ELSEVIER

BioSystems 52 (1999) 25–33

BioSystems

www.elsevier.com/locate/biosystems

Progress toward demonstration of a surface based DNA computation: a one word approach to solve a model satisfiability problem

Qinghua Liu [a], Anthony G. Frutos [a], Liman Wang [a], Andrew J. Thiel [a], Susan D. Gillmor [b], C. Todd Strother [a], Anne E. Condon [c], Robert M. Corn [a], Max G. Lagally [b], Lloyd M. Smith [a,*]

[a] *Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA*
[b] *Department of Materials Science and Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA*
[c] *Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA*

Abstract

A multi-base encoding strategy is used in a one word approach to surface-based DNA computation. In this designed DNA model system, a set of 16 oligonucleotides, each a 16mer, is used with the format 5′-FFF-FvvvvvvvvFFFF-3′ in which 4–8 bits of data are stored in eight central variable ('v') base locations, and the remaining fixed ('F') base locations are used as a word label. The detailed implementations are reported here. In order to achieve perfect discrimination between each oligonucleotide, the efficiency and specificity of hybridization discrimination of the set of 16 oligonucleotides were examined by carrying out the hybridization of each individual fluorescently tagged complement to an array of 16 addressed immobilized oligonucleotides. A series of preliminary hybridization experiments are presented and further studies about hybridization, enzymatic destruction, read out and demonstrations of a SAT problem are forthcoming.

Keywords: Surface-based; Hybridization; Multi-base encoding; Satisfiability (SAT); DNA computing

1. Introduction

Computer scientists rank computational problems in three classes: easy, hard, and uncomputable (Ouyang et al., 1997). One of the major achievements of computer science in the last two decades is the understanding that many important computational search problems are NP-complete and thus are unlikely to have efficient algorithms that solve the problem exactly (Garey and Johnson, 1979).

Algorithms that solve optimization versions of such problems exactly are suitable only for small-size instances in the worst case. 'Heuristic' algorithms such as simulated annealing (Kirkpatrick et al., 1983) often work well in prac-

* Corresponding author. Tel.: +1-608-263-2594; fax: +1-608-265-6780.
E-mail address: smith@chem.wisc.edu (L.M. Smith)

0303-2647/99/$ - see front matter
PII: S0303-2647(99)00029-5

tice, but offer no guarantee of obtaining an optimal solution. 'Approximation algorithms' guarantee a solution within some percentage of optimal, for example there are known efficient approximation algorithms for the Satisfiability (SAT) problem that output a truth assignment satisfying a number of clauses that is at least 75% of the number of clauses satisfied by the best possible truth assignment.

Recently Adleman (1994) showed that DNA can be used to solve a computationally hard problem, the Directed Hamiltonian Path problem (DHPP), and demonstrated the potential power of parallel, high-density computation by molecules in solution. This parallelism allows DNA computers to solve larger hard problems such as NP-complete problems in linearly increasing time, in contrast to the exponentially increasing time required by a Turing machine (Ouyang et al., 1997).

After Adleman initiated the field of DNA computing in 1994, Lipton (1995) proposed DNA experiments to solve the SAT problem. In 1997, Ouyang et al. (1997) presented a molecular biology-based experimental solution to the 'maximal clique' problem. This problem is in the same class (NP-complete) as the DHPP. The paper is significant because it shows a functional demonstration of improved design principles for DNA computing, and the use of living organisms (*Escherichia coli*) to read the answer of a computation by cloning.

All three works (Adleman 1994; Lipton 1995; Ouyang et al., 1997) use the tools of molecular biology (in this case, DNA as an information carrier, ligation, PCR, gel electrophoresis, magnetic bead separation, enzymatic digestion, etc. as operations), and all demonstrate the feasibility of carrying out computations at the molecular level. However, since they are all solution-based methods, they share the common problems of scale-up of this test tube-based approach for a number of reasons, including poor efficiencies in the purification and separation steps.

In contrast with the solution-based experiments of Adleman (1994), Lipton (1995), and Ouyang et al. (1997), surface-based DNA computations manipulate DNA strands that are immobilized on a surface using chemical linkers (Cai et al., 1997; Frutos et al., 1997; Liu et al., 1998; Smith et al., 1998; Liu et al., 1999). This means that a key operation used in solution-based DNA computations, that of selectively separating strands into separate test tubes, cannot be performed. Also the number of DNA strands involved in the computation is limited since the strands are restricted to two rather than three dimensions (Cai et al., 1997). (The number of strands that can fit on a 1 cm^2 planar surface is roughly 10^{12}.) Nevertheless, it is our premise that surface-based chemistry will be important to advances in DNA computation. Detailed discussions about the strategy for a surface-based approach to DNA computation are reported in Smith et al. (1998).

In this paper a one-word DNA experiment using a surface-based approach is designed and in progress as a demonstration of a small-scale prototype DNA computer with several 'operations' that could perform simple calculations such as the SAT problem (Garey and Johnson, 1979).

2. Operations on surfaces

A simple version of surface-based DNA computing uses three basic operations MARK, UNMARK, and DESTROY (Smith et al., 1998). In the MARK operation a less complex combinatorial mixture of DNA corresponding to the query would be added to the surface; the complementary strands would bind to form a duplex; thus MARKED strands would be duplexed, and UNMARKED strands would be single-stranded. This operation has been demonstrated with both single-base and multiple-base encoding strategies (Frutos et al., 1997; Liu et al., 1998; Smith et al., 1998).

The DESTROY operation consists of adding an exonuclease specific for single-stranded DNA. Every unmarked strand is destroyed, leaving on the surface only the MARKED DNA molecules. This operation has been demonstrated with three cycles of selective enzymatic destruction of unmarked (single-stranded) DNA words in the presence of marked (hybridized) oligonucleotides using $3' \rightarrow 5'$ single-strand-specific *E. coli* Exonuclease I (Frutos et al., 1997; Smith et al., 1998)).

The UNMARK operation consists of subjecting the surface to conditions under which hybrids dissociate into single strands. This is performed through a combination of increased temperature and addition of denaturants such as urea. Subsequent washing with a suitable buffer removes the free strands in solution and regenerates the DNA-modified surface (Smith et al., 1998).

After each cycle of MARK, DESTROY, and UNMARK, fewer molecules remain on the surface. Repeated queries constitute the DNA computation process, permitting subsets of the initial combinatorial space to be eliminated, leaving the desired solutions to the problem of interest.

The end result of a DNA computation is a surface on which DNA molecules exist, whose sequence encodes the solution to a combinatorial problem of interest. It is thus necessary to determine the sequence(s) of these surface-bound DNA molecules in order to ascertain the solution to the computational problem in question (this operation is called READOUT). Both conventional electrophoresis-based DNA sequencing and hybridization to word-specific addressed arrays have been studied (Gillmor et al., 1998; Wang et al., 1998).

3. One word SAT problem model system

A model system for demonstration of a one word SAT problem was designed, using the multiple base encoding strategy with 16mers attached to a chemically modified gold surface.

The exhaustive search algorithm for the SAT problem is as follows (Liu et al., 1999):

```
for each clause C do
  for each unnegated variable x_i in C do
    mark (i, 1)
  for each negated variable x_i in C do
    mark (i, 0)
  comment: all remaining solutions that set C
  to 'true' are marked
  DESTROY-unmarked (using single-strand
  exonuclease)
test-if-empty
```

Here, the initial set of strands represent all 2^n possible truth assignments to the variables $x_1,\ldots, x_n$ of the input formula. Mark (i, 1) means that all strands in which variable i is set to true are marked.

The DNA model system is a set of 16, 16-base oligonucleotides, which are attached to chemically modified gold surfaces as described previously (Frutos et al., 1997). The 16mers have the following design sequence 5′-FFFFvvvvvvvvFFFF-3′ in which the fixed sequences are 5′-HS-$(T)_{15}$-GCTTvvvvvvvvTTCG-3′. The variable sequences chosen are from the set of 108 8-mers previously identified (Frutos et al., 1997). Each member of this set differs from every other member of the set in at least 4 base locations. Table 1 shows the variable sequence regions (from 5′ to 3′) for the 16 oligonucleotides in the model system. Each panel of Table 1 contains the same sequences as the other panels, but each panel is different in that it shows the number of matches and mismatches and their distributions when each oligonucleotide in the set is hybridized with a particular complement. Blue letters represent perfect matches to the specific complements being tested, red represents mismatches, and black represents a partial match. For example, in panel 1 in the upper left corner, the complement considered will be 3′-CGAAgttgggttAAGC-5′, which is perfectly matched with the first sequence in blue, and is two base matched (two bases in black) with the second DNA molecule (AACCTGGT) and eight base mismatched (all in red) with the third DNA strand (ACCAAACC), and so on. When the complement 3′-CGAAttggaccaAAGC-5′ is considered, the perfectly matched oligonucleotide will be the second one in panel 2 at the upper left corner (blue in second sequence), and the DNA strand with the first sequence in that box will be two base matched (two are in black) with the complement, and so on.

The set of 16 oligonucleotides is used to encode four bits (four binary variables, $2^4 = 16$) of information. One of the possible encoding schemes is shown below. In this encoding scheme, the variable DNA sequences encode 4 binary variables x_1, x_2, x_3 and x_4.

We have proposed to solve the following specific SAT problem using the set of 16 oligonucleotides:

Table 1
Variable sequence regions for the 16 oligonucleotides

CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA
AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT
ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC
AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC
ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG
CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG
GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC
GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT
TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC
TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG
TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG
TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA
ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA
CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC
ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA
ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT
CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA	CAACCCAA
AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT	AACCTGGT
ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC	ACCAAACC
AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC	AGAGTCTC
ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG	ATATCGCG
CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG	CCAAGTTG
GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC	GGTTCAAC
GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT	GTTGGGTT
TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC	TATAGCGC
TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG	TCTCAGAG
TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG	TGGTTTGG
TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA	TTGGACCA
ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA	ACTGGTCA
CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC	CAGTTGAC
ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA	ATGCAGGA
ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT	ATCGAGCT

$(x_1 \vee x_2' \vee x_4) \wedge (x_3) \wedge (x_2') \wedge (x_4 \vee x_1')$

This particular problem was chosen arbitrarily; many other problems would be equally good as model SAT demonstration problems (example in Smith et al. (1998)). Note that the problem involves all four variables, and has four clauses. The problem can be solved easily by checking each of the 16 strands. However, successfully solving this problem would be the first demonstration of a SAT calculation by DNA computation using a surface-based approach. Solving the problem by DNA computation requires four cycles of MARK, DESTROY, and UNMARK (since there are four clauses), followed by READOUT.

DNA (probe No.)	Variable sequences	$x_1x_2x_3x_4$
S_0 (4)	CAACCCAA	0000
S_1 (85)	AACCTGGT	0001
S_2 (2)	ACCAAACC	0010
S_3 (29)	AGAGTCTC	0011
S_4 (24)	ATATCGCG	0100
S_5 (87)	CCAAGTTG	0101
S_6 (81)	GGTTCAAC	0110
S_7 (94)	GTTGGGTT	0111
S_8 (18)	TATAGCGC	1000
S_9 (11)	TCTCAGAG	1001
S_{10} (92)	TGGTTTGG	1010
S_{11} (79)	TTGGACCA	1011
S_{12} (108)	ACTGGTCA	1100
S_{13} (107)	CAGTTGAC	1101
S_{14} (73)	ATGCAGGA	1110
S_{15} (19)	ATCGAGCT	1111

In preparation for the DNA computation, combinatorial mixture of strands S_0–S_{15} would first be immobilized on a gold surface. In the first computational cycle, all strands which do not satisfy the first clause are destroyed, namely those two strands in which both x_1 and x_4 are set to 'false' and x_2 is set to 'true' (S_4 [0100] and S_6 [0110]). Specifically, this would be accomplished in the following manner: first, hybridize the sur face-bound $S_0 \ldots S_{15}$ to a mixture of complements C_0, C_1, C_2, C_3, C_5, C_7, C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15} (complements of S_0 to S_{15} are denoted as C_0 to C_{15}), at room temperature in 2 × SSPE/0.2% SDS buffer for 30 min., leaving S_4 and S_6 unhybridized. The surface is then subjected to *E. coli* Exonuclease I digestion for 3 h to selectively destroy surface-bound unmarked strands S_4 and S_6. Finally, the surface is re-generated by the UNMARK operation using 8.3 M urea (the 14 different hybridized duplexes would be denatured in this step to return the surface to single-stranded form without S_4 and S_6).

In the second computational cycle, the remaining seven strands (S_0 [0000], S_1 [0001], S_5 [0101], S_8 [1000], S_9 [1001], S_{12} [1100], and S_{13} [1101]) which do not satisfy the second clause are destroyed, namely those left on the surface in which x_3 is set to 'false'. Again, this would be accomplished as follows: first, hybridization of the 14 strands left on the surface from the first cycle to a mixture of the complements C_2, C_3, C_7, C_{10}, C_{11}, C_{14}, and C_{15}, followed by *E. coli* Exonuclease I digestion to selectively destroy surface-bound unmarked strands S_0, S_1, S_5, S_8, S_9, S_{12}, and S_{13}. Finally, the surface is re-generated by the UNMARK operation to return the surface-bound oligonucleotides S_2, S_3, S_7, S_{10}, S_{11}, S_{14} and S_{15} to their single-stranded form.

In the third computational cycle, the remaining three strands (S_7 [0111], S_{14} [1110], and S_{15} [1111]) which do not satisfy the third clause are destroyed, namely those left on the surface in which x_2 is set to 'true'. The implementation is quite similar to the two cycles described above, but with hybridization to the mixture of complements C_2, C_3, C_{10}, and C_{11}. The Exonuclease I digestion destroys unmarked strands S_7, S_{14} and S_{15}, and the final UNMARK operation leaves S_2, S_3, S_{10} and S_{11} single-stranded on the surface.

In the last cycle, the one strand (S_{10} [1010]) which does not satisfy the fourth clause is destroyed, namely the strand left on the surface in which x_4 is set to 'false' and x_1 is set to 'true'. This time, the remaining surface-bound strands are hybridized to complement C_2, C_3 and C_{11}; S_{10} which is unmarked is destroyed by Exonuclease I, followed by re-generation of the surface, leaving the three satisfying assignments, S_2 [0010], S_3 [0011], and S_{11} [1011] on the surface.

The final operation is READOUT. As mentioned previously, two approaches have been investigated: cycle sequencing (Blakesley, 1993), and PCR amplification (Mullis and Faloona, 1987; Erlich et al., 1991)) followed by hybridization to addressed arrays.

Before going into detail about the 'READOUT' operation, several details of the computational cycles should be addressed. In the four successive cycles of MARK, DESTROY, and UNMARK operations, both fluorescein-labeled and unlabeled complements are employed. Sometimes hybridization, destruction, and/or denaturation needs to be monitored, therefore fluorescently tagged complements need to be utilized. When no monitoring is needed, unlabeled complements are used. In addition, since cycle sequencing or PCR amplification is needed, appropriate flanking PCR-primer sequences should be designed into the set of complements for the final READOUT. If the material is cycle-sequenced, a suitable primer sequence, which can be the same as one of those employed for PCR amplification, should be included in the set of complements to permit eventual enzymatic sequence analysis.

If an addressed array of 16 spots is employed for READOUT, two primer sequences need to be incorporated into the set of complements for PCR amplification before the READOUT hybridization. Therefore, for a complement design suitable for both cycle sequencing and PCR amplification, two 20mer sequences should be added to the 16mer complements as shown below. The lower case sequences are the two primers, the upper case sequences are fixed word labels, and v is the complement region shown in Table 1:

5′-tatttttgagcagtggctccCGAAvvvvvvvvAAGCtagctatctacaagattcag-3′

For cycle sequencing READOUT, the primer is labeled either with one dye (FAM) or with four dyes (Fam-C, Joe-A, Tamra-G, and Rox-T). In the first case, each reaction (dideoxy C, dideoxy A, dideoxy G and dideoxy T) must be performed in a separate container and the products loaded in four separate lanes on the electrophoresis gel; in the latter case, the reactions are performed in four separate containers, but the products are pooled and analyzed in a single lane on the electrophoresis gel. 6% denaturing polyacrylamide gels can be used on an ABI DNA sequencer. The sequence of the primer is complementary to the 3′ end of the sequence shown above:

5′-dye-ctgaatcttgtagatagcta-3′

For PCR amplification and addressed array hybridization READOUT, the amplified materials should be fluorescently tagged and be complementary to one or more of the surface-bound DNA molecules. Therefore, the PCR primers must be designed accordingly. One primer should be biotinylated to permit strand separation after PCR amplification, as explained later. The other primer should be fluorescently tagged as shown below:

5′-biotin-ctgaatcttgtagatagcta-3′

5′-FAM-tatttttgagcagtggctcc-3′

After PCR amplification, the products are double-stranded 56mers with one strand biotinylated, and the other strand fluorescein labeled. Prior to addressed array readout, double-stranded PCR products are strand-separated using Dynabeads M-280 streptavidin (Dynal Prod.). The supernatant (containing FAM labeled strands) is desalted and concentrated on a Microcon 10 concentrator (Amicon, Beverly, MA). The concentrated fluorescently tagged single-stranded PCR products are to be applied to a 16-spot addressed array for READOUT.

4. Progress toward demonstration of one word SAT problem

All oligonucleotides designed above were synthesized by the University of Wisconsin Biotechnology Center on an ABI 308B or 391 DNA synthesizer. Glen Research 5′-thiol-modifier C6 and ABI 6-FAM were used for 5′-thiol modified and 5′-fluorescein-modified oligonucleotides, respectively. Prior to purification, thiol-modified oligonucleotides were deprotected as specified by Glen Research Corporation (1990). Before use, each oligonucleotide was purified by reverse-phase

A

92 TGGTTTGG	19 ATCGAGCT	18 TATAGCGC	4 CAACCCAA
87 CCAAGTTG	73 ATGCAGGA	11 TCTCAGAG	29 AGAGTCTC
85 AACCTGGT	79 TTGGACCA	108 ACTGGTCA	24 ATATCGCG
81 GGTTCAAC	94 GTTGGGTT	107 CAGTTGAC	2 ACCAAACC

B

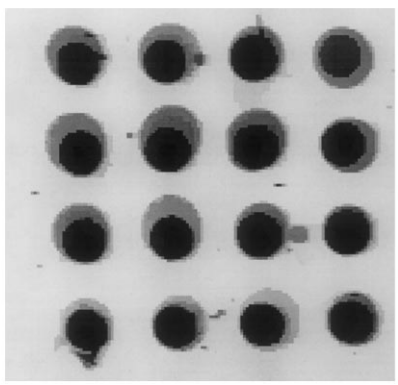

Fig. 1. Results obtained when a combinatorial mixture of 16 fluorescently tagged DNA targets is hybridized to an array of their 16 complements immobilized on a gold surface. Panel A shows the eight nucleotide variable regions and their probe numbers. Panel B shows the hybridization pattern obtained after washing at room temperature.

high performance liquid chromatography. The concentrations of the oligonucleotides were determined by UV spectrophotometry at 260 nm (Smith et al., 1987). All thiol oligonucleotides should be used immediately after purification. Gold substrates were prepared with self-assembled monolayers of 11-mercaptoundecanoic acid (MUA) and poly-L-lysine (PL) using the method described elsewhere (Liu et al., 1998). With the MUA and PL monolayers attached to the surface, the sample was irradiated with a 200 W mercury arc lamp for 3 h with a patterned mask (Jordan and Corn, 1997). The sample was then immersed in 1 mM *n*-octadecylmercaptan to create hydrophobic and hydrophilic regions on the surface (Gillmor et al., 1998). The hydrophobic surface barrier prevents any cross-contamination between the many different DNA strands spotted onto the hydrophilic portions of the slide. Detailed surface attachment reactions, hybridizations, and denaturations were carried out as described previously (Frutos et al., 1997).

The efficiency and specificity of hybridization discrimination of the set of 16 oligonucleotides were examined. Fig. 1 shows the results obtained when a combinatorial mixture of 16 fluorescently tagged DNA complements is hybridized to an array of the 16 surface-bound DNA molecules. The sequence of the eight nucleotide (nt) variable regions, along with their probe numbers within the set of 108 oligonucleotides, are shown in panel A, and the pattern of fluorescence binding at room temperature in panel B. Under the conditions employed, all members of the array show significant fluorescence binding, demonstrating the accessibility of the surface-bound oligonucleotides to hybridization.

An exhaustive study was conducted in which each individual fluorescently tagged complement (a total of 16) was hybridized to this array of 16 immobilized oligonucleotides to verify that only one spot is hybridized per complement, and at what stringency conditions perfect discrimination can be achieved.

Fig. 2 shows the results obtained when the surface-immobilized arrays are sequentially hybridized to each single fluorescently tagged oligonucleotide complement (CF). Sixteen successive hybridizations to CFs were performed. For example, '4CF' is the fluorescently tagged complement of probe 4 (3′-CGAAgttgggttAAGC-FAM 5′). For each hybridization step, all 16 perfectly matched pairs were observed. Nine of the 16 perfect matches (85-85CF, 24-24CF, 87-87CF, 81-81CF, 11-11CF, 79-79CF, 108-108CF, 73-73CF, and 19-19CF) in the set were perfectly discriminated after washing the surface in a 37°C buffer solution for 20 min. Also shown in the Figure is the fluorescence image obtained at room temperature (22°C) prior to washing at 37°C; at room temperature 85-85CF, 87-87CF and 73-73CF are the three perfectly matched pairs observed in which no mismatch hybridization was detected. In other words, perfect discrimination can be achieved at room temperature for the three pairs,

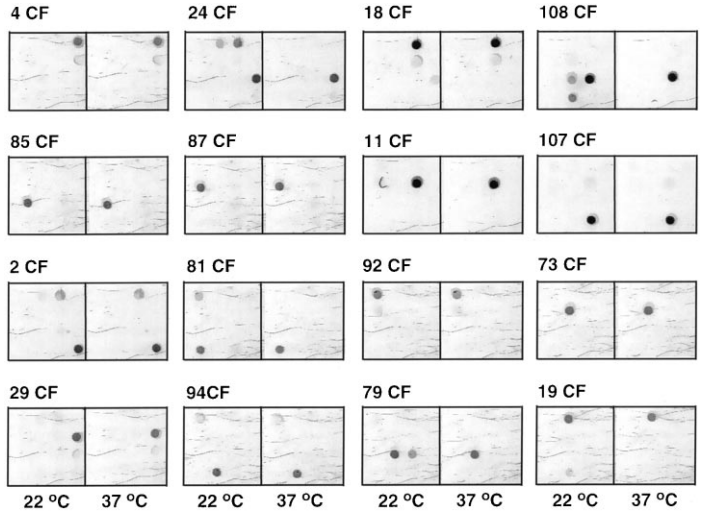

Fig. 2. Results obtained when an array of 16 oligonucleotides immobilized on a gold surface is successively hybridized to each single fluorescently tagged oligonucleotide complement (CF), perfectly complementary to each surface-bound probe. Two hybridization patterns obtained after washing at room temperature and 37°C are shown for each hybridization test.

and it can be obtained at 37°C for the other six pairs. Complete hybridization discrimination between 4-4CF, 2-2CF, 29-29CF, 94-94CF, 18-18CF, 92-92CF, and 107-107CF could not be achieved at 37°C. The sequences between the mismatched pair were compared (e.g. for mismatched pair 29-4CF, variable sequence region of the complement of 4: 3′-gt**t**ggg**t**t-5′ and that of immobilized oligonucleotide 29: 5′-ag**a**gt**c**tc-3′ are compared and only partial matches are found which are shown in the bolded regions). In the case of 94-94CF, 94CF is totally mismatched with probe 92 immobilized on the surface, but 92-94CF is observed (second column, last block). By examining the variable sequences between 94CF (3′-c**aa**ccc**aa**-5′) and probe 92 (5′-t**gg**ttt**gg**-3′), there are two pairs of g-a's in a row which is claimed as the stable mismatches (Li et al., 1991). Several mismatched pairs have multiple Gs which are known to form stable mismatches (Aboul-ela et al., 1985; Werntges et al., 1986; Ikuta et al., 1987). However, the reasons that the mismatched pairs are observed even at 40°C washing are not clear yet and further studies such as solution melting temperature measurements are underway now to determine if oligonucleotides other than the ones in the set of 16 are needed to achieve perfect discrimination.

5. Summary

A small-scale prototype of a DNA computer, with oligonucleotides immobilized on gold surfaces, is designed, and progress is being made toward testing this DNA computer with a one-word four-variable SAT problem. The experiments presented here have tested the hybridization discrimination and efficiency in the MARK operation. Other work is underway to test Exonuclease I destruction, PCR amplification and cycle sequencing conditions for readout. The long term benefit of this work will be a real-world demonstration of a surface-based DNA SAT calculation, as well as providing a foundation of ready-to-use operations for more complex DNA computations in the future.

References

Aboul-ela, F., Koh, D., Tinoco, I. Jr., Martin, F.H., 1985. Nucleic Acids Res. 13, 4811–4824.

Adleman, L.M., 1994. Science 266, 1021–1024.
Blakesley, R.W., 1993. Methods Mol. Biol. 23, 209–217.
Cai, W., Condon, A.E., Corn, R.M. et al., 1997. Proc. 1st Annu. Int. Conf. on Computational Molecular Biology (RECOMB97). ACM, pp. 67–74.
Erlich, H.A., Gelfand, D., Sininsky, J.J., 1991. Science 252, 1643–1651.
Frutos, A.G., Liu, Q., Thiel, A.J., et al., 1997. Nucleic Acids Res. 25, 4748–4757.
Garey, M.R., Johnson, D.S., 1979. Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman Press, New York.
Gillmor, S.D., Liu, Q., Wang, L. et al., 1998. Genetic Programming 1998, poster.
Glen Research Corporation, 1990, User Guide to DNA Modification and Labeling.
Ikuta, S., Takagi, K., Wallace, R.B., Itajura, K., 1987. Nucleic Acids Res. 15, 797–811.
Jordan, C.E., Corn, R.M., 1997. Anal. Chem. 69, 1449–1456.
Kirkpatrick, S., Gelatt, C.D., Vecchi, M., 1983. Science 220, 671–680.
Li, Y., Zon, G., Wilson, W.D., 1991. Biochemistry 30, 7566–7572.
Lipton, R.J., 1995. Science 268, 542–545.
Liu, Q., Frutos, A.G., Thiel, A.J., Corn, R.M., Smith, L.M., 1998. J. Comput. Biol. 5 (2), 269–278.
Liu, Q., Guo, Z., Condon, A.E., Corn, R.M., Lagally, M.G., Smith, L.M., 1999. Dimacs Ser. Discret. Math. Theor. Comput. Sci. 44, 123–132.
Mullis, K.B., Faloona, F.A., 1987. Methods Enzymol. 155, 335–350.
Ouyang, Q., Kaplan, P.D., Liu, S., Libchaber, A., 1997. Science 278, 446–449.
Smith, L.M., Kaiser, R.J., Sander, J.Z., Hood, L.E., 1987. Methods in Enzymology 155, 260–301.
Smith, L.M., Corn, R.M., Condon, A.E., et al., 1998. J. Comput. Biol. 5 (2), 255–267.
Wang, L., Liu, Q., Frutos, A.G., et al., 1998. Genetic Programming 1998, poster.
Werntges, H., Steger, G., Riesner, D., Fritz, H.J., 1986. Nucleic Acids Res. 14, 3773–3790.