Molecular associative memory built on DNA

Robert M. Nowak[a] and Jan J. Mulawka[a] and Andrzej Pucienniczak[b]

[a]Institute of Electronic Systems, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland;
[b]Institute of Biotechnology and Antibiotics, Staroscinska 5, 02-516 Warsaw, Poland

ABSTRACT

This paper describes an associative memory based on DNA strands practically build in laboratory. The method for suppressing DNA fragment amplification during polymerase chain reaction (PCR) was used. Such memory exhibits a number of advantages over Baum's associative molecular memory as well as traditional electronic implementations.

Keywords: Nanotechnology, DNA computing, self-assembly, associative memories

1. INTRODUCTION

In traditional computers based on electronics directly addressed memory is the most popular. Organisation of this memory resembles a vector. The memory cell corresponds to the element of the vector. Thus, the location of appropriate elements in the vector indicate the address. Such memory may have a large capacity. The other advantages of this model are the following: a short access time, an easy and regular structure. However, the main disadvantages are as follows: the address is a short number because it indicates a position in the vector; a specified address defines one and only one memory cell.

On the other hand, the associative memory or content addressable memory is a set of memory cells as compared to a vector in the previous model. The memory cell stores its address and data, thus it has two parts: the address or cue portion and the data portion. The specified address or cue can be associated with a single, many or no memory cells. Electronics associative memories have small capacity. Their application is limited by the addressing process in which the memory cell with the specified cue is found out. It needs to build as many comparators as there are cells in the memory. Such memories can be simulated in directly addressed memories by using additional software, but it increases access time.

The associative memory may be considered as an implementation of binary relation. The binary relation $\mathbf{R}$ is the subset of Cartesian product $\mathbf{C} \times \mathbf{D}$, where $\mathbf{C}$ is called the cue set and $\mathbf{D}$ is called the data set. The access process for $a \in \mathbf{C}$ returns set $\mathbf{X} \subseteq \mathbf{D}$: $\forall x \in \mathbf{X}: (a, x) \in \mathbf{R}$. For example $\mathbf{C} = \{a, b, c\}$, $\mathbf{D} = \{x, y\}$, $\mathbf{R} = \{(a, x), (b, x), (b, y)\}$. The access for cue a returns $\mathbf{X} = \{x\}$, and for cue b returns $\mathbf{X} = \{x, y\}$, for cue c returns $\mathbf{X} = \emptyset$(empty set).

The idea of building DNA associative memory is not new. Baum suggested constructing such a memory using DNA strands in Ref. 1 . Since then not much progress in practical implementation of DNA memory has been achieved, however the general concept of such memory is promising.

2. REALIZATION OF MEMORY BASED ON DNA

The DNA molecules and basic operations such as hybridization, denaturation, ligation, cutting, PCR are described in Ref. 2–4.

Memory described in this paper is composed of a set of memory cells where each cell is represented by a double stranded DNA molecule. Memory is created in a vessel. Each memory cell is built from a data portion and a cue or an address portion as illustrated in Fig. 1a. The cue portion has a constant length, which defines a number of

Further author information:
Robert M. Nowak: E-mail: R.M.Nowak@elka.pw.edu.pl
Andrzej Pucienniczak: E-mail: apl@iba.waw.pl

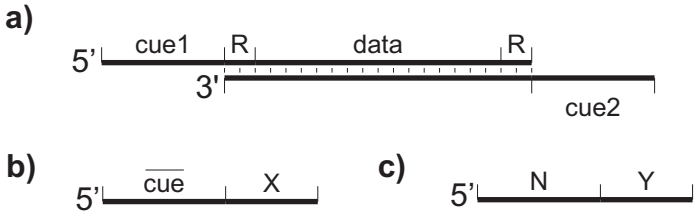

Figure 1. The molecules representing memory cell and other signals. a) Memory cell; ends store cues, central part stores data; b) input signal, molecule representing searched cue; c) temporary signal.

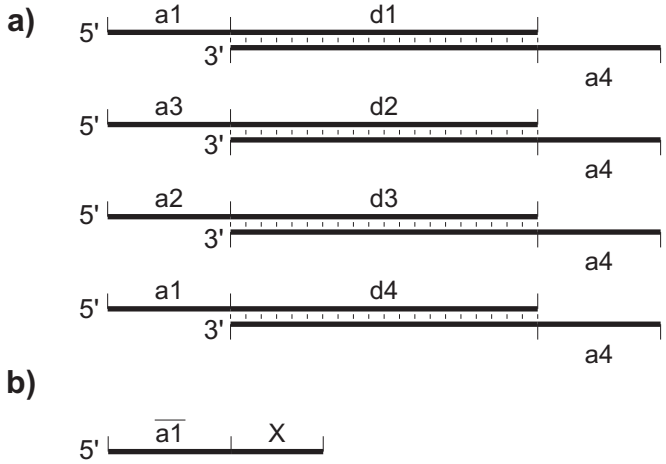

Figure 2. Molecular associative memory, an example. a)Memory consists of four memory cells and stores the data `d1`, `d2`, `d3`, `d4` respectively. b)Input signal `a1`.

possible cues (when the length is N up to 4^N different cues may be obtained). The central part of the molecule represents the stored information. This portion length is not defined. In the implementation described here the memory cell stores two cues, on both ends, as provided in Fig. 1a.

It should be mentioned, that access to a specified cell is possible by any of the stored cues. In typical situation the access to the memory cell is provided by one cue, thus sequence on one end stores this cue and sequence on the other end stores cue not used for addressing.

The searched cue is represented by a molecule called the input signal depicted in Fig. 1b. This molecule has the sequence complementary to the specified cue on its 5' end. Next, the molecule has the sequence denoted by `X`, a sequence different from any cue used in the memory.

Reading data from memory needs the molecules, called temporary signals, illustrated in Fig. 1c. These molecules are similar to the input signals. On the 5' end they have a sequence complementary to some cue, then the sequence `Y` (different from any cue) occurs. The mixture of temporary signals contains sequences complementary to any cue used in the molecular memory.

3. READING DATA FROM ASSOCIATIVE MEMORY

The associative memory of four memory cells is illustrated in Fig. 2a. The cue $a4$ is not used for addressing, thus this memory is the implementation of the relation $\mathbf{R} = \{(a1, d1), (a3, d2), (a2, d3), (a1, d4)\}$, where $\mathbf{C} = \{a1, a2, a3\}$, $\mathbf{D} = \{d1, d2, d3, d4\}$.

The input signal `a1` is depicted in Fig. 2b. As a result of retrieving it is possible to get a single, more or no molecules. In the given example the reading for $a1$ gets $\mathbf{X} = \{d1, d4\}$.

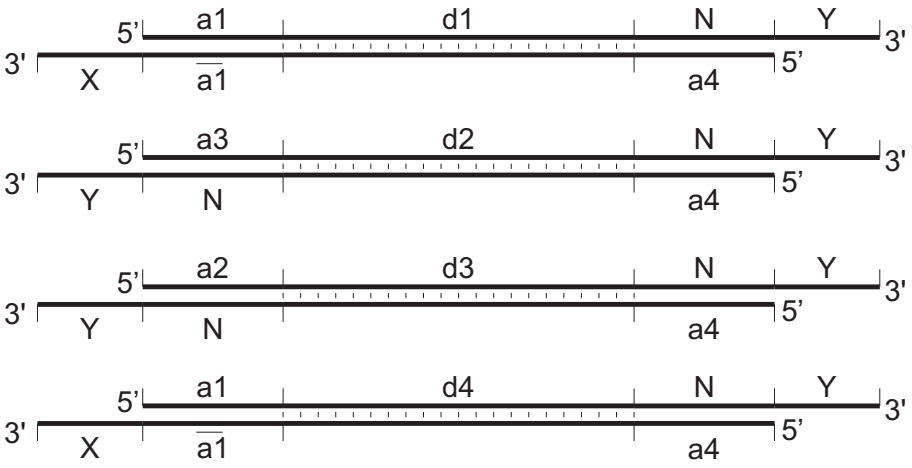

Figure 3. Reading process, memory after ligation. The molecules created from memory cells with specified cue have different sequences on their ends. The others have sequence `Y` on both ends.

The process of reading information from the memory may be described as retrieving from the vessel molecules with a specified cue part. This process is presented in Alg. 1.

```
1: adding the input signals to the vassel
   containing memory cells;
2: adding temporary signals;
3: ligation;
4: PCR with suppression.
```

Algorithm 1: Reading data from memory.

In the first step, to the vessel containing memory strings the input molecules are added. These molecules hybridize to the memory cells, which have a specified cue. This step is called the addressing step, because the proper memory cells are selected.

Then (step 2 Alg. 1), the temporary signals are added. These molecules hybridize to the rest of the memory cells - in places where the input signals have not hybridized.

Next (step 3 Alg. 1), the DNA ligase is added to the solution and the covalent bonds are created. The memory after this step is depicted in Fig. 3 The memory cells with a specified cue have sequence `X` on one sticky end and sequence `Y` on the other one. The others have sequence `Y` on both ends.

Finally (step 4 Alg. 1), the method of suppressing DNA fragment amplification during PCR described in Ref. 5, 6 is applied.

This method, called suppression PCR, is based on "pan-like" single stranded DNA structures that restrict amplification during PCR. The double stranded DNA molecules with different 5' ends (when their ends have different sequence of nucleotides) are amplified; the other DNA molecules are not. This method is illustrated in Fig. 4. As a result of this process the memory cells, which have a specified cue are withdrawn from the solution.

The suppression PCR makes possible to find out the molecules with different sticky ends in mixture containing huge number of molecules. The difference between ends can be very small, event one nucleotide. This property is useful for addressing process.

During the reading process the memory is changed. The next step is the refreshing process, which prepares the memory for the next reading. This process equalises the amount of different molecules (because after PCR the number of molecules representing the memory cells which were specified is larger than the number of the other molecules). The method described in Ref. 5 can be used. The input and the temporary signals are removed

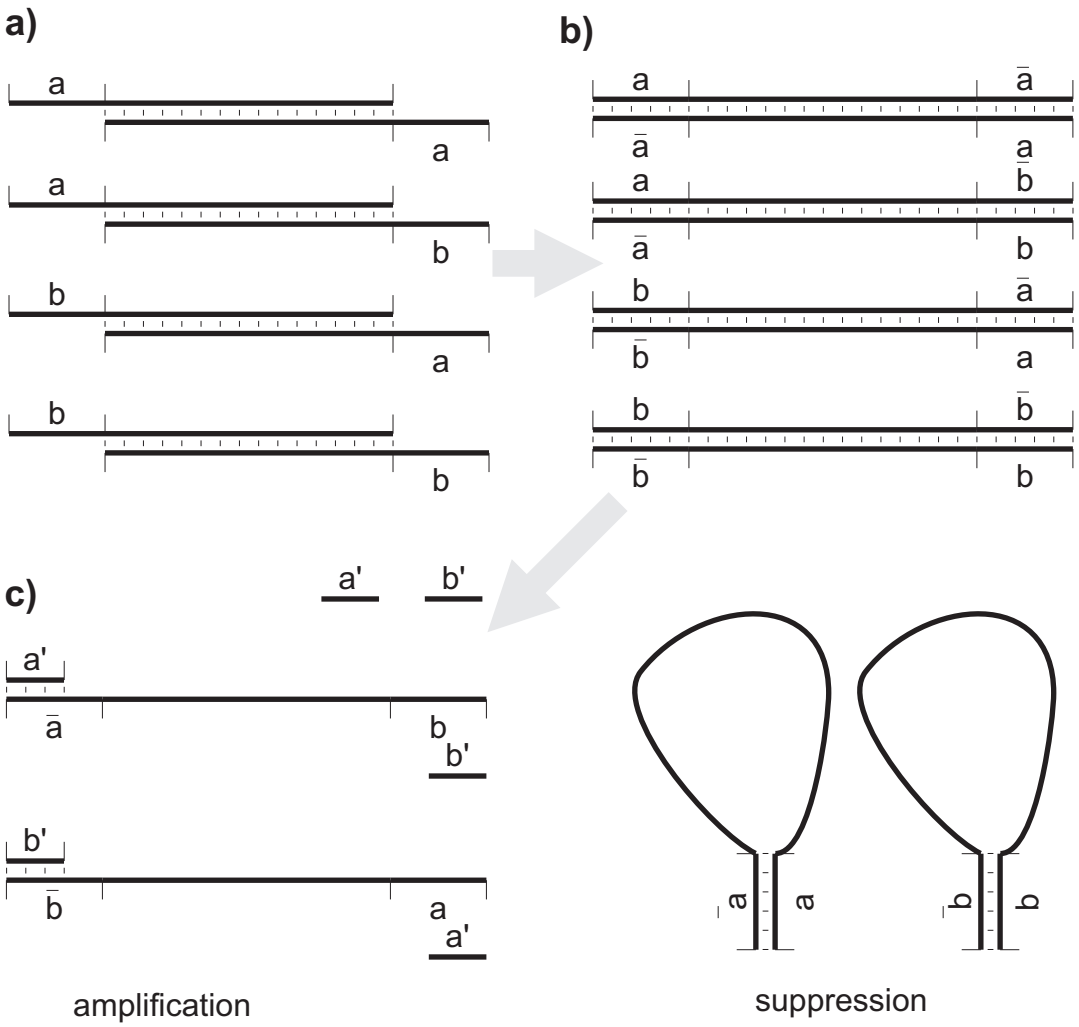

Figure 4. Method for suppressing DNA fragment amplification during PCR. a)Molecules with ends denoted by a or b, the second and third molecule have different ends; the first and the last have not. b)The first step - filling the ends. The both strings from first molecules have complementary sequences a and $\overline{\mathtt{a}}$. The strings from last molecules have sequences b end $\overline{\mathtt{b}}$. c)PCR with primers $\mathtt{a}'$ and $\mathtt{b}'$. The primers are shorter than a or b, thus primers hybridise to the target strands with second and third molecule. Some strands are amplified, others are not, because form "pan like" structures.

by cutting the molecules representing memory cells (with restriction enzymes) and next they are separated (for example in electrophoresis gel). This process restores memory and after that the memory is ready for the next reading.

4. GROUP ADDRESSING

In one reading process the mixture of different input signals can be used. In this case the memory cells with different cues are read.

When the access to the memory cell is provided by one cue (memory cell stores two cues), the set of output signals is sum of sets obtained for each cue.

For memory $\mathbf{R} = \{(a,x),(b,x),(b,y),(c,z)\}$, where $\mathbf{C} = \{a,b,c\}$, $\mathbf{D} = \{x,y,z\}$, reading for a key gives $\mathbf{X} = \{x\}$, reading for b key gives $\mathbf{X} = \{x,y\}$, and reading for c key gives $\mathbf{X} = \{z\}$. Reading for mixture cues $\mathbf{M} = \{a,b\}$ gives $\mathbf{X} = \{x,y\} = \{x\} \cup \{x,y\}$, for $\mathbf{M} = \{a,c\}$ gives $\mathbf{X} = \{x,z\} = \{x\} \cup \{z\}$.

When the access to the memory cell is provided by two cues, it is not true. Memory cells with both address which are in input mixture are not extracted (this property have the memory cells with the same cue on the both ends).

The group addressing can be achieved in other way: by using artificial nucleotides which do not bind selectively or by performing the reaction at a lower temperature at which not fully complementary nucleotides can hybridize. In this case a molecule representing a specified input signal can bind (step 1 Alg. 1) to memory cell stored similar (not equal) cue.

The group addressing increases the number of possible applications described memory.

5. EXPERIMENTAL VERIFICATION

To verify the method of realization associative memory some experiments were held in the genetic engineering laboratory. First, a simple model of such memory was designed using DNA strands. Next, retrieving data from the memory was demonstrated.

5.1. Building of memory

In experiments the memory set consisted of 5 memory cells. The length of the address part was 3. The 4 different cues were used, they differ on the second position. Each cell was created from the pUC19 plasmid (GenBank Accession No. M77789) by cutting with *Hinf*I restriction enzyme. The lengths of the memory cell molecules were different. It helped to demonstrate which molecules were obtained after the reading process, during identification on gel electrophoresis.

Name	Length	sequences on begin and end
1GG	517	AGTCCAAC − −− − − −GCCTG GGTTG − −− − − −CGGACTGA
2AC	396	AATCAGGG − −− − − −TCTTG GTCCC − −− − − −AGAACTCA
3GT	214	AGTCGACC − −− − − −TAATG GCTGG − −− − − −ATTACTTA
4CT	75	ACTCGCTG − − − − CACAG GCGAC − − − − GTGTCTTA
5AG	65	AATCGGCC− −CACTG GCCGG− −GTGACTGA

Table 1. Memory cells. The nucleotides determining differences between addresses are underlined.

The memory sequences are described in Tab. 1 and provided in Fig. 5. This memory performs the implementation of relation $\mathbf{R} = \{(G,d1),(A,d2),(C,d2),(G,d3),(T,d3),(C,d4),(T,d4),(A,d5),(G,d5)\}$. The cues are $(C) = \{A,C,G,T\}$, the stored information is $\mathbf{D} = \{d1,d2,d3,d4,d5\}$.

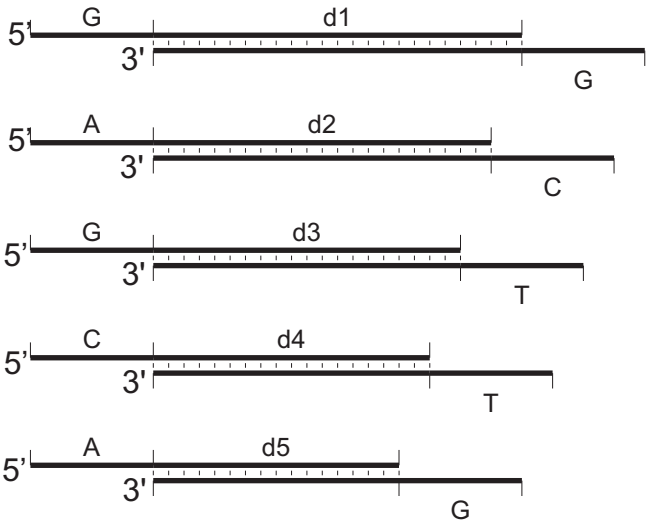

Figure 5. Memory used in experiments.

The molecule stored $d1$ is denoted as `1GG`, the molecule represented by $d2$ is denoted as `2AC`, for $d3$ is `3GT` and for $d4$ if `4CT`, and finally for $d5$ is `5AG`.

The input signals sequences are provided in Tab. 2. In experiments the temporary signals and primers for PCR are described in Tab. 3.

Name	Length	Sequence
`cueA`	18	`CCTTCATCCACCAACGTC` `TGGTTGCAGTTA`
`cueC`	18	`CCTTCATCCACCAACGTC` `TGGTTGCAGTGA`
`cueG`	18	`CCTTCATCCACCAACGTC` `TGGTTGCAGTCA`
`cueT`	18	`CCTTCATCCACCAACGTC` `TGGTTGCAGTAA`

Table 2. Input signals. The nucleotides determining differences between addresses are underlined.

Name	Length	Sequence
`TEMP`	32	`GGATGGTAGACGAAGGAACGCCGTTGTCGACG` `AACAGCTGCTNA`
`PRIMER1`	18	`CCTTCATCCACCAACGTC`
`PRIMER2`	18	`GGATGGTAGACGAAGGAA`

Table 3. Temporary signals.

5.2. Data reading from the memory

To verify the reading operation from the molecular associative memory, four independent experiments were performed. In each, a different molecule was used as the input signal. The diagram of gel electrophoresis is depicted in Fig. 6, the results are described in Tab. 4.

As follows in Fig. 6 there are two bands in line 2: one with the length of 267 bp, the second 128 bp. This means that the result consists of molecules `3GT` and `4CT`. The reading for the `T` cue returns $d3$ and $d4$. In

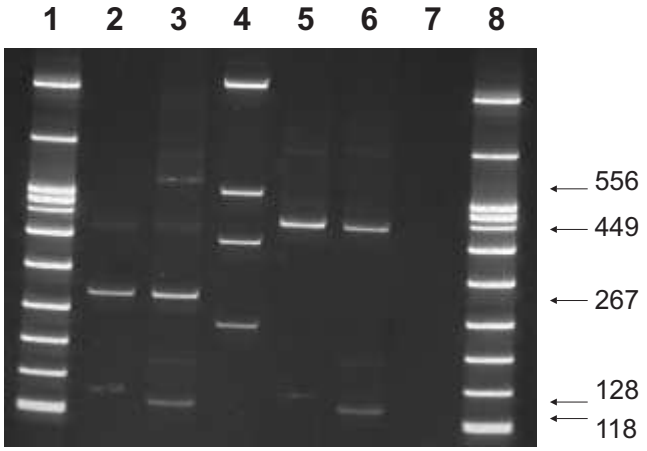

Figure 6. The gel electrophoresis diagram. Line 1 and 8: DNA markers 1444, 736, 501, 489, 476, 404, 331, 242, 190, 147, 111, 110, 67; line 2: reading for cue T; line 3: reading for cue G; line 4: pUC19 + *Hinf*I; line 5: reading for cue C; line 6: reading for cue A; line 7: test of supression.

line 3 the reading reaction for the input signal cue `G` was provided. The three bands represent molecules `1GG`, `3GT`, `5AG`. The memory cell `1GG` has the same address on both ends. It is seen as a thinner strip in the diagram (amplification is done only when the ends are different: in the case when on one end the input signal is hybridized and on the other it is a temporary molecule). The memory cells are $d1$, $d3$ and $d5$ respectively. In line 5 there are two bands: 128 and 449 bp, respectively representing molecules `2AC` ($d2$) and `4CT` ($d4$). These memory cells are returned in reading the `C` cue. In line 6 the results of the reading reaction with cue `A` as an input signal are shown. There are also molecules `2AC` - band 118 and `5AG` - band 449. They represent $d2$ and $d5$ respectively.

Input signal	Read cells	Band's length
`cueT`	`3GT`, `4CT`	267, 128
`cueG`	`1GG`, `3GT`, `5AG`	556, 267, 118
`cueC`	`2AC`, `4CT`	449, 128
`cueA`	`2AC`, `5AG`	449, 118

Table 4. Results of reading process for different input signals.

The above retrieved DNA molecules confirm that by using the presented method the process of reading data from associative memory is performed in the correct way.

5.3. Reaction details

In our experiments manipulations with DNA fragments were performed generally in the same manner as described in Ref. 6–8.

The memory cells were prepared by digestion of 2 μg of `pUC19` DNA with *Hinf*I restriction nuclease. An enzymatic reaction was performed for 2 hours in 20 μl reaction volume at 37 °C in the conditions recommended by the supplier of the enzyme (Amersham). The reaction mixture was supplemented with 180 μl TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), 20 μl 3 M sodium acetate, treated with phenol and a chloroform-isoamyl alcohol mixture (24:1), ethanol precipitated, dried in a vacuum and dissolved in 100 μl of TE.

With the aim to recover specific memory cell(s) from the mixture the adequate input signal, synthetic oligonucleotide, one from: `cueA`, `cueC`, `cueG` and `cueT`, and temporary molecule (`TEMP`) were mixed together with 1 μl of *Hinf*I digest, DNA ligase (1 μl, Amersham) was added and the reaction was performed at 16 °C for 2 hours.

After that, the PCR was performed in 50μl reaction mixture containing 5 μl 10 × PCR buffer, 0.1 μl ligation solution, 25 pM each `PRIMER1` and `PRIMER2` primers, 100 μM each dNTPs, 2.5 u Taq polymerase, and water to

50 μl. Thermal cycling was performed as follows: at 70 °C for 2 min, then Taq DNA polymerase was added, next 22 cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 60 s, finally 25 °C for 60 s. 10 μl of the PCR products were electrophoresed on 6% polyacrylamide gels (acrylamide: bisacrylamide = 59:1) with TAE buffer, stained in ethidium bromide at 0.5 $\mu g/L$ aqueous solution for 10 min. Image of the gel was made using a White/Ultraviolet Transiluminator. See Ref. 6 for details.

6. CONCLUSIONS

In this paper the implementation of associative memory based on DNA molecules was presented. The experiments demonstrated that the concept described in this paper works properly. Such memory could have a large capacity and constant access time.

It should be mentioned, that by storing information on the molecular level many advantages can be achieved. Among them are: very high scale of integration - the possibility of storing a lot of information in a small space, durability, reliability and the ability to associate.

Discussed memories can be extended easily and they can be realized with a great number of cells. It should be noted that in this case a single memory cell is very small and it does not require using comparators as in traditional associative memories. By using a molecular approach, the memory can store 1 bit of information in 1 cubic nanometre (this is unattainable for present electronic devices, they can store 1 bit in a cubic micrometer in direct addressed memories). However, it may be difficult to build an interface between electronic computers and big associative DNA memories. The memory build in a molecular manner does not loose information after some period of time (like an electronic dynamic memory). In this case the refreshing process is not necessary when information is not read from the memory. Thus, information can be stored in the memory for a longer period of time.

The method described in Ref. 1 has some disadvantages. Coding binary symbols in sequences of DNA decreases the density of stored information. The necessity of applying magnetic beads causes difficulties in cleavage of DNA strands. To use the strands as a cue for the next reading the molecules must be purified. In this approach special devices are required, which causes the method to be more expensive. It should be mentioned that the approach based on magnetic beads is less selective than the method using suppressing DNA fragment amplification during PCR.

More intensive research is required for better understanding of the properties of associative DNA memories. These devices may have a number of potential applications. They may be used in artificial intelligence to realize associative memories and expert systems. The other area of application is a molecular computer, which has many advantages over traditional computers.

REFERENCES

1. E. B. Baum, "Building an associative memory vastly larger than the brain," *Science* **268**, pp. 583–585, Apr. 28, 1995.
2. G. Păun, G. Rozenberg, and A. Salomaa, *DNA Computing: New Computing Paradigms*, Springer Verlag, Berlin, Heidelberg, New York, Sept. 1998.
3. M. Amos, *Theoretical and experimental DNA computation*, vol. -, pp. 614–630. World Scientific, 2001.
4. L. Kari, R. Kitto, and G. Gloor., "A computer scientist's guide to molecular biology," *Soft Computing* **5**(2), pp. 95–101, 2001.
5. C. Chenchik, A. Diatchenko, P. Siebert, S. Lukainov, K. Lukanov, N. Gurskaya, V. Tarabykin, and E. Sverdlov, "Method for suppressing dna fragment amplification during pcr," 1998.
6. A. Masny and A. Płuciennicza k, "Finerprinting of bacerial genomes by amplification of dna fragments surrounding rare restriction sites," *BioTechniques* **31**, 2001.
7. J. Sambrock, E. Fritsch, and T. Manitis, *Molecular cloning, a laboraory manual*, Cold Spring Harbor Laboratory Press, 1989.
8. C. Yanish-Person, J. Vieira, and J. Messing, "Improved m13 phage cloning vectors and host strains: nucleotide sequences of m13mp18 and puc19 vectors," *Gene* **33**, pp. 103–119, 1985.