DNA Nanotechnology Hot Paper

International Edition: DOI: 10.1002/anie.201502971
German Edition: DOI: 10.1002/ange.201502971

Enzymatic Method for the Synthesis of Long DNA Sequences with Multiple Repeat Units**

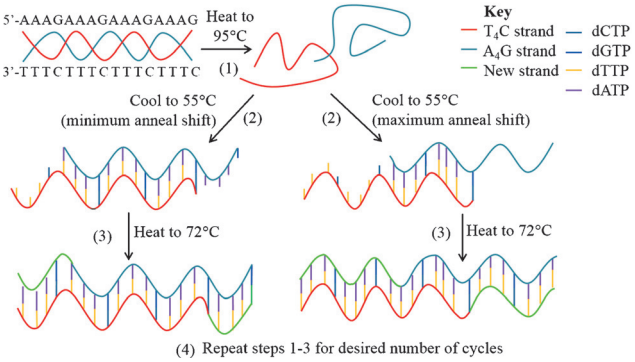
*Colette J. Whitfield, Andrew T. Turley, Eimer M. Tuite, Bernard A. Connolly, and Andrew R. Pike**

Abstract: *A polymerase chain reaction (PCR) derived method for preparing long DNA, consisting of multiple repeat units of one to ten base pairs, is described. Two seeding oligodeoxynucleotides, so-called oligoseeds, which encode the repeat unit and produce a duplex with 5′-overhangs, are extended using a thermostable archaeal DNA polymerase. Multiple rounds of heat–cool extension cycles, akin to PCR, rapidly elongate the oligoseed. Twenty cycles produced long DNA with uniformly repeating sequences to over 20 kilobases (kb) in length. The polynucleotides prepared include $[A]_n/[T]_n$, $[AG]_n/[TC]_n$, $[A_2G]_n/[T_2C]_n$, $[A_3G]_n/[T_3C]_n$, $[A_4G]_n/[T_4C]_n$, $[A_9G]_n/[T_9C]_n$, $[GATC]_n/[CTAG]_n$, and $[ACTGATCAGC]_n/[TGACTAGTCG]_n$, indicating that the method is extremely flexible with regard to the repeat length and base sequence of the initial oligoseeds. DNA of this length (20 kb ≈ 7 μm) with strictly defined base reiterations should find use in nanomaterial applications.*

The double-helical structure of DNA and the simple Watson–Crick base-pairing rules that determine helix formation[1] have led to its increasing use as a nanomaterial. Multiple applications have been reported, including nanoscale conducting wires,[2,3] in drug transport and target-driven medicine,[4,5] and as highly ordered one- and two-dimensional nanomaterials with functionalities at specific sites.[6] Developments in DNA origami[7,8] provide the ability to control the self-assembly of DNA into specific and intricate shapes with useful modifications in precise locations.[9]

Therefore, one key to using DNA in nanoscale engineering is the ability to prepare long sequences that consist of multiple repeat units, which is currently best achieved using the slippage reaction.[10–12] This provides a route towards the fabrication of long DNA that has a uniform repeat sequence covering many nanometers. Initial work demonstrated the extension of several two- and three-base repeats to a maximum of 700 base pairs (235 nm) by incubation with *Escherichia coli* DNA polymerase I.[10] A further advance showed that poly(dG).poly(dC) could be slipped to 10000 base pairs (3.4 μm) using the *E. coli* DNA polymerase I Klenow fragment.[13,14] More recent work investigated the range of sequences that could be extended by the slippage mechanism, but did not report on the lengths of the final products.[15] Commonly, the repeat unit must be able to form an internal loop for slippage to occur, limiting the number of reiterated polymers that can be prepared.[16] The level of control over the length of the DNA produced by the slippage approach relies on the quenching of the polymerase enzyme at a chosen time. These limitations prompted our investigations into an alternative method that affords DNA that consists of more complex repeat units, with lengths that can be controlled by the cyclical approach, common to the polymerase chain reaction (PCR).

PCR is routinely used in the biosciences to amplify long tracts of DNA.[17] However, products are usually limited to copies of naturally occurring DNA, and therefore, it is difficult to prepare polymers with repeating units. A general method that is capable of accurately extending short starting oligodeoxynucleotide repeat elements would be a valuable addition to the toolbox of DNA nanoengineers; such a protocol is outlined in Scheme 1. The starting materials are two short fully complementary oligodeoxynucleotides (ca. 20 bases). These oligoseeds encode the repeat element to be extended; in Scheme 1, the initial duplex has the sequence



Scheme 1. PCR-derived heat–cool cycles to extend a short oligoseed containing a repetitive element. Step 1: Melting of the duplex at 95 °C to form two single strands. Step 2: Annealing of complementary regions at 55 °C to give "shifted" duplexes with 5′ single-stranded extensions. Both a small (left) and a large (right) shift are illustrated; for the large shift, eight base pairs are considered as the minimum for stable duplex formation. Step 3: Incubation at 72 °C for the polymerase-catalyzed extension of the 5′ overhangs. Step 4: Repetition of steps 1 to 3 until the desired length of DNA is obtained.

[*] C. J. Whitfield, Prof. B. A. Connolly
Institute for Cell and Molecular Biosciences
University of Newcastle upon Tyne
Newcastle upon Tyne, NE2 4HH (UK)
A. T. Turley, Dr. E. M. Tuite, Dr. A. R. Pike
School of Chemistry, University of Newcastle upon Tyne
Newcastle upon Tyne, NE1 7RU (UK)
E-mail: andrew.pike@ncl.ac.uk

[**] This research was supported by the Biotechnology and Biological Sciences Research Council (BBSRC).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201502971.

Table 1: Theoretical minimum and maximum DNA extension products by repeat unit length ($n = 2$–10).

No. of cycles	$[AG]_{10}/[TC]_{10}$ ($n = 2$)		$[A_2G]_7/[T_2C]_7$ ($n = 3$)		$[A_3G]_5/[T_3C]_5$ ($n = 4$)		$[A_4G]_4/[T_4C]_4$ ($n = 5$)		$[A_9G]_2/[T_9C]_2$ ($n = 10$)	
	min.[a]	max.[b]	min.[a]	max.[b]	min.[a]	max.[b]	min.[a]	max.[b]	min.[a]	max.[b]
0	20	20	21	21	20	20	20	20	20	20
5	30	392	36	393	40	392	45	330	70	330
10	40	11152	51	11151	60	11152	70	10130	120	10130
15	50	36152	66	36141	80	36152	95	35130	170	35130
20	60	61152	81	61131	100	61152	120	60130	220	60130

The algorithm used to determine the theoretical minimum and maximum product lengths assumes that 1) at least eight base pairs must exist to give a duplex stable enough for elongation, that is, the minimal stable overlap is eight for $[AG]_{10}/[TC]_{10}$, nine for $[A_2G]_7/[T_2C]_7$, eight for $[A_3G]_5/[T_3C]_5$, and ten for $[A_4G]_4/[T_4C]_4$, and that 2) the polymerase completely fills any single-stranded regions. [a] Minimum theoretical length (base pairs) = original length + ($y \times n$). [b] Maximum theoretical length (base pairs) = original length $\times 2^y$ − [(min. stable overlap $\times 2^y$) − min. stable overlap], where n = number of base pairs in the repeat unit and y = number of cycles.

$[A_3G]_5/[T_3C]_5$, that is, a repeat of four base pairs ($n = 4$). Heating denatures the duplex, and cooling allows for re-hybridization. Although the most stable duplex will be the starting 20-mer, structures displaced by an integral value of the repeat length are possible. For the example given in Scheme 1, the minimum possible shift is four bases (1 × repeat length). The maximum displacement cannot be rigorously stated; however, assuming that at least eight base pairs are needed for stable duplex formation, a shift of eight bases (2 × repeat length) is a possibility. All shifted structures possess 5′ single-strand overhangs and, as such, are DNA polymerase substrates. Filling in by the polymerase extends the original seeds while exactly maintaining the repeat unit. Repeating the heat–cool cycles, akin to PCR, rapidly elongates the seeds. The theoretical lengths attainable are shown in Table 1, assuming that an overlap of at least eight base pairs is required to generate an elongation-competent duplex and that the polymerase is fully capable of filling the overlaps produced. If only a minimal shift takes place, the seeds grow in a slow linear fashion; however, with maximal displacement, rapid expansion is possible. The greatest theoretical length is unlikely to be achieved as maximal displacement at each stage is improbable and the polymerase may lack the capacity to fill the large gaps present in later cycles. Furthermore, selecting more than eight base pairs for stable duplex formation also reduces the lengths that may be realizable. Nevertheless, Table 1 clearly illustrates the potential of this method to generate long DNA segments with multiple repeat units. The heat–cool cycles that are necessary for oligoseed extension demand a thermostable DNA polymerase, and a *Thermococcus gorgonarius* family B polymerase (Tgo-Pol) mutant, Z3, was used.[18,19] Z3 is disabled in the 3′→5′ proof-reading exonuclease (exo−) and has an insertion in the fingers domain—mutations that confer the ability to read non-standard bases in the template strand.

To gain an idea of the maximum DNA lengths that could be achieved, extension was initially investigated with Tgo-Pol Z3 (exo−) and $[A_3G]_5/[T_3C]_5$ as the starting oligoseed. DNA of escalating size builds up as the number of heat–cool cycles increases, with products of over 20000 bases (20 kb) visible after 15 cycles. The maximum size range was generated after around 20–25 cycles, and further rounds of up to 40 cycles did not lead to longer products (Figure 1; see also the Supporting Information, Figures S1 and S2). The addition

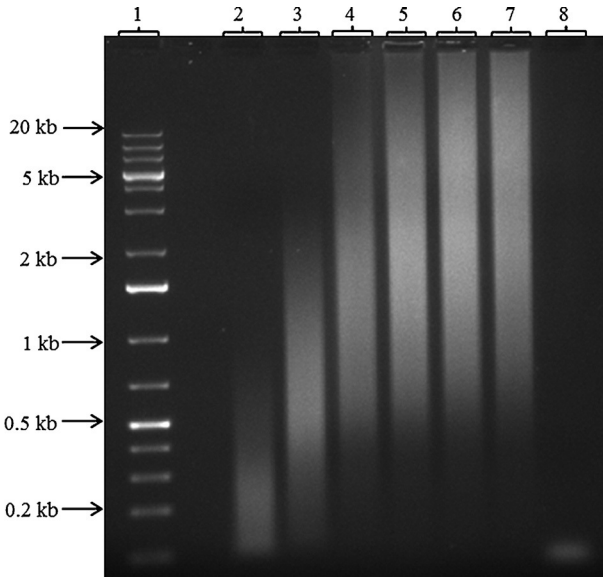


Figure 1. The extension products produced from an $[A_3G]_5/[T_3C]_5$ seed duplex during heat–cool cycling were analyzed on 0.3% SeaKem Gold Agarose gel. Lane 1: DNA ladder; lane 2: 5 cycles; lane 3: 10 cycles; lane 4: 15 cycles; lane 5: 20 cycles; lane 6: 25 cycles; lane 7: 30 cycles; lane 8: no polymerase (control).

of a fresh aliquot of 2′-deoxyribonucleoside triphosphates (dNTPs) also failed to increase product size, showing that the depletion or degradation of the dNTPs is not limiting the attainable length (Figure S3). An Image J analysis plot shows that the mean DNA product size increases between cycles 2 and 20 (Figures S4 and S5). Therefore, the ability to stop the reaction after any number of rounds affords some control over the length of the DNA produced. In all instances, the extension products appeared as an elongated smeared band covering a size range of approximately 0.5 to 20 kb. This dispersion undoubtedly arises from the multiple modes in which intermediate fragments can anneal, giving differently sized single-stranded extensions and, consequently, products of various lengths. The smallest size after 15 cycles, approximately 0.5 kb, is greater than the theoretical minimum, and as expected, the longest DNA segment (20 kb) did not reach the theoretical limit of 36 kb, but it is still of a suitable length for nanomaterial applications. However, access to DNA products with a narrower size distribution is required, and so a further gel-purification step was employed.

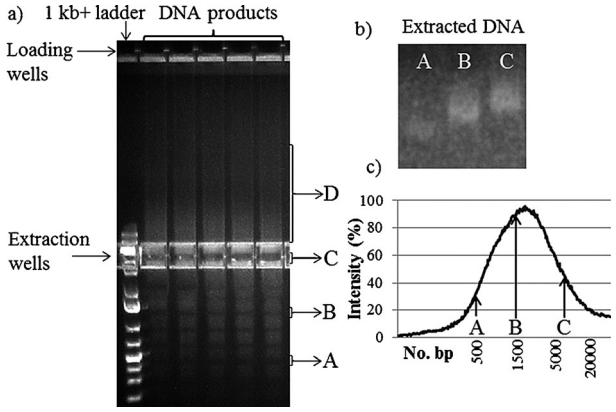


Figure 2. a) Lonza FlashGel DNA system depicting the DNA extraction process of samples A, B, and C and the unextracted DNA (D). b) Lonza gel analysis of the extracted DNA products A–C. c) Image J analysis of lane 4, Figure 1, showing the product length distribution as a percentage of the highest level of intensity in the lane.

The Lonza FlashGel DNA system enables the simple purification of DNA according to length[20] and was used to enrich products to a more limited size range. DNA with a size distribution of 0.5–20 kb was applied to multiple lanes of the Lonza gel, and selected aliquots of DNA were removed from the extraction wells (Figure 2). As an illustration, regions corresponding to DNA of 0.5 kb (sample A), 1.5 kb (sample B), and 7 kb (sample C; based on comparison with the standard ladder) were extracted from the gel. The successful removal of samples A and B is confirmed by the gaps in the elongated smeared band running beyond the extraction wells (Figure 2a). Sample C has just entered the extraction wells, and DNA at position D (> 10 kb) is pre-purification, still to reach the wells. The three samples were subsequently analyzed on a second Lonza gel, showing much tighter bands (Figure 2b). Before purification, the total amount of all extended DNA products from a typical starting oligoseed concentration of 6 ng μL^{-1} was estimated from the electronic absorption at 260 nm[21] and was typically 200 ng μL^{-1}. The amounts in the purified samples varied in line with the Image J analysis plot (0.5 kb, sample A: 9 ng μL^{-1}; 1.5 kb, sample B: 34 ng μL^{-1}; and 7 kb, sample C: 11 ng μL^{-1}; see Table S2). In each instance, the DNA products A, B, and C were characterized by atomic force microscopy (AFM) and fluorescence microscopy.

AFM confirmed the presence of ds DNA in samples A and B. A 5 μL drop of each sample was spotted onto a freshly cleaved mica surface, and the DNA was immobilized by molecular combing. Typical AFM images are shown in Figure 3a and b. The mean height of the DNA across all observed images was 0.7 ± 0.2 nm, which is in agreement with previous AFM reports of DNA duplex dimensions.[22,23] A length analysis was performed on each sample and indicated that the mean observed lengths, 175 ± 18 nm (0.5 kb) and 420 ± 55 nm (1.2 kb), were close to the DNA lengths expected for samples A and B, respectively (for details, see the Supporting Information).

Labelling of sample C with YOYO-1, a fluorescent intercalator,[24] allowed for spectroscopic characterization of the

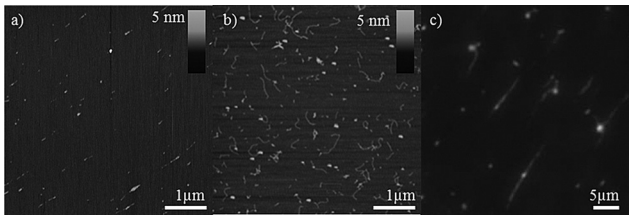


Figure 3. a) AFM image of an enriched sample A (500 bp) on mica. b) AFM image of an enriched sample B (1500 bp) on mica. c) Fluorescence microscopy image of an enriched sample C (7000 bp) on a glass slide.

enriched 7 kb band. The fluorescence micrograph in Figure 3c shows DNA strands that are 2.5–4 μm long, which corresponds to 7–12 kb. In, similar studies, other YOYO-1-labelled samples also appeared to be longer than expected owing to limitations of the optics of the fluorescence microscope (see Figure S9 and Table S5),[25] but in each case, the lengths of the observed DNA strands were in good agreement (± 1.6 μm) with the targeted extraction lengths.

Having established that the heat–cool cycling can extend the duplex $[A_3G]_5/[T_3C]_5$ (repeat length $n = 4$), alternative seeds were explored. The extension products observed with $[AG]_{10}/[TC]_{10}$ ($n = 2$), $[A_2G]_7/[T_2C]_7$ ($n = 3$), and $[A_4G]_4/[T_4C]_4$ ($n = 5$) as the primer templates are compared in Figures 4 and S10. These two-, three-, and five-base repeat units all gave extension products of comparable lengths to the four-base oligoseed and again showed efficient extension to more than 20 kb after 20 cycles. The products from the extension of the $[A_4G]_4/[T_4C]_4$ seed after five cycles were sequenced (Figure 4b, for more details see Figures S11 and S12), and 305 bases of AAAAG repeat units were identified, which correlates well with the 330 bp product expected after five cycles (see Table 1).

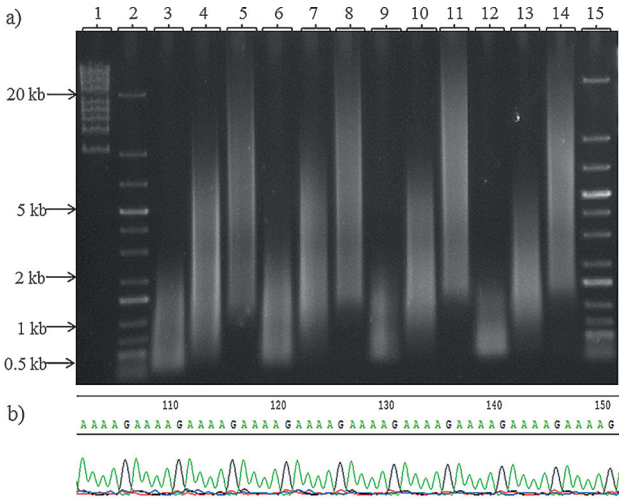


Figure 4. a) DNA extension products from $[AG]_{10}/[TC]_{10}$ (lanes 3–5), $[A_2G]_7/[T_2C]_7$ (lanes 6–8), $[A_3G]_5/[T_3G]_5$ (lanes 9–11), and $[A_4G]_4/[T_4C]_4$ (lanes 12–14) after 5, 10, and 20 cycles, respectively, were analyzed on 0.3% SeaKem Gold Agarose gel. Lanes 1, 2, and 15 show DNA ladders. b) A segment of the DNA sequencing data of the DNA extension product of $[A_4G]_4/[T_4C]_4$ after 5 cycles (Figure 4a, lane 12).

To increase potential applications, the diversity in the DNA repeat sequences was expanded. The simplest seed, the repeating polymer of $[A]_{20}/[T]_{20}$ ($n=1$), could be extended to give very long polymers (Figure S13) comparable to the equivalent product obtained by the slippage reaction.[13] The 20 bp oligoseed $[A_9G]_2/[T_9C]_2$ ($n=10$) could also be elongated, but yielded shorter products of up to approximately 2 kb. The ability to extend repeats of the sequence type $[A_xG]/[T_xC]$, where $x=1$ to 10, affords control over the base-pair density, that is, the ability to locate individual G:C pairs at defined distances from each other. Thermal cycling also succeeds with sequences that contain all four bases as shown with $[GATC]_5/[CTAG]_5$, ($n=4$) and $[ACTGATCAGC]_2/[TGACTAGTCG]_2$ ($n=10$; see Figure S14). Analysis of the DNA melting temperature outlined the expected increase in melting temperature from the oligoseed to the product (Figure S15).

The experiments described thus far have been carried out with Tgo-Pol Z3 (exo−), a mutated version with improved ability to copy non-standard bases.[18] Therefore, this polymerase is potentially useful for a longer-term aim, namely the incorporation of modified bases into repeating oligodeoxynucleotides for the preparation of future hybrid DNA nanomaterials. Nevertheless, all extensions described in this publication, which involve only the four standard bases, can be efficiently executed with a standard, commercially available archaeal DNA polymerase, and the method is not dependent on an exotic enzyme. This was demonstrated by comparing the extension efficiencies of the wild-type Tgo-Pol Z3 (exo−) DNA polymerase and the *Pyrococcus furiosus* DNA polymerase (Pfu-Pol exo−). After 10 cycles, the wild-type Pfu-Pol exo− polymerase was slightly more efficient than the Z3 mutant (Figure S16). However, the difference in efficiency between the DNA polymerases is minimal, and as Z3 is more accepting of modified bases, it shows promise for further applications. In a second comparison between the exo− and exo+ variants of Pfu-Pol, the exo+ variant was unable to extend the oligoseeds by the heat–cool method described here (Figure S17). Finally, this heat–cool cycling method was compared with the previously described isothermal slippage approach.[13] For oligoseeds of $n=2$ to 5, the length of the product produced was consistently higher when our cycling procedure was applied (Figure S18–S20), suggesting that this method, at least with archaeal exo− polymerases, is more suitable for the assembly of long repeating DNA sequences.

In summary, the elongation procedure described here provides a flexible method for preparing long DNA sequences with multiple repeating units, and is compatible with very many oligoseed sequences. This offers an alternative technique to the established isothermal slippage route towards long DNA. With both the cycling and slippage methods, the DNA product size distribution necessitates further gel enrichment. However, especially with the Tgo-Pol Z3 (exo−) polymerase, our method affords a route towards the introduction of multiple, uniformly spaced modified bases to give DNA with properties useful in nanomaterial fabrication, that could be used as metal-binding sites for the assembly of conducting nanowires.

Keywords: DNA · DNA structures · nanotechnology · polymerase chain reaction

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 8971–8974
Angew. Chem. **2015**, *127*, 9099–9102

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
[2] S. A. Farha Al-Said, R. Hassanien, J. Hannant, M. A. Galindo, S. Pruneanu, A. R. Pike, A. Houlton, B. R. Horrocks, *Electrochem. Commun.* **2009**, *11*, 550–553.
[3] G. I. Livshits et al., *Nat. Nanotechnol.* **2014**, *9*, 1040–1046.
[4] V. Bagalkot, O. C. Farokhzad, R. Langer, S. Jon, *Angew. Chem. Int. Ed.* **2006**, *45*, 8149–8152; *Angew. Chem.* **2006**, *118*, 8329–8332.
[5] E. S. Andersen et al., *Nature* **2009**, *459*, 73–76.
[6] T. Wang, D. Schiffels, S. Martinez Cuesta, D. K. Fygenson, N. C. Seeman, *J. Am. Chem. Soc.* **2012**, *134*, 1606–1616.
[7] P. Rothemund, *Nature* **2006**, *440*, 297–302.
[8] B. Wei, M. Dai, P. Yin, *Nature* **2012**, *485*, 623–626.
[9] A. C. Pearson, J. Liu, E. Pound, B. Uprety, A. T. Woolley, R. C. Davis, J. N. Harb, *J. Phys. Chem. B* **2012**, *116*, 10551–10560.
[10] C. Schlötterer, D. Tautz, *Nucleic Acids Res.* **1992**, *20*, 211–215.
[11] M. J. Hartenstine, M. F. Goodman, J. Petruska, *J. Biol. Chem.* **2000**, *275*, 18382–18390.
[12] J. Petruska, M. J. Hartenstine, M. F. Goodman, *J. Biol. Chem.* **1998**, *273*, 5204–5210.
[13] A. B. Kotlyar, N. Borovok, T. Molotsky, L. Fadeev, M. Gozin, *Nucleic Acids Res.* **2005**, *33*, 525–535.
[14] H. Klenow, I. Henningsen, *Proc. Natl. Acad. Sci. USA* **1970**, *65*, 168–175.
[15] L. M. Chi, S. L. Lam, *J. Phys. Chem. B* **2012**, *116*, 1999–2007.
[16] T. Kato, X. Liang, H. Asanuma, *Biochemistry* **2012**, *51*, 7846–7853.
[17] L. Raeymaekers, *Genome Res.* **1995**, *5*, 91–94.
[18] S. K. Jozwiakowski, B. A. Connolly, *ChemBioChem* **2011**, *12*, 35–37.
[19] B. D. Biles, B. A. Connolly, *Nucleic Acids Res.* **2004**, *32*, e176.
[20] M. Riley, H. White, FlashGel System for DNA Recovery, Resource Notes (App Note) **2009**, *7*, 17–20.
[21] A. V. Tataurov, Y. You, R. Owczarzy, *Biophys. Chem.* **2008**, *133*, 66–70.
[22] S. M. D. Watson, N. G. Wright, B. R. Horrocks, A. Houlton, *Langmuir* **2010**, *26*, 2068–2075.
[23] S. J. T. Van Noort, K. O. Van der Werf, B. G. De Grooth, N. F. Van Hulst, J. Greve, *Ultamicroscopy* **1997**, *69*, 17–27.
[24] M. L. Bennink, O. D. Schärer, R. Kanaar, K. Sakata-Sogawa, J. M. Schins, J. S. Kanger, B. G. de Grooth, J. Greve, *Cytometry Part A* **1999**, *36*, 200–208.
[25] T. F. Chan, C. Ha, A. Phong, D. Cai, E. Wan, L. Leung, P. Y. Kwok, M. Xiao, *Nucleic Acids Res.* **2006**, *34*, e113.

Received: March 31, 2015
Published online: June 10, 2015

