed towards the total synthesis of a non-covalently-linked heterodimeric bHLH-ZIP protein mimic by convergent chemical ligation of unprotected peptide segments. Furthermore, various artificial models of homodimeric and heterodimeric bZIP TFs have been developed by several groups[16–27], through substitution of the dimerization domain by a synthetically accessible linker to which the basic regions were attached. We previously contributed to the area with the synthesis of simplified bHLH-ZIP TF models which involved the substitution of the dimerization domain by a steroid-based moiety[28] (Figure **2**).

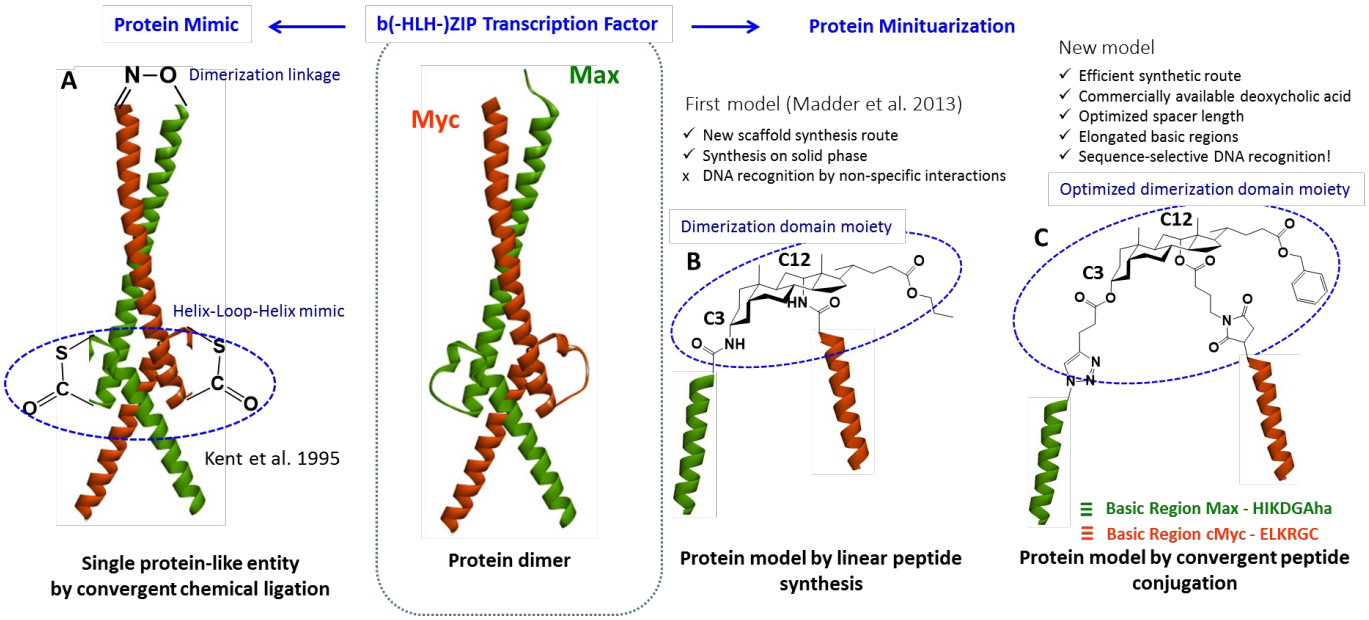

Figure **2**. Illustration of the first mimic of bHLH-ZIP TF by Kent (**A**) and our cMyc/Max mimics of the DNA binding domain of bHLH-ZIP TF (**B** and **C**).

Steroidal building blocks have been extensively studied in drug design due to their amphiphilic nature, bioavailability and tendency to enhance peptide biostability[29–32]. In addition, due to their rigid core, suitable dimensions (in terms of distance between the different attachment points on the steroid scaffold) for major groove binding increase the helicity of the appended peptides and they can thus be envisaged to substitute the leucine zipper domain as an anchoring point for the basic region of bHLH ZIP TF. While various other scaffolds have been evaluated as artificial dimerization domains [33–36], the steroid offers rigidity, proper dimensions for major groove recognition as well as the capability of increasing cell-uptake. Indeed, we have shown with previous models that the steroid moiety within peptide-steroid conjugates substantially increases their cellular uptake[23]. Our previously synthesized artificial peptidosteroid-based mimic of the DNA binding domain of the c-Myc-Max oncoprotein suffered from unsuccessful DNA binding due to the lack of an α-helix conformation and suitable orientation of the appended peptides.[28] It was shown that the direct conjugation of the basic region peptides to the steroid scaffold impedes the correct dimerization of the peptides due to the rigidity and steric encumbrance imposed by the steroid moiety which could be derived with the aid of molecular modeling from the high resolution of the protein structure by Burley et al[7] (PDB-ID: 1NKP).

We now report on the synthesis of an optimized model of the DNA binding domain of the cMyc/Max-bHLH ZIP TF able to bind specifically to the non-palindromic target DNA sequence. We decided to introduce spacers between the dimerization moiety and the appended peptides in addition to using an elongated DNA recognition domain. We considered the elongation of the basic region peptides at the C-terminus corresponding to at least four amino acids or approximately one helical turn, which should be added to the 16-17-residue c-Myc/Max peptides respectively. A linker was additionally foreseen in the final peptide-steroid conjugate in order to provide the peptides with the required flexibility needed for insertion into to the major groove of the dsDNA. Glycine was coupled as second residue to increase the space between the scaffold and the peptide as reported in previous models of bZIP TFs. Moreover, a convergent solution phase approach was applied for the appendage of the peptides onto the scaffold, as a more efficient way to synthesize such peptide conjugates, in comparison with the previously developed solid phase-based linear strategy. In order to dimerize the peptides c-Myc and Max onto the scaffold, an orthogonally functionalized deoxycholic acid scaffold derivative was synthesized to allow the convergent incorporation of both basic region peptides (Figure **3**). Commercially available deoxycholic acid (**1**) was chosen as starting material as its concave structure allowed the two alcohol functionalities to be selectively esterified facilitating convergent conjugation of two different peptide sequences[37]. In a first step, the carboxylic acid functionality was protected with benzyl bromide to facilitate HPLC monitoring of the conjugation reactions (**2**). Furthermore, as reported previously, aliphatic linkers have proven useful to ensure DNA binding for this type of TF mimics[23]. These spacers between the peptides and the scaffold should provide sufficient length and flexibility to allow accommodation of the peptides into the DNA major groove. Therefore, we chose 4-maleimido-butyric acid and pentynoic acid as linkers to ensure the flexibility needed for DNA recognition through the major groove. Selective esterification with pentynoic acid at C3 was carried out using a catalytic amount of DMAP and EDC·HCl as coupling reagent (**3**). Subsequent functionalization of the alcohol at C12 with Boc protected gamma-aminobutyric acid was performed with EDC and excess of DMAP to allow the reaction to reach completion. Finally, Boc deprotection and transformation of the obtained amine into a maleimide resulted in hetero-difunctionalized scaffold (**6**). Cu catalyzed azide alkyn cycloaddition click and cysteine-maleimide conjugation strategies were chosen for peptide conjugation in view of their generality and orthogonality. For this purpose, the c-Myc and Max peptides sequences were modified at the C-terminus with cysteine and azido-homoalanine (Aha) respectively. The Myc/Max basic regions consist of a sequence of 16 amino acids that specifically recognize the E-box DNA sequence 5'- AG**CACGTG**CT -3'. The monomeric Myc (NVKRRTHNVLERQRRNELKRGC) and Max (ADKRAHHNALERKRRDHIKDGAha) sequences were synthesized using Fmoc-/tBu-SPPS on Rink-amide ChemMatrix resin.

ChemComm
View Article Online
DOI: 10.1039/C7CC01705G
COMMUNICATION

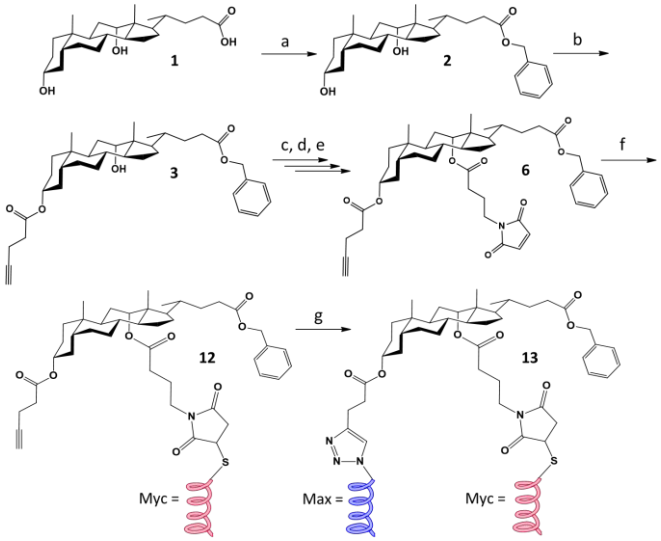

Figure **3**. Synthetic scheme towards final peptidosteroid conjugate **13**. a) DBU, Benzyl bromide, Dry DMF, 50°C, 24h (78%); b) Pentynoic acid, EDC·HCl, DMAP, dry DCM, rt, overnight (60%); c) Boc-Gaba-OH, EDC, DMAP, dry DCM, rt, overnight (40%); d) 20% TFA in DCM, 2h (99%); e) Maleic anhydride, dry toluene. 80°C, 24h (38%); f) H_2O/ACN 1:1, NH_4HCO_3 pH 6.5, rt, 2h; g) $Cu(CH_3CN)_4PF_6$, DMSO/H_2O, rt, 3h (Combined yield for step f and g: 24%).

The resulting Myc/Max basic regions, functionalized with a cysteine and an azide, were then conjugated to the central steroid scaffold resulting in the heterodimeric TF mimic. First, the conjugation at the C12 position was performed via thiol-maleimide conjugation through 1,4-Michael addition to a cysteine-contained peptide resulting in a succinimidyl thioether moiety (**12**). We decided to perform the conjugation at C12 in first place as the maleimide functional group could be hindered by the presence of a peptide at C3. Without isolation of the monopodal intermediate **12**, the second conjugation was performed via CuAAC at the C3 position of the alkyne-functionalized scaffold forming a triazole (**13**, Figure **3**).
The click conjugation was compatible with the presence of the peptide chain at C12. These and previous results show that the here applied CuAAC protocol is compatible with different peptide sequences. Both, the succinimidyl thioether and triazole moieties are generally accepted as stable linkages under physiological conditions.
In order to detect DNA binding affinity of peptide-steroid conjugate **13** an Electrophoretic Mobility Shift Assay (EMSA) using radioactive ^{32}P was performed (Figure **4**). Whereas our previously developed homodimeric GCN4 based mimics[23], when added to a 21-mer dsDNA containing the ATF/CREB recognition site, lead to a complete shift in the EMSA analysis, indicating full binding with only a 10-12 fold excess of peptide, the current Myc/Max models bind less good to the 26-mer target containing the E-box recognition-binding site in the middle requiring an at least 50-fold excess to detect any binding. Additionally, when using larger excesses of peptide, formation of higher order complexes can be observed (see supporting information). This observation is in line with the low specificity of the native cMyc/Max heterodimer towards the E-Box DNA sequence, presenting multiple binding sites[38–40]. Additional experiments were carried out using a 50-mer dsDNA target sequence, containing the E-box recognition-binding site in the middle (5'-CCATGGCGAGCGTCGCTACTAG**CACGTG**CTAGTAGGTGCGCTATCTAAGG-3').Increasing concentrations of compound **13** were added, at a constant 5 nM concentration of dsDNA and different ratios of dsDNA:peptide concentrations were tested to study binding of compound **13** to the 50-mer E-box DNA sequence (figure **4** A).

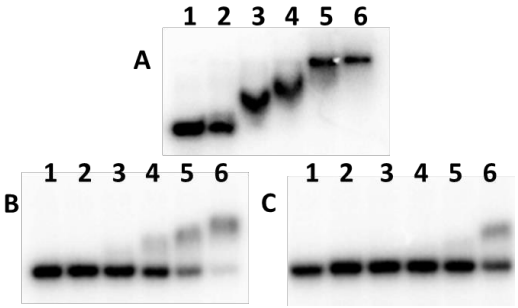

Figure **4**. A, B and C) EMSA titration of the peptide-steroid conjugate **13** to 5 nM of 5'-labeled ^{32}P-DNA target sequence (50-mer) containing the E-box. First lane in all the gels: dsDNA. Lanes 2-6 contain peptide concentrations from 0.125, 0.25, 0.5, and 1.0, up to 2.0 µM of **13**. B) EMSA titration in the presence of competitor dsDNA sequence 45-mer (5'-AGCAGAGGGCGTGGGGGAAAAGAAAAAAGATCCACCGGTCGCCAC-3') at 0.5 µM. C) EMSA titration in the presence of competitor dsDNA sequence 45-mer at 2 µM concentration.

It can now be noticed that, at a ratio of 1:50 dsDNA:peptide complete binding can be observed in the gel, evident by the presence of an up-shifted band from the dsDNA band (Figure **4**A, lane 3). Again higher-order species are visible at increased peptide:dsDNA ratios (Figure **4**A, lanes (5-6). Additional EMSA experiments were then performed to check specificity of construct **13** towards the recognition site of the dsDNA. Therefore, two competitors were used: 1) a dsDNA sequence (45-mer), and 2) a ssDNA sequence (15-mer) both of which are unrelated sequences. Competitor sequences were added at different concentrations. Whereas binding is clearly less strong than for our homodimer GCN4 mimics and adding competitor DNA influences the binding pattern, our results reveal that construct **13** can still bind specifically to the target sequence to a certain extent, even in presence of a high excess of competitor dsDNA (figure **4** B and C). It was however also shown that the concentration of peptide required for dsDNA binding was very high as indicated by the fact that the dsDNA band is still present at 2 µM of peptide, which was the concentration required for complete dsDNA binding without competitor dsDNA. Further EMSA experiments were performed using 3 µM (Figure S18) and 12 µM (Figure S19) of ssDNA as competitor (5'-CCTTTTCTTTTTTCT-3'). The used concentrations ensure the same number of competing negative charges for both types of competitors. Peptide binding was observed even when 1-4 fold excess of the ssDNA was used. As similar results were observed for both types of competitor, this does support that the peptide is specific for the target sequence, but only at high excess concentration of the peptide (ratio 1:400). This can be explained by the existence of random electrostatic interactions between the peptide and the competitors. Although non-specific interactions can be observed, the results of the experiment clearly reflect the formation of a specific dsDNA/dimeric peptide construct albeit with lower affinity as compared to our previously reported homodimeric models.

To conclude, we show the possibility of selective functionalization of steroid-based scaffolds to synthesize heterodimeric peptide-based conjugates, in principle allowing application of current methodology to other families of TF, such as bZIP Fos/Jun TF. Moreover, our design overcomes the difficulties observed when decorating building blocks by SPPS, since it is based solely on conjugation strategies in solution phase.

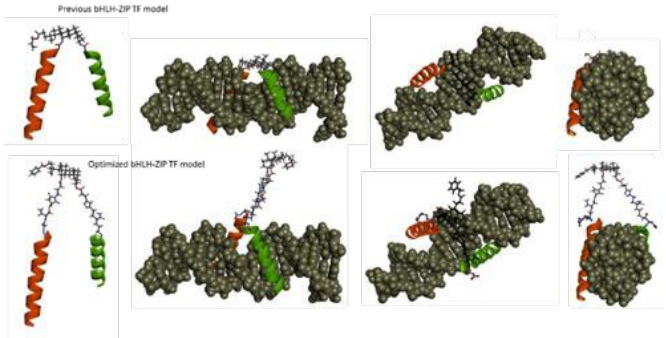

Figure **5**. Visualization models of our previous peptide-steroid conjugate and the optimized conjugate **13** in complex with its binding sequence from different perspectives.

In addition, the potential of a reduced-size bHLH-ZIP cMyc/Max TF to recognize its dsDNA binding sequence was demonstrated. The results show that an optimized and miniaturized version of the protein contains the necessary features to form a complex with dsDNA in a sequence-specific manner (Figure **5**).

Acknowledgements

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 is 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) from Colombia, grant 02007/24122010.

Notes and references

[a]Organic and Biomimetic Chemistry Research Group. Department of Organic and Macromolecular Chemistry. Ghent University. Krijgslaan 281, S4, B-9000. Gent, Belgium. E-mail: annemieke.madder@ugent.be.
[b]Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Stockholm, Sweden. E-mail: edvard.smith@ki.se

†Electronic Supplementary Information (ESI) available: Experimental details for scaffold and peptide synthesis, peptide to scaffold conjugation and EMSA analysis. See DOI: 10.1039/c000000x/

1. J. E. Darnell Jr., *Nat. Rev. Cancer*, 2002, **2**, 740–749.
2. T. A. Libermann and L. F. Zerbini, *Curr. Gene Ther.*, 2006, **6**, 17–33.
3. M. I. Barrasa, P. Vaglio, F. Cavasino, L. Jacotot, and A. J. M. Walhout, *BMC Genomics*, 2007, **8**, 21.
4. Z. Xie, S. Hu, J. Qian, S. Blackshaw, and H. Zhu, *Cell. Mol. Life Sci.*, 2011, **68**, 1657–1668.
5. E. Pazos, J. Mosquera, M. E. Vázquez, and J. L. Mascareñas, *ChemBioChem*, 2011, **12**, 1958–1973.
6. C. Grandori, S. M. Cowley, L. P. James, and R. N. Eisenman, *Annu. Rev. Cell Dev. Biol.*, 2000, **16**, 653–699.
7. S. K. Nair and S. K. Burley, *Cell (Cambridge, MA, U. S.)*, 2003, **112**, 193–205.
8. V. H. Cowling and M. D. Cole, *Semin. Cancer Biol.*, 2006, **16**, 242–252.
9. N. Kuramoto, K. Ogita, and Y. Yoneda, *Jpn. J. Pharmacol.*, 1999, **80**, 103–109.
10. E. M. Blackwood and R. N. Eisenman, *Sci. (Washington, D. C., 1883-)*, 1991, **251**, 1211–1217.
11. M. Vita and M. Henriksson, *Semin. Cancer Biol.*, 2006, **16**, 318–330.
12. C. V Dang, *Cell (Cambridge, MA, U. S.)*, 2012, **149**, 22–35.
13. T. C. He, A. B. Sparks, C. Rago, H. Hermeking, L. Zawel, L. T. da Costa, P. J. Morin, B. Vogelstein, and K. W. Kinzler, *Science*, 1998, **281**, 1509–12.
14. L. E. Canne, A. R. Ferre- D'Amare, S. K. Burley, and S. B. H. Kent, *J. Am. Chem. Soc.*, 1995, **117**, 2998–3007.
15. A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff, and S. K. Burley, *Nat.*, 1993, **363**, 38–45.
16. R. V Talanian, C. J. McKnight, and P. S. Kim, *Sci. (Washington, D. C., 1883-)*, 1990, **249**, 769–771.
17. B. Cuenoud and A. Schepartz, *Sci. (Washington, D. C., 1883-)*, 1993, **259**, 510–513.
18. M. Ueno, A. Murakami, K. Makino, and T. Morii, *J. Am. Chem. Soc.*, 1993, **115**, 12575–12576.
19. A. M. Caamano, M. E. Vazquez, J. Martinez-Costas, L. Castedo, and J. L. Mascarenas, *Angew. Chem., Int. Ed.*, 2000, **39**, 3104–3107.
20. J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vazquez, and J. L. Mascarenas, *Nat. Commun.*, 2013, **4**, 1874–1878.
21. L. L. G. Carrette, T. Morii, and A. Madder, *European J. Org. Chem.*, 2014, **2014**, 2883–2891.
22. Y. Ruiz García, J. Zelenka, V. Pabon, A. Iyer, M. Buděšínský, T. Kraus, C. I. E. Smith, and A. Madder, *Org. Biomol. Chem.*, 2015.
23. Y. Ruiz García, A. Iyer, D. Van Lysebetten, Y. Vladimir Pabon, B. Louage, M. Honcharenko, B. G. De Geest, C. I. Edvard Smith, R. Strömberg, and A. Madder, *Chem. Commun.*, 2015, 17552–17555.
24. A. Jiménez-Balsa, E. Pazos, B. Martínez-Albardonedo, J. L. Mascareñas, and M. E. Vázquez, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 8825–8829.
25. J. Mosquera, M. I. Sánchez, M. E. Vázquez, and J. L. Mascareñas, *Chem. Commun.*, 2014, **50**, 10975-10978.
26. G. A. Bullen, J. H. R. Tucker, A. F. A. Peacock, J. P. Lee, S. C. Harrison, K. Struhl, J. Busa, D. Podhradský, J. Imrich, D. A. Holt, M. Gilman, and T. Clackson, *Chem. Commun.*, 2015, **51**, 8130–8133.
27. T. Morii, Y. Saimei, M. Okagami, Keisuke Makino, and Y. Sugiura, *J. Am. Chem. Soc.*, 1997, **119**, 3649–3655.
28. D. Verzele and A. Madder, *Eur. J. Org. Chem.*, 2013, **2013**, 673–687.
29. T. Vorherr, *PharmaChem*, 2008, **7**, 13–16.
30. X.-Z. Lai, Y. Feng, J. Pollard, J. N. Chin, M. J. Rybak, R. Bucki, R. F. Epand, R. M. Epand, and P. B. Savage, *Acc. Chem. Res.*, 2008, **41**, 1233–1240.
31. E. Virtanen and E. Kolehmainen, *Eur. J. Org. Chem.*, 2004, 3385–3399.
32. C. A. Bode, C. P. Muller, and A. Madder, *J. Pept. Sci.*, 2007, **13**, 702–708.
33. X. Li, J. Guo, J. Asong, M. A. Wolfert, and G. J. Boons, *J. Am. Chem. Soc.*, 2011, **133**, 11147–11153.
34. L. N. Feldborg, R. I. Jølck, and T. L. Andresen, *Bioconjug. Chem.*, 2012, **23**, 2444–2450.
35. L. Guerrini, E. Pazos, C. Penas, M. E. Vázquez, J. L. Mascareñas, and R. A. Alvarez-puebla, *J. Am. Chem. Soc.*, 2013, **135**, 10314–10317.
36. E. Pazos, C. Portela, C. Penas, M. E. Vazquez, and J. L. Mascarenas, *Org. Biomol. Chem.*, 2015, **13**, 5385–5390.
37. M. G. Simpson, M. Pittelkow, S. P. Watson, and J. K. M. Sanders, *Org. Biomol. Chem.*, 2010, **8**, 1181–1187.
38. S. Adhikary and M. Eilers, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 635–645.
39. S. Cawley, S. Bekiranov, H. H. Ng, P. Kapranov, E. A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A. J. Williams, R. Wheeler, B. Wong, J. Drenkow, M. Yamanaka, S. Patel, S. Brubaker, H. Tammana, G. Helt, K. Struhl, and T. R. Gingeras, *Cell*, 2004, **116**, 499–509.
40. C. Grandori, J. Mac, F. Siëbelt, D. E. Ayer, and R. N. Eisenman, *EMBO J.*, 1996, **15**, 4344–57.

View Article Online
DOI: 10.1039/C7CC01705G