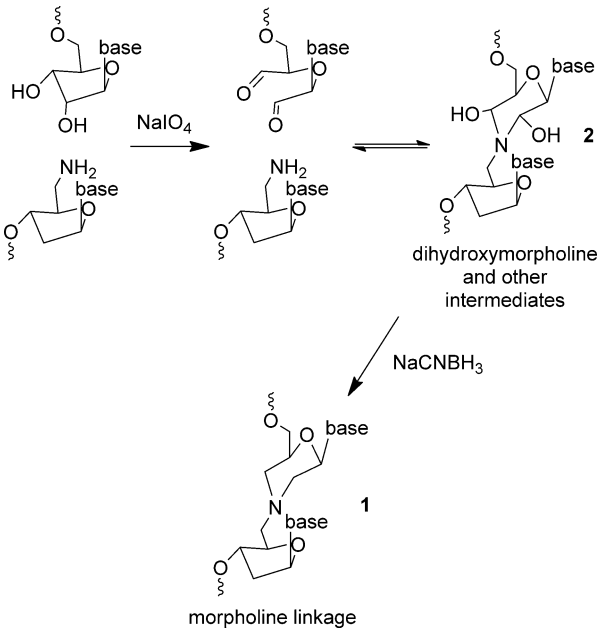
DOI: 10.1002/cbic.201200167

Nonenzymatic Ligation of DNA with a Reversible Step and a Final Linkage that Can Be Used in PCR

Aaron E. Engelhart,[a, c] Brian J. Cafferty,[a] C. Denise Okafor,[a] Michael C. Chen,[a] Loren Dean Williams,[a] David G. Lynn,[b] and Nicholas V. Hud*[a]

DNA is now used for purposes well beyond its natural information-carrying capacity (e.g., in nanotechnology, as a catalyst, and as a recognition element).[1] As a result, this polymer appears in applications as varied as sensors, therapeutic agents, and imaging technologies.[1e,2] Synthetic procedures for the nonenzymatic coupling of oligonucleotides that are compatible with self-assembly in aqueous solution, allowing selection of thermodynamically favored assemblies, have further expanded the possible uses of DNA.[3] However, most methods used for the nonenzymatic polymerization of nucleic acids lack a means to allow thermodynamic selection, involve synthetic procedures that are inaccessible to most laboratories, or result in formation of a polymer linkage that polymerase enzymes cannot read, thereby limiting the accessibility and utility of existing ligation methods. Here, we report a DNA ligation system that: 1) Is accessible using commercially available reagents; 2) includes a reversible step that can allow selection of a thermodynamically favored product; 3) proceeds in a template-selective mode at lower concentrations; 4) proceeds in an untemplated mode at higher substrate concentrations; and 5) produces a linkage that can be tolerated in a template strand by a number of thermophilic and mesophilic polymerases, allowing product sequence amplification by PCR.

We have investigated the DNA duplex and DNA polymerase compatibilities of the morpholine backbone linkage (**1**) that can be formed by the reaction of an oligonucleotide containing a 5′-amino residue with an oligonucleotide containing a periodate-oxidized 3′-ribose (Scheme 1). It has been known for some time that the 2′,3′-dialdehydes generated by periodate oxidation of ribonucleotides react in water with alkylamines to produce, upon reduction, a hydrolytically stable morpholine.[4] Wincott and co-workers have previously shown the utility of this amine-dialdehyde ligation reaction within an RNA hairpin loop,[5] and other investigators have used similar reactions to extend and conjugate chemical groups onto the ends of nucleic acids.[5–6] However, to the best of our knowledge, the effect of a morpholine linkage on DNA duplex stability and in a polymerase template strand has not been reported. The linkage investigated here differs substantially from the well-known phosphorodiamidate morpholino linkage[7] in that linkage **1** lacks a phosphorodiamidate group between the morpholine ring and the residue on its 3′ side (Scheme 1).



Scheme 1. Ligation of 3′-ribo-oligonucleotides and 5′-amino-oligonucleotides by morpholine ring formation.

To explore the potential for linkage **1** and the related unreduced linkage **2** (Scheme 1) to support duplex formation within otherwise naturally linked DNA, the pentamer d(GAGT)rC (**Ald5**) was incubated with $NaIO_4$ (to generate the 2′,3′-dialdehyde) in the presence of 5′-amino-d(TAAGC) (**Am5**) and the decamer d(GCTTAGACTC) (**Temp10**), which could serve as a template for the ligation of **Ald5** and **Am5**. These oligonucleotide lengths were chosen such that the majority of **Ald5** and **Am5** would hybridize with **Temp10** under the conditions of incubation (1 mM in each strand, 5 °C),[8] with the hybridized complex being further stabilized if either linkage **1** or **2** is compatible with a DNA duplex.

The stabilizing effect of linkage **2** is demonstrated by the fact that, in the absence of a reducing agent, **Ald5**, **Am5**, and **Temp10** form a complex, which, when diluted 500-fold immediately before analysis, exhibits a cooperative melting transition (T_m) of $\geq$20 °C (see Figure S1 in the Supporting Informa-

[a] *Dr. A. E. Engelhart, B. J. Cafferty, C. D. Okafor, M. C. Chen, Prof. L. D. Williams, Prof. N. V. Hud*
School of Chemistry and Biochemistry, Georgia Institute of Technology
Atlanta, GA 30332-0400 (USA)
E-mail: hud@chemistry.gatech.edu

[b] *Prof. D. G. Lynn*
Departments of Chemistry and Biology, Emory University
Atlanta, GA 30322-1003 (USA)

[c] *Dr. A. E. Engelhart*
Present address: NPP Fellow, NASA Astrobiology Institute
Center for Computational and Integrative Biology, Massachusetts General Hospital
Boston, MA 02114 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201200167: detailed methods, DNA sequences, sequencing results, thermal denaturation analyses, and ligation kinetics data.

tion for melting traces). The T_m of $\geq 20\,^{\circ}C$ is reported as a lower bound, because the complex did not reassemble after heating and cooling back to 5 °C. We note that the individual melting temperatures for unlinked **Ald5** and **Am5** with **Temp10** are predicted to be less than 0 °C in the conditions of the diluted sample.[8] Thus, our observation of a melting transition at 20 °C is fully consistent with the dissociation of a reversibly linked **Ald5**-**2**-**Am5** strand from the **Temp10** template, followed by the hydrolysis of **Ald5**-**2**-**Am5** back to unlinked **Ald5** and **Am5**.

To examine if the *irreversible* linkage **1**, formed upon the reduction of linkage **2**, is compatible with a DNA duplex, the reducing agent $NaCNBH_3$ was added to a solution containing **Ald5**, **Am5**, and **Temp10**, at a temperature and oligonucleotide concentrations (5 °C; 1 mM in each strand) for which the formation of **Ald5**-**2**-**Am5** had already been verified by melting studies. Following sufficient time for linkage reduction (i.e., 24 h) and dilution to a final concentration of 2 μM in each strand, thermal denaturation of this sample revealed a cooperative transition at 24 °C, which is only 3 °C less than the T_m observed for the duplex formed by **Temp10** with a complementary strand containing all phosphodiester linkages. Unlike the pre-reduction duplex, the hydrolytically stable **Ald5**-**1**-**Am5** showed duplex reformation with **Temp10** during sample cooling (Figure S1). These observations demonstrate that a single substitution of the reduced morpholine linkage is well tolerated within a DNA duplex.

We next examined the sequence specificity of this template-directed ligation reaction. The addition of a stoichiometric amount of **Temp10** to reaction mixtures containing 10 μM **Am5** and **Ald5** resulted in a 20-fold enhancement in the rate of **Am5**-**1**-**Ald5** formation relative to the rate of spontaneous **Am5**-**1**-**Ald5** formation that is observed at substrate concentrations of 10 μM and higher (Figure 1). When **Ald5** was replaced in the reaction with d(GAAT)rC (**Ald5MM**), a strand that contains a single mismatch when bound to **Temp10**, the addition of **Temp10** resulted in a **Am5MM**-**1**-**Ald5** formation rate essentially identical to that of a control reaction lacking a template strand (Figure 1). Additionally, in a competition experiment containing both **Ald5** and **Ald5MM** in the presence of **Temp10**, **Ald5** ligated with **Am5** with approximately sevenfold higher yield than **Ald5MM** after two days of reaction (Figure S2). When the competition experiment was performed without **Temp10**, a similar rate was observed for **Am5** ligation with **Ald5** and **Ald5MM** (i.e., the rate of untemplated substrate coupling at 10 μM substrate concentration) (Figure S2). When these reactions were carried out with **Am5**, **Ald5**, and **Ald5MM** at 1 μM in the absence of **Temp10**, no ligation products were detected after two days (Figure S2).

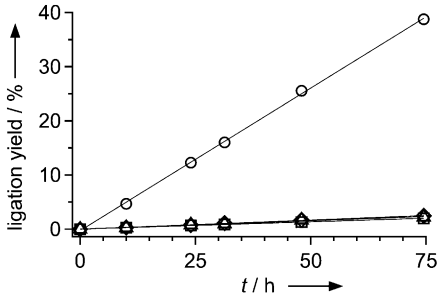


Figure 1. Comparison of morpholine-linked product formation rates in templated and untemplated reactions, with or without an internal mismatch in the dialdehyde substrate. **Ald5-1-Am5** with **Temp10** present (○), **Ald5MM-1-Am5** with **Temp10** present (△), **Ald5-1-Am5** untemplated (◇), **Ald5MM-1-Am5** untemplated (□). Reactions were at 0 °C, with oligonucleotides at 10 μM in strand. Oligonucleotides were reacted with $NaIO_4$ (2.5 mM) for 30 min before the addition of $NaCNBH_3$ (25 mM). See the Supporting Information for additional experimental details.

Having verified sequence selectivity by this ligation system in a template-directed reaction, we next determined whether the product of a ligation reaction could serve as a template for a naturally occurring DNA polymerase. For these experiments, a ribose-terminated 23-mer (**Ald23**) and a 5′-amino 49-mer (**Am49**) were prepared, along with oligonucleotides that could serve to hybridize to **Ald23** and **Am49**, bringing the ribose and amino termini in close proximity. Using these oligonucleotides and the chemical steps described above for aldehyde generation and linkage reduction, **Ald23**-**1**-**Am49** was generated. A number of thermophilic and mesophilic polymerases proved capable of extending a radiolabeled primer when annealed to the morpholine-containing **Ald23**-**1**-**Am49** (Figure 2).

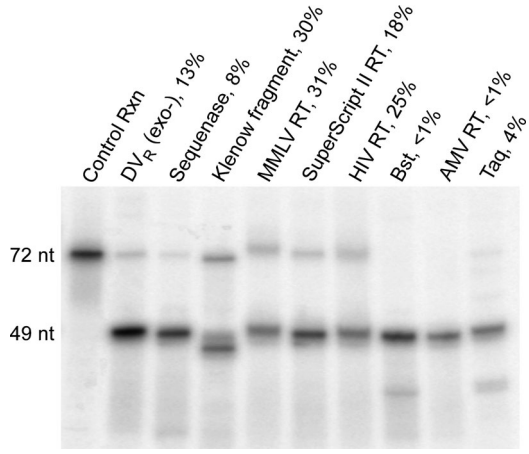


Figure 2. Autoradiograph of PAGE-separated products resulting from primer extension by selected enzymes with **Ald23**-**1**-**Am49** as a template. Yields of the full-length 72-mer product for each enzyme are given after enzyme names. Control Rxn is product of the all phosphate-linked template by DV_R (*exo*-). See the Supporting Information for experimental details.

This result was obtained irrespective of whether **Ald23**-**1**-**Am49** was generated in a templated reaction, using a full-length template (**Temp72**) or a shorter template (**Splint21**) that formed a 21-mer duplex between **Am49** and **Ald23**, or by the spontaneous higher-concentration ligation reaction of **Am49** and **Ald23** (i.e., in the absence of a template). When a mixture of **Ald23, Am49**, and **Splint21** was employed without periodate treatment (i.e., without dialdehyde generation), no 72-mer product was observed, indicating that overlap extension was not providing a false positive by isothermal overlap extension (Figure S3).

For all primer-extension reactions with **Ald23**-**1**-**Am49** as a template, the primer was not fully extended in 100% yield

for any polymerase. A significant amount of about 49 nt product was generated, consistent with all of the polymerases tested pausing, or even stopping, at or near the site of the morpholine linkage (Figure 2). Quantification of full-length product yield revealed a range of propensity to read through linkage **1**, from no detectable full-length product (Bst, AMV RT) to more than 25% (Klenow fragment, MMLV RT, HIV RT). The production of an appreciable fraction of abortive products is not necessarily problematic, however, given that, in most cases, the fraction of primer fully extended was at least 4%, and PCR could be used as a subsequent reaction to amplify the full-length product.

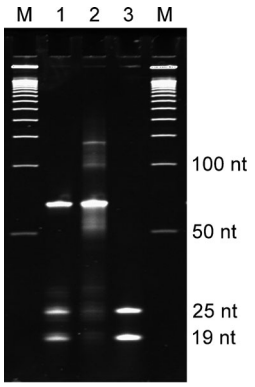


Figure 3. PAGE analysis of DV_R (*exo*-) amplifications of **Ald23-1-Am49** (lane 1), an all phosphate-liked template with a sequence identical to **Ald23-1-Am49** (lane 2), and control reaction with PCR primers and no template (lane 3). Marker bands (lane M) are 50 nt ladders. 25 and 19 nt bands represent PCR primers. Reactions contained 1 nM template, 2 μM of each primer, 200 μM dNTPs, 2×ThermoPol buffer. Thermal cycle: 4 min at 95 °C, 30 cycles of two-step PCR (30 s at 95 °C, 60 s at 72 °C), and a 5 min final extension at 72 °C. Template and primer sequences are provided in the Supporting Information.

One enzyme that produced considerable full-length product with the morpholine-containing **Ald23-1-Am49**, Deep $Vent_R$ (*exo*-) DNA polymerase (DV_R (*exo*-), New England Biolabs), is a high-fidelity thermophilic polymerase routinely used in PCR. Using DV_R (*exo*-) with **Ald23-1-Am49** and primers designed to yield a 64 bp product, a PCR reaction resulted in a one-pot readout-amplification reaction, as indicated by the observation of a product of expected length and a gel mobility identical to the PCR product of an all-phosphate-linked template of identical sequence to **Ald23–1-Am49** (Figure 3).

To investigate the fidelity of template read-through across the morpholine linkage, the 64 bp PCR product described above was subjected to a second round of PCR (using Phusion high-fidelity polymerase, New England Biolabs) that added the M13 17-mer forward and reverse primer sites for sequencing. The resulting 98 bp product was PAGE-purified and sequenced. The isolated dsDNA retained the initial **Ald23-1-Am49**-derived template sequence with no mutations, including at or near the morpholine linkage site (Figure S5). This result was obtained regardless of whether the original **Ald23-1-Am49** template strand was produced by template-directed ligation with **Splint21** or by spontaneous ligation, illustrating that both the untemplated and templated reaction are capable of generating templates suitable for PCR amplification.

Given the substantial difference between the structure of the morpholine linkage explored in the present study, **1**, and the natural phosphate-deoxyribose linkage (i.e., four atoms along the backbone linkage versus six, and removal of the phosphate charge), it was somewhat surprising to find that linkage **1** can be tolerated in a template strand by several natural polymerases. To gain insight into how linkage **1** is accommodated within a DNA duplex, molecular dynamics simulations were performed on a model duplex of **Ald5-1-Am5·Temp10** using the AMBER11 force field.[9] The initial conformation of the duplex was canonical B-form. The morpholine ring of linkage **1** was set to neutral charge (i.e., not protonated), as the analogous *N*-methylmorpholine has a pK_b value of 7.4,[10] suggesting that the morpholine linkage would be predominantly neutral in the ThermoPol buffer (pH 8.8). Helical parameters obtained with the CURVES+ program[11] for the structures adopted over more than 50 ns of simulation revealed that duplex **Ald5-1-Am5·Temp10** adopts two main structures during the simulation. In the first structure, residues near linkage **1** maintain a helical conformation similar to the B form; in the second, the duplex is locally unwound (Figures S6–S9). Base pairing is maintained to a similar extent in both conformations, and the calculated free energies and entropies of the two structures are within one standard deviation of each other (Supporting Information). These results support the observation that linkage **1** can be maintained in a B-form duplex and potentially assume alternative conformations as well.

Towards finding nonenzymatic aqueous DNA ligation chemistries that form linkages capable of being utilized by polymerases, Brown and co-workers recently demonstrated that a modified oligonucleotide containing a single neutral triazole linkage, a product of "click" Huisgen cycloaddition,[12] can be amplified by PCR and replicated in vivo.[13] In their development of ligation chemistries, Brown and co-workers observed that their first generation linkage suffered from nucleotide deletion at the linkage site after PCR amplification, a result that was interpreted as being symptomatic of a linkage that was too rigid.[14] This body of work by Brown and co-workers illustrates the importance of a modification's ability to structurally mimic that of the natural phosphodiester linkage. As suggested by our modeling studies, while linkage **1** does differ markedly from the canonical backbone linkage, it appears to, like the second-generation linkage of Brown and co-workers, be able to adopt a structure that is acceptable for use by a polymerase.

Finally, the polymerases used in our single primer extension studies showed various abilities to tolerate a morpholine linkage in a template strand (Figure 2). We see some correlations between our polymerase-screening study and those that have been previously reported for the reading of templates that contain other unnatural linkages. For example, Szostak and co-workers reported that MMLV RT, Sequenase and SuperScript II RT were among the polymerases with the greatest ability to read templates with sequential residues of threose-nucleic acid, or TNA, whereas the polymerase AMV RT was one of the least able to read TNA.[15] We observe the same grouping of these polymerases for their ability to read through a morpholine linkage. In contrast, the polymerase Bst, which was also reported to read a TNA template, did not show appreciable read-through of the morpholine linkage. In another study that utilized templates containing 2′,5′-linked DNA, Switzer and co-workers reported that HIV RT and Klenow (*exo*-) worked better

than AMV RT and Taq.[16] These results are also consistent with the ranking of these enzymes for toleration of a morpholine linkage. Taken together, these comparisons of polymerase activity with templates containing nonnatural linkages provide additional evidence that some polymerases are intrinsically more accepting of nonnatural linkages than others, and that some polymerases may be very selective regarding which unnatural linkages will be tolerated.

In conclusion, we have demonstrated that a morpholine backbone linkage, introduced into an otherwise natural DNA strand, can be tolerated in the template strand of a PCR reaction with a naturally occurring polymerase. This linkage can be generated in a one-pot reaction, using commercially available reagents, enhancing the general accessibility of this chemistry. Several features of the morpholine linkage system discussed here promise to make this chemistry, and related systems, useful for a wide range of applications. In particular, a covalent linkage that is formed in water and is reversible until reduced, provides the ability to select thermodynamically favored products during nucleic acid self-assembly.[3] Additionally, the enzyme-free nature of this ligation system has the potential to be applied in non-duplex regions of nucleic acids (e.g., in loops, triplexes, and G quadruplexes), such as during dynamic combinatorial reactions, followed by post-reduction read-through by a polymerase.

Acknowledgements

We thank Gary B. Schuster, Joshy Joseph and Elizabeth Kuruvilla for assistance with oligonucleotide synthesis. This research was supported by NASA Exobiology [NNX08A014G], the NSF-NASA Center for Chemical Evolution [CHE-1004570], and the NASA Astrobiology Institute.

Keywords: DNA ligation · dynamic combinatorial chemistry · nucleic acids · polymerase chain reaction · reversible covalent bond formation

[1] a) N. C. Seeman, *Trends Biochem. Sci.* **2005**, *30*, 119 – 125; b) F. A. Aldaye, A. L. Palmer, H. F. Sleiman, *Science* **2008**, *321*, 1795 – 1799; c) S. K. Silverman, *Angew. Chem.* **2010**, *122*, 7336 – 7359; *Angew. Chem. Int. Ed.* **2010**, *49*, 7180 – 7201; d) L. C. H. Poon, S. P. Methot, W. Morabi-Pazooki, F. Pio, A. J. Bennet, D. Sen, *J. Am. Chem. Soc.* **2011**, *133*, 1877 – 1884; e) E. J. Cho, J. W. Lee, A. D. Ellington, *Annu. Rev. Anal. Chem.* **2009**, *2*, 241 – 264.

[2] a) E. W. M. Ng, D. T. Shima, P. Calias, E. T. Cunningham, D. R. Guyer, A. P. Adamis, *Nat. Rev. Drug Discovery* **2006**, *5*, 123 – 132; b) M. N. Stojanovic, D. M. Kolpashchikov, *J. Am. Chem. Soc.* **2004**, *126*, 9266 – 9270.

[3] a) X. Li, A. F. Hernandez, M. A. Grover, N. V. Hud, D. G. Lynn, *Heterocycles* **2011**, *82*, 1477 – 1488; b) Z. Y. J. Zhan, D. G. Lynn, *J. Am. Chem. Soc.* **1997**, *119*, 12420 – 12421.

[4] a) B. S. Ermolinsky, S. N. Mikhailov, *Bioorg. Khim.* **2000**, *26*, 483 – 504; b) S. Kurata, T. Ohtsuki, T. Suzuki, K. Watanabe, *Nucleic Acids Res.* **2003**, *31*, e145; c) R. E. Hileman, K. M. Parkhurst, N. K. Gupta, L. J. Parkhurst, *Bioconjugate Chem.* **1994**, *5*, 436 – 444; d) J. Rinke, R. Brimacombe, *Mol. Biol. Rep.* **1978**, *4*, 153 – 156; e) J. A. Hunt, *Biochem. J.* **1965**, *95*, 541 – 551.

[5] L. Bellon, C. Workman, J. Scherrer, N. Usman, F. Wincott, *J. Am. Chem. Soc.* **1996**, *118*, 3771 – 3772.

[6] a) Z. F. Wang, L. Chen, S. F. Bayly, P. F. Torrence, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1357 – 1360; b) L. H. Zhou, E. R. Civitello, N. Gupta, R. H. Silverman, R. J. Molinaro, D. E. Anderson, P. F. Torrence, *Bioconjugate Chem.* **2005**, *16*, 383 – 390; c) N. Kojima, M. Sugino, A. Mikami, K. Nonaka, Y. Fujinawa, I. Muto, K. Matsubara, E. Ohtsuka, Y. Komatsu, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5118 – 5121; d) N. M. Bell, R. Wong, J. Micklefield, *Chem. Eur. J.* **2010**, *16*, 2026 – 2030; e) O. Y. Wong, P. I. Pradeepkumar, S. K. Silverman, *Biochemistry* **2011**, *50*, 4741 – 4749.

[7] J. Summerton, D. Weller, *Antisense Nucleic Acid Drug Dev.* **1997**, *7*, 187 – 195.

[8] J. SantaLucia, D. Hicks, *Annu. Rev. Bioph. Biom.* **2004**, *33*, 415 – 440.

[9] a) D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, *AMBER11*, University of California, San Francisco, **2010**; b) F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821 – 7839.

[10] H. K. Hall, Jr., *J. Am. Chem. Soc.* **1957**, *79*, 5441 – 5444.

[11] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, *Nucleic Acids Res.* **2009**, *37*, 5917 – 5929.

[12] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.* **2001**, *113*, 2056; *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.

[13] A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli, T. Brown, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11338 – 11343.

[14] A. H. El-Sagheer, T. Brown, *J. Am. Chem. Soc.* **2009**, *131*, 3958 – 3964.

[15] J. C. Chaput, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 9274 – 9275.

[16] S. Sinha, P. H. Kim, C. Switzer, *J. Am. Chem. Soc.* **2004**, *126*, 40 – 41.

Received: March 7, 2012
Published online on May 3, 2012

