

# Enzymatic Synthesis of Nucleic Acids with Defined Regioisomeric 2'-5' Linkages

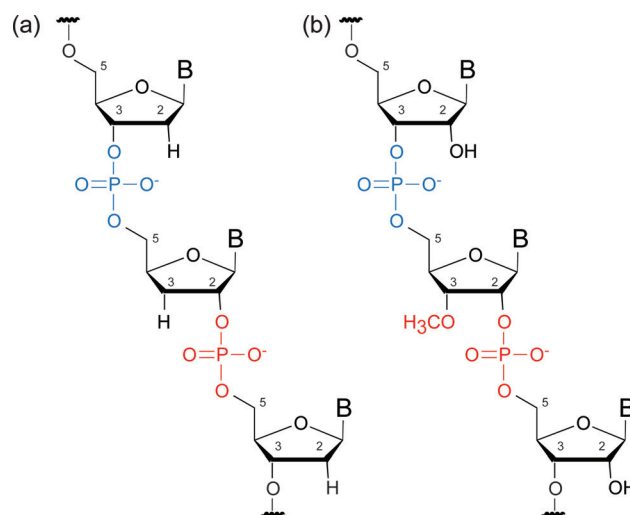
Christopher Cozens, Hannes Mutschler, Geoffrey M. Nelson, Gillian Houlihan, Alexander I. Taylor, and Philipp Holliger\*

**Abstract:** Information-bearing nucleic acids display universal 3'-5' linkages, but regioisomeric 2'-5' linkages occur sporadically in non-enzymatic RNA synthesis and may have aided prebiotic RNA replication. Herein we report on the enzymatic synthesis of both DNA and RNA with site-specific 2'-5' linkages by an engineered polymerase using 3'-deoxy- or 3'-O-methyl-NTPs as substrates. We also report the reverse transcription of the resulting modified nucleic acids back to 3'-5' linked DNA with good fidelity. This enables a fast and simple method for "structural mutagenesis" by the position-selective incorporation of 2'-5' linkages, whereby nucleic acid structure and function may be probed through local distortion by regioisomeric linkages while maintaining the wild-type base sequence as we demonstrate for the 10–23 RNA endonuclease DNazyme.

Genetic information storage and propagation in biology is based on nucleic acids with uniform 3'-5' phosphodiester linkages. In contrast, 2'-5' linked RNA oligoadenylates<sup>[1]</sup> and mixed 2'-5'/3'-5' dinucleotides (e.g. Gp(2'-5')Ap(3'-5'))<sup>[2]</sup> are involved in innate immune signaling but not genetic information transfer. Sporadic 2'-5' linkages occur in non-enzymatic RNA synthesis<sup>[3]</sup> and have been proposed to facilitate primordial RNA replication and evolution via transient duplex destabilization<sup>[3d]</sup> while retaining overall RNA folding and function.<sup>[4]</sup> Duplex destabilization by sporadic 2'-5' linkages has been examined in detail<sup>[5]</sup> and results from weakened Watson–Crick base pairing and base stacking caused by lateral displacement of the base and adoption of non-canonical C2' *endo* puckering.<sup>[6]</sup> Nevertheless, 2'-5' linked nucleic acids specifically hybridize with complementary 3'-5' RNA and 2'-5' RNA and DNA, albeit more weakly.<sup>[5a,d,7]</sup> Thus, despite their destabilizing influence on duplex structures, we reasoned that 2'-5' linked nucleic acids should be able to both encode and transmit genetic information to the canonical 3'-5' linked nucleic acids. Furthermore, their non-

canonical backbone conformations might be harnessed to expand the structural and functional space of nucleic acid ligands and enzymes.

2'-5' linkages are accessible through solid-phase synthesis, but this is time- and cost-intensive. Herein we report a rapid, inexpensive and scalable strategy for synthesizing defined 2'-5' linked DNA and RNA regioisomers by an engineered DNA polymerase capable of forming both canonical 3'-5' and non-canonical 2'-5' linkages in either DNA or RNA (Figure 1).



**Figure 1.** Structure of partially substituted a) 2'-5' DNA and b) 2'-5' RNA, with 3'-O-methyl groups as synthesized by polymerase TGLLK.

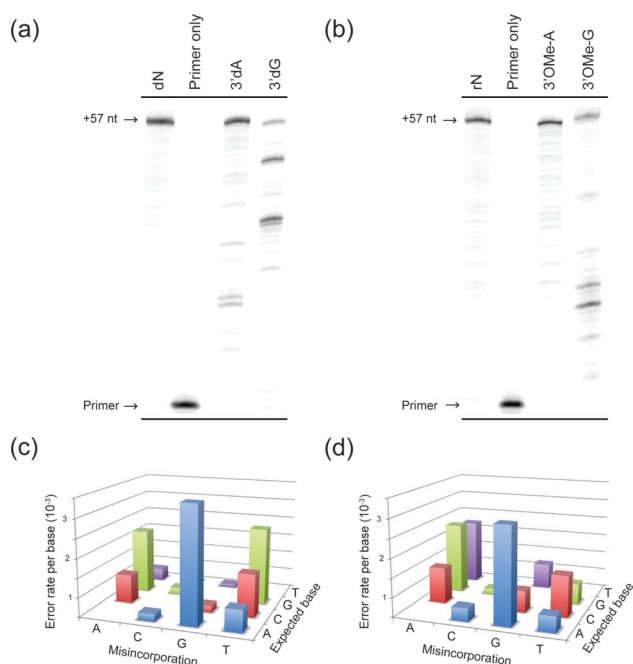
We first screened a panel of engineered<sup>[8]</sup> and commercial polymerases for DNA synthesis in which dGTP was completely replaced by 3'-deoxy-GTP (3'dGTP). The engineered polymerases TgoT<sup>[8b]</sup> (a Tgo mutant comprising V93Q,<sup>[9]</sup> D141A, E141A, A485L<sup>[10]</sup>) and TGK (TgoT: Y409G, E664K, previously described for primer-dependent RNA synthesis<sup>[8a]</sup>) as well as the related commercial polymerase Vent(*exo-*) were capable of full-length DNA synthesis with mixed 3'-5'/2'-5' linkages (Figure S1 in the Supporting Information). However, none of these efficiently synthesized RNA with mixed 3'-5'/2'-5' linkages when GTP was replaced by 3'-O-methyl-GTP (3'OMe-GTP). To this end, we prepared a TGK variant with two further mutations known to expand the polymerase substrate spectrum (I521L,<sup>[8b]</sup> F545L<sup>[11]</sup>) yielding polymerase TGLLK (TgoT: Y409G, I521L, F545L,

[\*] Dr. C. Cozens,<sup>[†]</sup> Dr. H. Mutschler,<sup>[†]</sup> G. M. Nelson, Dr. G. Houlihan, Dr. A. I. Taylor, Dr. P. Holliger  
Medical Research Council Laboratory of Molecular Biology  
Francis Crick Avenue, Cambridge, CB2 0QH (UK)  
E-mail: ph1@mrc-lmb.cam.ac.uk

[†] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201508678>.

© 2015 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



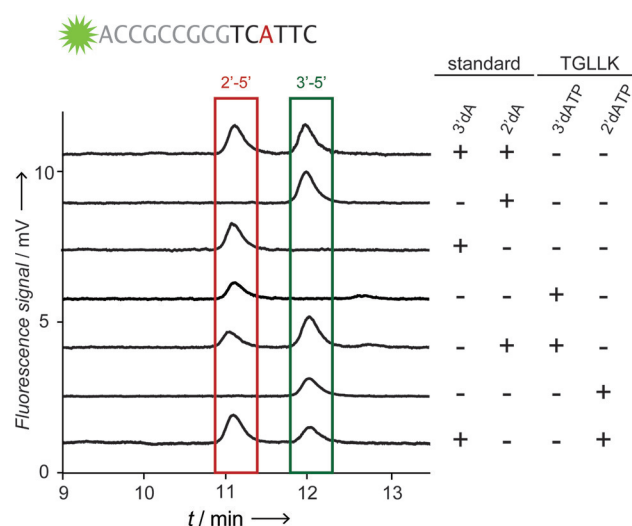
**Figure 2.** Enzymatic synthesis of partially substituted a) 2'-5' DNA b) 3'-5' RNA by TGLLK on a 57 nt template (TempN, see Supporting Information) encoding all possible dinucleotide combinations. Reactions in (a) involve dNTPs apart from 3'dA/G as indicated. Reactions in (b) involve NTPs apart from 3'OMeA/G as indicated. c), d) Error spectra of c) 3'dG/dHTP synthesis (aggregate misincorporation rate  $5.08 \times 10^{-4}$ ) and d) 3'OMe-A/BTP synthesis (aggregate misincorporation rate  $7.18 \times 10^{-4}$ ). The columns show the misincorporation frequency for each incorrect nucleotide.

E664K; Figure S2), which proved effective at synthesizing both DNA and RNA with defined 2'-5' linkages (Figure 2).

TGLLK is capable of fully replacing purine dNTPs and NTPs with their respective 3'deoxy (3'dATP, 3'dGTP) or 3'-O-methyl analogues (3'OMe-ATP, 3'OMe-GTP) during processive synthesis (Figure 2) even though these normally act as potent chain terminators (e.g. Cordycepin, 3'dATP<sup>[12]</sup>). Activity with 3'-analogues of pyrimidine nucleotides or a combination of any two nucleotide analogues was poor, although some full-length synthesis was possible with 3'dUTP or 3'OMe-ATP/3'OMe-GTP (Figure S3). Pausing is observed mostly at sequential 3'dNTP insertion sites, presumably because consecutive 2'-5' linkages weaken duplex stability and cause progressive conformational distortions relative to canonical 3'-5' helices.<sup>[6c,13]</sup>

To ensure that full-length synthesis is not the result of misincorporation utilizing only the dNTP substrates (as has been observed for some polymerases<sup>[14]</sup>), we sought to verify the presence of 2'-5' linkages using an HPLC assay, which distinguishes 2'-5' and 3'-5' phosphodiester linkages by mobility.<sup>[15]</sup> This assay unambiguously confirmed the ability of TGLLK to insert specific 2'-5' linkages into DNA using 3'dATP (Figure 3).

Reverse transcription of 2'-5' substituted DNA and 3'OMe-RNA back to all 3'-5' DNA using Taq DNA polymerase (Taq) and Avian Myeloblastosis Virus RT (AMV RT) respectively (as previously reported<sup>[16]</sup>) allowed us to deter-



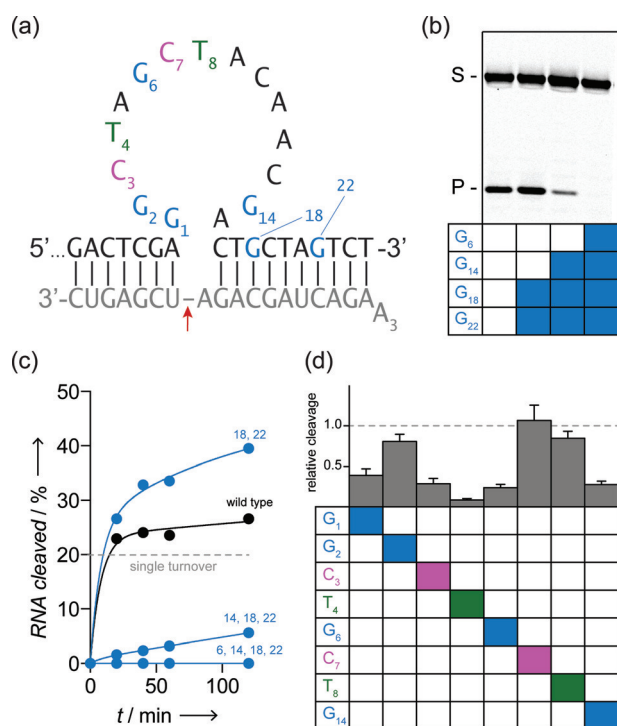
**Figure 3.** HPLC analysis of a fluorescent oligonucleotide that was synthesized by TGLLK either using 2'dATP or 3'dATP. Additional traces are corresponding chemically synthesized 2'-5' (3'dA)/3'-5' (2'dA) standards and an enzymatically synthesized 2'dA control. Standards and polymerase-synthesized oligonucleotides were mixed in equimolar ratio and separated by HPLC under denaturing conditions.

mine aggregate fidelity of the information transfer through 2'-5' linkages in both DNA and RNA by deep sequencing, revealing misincorporation frequencies ranging from  $8 \times 10^{-3}$  for 3'dATP to  $2 \times 10^{-4}$  for 3'OMe-GTP (Figure 2 c,d; Table S1).

Next, we set out to probe the impact of 2'-5' linkages on nucleic acid function. Limited, sporadic substitution (< 25%) of randomly distributed 2'-5' linkages is compatible with function in some ribozymes and aptamers as polyclonal populations.<sup>[4]</sup> We reasoned that a better understanding of the functional impact of 2'-5' linkages might be gained from site-specific insertion of regioisomeric linkages. Such a “structural mutagenesis” approach would not replace functional groups—like conventional mutagenesis—but instead alter their three-dimensional positioning, potentially allowing a novel interrogation of the roles of different nucleotides in nucleic acid structure and function.

We chose the well-studied 10–23 RNA-endonuclease DNazyme<sup>[17]</sup> (Figure 4a) to validate this “structural mutagenesis” approach and first synthesized 10–23 with progressively more 2'-5' linked purines (Figure 4 and Figure S4). This “primer scanning” approach allows a quick determination of obligatory 3'-5' linkages sensitive to distortion, neutral positions tolerating both 3'-5' and 2'-5' geometry, and any positions where 2'-5' introduction is beneficial. We identified nucleotide positions broadly falling into each category (Figure 4). For example, structural mutagenesis of the distal RNA-binding arm revealed that introducing 2'-5' linkages after G18 and G22 results in a modest activity gain (Figure 4b,c), presumably because of reduced product inhibition.<sup>[18]</sup>

Previous analysis had shown that G14 is highly sensitive to mutation as changes to A, C, or T result in around a 20-fold loss of activity and even conservative substitution of G with inosine or 2-aminopurine results in a greater than 10-fold



**Figure 4.** Structural mutagenesis of enzymatically synthesized 10–23 DNAzyme variants. a) 10–23 DNAzyme/substrate pair used in this study. b) Cleavage of the fluorescent RNA (200 nM) after 1 h at 37°C by different DNAzyme variants (40 nM) containing 2'-5' linkages downstream of G6 or G14 shown in panel (a). Blue boxes indicate 2'-5' linkages after 3'dG. c) Cleavage reaction progress of DNAzymes with increasing amounts of 3'dG substitutions. d) Cleavage after 1 h by different DNAzyme variants containing single 2'-5' linkages produced by position-selective enzymatic synthesis.

drop in activity.<sup>[19]</sup> In contrast, we find that structural mutagenesis of G14, whereby guanine functional groups are retained but a geometric perturbation of the active site is introduced,<sup>[20]</sup> results in a comparatively mild reduction in 10–23 activity (around fivefold; Figure 4b,c).

Primer scanning using 3'dA revealed modest activity gains when binding-arm positions were mutated (A21); inserting 2'-5' linkages after all As at once resulted in complete loss of activity (Figure S4). Restoring the 3'-5' linkage after A5 rescued limited activity and restoring them after A9, A11, and A15 resulted in further progressive activity gains.

To obtain a higher resolution picture of the effects of 2'-5' linkages on 10–23 activity, we investigated the effects of discrete 2'-5' linkages using a “pulse-chase” approach similar to the recently described position-selective labeling of RNA (PLOR<sup>[21]</sup>) method, which we adapted to a simple one-pot reaction using TGLLK polymerase. Briefly, primer, template, and TGLLK were incubated with a single 3'dNTP to introduce one defined 3' deoxynucleotide (pulse) and then synthesis was completed with unmodified dNTPs (chase), generating a position-selective 2'-5' linkage (Figure S5).

We inserted position-selective 2'-5' linkages after several nucleotides in the catalytic core (G1, G2, C3, T4, G6, C7, T8, and G14, Figure 4d) known to affect 10–23 function. Several positions (G2, C7, T8) were largely unaffected by insertion of

a 2'-5' linkage (>75% wild type activity), suggesting a high tolerance to conformational distortions, while others (G1, C3, G6, G14) displayed more substantial drops in activity (20–40% wild type activity). One position (T4) proved highly sensitive to insertion of a 2'-5' linkage (<10% wild type activity).

These data suggest that T4 and the region C3–G6 in general are sensitive to structural mutagenesis, as well as to base substitutions,<sup>[19]</sup> base deletions,<sup>[22]</sup> or replacement with abasic sites.<sup>[23]</sup> This may be because the essential Mg<sup>2+</sup> ion coordinated by the T4–A5 backbone linkage is displaced (Figure S6).<sup>[24]</sup> We also find position G2, which is highly intolerant to base substitution,<sup>[19]</sup> is largely insensitive to the backbone distorting effects of 2'-5' linkages. Thus, our results show that structural mutagenesis through site-specific insertion of 2'-5' linkages can probe novel aspects of nucleic acid structure and function and may be particularly useful for the identification of backbone-mediated metal-ion coordination sites.

In summary, we have described a novel engineered polymerase capable of template-dependent synthesis of DNA and RNA with regioisomeric 2'-5' linkages. In the case of RNA synthesis, these are currently associated with the concomitant insertion of a 3'O-methyl group. Future approaches may allow the incorporation of NTPs with a cleavable 3' group, whereby mild deprotection would yield “pure” 2'-5' RNA in a manner analogous to cyclic reversible termination sequencing technologies<sup>[25]</sup> or deacetylation of 2'-O-acetylated RNA.<sup>[15]</sup>

We also describe a rapid method for position-selective structural mutagenesis of nucleic acids, whereby structure, conformation, and activity may be altered through insertion of regioisomeric backbone distortions, or indeed any unnatural nucleotide. While currently limited by the inefficient incorporation of 3'deoxy and 3'O-methyl pyrimidines as well as multiple substitutions, the described methodologies allow rapid, template-directed synthesis and reverse transcription of nucleic acid polymers containing mixed 2'-5'/3'-5' backbone linkages. This procedure forms the foundation for future in vitro evolution experiments and a new approach for expanding the structural and functional repertoire of nucleic acid enzymes, ligands, and sensors.

## Acknowledgements

This work was supported by a FEBS long-term fellowship (H.M.) and by the Medical Research Council (program no. U105178804).

**Keywords:** DNAzyme · nucleic acids · nucleotides · polymerase · regioselectivity

**How to cite:** *Angew. Chem. Int. Ed.* **2015**, *54*, 15570–15573  
*Angew. Chem.* **2015**, *127*, 15791–15794

- [1] R. Hartmann, J. Justesen, S. N. Sarkar, G. C. Sen, V. C. Yee, *Mol. Cell* **2003**, *12*, 1173–1185.
- [2] A. Ablasser, M. Goldeck, T. Cavlar, T. Deimling, G. Witte, I. Rohl, K. P. Hopfner, J. Ludwig, V. Hornung, *Nature* **2013**, *498*, 380–384.
- [3] a) R. Lohrmann, L. E. Orgel, *Tetrahedron* **1978**, *34*, 853–855; b) A. Kanavarioti, L. F. Lee, S. Gangopadhyay, *Origins Life Evol. Biospheres* **1999**, *29*, 473–487; c) P. A. Monnard, A. Kanavarioti, D. W. Deamer, *J. Am. Chem. Soc.* **2003**, *125*, 13734–13740; d) J. W. Szostak, *J. Syst. Chem.* **2012**, *3*, 2.
- [4] A. E. Engelhart, M. W. Powner, J. W. Szostak, *Nat. Chem.* **2013**, *5*, 390–394.
- [5] a) T. P. Prakash, K. E. Jung, C. Switzer, *Chem. Commun.* **1996**, 1793–1794; b) T. L. Sheppard, R. C. Breslow, *J. Am. Chem. Soc.* **1996**, *118*, 9810–9811; c) P. A. Giannaris, M. J. Damha, *Nucleic Acids Res.* **1993**, *21*, 4742–4749; d) M. Wasner, D. Arion, G. Borkow, A. Noronha, A. H. Uddin, M. A. Parniak, M. J. Damha, *Biochemistry* **1998**, *37*, 7478–7486.
- [6] a) B. J. Premraj, N. Yathindra, *J. Biomol. Struct. Dyn.* **1998**, *16*, 313–328; b) L. Li, J. W. Szostak, *J. Am. Chem. Soc.* **2014**, *136*, 2858–2865; c) J. Sheng, L. Li, A. E. Engelhart, J. Gan, J. Wang, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3050–3055.
- [7] a) K. E. Jung, C. Switzer, *J. Am. Chem. Soc.* **1994**, *116*, 6059–6061; b) H. Robinson, K. E. Jung, C. Switzer, A. H. J. Wang, *J. Am. Chem. Soc.* **1995**, *117*, 837–838.
- [8] a) C. Cozens, V. B. Pinheiro, A. Vaisman, R. Woodgate, P. Holliger, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8067–8072; b) V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science* **2012**, *336*, 341–344.
- [9] M. J. Fogg, L. H. Pearl, B. A. Connolly, *Nat. Struct. Biol.* **2002**, *9*, 922–927.
- [10] A. F. Gardner, W. E. Jack, *Nucleic Acids Res.* **1999**, *27*, 2545–2553.
- [11] N. Ramsay, A. S. Jemth, A. Brown, N. Crampton, P. Dear, P. Holliger, *J. Am. Chem. Soc.* **2010**, *132*, 5096–5104.
- [12] a) V. D. Axelrod, R. M. Vartikyan, V. A. Aivazashvili, R. S. Beabealashvili, *Nucleic Acids Res.* **1978**, *5*, 3549–3563; b) D. C. Montefiori, R. W. Sobol, Jr., S. W. Li, N. L. Reichenbach, R. J. Suhadolnik, R. Charubala, W. Pfeleiderer, A. Modliszewski, W. E. Robinson Jr., W. M. Mitchell, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7191–7194.
- [13] B. J. Premraj, P. K. Patel, E. R. Kandimalla, S. Agrawal, R. V. Hosur, N. Yathindra, *Biochem. Biophys. Res. Commun.* **2001**, *283*, 537–543.
- [14] C. Gloeckner, K. B. M. Sauter, A. Marx, *Angew. Chem. Int. Ed.* **2007**, *46*, 3115–3117; *Angew. Chem.* **2007**, *119*, 3175–3178.
- [15] F. R. Bowler, C. K. Chan, C. D. Duffy, B. Gerland, S. Islam, M. W. Powner, J. D. Sutherland, J. Xu, *Nat. Chem.* **2013**, *5*, 383–389.
- [16] a) J. R. Lorsch, D. P. Bartel, J. W. Szostak, *Nucleic Acids Res.* **1995**, *23*, 2811–2814; b) S. Sinha, P. H. Kim, C. Switzer, *J. Am. Chem. Soc.* **2004**, *126*, 40–41.
- [17] S. W. Santoro, G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4262–4266.
- [18] M. J. Cairns, A. King, L. Q. Sun, *Nucleic Acids Res.* **2003**, *31*, 2883–2889.
- [19] Z. Zaborowska, J. P. Furste, V. A. Erdmann, J. Kurreck, *J. Biol. Chem.* **2002**, *277*, 40617–40622.
- [20] B. J. Premraj, S. Raja, N. Yathindra, *Biophys. Chem.* **2002**, *95*, 253–272.
- [21] Y. Liu, E. Holmstrom, J. Zhang, P. Yu, J. Wang, M. A. Dyba, D. Chen, J. Ying, S. Lockett, D. J. Nesbitt, A. R. Ferre-D'Amare, R. Sousa, J. R. Stagno, Y. X. Wang, *Nature* **2015**, *522*, 368–372.
- [22] Z. Zaborowska, S. Schubert, J. Kurreck, V. A. Erdmann, *Febs Lett.* **2005**, *579*, 554–558.
- [23] B. Wang, L. Q. Cao, W. Chiuman, Y. F. Li, Z. Xi, *Biochemistry* **2010**, *49*, 7553–7562.
- [24] B. Nawrot, K. Widera, M. Wojcik, B. Rebowska, G. Nowak, W. J. Stec, *FEBS J.* **2007**, *274*, 1062–1072.
- [25] J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D. H. Kim, M. Sano Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J. J. Russo, N. J. Turro, J. Ju, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9145–9150.

Received: September 16, 2015

Published online: November 3, 2015