Combining Randomness and a High-Capacity DNA Memory

Atsushi Kameda[1], Satoshi Kashiwamura[2], Masahito Yamamoto[1,2],
Azuma Ohuchi[1,2], and Masami Hagiya[1,3]

[1] Japan Science and Technology Corporation (JST-CREST)
[2] Graduate School of Information Science and Technology, Hokkaido University
[3] Graduate School of Information Science and Technology, University of Tokyo
hagiya@is.s.u-tokyo.ac.jp

Abstract. In molecular computing, it has long been a central focus to realize robust computational processes by suppressing the randomness of molecular reactions. To this end, several methods have been developed to control hybridization reactions of DNA molecules by optimizing DNA sequences and reaction parameters. However, another direction in the field is to take advantage of molecular randomness rather than avoid it. In this paper, we show that randomness can be useful in combination with a huge-capacity molecular memory, and demonstrate its application to an existing technology — DNA ink.

Keywords: DNA memory, molecular memory, DNA ink, randomness.

1 Introduction

A central focus in the field of molecular computing is to suppress randomness of molecular reactions to realize robust computational processes. Methods that employ this strategy include designing precise DNA sequences based on energy predictions of DNA secondary structures, and carefully controlling DNA hybridization by tuning reaction conditions such as temperature and salt concentration.

For example, errors in the algorithmic self-assembly of DNA tiles occur because sticky ends that are not completely complementary may hybridize with nonzero probability. Suppressing such errors is a central challenge in DNA nanotechnology. In addition to designing DNA sequences that reduce error probability, new machineries such as proof-reading and self-healing tiles have been proposed [3,12].

Some molecular machines, such as Yurke's tweezers [16] or our photo-regulated hairpin machine [14], change their conformation according to inputs from the external environment. As a result, they can make state transitions, move toward a specified direction, or produce outputs to the environment. However, conformational change cannot occur with 100% probability, and thus to design robust molecular machines, it is crucial to design DNA sequences that promote intended conformational change and prohibit unintended changes. In the case of

M.H. Garzon and H. Yan (Eds.): DNA 13, LNCS 4848, pp. 109–118, 2008.

DNA, because transformation of secondary structures roughly determines conformational change, predicting the energy landscape of secondary structures is extremely important [10].

However, applications can also take advantage of randomness. In the seminal work by Adleman [1], which initiated the field of DNA computing, the random generation of paths in a directed graph was achieved via random hybridization reactions. However, random generation was not essential, but only substituted *complete enumeration,* which was required for solving the Hamiltonian path problem and should be implemented by molecules. Therefore, our goal was to create an application in which molecular randomness is essential.

In previous work, we developed a huge-capacity DNA memory, in which each molecular address consists of 6 hexadecimal digits, and the entire address space is about 16.8 million (words) [7,9]. The whole memory is managed as a solution of about 1μl, using standard experimental techniques such as PCR. In the present work, we employ molecular randomness caused by statistical fluctuations in a sample with a low copy number of molecules. By simply diluting the solution, one can easily obtain a unique memory state that cannot be replicated. Moreover, such a memory state can be amplified by PCR. Therefore, by combining randomness and a huge-capacity memory, it is possible to construct DNA ink, for example, which can never be reconstructed.

Here, we briefly describe our huge-capacity DNA memory, called Nested Primer Molecular Memory (NPMM). Then we summarize the concept of DNA ink and explain how to produce it using our proposed strategy. Finally, we report the results of a preliminary experiment we conducted to demonstrate the feasibility of our method.

2 NPMM

We developed the NPMM under the JST CREST Molecular Memory Project [9]. Our goal was to overwhelm the random pool of size 2^{20} (for solving a 20-variable SAT problem) realized by Adleman's group in 2002 [2], and construct a DNA memory that can easily be managed by well established and standard experimental techniques, such as PCR [7].

The result was a DNA memory composed of memory molecules with the structure shown in Fig. 1 [15]. A memory molecule is a double strand of DNA consisting of data at its center surrounded by 3-digit addresses. Because each digit is chosen from among 16 sequences, and the address of each memory molecule is 6 digits long, the whole address space is about 16.8 million (words), i.e., the whole set of 6-digit hexadecimal numbers.

The DNA memory is managed as a solution of about 1μl, as shown in Fig. 2.

To access each memory molecule, PCR (the established method for copying DNA) is repeated, using address digits as *nested primers.* First, PCR is performed using the outermost two digits denoted by `C` (`CL` and `CR`) as primers, so

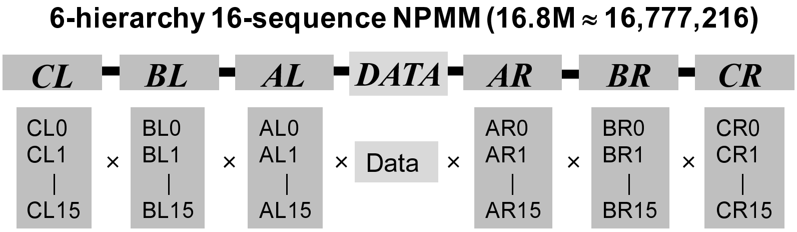

Fig. 1. The structure of a memory molecule in NPMM

Fig. 2. NPMM in a 1-μl solution

that the molecules with CLi and CRj specified in their address are extracted and amplified. Then the process is repeated for BL and BR (BLi and BRj), and AL and AR (ALi and ARj).

It took a few years to construct the 16.8-million-address DNA memory [15] because we had to develop new technologies or refine existing ones. A summary follows.

Designing 16 sequences for each of 6 digits plus bridge sequences: We designed 16 different 20-mer sequences for each of the 6 address digits, i.e., 96 different sequences in total. In addition, we designed bridge sequences, which are placed between adjacent address digits and also used to bridge data and the innermost digits. These sequences should not hybridize with one another, but only with their complementary counterparts. We developed a new design method for these sequences, called a *two-step search* [8].

Protocol for constructing memory molecules for each address: Molecules of 16.8 million addresses should be synthesized as uniformly as possible. NPMM is constructed in three steps: concatenating address digits with data from the innermost to the outermost, performing PCR on DNA molecules whose bridge sequences are complementary, and hybridizing these together. Regarding uniformity, the ratio between the most and least concentrated types of molecules was estimated between 2 and 3 during each step.

Protocol for accessing memory molecules with the specified address: To selectively amplify the molecules with a specified address, we should optimize various reaction parameters for PCR, such as the number of cycles. As a result, we successfully accessed a single address out of 16.8 million addresses. To date, we have accessed 16 different addresses.

Consequently, each address in the DNA memory consists of about 200 to 250 molecules. Using a mathematical model of our PCR reaction, we determined that the current capacity is almost maximal for correctly accessing memory molecules by their addresses.

3 DNA Ink

To the best of our knowledge, the concept of DNA ink was first proposed and tested by Tsujii *et al.* in 2001 [5,6]. DNA molecules of a given sequence are diluted in ink; after the ink is applied to paper, then extracted, the DNA molecules in the ink can be analyzed to determine the sequence. DNA ink can be used for various purposes related to encryption, steganography, and authentication. For example, a secret key for encrypted communication can be sent in DNA ink. Contracts can be signed using DNA ink, and paper money can be printed with DNA ink to avoid counterfeiting. Brand-name products, such as Chanel or Prada, can be made of strings dyed with DNA ink to ensure authenticity. It is also possible to spray DNA ink over brand-name foods, such as Kobe beef, because DNA is completely harmless in foods.

Recently, Suyama *et al.* developed a practical DNA ink under the JST CREST Molecular Memory Project. The ink is composed of about 300 pairs of orthonormal sequences, although it could be expanded to more than 10,000 sequences, and thus can store 300 bits of information by including or excluding specific sequences. The orthonormal sequences employed in their DNA ink were originally developed for DNA computing [13,11] and gene expression profiling [4].

The original DNA ink developed by Tsujii *et al.* [5,6] was limited in that it required DNA sequencing to analyze the extracted ink. More importantly, the contents of the ink could be exposed by attaching primers to DNA molecules and amplifying them by PCR. Suyama *et al.* solved these problems by mixing DNA ink with dummy DNA (which has primers in common with true DNA and serves to mask the latter) and noise DNA (which consists of unused orthonormal sequences that hinder whole DNA amplification). Suyama *et al.* also prepared a detection kit for their DNA ink, which contains dummy primers so that even if the kit is analyzed, the true primers will not be revealed. To detect the primers, one must use a special DNA chip, which itself also contains dummy probes so that if the chip alone is analyzed, the true probes will not be revealed.

Suyama *et al.* succeeded in recovering the correct information in DNA ink extracted from paper using their detection kit and special DNA chip.

4 DNA Ink Constructed by Inducing Randomness at the Molecular Level

With our method, molecular randomness is introduced by a very simple reaction: diluting the solution. For example, if we dilute a given solution twice and dispense the results into two tubes, A and B, then a single molecule in the original solution is either in tube A or in tube B, with a probability of 1/2. If the original solution contained 100 molecules and was diluted 100 times, then the probability that the resulting tube contains one or more molecules is about $1 - e^{-1} = 0.63$, because the number of chosen molecules roughly follows the Poisson distribution with the average value 1 and the probability that no molecule is chosen is approximately equal to e^{-1}. Similarly, if the original solution was diluted 200 times, then the probability that the resulting tube contains one or more molecules is about $1 - e^{-0.5} = 0.39$.

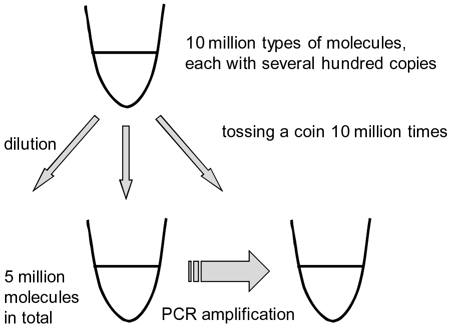
10 million types of molecules, each with several hundred copies
dilution
tossing a coin 10 million times
5 million molecules in total
PCR amplification

Fig. 3. DNA ink made by inducing randomness at the molecular level

Assume that a given solution contains 10 million types of molecules, each with several hundred copies (Fig. 3; note that NPMM actually has 16.8 million types of molecules, each with about 200 copies), and it is diluted to 5 million molecules. The probability that at least one molecule of any given type remains in the diluted solution is about $1 - e^{-0.5} = 0.39$, as described above. This process of dilution is akin to tossing a coin 10 million times, when one side of the coin turns up with a probability of 0.39.

After dilution, the solution is uniformly amplified by PCR. In NPMM, amplification requires common outermost primers in each memory molecule, or all outermost digits as primers.

After amplification, each molecule that survived should have a reasonable number of copies. Therefore, even if the amplified solution is further diluted, those copies will remain. Let us call the solution obtained by PCR *the master ink.* Each solution obtained by diluting the master ink should have the same ingredients as those of the master ink. Therefore, each diluted solution can be used as DNA ink corresponding to the master ink. The molecules extracted from inked paper are then compared with those in the master ink. To this end, DNA ink should contain an enough number of copies of each molecule.

The contents of the master ink can never be revealed completely; checking for each type of molecule among millions of molecules is prohibitively time-consuming. However, correctly verifying whether a given DNA ink was copied from the master ink is much simpler, by checking for a reasonable number of addresses in the given DNA ink and the master ink. In this way, DNA ink can be created by combining molecular randomness with a huge-capacity DNA memory. Note that the contents of the master ink are not known even to the creator of the ink.

One note of caution: if the DNA ink extracted from paper is amplified as a whole, then the forged ink can be used as if it were obtained from the master ink. To avoid this problem, one can apply various methods as in Suyama's DNA ink. For example, one can add short junk DNA molecules as noise, which are more easily amplified by PCR.

5 Preliminary Experiment

To test the feasibility of our method, we conducted a preliminary experiment. The overall setting of the experiment is shown in Fig. 4. We employed a partial instance of NPMM obtained by completing the first two steps in the construction process. This partial instance had four address digits, each with 10 sequences. Thus, we had a total of 10,000 addresses, each with about 3×10^5 copies in a 1-μl solution. Because molecules in this memory have common bridge sequences at both ends, they can be uniformly amplified using the bridge sequences as primers. This greatly simplified the experiment.

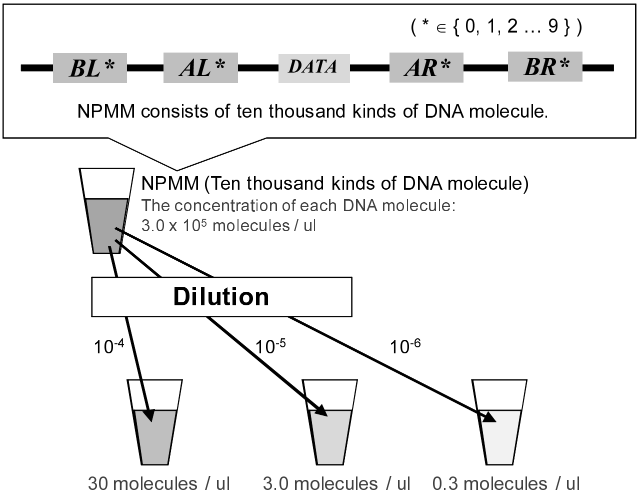
(* ∈ { 0, 1, 2 ... 9 })
BL*
AL*
DATA
AR*
BR*
NPMM consists of ten thousand kinds of DNA molecule.
NPMM (Ten thousand kinds of DNA molecule)
The concentration of each DNA molecule:
3.0 x 10^5 molecules / ul
Dilution
10^-4
10^-5
10^-6
30 molecules / ul
3.0 molecules / ul
0.3 molecules / ul

Fig. 4. Setting of the preliminary experiment

A. First PCR amplification

1. In a PCR reaction tube, prepare the following.
- distilled water 4.4μl
- 10 x PCR buffer 1.0μl
- 2 mM dNTP mix 1.0μl
- 25 mM $MgSO_4$ 0.4μl
- diluted NPMM solution 1.0μl (dilution ratios: 10^{-4}, 10^{-5} and 10^{-6})
- 5 μM Primer [BL0] and c[BR0] 1.0μl each ('c' denotes complementary sequence)
- KOD plus DNA polymerase (TOYOBO) 0.2μl
- total volume 10μl

2. Perform 20 cycles of PCR
- preheat (94°C for 2 min.)

↓
- denature (94°C for 20 sec.)
- anneal and extension (65°C for 10 sec.)

(Repeated 20 cycles)
PCR was performed by PTC-200 peltier thermal cycler (MJ Research).

B. Secondary PCR

1. Dilute the first PCR products by the ratio 10^{-3}.
2. In a 96-well PCR reaction plate, prepare the following.
- distilled water 3.4μl
- 10 x PCR buffer 1.0μl
- 2 mM dNTP mix 1.0μl
- 25 mM $MgSO_4$ 0.4μl
- Diluted first PCR product 1.0μl
- 10 pM Primer set 1.0μl each
- KOD plus DNA polymerase (TOYOBO) 0.2μl
- 1/1000 diluted SYBR green I 1.0μl
- total volume 10μl

The primer pairs used in the experiment were ([AL0] and c[AR0]), ([AL0] and c[AR1]), ([AL1] and c[AR0]), and ([AL1] and c[AR1]).

2. Perform 40 cycles of realtime PCR.
- preheat 95°C for 2 min.

↓
- denature 94°C for 20 sec.
- anneal and extension 65°C for 10 sec.

(Repeated 40 cycles)

3. After PCR, perform melting curve analysis.

The confirmation of the existence of each address was made by the results of amplification check comprised of measuring the SYBR green I fluorescence and analyzing the melting curve. Realtime PCR was performed by DNA Engine OPTICON 2 (MJ Research).

Fig. 5. Details of the PCR protocol

The details of the PCR protocol employed in the experiment are described in Fig. 5.

Results: We diluted the solution by 10^{-4}, 10^{-5}, and 10^{-6}, and then checked for four addresses. First, we performed an initial PCR using the two digits BL0BR0 as primers, then performed a second PCR using AL0AR0, AL0AR1, AL1AR0, and AL1AR1 as primers. This trial was repeated eight times for each set of primers (Fig. 6).

At dilutions of 10^{-4} and 10^{-5}, we detected molecules for each of the four addresses. Note that the probability of having one or more molecules in the diluted solution is about $1-e^{-30.0} = 0.999999 = 1$ for 10^{-4} and about $1-e^{-3.0} = 0.95$ for 10^{-5}.

	Address	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
10^-6	AL0AR0	□	□	□	□	□	□	□	■	□	□	□	■	□	■	□	■	□	□	□	■	■	■	■	■	■	■	■	□	■	■	□	□	□	□	□
	AL0AR1	■	□	□	□	■	□	□	■	□	■	□	□	■	□	□	□	□	□	■	□	□	□	□	■	□	□	□	□	■	■	□	□	□	□	□
	AL1AR0	■	■	□	□	□	□	□	■	■	□	□	□	■	■	□	■	■	■	□	■	■	■	■	■	■	□	□	□	■	■	□	□	□	■	□
	AL1AR1	■	□	■	□	■	□	□	■	□	■	□	□	■	□	□	□	□	□	□	□	□	□	■	□	□	□	□	□	□	■	■	■	□	■	□
10^-5	AL0AR0	■	■	■	■	■	■	■	■																											
	AL0AR1	■	■	■	■	■	■	■	■																											
	AL1AR0	■	■	■	■	■	■	■	■																											
	AL1AR1	■	□	■	■	■	■	■	■																											
10^-4	AL0AR0	■	■	■	■	■	■	□	■																											
	AL0AR1	■	■	■	■	■	■	■	■																											
	AL1AR0	■	■	■	■	■	■	□	■																											
	AL1AR1	■	■	■	■	■	■	■	■																											

■ : The address exists.
□ : The address does not exist.

	Address	Exists	Does not exist	Total	The probability of existence of the address
10^-6	AL0AR0	14	21	35	0.400
	AL0AR1	9	26	35	0.257
	AL1AR0	18	17	35	0.514
	AL1AR1	10	25	35	0.286
10^-5	AL0AR0	8	0	8	1.000
	AL0AR1	8	0	8	1.000
	AL1AR0	8	0	8	1.000
	AL1AR1	7	1	8	0.875
10^-4	AL0AR0	7	1	8	0.875
	AL0AR1	8	0	8	1.000
	AL1AR0	7	1	8	0.875
	AL1AR1	8	0	8	1.000

Fig. 6. Results of the preliminary experiment

At a dilution of 10^{-6}, only a random detection of addresses was achieved. Therefore, we conducted more trials at this dilution, 35 in total. As a result, we obtained the detection ratios for the four addresses: 14/35=0.400, 9/35=0.257, 18/35=0.514, and 10/35=0.286. The probability of having one or more molecules in the diluted solution was estimated at about $1 - e^{-0.3} = 0.26$ in this case. Although some correlation existed among addresses (for ARj in particular), the resulting pattern was nearly random.

It may be questionable whether PCR succeeded in amplifying a single molecule because the actual number of molecules that remained in the diluted solutions was not known. We carefully excluded wrong PCR products in the experiment. However, we plan to perform DNA sequencing to assess the correctness of the amplified solution.

One of the diluted solutions was amplified by the outermost primers, and the result was further diluted. We checked each address in each diluted solution, and verified that the detection pattern was uniform, which indicates that the detection pattern using the outermost primers was correctly amplified by PCR.

6 Concluding Remark

Molecular authentication is considered to have a potential of becoming a killer application of DNA and molecular computation. On the other hand, constructing a random pool of DNA molecules has long been an active research topic in the field. This paper proposed to apply achievements of the latter research to the former application by inducing randomness at the molecular level.

We plan to eventually apply the 16.8-million-address molecular memory to DNA ink. However, there is much to be done in analyzing the results of the preliminary experiments. As noted above, the correctness of the PCR products should be confirmed by DNA sequencing.

Acknowledgments

The work presented in this paper was partially supported by Grand-in-Aid for Scientific Research on Priority Area No.14085202, Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Leonard, M.: Adleman: Molecular Computation of Solutions to Combinatorial Problems. Science 266, 1021–1024 (1994)
2. Braich, R.S., Chelyapov, N., Johnson, C., Rothemund, P.W.K., Adleman, L.: Solution to a 20-Variable 3-SAT Problem on a DNA Computer. Science 296, 499–502 (2002)
3. Fujibayashi, K., Murata, S.: A Method of Error Suppression for Self-assembling DNA Tiles. In: Ferretti, C., Mauri, G., Zandron, C. (eds.) DNA Computing. LNCS, vol. 3384, pp. 113–127. Springer, Heidelberg (2005)
4. Gotoh, O., Sakai, Y., Mawatari, Y., Gunji, W., Murakami, Y., Suyama, A.: Normalized molecular encoding method for quantitative gene expression profiling. In: Carbone, A., Pierce, N.A. (eds.) DNA Computing. LNCS, vol. 3892, p. 395. Springer, Heidelberg (2006)
5. Itakura, Y., Hashiyada, M., Nagashima, T., Fukuyama, M.: Validation Experiment Report on DNA Information for Personal Identification (Part I). Technical Report of IEICE ISEC2001-12 (2001-05), The Institute of Electronics, Information and Communication Engineers, pp. 1–8 (2001) (in Japanese)

6. Itakura, Y., Hashiyada, M., Nagashima, T., Tsuji, S.: Validation Experiment Report on DNA Information for Personal Identification (Part II). Technical Report of IEICE ISEC2001-13 (2001-05), The Institute of Electronics, Information and Communication Engineers , 9–16 (2001) (in Japanese)
7. Kashiwamura, S., Yamamoto, M., Kameda, A., Shiba, T., Ohuchi, A.: Hierarchical DNA Memory Based on Nested PCR. In: Hagiya, M., Ohuchi, A. (eds.) DNA Computing. LNCS, vol. 2568, pp. 112–123. Springer, Heidelberg (2003)
8. Kashiwamura, S., Kameda, A., Yamamoto, M., Ohuchi, A.: Two-Step Search for DNA Sequence Design. IEICE E87-A(6), 1446–1453 (2004)
9. Kashiwamura, S., Yamamoto, M., Kameda, A., Shiba, T., Ohuchi, A.: Potential for Enlarging DNA Memory: The Validity of Experimental Operations of Scaled-up Nested Primer Molecular Memory. BioSystems 80, 99–112 (2005)
10. Kubota, M., Hagiya, M.: Minimum Basin Algorithm: An Effective Analysis Technique for DNA Energy Landscapes. In: Ferretti, C., Mauri, G., Zandron, C. (eds.) DNA Computing. LNCS, vol. 3384, pp. 202–214. Springer, Heidelberg (2005)
11. Nitta, N., Suyama, A.: Autonomous biomolecular computer modeled after retroviral replication. In: Chen, J., Reif, J.H. (eds.) DNA Computing. LNCS, vol. 2943, pp. 203–212. Springer, Heidelberg (2004)
12. Soloveichik, D., Winfree, E.: Complexity of Compact Proofreading for Self-assembled Patterns. In: Carbone, A., Pierce, N.A. (eds.) DNA Computing. LNCS, vol. 3892, pp. 305–324. Springer, Heidelberg (2006)
13. Suyama, A.: Programmable DNA computer with application to mathematical and biological problems. Preliminary Proceedings of the Eighth International Meeting on DNA Based Computers, 91 (2002)
14. Takahashi, K., Yaegashi, S., Asanuma, H., Hagiya, M.: Photo- and Thermoregulation of DNA Nanomachines. In: Carbone, A., Pierce, N.A. (eds.) DNA Computing. LNCS, vol. 3892, pp. 336–346. Springer, Heidelberg (2006)
15. Yamamoto, M., Kashiwamura, S., Ohuchi, A.: DNA Memory with 16.8M addresses. DNA13 (submitted 2007)
16. Yurke, B., Turberfield, A.J., Mills Jr., A.P., Simmel, F.C., Neumann, J.L.: A DNA-fuelled molecular machine made of DNA. Nature 406, 605–608 (2000)