Arrayed Primer Extension on DNA Microchips (APEX). Molecular Computation of Satisfaction (SAT) Problems

Michael Pirrung[1] pirrung@chem.duke.edu
Richard Connors[1] connors@pharmacop.com
Amy Odenbaugh[1] alo@chem.duke.edu
Michael Montague-Smith[1] mpms@chem.duke.edu
Nathan Walcott[2]
Jeff Tollett[2] jtollett@bcm.tmc.edu

[1] Department of Chemistry, Levine Science Research Center, Duke University, Durham, NC 27708, USA
[2] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

Keywords: DNA computing, primer extension, satisfaction problems

1 Introduction

A DNA computer with the capability of solving NP-complete problems in polynomial time using the Arrayed Primer EXtension (APEX) method was demonstrated experimentally. APEX is a high fidelity, surface-based method of nucleic acid analysis based on DNA polymerase extension of primer-template complexes on DNA microchips. In our algorithm, surface-bound primers and templates represent solutions and clauses of a Boolean formula. The method was applied to several Satisfiability problems, including a 3SAT.

2 Method and Results

Adelman [1] described an algorithm that uses DNA to solve a Hamiltonian path problem by exploiting the massive parallelism accessible through the combinatorial ligation of DNA to generate all possible solutions. He then applied the tools of molecular biology to sort through myriad DNA strands until a solution was found. His method is the first practical *polynomial-time* algorithm that conducts a *parallel* exhaustive search. Since this original work, many theoretical studies on DNA computing have appeared, and other experimental DNA computations, also in graph theory, have been reported [2].

Implementing DNA computing algorithms is fraught with problems. These include physical losses, errors in DNA manipulations, and mishybridization. We recognized that the Arrayed Primer EXtension (APEX) method under development in our laboratory [4] could be applied to solve these difficulties. Primers are attached at microscopic sites on a solid surface to form a DNA chip. Primer-template complexes receive a single dideoxynucleotide terminator bearing a fluorescent tag catalyzed by a DNA polymerase only when in a perfect duplex. Imaging of the chip identifies those primers that had complementary sequences in the template. Primers represent values of variables, and templates query for those variables in clauses of the Boolean equation. The array/primers are the *same* for all n-variable problems ($\equiv$hardware), and the templates are dictated by the equation to be solved ($\equiv$software). This method was used to implement an error-resistant version of an algorithm devised by Lipton that uses the massive parallelism of combinatorial DNA ligation to solve satisfaction problems (SAT) [3]. The SAT asks, given a set of Boolean clauses $C = \{c_1, c_2, \ldots, c_m\}$ on a finite set U

of variables such that $|c_i| = 2$ for $1 \leq i \leq m$, is there a truth assignment for U that satisfies all the clauses in C?

2.1 Possible values $\{0, 1\}$ of x and y are represented by discrete oligonucleotides. The value of x or y is encoded by the location of its representative oligonucleotide on the surface of the slide. An oligonucleotide encoding $x = 0$ is placed on the upper half and the $x = 1$ oligonucleotide is placed on the lower half. The slide is rotated 90° and the process is repeated with the $y = 0$ and $y = 1$ oligonucleotides.

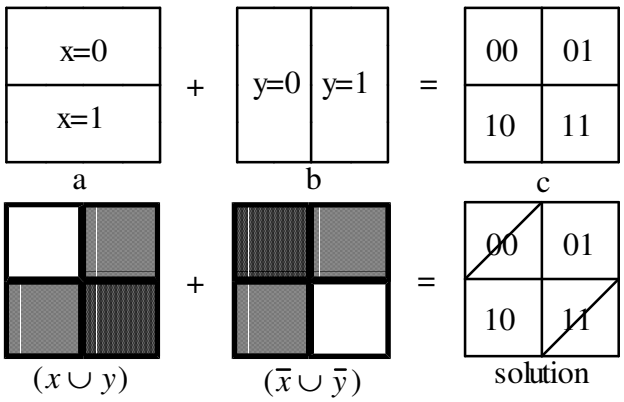

Figure 1: APEX Computing.

2.2 An APEX reaction is conducted for each clause in the SAT, using complementary DNA templates encoding 1 for normal variables (i.e., the complement of $x = 1$ for x) and 0 for negations (i.e., the complement of $x = 0$ for $\bar{x}$). The first round of this procedure produces a slide in which all regions that satisfy clause 1 $(x \cup y)$ are colored. The second round of this procedure will produce a slide in which all regions that satisfy clause 2 $(\bar{x} \vee \bar{y})$ are colored.

2.3 The dye-free regions from each round of primer extension are eliminated from the solution(s) to the SAT. This method has also been applied to a four-variable 2SAT, and a three-variable 3SAT.

3 Conclusion

With this DNA computing algorithm, as the number of variables (n) in the SAT increases, the number of oligonucleotides and manipulations required to solve the problem increases polynomially (by $4n$), while the number of possible solutions increases exponentially (by 2^n).

References

[1] Adleman, L.M., Molecular computation of solutions to combinatorial problems, *Science*, 266(5187):1021–1024, 1994.

[2] Ouyang, Q., Kaplan, P.D., Liu, S. and Libchaber, A., DNA solution of the maximal clique problem, *Science*, 278(5337):446–449, 1997.

[3] Lipton, R.J., DNA solution of hard computational problems, *Science*, 268(5210):542–545, 1995.

[4] Shumaker, J.M., Metspalu, A. and Caskey, C.T., Mutation detection by solid phase primer extension, *Hum Mutat.*, 7(4):346–354, 1996.