The Arrayed Primer Extension Method for DNA Microchip Analysis. Molecular Computation of Satisfaction Problems

**Michael C. Pirrung,*,† Richard V. Connors,† Amy L. Odenbaugh,†
Michael P. Montague-Smith,† Nathan G. Walcott,‡ and Jeff J. Tollett‡**

*Contribution from the Department of Chemistry, Levine Science Research Center, Duke University,
Durham, North Carolina 27708-0317, and the Department of Molecular and Human Genetics,
Baylor College of Medicine, Houston, Texas 77030*

Received July 9, 1999

Abstract: A high fidelity, surface-based method of nucleic acid analysis has been developed based on DNA polymerase extension of primer−template complexes on DNA microchips. The ability of the method to discriminate against mismatches and provide an almost "digital" signal recommended it for molecular computation. A DNA computer with the capability of solving nondeterministic polynomial time (NP)-complete problems (those whose time−complexity function rises exponentially with the problem size) in polynomial time using this Arrayed Primer EXtension (APEX) method was experimentally demonstrated. An algorithm involving extension of surface-bound primer−template complexes, representing solutions and clauses of a Boolean formula, is described for the solution of two-, three-, and four-variable satisfiability (SAT) problems, including a 3SAT, exploiting the theoretical concepts of Lipton. A discussion of the principles of nondeterministic computing with APEX is also provided.

Introduction

The use of microarrays of nucleic acids ("DNA chips") is revolutionizing many aspects of genetic analysis,[1−11] diagnostics,[12] drug discovery,[13] and basic structural biology.[14] Generally, the use of these arrays involves the hybridization (Figure 1) of a labeled derivative of an analyte nucleic acid ("target") to chip-bound DNA ("probes"), followed by detection of the label by an imaging or scanning process. One drawback of this type of assay is the need for labeling of the target. However, when a specific locus is to be analyzed from a high-complexity, biologically derived nucleic acid sample, amplification is required and labeling may be performed as part of the polymerase chain reaction (PCR) step. Another concern is the fidelity of hybridization. Ideally, full signal would be observed from all perfect hybrids and no signal from hybrids including even single base mismatches. A priori, this goal seems unattainable simply because of the plethora of probes on a chip and their *simultaneous* hybridization. The expected divergence of the melting temperature of the hybrids formed between a single target at its various subsites and different probes of varying sequence should mean that each probe site would have unique optimum hybridization conditions (solvent/temperature) for such discrimination. This problem has been recognized since the earliest considerations of the concept of chip-based sequencing-by-hybridization (SBH),[15] and various strategies to circumvent it by leveling the stabilities of hybrids of different GC-content have been proposed. Other factors (probe length, possible internal secondary structure in both probe and target, etc.) can also affect the strength of hybridization. Chip-based hybridization has been subject to detailed biophysical analysis, and a number of unique traits of hybridization to high-density probe arrays (defined as those in which the spacing between individual probe molecules is comparable to the molecular dimension) have been found that somewhat mitigate these concerns.[16] Clearly, though, this issue is complicated and requires internal controls, high redundancy, and statistical/computer analysis of the data to draw valid conclusions from a particular hybridization experiment.

We are developing a method for DNA chip analysis involving the DNA polymerase-based solid-phase extension of primer−template complexes with labeled terminators, called APEX

* To whom correspondence should be addressed.
† Department of Chemistry.
‡ Department of Molecular and Human Genetics.
(1) Pease, A. C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. P.; Fodor, S. P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5022−5026.
(2) Woychik, R. P.; Klebig, M. L.; Justice, M. J.; Magnuson, T. R.; Avrer, E. D. *Mutat. Res.* **1998**, *400*, 3−14.
(3) Lockhart, D. J.; Mack, D. H.; Chee, M. S.; Fan, J.-B.; Wodicka, L.; Fodor, S. P. A. *Nucleic Acids Symp. Ser. 38* (Advances in Gene Technology: Molecular Biology in the Conquest of Disease); 1998; pp 11−12.
(4) Ramsay, G. *Nat. Biotechnol.* **1998**, *16*, 40−44.
(5) Marshall, A.; Hodgson, J. *Nat. Biotechnol.* **1998**, *16*, 27−31.
(6) Nollau, P.; Wagener, C. *Clin. Chem. (Washington, D. C.)* **1997**, *43*, 1114−1128.
(7) Southern, E. M. *Trends Genet.* **1996**, *12*, 110−115.
(8) Sapolsky, R. J.; Hsie, L.; Berno, A.; Ghandour, G.; Mittmann, M.; Fan, J. B. *Genet. Anal.* **1999**, *14*, 187−192.
(9) Kozian, D. H.; Kirschbaum, B. J. *Trends Biotechnol.* **1999**, *17*, 73−78.
(10) Gerhold, D.; Rushmore, T.; Caskey, C. T. *Trends Biochem. Sci.* **1999**, *24*, 168−173.
(11) Lander, E. S. *Nat. Genet.* **1999**, *21*, 3−4. Hacia J. G. *Nat. Genet.* **1999**, *21*, 42−47.
(12) Whitcombe, D.; Newton, C. R.; Little, S. *Curr. Opin. Biotechnol.* **1998**, *9*, 602−608.
(13) Debouck, C.; Goodfellow, P. N. *Nat. Genet.* **1999**, *21*, 48−50.
(14) Sohail, M.; Akhtar, S.; Southern, E. M. *RNA* **1999**, *5*, 646−655.
(15) Maskos, U.; Southern, E. M. *Nucleic Acids Res.* **1993**, *21*, 4663.
(16) Forman, J. E.; Walton, I. D.; Stern, D.; Rava, R. P.; Trulson, M. O. *ACS Symp. Ser.* **1998**, *682*, 206−228.
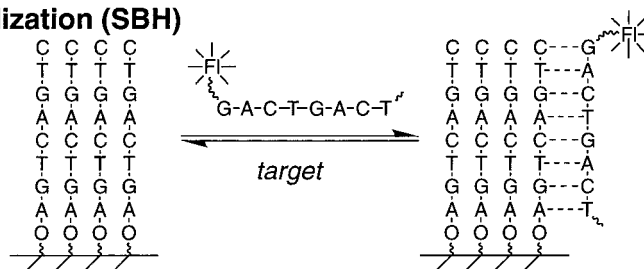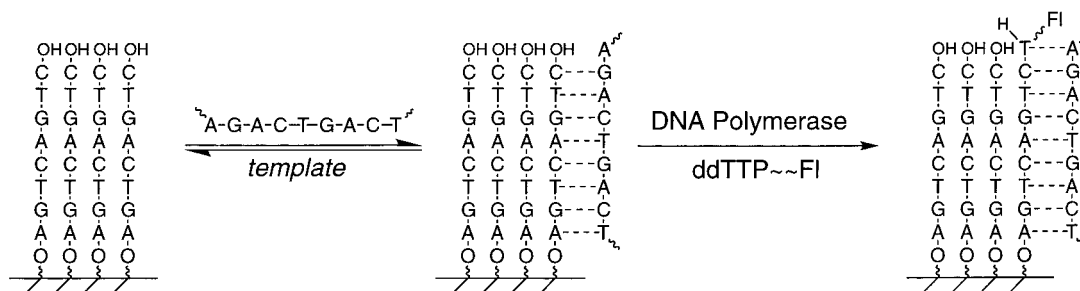
Sequencing-by-Hybridization (SBH)

Arrayed Primer Extension (APEX)

Figure 1. Hybridization on DNA chips is simply mass action driven, and the target must bear a label. APEX involves a preequilibrium followed by an irreversible enzymatic reaction, both steps of which are sensitive to mismatches.
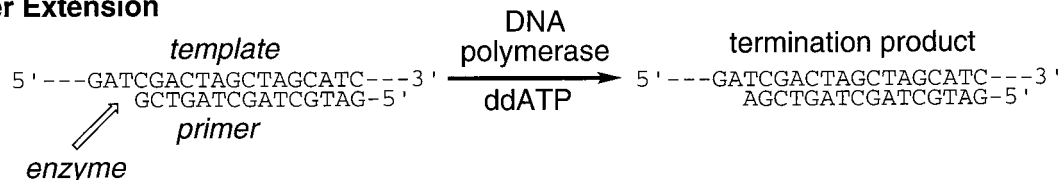
Primer Extension

Figure 2. The APEX method for detection of single nucleotide polymorphisms is related to Sanger dideoxynucleotide sequencing. Individual primers address a site of interest by hybridization to analyte DNA. Detection of surface-based hybridization is based on the DNA polymerase-catalyzed incorporation of a labeled terminator nucleotide. The identity of the added base provides an additional nucleotide of sequence information.

(Arrayed Primer Extension; Figure 1).[17−20] Primer extension methods were developed earlier for the analysis of single-nucleotide polymorphisms (SNPs). These methods are based on template-dependent extension of DNA primers bound to a solid phase with a labeled dideoxynucleotide terminator, followed by detection of the label so added. The method is APEX when applied to an array of primers immobilized on a DNA microchip. Compared with hybridization-based DNA chip analysis, APEX offers a high signal-to-noise (S/N) ratio and consequent high fidelity in sequencing. A novel method of DNA chip preparation and use in APEX has been recently described.[20]

The APEX method is distinguished from hybridization by a preequilibrium step followed by an essentially irreversible addition of a dideoxyoligonucleotide terminator bearing a label. This method offers a number of advantages over hybridization-based DNA chip analysis. Both steps in APEX are dependent on the match between primer and template. Because of the requirement of the polymerase for a perfect hybrid near the incorporation site (3′ end of the primer), only mismatches distal to the 3′ end can lead to extension and thereby contribute to background in APEX. The covalent attachment of the fluorescent tag to priming sites in APEX permits very stringent washing steps, and no labeling of analyte nucleic acid is required. Further, because the identity of the terminator incorporated at each priming site can be determined, either by conducting one chip-reaction for each labeled terminator or by using different dyes for each terminator, one additional base of sequence information (adjacent to the primer−template complex) is obtained (Figure 2). Unlike de novo sequencing with the Sanger method, APEX requires prior knowledge of the nucleic acid to be analyzed because primers must be chosen to address each analysis site. Effectively, APEX performs single-nucleotide Sanger chain-termination sequencing. Because each primer addresses only one base in the template, one primer is required for each site in the DNA to be analyzed. The current ability to prepare DNA chips of high complexity makes sequencing of genes of thousands of nucleotides accessible with APEX. Thus, a nested, overlapping set of primers permits full analysis of a given region (Figure 3). Each single nucleotide site in the template is involved in extensions with multiple primers, giving high redundancy in data collection. A single variation in the template not only affects the identity of the base added in the extension of primer **A**, it causes a mismatch in the hybridization of primers **B, C, D**, etc., preventing extension. A minor disadvantage of APEX is that

(17) Shumaker, J. M.; Metspalu, A.; Caskey, C. T. *Hum. Mutat.* **1996**, 7, 346−354.

(18) Shumaker, J. M.; Metspalu, A.; Caskey, C. T. *Laboratory Protocols for Mutation Detection*; Landegren, U., Ed.; Oxford University: Oxford, UK, 1996; pp 93−95.

(19) Tollett, J. J.; Kurg, A.; Shah, A.; Roa, B. B.; Richards, C. S.; Nye, S. H.; Pirrung, M.; Metspalu, A.; Shumaker, J. M. *Am. J. Hum. Gen.* **1997**, 61(Suppl. S), 1322. Metspalu, A.; Saulep, H.; Tonisson, N.; Soo, V.; Kurg, A.; Shumaker, J. M. *Am. J. Hum. Genet.* **1997**, 61(Suppl. S), 1301. Kurg, A.; Tollett, J. J.; Shah, A.; Roa, B. B.; Richards, C. S.; Nye, S. H.; Pirrung, M.; Metspalu, A.; Shumaker, J. M. *Am. J. Hum. Genet.* **1997**, 61(Suppl. S), 68.

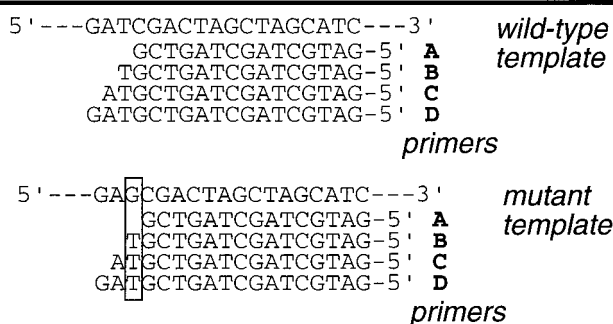(20) Pirrung, M. C.; Davis, J. D.; Odenbaugh, A. L. *Langmuir*, in press.

```
5'---GATCGACTAGCTAGCATC---3'        wild-type
        GCTGATCGATCGTAG-5'  A       template
        TGCTGATCGATCGTAG-5'  B
        ATGCTGATCGATCGTAG-5'  C
        GATGCTGATCGATCGTAG-5'  D
                                    primers

5'---GAGCGACTAGCTAGCATC---3'        mutant
        GCTGATCGATCGTAG-5'  A       template
        TGCTGATCGATCGTAG-5'  B
        ATGCTGATCGATCGTAG-5'  C
        GATGCTGATCGATCGTAG-5'  D
                                    primers
```

Figure 3. An individual site within a template is addressed by a nested set of primers in two ways. Primer **A** addresses the T based on the incorporation of the complementary ddA. Primers **B**−**D** also address this position because their incorporation sites are close to this base. In a mutant template, the **A** primer is still extended, but by a different base (C). Primers **B**−**D** cannot be extended because of the G−T mismatch (boxed).

chips cannot be reused because the primers have been terminated. Unknown mutations in a known sequence (deletions, transversions, and up to two base insertions, based on APEX reactions on both strands) can be readily found with APEX. Other types of genetic analysis (scoring SNPs, gene expression analysis) are also readily envisioned for APEX.

Other methods for genetic analysis based on solid-phase primer extension have been reported. The mini-sequencing method[21−23] involves attachment of the *template* to the support, hybridization, and enzymatic incorporation of terminator nucleotide triphosphates bearing radioisotopes. Genotyping of single genetic loci is based on the ratio of radioactivity incorporated into two allele-specific primers and readily identifies heterozygous, homozygous wild-type and homozygous mutant alleles. A recent adaptation of mini-sequencing to an array format[24] utilized as templates amplified genomic DNA fragments for nine disease mutations. Compared with hybridization with allele-specific probes in the same format, mini-sequencing showed one order of magnitude higher power of discrimination between homozygous and heterozygous genotypes. Mini-sequencing by primer extension reactions on gel-based chips has also recently been reported.[25] Genetic Bit Analysis (GBA)[26] attaches primers to macroscopic supports such as microtiter plates for SNP determination/genotyping by primer extension. This method has recently been adapted to a macro-array format[27] and applied to the detection of a rifampin-resistant mutation in the rpoB gene *of M. tuberculosis.* Primer extension can be coupled to mass spectrometry to enhance hybridization assays.[28] Finally, although it is not strictly a primer extension reaction, the fidelity of hybridization in a chip format has been improved by a similar enzymatic processing of a hybrid, in this case by ligation.[29]

Figure 4. Phosphorothioate/bromoacetamide immobilization chemistry for DNA chip fabrication.

The fabrication of DNA arrays can be accomplished by spotting of presynthesized oligonucleotides or by in situ synthesis. Because polymerases operate on the 3′ end of primers, they must be attached to the chip through either their 5′ ends or an internal site for use in APEX. Classical methods of solid-phase DNA synthesis attach the first nucleotide at it 3′ end and proceed toward the 5′ end. This method enables ready incorporation of linking functionalities (e.g., amino, disulfide) at the 5′ end, which is also a common option in commercial DNA synthesis. New methods for oligonucleotide attachment are under continuing development.[30] Oligonucleotides derivatized in this way can be attached to chips functionalized with appropriate complementary functional groups to present their 3′ ends for primer extension or other enzymatic processing. Our laboratory has recently developed an efficient and rapid method for oligonucleotide immobilization based on bromoacetylsilane-derivatized glass slides reacting with oligonucleotide 5′-phosphorothioates (Figure 4), a method used in this work.[20] In situ synthesis of primers useful for APEX requires synthesis from the 5′ to the 3′ end, which is opposite to the conventional method. This synthesis has been accomplished with nucleoside 3′-dimethoxytrityl/5′-phosphoramidites,[31] which are now commercially available, though this method has yet to be applied in a spatially directed way to chip preparation. Our laboratory has also developed a photochemical 5′→3′ synthesis method useful for this purpose.[32]

This research focuses on some simple characteristics of APEX reactions and their specific application to a nonbiological problem, the use of DNA for molecular computation.[33] The opportunity to use manipulations of DNA molecules as a massively parallel computer has been exposed by a number of theoretical studies[34,35] and two experimental demonstrations. Adleman[36] solved a Hamiltonian circuit problem and Ouyang

(21) Syvänen, A. C.; Ikonen, E.; Manninen, T.; Bengtstrom, M.; Söderlund, H.; Aula, P.; Peltonen, L. *Genomics* **1992**, *12*, 590−595.

(22) Syvänen, A. C.; Aalto-Setälä, K.; Harju, L.; Kontula, K.; Söderlund, H. *Genomics* **1990**, *8,* 684−692.

(23) Syvänen, A. C.; Landegren, U. *Hum. Mutat.* **1994**, *3*, 172−179.

(24) Pastinen, T.; Kurg, A.; Metspalu, A.; Peltonen, L.; Syvänen, A. C. *Genome Res.* **1997**, *7*, 606−614.

(25) Dubiley, S.; Kirillov, E.; Mirzabekov, A. *Nucleic Acids Res.* **1999**, *27*, e19.

(26) Nikiforov, T. T.; Rendle, R. B.; Goelet, P.; Rogers, Y.-H.; Kotewicz, M. L.; Anderson, S.; Trainor, G. L.; Knapp, M. R. *Nucl. Acids Res.* **1994**, *22*, 4167−4175.

(27) Head, S. R.; Parikh, K.; Rogers, Y. H.; Bishai, W.; Goelet, P.; Boyce-Jacino, M. T. *Mol. Cell. Probes* **1999**, *13*, 81−87.

(28) Broude, N. E.; Sano, T.; Smith, C. L.; Cantor, C. R. *Proc. Natl. Acad. Sci. U. S.A.* **1994**, *91*, 3072.

(29) Gunderson, K. L.; Huang, X. C.; Morris, M. S.; Lipshutz, R. J.; Lockhart, D. J.; Chee, M. S. *Genome Res.* **1998**, *8*, 1142−1153.
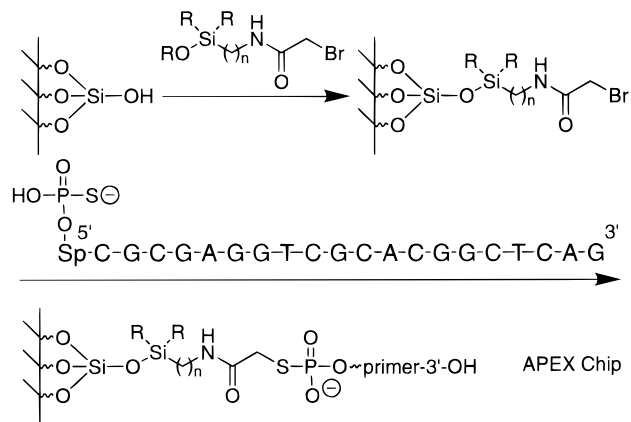
(30) Rogers, Y. H.; Jiang-Baucom, P.; Huang, Z. J.; Bogdanov, V.; Anderson, S.; Boyce-Jacino, M. T. *Anal. Biochem.* **1999**, *266*, 23−30. Beier, M.; Hoheisel, J. D. *Nucleic Acids Res.* **1999**, *27*, 1970−1977.

(31) Horne, D. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435−2437.

(32) Pirrung, M. C.; Fallon, L.; Lever, D. C.; Shuey, S. W. *J. Org. Chem.* **1996**, *61*, 2129.

(33) This work has been reported in preliminary form: Pirrung, M. C.; Connors, R. V.; Montague-Smith, M. P. *FASEB J.* **1997**, *11* (Suppl. S), A1214.

(34) Lipton, R. J. *Science* **1995**, *268*, 542.

(35) Oliver, J. S. *J. Mol. Evol.* **1997**, *45*, 161. Arita, M.; Suyama, A.; Hagiya, M. *Genetic Programming 1997*, Proceedings of the 2nd Annual Conference, Morgan Kaufmann: San Francisco, 1998; p 457. Rose, J. A.; Gao, Y. A.; Garzon, M.; Murphy, R. C.; Deaton, R.; Franceschetti, D. R.; Stevens, S. E. *Genetic Programming 1997*, Proceedings of the 2nd Annual Conference, Morgan Kaufmann: San Francisco, 1998; p 479.

et al.[37] solved a maximal clique problem. These studies are of interest because they provide the possibility to address computationally intractable problems, the so-called NP-complete problems (*N*ondeterministic *P*olynomial time, those whose time−complexity function rises exponentially with the size of the problem).[38,39] The approach of these workers involves generating a set of all possible solutions and selecting those that fulfill relevant criteria using techniques common in molecular biology. Although the power of this approach to such computationally hard problems as code breaking (the US DES), chemical structure generation, or predicting the folding of large RNAs[40] is apparent, significant limitations are also apparent. Among the concerns[41] that have been raised is the error rate, which in many cases has been linked to mishybridization.[42] Efforts to minimize this problem have included an optimal DNA encoding of solutions.[43,44] Another concern is imperfection and losses in the manual manipulation of DNA strands required in techniques such as PCR, gel separation, and affinity purification. A surface-based approach to DNA computation has been offered by Smith and Corn to address some of these difficulties, as well as to advance a novel computing algorithm.[44] Because it has been shown that any of the NP-complete problems (e. g., Hamiltonian circuit, Partition, Clique, Vertex cover, three-dimensional matching, and three-satisfiability (3SAT)) can be polynomially transformed into the others (vide infra), a method that solves a 3SAT problem is in principle applicable to any NP-complete problem.

Lipton devised a theoretical algorithm that uses the massive parallelism of combinatorial DNA ligation (Adleman's approach) to solve satisfaction problems by converting the SAT into a graph problem to which Adleman's algorithm could be applied. The SAT asks, given a set of clauses $C = \{c_1, c_2,...c_m\}$ 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? Lipton examined an SAT of the form:

$$F = (x \vee y) \wedge (\bar{x} \vee \bar{y}) \qquad (1)$$

In this equation there are two clauses, $(x \vee y)$ and $(\bar{x} \vee \bar{y})$. The variables x and y are Boolean and can assume values of 0 (false) and 1 (true); $\vee$ is the logical OR operation, which dictates that $(x \vee y) = 0$ only if $x = y = 0$; $\wedge$ is the logical AND operation, which dictates that $(x \vee y) = 1$ only if $x = y = 1$; $\bar{x}$
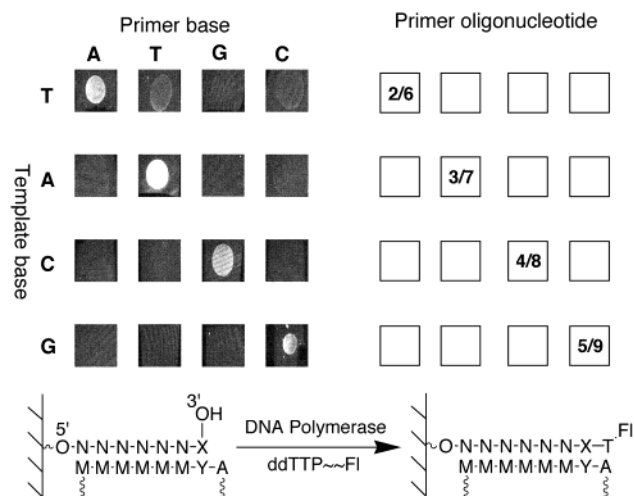

Figure 5. Experiment addressing the fidelity of primer extension. Primers differing only at the X position and templates differing only at the Y position were prepared, and the primers were arrayed in columns. The templates were applied to those primers in rows. Thus, at each priming site, a different combination of X and Y (of the 16 possible) is formed. Only when complementary base pairs are formed is extension signal observed. The sequences of the primers are given Table 1.

is the negation of x, which dictates that $\bar{x} = 0$ if $x = 1$, and $\bar{x} = 1$ if $x = 0$. The SAT problem is to assign values to x and y such that $F = 1$ (F is true).[45] Lipton's proposed algorithm employs combinatorial DNA ligation to simultaneously create all possible solutions to this SAT, encoded in 15 nucleotide DNA sequences to maximize hybridization fidelity. SAT problems are identified by the number of literals (variables) per clause. Thus, the SAT of eq 1 is a 2SAT because it has two literals per clause.

Solutions of this SAT and two others, including a 3SAT, have been implemented experimentally using, instead of random ligation, the much more error-resistant APEX process, in which a single base addition faithfully discriminates among DNA strands in a primer−template set.

Results

An example of the utility of APEX in discriminating single nucleotide polymorphisms is shown in Figure 5. Four oligonucleotides differing in only the nucleotide at their (free) 3′ ends were arrayed. Were this array treated with a target complementary to one of these probes, no difference in hybridization at the four probe sites would be expected. However, when this array is treated with polymerase and a fluoresceinated terminator, specific labeling of only the primer with perfect complementarity to the template is observed. Earlier work has shown that catalysis by the polymerases used in APEX is very sensitive to mismatches at the 3′ end of the primer. This sensitivity is likely related to specific recognition of the last five base pairs of the primer−template complex by the polymerase, as recently demonstrated by crystallography.[46]

The almost digital signals obtained in this experiment and many like it strongly suggested the applicability of APEX to DNA molecular computation. We therefore developed the following DNA chip-based design that permits the solution *in*

(36) Adleman, L. M. *Science* **1994**, *266*, 1021.

(37) Ouyang, Q.; Kaplan, P. D.; Liu, S.; Libchaber, A. *Science* **1997**, *278*, 446.

(38) Garey, M. R.; Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-completeness*; Freeman: San Francisco, 1979.

(39) A discussion of the principles of NP-complete in this context is available from our web site: *http://www.chem.duke.edu/~pirrung*.

(40) Gaspin, C.; Westhof, E. *Advances in Molecular Bioinformatics*; Schulze-Kremer, S., Ed.; IOS: Amsterdam, 1994; p 103. Nuzillard, J.-M.; Naanaa, W.; Pimont, S. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1068−1073.

(41) Gibbons, A.; Amos, M.; Hodgson, D. *Curr. Opin. Biotechnol.* **1997**, *8*, 103. Hartmanis, J. *Bull. Eur. Assn. Theor. Comput. Sci.* **1995**, *55*, 136.

(42) Deaton, R.; Garzon, M.; Murphy, R. C.; Franceschetti, D. R.; Rose, J. A.; Stevens, S. E. *Genetic Programming 1997*, Proceedings of the 2nd Annual Conference; Morgan Kaufmann: San Francisco, 1998; p 463.

(43) Garzon, M.; Neathery, P.; Deaton, R.; Murphy, R. C.; Franceschetti, D. R.; Stevens, S. E. *Genetic Programming 1997*, Proceedings of the 2nd Annual Conference; Morgan Kaufmann: San Francisco, 1998; p 472. Baum, E. *DNA based computers*; Lipton, R. J.; Baum, E. B., Eds.; American Mathematical Society: Providence, Rhode Island, DIMACS *27* 1997, p 122.

(44) Frutos, A. G.; Liu, Q.; Thiel, A. J.; Sanner, A. M. W.; Condon, A. E.; Smith, L. M.; Corn, R. M. *Nucleic Acids Res.* **1997**, *25*, 4748−4757. Smith, L. M.; Corn, R. M.; Condon, A. E.; Lagally, M. G.; Frutos, A. G.; Liu, Q.; Thiel, A. J. *J. Comput. Biol.* **1998**, *5*, 255−267. Liu, Q.; Frutos, A. G.; Thiel, A. J.; Corn, R. M.; Smith, L. M. *J. Comput. Biol.* **1998**, *5*, 269. Frutos, A. G.; Smith, L. M.; Corn, R. M. *J. Am. Chem. Soc.* **1998**, *120*, 10 277.

(45) Such problems are ubiquitous. Versions of SAT best known in the lay literature include the "pork-chop problem" and " barber-shop paradox" devised by Lewis Carroll: *Symbolic logic: part I, Elementary* 1896, 5th Ed.; C. N. Potter: dist. Crown Publishers: New York, 1977.
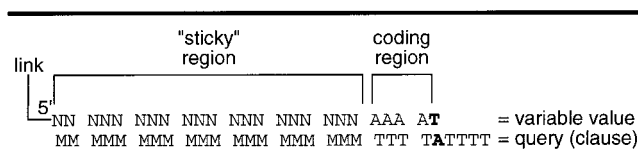
APEX Computation of Satisfaction Problems

J. Am. Chem. Soc., Vol. 122, No. 9, 2000 **1877**

Figure 6. Encoding scheme for primers and templates used in APEX computation. A complementary 18−24-nt constant region in each provides a perfect duplex in the primer−template complex. A 5-nt coding region is located at the 3′ end of the primer, where polymerases show high discrimination against mismatches. Complements to this coding region are at the 5′ end of the template. Ts follow this region, leading to the addition of ddATP to the primer strand. Each query or clause requires one to several of these templates.

polynomial-time of two-, three-, and four-variable SAT. Surface-bound primer−template complexes represent solutions and clauses of a Boolean formula. Extension by DNA polymerase identifies possible solutions based on the perfect match in the coding region of each primer. The variable encoding used to implement Lipton's algorithm for APEX solution of satisfaction problems is shown schematically in Figure 6. All primers, representing values of variables, and templates, representing queries (clauses), have two distinct regions, a common "sticky" region designed to maximize the affinity of primer for template and minimize secondary structure, and a variable region. Primers also bear either 5′-aminolink or 5′-phosphorothioate groups for attachment to glass, in the latter case using our novel protocol. The 5 nt variable region appended to the constant sequence partially encodes the value of its variable. The variable region only partially encodes the variable value because the balance of this information is encoded in the location of the primer. With a variable region of this length, 4^5 or 1024 combinations can be encoded, corresponding to 10 Boolean variables.

The logic of solving an equation such as eq 1 with APEX is as follows. Primers encoding the values of each variable (≡literal) are combinatorially tiled onto the surface (Figure 7) so that all possible assignments for the variables are represented.[47] These assignments can be addressed by in situ synthesis using a binary masking scheme[48] or other methods. Each zone contains multiple oligonucleotides (reflecting the parallelism of the method), with each oligonucleotide representing the value of a different variable. The array/primers are the *same* for all problems of *n* or fewer variables (equivalent to hardware). The templates used are dictated by the equation to be solved, however (equivalent to software). One APEX is conducted for each clause in the equation, with templates included according to the presence of their encoded variables (or complements) in the clause. The number of templates in each APEX therefore equals the number of literals per clause. The APEX signals indicate zones containing primers complementary to one or more of the templates, and potentially encoding the solution to the SAT. A zone encoding a valid solution for the SAT must be extended in *each* APEX, however, reflecting the logical AND of the SAT. The logical OR of the SAT is reflected in the inclusion of multiple templates in each APEX.

(46) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391*, 304−307.

(47) For *n* Boolean variables, there are 2^n variable assignments, which can be prepared through combinatorial methods in 2*n* processes. Such a surface can be used in the solution of an equation involving *n* variables (or fewer).

(48) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Solas, D.; Liu, A. *Science* **1991**, *251*, 767.
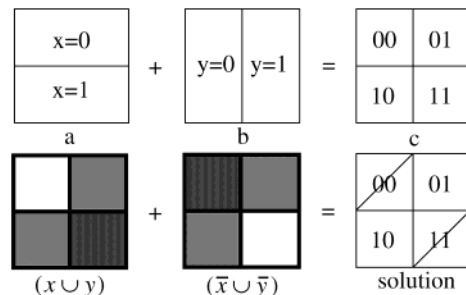

Figure 7. SAT Protocol: (1) Prepare a spatially addressable array of DNA primers (c) encoding all possible solutions to the SAT. Each possible value {0,1} of *x* and *y* is represented by a discrete oligonucleotide. The value of the variable (*x* or *y*) is encoded by the location of its representative oligonucleotide on the surface of a microscope slide. For example, the oligonucleotide encoding $x = 0$ is placed on the upper half of the slide, whereas the $x = 1$ oligonucleotide is placed on the lower half. The slide is then rotated 90° and the process is repeated with the $y = 0$ and $y = 1$ oligonucleotides.[55] (2) Conduct an APEX reaction with a dye-labeled nucleotide terminator and DNA complements 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 $\bar{x}$) for each bracketed clause in the SAT. The first round of this procedure will produce 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 ($\bar{x} \cup \bar{y}$) are colored. (3) Eliminate from consideration dye-free regions containing DNA sequences that do not satisfy the SAT. (4) Decode remaining regions that provide the solution to the SAT.
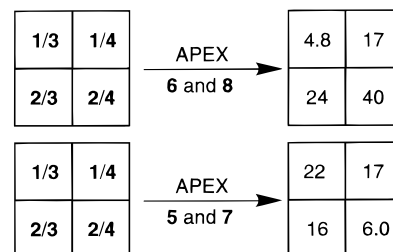

Figure 8. Two-bit SAT. An array was prepared from primers **1**−**4** in accord with the protocol given in Figure 7. APEX reactions were conducted with primers **6** and **8**, leading to the S/N ratios shown in each priming site. A duplicate array was used in APEX reactions with primers **5** and **7**, giving the S/N ratios shown.

The experimental study of this two-variable SAT used primers **1**−**4**, which were immobilized on glass slides coated with *N*-(3-dimethylethoxysilyl)propyl bromoacetamide by our thin film method. The following four-element grid was created in which two oligonucleotides are spotted in each quadrant, as a 1:1 mixture of 1 mM oligonucleotides, creating a 'two-bit' experiment where each oligonucleotide can represent a bit. These

code	oligonuleotides
00	**1/3**
01	**1/4**
10	**2/3**
11	**2/4**

spotting solutions were applied to slides as shown in Figure 8. Two templates, oligonucleotides **6** and **8** or **5** and **7**, [for the other two possible clauses, (*x* or negation *y*) and (negation *x* or *y*), **6** and **7** or **5** and **8** would be used] are used in each APEX reaction mixture. For the first clause of the equation, templates
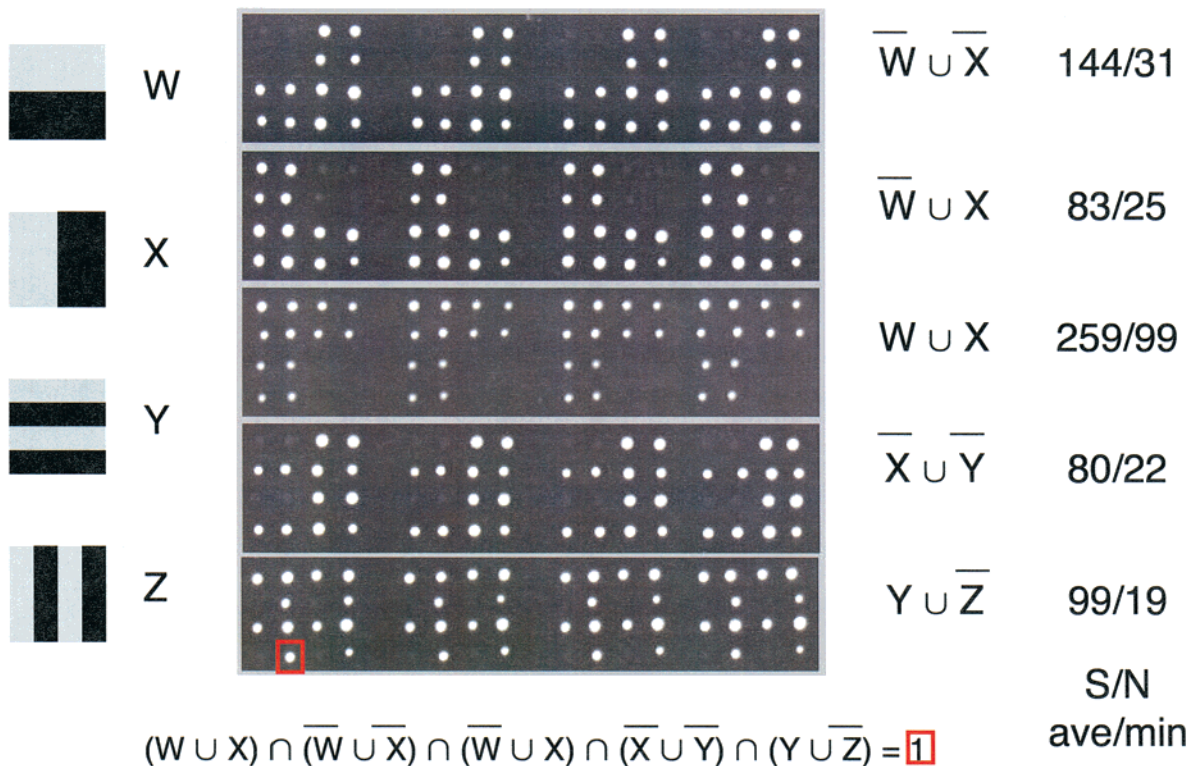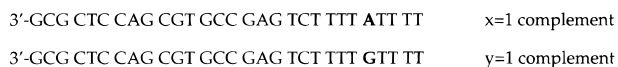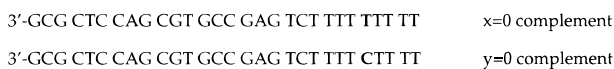
Figure 9. Four-bit SAT. An array of DNA primers was prepared in quadruplicate from eight oligonucleotides, **1**−**4** and **9**−**12**. The key on the left shows the distribution of primers for each variable (grey = 1, black = 0). Four primers were pooled for each array element, yielding 16 sites with unique combinations of the TRUE and FALSE state primers for four variables. Approximately 10 nL of 100 μM oligonucleotides solution was deposited per spot, with spot diameters of 300 μm spaced on 1-mm centers. After one APEX for each clause of the SAT, using two templates (as shown to the right of the images) selected from among **5**−**8** and **13**−**16**, measurements of S/N ratio were made at each spot in the array. The numbers given represent the ratios of the signal at TRUE primer extension sites to FALSE sites. The average S/N ratio (#:1) observed on all four arrays and the minimum S/N ratio on one array are given. The unique solution, boxed in red, is ($w = 0$, $x = 1$, $y = 0$, $z = 0$).

sequences were as follows:

3′-GCG CTC CAG CGT GCC GAG TCT TTT **A**TT TT $x=1$ complement

3′-GCG CTC CAG CGT GCC GAG TCT TTT **G**TT TT $y=1$ complement

For the second clause of the equation, templates sequences were as follows:

3′-GCG CTC CAG CGT GCC GAG TCT TTT **T**TT TT $x=0$ complement

3′-GCG CTC CAG CGT GCC GAG TCT TTT **C**TT TT $y=0$ complement

One of the quadrants should contain no primer complementary to either template, two of the quadrants should contain one primer complementary to template, and one quadrant contains two primers complementary to template. Thus, after the APEX reaction, one quadrant should not be fluorescent, two quadrants should contain spots of intermediate fluorescence, and one quadrant should contain a spot of highest fluorescence. The average S/N[49] from replicate ($n = 15$) two-bit experiments with oligonucleotides **6** and **8** in the APEX reaction are shown. A two-bit experiment with oligonucleotides **5** and **7** gives the average S/N shown ($n = 18$). The results are all comparable and excellent, fully consistent with Figure 7, and provide solutions to the SAT of (x,y) = (1,0) and (0,1).

A more challenging four-variable SAT (eq 2) was then addressed using 16 synthetic oligonucleotides.

$$(w \cup x) \cap (\bar{w} \cup \bar{x}) \cap (\bar{w} \cup x) \cap (\bar{x} \cup \bar{y}) \cap (y \cup \bar{z}) = 1 \quad (2)$$

Quadruplicate arrays of DNA primers were prepared from eight oligonucleotides **1**−**4** and **9**−**12**, and eight templates were pooled pairwise for use as queries. Only one priming site is labeled in all five APEX reactions, which reflect by their template composition the variables in each of the clauses of the SAT. The surviving priming site encodes the solution to this SAT [(w,x,y,z) = (0,1,0,0)] (Figure 9).
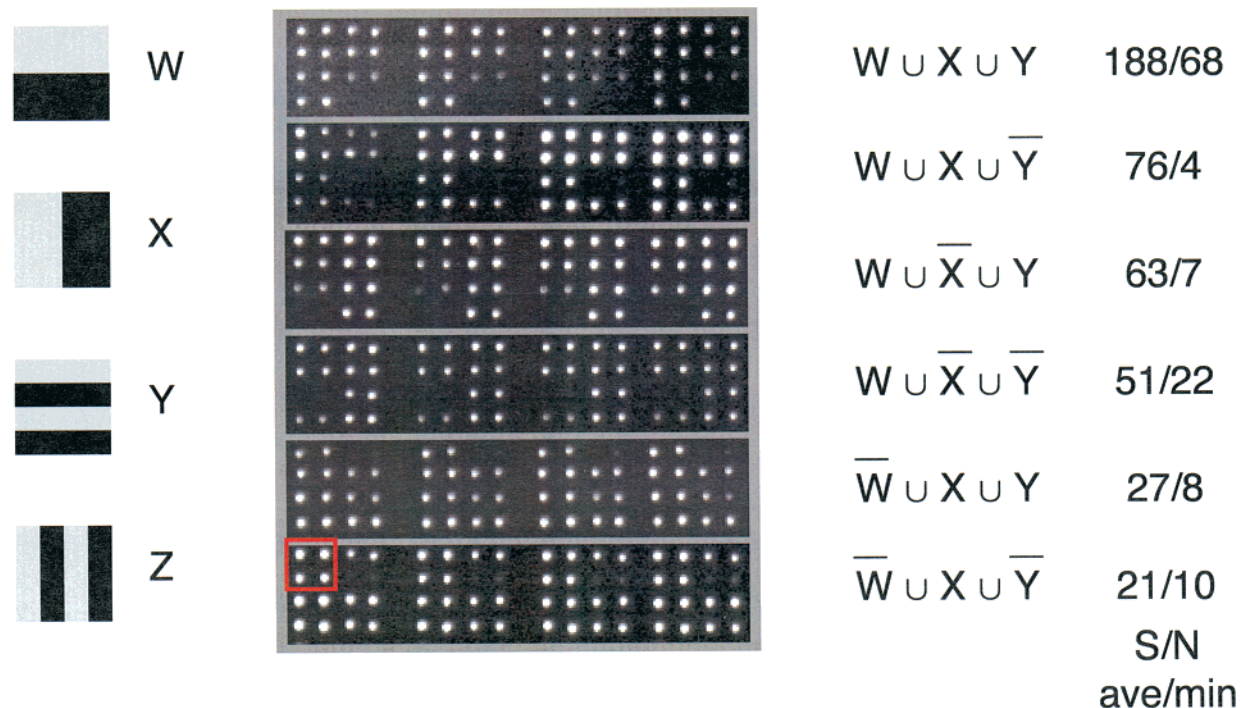
Although this was a more powerful demonstration of the ability of APEX to faithfully perform DNA computations, this four-bit SAT problem is not actually a member of the most challenging class of problems, the NP-complete. A 3SAT (eq 3), which is distinguished from eq 2 because it has three literals per clause and *is* a member of NP-complete, was then computed.

$$(w \cup x \cup y) \cap (w \cup x \cup \bar{y}) \cap (w \cup \bar{x} \cup y) \cap (w \cup \bar{x} \cup \bar{y}) \cap$$
$$(\bar{w} \cup x \cup y) \cap (\bar{w} \cup x \cup \bar{y}) = 1 \quad (3)$$

Because this 3SAT is a function of only w, x, and y, it is a three-variable (or three-bit) problem. Therefore, it requires templates for only those three variables (six total). A four-bit DNA array similar to that from the four-variable SAT just mentioned was used with these templates. This computation provided two solutions, (w,x,y) = (1,1,0) and (1,1,1) (Figure 10). As eq 3 is not a function of z, either value for z is a valid solution to the equation, and signals for both variable assignments for z are observed.

Discussion

Polynomial-Time versus Exponential-Time. Complexity theory concerns itself with different types of computing

(49) This value was obtained by measuring by confocal fluorescence microscopy the intensity of fluorescence in the spot as well as in a dark area a short distance away from the spot and taking their ratio.

$$(W \cup X \cup Y) \cap (W \cup X \cup \overline{Y}) \cap (W \cup \overline{X} \cup Y) \cap (W \cup \overline{X} \cup \overline{Y}) \cap (\overline{W} \cup X \cup Y) \cap (\overline{W} \cup X \cup \overline{Y}) = 1$$

Figure 10. Three-bit 3SAT. Following a protocol similar to that in Figure 9, six slides were incubated using three templates specific for each clause, as shown to the right of the images. Each clause eliminates two spots from the set of 16, so that the commonly labeled elements represent the solution boxed in red ($w = 1$, $x = 1$, $y = 0 \cup 1$, $z = 0 \cup 1$). In this 3SAT, there are no constraints on z, so both values satisfy the expression. The S/N data are tabulated to the right of each image, as in Figure 9. The lower minimum S/N numbers on several of the conditions are likely due to carry-over during the array printing process. The minimum value of 68 for the first clause is more representative of usual APEX data because, in this case, the FALSE (unlabeled) spots are the first to be printed.

problems and their tractability. The degree to which the difficulty of a problem increases as its size increases is the central criterion by which complexity theorists judge whether the problem is intractable. Usually a *problem* is described by a general description of its parameters or variables and the properties its solution is required to satisfy. An *instance* of a problem is obtained by specifying values for the variables. The time requirements of an algorithm designed to solve the problem are expressed in terms of a single variable n, representing the *size*, which reflects the amount of input data needed to describe the instance. The *time complexity function* for an algorithm, which is a function of the difficulty of the problem, expresses the maximum amount of time required to solve a problem instance of a given size. Algorithms/problems/computations have been organized into two basic classes, *polynomial-time* and *exponential-time*, based on the relationship between size and time complexity function. With polynomial-time computations, the time complexity function is a polynomial function of problem size. Scientists have historically preferred to work with "easy," polynomial-time problems. Many scientific laws are expressed as a relationship between two or more measured quantities, reaction kinetics, for instance. Much more troublesome are the exponential-time computations, for which the time complexity function is an exponential function of problem size. This type of problem is commonly encountered in molecular or quantum mechanical calculations where the computation time rises astronomically as the number of atoms and bonds increases. Similarly, the challenge of predicting the secondary and tertiary structures of proteins is a problem that becomes insurmountable as the number of peptide residues increases. The distinction between problems whose size increases with n, n^2, n^3, etc.

(polynomial-time) and those whose size increases with 2^n, 3^n, etc. (exponential-time) is readily appreciated. The polynomial-time/exponential-time dichotomy is a fundamental organizing principle in complexity theory because of the chasm that opens between polynomial-time and exponential-time complexity functions as problem instance size (n) increases. From the perspective of an algorithm designer or computer programmer, polynomial-time problems are easy to solve, meaning their solutions are arrived at relatively quickly. Conversely, although exponential-time problems may be manageable with a few variables, they quickly become intractable as the number of variables increases, and the computing time required to find a solution becomes prohibitive. Most exponential-time algorithms are variants of exhaustive searches, whereas polynomial-time algorithms are devised based on a deep understanding of the structure of a problem. There is consensus among computer scientists that a problem is not considered well-solved until a polynomial-time algorithm is designed for it and is considered *intractable* if no polynomial-time algorithm can possibly solve it. Combinatorial problems are an example of exponential-time scientific problems.

NP-Complete. A class of problems that has attracted considerable interest among complexity theorists is NP-complete; these problems are notoriously difficult to solve. Most cryptographic systems are based on NP-complete problems. Until the recent advent of molecular (DNA) computing, no efficient algorithm was available that could solve NP-complete problems. In a landmark paper, Cook proved[50] that the SAT is

(50) Cook, S. *The Complexity of Theorem Proving Procedures*, Proceedings of the 3rd Annual ACM Symposium on the Theory of Computing; Association of Computing Machinery: New York, 1971.

Table 1. Oligonucleotide Sequences Synthesized in this Study, Listed from 5′ to 3′[a]

oligonucleotide number	sequence	variable assignment
1	JSp-d(CG CGA GGT CGC ACG GCT CAG AAA A**A**)	x = 0
2	JSp-d(CG CGA GGT CGC ACG GCT CAG AAA A**T**)	x = 1
3	JSp-d(CG CGA GCT CGC ACG GCT CAG AAA A**G**)	y = 0
4	JSp-d(CG CGA GCT CGC ACG GCT CAG AAA A**C**)	y = 1
5	d(TTT T**TT** TTT CTG AGC CGT GCG ACC TCG CG)	x = 0 comp
6	d(TTT T**AT** TTT CTG AGC CGT GCG ACC TCG CG)	x = 1 comp
7	d(TTT T**CT** TTT CTG AGC CGT GCG ACC TCG CG)	y = 0 comp
8	d(TTT T**GT** TTT CTG AGC CGT GCG ACC TCG CG)	y = 1 comp
9	JSp-d(CG CGA GGT CGC ACG GCT CAG AAA **TA**)	
10	JSp-d(CG CGA GGT CGC ACG GCT CAG AAA **TT**)	
11	JSp-d(CG CGA GGT CGC ACG GCT CAG AAA **TG**)	
12	JSp-d(CG CGA GGT CGC ACG GCT CAG AAA **TC**)	
13	d(TTT T**TA** TTT CTG AGC CGT GCG ACC TCG CG)	
14	d(TTT T**AA** TTT CTG AGC CGT GCG ACC TCG CG)	
15	d(TTT T**CA** TTT CTG AGC CGT GCG ACC TCG CG)	
16	d(TTT T**GA** TTT CTG AGC CGT GCG ACC TCG CG)	

[a] J denotes phosphorothioate [S—PO$_2$—O—], Sp denotes Spacer-18, [—(CH$_2$CH$_2$O)$_6$—PO$_3$—]. Encoding nucleotides are shown in bold.

NP-complete (Cook's theorem), and therefore, *as hard or harder than any other problem in NP*. Since then, hundreds of problems have been classified as NP-complete by showing that they can be "polynomially transformed"[51] into SAT. After Cook's discovery, Karp compiled[52] a list of 21 NP-complete problems that he classified by polynomial transformation to SAT. Of these 21 problems, six have emerged as the core set of NP-complete problems with which most others are proven NP-complete; the six basic *NP-complete* problems are:

(1) 3-Satisfiability (3SAT)

(2) 3-Dimensional Matching (3DM)

(3) Vertex cover (VC)

(4) Clique

(5) Hamiltonian circuit (HC)

(6) Partition

These six problems form an equivalence class, that is, they have been proven equivalent. Furthermore, a polynomial-time algorithm that solves HC can solve 3SAT, 3DM, VC, Clique, Partition or any other NP-complete problem as well. Considering these equivalence properties of the class NP-complete, the importance of earlier accomplishments in experimental DNA computation becomes clear; that is, these algorithms or variants could (in principle) be used to solve any NP-complete problem. However, an approach to actually accomplishing such computations has not been transparent, and experimental progress has been slow despite many theoretical proposals. For example, Lipton modified Adelman's approach, devising an algorithm that uses the massive parallelism of combinatorial DNA ligation to solve SATs by converting the SAT into a graph problem to which his algorithm could be applied. This process is analogous to the mechanism by which new problems are proven to be NP-complete; that is, polynomial transformation of the new problem, about which little is known, into the old one, which is well-characterized. Although it was not supported by experimental data, Lipton's algorithm is noteworthy because it solves SAT, the hardest problem in NP. Moreover, the SAT is the only problem in NP that has been rigorously proven to be NP-complete.

The algorithm described here to perform a 3SAT has reduced to practice the third instance of DNA computation of an NP-complete problem, the first two being the HC and Clique. It offers the advantage that manipulation of DNA on a surface can be more efficient than in solution, and is amenable to automation in the oligonucleotide synthesis/immobilization, enzymatic extension, and readout stages. Spatial addressability offers immediate access to the solution of the SAT, rather than requiring a further sequencing process as in the Smith—Corn method. However, two-dimensional surface-based DNA computation currently offers a limited scale of computation ($10^9 - 10^{12}$ strands, corresponding to a ~14 vertex HC graph). As the number of variables (*n*) in the SAT increases, the number of oligonucleotides and manipulations required to solve the problem increases polynomially (by 4*n*), while the number of possible solutions increases exponentially (by 2^n). This algorithm benefits from the same exponential increase in complexity exhibited by chemical combinatorial library methods as the number of elementary units increases polynomially.[53] Considering that a plot of variable number versus time complexity function for *any* exponential-time problem rises steeply without limit, one can always find instances of a DNA computation problem that require an unreasonable amount of DNA to compute. Such rejoinders miss the point of current DNA-based computing research. Although obviously none of the extant methods can solve *all* exponential-time problems, they do redefine the limits of *demonstrable* computation and lead to the development of new paradigms of computation. The polynomial-time algorithm (P) for 3SAT developed here solves an exponential-time (NP) problem and provides a new tool for DNA-based computing.

Materials and Methods

Microscopy. Confocal laser-scanning epifluorescence microscopy was performed with a BioRad MRC-1000, Zeiss Axioscope (10× objective), and a Kr/Ar laser (model 5470K-00C-2B, Ion Laser Technology, Salt Lake City, Utah) at an excitation wavelength of 488 nm. Quantitative S/N ratios were obtained in the photon counting mode under a specified laser intensity (%L), iris aperture (I), gain (G) [most images were obtained with a multiplier of 16], and scan speed (SS) by measuring the average pixel intensity of the slide background (areas that had not been exposed to any oligonucleotide) and a representative and/or large area of each spot using software supplied with the

(51) This term encompasses the concept that different problems can be proved to be mathematically equivalent, and that an algorithm that can solve one such problem can be applied to any problem to which it is equivalent. Further amplification is beyond the scope of the present paper, but is available in the Supporting Material.

(52) Karp, R. M. "Reducability among combinatorial problems" In Miller, R. E.; Thatcher, J. W. *Complexity in Computer Computations*, Plenum: New York, 1972; pp 85—103.

(53) Pirrung, M. C. *Chemtracts: Org. Chem.* **1995**, *8*, 3—8.

microscope. Whole slide scanning was performed by Total Internal Reflection Fluoresence (TIRF) using a red laser diode (10mW, 635 nm) and imaged through a band-pass filter onto a cooled charge-coupled device (CCD) camera.

DNA Synthesis. Spacer Phosphoramidite 18, Phosphorylating Agent II, 3*H*-1,2-benzodithiole-3-one-1,1-dioxide (Beaucage reagent[54]), Poly-Pak cartridges, and Poly-Pak II cartridges were from Glen Research (Sterling, Virginia). Oligodeoxyribonucleotides were synthesized using solid-phase phosphoramidite methodology on an Applied Biosystems 392 DNA/RNA automated synthesizer. Table 1 lists the synthetic oligonucleotide sequences, all of which include a 5′-Spacer Phosphoramidite 18 at the penultimate 5′-position and which are terminated with a 5′-phosphorothioate. These oligonucleotides were synthesized using normal synthesis cycles for the conventional nucleoside phosphoramidites and the Spacer Phosphoramidite 18. The phosphorothioate was generated using the protocol provided with the reagent. The oligonucleotides were purified via Poly-Pak or Poly-Pak II cartridges. After the purified oligonucleotides were dried, the yield was determined by dissolving a known percentage of the oligonucleotide in deionized water and measuring the absorbance of the sample at 260 nm.

Oligonucleotide Spotting. 5′-Phosphorothioate oligonucleotides were dissolved in deionized water at 1−0.125 mM concentration. The oligonucleotide solutions were spotted onto silanized slides in 0.4−0.2-μL aliquots from a 10-μL pipet tip. These spots were ∼2 mm in diameter. Smaller spots could be made using elongated glass micropipets (∼0.30−0.18 mm in diameter). A capillary-based printer was also used to array 5′-phosphorothioate-linked and 5′-aminohexanoate-linked primers onto bromoacetamido- and epoxysilane-derivatized microscope slides, respectively. The spotted slides were placed in 50-mL plastic tubes containing 1.0−0.2 mL of deionized water to keep the chamber humid, thereby keeping the oligonucleotide spots from drying. The slides were removed from the humid chambers after 1 h and rinsed thoroughly under a stream of deionized water. The slides were placed in a rack and allowed to air-dry in a dark, dust-free container.

APEX Reactions. PCR buffer, ddCTP, ddGTP, dTTP, and Amplitaq Polymerase were from Perkin-Elmer. Fluorescein−ddATP was from NEN. Slides were fitted with a 50-μL cover well (Aldrich-Sigma). One-hundred microliters of an an APEX reaction mix [15.0 μL of 10X PCR buffer; 0.4 μL of ddCTP (10 mM); 0.4 μL of ddGTP (10 mM); 0.4 μL of dTTP (10 mM); 1.0 μL of fluorescein−ddATP (1 mM); 2.0 μL of Amplitaq Polymerase (5 U/μL); 15.0 μL of 25 mM MgCl₂; 12.5 μL of template oligonucleotide (0.165 mM); 103.8 μL of deionized water] was placed between the cover well and the slide using a 1-mL tuberculin syringe. The slide was then subjected to 20 thermal cycles. One cycle consisted of heating the slides to 90 °C for 1 min, cooling to 37 °C for 1 min, and heating to 70 °C for 2 min. At the end of 20 cycles, the slides were held at 0 °C for 1−20 h. The cover well was removed, and the APEX reaction mix was thoroughly rinsed from the slide under a stream of deionized water. After the slide was allowed to air-dry, 24 μL of pH 10 buffer was placed on the slide, and a cover slip was applied. The S/N ratios of the fluorescence of the fluorescein-ddATP-tagged spots were determined as before, using the confocal microscope.

Two-Variable SAT. APEX reactions were conducted with one of two different template solutions. Solution A contained 36 μL of 10X PCR buffer, 5.8 μL of fluorescein−ddATP (1 mM), 4.8 μL of Amplitaq Polymerase (5 U/μL), 36 μL of MgCl₂, 30 μL of 0.165 mM 5′-TTT TAT TTT CTG AGC CGT GCG ACC TCG CG, and 30 μL of 0.165 mM 5′-TTT TGT TTT CTG AGC CGT GCG ACC TCG CG (the

complements to the $x = 1$ and $y = 1$ oligonucleotides) in 213.6 μL of deionized water. Solution B contained 36 μL of 10X PCR buffer, 5.8 μL of fluorescein−ddATP (1 mM), 4.8 μL of Amplitaq Polymerase (5 U/μL), 36 μL of MgCl₂, 60 μL of 0.165 mM 5′-TTT TTT TTT CTG AGC CGT GCG ACC TCG CG, and 30 μL of 0.165 mM 5′-TTT TCT TTT CTG AGC CGT GCG ACC TCG CG (the complements to the $x = 0$ and $y = 0$ oligonucleotides) in 213.6 μL of deionized water. The fluoresceinated spots were detected with the confocal laser-scanning microscope (4X objective), as shown in a representative image of the first clause (four spots that were exposed to APEX solution A) of this two-variable SAT experiment at the top of Figure S1. An S/N ratio for each of the four areas was obtained by comparing the pixel intensity of a representative area of the detectable spot or a representative area where a phosphorothioate solution had been applied to the pixel intensity of the area in the middle of the spotted pattern that had not had oligonucleotide attached. The S/N ratios are given adjacent to the micrograph. It is worth noting that the signal in the SE quadrant is about twice that of the signal in the NE and SW quadrants, showing that there is also a direct effect of the primer occupancy on the APEX signal. A representative micrograph and S/N ratios of the spots from a slide that had been exposed to APEX solution B is given at the bottom of Figure S1.

Four-Variable SAT. A capillary-based printer arrayed eight 5′-phosphorothioate-linked oligonucleotide primers onto either epoxysilane or bromoacetamido-derivatized microscope slides. Four primers were pooled for each array element, yielding 16 sites with unique combinations of the TRUE and FALSE state primers for four variables. The oligonucleotides were diluted into ddH₂O to a final concentration of 100 μM. Approximately 10 nL of solution was deposited per spot, with spot diameters of 300 μm spaced on 1 mm centers.

The primer (variable) sequences are:

5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **AA**)	$w = 0$
5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **AT**)	$w = 1$
5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **AG**)	$x = 0$
5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **AC**)	$x = 1$
5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **TA**)	$y = 0$
5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **TT**)	$y = 1$
5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **TG**)	$z = 0$
5′-SO₂P-Sp-d(CG CGA GGT CGC ACG GCT CAG AAA **TC**)	$z = 1$

An APEX reaction mix (50 μL) was added containing 5 μL of 10X Thermosequenase buffer, 1 μL of Cy5-ddTTP (50 μM), 3 μL of ddATP/ddCTP/ddGTP (50 μM), 2 U of Thermosequenase, 1 μL each of oligonucleotide template (100 μM), and the balance ddH₂O.

The template sequences are:

5′-d(TTT **TTT** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{w} = 0$
5′-d(TTT **TAT** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{w} = 1$
5′-d(TTT **TCT** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{x} = 0$
5′-d(TTT **TGT** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{x} = 1$
5′-d(TTT **TTA** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{y} = 0$
5′-d(TTT **TAA** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{y} = 1$
5′-d(TTT **TCA** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{z} = 0$
5′-d(TTT **TGA** TTT CTG AGC CGT GCG ACC TCG CG)-3′	$\overline{z} = 1$

The reaction mix was spread over the microarray and incubated in the open on a 48 °C heat block for 10 min. Following incubation, the slides were washed in hot (95 °C) water for 5 min and then dried for

(54) Iyer, R. P.; Egan, W.; Regan, J. B.; Beaucage, S. L. *J. Am. Chem. Soc.* **1990**, *112*, 1253−1254.

(55) Because spotting was used instead of in situ synthesis, oligonucleotide solutions of $x = 0$ and $y = 0$ could be mixed together and immobilized in the NW quadrant. Similarly, mixtures of oligos $x = 0 + y = 1$, $x = 1 + y = 0$ and $x = 1 + y = 1$ were spotted in the NE, SW, and SE quadrants, respectively.

imaging. The labeled slides were excited by Total Internal Reflection Fluoresence (TIRF) using a red laser diode (10 mW, 635 nm) and imaged through a band-pass filter onto a cooled CCD camera.

Three-Variable 3SAT. Similar to the experiment just described, 5′-amino terminal oligonucleotide primers were attached to epoxysilanized glass slides in a quadruplicate array format. The primer sequences for this experiment were chosen based on our empirical experience with them in other contexts. In this case, ∼4 nL of material was deposited as 200-μm diameter spots on 500-μm centers. The oligonucleotides were diluted into 100 mM NaOH to a final concentration of 50 μM. Following a reaction protocol similar to that already described, six slides were incubated using templates specific for each clause.

The primer (variable) sequences are:

5′-aminolink-Sp-d(AAT ACC AGG TTC ACC CTT GAA AAG **A**)-3′	W=0
5′-aminolink-Sp-d(GAG AAT CTG GTG TGA GGA CCT GCC **T**)-3′	W=1
5′-aminolink-Sp-d(CCA AGA TGG TCT TCC GGT TGC CCC **C**)-3′	X=0
5′-aminolink-Sp-d(ACT CCA TCT CAA GCA TCG ATG TCA **A**)-3′	X=1
5′-aminolink-Sp-d(TTG GTG TGT CCC CTG GCC AGC TTC **A**)-3′	Y=0
5′-aminolink-Sp-d(CAT CTC TGA CTG GCT GCC AAC ACT **C**)-3′	Y=1
5′-aminolink-Sp-d(AGT TGA AAG ATC GTC AGA AAA ATC **C**)-3′	Z=0
5′-aminolink-Sp-d(GGT TTG GTC TAA GTT GCT GAT TAC **C**)-3′	Z=1

The template sequences are:

5′-d(GAA **A**GG CAG GTC CTC ACA CCA GAT T)-3′	W=1
5′-d(CTA **T**CT TTT CAA GGG TGA ACC TGG T)-3′	W=0
5′-d(CCA **T**TG ACA TCG ATG CTT GAG ATG G)-3′	X=1
5′-d(AAA **G**GG GGC AAC CGG AAG ACC ATC T)-3′	X=0
5′-d(CAA **G**AG TGT TGG CAG CCA GTC AGA G)-3′	Y=1
5′-d(TCA **T**GA AGC TGG CCA GGG GAC ACA C)-3′	Y=0
5′-d(CCA **G**GT AAT CAG CAA CTT AGA CCA A)-3′	Z=1
5′-d(CCA **G**GA TTT TTC TGA CGA TCT TTC A)-3′	Z=0

Supporting Information Available: A micrograph and schematic of the two-bit SAT computation and a white paper presenting the principles of nondeterministic computing and providing more background concerning the APEX computing algorithm (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA992392J