DNA Nanotechnology

DOI: 10.1002/ange.201107156

A Molecular Cryptosystem for Images by DNA Computing**

*Sivan Shoshani, Ron Piran, Yoav Arava, and Ehud Keinan**

Image encryption by immobilized DNA molecules on chips could offer significant advantages, which include vast parallelism, immense information density, high chemical stability, and energy efficiency. Several theoretical or computer simulated models for encryption and steganography of texts have been proposed on the basis of DNA molecules,[1–4] and a few DNA-based models relevant to alphanumeric information have been realized.[5–10] DNA methods were also proposed for commercial encryption of signatures[11,12] and for information storage.[13–17] In contrast, image encryption has received very little attention.[18,19] Although DNA computing procedures were employed in theoretical schemes,[20–22] no molecular encryption of images has been tested experimentally, and certainly not by methods that involve molecular computing.

Herein, we report the use of parallel computing with molecular finite-state automata and fluorescently labeled DNA molecules for deciphering two different images. Both logos of the Technion and The Scripps Research Institute were encrypted on a single DNA chip. To decipher any of these images, a mixture of input molecules was processed by a molecular finite-state automaton, which led to image visualization by fluorescent, surface-bound output molecules. The huge diversity of potentially encrypted images that is offered by this strategy stems from current DNA microarray technology, with millions of printed pixels per chip, along with a plethora of possible input molecules and a variety of DNA-based automata.

Our approach to meet the challenge of image encryption was based on our two-symbol-two-state finite automata[23–27] as a mathematical model for information retrieval by using DNA molecules (Figure 1 A). We prepared two input molecules **In1** and **In2** in the form of a linear double-stranded (ds)DNA (Figure 1 B) that contain a string of symbols, which are six base pairs (bp) each. **In1** comprises the symbols **ab**, whereas **In2** contains the string **aa**. Each symbol-string was followed by a six bp terminator segment (**t**) and a long single-stranded (ss)DNA tail (48–56 bases). This tail was designed to be uniquely complementary to one of the connecting molecules **Con1** and **Con2**, to avoid self-hybridization, secondary structures, and cross-hybridization with incorrect connecting molecules. The connecting molecules, in the form of ssDNA, were intended to localize the output molecules on the microarray glass slide. Again, an important part of their design considered the need to prevent self-hybridization, secondary structures and cross-hybridization with incorrect molecules.

The two-symbol-two-state automata were programmed by the choice of subsets from the complete library of eight transition molecules. Each such molecule, which represents a transition rule, is comprised of three parts: a recognition site for the type II endonuclease *Fok*I, a four-base sticky end, and a spacer of between one and five bp between them (Figure 1 C). Automaton 1 (**A1**, Figure 1 A) includes transition molecules **TM2**, **TM3**, **TM6**, and **TM7**, whereas automaton 2 (**A2**), comprises **TM1**, **TM4**, **TM6**, and **TM8**. The two dsDNA detection molecules **DM0** and **DM1** (Figure 1 D) contain a four-base sticky end, which is complementary to one of the two four-base sticky ends that are generated upon restriction of the terminator domain at the final step of the computation.

Each deciphering mixture was generated by a computing process, which was carried out by mixing both input molecules **In1** and **In2**, the appropriate set of transition molecules (software), and a buffered solution that contained adenosine triphosphate (ATP), *Fok*I, and ligase (hardware). Mixing the above components results in an autonomous cascade of chemical events that includes repetitive cycles of DNA restriction, hybridization, and ligation. Restriction of the input by *Fok*I represents reading an input symbol and determining the new internal state of the automaton. Each of the six bp symbols, as well as the terminator, can be cleaved by this endonuclease either at the beginning of the symbol or two bp deeper into that domain. The two restriction modes represent the two internal states, S_0 and S_1 respectively. Each mode is dictated by the distance between the recognition site of *Fok*I and the cleavable symbol, as instructed by the newly incorporated **TM**. After all symbols are digested, the six bp terminator segment is restricted either at its beginning or two bases deeper, which creates one of two possible four-base sticky ends. One sticky end, 5′-ATAC, represents the final state S_0, whereas the other, 5′-ACAT, represents S_1. Thus, each of these sticky ends can ligate to one of the detection molecules **DM0** or **DM1**. For example, processing of **In1** with

[*] S. Shoshani, Dr. R. Piran, Prof. E. Keinan
Schulich Faculty of Chemistry, Technion–Israel Institute of Technology, Technion City, Haifa 32000 (Israel)
E-mail: keinan@technion.ac.il
Prof. Y. Arava
Faculty of Biology, Technion–Israel Institute of Technology (Israel)
Prof. E. Keinan
Department of Molecular Biology and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)

[**] This work was supported by the National Science Foundation under Grant No. 0523928 and the Russell Berrie Nanotechnology Institute (RBNI). We also thank the Israel-US Binational Science Foundation and the Skaggs Institute for Chemical Biology. S.S. thanks the Irwin and Joan Jacobs Foundation, the Fine Foundation, the RBNI, and the Ministry of Science and Technology for graduate school fellowships. E.K. is incumbent of the Benno Gitter & Ilana Ben-Ami Chair of Biotechnology, Technion. We thank Dr. Miriam Kott-Gutkowski, Head of the MicroArray Service Laboratory in the Core Reasearch Facility at the Faculty of Medicine—Ein Karem, Hebrew University of Jerusalem.

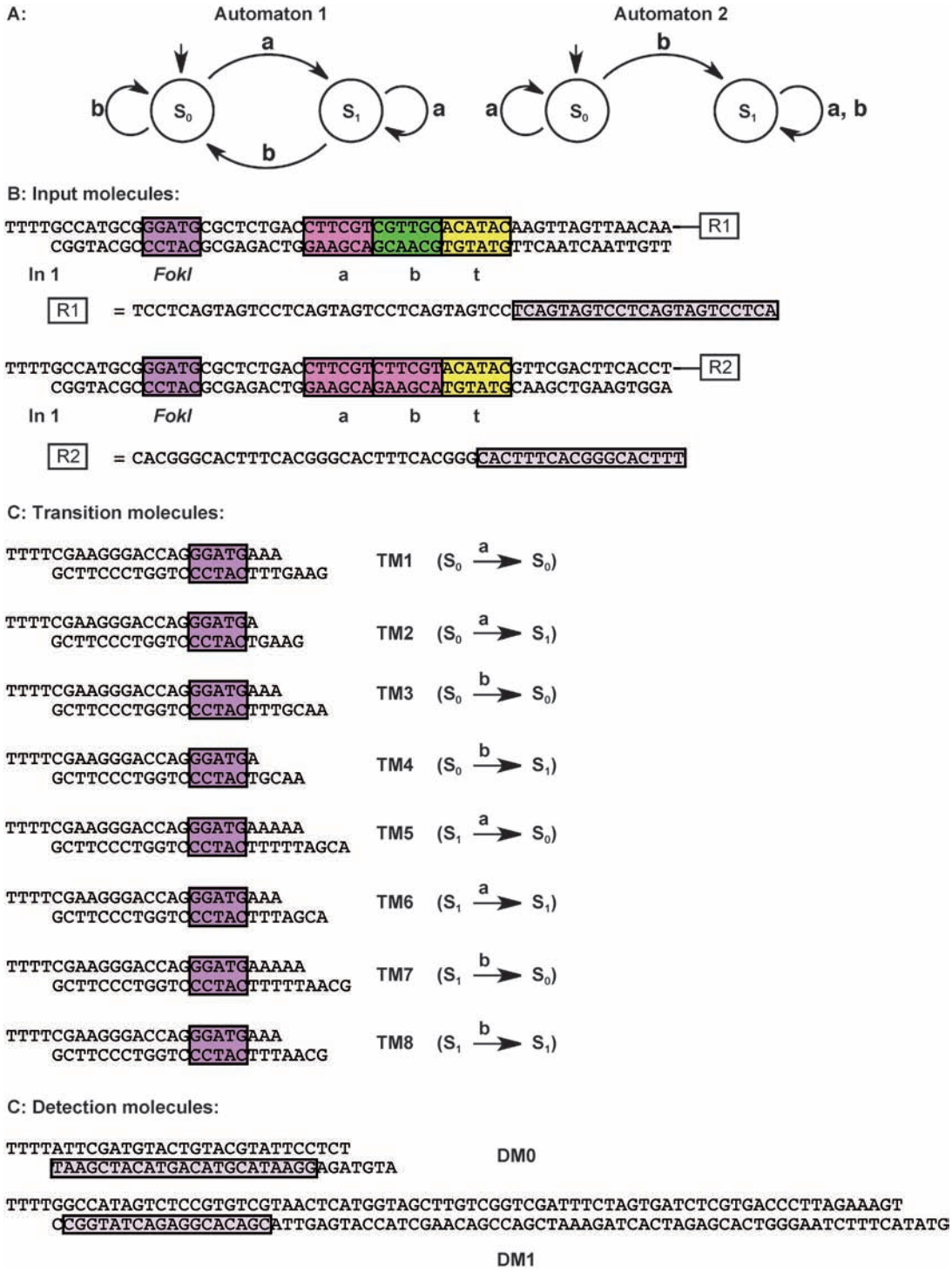

Figure 1. A) The two-symbol-two-state finite automata, the straight arrow represents the initial state. B) The input molecules with the area that is complementary to the primer being highlighted in gray. C) The transition molecules each represent one transition rule (shown in parentheses). D) The detection molecules with the area that is complementary to the primer being highlighted.

A1 (Figure 2) produced a molecule that was ligated to the appropriate detection molecule **DM0**.

To amplify the output signal and label it with a visible tag, the resulting DNA mixture was amplified by asymmetric PCR by using a fluorescently labeled primer.[28] One of the two primer binding sites is located on each detection molecule, whereas the second is located on each input molecule near to the terminator segment (Figure 1 B). Consequently, having two inputs and two detection molecules requires four unique primers, which are all designed to prevent false amplification. Employment of a mixture that contained all four primers assures the amplification of any possible output that results

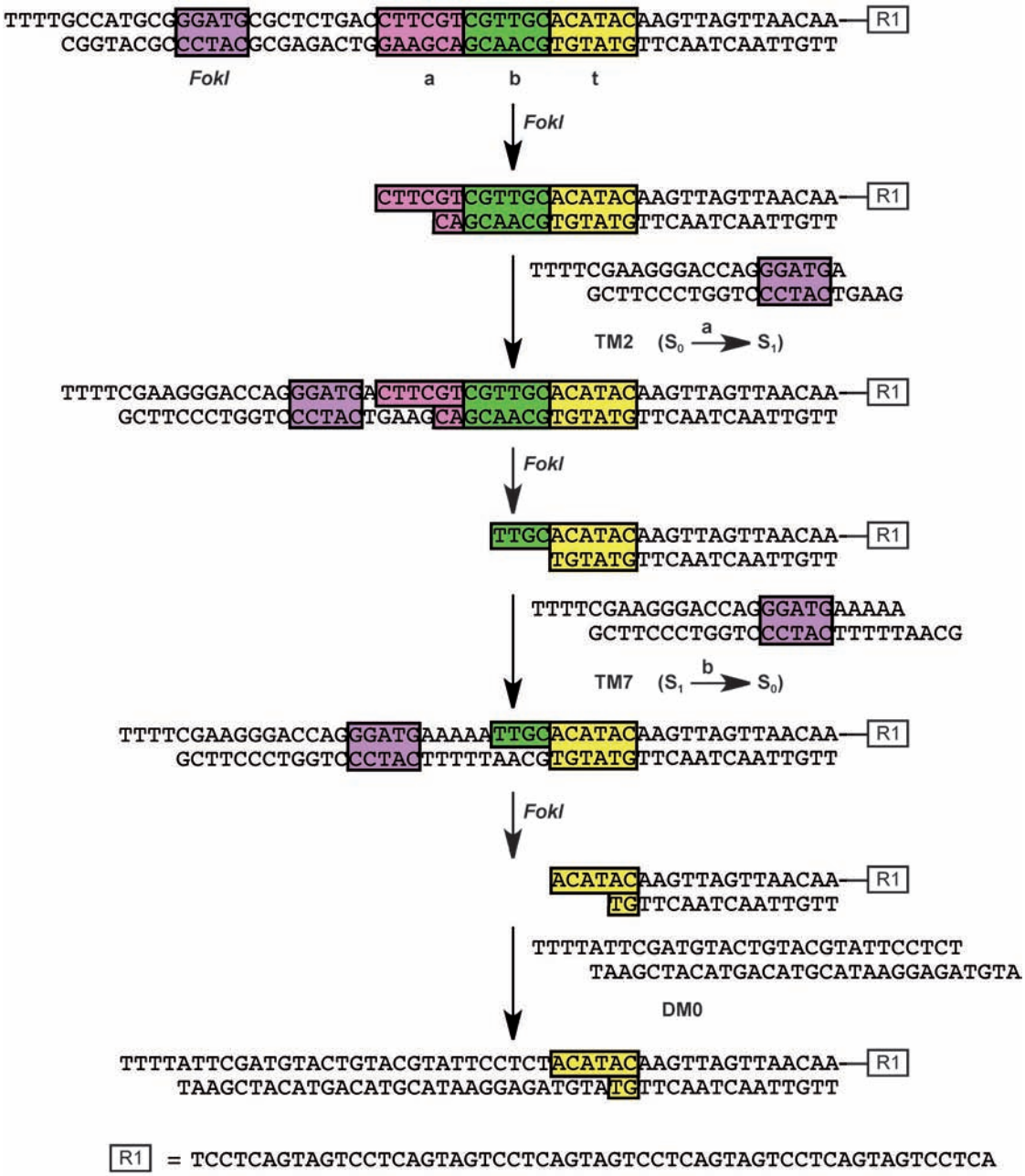

Figure 2. Details of the computing cascade with **In1** and **A1**. In the first event, *Fok*I restricts the **a** symbol outside of this segment, which represents S_0. Ligation of the truncated input with the appropriate transition molecule **TM2** switches the automaton to S_1. Another restriction inside the **b** segment represents S_1. Ligation to **TM7** represents switching to S_0. Finally, restriction at the terminator segment creates an appropriate sticky end for ligation to the correct detection molecule **DM0**.

from hybridization of a digested input with a detection molecule. Tagging the primer that is complementary to **DM1** with a green fluorescent dye results in fluorescent labeling of all computation outputs that represent the final state S_1.

We encrypted two overlapping images, the logos of the Technion and The Scripps Research Institute, on a single DNA chip that has 74×74 dots (5476 pixels). Each pixel contains either one of the two connecting molecules or a mixture of both (Figure 3). To decipher any of these images, a mixture of both input molecules was subjected to the computation described above with one automaton. Addition of the resultant mixture to the encrypted chip led to hybridization between all output molecules and the appropriate surface-bound connecting molecules. A laser scanner visualized the resultant green fluorescent image. The use of the first automaton **A1** revealed the logo of the Technion, whereas **A2** deciphered the logo of The Scripps Research Institute (Figure 4).

In principle, the computation step could be performed with surface-bound input molecules[24,29,30] rather than with soluble inputs and subsequent hybridization to surface-bound connectors. Our design, which is based on the latter approach

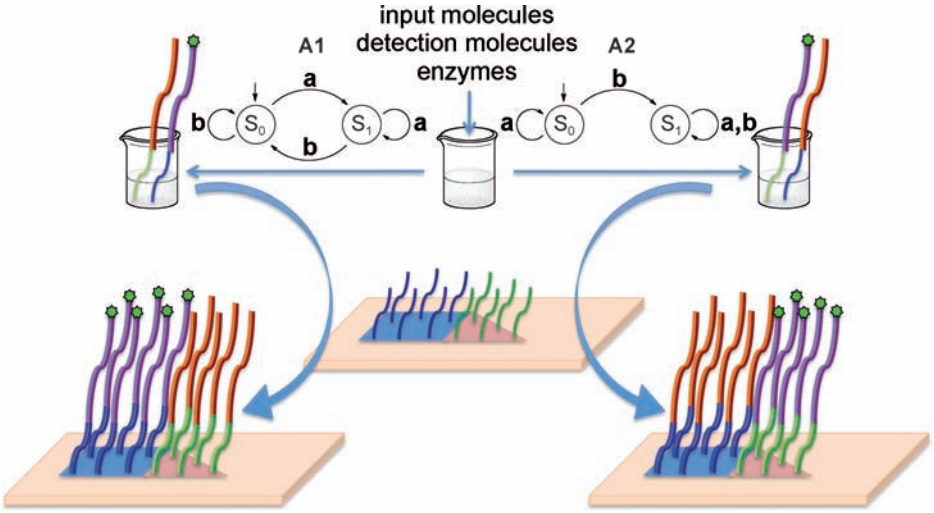

Figure 3. The general strategy for deciphering the encrypted images. The encrypted images are represented by the connecting molecules (blue and green, center). A given mixture of inputs (top center) is processed by either **A1** (top left) or **A2** (top right) to produce fluorescently labeled outputs that hybridize with either the blue or green connecting molecules (bottom left and right), respectively.

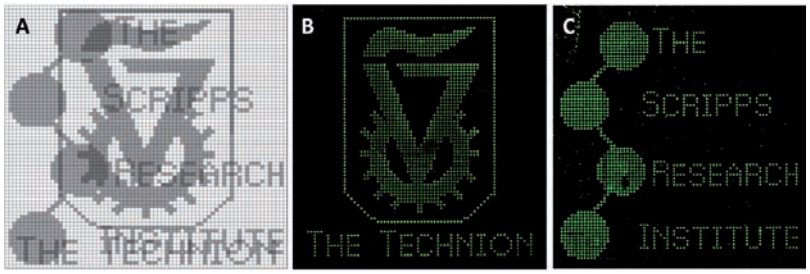

Figure 4. A) Planned coencryption of two images on a single DNA chip. Pixels showing the Scripps logo in gray (bound to **Con1**), those of the Technion logo in diagonal parallel lines (bound to **Con2**), and those representing both are in dotted areas (bound to equal proportions of **Con1** and **Con2**). Actual outputs from deciphering with B) **A1** and C) **A2**. The round black defects are the result of air bubbles on the chip.

and uses homogeneous mixtures, affords significant advantages. This practice enhances the encryption complexity because deciphering is dependent on three elements, each of which has its own diversity: the automaton mixture, the mixture of input molecules, and the chip-bound connecting molecules. Furthermore, since both endonuclease and ligase work more efficiently in homogeneous solution where all of the substrates are freely accessible to these enzymes, this practice offers improved kinetics, reaction yields, fidelity, and image resolution.

In conclusion, the technology described above is a powerful method for image encryption by parallel computing with all input molecules processed in the same solution by the same automaton. Even with the currently employed two final states, represented by green and black in the images, and with the relatively small array of 5476 pixels, the number of different images that may be accommodated on the chip can be 2^{5476} or 10^{1648}. Considering the fact that current microarray technology allows for printing millions of pixels on a single chip, the numbers of possible images would become very large. These numbers could be further increased if more than two inputs with an equal number of connecting molecules were employed. Moreover, for the construction of an image in full color, three different dyes are sufficient, thus, this enhances the diversity of potential images. Increasing the number of different automata can further enhance this diversity. Another dimension of diversity can be achieved by applying different reaction conditions, which include buffer, temperature, and pH, as well as the identity of enzymes employed. The fact that multiple images can be encrypted on the same chip can be used for a higher level of encryption, in which each individual image is meaningless but their linear combination represents the complete picture.

On the basis of current technology, it would be extremely difficult to decipher the sequence of a single DNA input on a specific pixel. This task becomes essentially impossible upon reinforcing the encryption by methods of steganography, such as diluting the input by additional inputs or nonrelevant DNA molecules that contain either irrelevant or false information. Even if the sequences of all possible input molecules became available to an attacker, it would be impossible to decipher their specific location on the chip in the absence of the appropriate automaton. Future work along these lines may lead to powerful cryptosystems at the interface between molecular computing and DNA microarray technologies.

Experimental Section

Materials. Bovine serum albumin (BSA), sodium dodecyl sulfate (SDS), NaCl, sodium citrate, and ATP were all purchased from Sigma–Aldrich.

General molecular biology techniques. Annealing of complementary strands was performed by heating a solution of both strands to 95 °C for 10 min and then slowly cooling to room temperature. Enzymes were inactivated by heating their solution to 85 °C for 25 min and then slowly cooling to room temperature.

Inputs, transition molecules, and detection molecules. All oligonucleotides were purchased from Syntezza Biosciences (Israel) and Integrated DNA Technologies (IDT). The input molecules **In1** (**ab**) and **In2** (**aa**), transition molecules **TM1**–**TM8**, detection molecules **DM0** and **DM1**, and molecular connectors **Con1** and **Con2** were all either dsDNA or ssDNA, as shown in Figure 1.

Computing procedure. Computation mixtures were prepared by mixing **In1**, **In2**, **DM0**, and **DM1** with the appropriate selection of transition molecules. **A1**, comprised four transition molecules: **TM2**, **TM3**, **TM6**, and **TM7**, whereas **A2**, comprised a different combination

of four molecules: **TM1**, **TM4**, **TM6**, and **TM8**. These mixtures were used for the computing procedure as follows: **In1** and **In2** (5 pmol each) were mixed with the transition molecules (15 pmol each), ATP (12 μL of a 10 mM stock solution in double distilled water (DDW), *Fok*I (12 units, New England BioLabs(NEB) and T4 DNA ligase (120 units, NEB) in Buffer 4 (NEB) in a total volume of 120 μL. The mixtures were incubated at 18 °C for 50 min, then the enzymes were immediately inactivated and the mixture was purified with a QIAquick Gel Extraction Kit (Qiagen) before dissolving the purified residue in 30 μL of DDW.

Each computation mixture underwent asymmetric PCR amplification with a set of four primers. The two primers **ForDM0** (5′-ATTCGATGTACTGTACGTATTCC) and **ForDM1** (5′-Cy3-GCCATAGTCTCCGTGTCG) were used in 30-fold excess over the other two primers **RevI1** (5′-TGAGGACTACTGAGGACTACTGA) and **RevI2** (5′-AAAGTGCCCGTGAAAGTG). The amplification was carried out by mixing the computation mixture (0.5 μL) with PCR Master Mix (Thermo Scientific, 12.5 μL) and all four primers: **RevI1** and **RevI2** (0.15 μM each), **ForDM0** and **ForDM1** (4.5 μM each), in a total volume of 25 μL. These mixtures were heated to 95 °C for 2 min, then underwent 17 repeated cycles of PCR (heating to 94 °C for 20 s, cooling to 58.5 °C for 30 s, heating to 72 °C for 1 min, and heating to 72 °C for 10 min).

Hybridization to the microarray.

Pre hybridization. The glass slides were incubated for 1 h at 42 °C in a preheated, prehybridization solution that contained 5X saline-sodium citrate (SSC) buffer (20X SSC buffer was prepared from 3 M NaCl and 0.3 M sodium citrate), SDS (0.1 %), and BSA (0.1 mg mL^{-1}). The slides were washed twice for 5 min in 0.1X SSC buffer at room temp. The slides were then washed in DDW for 30 s at room temperature and dried by centrifugation at 1750 rpm for 2 min.

Hybridization. Hybridization solution was comprised of formamide (20 %), 5X SSC, SDS (0.1 %), sheared salmon sperm DNA (1 mg mL^{-1}), and the computation solution (4.75 μL) in a total volume of 20 μL. The hybridization solution was heated to 70 °C for 5 min, then centrifuged at 13000 rpm for 2 min. The hybridization solution (16 μL) was placed onto the printed area, which was then covered by a cover glass. Pronto! solution (20 μL, Corning) was placed on random spots on the slide, which was then incubated at 42 °C for 6 h.

Post hybridization. The slide was incubated at room temperature in a solution comprising 2X SSC and SDS (0.1 %), then preheated to 42 °C for 5 min and subjected to a second incubation at 42 °C for 5 min. The slide was incubated in a solution of 0.1X SSC and SDS (0.1 %) for 5 min at room temperature, then agitated for 5 min at room temperature in a solution of 0.1X SSC, incubated in 0.01X SSC for 10 s, and finally dried by centrifugation at 1750 rpm for 2 min.

Microarray printing. The slides were prepared on a MicroGrid II compact (Digilab) with TAS Application Suite software. The size of the printed area was 1.98 mm^2 with 270 μm gaps between pixels. The chamber temperature was 22 °C at 50 % humidity. Cross-linking was performed after 2 h of incubation in the chamber (300 mJ, Bio-Link crosslinker, Vilber Lourmat).

Image scanning. The slides were scanned with a GenePix 4000B scanner (Axon Instruments). Images were processed with GenePix Pro 5.1 software.

Received: October 9, 2011
Revised: January 9, 2012
Published online: February 3, 2012

Keywords: cryptosystems · DNA · fluorescent probes · microarrays

[1] G. C. Smith, C. C. Fiddes, J. P. Hawkins, J. P. L. Cox, *Biotechnol. Lett.* **2003**, *25*, 1125 – 1130.
[2] D. Heider, A. Barnekow, *BMC Bioinf.* **2007**, *8*, 176 – 185.
[3] L. Morford, *J. Theor. Biol.* **2011**, *273*, 100 – 102.
[4] H. J. Shiu, K. L. Ng, J. F. Fang, R. C. T. Lee, C. H. Huang, *Inf. Sci.* **2010**, *180*, 2196 – 2208.
[5] C. T. Clelland, V. Risca, C. Bancroft, *Nature* **1999**, *399*, 533 – 534.
[6] A. Leier, C. Richter, W. Banzhaf, H. Rauhe, *BioSystems* **2000**, *57*, 13 – 22.
[7] K. Tanaka, A. Okamoto, I. Saito, *BioSystems* **2005**, *81*, 25 – 29.
[8] X. J. Lai, M. X. Lu, L. Qin, J. S. Han, X. W. Fang, *Sci. China Inf. Sci.* **2010**, *53*, 506 – 514.
[9] K. W. Kim, V. Bocharova, J. Halámek, M. K. Oh, E. Katz, *Biotechnol. Bioeng.* **2011**, *108*, 1100 – 1107.
[10] M. A. Palacios, E. Benito-Peña, M. Manesse, A. D. Mazzeo, C. N. LaFratta, G. M. Whitesides, D. R. Walt, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16510 – 16514.
[11] M. Arita, Y. Ohashi, *Biotechnol. Prog.* **2004**, *20*, 1605 – 1607.
[12] J. P. L. Cox, *Analyst* **2001**, *126*, 545 – 547.
[13] P. C. Wong, K. Wong, H. Foote, *Commun. ACM* **2000**, *46*, 95 – 98.
[14] C. Bancroft, T. Bowler, B. Bloom, C. T. Clelland, *Science* **2001**, *293*, 1763c – 1765c.
[15] J. P. L. Cox, *Trends Biotechnol.* **2001**, *19*, 247 – 250.
[16] N. Yachie, K. Sekiyama, J. Sugahara, Y. Ohashi, M. Tomita, *Biotechnol. Prog.* **2007**, *23*, 501 – 505.
[17] J. H. Choy, J. M. Oh, M. Park, K. M. Sohn, J. W. Kim, *Adv. Mater.* **2004**, *16*, 1181 – 1184.
[18] S. Zhou, Q. Zhang, X. Wei in *Proceedings of the 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems* (ICIS), **2010**, *3*, pp. 315 – 319.
[19] Q. Zhang, L. Guo, X. Wei, *Math. Comput. Model.* **2010**, *52*, 2028 – 2035.
[20] D. Boneh, C. Dunworth, R. J. Lipton in Proceedings *of the DIMACS Workshop on DNA Compunting* (Eds.: R. J. Lipton, E. B. Baum), Am. Math. Soc., Providence, **1996**, pp. 37 – 66.
[21] L. M. Adleman, P. W. K. Rothemund, S. Roweis, E. Winfree, *J. Comput. Biol.* **1999**, *6*, 53 – 63.
[22] A. Gehani, T. LaBean, J. Reif in *DNA-based Computers*, Am. Math. Soc., Providence, **2000**, *54*, 233 – 249.
[23] Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, E. Shapiro, *Nature* **2001**, *414*, 430 – 434.
[24] M. Soreni, S. Yogev, E. Kossoy, Y. Shoham, E. Keinan, *J. Am. Chem. Soc.* **2005**, *127*, 3935 – 3943.
[25] E. Kossoy, N. Lavid, M. Soreni-Harari, Y. Shoham, E. Keinan, *ChemBioChem* **2007**, *8*, 1255 – 1260.
[26] S. Shoshani, S. Wolf, E. Keinan, *Mol. Biosyst.* **2011**, *7*, 1113 – 1120.
[27] S. Shoshani, T. Ratner, R. Piran, E. Keinan, *Isr. J. Chem.* **2011**, *51*, 67 – 86.
[28] U. B. Gyllensten, H. A. Erlich, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 7652 – 7656.
[29] Q. Liu, L. Wang, A. G. Frutos, A. E. Condon, R. M. Corn, L. M. Smith, *Nature* **2000**, *403*, 175 – 179.
[30] X. Su, L. M. Smith, *Nucleic Acids Res.* **2004**, *32*, 3115 – 3123.

