DNA Nanotechnology

DOI: 10.1002/ange.201206201

Combinatorial Displacement of DNA Strands: Application to Matrix Multiplication and Weighted Sums**

*Anthony J. Genot, Jonathan Bath, and Andrew J. Turberfield**

Programming interactions through DNA base sequence design underlies the use of synthetic oligonucleotides in molecular computing.[1] Toehold-mediated strand displacement[2] is often used to orchestrate hybridization reactions that compute. In this mechanism, a DNA strand is displaced from a duplex by an invading strand which hybridizes first to an overhanging single-stranded "toehold" domain. Toehold hybridization provides a thermodynamic driving force and accelerates strand displacement. The mechanism has applications in molecular computation,[2e,3] DNA-templated chemistry,[4] autonomous machinery,[5] detection of specific nucleic acid sequences,[6] and in the actuation of DNA structures.[7]

In conventional strand-displacement systems, the toehold and displacement domains are covalently linked during synthesis. However, particularly in large systems where it is necessary to control many competing interactions, it may be desirable to reprogram the function of some strands without having to re-synthesize the entire system;[2a] dynamic reprogramming of strand interactions may also add computational power. For example, DNA-templated chemistry applied to combinatorial drug discovery may require control of interactions within a large library of components.[8] DNA logic circuits that could classify gene expression, by searching for patterns specific to a pathology,[9] would rely on operations such as matrix multiplication in which, for an n-dimensional system, the number of intermediate computations grows as n^3.

Herein, we present a "combinatorial displacement" mechanism in which toehold and displacement domains are dynamically and combinatorially linked to form functional displacing complexes. This mechanism considerably reduces the number of strands that must be synthesized: $2n$ strands can be programmed to form complexes to invade n^2 substrates. Combinatorial displacement has a similar structure to matrix multiplication and is well suited to the implementation of linear operations. We demonstrate Boolean matrix multiplications and the computation of weighted sums using systems of DNA strands that grow only linearly with the number of inputs or outputs.

The principle of combinatorial displacement derives from the observation that the toehold domain and the displacement domain need not be adjacent, nor covalently linked.[2a,b] These domains can be carried by two separate strands,[2a,b] the toehold-domain and displacement-domain strands, which form a functional displacing complex when linked by hybridization of complementary linking domains (Figure 1). In our method, 17-nucleotide (nt) linking domains hybridize quickly (approximately $10^6\,\text{M}^{-1}\,\text{s}^{-1}$)[10] to form displacing complexes in which the toehold domain (10 nt) and displacement domain (16 nt) are separated by a flexible linker of two T residues and the diameter of the linking duplex (approximately 2 nm).[2b] A DNA design software package, NUPACK,[11] was used to design sequences that minimize the secondary structure and interactions between unrelated linking domains. We measured the kinetics of combinatorial displacement using reporter complexes (Figure 1): when the displacing complex invades the reporter the dual-labeled output strand is displaced and adopts a more compact conformation, in which its fluorescence is quenched (Figure S1). The displacing complex is capable of invading a complementary substrate through a toehold-mediated strand-displacement mechanism to form a three-way DNA junction at a rate of approximately $4\times10^4\,\text{M}^{-1}\,\text{s}^{-1}$ (Supporting Information, Figure S3).

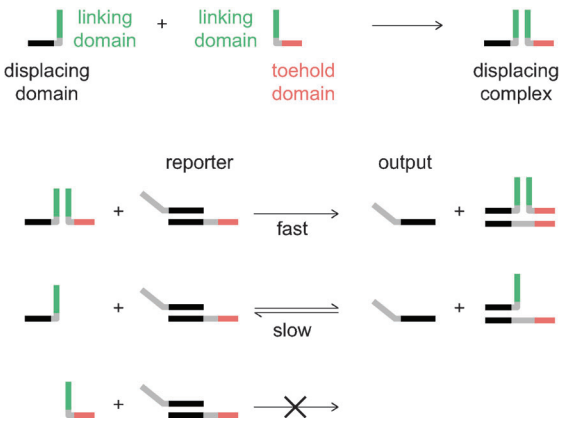

Figure 1. Combinatorial strand displacement. When the toehold and displacement domains are linked by hybridization of linking domains (green), they form a complex capable of invading a two-strand reporter complex and rapidly displacing an output strand. A strand having only a displacement domain (black) cannot displace the output rapidly, and a strand having only a toehold domain (red) cannot displace it at all.

[*] Dr. A. J. Genot, Dr. J. Bath, Prof. A. J. Turberfield
Oxford University
Department of Physics, Clarendon Laboratory
South Parks Road, Oxford OX1 3PU (UK)
E-mail: a.turberfield@physics.ox.ac.uk

[**] We thank N. Pierce and C. Steenberg for early access to the NUPACK designer and the DNA group in Caltech for helpful comments. This research was supported by EPSRC Grants EP/G037930/1, EP/P501709/1 and EP/P503876/1, a Royal Society Wolfson Research Merit Award (A.J.T.) and the Ecole Polytechnique and Merton College (A.J.G.).

Supporting information for this article (experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201206201.

Only a complete displacing complex, containing both toehold and displacing domains, is capable of rapid strand displacement (Figure S3); the leakage reaction of a displacement-domain strand with no associated toehold, or an incorrect toehold, is 400-fold slower ($100\,\text{M}^{-1}\text{s}^{-1}$; Figure S3).[10,12] The cooperativity of combinatorial displacement gives a very simple AND gate, in which the inputs are the toehold and displacement strands and the output is the strand being displaced.

Once formed, the displacing complex is stable on the experimental time scale.[10] Two displacing complexes with identical linking domains can exchange their domain strands through nucleation and migration of a four-way junction, but exchange is very slow in the presence of Mg^{2+} ions and in the absence of a toehold.[13] The linking duplex can be invaded by a strand with a complementary linking domain (blunt-end strand displacement): this process, involving nucleation and migration of a three-way branch junction, is faster, but the rate constant is still only approximately $1\,\text{M}^{-1}\text{s}^{-1}$.[10,12]

This modular system allows displacing complexes to be generated combinatorially. One toehold-domain strand can be linked to several different displacement-domain strands to create complexes capable of displacing different target strands. Similarly, displacement-domain strands can be linked to multiple toeholds. This gives rise to the combinatorial power of the mechanism: n toehold-domain strands and n displacement-domain strands can generate up to n^2 displacing complexes. Once a displacing complex is formed by hybridization of linking domains, the sequence of the linking domain has no further functional significance: displacing complexes with different linking domains, but identical toehold and displacement domains, can invade the same substrate. This property could be used to compute OR functions.

To demonstrate the power of combinatorial displacement, we computed the Boolean product of a matrix M and a vector X (Figure 2). The algorithm builds on previous observations that matrix multiplication and DNA hybridization share a similar structure (Figure 2 a,b).[14] The value of each variable is encoded by the presence (TRUE) or absence (FALSE) of the corresponding strand. (Matrices, vectors, and their elements, that is, mathematical symbols, are in Roman font, and the names of corresponding DNA strands and their component domains are italicized.) An element M_{ik} in row i and column k of matrix M is represented by a displacement-domain strand M_{ik} of structure $d_i b_k$, where d_i is a displacement domain and b_k a linking domain. An element X_{k} in the row k of the vector X is represented by a toehold-domain strand X_k of the form $\bar{b}_k t_1$, where $\bar{b}_k$ is a complementary linking domain and t_1 is a toehold domain. Matrix strands M_{ik} bind to complementary vector strands X_k to form displacing complexes $d_i b_k \bullet \bar{b}_k t_1$ that, together, represent the ith element of the output vector. These complexes displace output strands Y_i from the corresponding reporter complex (with domain structure $\bar{t}_1 \bar{d}_i \bullet d_i$), so the results of the computation can be read out by measuring the intensities of output strand fluorescence. This set of operations implements Boolean matrix multiplication: the concentration of complexes capable of displacing output Y_i is non-zero if and only if $\Sigma_k[M_{ik}][X_k] \neq 0$.

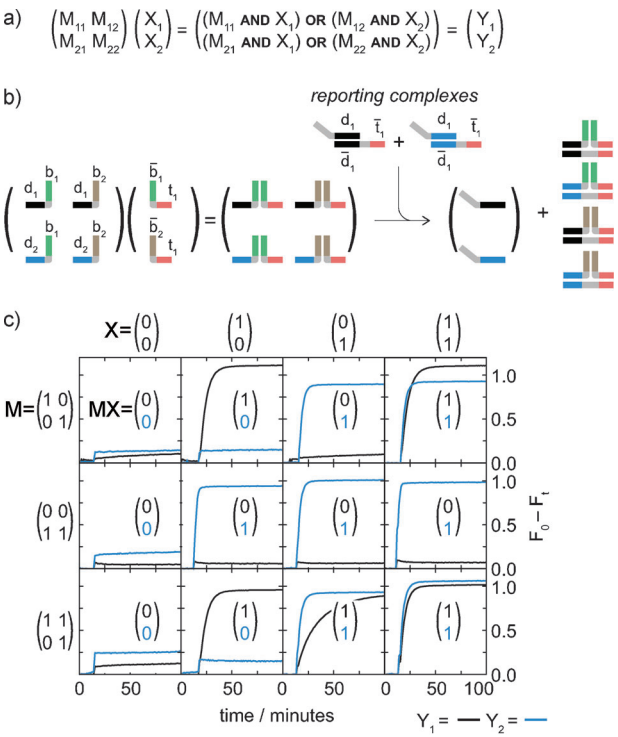

Figure 2. Boolean matrix multiplication. a) Rules of Boolean matrix multiplication. b) Implementation using combinatorial displacement. Matrix and vector elements are encoded by using the presence or absence of specific DNA strands to represent TRUE (1) and FALSE (0), respectively. Matrix elements are represented by displacement-domain strands, vector elements by toehold-domain strands. Matrix and vector strands bind to form displacing complexes whose concentrations are determined by the rules of Boolean multiplication. Reporter complexes incorporate output strands Y_1 and Y_2 which are covalently modified with a fluorophore and quencher, held apart by hybridization in the reporter complex: displacement of the output strand decreases their separation, leading to a decrease in fluorescence. c) Computation of 12 products. Reactions containing both reporter complexes (10 nM) were incubated for 10 min before addition of the matrix and vector strands required for the computation (66 nM). The vertical axis shows the variation of fluorescence F_0-F_t, which is proportional to the quantity of the output strand displaced. All traces from the same fluorescence channel are normalized by the same constant, so that full displacement of an output leads to a signal of approximately 1.

The algorithm can be generalized to multiply two matrices, using strands $d_i b_j$ for the left matrix M and strands $\bar{b}_j t_k$ for the right matrix N: $(\text{MN})_{\text{ik}}$ is represented by the sum of displacing complexes $\sum_j d_i b_j \bullet \bar{b}_j t_k$. The algorithm is economical: evaluation of the Boolean product of two $n \times n$ matrices requires evaluation of n^3 AND operations, but the number of strands needed for the computation scales as n^2. One fluorescent reporter complex is required for each element of the product vector or matrix. We focused on the multiplication of a 2×1 vector by a 2×2 matrix which requires only two fluorescence channels. To read out larger matrices, alternative methods such as DNA micro-arrays, gel electrophoresis,[14b] or plate readers should be considered.

Figure 2c shows the results of 12 separate experiments, corresponding to the multiplication of four vectors by three matrices. In these experiments, assembly of displacing complexes (computation) takes place one or two orders of magnitude more quickly than displacement of the fluorescent reporter strands (readout). The time for the reporter fluorescence signal to saturate ranges from 100 s to 1000 s. The strand displacement rate for the reporter complexes is strongly sequence dependent.[10] For example, in the calculation of $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$, the displacement of Y_2 by the complex $M_{22}{\cdot}X_2$ is 1.4-fold faster than displacement of Y_1 by $M_{11}{\cdot}X_1$ (Figure 2c). Competition between all matrix strands within a column for hybridization to the corresponding vector strand reduces the individual concentrations of the corresponding displacing complexes and thus influences the rates of their second-order displacement reactions. This competition is also sequence-dependent: although pairs of matrix strands M_{1k} and M_{2k} have identical linking domains, their rates of binding to X_k can be different as their different toehold domains introduce different secondary structures and interactions with X_k. Consider the product $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$: rate constants for binding of M_{12} and M_{22} to X_2 differ, with the result that three times as many complexes $M_{22}{\cdot}X_2$, capable of displacing Y_2, are produced as complexes $M_{12}{\cdot}X_2$, capable of displacing Y_1 (Figure S5). The combined effect of the different concentrations of displacing complexes and their different reaction rates is that the fluorescence change that reports the calculation of $M_{12}X_2$ is four times slower than that corresponding to $M_{22}X_2$. Competition between matrix strands was used to perform analog computations below.

The primary focus of molecular computation is digital computation, in which a Boolean input is encoded by the presence or absence of a specific input strand[15] or by the presence of one of two strands representing the possible values of the input.[3e] Digital computation of a function of one or more continuous variables is a formidable challenge.[14a] If high precision is not necessary, analog computation that directly processes DNA strand concentrations should be simpler, faster, and more robust.[14a] We next demonstrated an analog circuit which takes as inputs the concentrations of a set of n DNA strands X_k and returns an output strand Y_1 whose concentration is proportional to their weighted sum: $[Y_1]=\Sigma_k w_k[X_k]$. A weighted sum of digital inputs can be computed using, for example, seesaw gates;[3d] for continuous inputs, systems based on a damped catalyst and competitive hybridization have been proposed.[2f,16] We show how weights between 0 and 1 can be implemented using the competitive hybridization of displacement-domain strands. Consider the method for Boolean matrix multiplication described above applied to the multiplication of an n-component vector by a $2\times n$ matrix, where the concentrations of matrix strands are adjusted to correspond to the weights: $[M_{1k}]=A\,w_k$; $[M_{2k}]=A\,(1-w_k)$, where A is much greater than the concentration of any vector strand. If the competition between matrix strands M_{1k} and M_{2k} is unbiased (if the rate constants for the two reactions are the same) then the outputs of the matrix multiplication algorithm described above are represented by the following strand concentrations: $[Y_1]=\Sigma_k w_k[X_k]$; $[Y_2]=\Sigma_k(1-w_k)[X_k]$. In practice, it is necessary to adjust these concentrations to compensate for the different rates of binding to X_k of pairs of matrix strands M_{1k} and M_{2k} (Supporting Information).

Figure 3 shows the output of a bias-corrected circuit that calculates the weighted sum of the components of a specific two-vector. This weighted-sum circuit uses only one reporter,

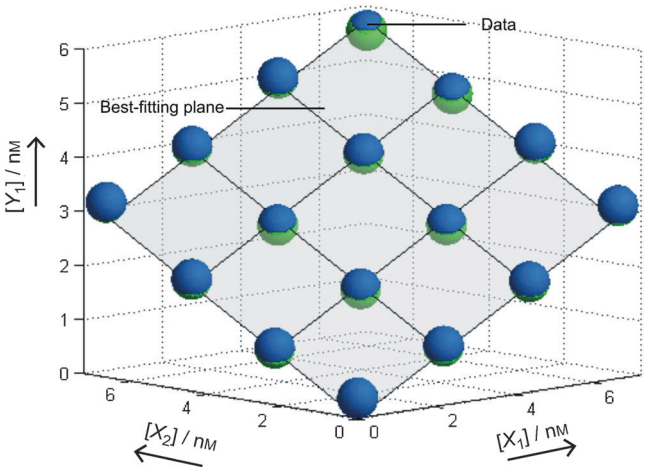

Figure 3. Calculation of a weighted sum, the average of inputs X_1 and X_2. The gray plane is the best fit to the measured output concentration $[Y_1]$ inferred from fluorescence measurements 11 h after adding the inputs and weight strands. [Reporter complex] = 10 nM, $[M_{1k}]+[M_{2k}]=16$ nM.

that accepts displacing complexes $d_1b_k \bullet \bar{b}_k t_1$ releasing strand Y_1. With both weights set to 0.5, a least-squares regression shows that the output concentration varies linearly with the input concentrations ($R^2=0.99$), and that the fitted weights ($w_1=w_2=0.45\pm10\%$) are within 10% of the programmed weights. Errors in the output of the circuit comprise measurement errors (including pipetting and fluorescence), calibration, and systematic errors. We estimate measurement and calibration errors to be approximately 15% (10% deviation from the programmed weights and 10% uncertainty on the weights themselves). There are systematic errors inherent in the design of the circuit: even in the absence of any input, weight strands can slowly displace the output strand from the substrate through blunt-end strand displacement.[10,12] The magnitude of this leakage can be estimated to be approximately 10% of the average circuit output, corresponding to the measured output for $[X_1]=[X_2]=0$ nM (Figure 3).

This circuit illustrates the typical trade-off between analog and digital circuits.[17] The physical errors of our analog circuit could be reduced, but not suppressed entirely, by careful calibration. In contrast, a digital circuit can compute with arbitrary precision but requires more resources than an analog circuit for low-precision computation, in part because of the overhead inherent in digital signal restoration.

We have introduced a mechanism that allows combinatorial generation of complexes capable of toehold-mediated strand displacement and can be used to perform matrix multiplication. This mechanism is well suited to implement

linear operations such as the calculation of weighted sums and pattern recognition.[9,18]

Received: August 2, 2012
Published online: December 3, 2012

Keywords: DNA recognition · DNA structures · molecular computation · nanotechnology · strand displacement

[1] a) *Biomolecular Information Processing*, Wiley-VCH, Weinheim, **2012**; b) Y. Benenson, *Nat. Rev. Genet.* **2012**, *13*, 455 – 468; c) N. Jonoska, N. C. Seeman, *Interface Focus* **2012**, *2*, 504 – 511.

[2] a) X. Chen, *J. Am. Chem. Soc.* **2012**, *134*, 263 – 271; b) A. J. Genot, D. Y. Zhang, J. Bath, A. J. Turberfield, *J. Am. Chem. Soc.* **2011**, *133*, 2177 – 2182; c) B. Yurke, A. P. Mills, *Genet. Prog. Evol. Machines* **2003**, *4*, 111 – 122; d) B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, J. L. Neumann, *Nature* **2000**, *406*, 605 – 608; e) D. Y. Zhang, *J. Am. Chem. Soc.* **2011**, *133*, 1077 – 1086; f) D. Y. Zhang, G. Seelig, *Nat. Chem.* **2011**, *3*, 103 – 113.

[3] a) B. M. Frezza, S. L. Cockroft, M. R. Ghadiri, *J. Am. Chem. Soc.* **2007**, *129*, 14875 – 14879; b) K. Oishi, E. Klavins, *IET Syst. Biol.* **2011**, *5*, 252 – 260; c) A. Phillips, L. Cardelli, *J. R. Soc. Interface* **2009**, *6*, S419 – S436; d) L. Qian, E. Winfree, *Science* **2011**, *332*, 1196 – 1201; e) L. Qian, E. Winfree, J. Bruck, *Nature* **2011**, *475*, 368 – 372; f) G. Seelig, D. Soloveichik, D. Y. Zhang, E. Winfree, *Science* **2006**, *314*, 1585 – 1588; g) D. Soloveichik, G. Seelig, E. Winfree, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5393 – 5398.

[4] a) Y. He, D. R. Liu, *J Am Chem. Soc* **2011**, *133*, 9972 – 9975; b) M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, R. K. O'Reilly, A. J. Turberfield, *J. Am. Chem. Soc.* **2012**, *134*, 1446 – 1449; c) M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, A. J. Turberfield, R. K. O'Reilly, *Angew. Chem.* **2010**, *122*, 8120 – 8123; *Angew. Chem. Int. Ed.* **2010**, *49*, 7948 – 7951; d) P. J. Milnes, M. L. McKee, J. Bath, L. J. Song, E. Stulz, A. J. Turberfield, R. K. O'Reilly, *Chem. Commun.* **2012**, *48*, 5614 – 5616.

[5] a) S. J. Green, J. Bath, A. J. Turberfield, *Phys. Rev. Lett.* **2008**, *101*, 238101; b) R. A. Muscat, J. Bath, A. J. Turberfield, *Nano Lett.* **2011**, *11*, 982 – 987; c) T. Omabegho, R. Sha, N. C. Seeman, *Science* **2009**, *324*, 67 – 71; d) A. J. Turberfield, J. C. Mitchell, B. Yurke, A. P. Mills, M. I. Blakey, F. C. Simmel, *Phys. Rev. Lett.* **2003**, *90*, 118102; e) S. F. J. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama, A. J. Turberfield, *Nat. Nanotechnol.* **2012**, *7*, 169 – 173; f) S. F. J. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama, A. J. Turberfield, *Nat. Nanotechnol.* **2011**, *6*, 166 – 169.

[6] a) H. M. Choi, J. Y. Chang, A. Trinh le, J. E. Padilla, S. E. Fraser, N. A. Pierce, *Nat. Biotechnol.* **2010**, *28*, 1208 – 1212; b) B. L. Li, A. D. Ellington, X. Chen, *Nucleic Acids Res.* **2011**, *39*, e110; c) J. M. Picuri, B. M. Frezza, M. R. Ghadiri, *J. Am. Chem. Soc.* **2009**, *131*, 9368 – 9377; d) D. Y. Zhang, S. X. Chen, P. Yin, *Nat. Chem.* **2012**, *4*, 208 – 214.

[7] a) F. A. Aldaye, H. F. Sleiman, *J. Am. Chem. Soc.* **2007**, *129*, 13376 – 13377; b) J. Elbaz, Z. G. Wang, F. A. Wang, I. Willner, *Angew. Chem.* **2012**, *124*, 2399 – 2403; *Angew. Chem. Int. Ed.* **2012**, *51*, 2349 – 2353; c) E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree, F. C. Simmel, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, E784 – E793; d) R. P. Goodman, M. Heilemann, S. Doose, C. M. Erben, A. N. Kapanidis, A. J. Turberfield, *Nat. Nanotechnol.* **2008**, *3*, 93 – 96; e) S. M. Douglas, I. Bachelet, G. M. Church, *Science* **2012**, *335*, 831 – 834.

[8] M. W. Kanan, M. M. Rozenman, K. Sakurai, T. M. Snyder, D. R. Liu, *Nature* **2004**, *431*, 545 – 549.

[9] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, T. R. Golub, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 15149 – 15154.

[10] D. Y. Zhang, E. Winfree, *J. Am. Chem. Soc.* **2009**, *131*, 17303 – 17314.

[11] a) J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, N. A. Pierce, *J. Comput. Chem.* **2011**, *32*, 170 – 173; b) J. N. Zadeh, B. R. Wolfe, N. A. Pierce, *J. Comput. Chem.* **2011**, *32*, 439 – 452.

[12] L. P. Reynaldo, A. V. Vologodskii, B. P. Neri, V. I. Lyamichev, *J. Mol. Biol.* **2000**, *297*, 511 – 520.

[13] I. G. Panyutin, P. Hsieh, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2021 – 2025.

[14] a) A. P. Mills, B. Yurke, P. M. Platzman, *Biosystems* **1999**, *52*, 175 – 180; b) J. S. Oliver, *J. Mol. Evol.* **1997**, *45*, 161 – 167.

[15] M. N. Stojanovic, D. Stefanovic, *Nat. Biotechnol.* **2003**, *21*, 1069 – 1074.

[16] a) H. W. Lim, S. H. Lee, K. A. Yang, J. Y. Lee, S. I. Yoo, T. H. Park, B. T. Zhang, *Biosystems* **2010**, *100*, 1 – 7; b) D. Zhang, G. Seelig in *DNA16* (Eds.: Y. Sakakibara, Y. Mi), Springer, Berlin, **2010**, pp. 176 – 186.

[17] R. Sarpeshkar, *Neural Comput.* **1998**, *10*, 1601 – 1638.

[18] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander, T. R. Golub, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2907 – 2912.

