VOLUME 122, NUMBER 31
AUGUST 9, 2000
© Copyright 2000 by the
American Chemical Society

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Multiple Word DNA Computing on Surfaces

Liman Wang,[†] Qinghua Liu,[†,§] Robert M. Corn,[†] Anne E. Condon,[‡,#] and Lloyd M. Smith[†,*]

Contribution from the Department of Chemistry, Department of Computer Science, University of Wisconsin−Madison, 1101 University Avenue, Madison, Wisconsin 53706

Received March 22, 2000

Abstract: The enzymatic manipulation of DNA molecules immobilized on a surface that each contain linked, multiple "DNA words" is demonstrated, with applications to DNA computing. A new DESTROY operation to selectively remove unmarked DNA strands from surfaces, consisting of polymerase extension followed by restriction enzyme cleavage, has been developed for multiple-word DNA computing. DNA polymerase is used to extend DNA primers hybridized to DNA strands that are covalently attached to a chemically modified gold thin film. The efficiency of this surface polymerase extension reaction is >90%, as determined by removal of the extended DNA molecules from the surface followed by gel electrophoretic analysis. Complete extension of the DNA strands creates a *Dp*n II restriction enzyme site in the duplex DNA; these molecules may then be cleaved from the surface by addition of *Dp*n II, with an efficiency exceeding 90%. DNA molecules may be protected from such destruction by hybridization of a peptide nucleic acid (PNA) oligomer to one of the words. The hybridized PNA blocks polymerase extension, thereby preventing formation of the restriction site and consequent strand cleavage. The utility of these operations for DNA computing is demonstrated by solving a small (2-bit) Satisfiability problem in which information was encoded in two tandem words.

I. Introduction

The field of DNA computing was initiated in 1994 by Adleman,[1] who proposed that the tools of molecular biology could be used to solve instances of difficult mathematical problems known as NP-complete problems.[2] We have adapted these ideas to combinatorial mixtures of DNA molecules attached to surfaces in an unaddressed format and have performed logical manipulations of sets of data by the hybridization and enzymatic manipulation of the attached oligonucleotides. In a recent paper,[3] we solved a 4-variable 3-Satisfiability (SAT) problem using a brute-force search algorithm applied to a mixture of 16 distinct DNA strands attached to chemically modified gold thin films. Three "primitive" operations were employed: "MARK", in which subsets of the DNA strands are tagged (marked) by the hybridization of complementary strands; "DESTROY", in which DNA strands that are not marked are removed from the surface; and "UNMARK", in which marked molecules are untagged by removing hybridized complements.[4] To scale-up this approach to solve larger problems, a larger combinatorial set of DNA molecules is required, encoding more possible solutions to the computational problem. A previous

* Corresponding author. E-mail: smith@chem.wisc.edu.
† Department of Chemistry, University of Wisconsin−Madison.
§ Present Address: 10210 Genetic Center Drive, Mail Stop #6, Gen-Probe, Inc., San Diego, CA 92121.
‡ Department of Computer Science, University of Wisconsin−Madison.
# Present Address: The Department of Computer Science, University of British Columbia, 201-2366 Main Mall, Vancouver, BC V6T1Z4.

(1) Adleman, L. M. *Science* **1994**, *266*, 1021−1024.
(2) Garey, M. R.; Johnson, D. S. *Computers and Intractability: A guide to the Theory of NP−Completeness*; W. H. Freeman and Company: New York, 1979.
(3) Liu, Q.; Wang, L.; Frutos, A. G.; Condon, A. E.; Corn, R. M.; Smith, L. M. *Nature* **2000**, *403*, 175−179.
(4) 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.

10.1021/ja0010195 CCC: $19.00 © 2000 American Chemical Society
Published on Web 07/20/2000

study[5] has described a word design strategy for DNA computing on surfaces, which utilizes 16-base oligonucleotides, or DNA "words", to encode information. The 16mer DNA words contain 8 fixed word label bases and 8 variable bases which encode the information in each word. In the multiple-word strategy,[4,6] the word label sequence is the same for every 16mer in a given word set; additional word sets are generated by varying the word label sequence. By linking different word sets together to form multiple-word DNA strands, large combinatorial mixtures can be created to encode the set of possible solutions to large computational problems. We have shown previously[6] that the enzyme T4 DNA ligase can be utilized to create and manipulate such linked multiple-word DNA strands. However, the DESTROY operation developed previously for single-word DNA computing is not suitable for use in multiple-word DNA computing. In this paper we describe the development of a new DESTROY operation suitable for use in multiple-word DNA computing, and demonstrate how to use this operation to solve a small Satisfiability problem with multiple words.

The DESTROY operation for single-word DNA computing consists of adding an exonuclease specific for single-stranded DNA.[3,5] After the MARK operation, each unmarked strand is in a single-stranded form. It is then destroyed by single-strand specific enzyme *E. coli* exonuclease I, leaving on the surface only the marked double-stranded DNA molecules. In the multiple-word strategy, the words are "linked" together when synthesized, and the resultant multiple-word DNA strands are attached to the surface; thus, a single surface-bound oligonucleotide might be comprised of several consecutive words from different word sets. In this case, the MARK operation may target a single word within the multiple-word oligonucleotide, resulting in a short double-stranded region (the word which was marked) flanked by single-stranded DNA (the words that have not been marked). Treatment of such a structure with the single-strand specific *E. coli* exonuclease I will result in destruction of single-stranded DNA from the 3′ terminus, an unwanted result if an internal word was marked.

Accordingly, an alternative strategy for the DESTROY operation for multiple-word DNA computing has been developed here, consisting of polymerase extension followed by restriction enzyme cleavage (Figure 1). Each surface immobilized DNA strand has a common primer site at the 3′ terminus and a specific restriction site near its spacer region close to the surface. In the MARK operation, a primer oligonucleotide is combined with the MARK oligonucleotides, to form a duplex region at the 3′ terminus of the immobilized strand. After the MARK operation, if there is no duplex region formed below the primer site, the polymerase will extend the primers to the downstream spacer region to form a double-stranded restriction enzyme cleavage site. A subsequent restriction digestion step will cleave these DNA duplexes from the surface. In contrast, for any DNA strands that were marked (formed a duplex) in the MARK operation, the hybridized complement will block the polymerase extension, preventing double-stranded restriction sites from being formed near the spacer region. Therefore, after the DESTROY operation, only marked DNA molecules will remain on the surface. In practice, it was found that successful development of this operation required the use of a peptide nucleic acid (PNA)[7−9] oligonucleotide analogue in the MARK operation rather than a standard unmodified oligonucleotide, to avoid their displacement by the polymerase enzyme during the strand extension reaction.

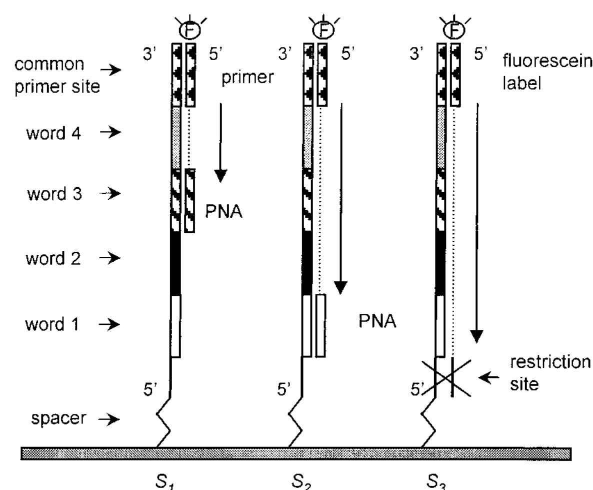

Figure 1. Overview of the DESTROY operation for multiple-word DNA computing. The figure shows three four-word DNA strands on the surface. Strands S_1 and S_2 are marked by complements at word 3 and word 1, respectively, whereas strand S_3 is not marked by any complements. Therefore, the DNA polymerase extends the primer annealed to strand S_3 to form a double-stranded restriction enzyme cleavage site near the spacer region. Strand S_3 is thus cleaved by the restriction enzyme, whereas the polymerase extensions are blocked on both strands S_1 and S_2 by marked complements. Peptide nucleic acid (PNA) oligonucleotide analogues were used in the MARK operation rather than standard unmodified oligonucleotides to avoid their displacement by the polymerase enzyme during the strand extension reaction.

These operations for multiple-word DNA computing were developed and tested upon a small example of the Satisfiability (SAT) problem.[2] The SAT problem is one of the first NP-complete search problems described. The simple example of the SAT problem we examined in the present work is $(x \vee \bar{y}) \wedge (\bar{x} \vee y)$. (Note that this small example is a 2-SAT problem (two variables or fewer per clause), which is not NP-complete; however, as we have shown previously,[3] these methods are readily applied as well to 3-SAT problems, which are NP-complete.) The variables x and y are Boolean logic variables which can hold only one of two possible values, 0 (false) and 1 (true). This example consists of two clauses separated by the logical AND operation (denoted by "$\wedge$"; $x \wedge y = 1$ if and only if $x = y = 1$); within each clause, Boolean variables are separated by the logical OR operation (denoted by "$\vee$"; $x \vee y = 0$, if and only if $x = y = 0$). The problem is to find whether there are values for the variables that simultaneously satisfy each clause in a given instance of the problem. $\bar{x}$ denotes the "negation" of x ($\bar{x} = 0$ if and only if $x = 1$, and $\bar{x} = 1$ if and only if $x = 0$). Each of the two variables can be either true or false and thus there are a total of 2^2 or 4 candidate solutions.

It may be noted that although the primary emphasis of this paper is in the area of surface-based DNA computing,[3,10−12] the chemical and biochemical procedures being developed are

(5) Frutos, A. G.; Liu Q.; Thiel A. J.; Sanner, A. M. W.; Condon, A. E.; Smith, L. M.; Corn, L. M. *Nucleic Acids Res.* **1997**, *25*, 4748−4757.
(6) Frutos, A. G.; Smith, L. M.; Corn, R. M. *J. Am. Chem. Soc.* **1998**, *120*, 10277−10282.
(7) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497−1500.
(8) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *114*, 1895−1897.
(9) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freler, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993**, *365*, 566−568.
(10) Morimoto, N.; Arita, M.; Suyama, A. *Proc. DIMACS: DNA Based Computers (III)* **1997**, 83−92.
(11) Yoshida, H.; Suyama, A. *Preliminary Proc. DIMACS: DNA Based Computers (V)* **1999**, 9−20.

likely to find substantial applicability in other areas as well. There is presently widespread interest and activity in the development and use of large arrays of proteins and nucleic acids on surfaces, permitting highly parallel analyses of RNA, protein, and small molecule binding and other activities in biological systems.[13-22] The ability to enzymatically manipulate surface-bound biomolecules substantially extends the power and versatility of such array-based methodologies.

II. Experimental Section

A. Materials. The chemicals 11-mercaptoundecanoic acid (MUA) (Aldrich), poly(L-lysine)hydrobromide (PL) (Sigma), sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (SSMCC) (Pierce), urea (United States Biochemical), and triethanolamine hydrochloride (TEA) (Sigma) were all used as received. Gold substrates were manufactured by Evaporated Metal Films (Ithaca, NY). They were prepared by evaporating 50 Å of chromium followed by 1000 Å of gold onto glass slides. Millipore filtered water was used for all aqueous solutions and rinsing. All peptide nucleic acid (PNA) oligomers were synthesized by Perseptive Biosystems (Framingham, MA). All oligonucleotides were synthesized by the University of Wisconsin Biotechnology Center (Madison, WI). Glen Research 5′-Thiol Modifier C6 and 6-FAM were used for 5′-thiol-modified and 5′ fluorescein-modified oligonucleotides, respectively. Glen Research spacer phosphoramidite 18 (S18) was utilized as a spacer. Surface-bound oligonucleotides have the structures shown in Figure 2 and Table 1. The primer oligonucleotide employed has its structure given in the legend to Figure 2. Table 1 gives the sequences of the 4 PNAs employed. Prior to purification, thiol-modified oligonucleotides were deprotected as outlined by Glen Research Corp.[23] Before use, each oligonucleotide was purified by reverse-phase binary gradient elution HPLC (Shimadzu SCL-6A). All free thiol oligonucleotides are stored under an inert nitrogen atmosphere to prevent thiol oligonucleotides from being oxidized to form disulfide dimers. DNA concentrations were verified prior to use with an HP8453 UV−vis spectrophotometer.

B. DNA Surface Attachment Chemistry. DNA oligonucleotides were immobilized onto gold thin films via a four-step chemical modification described elsewhere.[24] Briefly, a gold thin film was modified with a monolayer of the alkanethiol 11-mercaptoundecanoic acid (MUA), followed by the electrostatic adsorption of a poly-L-lysine (PL) monolayer. These steps create an amine-terminated surface that can then be reacted with the heterobifunctional linker sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (SSMCC), creating a thiol-reactive, maleimide-terminated surface. 5′-Thiol-modified DNA strands were covalently attached to this maleimide-terminated surface by placing a 0.5 μL drop of a solution containing 1 mM DNA onto the surface and reacting for at least 12 h in a humid environment to prevent evaporation. The DNA was in a pH 7, 100 mM triethanolamine (TEA) buffer when used in the surface attachment reactions. The drops of DNA spread out on the surface to a diameter of ~2−3 mm. After exposure to the DNA solution, the surface was rinsed with water and soaked for at least 1 h in 2xSSPE/0.2%SDS (pH 7.4, consists of 300 mM NaCl, 20 mM sodium phosphate, 2 mM EDTA, and 6.9 mM sodium dodecyl sulfate) at 37 °C. From previous measurements,[24] the DNA strand surface density was estimated to be 5×10^{12} molecules/cm^2.

C. Surface Hybridization. Hybridization to the attached DNA strands was accomplished by exposure of the surface to a 2 μM solution of 5′-fluorescein-labeled oligonucleotides in 2xSSPE/0.2%SDS buffer. A 30 μL drop of this solution was placed onto the gold surface and then spread over the entire surface by placing a clean coverslip on top of the sample. Hybridization adsorption was allowed to proceed in a humid environment in the dark at room temperature for 30 min with DNA oligomers, or 1 h with PNA oligomers. After hybridization, the sample was immersed in a beaker of 2xSSPE/0.2%SDS buffer at room temperature for 10 min. The sample was then placed face down on top of a glass scanner tray with a droplet of 2xSSPE/0.2%SDS buffer between the gold surface and tray and then scanned with a FluorImager 575 (Molecular Dynamics, Sunnyvale, CA). Removal of hybridized complementary molecules (referred to as "UNMARK") was accomplished by immersing the sample in 8.3 M urea at 37 °C for 15 min.

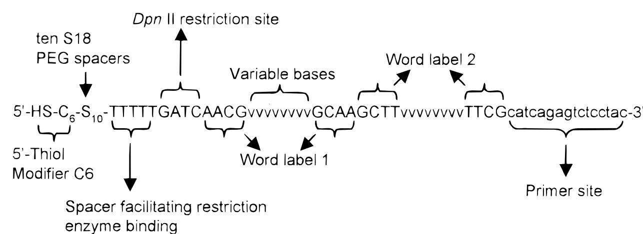

Figure 2. Sequence design of the surface-bound multiple-word oligonucleotides. 5′-HS-C6 is 5′-thiol-modifier C6 which reacts with the maleimide-functionalized gold surfaces; S_{10} corresponds to ten S18 PEG spacers, which separate the hybridizing sequence from the solid support; GATC is the recognition site for the restriction enzyme *Dpn* II; TTTTT is inserted here as a spacer to enhance the efficiency of the restriction enzyme cleavage step. This reflects the fact that additional bases flanking restriction enzyme recognition sites are often required when the cleavage occurs close to the end of the double-stranded DNA substrate being cleaved.[32] AACG...GCAA and GCTT...TTCG sequences are the "word labels" used to target hybridization to a particular word, and the variable sequence denoted by "vvvvvvvv" is used to encode information. "catcagagtctcctac" is a primer site for the polymerase extension reaction. The primer sequence employed was 5′-6-FAM-gtaggagactctgatg 3′.

Table 1. Two-Word DNA Sequence Employed and Information Encoding Scheme[a]

variable	Boolean value	word sequence (5′-3′)	PNA complements (5′-3′)
x	0	AACGcaacccaaGCAA	TTGCttgggttg
x	1	AACGgttgggttGCAA	CaacccaacCGT
y	0	GCTTtggtttggTTCG	CGAAccaaacca
y	1	GCTTaccaaaccTTCG	AAggtttggtAA

strand [*xy*]	two-word DNA sequence (5′-3′)
00	AACGcaacccaaGCAAGCTTtggtttggTTCG
01	AACGcaacccaaGCAAGCTTaccaaaccTTCG
10	AACGgttgggttGCAAGCTTtggtttggTTCG
11	AACGgttgggttGCAAGCTTaccaaaccTTCG

[a] The four strands [00], [01], [10], and [11] encode 2 bits (2^2) of information (variables *x* and *y*). Word labels are capitalized, and variable sequences are in lower case.

(12) Pirrung, M. C.; Connors, R. V.; Odenbaugh, A. L.; Montague-Smith, M. P.; Walcott, N. G.; Tollett, J. J. *J. Am. Chem. Soc.* **2000**, *122*, 1873−1882.
(13) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. *Science* **1991**, *251*, 767−773.
(14) 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.
(15) Taillon-Miller, P.; Piernot, E. E.; Kwok, P. Y. *Genome Res.* **1999**, *9*, 499−505.
(16) Marth, G. T.; Korf, I.; Yandell, M. D.; Yeh, R. T.; Gu, Z. J.; Zakeri, H.; Stitziel, N. O.; Hillier, L.; Kwok, P. Y.; Gish, W. R. *Nat. Genet.* **1999**, *23*, 452−456.
(17) Case-Green, S. C.; Mir, K. U.; Pritchard, C. E.; Southern, E. M. *Curr. Opin. Chem. Biol*. **1998**, *2*, 404−410.
(18) Hall, J. M.; LeDuc, C. A.; Watson, A. R.; Roter, A. H. *Genome Res*. **1996**, *6*, 781−790.
(19) Debouck, C.; Goodfellow, P. N. *Nat. Genet. Suppl*. **1999**, *21*, 48−50.
(20) Marshall, W. S.; Boymel, J. L. *Drug Discov. Today* **1998**, *3*, 34−42.
(21) Drmanac, R.; Drmanac, S.; Strezoska, Z.; Paunesku, T.; Labat, I.; Zeremski, M.; Snoddy, J.; Funkhouser, W. K.; Koop, B.; Hood, L.; Crkvenjakov, R. *Science* **1993**, *260*, 1649−1653.
(22) Broude, N. E.; Sano, T.; Smith, C. L.; Cantor, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3072−3076.
(23) Glen Research Corporation 1990 User Guide to DNA Modification and Labeling.
(24) Jordan, C. E.; Frutos, A. G.; Thiel, A. J.; Corn, R. M. *Anal. Chem.* **1997**, *69*, 4939−4947

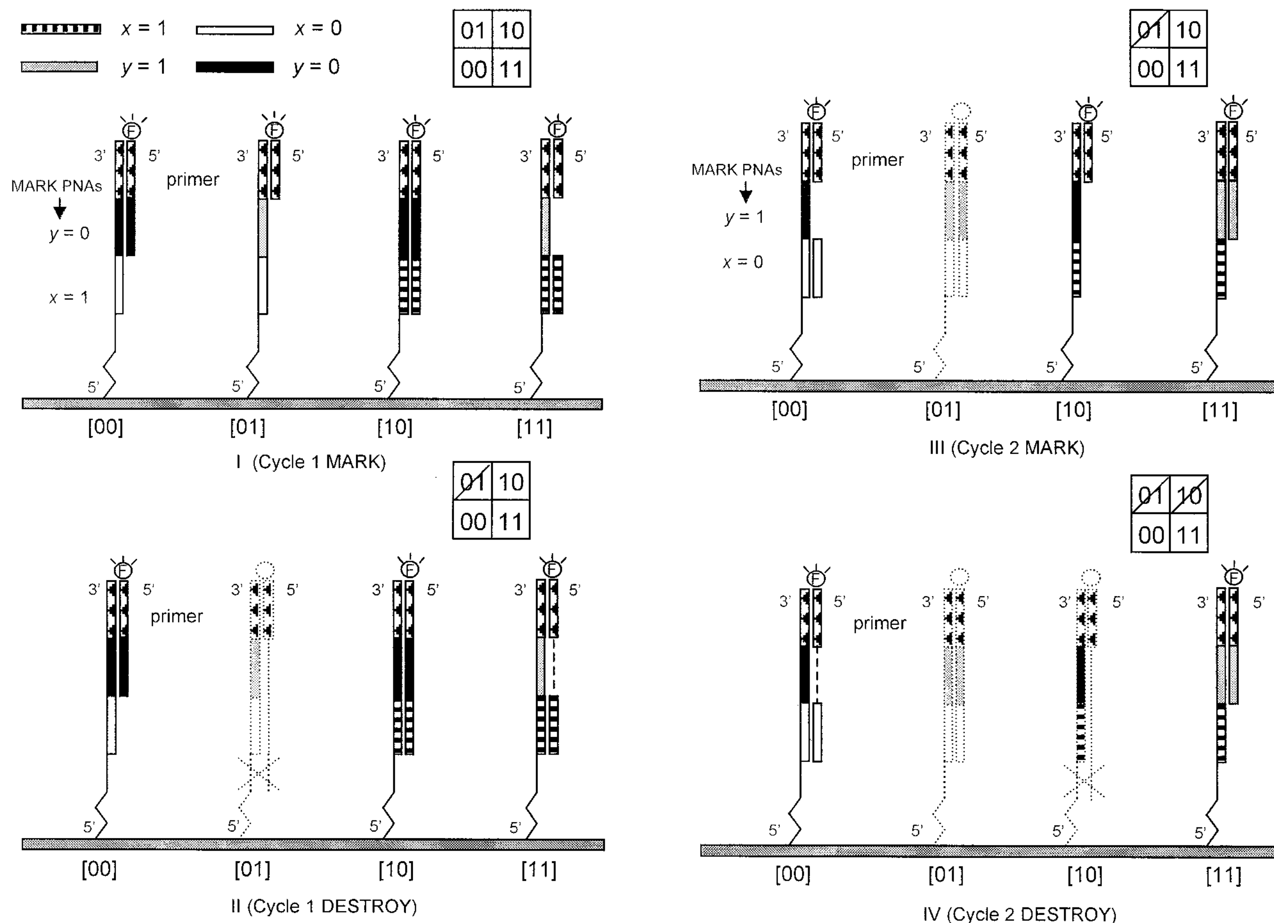

Figure 3. Illustration of the DESTROY operation in two cycles. In Cycle 1, strands [00], [10], and [11] were marked by PNA complements $x = 1$ and $y = 0$ (I). Polymerase extended the primer hybridized to strand [01] to form a double-stranded restriction site; however, strands [00], [10], and [11] were all blocked from full polymerase extension by the bound PNAs. Strand [01] was cleaved by the restriction enzyme in Cycle 1 (II). In Cycle 2, strands [00] and [11] were marked by PNA complements $x = 0$ and $y = 1$ (strand [01] should also be marked, but it was destroyed in Cycle 1) (III), polymerase extended the primer on strand [10] to form a restriction site near the spacer region. Strand [10] was then destroyed by the restriction enzyme in Cycle 2 (IV). Two strands [00] and [11] remained on the surface after two cycles of MARK and DESTROY operations. The top of panel I shows a sketch of the positions of the surface-immobilized DNA strands, their corresponding Boolean values, and the patterns employed to represent the four different oligonucleotides in the figure.

D. Surface Polymerase Extension. Polymerase extension reactions were performed by reacting the surface with 100 μL of a solution containing 0.1 U/μL Deep Vent (exo$^-$) DNA polymerase, 200 μM of each dNTP, and 100 μg/mL BSA in 1X ThermoPol Reaction Buffer (all from New England Biolabs) which consisted of 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris−HCl (pH 8.8, at 25 °C), 2 mM $MgSO_4$, and 0.1% Triton X-100. The reaction was allowed to proceed at 37 °C for 2 h.

E. Surface Restriction Enzyme Digestion. The surface was reacted with 100 μL of a solution containing 0.2 U/μL *Dpn* II restriction enzyme and 100 μg/mL BSA in 1X NEBuffer *Dpn* II (pH 6.0, at 25 °C) (all from New England Biolabs) which consisted of 100 mM NaCl, 50 mM Bis Tris−HCl, 10 mM $MgCl_2$, and 1 mM DTT. The reaction was allowed to proceed at 37 °C for 2 h.

F. Surface Polymerase Extension Efficiency. Following the surface polymerase reaction, all extended complements were desorbed from the surface by assembling the sample in a GeneAmp *In Situ* PCR System 1000 (Perkin-Elmer, Applied Biosystems, Foster City, CA) containing 100 μL of deionized water and heating at 95 °C for 10 min. The solution was collected, concentrated to a volume of ∼5 μL using a Microcon 3 Centrifugal Filter (Millipore, Bedford, MA), and loaded and electrophoresed on a 20% denaturing polyacrylamide gel containing 7 M urea. Primers were tagged with fluorescein to permit fluorescence detection and quantification. After electrophoresis the gel was imaged using a Molecular Dynamics FluorImager 575, and the relative intensities of the various bands were determined using the ImageQuaNT 4.1 software (Molecular Dynamics, Sunnyvale, CA). Polymerase extension efficiency was defined as the ratio of the fluorescence intensity of the extension product to the sum of the fluorescence intensities of the extension product and the remaining unextended primer.

III. Results and Discussions

The design of the surface-bound DNA oligonucleotides employed here is shown in Figure 2. Table 1 shows the four 8mer variable sequences employed (in lower case), which were chosen from a previously described set of 108 possible sequences,[5] along with the encoding scheme utilized to represent the possible values of the SAT variables x and y using two tandem words with different word labels. The four two-word combinations shown, as incorporated into the overall oligonucleotide structure shown in Figure 2, are denoted below by "[00]", "[01]", "[10]", and "[11]", respectively. Their spatial arrangement on a 4-element addressed array is shown in Figure 3. Solving each clause of the SAT problem requires one cycle of MARK, DESTROY, and UNMARK, and thus 2 cycles were employed to solve the 2-variable 2-SAT problem $(x \vee \bar{y}) \wedge (\bar{x} \vee y)$ (Figure 3). The goal of the first computational cycle is to destroy all DNA molecules which do not satisfy the first clause $(x \vee \bar{y})$. This is achieved by hybridizing to the surface those PNA oligomers (see below) that are complementary to the molecules which do satisfy the clause, and then destroying the

remaining (unmarked) molecules. Only one sequence does not satisfy the first clause, namely that for which x is set to 0 and y is set to 1. Thus in Cycle 1 the PNA complements of $x = 1$ and $y = 0$ were combined with primer and hybridized (in the MARK operation) to the surface. After the MARK operation, DNA polymerase Deep Vent (exo$^-$) was added to the surfaces. Since there was no duplex region formed below the primer site of strand [01], the polymerase extended the primers annealed to strand [01] to the downstream spacer region to form a double-stranded restriction enzyme site, which was then recognized by the restriction enzyme *Dpn* II in a subsequent restriction digestion step. Therefore, strands [01] were cleaved (destroyed) from the surface in Cycle 1. In contrast, in the other DNA strands (strands [00], [10], and [11]) that were marked (formed a duplex) by either the complements of $x = 1$ or $y = 0$, the hybridized PNA complements blocked the polymerase extension, preventing double-stranded restriction sites from being formed near the spacer region. Therefore, they remained after the DESTROY operation. The surface was then regenerated by the UNMARK operation to return the remaining surface-bound oligonucleotide strands [00], [10], and [11] to single-stranded form. In Cycle 2, the complements of $x = 0$ and $y = 1$ were combined with primer and hybridized to the surface. After this MARK operation, polymerase extended the primers to form a double-stranded restriction site if the strand was not marked. Strand [10] was not marked in this cycle, and accordingly it was destroyed in Cycle 2. After these two cycles of DNA computing, the DNA molecules left on the surface were strands [00] and [11]. These strands correspond to the correct answers to the SAT problem $(x \vee \bar{y}) \wedge (\bar{x} \vee y)$. The experimental results corresponding to the process outlined above are shown in Figure 4.

Quantification of the residual fluorescence intensity for strand [01] after Cycle 1 and of strand [10] after both cycles, as measured by the intensity of fluorescent primer subsequently hybridized to the surface, showed approximately 10% of the fluorescence signal remained compared to the levels prior to the computing operations. The overall efficiency of the DESTROY operation is thus approximately 90%. This inefficiency in the overall DESTROY operation can derive in principle from inefficiencies for any of the three procedures which comprise the DESTROY operation, i.e., primer hybridization, polymerase extension, and restriction enzyme cleavage. With respect to primer hybridization, as no significant increase in signal was obtained by increasing primer concentration or hybridization time, it is likely that the hybridization efficiency is near unity, i.e., all of the available surface binding sites on the surface are fully occupied under the conditions employed. The inefficiency of polymerase extension may be due to incomplete removal of the base-protecting groups during oligonucleotide synthesis. Residual base-protecting groups on the oligonucleotides might block the polymerase enzyme, causing false terminations. Both in solution phase and on the surface, the efficiency of the polymerase extension step as determined from the ratio of the extended and unextended primers (see Experimental Section) was measured to be >90%, and the efficiency of the *Dpn* II digestion step was measured to be >95% in solution,[25] and about 90% on the surface (data not shown). Further work will be required to improve the efficiency of these enzymatic reactions on surfaces.

During development of the surface reaction it was found that spacer length had a strong effect upon the polymerase extension efficiency. The use of ten S18 spacer moieties rather than 5 or fewer provided a substantially greater efficiency (data not shown). One possible explanation for this behavior is that the surface morphology or chemistry interferes with the polymerase binding and/or extension reaction, and the longer spacer serves to reduce this interaction. It was also noted that the polymerase extension step produces a significant increase in the surface fluorescence intensity (in Figure 4, compare panels a and b (2.3-fold increase) or d and e (0.7-fold increase)). Control experiments in solution produced a similar fluorescence increase (data not shown), showing that the effect is not surface-dependent. The reason for this increase in fluorescence is not known, although it presumably relates to the well-known environmental dependence of fluorescence emission processes.

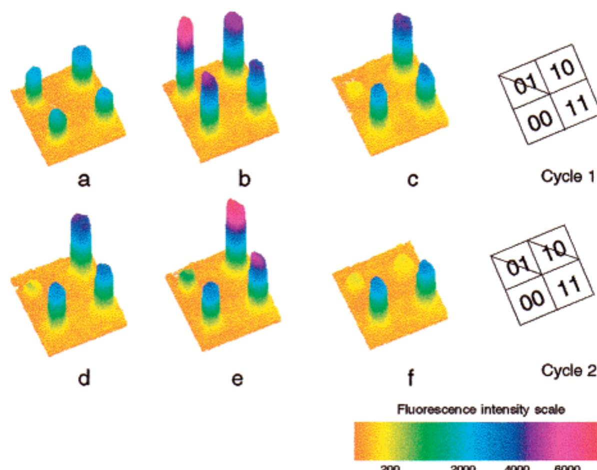

Figure 4. Three-dimensional plot of the fluorescence intensities on a surface as it passes through the steps of a DNA computation. A gold surface was prepared with four spots corresponding to the four DNA strands of structure shown in Figure 2 and Table 1 and arranged as depicted in Figure 3. Each spot is approximately 2 mm in diameter. It was exposed to a solution containing a mixture of fluorescein-labeled primer and PNA complements $x = 1$ and $y = 0$ (see Table 1) and washed giving the image shown in panel a. After the polymerase extension step the fluorescence image shown in panel b was obtained (see text for a discussion of the fluorescence intensity increase observed). The surface was then incubated with restriction enzyme, denatured with urea to remove all the complements, and again hybridized with fluorescein-labeled primer and imaged to yield the result shown in panel c. This concluded the first Cycle of the DNA computing operations. The single-stranded surface was regenerated by denaturing all the hybrids. In Cycle 2, a mixture of fluorescein-labeled primer and PNA complements $x = 0$ and $y = 1$ was hybridized to the surface, giving the image in panel d. Polymerase extension was repeated to yield panel e, followed by restriction enzyme digestion, denaturation, and rehybridization with the fluorescent primers to yield panel f. In the DESTROY operation, >90% of the unmarked strands were removed as ascertained by the measurement of the remaining fluorescence intensity after the operation. A diagram depicting the progression of the SAT problem is provided beside each set of three panels showing the experimental results for the two cycles. The three-dimensional plot was generated using NIH Image version 1.61 software (National Institute of Health, Bethesda, Maryland: http://rsb.info.nih.gov/nih-image/download.html).

In the process of developing this multiple-word DESTROY operation, it was found that when normal oligodeoxynucleotide complements were employed for the MARK operation, they were displaced by the polymerase enzyme during the polymerase extension reaction, compromising the ability of the DESTROY operation to act only upon specifically marked strands. This was found to be a problem for a wide variety of different polymerases examined, including several noted for their low

(25) Wang, L.; Liu, Q.; Frutos, A. G.; Gillmor, S. D.; Thiel, A. J.; Strother, T. C.; Condon, A. E.; Corn, R. M.; Lagally, M. G.; Smith, L. M. *Biosystems* **1999**, *52*, 189−191.

propensity for strand displacement. It was then found that peptide nucleic acids (PNAs) employed in place of the MARK DNA oligonucleotides formed more stable duplexes that were not disrupted by the polymerase extension reaction. Accordingly, the multiple-word DESTROY operation described here employs PNA oligomers in the MARK operation. The use of PNA in a sticker-based model for DNA computation has been proposed previously.[26] The motivation for use of PNAs here arises from three of their characteristics: (a) normal PNAs are not extensible by the DNA polymerases;[27,28] (b) PNA−DNA duplexes are more stable than DNA−DNA duplexes, due to the absence of inter-strand charge repulsion (unlike the charged sugar−phosphate backbone of normal DNA, PNAs have a neutral peptide backbone[9,29]), which makes the blocking of polymerase extension by PNA complements more efficient than that by DNA complements; and (c) PNA−DNA hybrids are stable in a low salt environment, which is provided by most of the enzyme buffers. Unfortunately, the use of PNAs is also accompanied by some disadvantages, as follows: (a) the synthesis of PNAs is less general than the synthesis of normal oligodeoxynucleotides, with particular difficulties for purine-rich sequences;[30] (b) because of the absence of backbone charges, PNAs can form stable internal secondary structures even as short oligomers; (c) the PNA−PNA interaction is much stronger than the DNA−PNA interaction, which means that care must be taken to select PNA sequences that avoid PNA-PNA interactions; and (d) the cost of PNA synthesis is much higher than the cost of DNA synthesis.

Initially 16mer PNA sequences were employed for this study. However, difficulties were encountered in their use (poor blocking of the polymerase extension) which we hypothesized as being due to formation of the intramolecular secondary structures alluded to above. Accordingly, shorter 12mer PNAs were tried, and these were found to work very well in blocking the strand extension reaction and were therefore used henceforth. Considerable thought went into the design of the primer sequence and the particular word sequences employed for these experiments. They were chosen from a previously described set of 108 sequences[5] differing by at least 4 bases out of 8 to provide good hybridization discrimination between one another, and the sequence candidates were screened using the online DNA folding program Mfold version 3.0 (http://mfold2.wustl.edu/~mfold/dna/form1.cgi) to predict possible hairpin structures and permit elimination of those word candidates with a propensity to form such structures. A C++ program was written and utilized to calculate the number of "sliding matches" between the two sequences, that is, possible hybridization binding sites for the primer oligonucleotides or PNAs, formed at the junction between adjacent words or between a word and the adjacent primer binding site or other sequences. Word sequences that led to the formation of such secondary binding sites were not utilized, and sequences that could lead to hybridization between the primer oligonucleotide and the PNAs were also discarded. This work on the design of the words employed served to minimize the chance of significant secondary structure formation or undesired binding events, both of which are important to good performance in the DNA computing operations.

In this work two enzymatic procedures, DNA polymerase-mediated strand extension (using the polymerase Deep Vent (exo$^-$)) and restriction digestion by *Dpn* II, have been employed in tandem to implement a multiple-word DESTROY operation for DNA computing on surfaces. The efficiency of the DESTROY operation for multiple-word DNA computing (90%) is comparable with the previously reported efficiency of the DESTROY operation for single-word DNA computing (94%).[5] The utility of the approach was demonstrated on a small example of the SAT (Satisfiability) problem. This implementation of multiple-word DNA computing is critical to scale-up of the DNA computing process to larger problems, as it allows larger combinatorial sets of molecules to be generated and manipulated than is possible with a single-word approach. If 108 different word sequences were employed for each of 6 words, the corresponding search space would consist of $108^6 = 1.6 \times 10^{12}$ possibilities, corresponding to slightly more than 40 bits ($2^{40} = 1.1 \times 10^{12}$). The oligonucleotide length required to encode this much information would be $(16 \times 6) + 16 + 9 = 121$ nucleotides, (6 words, 1 primer site, 4 base restriction site + adjacent 5 base spacer), within the range of current oligonucleotide synthesis capabilities.[3] The number of MARK and DESTROY operations required to solve the problem grows polynomially with the number of variables,[31] in contrast to the search space which grows exponentially. The multiple-word DESTROY operation described in this paper will enable the development of such surface-based DNA computational devices.

Acknowledgment. The authors would like to thank Dr. Anthony G. Frutos for useful discussions. This work was supported by the Defense Advanced Research Projects Agency (DARPA) and the National Science Foundation.

JA0010195

(26) Roweis, S.; Winfree, E.; Burgoyne, R.; Chelyapov, N. V.; Goodman, M. F.; Rothemund, P. W. K.; Adleman, L. M. *J Comput. Biol.* **1998**, *5*, 615−629.

(27) Hyrup, B.; Nielsen P. E. *Bioorgan. Med. Chem.* **1996**, *4*, 5−23.

(28) Lutz, M. J.; Benner, S. A.; Hein, S.; Breipohl, G.; Uhlmann, E. *J. Am. Chem. Soc.* **1997**, *119*, 3177−3178.

(29) Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P. E.; Nordén, B.; Gräslund, A. *J. Am. Chem. Soc*. **1996**, *118*, 5544−5552.

(30) PNA Sequence Guide, PerSeptive Biosystems, Framingham, MA, 1998.

(31) Cai, W.; Condon, A. E.; Corn, R. M.; Glaser, E.; Fei, Z.; Frutos, T.; Guo, Z.; Lagally, M. G.; Liu, Q.; Smith, L. M.; Thiel, A. J. *Proceedings of the 1st Annual International Conference on Computational Molecular Biology* (RECOMB97) 67-74 (Association for Computing Machinery, New York, 1997).

(32) New England Biolabs Inc. 1998/99 Catalog, Beverly, MA, pp 258−259.