

Research Article

Biomolecular computation with molecular beacons for quantitative analysis of target nucleic acids

Hee-Woong Lim^{a,c,1,3}, Seung Hwan Lee^{a,b,3}, Kyung-Ae Yang^{a,2}, Suk-In Yoo^{a,c},
Tai Hyun Park^{a,b}, Byoung-Tak Zhang^{a,c,*}

^a Center for Biointelligence Technology, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea

^b School of Chemical and Biological Engineering, Bio-MAX Institute, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-744, Republic of Korea

^c School of Computer Science and Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea

ARTICLE INFO

Article history:

Received 24 June 2011

Received in revised form 23 June 2012

Accepted 11 September 2012

Keywords:

Biomolecular computation

Molecular beacon

Quantitative analysis

Weighted-sum

ABSTRACT

Molecular beacons are efficient and useful tools for quantitative detection of specific target nucleic acids. Thanks to their simple protocol, molecular beacons have great potential as substrates for biomolecular computing. Here we present a molecular beacon-based biomolecular computing method for quantitative detection and analysis of target nucleic acids. Whereas the conventional quantitative assays using fluorescent dyes have been designed for single target detection or multiplexed detection, the proposed method enables us not only to detect multiple targets but also to compute their quantitative information by weighted-sum of the targets. The detection and computation are performed on a molecular level simultaneously, and the outputs are detected as fluorescence signals. Experimental results show the feasibility and effectiveness of our weighted detection and linear combination method using molecular beacons. Our method can serve as a primitive operation of molecular pattern analysis, and we demonstrate successful binary classifications of molecular patterns made of synthetic oligonucleotide DNA molecules.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Many studies during the last decades have shown great potential of biomolecular computing not only as a novel computing paradigm (Banzhaf et al., 1996; Seeman et al., 1998; Henkel et al., 2007) or as a new technique for tackling computationally intractable problems (Chen and Yang, 2010; Zhang and Liu, 2011) but also as a useful tool for biological applications (Mills, 2002; Rinaudo et al., 2007; Benenson, 2009). Direct computation on biochemical information on a molecular level provides us with an interface to manipulate and analyze biological systems for therapeutic or diagnostic purposes more effectively. To this end, various approaches have been proposed including Boolean logical circuits (Stojanovic et al., 2002;

Macdonald et al., 2006; Seelig et al., 2006; Zhang et al., 2007), arithmetic operations (Oliver, 1997; Mills et al., 1999, 2001; Lim et al., 2010), molecular learning methods (Lim et al., 2003; Chen et al., 2005; Zhang and Jang, 2005; Zhang, 2008), DNA nanodevices (Simmel and Dittmer, 2005; Reif and LaBean, 2007; Ramlan and Zauner, 2011) and output detection devices (Henkel et al., 2005; Lee et al., 2008a). However, many of those works are DNA versions of Boolean logic where all information is encoded in binary levels and more sophisticated analog primitive operations are required to fully utilize the potential of direct computation in molecular level because primary target data are originally analog. Fluorescence-based quantitative assay (Kricka, 2002; Waggoner, 2006) is not only one of the basic wet-lab techniques in biology but also an excellent and versatile candidate for a biomolecular computing component. With an extensive range of applications, various materials and substrates have been developed (Mirkin et al., 1996; Storhoff et al., 2004; Jamieson et al., 2007) and lots of variations have been proposed coupled with enzymatic amplification methods (Heid et al., 1996; Tasara et al., 2003; Tani et al., 2007). However, those conventional methods were limited to single target detection or multiplexed detection, where each fluorescence signal reflects only a single target quantity.

Here, we propose a molecular beacon-based biomolecular computing method for quantitative detection and analysis of target nucleic acids. Molecular beacons are very efficient and useful

* Corresponding author at: Center for Biointelligence Technology, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea.
Tel.: +82 2 880 1833.

E-mail addresses: hwl@ailab.snu.ac.kr (H.-W. Lim), skulsh78@snu.ac.kr (S.H. Lee), kayang@bi.snu.ac.kr (K.-A. Yang), siyoo@ailab.snu.ac.kr (S.-I. Yoo), thpark@snu.ac.kr (T.H. Park), btzhang@bi.snu.ac.kr (B.-T. Zhang).

¹ Present address: School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

² Present address: Department of Medicine, Columbia University, New York, NY 10032, USA.

³ Both authors equally contributed to this work.

Table 1
Target and molecular beacon sequences.

Name	Sequence ^a
X ₁	ggtcaggctgCCTGGAAGGAGGCGAGAGTgtcttcgagaa
X ₂	ccacatcgccTGCTTCACCTGCCACCAGTgtcgcgagcgc
X ₃	gcagacctcgCCTGTGGCAGATGCACCGAATatttcttcaaa
X ₄	ttggtatagaTGAAGAGCAAGGCCCTCAGTatataaagtgt
B ₁ /B ₁ ^c	Cy5/Cy3-cgcgaACACTCTCGCTCCTCCAGTcgcg-BHQ2
B ₂ /B ₂ ^c	Cy5/Cy3-cgcgaACACTCTCGCTCCTCCAGTcgcg-BHQ2
B ₃ /B ₃ ^c	Cy5/Cy3-cgcgaATTCCGTCATCTGCCACAGGtgcg-BHQ2
B ₄ /B ₄ ^c	Cy5/Cy3-cgcgaACCTGAGGGCCTTGTCTTCAtcgcg-BHQ2

^a Uppercase domains in X_i denote beacon-binding parts. Uppercase and lowercase domains in B_i and B_i^c denote loop and stem regions, respectively. All the sequences are in the 5' to 3' direction.

materials for detecting and quantifying specific target nucleic acids (Tyagi and Kramer, 1996; Vet et al., 1999; Kim et al., 2007), and a theoretical DNA computing model has been proposed based on this as well (Xiaohui et al., 2009). Especially, they emit fluorescence only upon hybridization with target molecules, so assays using molecular beacons do not require a separation step of unbound beacons. Unlike the conventional assays, our method enables us to simultaneously detect and analyze multiple targets *in vitro*, where each target species is assigned a respective weight and their linear combination is computed. The computation is performed via competitive hybridization reactions between target nucleic acid molecules and molecular beacon mixtures, and then the outputs are detected as fluorescence signals. Our method can also be used as a primitive operation of a molecular pattern classifier. A series of experimental results are demonstrated, including weighted detection of a single target, a linear combination of two targets and binary classifications of molecular patterns made of four synthetic oligonucleotide DNAs. This paper is an extended written version of the oral presentation given at the 14th International Meeting on DNA Computing (Lee et al., 2008b).

2. Experimental details

2.1. Target and molecular beacon preparation

Four oligonucleotide DNA sequences, X_i (i = 1, 2, 3, 4), as targets and four pairs of molecular beacons, B_i and B_i^c (i = 1, 2, 3, 4), were designed from the gene transcripts, fumarylacetoacetate (FAH), zyxin (ZYG), c-myc (MYB) and proteasome subunit alpha type 6 (PSMA6), respectively (Table 1). The two beacons, B_i and B_i^c, have the same sequences and differ only by the attached dyes. In designing, thermodynamic properties were considered to avoid cross-homology between X_i's and crosstalk between X_i and B_j (B_j^c), where i ≠ j. The melting temperatures of target–beacon duplexes were also designed to be similar to each other. DINAMelt server (Markham and Zuker, 2005) was used for predicting and designing beacon structures. Target sequences were custom-made by Bioneer (Daejeon, Korea). Molecular beacons were custom-made by Integrated DNA Technologies, Inc. (Coralville, IA, USA), where Cy5 and Cy3 were used to label B_i and B_i^c, respectively, and BHQ2 was used as a quencher. Each of 12 sequences was brought to a stock concentration at 100 μM in distilled water and stored at –20 °C.

2.2. Hybridization and fluorescence signal detection

The hybridization reactions were performed in 50 μl reaction buffer containing 3.5 mM MgCl₂, 50 mM KCl and 10 mM Tris–HCl (pH 8.0). Target and beacon concentrations were varied according to experimental settings. A reaction buffer containing targets and beacons was incubated at 95 °C for 3 min and

the temperature was steadily lowered to 10 °C by 0.5 °C/min using a thermal cycler (iCycler, Bio-Rad, Hercules, CA, USA). After the hybridization, fluorescence signal intensities were measured using a computer-controlled fluorescence plate reader (GENios Pro, Tecan, Mannedorf, Switzerland) at the wavelength of 590 nm and 670 nm for Cy3 and Cy5, respectively. The two intensities were adjusted to the same level by detecting fluorescence signals from equal amount of Cy3-modified (and Cy5) linear probes.

3. Results and discussion

3.1. Weighted detection and analysis of target nucleic acids

Our weighted detection and analysis method enables us to detect multiple targets and compute a linear combination of their quantities simultaneously. The detection and computation are performed via a set of competitive hybridization reactions between target molecules and molecular beacon mixtures in a test tube, and the results are detected as fluorescence signals. Quantitative assays are very basic and essential techniques in biology. They are used for producing quantitative information on target biomarkers to be used in a subsequent analysis procedure. For example, distinguishing a diseased tissue from healthy ones is often performed via a mathematical function of target biomarker concentrations such as RNAs or microRNAs (Golub et al., 1999; Lu et al., 2005). In those works, fluorescence-based quantitative assays allow us to measure the expression levels of targets such as genomic DNAs, RNAs or microRNAs and produce raw input data for the mathematical function. Conventional assays are either single target detection or multiplexed detection (Fig. 1a). Even in the multiplexed detection, each target is detected independently as a fluorescence signal at a specific wavelength whose intensity is proportional to the target quantity. In this scheme, however, detection and analysis are separately performed in different phases, *in vitro* and *in silico*, respectively.

To describe the method in more detail, let us consider a set of n nucleic acid targets, X_i (i = 1, ..., n). Unlike conventional assays, our method uses a pair of molecular beacons (Tyagi and Kramer, 1996; Vet et al., 1999), B_i and B_i^c, for each X_i. B_i and B_i^c have the same nucleotide sequences and differ only by the fluorescent dyes at 5' ends, where B_i's are labeled with Cy5 and B_i^c's are labeled with Cy3 for all i. In a test tube, B_i and B_i^c competitively bind to a common target, X_i, to form hetero-duplexes, H_i and H_i^c, respectively, as follows:



Assuming excess amounts B_i and B_i^c, complete reaction of X_i and thermodynamic symmetry of the two competitive reactions because of the same nucleotide sequences, then the resulting amounts of hetero-duplexes, h_i and h_i^c, after the reaction are totally determined by the initial configuration. Specifically, the initial amount of X_i and the mixing ratio of B_i and B_i^c are:

$$h_i = \frac{x_i b_i}{b_i + b_i^c}, \quad h_i^c = \frac{x_i b_i^c}{b_i + b_i^c}, \quad (2)$$

where b_i, b_i^c and x_i denote the initial amount of the reactants, B_i, B_i^c and X_i, respectively. When all these hybridization reactions occur in a test tube, we can formulate two linear combinations of targets,

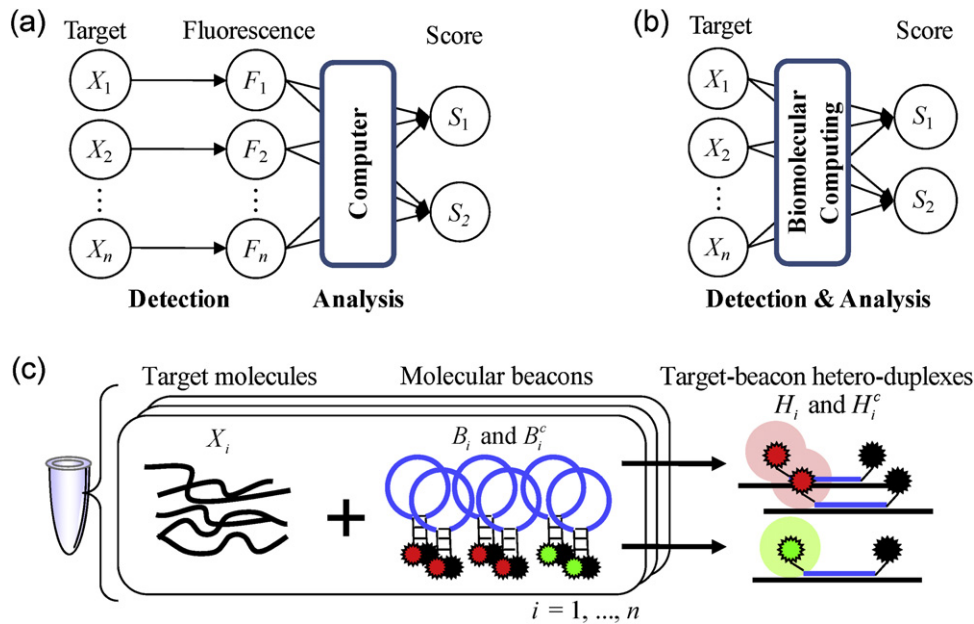


Fig. 1. Quantitative detection and analysis. (a) Conventional quantitative detection assay. Conventional assays are either single target detection or multiplexed detection, where each fluorescence signal represents corresponding single target only. Detected fluorescence intensities are given to a mathematical function as inputs and analyzed to produce index information that can be used as scores for the final decision-making. In this procedure, the detection step and the analysis step are separated. (b) Our quantitative detection and analysis method. Using biomolecular computation, the analysis step can be integrated into the detection step, where both are performed simultaneously on a molecular level to produce index information directly. (c) Weighted detection using molecular beacons. For each target, X_i , a pair of molecular beacons, B_i and B_i^c , are used as probes, where B_i 's (B_i^c) are labeled with the same fluorophore for all $i = 1, \dots, n$. B_i and B_i^c have the same nucleotide sequences but differ by the attached fluorophore so that they competitively bind to X_i to form hetero-duplexes, H_i and H_i^c , respectively. The resulting amounts of H_i and H_i^c after the reaction depend on the initial quantity of X_i and the mixing ratios of B_i and B_i^c . When these competitive reactions occur in a test tube for all i , the detected fluorescence signals form linear combinations of target quantities.

$s(\mathbf{x}; \mathbf{w})$ and $s^c(\mathbf{x}; \mathbf{w})$, where $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{w} = (w_1, w_2, \dots, w_n)$, from the reactant–product quantitative relations:

$$s(\mathbf{x}; \mathbf{w}) \equiv \sum_{i=1}^n h_i = \sum_{i=1}^n x_i w_i \quad (3)$$

$$s^c(\mathbf{x}; \mathbf{w}) \equiv \sum_{i=1}^n h_i^c = \sum_{i=1}^n x_i (1 - w_i),$$

where

$$w_i = \frac{b_i}{b_i + b_i^c}. \quad (4)$$

In this way, we can assign an arbitrary weight, $w_i \in [0, 1]$, to each target and compute their linear combinations, $s(\mathbf{x}; \mathbf{w})$ and $s^c(\mathbf{x}; \mathbf{w})$, and then detect their results as fluorescence signals of Cy5 and Cy3, respectively. Note that $s(\mathbf{x}; \mathbf{w})$ and $s^c(\mathbf{x}; \mathbf{w})$ are complementary to each other in the sense that their sum always equals the sum of x_i .

The competitive hybridization is a thermodynamically well-established technique and widely used in two-dye microarray systems (Lipshutz et al., 1999). Previously, we have also used competitive hybridization reactions to implement a molecular pattern classifier (Lim et al., 2010). However, it required affinity bead separation to remove unbound probes before the final output detection, which was the main source of error and brought about a significant increase of overall computation time (Khodor and Gifford, 1997). Using molecular beacons helps us to overcome these drawbacks. Molecular beacons do not fluoresce in the absence of targets by keeping the fluorophore and quencher in close proximity; but, when hybridized with target strands, they undergo a conformational change to become fluorescent. Therefore, using molecular beacons does not require an additional step to remove unbound probes before signal detection.

3.2. Experimental results: weighted detection of single target

As mentioned earlier, our weighted detection scheme assumes thermodynamic symmetry of two competitive hybridization reactions between targets and differentially labeled molecular beacons. As a proof-of-principle experiment, we performed a series of hybridization reactions using single targets and a pair of corresponding beacons under various target levels and mixing ratios of the beacons to validate our assumption. A target DNA, X_4 , and a pair of beacons, B_4 and B_4^c , in Table 1 were used in the experiment. Target levels, x , were 2, 5 and 8 pmol per reaction. The total amount of B_4 and B_4^c was set to be 50 pmol constantly; but their mixing ratio was varied to be (1:9), (3:7), (5:5), (7:3) and (9:1), i.e. $w = 0.1, 0.3, 0.5, 0.7$ and 0.9 . We repeated the experiment three times (see Section 2 for experimental details), and Fig. 2 shows the average relative fluorescence unit (RFU) and standard errors. As can be seen in the figure, for a constant amount of target, the RFU showed symmetric and linear increase with the ratio of the corresponding beacon. And the RFU increased linearly with the target level as well.

3.3. Experimental results: linear combination of two targets

We also performed linear combinations of two targets under various weight settings. Two target DNAs, X_2 and X_3 , and two pairs of beacons, B_2/B_2^c and B_3/B_3^c , in Table 1 were used in the experiment. Target levels, x_2 and x_3 , were set to be 5 pmol equally and constantly. For each target, 50 pmol of a beacon pair was used, but their mixing ratio was varied to be (2:8), (4:6), (6:4) and (8:2), i.e. $w_i = 0.2, 0.4, 0.6$ and 0.8 . Consequently, a total of 16 combinations of hybridization reactions were performed. We repeated the experiment three times (see Section 2 for experimental details), and Fig. 3 shows the average computation results and standard errors. Fluorescence signals from Cy5 representing the results of $x_2 w_2 + x_3 w_3$ increased linearly with both w_2 and w_3 as shown in

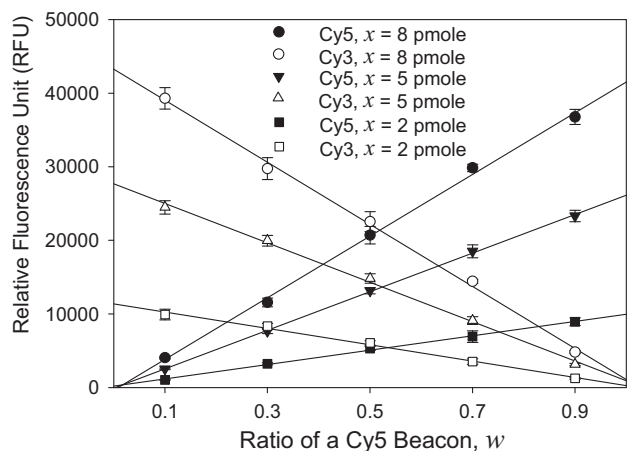


Fig. 2. Weighted detection of single targets. Resulting relative fluorescence unit (RFU) measured under various target levels, x , and mixing ratios, w , shows symmetric and linear increase with both x and w (and $1 - w$). This demonstrates the feasibility of weighted detection of a target using a differentially labeled molecular beacon mixture. The black solid symbols and the white symbols can be viewed as the computation results of xw and $x(1 - w)$, respectively. The correlation coefficients were 0.998, 0.998, 0.999, 0.997, 0.997 and 0.994 in the order of the legend.

Fig. 3a. Fluorescence signals from Cy3 representing the results of $x_2(1 - w_2) + x_3(1 - w_3)$ were also measured simultaneously from the same test tubes; they decreased linearly with both w_2 and w_3 as expected (Fig. 3b), i.e. increased with both $(1 - w_2)$ and $(1 - w_3)$. These symmetric results confirm the validity of the two-target linear combination.

3.4. Experimental results: application to molecular pattern classification

Many biological problems involve pattern classification tasks, i.e. assigning a class label to a given input pattern which is typically noisy. The quantitative information on target biomarkers naturally represents a real-valued vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ as a target pattern, each x_i of which is the molecular quantity of a biomarker X_i . As computing a linear combination of inputs plays a key role in many pattern classifiers, our weighted detection and analysis method can be used for biomolecular pattern classification. However, whereas the weights in conventional classifiers may be either positive or

Table 2
Target patterns in molecular pattern classification.^a

Group 1				Group 2					
Number	x_1	x_2	x_3	x_4	Number	x_1	x_2	x_3	x_4
1	7	3	7	1	6	3	5	7	5
2	5	3	5	1	7	1	5	7	3
3	5	3	5	7	8	1	3	7	1
4	5	1	3	7	9	1	7	5	5
5	3	1	1	7	10	1	5	5	3

^a Each row denotes the composition of molecular patterns used in the classification. Each column, x_i , denotes the amount (pmol) of the corresponding target, X_i , in the patterns.

negative, it is not tractable to represent negative weights explicitly by molecular quantities or ratios. Instead, we have developed a DNA-based molecular pattern classifier in our previous work (Lim et al., 2010), where the weights are represented as the mixing ratios of differentially labeled linear probes that is calculated from the original weights (see appendix for brief mathematical formulation). It is also possible to build an equivalent classifier by using molecular beacons and the two complementary linear combinations, $s(\mathbf{x}; \mathbf{w})$ and $s^c(\mathbf{x}; \mathbf{w})$ to obtain classification scores in the same way.

As a final application example, we built a binary linear classifier using our method and performed a classification of synthetic molecular patterns. We prepared two groups of molecular patterns made of four oligonucleotide DNAs, X_i ($i = 1, 2, 3, 4$), in Table 1, as a model system. Target patterns were prepared via the same procedure as in the previous work (Lim et al., 2010): An arbitrary hyperplane in a four-dimensional space was determined as a decision boundary between two groups with a normal vector $\mathbf{w}^0 = (0.9, 0.3, -0.6, -0.3)$; five points were selected from each side of the hyperplane as representative patterns in close proximity to the hyperplane; then X_i 's were mixed according to the coordinates as shown in Table 2. In the model system, we set the bias to be zero for convenience without loss of generality. This is reasonable because the bias term does not affect the general computation scheme.

We performed the classification task under two different settings: correct weights as a main task and incorrect weights as a control experiment. This model system is motivated by diagnostic situations under which multiple biomarkers are considered (Golub et al., 1999; Lu et al., 2005). Although even one species of biomarker can be an informative feature for diagnosis in some cases, it is essential to incorporate multiple biomarkers for more accurate diagnosis.

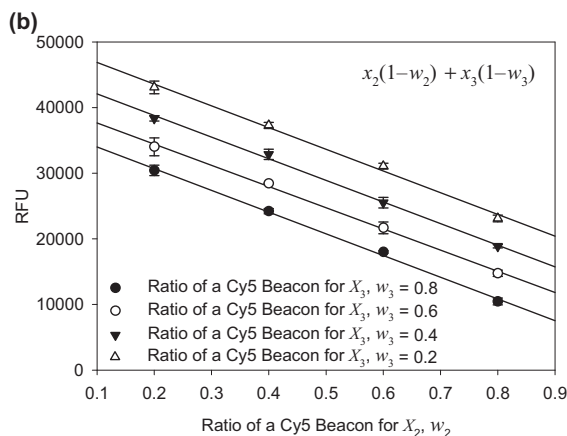
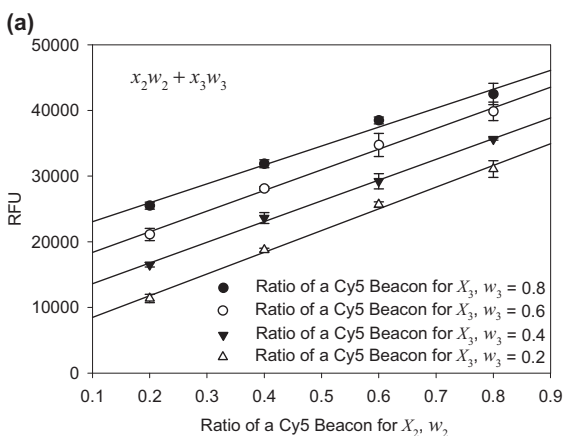


Fig. 3. Linear combinations of two targets. Two targets were weighted and summed, where the target levels, x_2 and x_3 , were set to be equal and constant and the weights were varied to be 0.2, 0.4, 0.6 and 0.8 for each target. A total of 16 combinations of computations were performed and repeated three times each. (a) RFU from Cy5 representing the computation results of $x_2w_2 + x_3w_3$ increases linearly with both w_2 and w_3 . (b) RFU from Cy3 representing the computation results of $x_2(1 - w_2) + x_3(1 - w_3)$ decreased linearly with both w_2 and w_3 . Higher ratios of Cy5 beacons mean lower ratios of Cy3 beacons. The RFU in both (a) and (b) linearly increases with the ratio of corresponding beacons. These two results were measured simultaneously from the same test tubes. The correlation coefficients were 0.989, 0.995, 0.998 and 0.995 in (a) and 0.998, 0.998, 0.997 and 0.994 in (b) in the order of the legend.

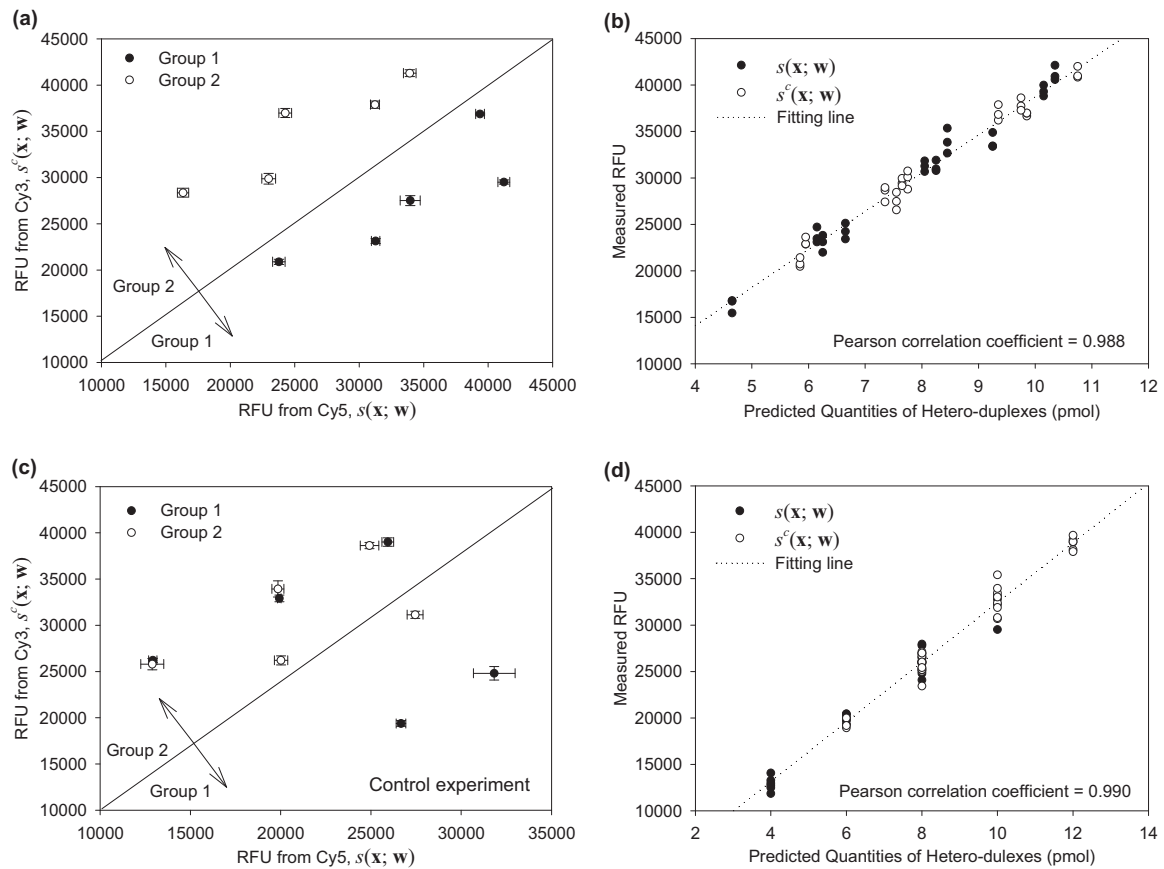


Fig. 4. Molecular pattern classification. (a) Plot of $s(\mathbf{x}; \mathbf{w})$ vs. $s^c(\mathbf{x}; \mathbf{w})$ showing the classification results. Two groups of patterns were successfully classified without overlap, where the diagonal solid line denotes a decision boundary between Groups 1 and 2. (b) Correlation between measured RFU in (a) and predicted computation results. As $s(\mathbf{x}; \mathbf{w})$ and $s^c(\mathbf{x}; \mathbf{w})$ are represented as the molecular quantities of target–beacon hetero-duplexes, the predicted computation results were obtained by calculating $\sum_{i=1}^4 x_i w_i$ and $\sum_{i=1}^4 x_i (1 - w_i)$ numerically. High correlation between predicted results and measured RFU supports the accuracy of the linear combinations as well as the soundness of the classification results. (c) Plot of $s(\mathbf{x}; \mathbf{w})$ vs. $s^c(\mathbf{x}; \mathbf{w})$ showing the classification results of a control experiment using incorrect weights. Two groups of patterns overlap across the decision boundary. This result demonstrates the importance of weights in pattern classifications. (d) Correlation between measured RFU and predicted computation results in the control experiment. Although the pattern classification was not successful in (c) as intended, the linear combinations are still accurate and show high correlation with the predicted results.

Furthermore, although many DNA-based logical analysis models have been developed (Macdonald et al., 2006; Yin et al., 2008), the binary quantitative information is often not sufficient because the relative importance of markers may be different from each other. Weights in a classifier represent such a relative importance of each target, *i.e.* marker, and thus appropriate weights are crucial information for correct classification.

First, we determined $\mathbf{w} = (0.95, 0.65, 0.2, 0.35)$ for a molecular pattern classifier from the above original normal vector, \mathbf{w}^0 , using (A.3). Four pairs of beacons, B_i/B_i^c ($i = 1, 2, 3, 4$), in Table 1 were mixed according to w_i for weight assignment. The weighted detection and analysis method was performed for each pattern and the results of $s(\mathbf{x}; \mathbf{w})$ and $s^c(\mathbf{x}; \mathbf{w})$ were measured as fluorescence signals (see Section 2 for experimental details). Fig. 4a shows the classification result, where the two groups of patterns were classified successfully without overlap. The diagonal line is the decision boundary between the two groups because decision-making is done by comparing the two scores, $s(\mathbf{x}; \mathbf{w})$ and $s^c(\mathbf{x}; \mathbf{w})$. We repeated the classification three times and still the computation results showed only small standard errors. As shown in Fig. 4b, the measured RFU also showed high correlation with the predicted molecular quantities of the corresponding target–beacon hetero-duplexes, which was obtained by calculating $\sum_{i=1}^4 x_i w_i$ and

$\sum_{i=1}^4 x_i (1 - w_i)$ numerically. The Pearson correlation coefficient was 0.988, which is higher than 0.96 in our previous work (Lim et al., 2010). This high correlation supports the soundness of the classification results as well as the computation accuracy of our method.

Then, we performed the second classification task using the same patterns but wrong weights as a control experiment. In this experiment, we ignored the original \mathbf{w}^0 and newly set $\mathbf{w}^0 = (1, 1, -1, -1)$. And \mathbf{w} for a molecular pattern classifier was determined to be $(1, 1, 0, 0)$ from the new \mathbf{w}^0 using (A.3). The experimental procedure and target patterns were the same with the first one except for the weights. Fig. 4c shows the classification result, where the two groups were overlapped across the decision boundary. Again, this result was intentional to show the importance and meaning of the weights in pattern classification. Mathematically speaking, weights in a binary linear classifier means an orientation of the decision boundary, *i.e.* a normal vector. In a diagnostic context, for example, each biomarker has its own priority for the diagnosis or classification depending on the distribution of its expression level according to the tissue types under consideration. If we ignore such relative importance and treat all the biomarkers equally, *i.e.* with same weights, it could lead to misdiagnosis as in Fig. 4c. However, still the measured RFU showed high correlation with the predicted

molecular quantities as shown in Fig. 4d, which suggests the accuracy of the molecular beacon-based computation itself.

It should be noted that the accuracy of the operation was significantly improved in comparison to the previous one (Lim et al., 2010). The main reason is that we removed the affinity bead separation step, which is the most time-consuming and vulnerable to human errors, from the computation procedure. Wet-lab techniques used in biomolecular computing take relatively long time unlike the operations in electronic computers. Moreover, the more human intervention in a biomolecular computer, the more potential of errors it has. Therefore, it is a critical issue to reduce the number of experimental steps in biomolecular computing. Molecular beacons are excellent substrates for this purpose.

4. Conclusion

The main idea of biomolecular computing has been to take advantage of innate massive parallelism in biochemical reactions. However, contemporary electronic computers are doing excellent jobs and their performance is growing fast. Instead, biomolecular computing still has a potential as an application to biological problems themselves, where biochemical information are naturally represented as biomolecules and direct computations are performed on the information in a molecular level. However, target bio-data to be processed are originally analog. Therefore, to fully utilize the advantage of the direct computation, we need appropriate primitive analog operators that can be a building block for large scale applications in conjunction with other biomolecular computing operations. To this end, we developed a molecular beacon-based biomolecular computing method for quantitative detection and analysis of target nucleic acids. Our method allows us not only to detect multiple nucleic acid targets but also compute a linear combination of their quantities by assigning an arbitrary weight on each target. We improved the accuracy and reduced the overall computation time by removing an affinity bead separation step from our previous work (Lim et al., 2010). A series of experimental results were presented to prove the underlying principles including the weighted detection of single targets and the linear combinations of two targets. Successful binary classifications of molecular patterns were also demonstrated for synthetic patterns of oligonucleotide DNA molecules.

Although we are using molecular beacons here, it should be noted that concrete physical mechanisms to implement our weighted detection and analysis method, in principle, are not necessarily restricted to molecular beacons. There could be various alternatives such as aptamer–proteins interaction, protein–protein interactions or hybridization reactions of various DNA derivatives as far as they satisfy the assumptions mentioned earlier. It is also important to note that both detection and analysis for pattern classification were done directly on a molecular level *in vitro*. Although we assumed that the weights are known for the target groups, the point is that we can use that information to prepare a weight-encoding molecular beacon mixture and apply it to the classification of target patterns without knowing their quantitative information explicitly. In a diagnostic context, determining appropriate weights and preparing molecular beacon mixtures can be viewed as manufacturing disease diagnosis kits; applying our method for pattern classification would correspond to actual diagnoses. More sophisticated techniques are yet to be developed and incorporated for the proposed method to be realized in clinical situations. Molecular signal amplifiers (Zhang and Seelig, 2010) to preprocess raw level information, more intuitive visualization techniques (Lee et al., 2008a), or DNA-based learning (Zhang and Kim, 2006; Lee et al., 2011) for learning weights for classifiers would be good examples.

Acknowledgments

This research was supported in part by Molecular Evolutionary Computing (MEC) project of MKE, in part by IT R&D Program of MKE/KEIT (10035348, mLIFE), in part by NRF grants of MEST (No. 2012-0005643, Videome; No. 2012-0005801, BrainNet), and in part by the BK21-IT program funded by Korean Government (MEST). KAY was supported by the Korea Research Foundation (KRF-2008-532-C00031). The ICT at Seoul National University provided research facilities for this study. The authors thank Russell J. Deaton for his careful proofreading and invaluable comments on this paper.

Appendix A. Mathematical formulation of a molecular pattern classifier

Here we briefly summarize the mathematical formulation of a molecular pattern classifier in our previous work (Lim et al., 2010) which builds the basis of the present work. A conventional binary linear classifier takes a real-valued vector, $\mathbf{x} = (x_0, x_1, \dots, x_n)$, as an input pattern, computes a linear combination of its elements and compares the result to a threshold for classification as follows:

$$g(\mathbf{x}; \mathbf{w}^0) = \begin{cases} G_1, & \text{if } \sum_{i=0}^n x_i w_i^0 \geq 0 \\ G_2, & \text{otherwise} \end{cases}, \quad (\text{A.1})$$

where G_1 and G_2 are class labels, $\mathbf{w}^0 = (w_0^0, w_1^0, \dots, w_n^0)$, w_0^0 is a bias term and $x_0 = 1$ as a dummy input for the bias. In this model, w_i^0 may be either positive or negative. In the meantime, a molecular pattern classifier computes two linear combinations of pattern elements, and then compares them for final decision-making as follows:

$$f(\mathbf{x}; \mathbf{w}) = \begin{cases} G_1, & \text{if } s(\mathbf{x}; \mathbf{w}) \geq s^c(\mathbf{x}; \mathbf{w}) \\ G_2, & \text{otherwise} \end{cases}, \quad (\text{A.2})$$

where $\mathbf{w} = (w_0, w_1, \dots, w_n)$ and w_0 is a bias term. Unlike the conventional classifier, w_i is always nonnegative, so that it can be represented by molecular mixing ratios. (A.2) can be made equivalent to (A.1) by setting w_i as follows.

$$w_i = \frac{1 + \lambda w_i^0}{2}, \quad i = 0, 1, \dots, n, \quad (\text{A.3})$$

where λ is a scaling constant to keep w_i in the range [0,1].

References

- Banzhaf, W., Dittrich, P., Rauhe, H., 1996. Emergent computation by catalytic reactions. *Nanotechnology* 7 (4), 307–314.
- Benenson, Y., 2009. Biocomputers: from test tubes to live cells. *Mol. Biosyst.* 5 (7), 675–685.
- Chen, J., Deaton, R., Wang, Y.Z., 2005. A DNA-based memory with *in vitro* learning and associative recall. *Natural Comput.* 4 (2), 83–101.
- Chen, R.C.S., Yang, S.J.H., 2010. Applying DNA computation to intractable problems in social network analysis. *Biosystems* 101 (3), 222–232.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S., 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286 (5439), 531–537.
- Heid, C.A., Stevens, J., Livak, K.J., Williams, P.M., 1996. Real time quantitative PCR. *Genome Res.* 6 (10), 986–994.
- Henkel, C.V., Bäck, T., Kok, J.N., Rozenberg, G., Spaink, H.P., 2007. DNA computing of solutions to knapsack problems. *Biosystems* 88 (1–2), 156–162.
- Henkel, C.V., Bladergroen, R.S., Balog, C.I.A., Deelder, A.M., Head, T., Rozenberg, G., Spaink, H.P., 2005. Protein output for DNA computing. *Natural Comput.* 4 (1), 1–10.
- Jamieson, T., Bakhshi, R., Petrova, D., Pocock, R., Imani, M., Seifalian, A.M., 2007. Biological applications of quantum dots. *Biomaterials* 28 (31), 4717–4732.
- Khodor, J., Gifford, D.K., 1997. The efficiency of sequence-specific separation of DNA mixtures for biological computing. In: Rubin, H., Harlan Wood, D. (Eds.),

- Preliminary Proceedings of Third Annual DIMACS Meeting for DNA Computers 1997, pp. 39–46.
- Kim, J.H., Chaudhary, S., Ozkan, M., 2007. Multicolour hybrid nanoprobe of molecular beacon conjugated quantum dots: FRET and gel electrophoresis assisted target DNA detection. *Nanotechnology* 18 (19), 195105.
- Kricka, L.J., 2002. Stains, labels and detection strategies for nucleic acids assays. *Ann. Clin. Biochem.* 39, 114–129.
- Lee, I.-H., Yang, K.-A., Lee, J.-H., Park, J.-Y., Chai, Y.G., Lee, J.-H., Zhang, B.-T., 2008a. The use of gold nanoparticle aggregation for DNA computing and logic-based biomolecular detection. *Nanotechnology* 19 (39), 395103.
- Lee, J.-H., Lee, S.H., Chung, W.-H., Lee, E.S., Park, T.H., Deaton, R., Zhang, B.-T., 2011. A DNA assembly model of sentence generation. *Biosystems* 106 (1), 51–56.
- Lee, S.H., Lim, H.-W., Yang, K.A., Yoo, S.-I., Zhang, B.-T., Park, T.H., 2008b. Weighted sum computation in vitro using differentially labeled molecular beacon mixture. In: Goel, A., et al. (Eds.), *Preliminary Proceedings of the 14th International Meeting on DNA Computing 2008*, p. 16.
- Lim, H.-W., Lee, S.H., Yang, K.A., Lee, J.Y., Yoo, S.-I., Park, T.H., Zhang, B.-T., 2010. *In vitro* molecular pattern classification via DNA-based weighted-sum operation. *Biosystems* 100 (1), 1–7.
- Lim, H.-W., Yun, J.-E., Jang, H.-M., Chai, Y.-G., Yoo, S.-I., Zhang, B.-T., 2003. Version space learning with DNA molecules. *Lect. Notes Comput. Sci. (DNA8)* 2568, 143–155.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., Lockhart, D.J., 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21 (Suppl. 1), 20–24.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebet, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R., 2005. MicroRNA expression profiles classify human cancers. *Nature* 435 (7043), 834–838.
- Macdonald, J., Li, Y., Sutovic, M., Lederman, H., Pendri, K., Lu, W., Andrews, B.L., Stefanovic, D., Stojanovic, M.N., 2006. Medium scale integration of molecular logic gates in an automaton. *Nano Lett.* 6 (11), 2598–2603.
- Markham, N.R., Zuker, M., 2005. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* 33, W577–W581.
- Mills Jr., A.P., 2002. Gene expression profiling diagnosis through DNA molecular computation. *Trends Biotechnol.* 20 (4), 137–140.
- Mills Jr., A.P., Turberfield, M., Turberfield, A.J., Yurke, B., Platzman, P.M., 2001. Experimental aspects of DNA neural network computation. *Soft Comput.* 5 (1), 10–18.
- Mills Jr., A.P., Yurke, B., Platzman, P.M., 1999. Article for analog vector algebra computation. *Biosystems* 52 (1/3), 175–180.
- Mirkin, C.A., Letsinger, R.L., Mucic, R.C., Storhoff, J.J., 1996. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* 382 (6592), 607–609.
- Oliver, J.S., 1997. Matrix multiplication with DNA. *J. Mol. Evol.* 45 (2), 161–167.
- Ramlan, E.I., Zauner, K.-P., 2011. Design of interacting multi-stable nucleic acids for molecular information processing. *Biosystems* 105 (1), 14–24.
- Reif, J.H., LaBean, T.H., 2007. Autonomous programmable biomolecular devices using self-assembled DNA nanostructures. *Commun. ACM* 50 (9), 46–53.
- Rinaudo, K., Bleris, L., Maddamsetti, R., Subramanian, S., Weiss, R., Benenson, Y., 2007. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* 25 (7), 795–801.
- Seelig, G., Soloveichik, D., Zhang, D.Y., Winfree, E., 2006. Enzyme-free nucleic acid logic circuits. *Science* 314 (5805), 1585–1588.
- Seeman, N.C., Wang, H., Yang, X., Liu, F., Mao, C., Sun, W., Wenzler, L., Shen, Z., Sha, R., Yan, H., Wong, M.H., Sa-Ardy, P., Liu, B., Qiu, H., Li, X., Qi, J., Du, S.M., Zhang, Y., Mueller, J.E., Fu, T.J., Wang, Y., Chen, J., 1998. New motifs in DNA nanotechnology. *Nanotechnology* 9 (3), 257–273.
- Simmel, F.C., Dittmer, W.U., 2005. DNA nanodevices. *Small* 1 (3), 284–299.
- Stojanovic, M.N., Mitchell, T.E., Stefanovic, D., 2002. Deoxyribozyme-based logic gates. *J. Am. Chem. Soc.* 124 (14), 3555–3561.
- Storhoff, J.J., Lucas, A.D., Garimella, V., Bao, Y.P., Muller, U.R., 2004. Homogeneous detection of unamplified genomic DNA sequences based on colorimetric scatter of gold nanoparticle probes. *Nat. Biotechnol.* 22 (7), 883–887.
- Tani, H., Kanagawa, T., Kurata, S., Teramura, T., Nakamura, K., Tsuneda, S., Noda, N., 2007. Quantitative method for specific nucleic acid sequences using competitive polymerase chain reaction with an alternately binding probe. *Anal. Chem.* 79 (3), 974–979.
- Tasara, T., Angerer, B., Damond, M., Winter, H., Dorhofer, S., Hubscher, U., Amacker, M., 2003. Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA. *Nucleic Acids Res.* 31 (10), 2636–2646.
- Tyagi, S., Kramer, F.R., 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* 14 (3), 303–308.
- Vet, J.A.M., Majithia, A.R., Marras, S.A.E., Tyagi, S., Dube, S., Poiesz, B.J., Kramer, F.R., 1999. Multiplex detection of four pathogenic retroviruses using molecular beacons. *Proc. Natl. Acad. Sci. USA* 96 (11), 6394–6399.
- Waggoner, A., 2006. Fluorescent labels for proteomics and genomics. *Curr. Opin. Chem. Biol.* 10 (1), 62–66.
- Xiaohui, H., Zhixiang, Y., Lingying, Z., Juan, H., 2009. Molecular beacon based DNA computing model for 0–1 programming problem. In: Xu, J. (Ed.), *Preliminary Proceedings of Fourth International Conference on Bio-inspired Computing, BICTA 2009*, pp. 1–5.
- Yin, P., Choi, H.M.T., Calvert, C.R., Pierce, N.A., 2008. Programming biomolecular self-assembly pathways. *Nature* 451 (7176), 318–U314.
- Zhang, B.-T., 2008. Hypernetworks: a molecular evolutionary architecture for cognitive learning and memory. *IEEE Comput. Intell. Mag.* 3 (3), 49–63.
- Zhang, B.-T., Jang, H.-Y., 2005. A Bayesian algorithm for *in vitro* molecular evolution of pattern classifiers. *Lect. Notes Comput. Sci. (DNA10)* 3384, 458–467.
- Zhang, B.-T., Kim, J.-K., 2006. DNA hypernetworks for information storage and retrieval. *Lect. Notes Comput. Sci. (DNA12)* 4287, 298–307.
- Zhang, D.Y., Seelig, G., 2010. DNA-based fixed gain amplifiers and linear classifier circuits. In: Sakakibara, Y., et al. (Eds.), *Preliminary Proceedings of the 16th International Conference on DNA Computing and Molecular Programming 2010*, pp. 127–134.
- Zhang, D.Y., Turberfield, A.J., Yurke, B., Winfree, E., 2007. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* 318 (5853), 1121–1125.
- Zhang, H., Liu, X., 2011. A CLIQUE algorithm using DNA computing techniques based on closed-circle DNA sequences. *Biosystems* 105 (1), 73–82.