Experimental Validation of the Transcription-Based Diagnostic Automata with Quantitative Control by Programmed Molecules

Miki Hirabayashi[1], Hirotada Ohashi[1], and Tai Kubo[2]

[1] Department of Quantum Engineering and Systems Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[2] Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan
miki@crimson.q.t.u-tokyo.ac.jp, ohashi@q.t.u-tokyo.ac.jp, tai.kubo@aist.go.jp

Abstract. Biomolecular computing using the artificial nucleic acid technology is expected to bring new solutions to various health problems. We focus on the noninvasive transcriptome diagnosis by salivary mRNAs and present the novel concept of transcription-based diagnostic automata that are constructed by programmed DNA modules. The main computational element has a stem shaped promoter region and a pseudo-loop shaped read-only memory region for transcription regulation through the conformation change caused by targets. Our system quantifies targets by transcription of malachite green aptamer sequence triggered by the target recognition. This algorithm makes it possible to realize the cost-effective and sequence-specific real-time target detection. Moreover, in the *in-vivo* therapeutic use, this transcription-based system can release RNA-aptamer drugs multiply at the transcription stage, different from the digestion-based systems by the restriction enzyme which was proposed previously. We verified the sensitivity, the selectivity and the quantitative stability of the diagnostic automata in basic conditions. Our approach will provide promising applications of autonomous intelligent systems using programmed molecules.

Keywords: Biomolecular computing, nucleic acid detection systems, DNA computing, molecular programming, autonomous diagnostic devices, molecular circuits.

1 Introduction

Salivary transcriptome diagnostics is expected as a novel clinical approach for early disease detection [1]. We describe herein a potential general approach to the rational construction of an intelligent sensor to detect the fluctuation of salivary mRNA biomarkers by transcriptional regulatory systems using molecular computation [2-9].

Recently, it was found that the combination of several disease-related salivary mRNA biomarkers yielded sensitivity (91%) and specificity (91%) to oral squamous cell carcinoma (OSCC) distinguishing from the controls [1]. The existing methods to

M.H. Garzon and H. Yan (Eds.): DNA 13, LNCS 4848, pp. 89–98, 2008.

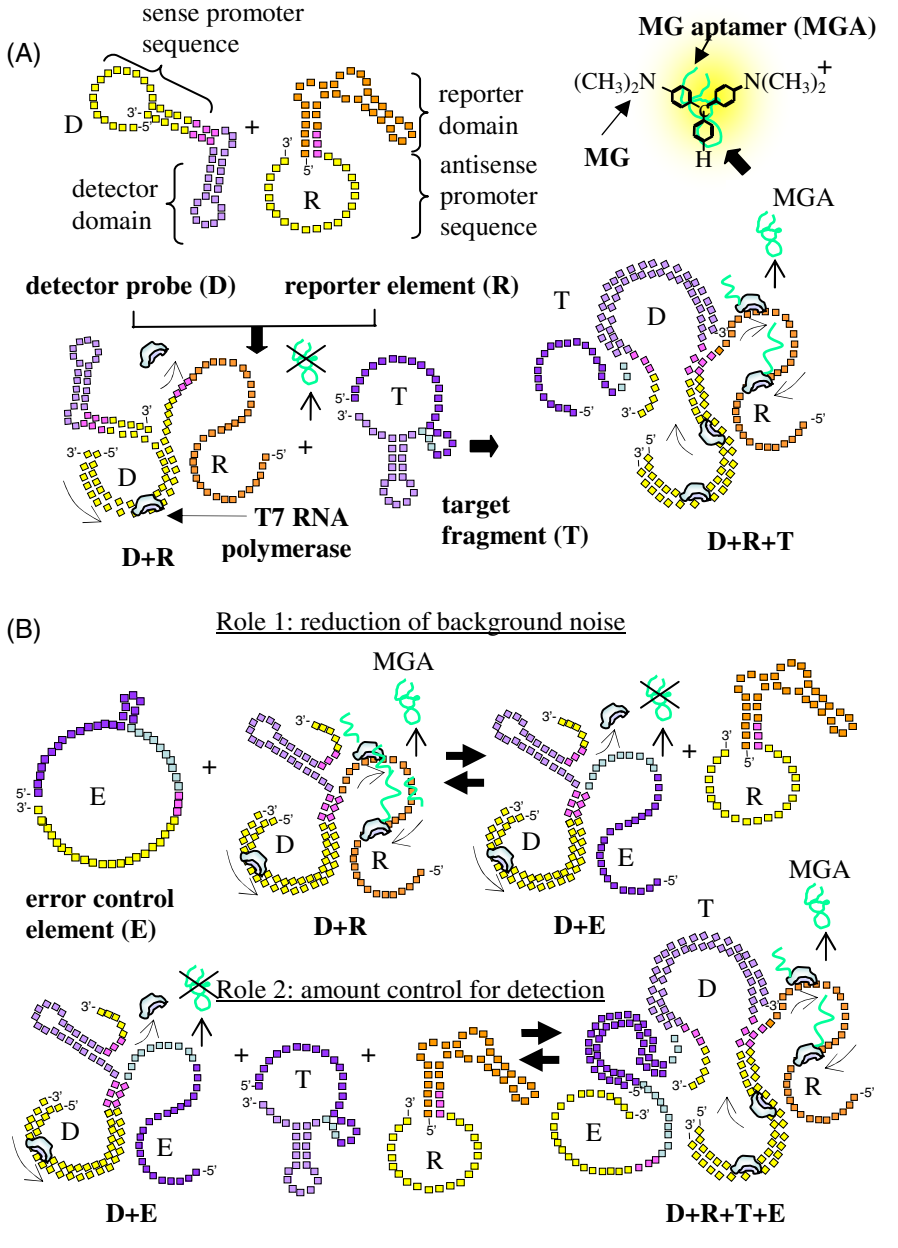

Fig. 1. (A) Operation principle of the detection system. (B) Expected main roles of error control elements.

quantify these transcripts have several problems. For example, three methods of quantitative polymerase chain reaction (qPCR), which is commonly employed to validate a subset of differently expressed transcripts identified by microarray analysis,

have some drawback and advantage [10]. The method through agarose gel electrophoresis is less expensive but less accurate than others. The method using sequence-specific fluorescent reporter probes is accurate and reliable, but expensive. The method using SYBR Green dye labels all double-stranded (ds)DNA including any unexpected PCR products, however, it has the advantage in the cost performance. Here we present the cost-effective and sequence-specific intelligent systems using molecular computation for the transcriptome diagnosis and demonstrate the concept of new diagnostic automata.

Our system consists of several computational elements (Fig. 1A). The detector probe has one stem for transcriptional regulation using the sense T7 promoter sequence and one read-only memory loop for target recognition. The stem is designed to significantly enhance their specificity in transcription regulation. The reporter element has an anti-sense promoter domain and a malachite green (MG) aptamer sequence domain. It can transcribe the aptameric sensor if the stem opens through the binding of a target oligonucleotide complementary to the memory sequence.

As for the computational elements, Stojanovic's group proposed molecular automata combined the standard oligonucleotide recognition element, a stem-loop molecular beacon with the MG aptamer itself as a part of computational elements and used on deoxyribozyme-based logic gates [11]. Here we utilize the transcription process of MG aptamer to detect the targets. As for the autonomous diagnostic and therapeutic systems using molecular computation, Shapiro's group demonstrated digestion-based therapeutic systems by the restriction enzyme (*Fok*I), which can release a single DNA molecule for therapy per therapeutic element according to diagnosis of biomarkers [9]. Our transcription-based system can transcribe RNA aptamer drugs [12] instead of the MG aptamer as much as needed under the control of programmed molecules, when it will be applied to the *in-vivo* treatment in the future. As for the additional computational elements, we introduced an error control element in order to increase the quantitative stability (Fig. 1B). This element has a target recognition domain and an inhibitor domain for detection probes, which was designed to reduce the background noise without inhibition of target recognition. Because our system adopts the conformation change to detect targets, we can control the conditions flexibly by using the additional programmed molecules such as this error control element.

We confirmed the significant sensitivity, the selectivity and the quantitative stability of in basic conditions. These proof-of-concept results will contribute to bring us the realization of autonomous intelligent diagnostic and therapeutic devices using molecular-scale computation.

2 Materials and Methods

2.1 Preparation of Oligonucleotides

DNA sequences of the oligonucleotide used for the construction of the three molecular computer components and two inputs are shown in Tables 1-3. The three components consist of the diagnostic system: detector probes, reporter elements, and error control elements. Oligonucleotides were custom-made and DNase/RNase free

HPLC purified by (Operon Biotechnologies, Tokyo, JAPAN) or (Hokkaido System Science, Sapporo, JAPAN) and used as received. Each sequence was folded using mFold server v 3.2 (URL: http://www.bioinfo.rpi.edu/applications/mfold) and visually examined to find sequences of low secondary structure.

Computational Elements. (1) detector probe: The detector probe is a detector of a target sequence. It has a stem shaped sense T7 promoter region for the transcription regulation of reporter molecules and a pseudo-loop shaped read-only memory region for the target detection (Table 1). The detector probe could receive information from target inputs at the memory domain and transfer signals to the promoter domain through the conformation change by opening the stem. **(2) reporter element:** The reporter element is a output-producing element. It has an anti-sense T7 promoter

Table 1. Single-stranded DNA models for computational elements

Name	DNA sequences (5'→3')	Length
Actin detector	5'-AGCTTAATACGACTCACTATAGGAC CTGAGGCTCTTTTCCAGCCTTTCCTAT AG-3'	54
MG aptamer reporter	5'-GGATCCATTCGTTACCTGGCTCTCGC CAGTCGGGATCCTATAGTGAGTCGTAT TAAGCT-3′	59
Actin error control	5'-TCTTGGGTATGGAATCCTGTGGAAA AAAAAAAAATCCTATAGTGAGTCGTA TTAAGCT-3'	56

Table 2. Profiling of computational elements

Name	T_m (calculated)	GC %
Actin detector domain	63.6 ℃	54.0
Actin error control domain	60.6 ℃	50.0
T7 promoter domain	57.0 ℃	37.5

Table 3. Single-stranded DNA models for input molecules

Name	DNA sequences (5'→3')	Length
βActin	5'-CCACAGGATTCCATACCCAAGAAGG AAGGCTGGAAAAGAGCCTCAGG-3'	47
IL8	5'-CACCGGAAGGAACCATCTCCATCCC ATCTCACTGTGTGTAAACATGACTTCC AAGCTG-3'	47

domain and an MG RNA aptamer sequence domain (Table 1). An MG aptamer increases the fluorescence of MG when bound [13-18], allowing us to know that the transcription occurs. The hybridization of a target at the memory region in the detector probe triggers the stem open and then the promoter region form a double strand with the reporter element. Consequently, the transcription of the MG aptamer

sequence is active and fluorescence is observed by the addition of MG. These successive reactions will enable us to recognize the existence of targets. **(3) error control element:** The error control element is a supporting element to control the computing cascade. It consists of the sense promoter domain and the target recognition domain for the reduction of the background noise and the introduction of the threshold in the transcription process (Table 1). When the target recognition domain of this element does not bind to inputs, the sense promoter region has more accessibility to the promoter module and inhibits the transcription of the MG aptamer sequence. Consequently, it is expected that the element can reduce the background noise and increase the quantitative stability.

Input Molecules. We used single-stranded (ss)DNAs to represent disease-related mRNA based on precedents in Ref. [9]. Two concentrations to represent mRNA levels: 0 μM for low level and 2-4 μM for high level at the detection stage. As disease-related biomarker models, *β*-actin and *IL8* mRNAs were selected based on reported cancer association [1]. The *β*-actin gene is a representative house-keeping gene and the transcript of *IL8* is one of salivary mRNA biomarkers for OSCC. DNA sequences used for the construction of the input models are shown in Table 3. These input ssDNA models include two recognition modules: one for detector probes and the other for error control elements.

2.2 Instrumental

Fluorescent spectra were taken on a microplate spectrofluorometer (Japan Molecular Devices, Tokyo, JAPAN, SpectraMax Gemini). Experiments were performed at the excitation wavelength (λ_{ex}) of 620 nm and emission wavelength (λ_{em}) scan of 650-700 nm. The spectra were exported to Microsoft Excel files.

2.3 Diagnostic Computations

Diagnostic computations consist of three steps: 1) mixing the detector probes for each input disease-related biomarker models and other computational elements and equilibrating them. 2) processing of the diagnostic string by T7 RNA polymerase supplementation. 3) quantifying of the fluorescence by MG supplementation.

Step 1. Control of DNA Hybridization. Detector probes, input molecules and reporter elements were mixed in that order and diluted in annealing buffer (50 mM NaCl, 100 mM HEPES pH 7.4) to 3 μM concentration each. The reaction mixtures were incubated for 22 h at 45 ^{o}C following denaturation at 94 ^{o}C for 2 min in a PCR machine block.

Step 2. Detection of Memory Recall. Hybridization mixture was subjected to transcription reaction using Ambion MEGAscript T7 Kit. The mixtures were incubated at 37 ^{o}C for up to 6 hours.

Step 3. Observation of MG Binding. Two μL of the reaction mixtures and MG were mixed in binding buffer (50 mM Tris-HCl, pH = 7.4, 50 mM $MgCl_2$, 1mM NaCl) with the final concentration of 10 μM of MG and the fluorescent spectra were taken.

3 Results

We investigated fundamental properties of our transcription-based diagnostic systems: sensitivity, selectivity, quantitative stability and scalability. The sensitivity is served by the stem shaped promoter region and the selectivity is served by the pseudo-loop shaped recognition domain for the target sequence in the detector probe. The stable quantitative scalability is realized by introduction of the error control element.

3.1 Sensitivity and Selectivity

Figure 2 shows fluorescence time scans at the transcription stage using the detector probe (D), the reporter element (R), the target ssDNA (T), and the non-target ssDNA (NT). Each spectrum and data point represents the average of ten consecutive scans at λ_{em} = 675 nm. It is confirmed that the system can generate about two-fold fluorescence when it recognizes targets (Fig. 2A). Nonzero fluorescence without targets is attributed to the fact that the reporter probe itself has a function as an opener of the stem-shaped promoter region and induces the transcription of reporter molecules. By the reducing of this background noise, the sensitivity would be further improved.

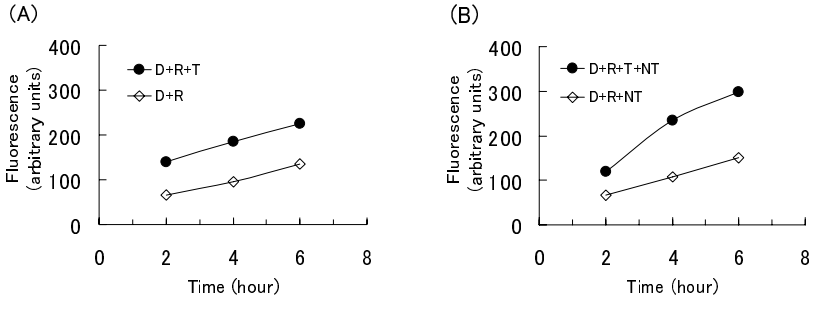

Fig. 2. Fluorescence time scans at the transcription stage. (A) Sensitivity of reporter probes. (B) Selectivity of detector probes in mixed conditions. Each data point represents the average of ten consecutive scans at λ_{em} = 675 nm.

Figure 2B shows that the detector probes can recognize target fragments in the mixture conditions basically. However, it seems that the increase of molecular species or the total amount of oligonucleotides introduces decrease of temporal quantitative stability in the amount of fluorescence increase.

3.2 Quantitative Stability and Scalability

To find the solution of the quantitative stability problems, we introduced an error control element to the system.

Figure 3 shows the increase of quantitative stability and controllability by the error control element (E). Figures 3A and 3B show that the error control element exhibits

the fluorescent reduction effects without losing the ability of fluorescence recovery and the selectivity although the sensitivity is not enough compared with Fig. 2. On the other hand, Fig. 3C shows that half standard detector probes can not decrease the background noise without losing the ability of fluorescence recovery. In Fig. 3D, the error control element exhibits the fluorescent reduction effects, which are dependent on the amount of elements. The case of half amount of standard error control elements reduces basic fluorescence as in the case of half amount of standard detector probes without losing the ability of fluorescence recovery. The case of full standard error control elements can decrease the larger amount of basic fluorescence and shows no fluorescent recovery by targets. These features make it possible to set threshold for the detectable amount of targets by adjustment of error control elements. In addition, the case of half standard error control elements with targets in Fig. 3B shows the temporal quantitative stability in the amount of fluorescence increase compared with Fig. 2B and shows the improvement of sensitivity compared with the Fig. 3A. This sensitivity improvement may be brought by the promotion of the programmed reaction due to entropy increase. Moreover it is expected that the increase of reaction efficiency by using higher level of reaction mixture also improves the system sensitivity.

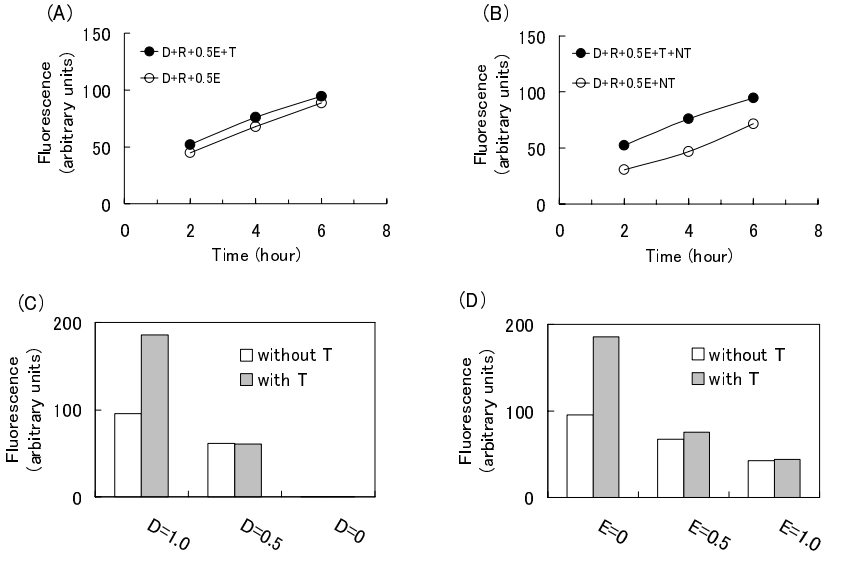

Fig. 3. Quantitative stability and controllability tests using error control elements. (A) Noise reduction effects. (B) Examination of selectivity and quantitative stability. Each data point represents the average of ten consecutive scans at λ_{em} = 675 nm. (C) Background noise control by adjustment of the detector probe. (D) Background noise control by the error control element. Fluorescence at λ_{em} = 675 nm measured at 6 hours after the transcription starts. Each datum represents the average of ten consecutive measurements.

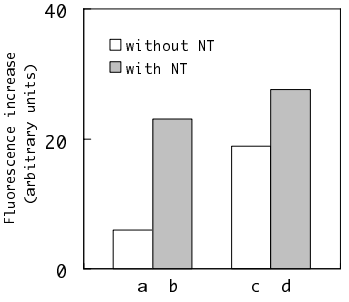

Fig. 4. Scalability of sensitivity with error control elements. Fluorescence increase by target detection at λ_{em} = 675 nm measured at 6 hours after the transcription starts. Low concentration: D+R+0.5E (a) or D+R+0.5E+NT (b) for T, and high concentration: 2D+2R+E (c) or 2D+2R+E+2NT (d) for 2T. Each datum represents the average of ten consecutive measurements.

Figure 4 shows that higher concentration improves the system sensitivity. This may be because the computational elements are short oligonucleotides and therefore the initiation becomes rare events when the concentration is low. Scalability remains intact in the relative complex cases with NT. This shows that the system can detect higher concentration of targets by using higher concentration of probes. The system sensitivity is improved with an increase in entropy as in the case of Fig. 3B. This phenomenon was observed through the preliminary experiments. We expect that this feature provides the promising consideration toward the practical use.

We showed that the error control element can adjust the detection amount of targets and improve the system stability for quantitative detection. These results indicate that it is possible that we set the threshold of detection amount of targets and perform the quantitative stable detection by choosing optimized combination and concentration of appropriate programmed molecules such as detector probes and error control elements.

4 Discussion

We tested the new molecular computation algorithm for the transcriptome diagnosis and confirmed that it can supply the system which shows the significant sensitivity, selectivity and quantitative stability in mixed conditions, in which none, one, or two input disease-related biomarker models are present.

Our transcription-based diagnostic automata have the following three remarkable features over existing quantitative methods for targets.

(1) flexibility in programming: For the target detection, our system utilizes the transcriptional regulation based on the conformational change of the detector probe that is triggered by the target recognition. This enables the flexible control by additional programmed molecules such as proposed error control elements. Thus the construction of complicated and intelligent automata becomes possible.

(2) cost-effective and sequence-specific real-time detection: The introduction of the MG-RNA-aptamer transcription for the fluorescence detection realizes the cost-effective real-time observation. Moreover the transcription regulated by the specific target sequence in the pseudo-loop shaped read-only memory region enables the sequence-specific detection. This is a substantial feature for the parallel quantitative diagnostic operation.

(3) potential ability for therapeutic automata: Because we adopts the transcription-based diagnostic systems instead of the digestion-based systems by the restriction enzyme, the controlled multiple release of RNA aptamer drugs is possible when this system is applied to the *in-vivo* therapeutic use. Our system has the prominent potential ability to organize the *in-vivo* therapeutic automata.

In this paper, we demonstrated the new concept of transcription-based diagnostic automata. By additional programming, we will be able to detect the target combinations of up-regulated disease-related biomarkers. For example, it was reported that by using the combination of up-regulated salivary mRNAs: *IL8* (24.3-fold), *SAT* (2.98-fold) and *H3F3A* (5.61-fold) for OSCC prediction, the overall sensitivity is 90.6% [1]. The introduction of molecular gates to our systems will enable this kind of autonomous one-step diagnosis for OSCC. In the near future, the accurate and reliable control by programmed molecules may offer the easy self-diagnostic tool using saliva.

These results would bring us one step closer to the realization of new intelligent diagnostic automata based on biomolecular computation.

References

1. Li, Y., John, M.A.R.St., Zhou, X., Kim, Y., Sinha, U., Jordan, R.C.K., Eisele, D., Abemayor, E., Elashoff, D., Park, N.-H., Wong, D.T.: Salivary transcriptome diagnostics for oral cancer detection. Clin. Cancer Res. 10, 8442–8450 (2004)
2. De Silva, A.P., McClenaghan, N.D.: Molecular-scale logic gates. Chemistry 10, 574–586 (2004)
3. Braich, R.S., Chelyapov, N., Johnson, C., Rothemund, P.W.K., Adleman, L.: Solution of a 20-variable 3-SAT problem on a DNA computer. Science 296, 499–502 (2002)
4. Winfree, E., Liu, F.R., Wenzler, L.A., Seeman, N.C.: Design and self-assembly of two-dimensional DNA crystals. Nature 394, 539–544 (1998)
5. Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Livneh, Z., Shapiro, E.: Programmable and autonomous computing machine made of biomolecules. Nature 414, 430–434 (2001)
6. Wang, H., Hall, J.G., Liu, Q., Smith, L.M.: A DNA computing readout operation based on structure-specific cleavage. Nat. Biotechnol. 19, 1053–1059 (2001)
7. Saghatelian, A., Voelcker, N.H., Guckian, K.M., Lin, V.S.-Y., Ghadiri, M.R.: DNA-based photonic logic gates: AND, NAND, and INHIBIT. J. Am. Chem. Soc. 125, 346–347 (2003)
8. Okamoto, A., Tanaka, K., Saito, I.: DNA logic gates. J. Am. Chem. Soc. 126, 9458–9463 (2004)
9. Benenson, Y., Gil, B.-D.U., Adar, R., Shapiro, E.: An autonomous molecular computer for logical control of gene expression. Nature 429, 423–429 (2004)
10. Ginzinger, D.G.: Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. Exp. Hematol. 30, 503–512 (2002)

11. Kolpashchikov, D.M., Stojanovic, M.N.: Boolean control of aptamer binding states. J. Am. Chem. Soc. 127, 11348–11351 (2005)
12. Bunka, D.H.J., Stockley, P.G.: Aptamers come of age – at last. Nat. Rev. Microbiol. 4, 588–596 (2006)
13. Babendure, J.R., Adams, S.R., Tsien, R.Y.: Aptamers switch on fluorescence of triphenylmethane dyes. J. Am. Chem. Soc. 125(48), 14716–14717 (2003)
14. Baugh, C., Grate, D., Wilson, C.: 2.8 A° crystal structure of the malachite green aptamer. J. Mol. Biol. 301(1), 117–128 (2000)
15. Famulok, M.: Chemical biology: Green fluorescent RNA. Nature 430(7003), 976–977 (2004)
16. Grate, D., Wilson, C.: Laser-mediated, site-specific inactivation of RNA transcripts. Proc. Natl. Acad. Sci. USA 96(11), 6131–6136 (1999)
17. Stojanovic, M.N., Kolpashchikov, D.M.: Modular aptameric sensors. J. Am. Chem. Soc. 126(30), 9266–9270 (2004)
18. Hirabayashi, M., Taira, S., Kobayashi, S., Konishi, K., Katoh, K., Hiratsuka, Y., Kodaka, M., Uyeda, T.Q.P., Yumoto, N., Kubo, T.: Malachite green-conjugated microtubules as mobile bioprobes selective for malachite green aptamers with capturing/releasing ability. Biotechnol. Bioeng. 94(3), 473–480 (2006)