4748–4757 *Nucleic Acids Research, 1997, Vol. 25, No. 23*

Demonstration of a word design strategy for DNA computing on surfaces

Anthony G. Frutos, Qinghua Liu, Andrew J. Thiel, Anne Marie W. Sanner, Anne E. Condon, Lloyd M. Smith and Robert M. Corn*

Departments of Chemistry and Computer Science, University of Wisconsin–Madison, 1101 University Avenue, Madison, WI 53706, USA

Received August 27, 1997; Revised and Accepted October 8, 1997

ABSTRACT

A strategy for DNA computing on surfaces using linked sets of 'DNA words' that are short oligonucleotides (16mers) is proposed. The 16mer words have the format 5′-FFFFvvvvvvvvFFFF-3′ in which 4–8 bits of data are stored in 8 variable ('v') base locations, and the remaining fixed ('F') base locations are used as a word label. Using a template and map strategy, a set of 108 8mers each of which possesses at least a 4 base mismatch with the complements to all the other members of the set (4bm complements) are identified for use as a variable base sequence set. In addition, sets of 4 and 12 word labels of the form ABCD....DCBA that are respectively 8bm and 6bm complements with each other are identified. The 16mers are chosen to have a G/C content of 50% in order to make the thermodynamic stability of the perfectly matched hybridized DNA duplexes similar; a simple pairwise additive method is used to estimate the perfect match and mismatch hybridization thermodynamics. A series of preliminary experiments are presented that use small arrays of 16mers attached to chemically modified gold surfaces and fluorescently labeled complements to study the hybridization adsorption and enzymatic manipulation of the oligonucleotides.

INTRODUCTION

In 1994, Adleman (1) proposed to use a combination of DNA combinatorial chemistry and enzymatic manipulation reactions to solve instances of NP-complete problems. In a recent set of papers we have adapted these ideas to DNA molecules attached to surfaces (2–5), and have proposed to perform logical manipulations of large sets of data by the hybridization and enzymatic manipulation of the attached oligonucleotides. In these experiments, combinatorial mixtures of DNA molecules are attached to chemically modified surfaces, and subsets of this mixture are 'marked' or identified by the hybridization adsorption of complementary DNA molecules. A single-strand-specific exonuclease is then used to destroy all unmarked DNA, and the process is repeated until only a few DNA molecules representing the solutions to a mathematical problem remain on the surface. In our initial experiments, a set of 32 15-base oligonucleotides ('15mers') was used to store 5 bits of information for a 5-variable satisfiability (SAT) calculation (4). In this paper, we propose a word design strategy in which a large amount of data can be stored in linked sets of short 16-base oligonucleotides (16mers) or 'DNA words' that can hold 4–8 bits of information. By linking these DNA words together, the longer DNA molecules required to make large combinatorial sets used in logical calculations can be created, while keeping the DNA chemistry regular and reliable on a much smaller length scale.

What errors can result from the marking of the attached DNA molecules by hybridization adsorption? Our proposed computation strategy assumes that any member of a combinatorial set of DNA molecules can be identified by hybridization to its perfect complement, i.e., the oligomer sequence that will form a DNA duplex in which all of the bases in the original molecule are hydrogen bonded to the correct complementary base. We denote this pair of molecules as the 'perfect match'; all other possible imperfect matches will lead to errors and are defined as '*n*-base mismatches' (*n*bm) depending upon the number of incorrect base pairs (e.g., 4bm for $n = 4$). A set of molecules in which all mismatches are greater than or equal to *n* is denoted as *n*bm complements (e.g., 4bm complements for $n \geq 4$). The stability of DNA hybridization increases with the number of base pairs in the oligomer, so that a longer DNA molecule (50mer or greater) will bind to its perfect match and any 1bm complement with approximately the same strength. For this reason, we propose to use sets of shorter oligomers (16mers) that are linked together either chemically or enzymatically. The 16mers have the following design:

$$5'\text{-FFFFvvvvvvvvFFFF-}3' \qquad \mathbf{1}$$

where the 8 bases labelled 'F' are denoted as the 'word label' and are the same for every 16mer in a word subset, and the 8 bases labelled 'v' are the 'variable' bases that code the data contained in each of the words. In order to keep the strength of binding similar for all of the perfect matches in the combinatorial set, the G/C content of the 16mers is fixed at 50%. The linked sets of 16mers will be attached to a surface to create binding sites for the correct complementary 16mers.

How much information can be stored in these linked sets? The answer depends upon the number of bits that can be stored in a word. There are $4^8 = 65\,536$ or '64K' possible 8mers that can be

*To whom correspondence should be addressed. Tel: +1 608 262 1562; Fax: +1 608 262 0453; Email: corn@chem.wisc.edu

used for both the variable bases and word labels; restricting the G/C content to 50% reduces that number to 17 920. Our initial goal is to create a combinatorial set of 64K molecules by linking 4 different words that each store 4 bits of information in the variable base pairs. These linked sets of molecules must be distinguishable by hybridization adsorption with the perfectly complementary word molecules. Because our previous results (4) indicate that it would be difficult to completely discriminate between two 16mers that differ by only 1 base, we have devised a more robust strategy in which sets of variable 8mer sequences all differ in at least 4 base locations (i.e., 8mers that are 4bm complements to all of the other molecules in the complementary oligonucleotide set). Other researchers have also designed and used sets of multiple mismatching oligonucleotides for other applications (6).

The various design issues for the DNA word strategy are addressed in the subsequent sections of this paper. In the next section, we show how to generate a set of 108 8mers for use in the variable base region that are at 4bm complements using a template and map strategy. The role of matches with molecules within the original combinatorial set (denoted as 'reversals') is also discussed. The appropriate design of the fixed word labels is examined in the section Word label set selection, including how to reduce the possibility of the hybridization of two 16mers that are not in registry (denoted as 'slide matches'). In the Experimental considerations section, the chemistry that we have developed to attach the DNA words onto chemically modified gold surfaces is presented.

Using the results of these next sections, we examine in subsequent sections, some test sets of 16mers that can be used for DNA computations on surfaces. In the 4bm Word set hybridization adsorption experiments section, fluorescently tagged complements are used to examine the hybridization behavior of small arrays of words attached to the chemically modified gold surfaces; the stabilities of the 4 base mismatches in the test sets are analyzed using simplified mismatch thermodynamics calculations. In the Word label hybridization adsorption experiments section, we test the hybridization behavior of DNA words containing the same internal bases but different word labels, and finally, in the Selective enzymatic destruction experiments section we demonstrate that unmarked (single-stranded) 16mers can be enzymatically destroyed on the surface in the presence of marked (hybridized) oligonucleotides.

VARIABLE BASE (8MER) SET GENERATION

Statement of the problem

The problem addressed in this section is that of finding a large set S of 8mers for use in the variable base region of 16mer DNA words, as described in the Introduction. As a reminder, there are a total of 64K (4^8) different 8 base oligonucleotides, and if we require that 50% of the 8 bases be either G or C (so that each perfectly matched duplex contains the same number of hydrogen bonds), this number is reduced to 17 920. The additional requirements of a good set S are that (i) no two 8mers in the set should hybridize with each other's complements (i.e., hybridization adsorption should only occur between a word and its perfectly matched complement), and (ii) no two 8mers in the set should hybridize with each other (this could be important, for example, in the process of hybridizing surface-bound words with a combinatorial set of complements). The larger S is, the more bits of information that can be encoded in each DNA word.

Roughly speaking, the likelihood of hybridization between two 8mers decreases as the number of mismatches between them increases. This suggests that if x and y are any two 8mers in the set S, then the following two properties should hold: (i) the Watson–Crick complement of x should differ from the Watson–Crick complement of y in many bases; and (ii) y should differ from the Watson–Crick complement of x in many bases. This latter constraint should hold even if x and y are identical, to avoid hybridization between two copies of a 'solution' on the surface.

These constraints can be expressed combinatorially, where for this paper the word 'many' in the constraints means '4'. Consider x and y as 'strings' over the alphabet {A, C, G, T}, where the left and right ends of a string represent the 5′ and 3′ ends, respectively, of the corresponding DNA strand. Let z^R denote the reverse or 'reversal' of a string z, and let z^C denote the 'complement' of z, obtained by replacing each A in the string by a T and vice versa, and by replacing each C in the string by a G and vice versa. For the strings z^R and z^C, the left and right ends of a string represent the 3′ and 5′ ends respectively. The problem considered in this section can then be succinctly expressed as follows:

Problem: find a large set S of strings (words) of length 8 over the alphabet {A, C, G, T} with the following properties: (i) each word in S has 4 symbols from {G,C}; (ii) each pair of distinct words x and y in S differ in at least 4 positions; (iii) each pair of words x and y (where x and y may be identical) are such that x^C and y^R in S differ in at least 4 positions.

A set that conforms to property (ii) is referred to as a set of '4bm complements', and a set that conforms to property (iii) is referred to as a set of '4bm reversals.' Finding a maximum-sized subset of the 17 920 8mers that are 4bm complements and reversals is an instance of a well-known NP-hard problem, namely the independent set problem, for which there is no known efficient algorithm (7). Instead, a heuristic method for finding a large subset is developed here.

Solution set of 108 8mers

We have found a set S of size 108 that satisfies the three properties outlined above using the following 'template-map' strategy. A 'template' t is defined as an 8-string over the alphabet {A,C}, and a 'map' m is defined as an 8-string over {0,1}. Each template-map pair (t,m) uniquely describes an 8-string s over the alphabet {A, C, G, T} in the following way: for each 1-bit in the map, change the corresponding bit in the template to its complement, and for each 0-bit, leave the template unchanged. For example, as shown in Figure 1, the template-map pair (AACCACCA, 10100101) describes the 8mer TAGCAGCT.

In the template-map strategy, a set T of templates is found that satisfies all of the conditions of the problem. Call such a set of templates a 'conflict-free' template set. If t and t′ are two templates that are conflict-free, then given any pair of maps m and m′, it will always be the case that the strings described by (t, m) and (t′, m′) satisfy constraints (ii) and (iii). Then independently for each template t in T, a set M(t) of maps is found such that the set of strings described by [t, M(t)] satisfies the conditions of the problem, and M(t) is as large as possible. Because the set of templates is conflict-free, the union over all templates t in T of the sets described by [t, M(t)] is a solution to the problem.

Figure 1 shows the two template-map sets that are used to generate the set of 8mers that are 4bm complements and 4bm reversals. The first set has 6 templates and 16 maps, and the

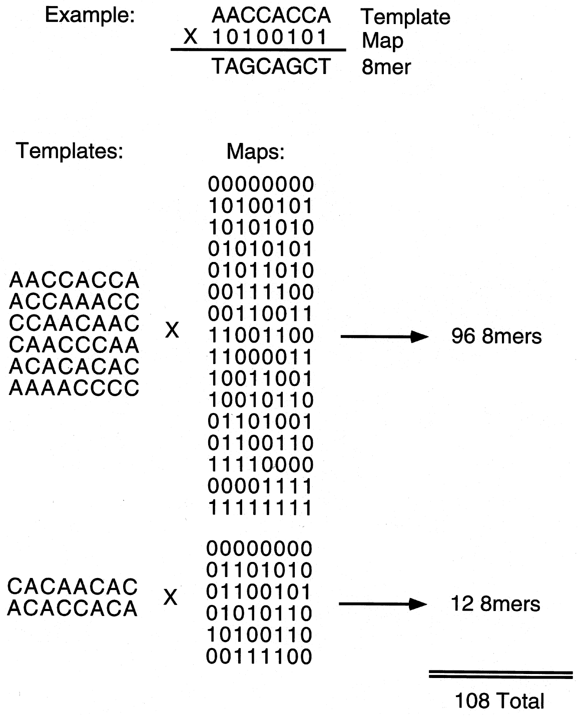

Figure 1. Template-map sets used to generate a set of 108 8mers that contain 50% G/C content and are 4bm complements and reversals. 8mers are generated by crossing each template with each map using the following rule: for each position in a map with a 1, the corresponding position in the template is changed to the complementary base, while for each position in a map with a 0, the corresponding position in the template is unchanged.

second set has 2 templates and 6 maps to yield a total of (6 × 16) + (2 × 6) = 108 distinct 8mers. An analysis of this set shows that of the 108 × 107 = 11 556 possible complementary mismatches, 2140 are 4bms, 1536 are 5bms, 4416 are 6bms, 1536 are 7bms and 1928 are 8bms. The methodology for selection of these templates and maps is described in detail below.

Map selection

A key subproblem, then, is to find as large as possible a map set M for a given template t that satisfies property (i). It turns out that the size of the optimal map depends on the symmetries in the template. There are two possibilities: (a) If a template is a palindrome, that is, it is the same backwards as forwards (for example, the templates AACCCCAA or ACCAACCA), then the largest map set we have found has size 6. One such map set is listed in Figure 1. (b) If the template t is not a palindrome, it must differ in at least 4 places from its reverse t^R (we explain why in the next paragraph). From this it follows easily that if t is a non-palindromic template and m and m′ are any maps, then the strings (t, m) and (t, m′) satisfy property (iii). Therefore if t is a template satisfying property (i) and M is any map set such that the set of strings described by (t, M) satisfies property (ii), then this set automatically satisfies all three properties of the problem. The largest map set we have found satisfying property (ii) has size 16. Again, one such map set is listed in Figure 1.

We now explain why a non-palindromic template t must differ in at least 4 places from its reverse t^R. Suppose that t = xy where x and y are strings of length 4, in which case $t^R = (y^R)(x^R)$. The number of 'mismatches' between the pair x and y^R equals the number of mismatches between y and x^R; therefore it is sufficient to show that if the number of mismatches between x and y^R is greater than 0 (i.e., t is non-palindromic), it must be at least 2. There are three cases, depending on the number of A's in x. If the number of A's in x is 0 or 4, then the number of C's in y is 4 or 0 (respectively) and so there are 4 mismatches between x and y^R. If the number of A's in x is 1, then y^R contains 3 A's, at least two of which must mismatch with the C's in x. For example, if x = CACC and y^R = CAAA, then the last two C's of x mismatch the last two A's of y^R. If the number of A's in x is 3, the argument is symmetric. The third case is when the number of A's in x is 2, and also the number of A's in y^R is 2, for example, if x = CAAC and y^R = CACA. In this case, since x and y^R are not equal, there will always be a position in which x contains an A and y^R a C, and another position in which x contains a C and y^R an A, resulting in at least two mismatches between x and y^R.

Template selection

The second subproblem is to create a good conflict-free template set T. Clearly it is desirable to have non-palindromic templates in the template set T. More precisely, if A is the number of palindromic templates and B is the number of non-palindromic templates in T, then the goal is to maximize 6A+16B. The best template sets we have found contain 2 palindromic templates and 6 non-palindromic templates. One such template set is shown in Figure 1. As mentioned above, the set of strings S resulting from this set of templates and the maps from above has size 6 × 2 + 16 × 6 = 108. How was this template set identified? First, template sets containing AAAACCCC were considered. Any other template in the set must be of the form xy where both x and y have 4 symbols, 2 of which are A's and 2 of which are C's (otherwise the template will 'conflict' with AAAACCCC). We say that a 4-string x appears in a template t if t = xy or t = yx for some y. There are 6 possible 4-strings with 2 A's and 2 C's:

$$\{\text{CACA, ACAC; ACCA, CAAC; CCAA, AACC}\} \qquad \mathbf{2}$$

Here, these 6 possible strings are grouped into pairs, where in each pair one string is obtainable from the other by replacing A's with C's and vice versa. For convenience, given a string x of A's and C's, we denote by x^S the string obtained by replacing all the A's in x by C's and vice versa.

We will show that the best conflict-free template set containing AAAACCCC has 6 non-palindromic strings and 2 palindromic strings. The following example illustrates the main idea. Suppose a 4-string X appears in a template t = xy. Then if x appears on the left side of any other template t′, then t′ must be x(y^S), in order that the pair t, t′ satisfy property (ii). The string x cannot appear on the left side of any other template; moreover, if x is palindromic then x also cannot appear on the right side of any template, in order that property (iii) holds.

In extending the reasoning used in this example, it is useful to consider separately those 4-strings that are palindromes (namely ACCA and CAAC) and those that are not (namely CACA, ACAC, CCAA and AACC). (i) Suppose that x is a palindrome and x appears more than twice in templates of T. It can be shown that in this case, without loss of generality the set T contains the strings xx, x(x^S) and (x^S)(x^S), and these are all the strings in T involving x and (x^S). Two of the three of these are palindromic. (ii) Suppose that x is not a palindrome and x appears more than twice in templates of T. In this case, without loss of generality the strings in T involving x and x^S are xx, $x(x^S)$ and $(x^S)x$. Again, two of the three are palindromic. (iii) Finally suppose string x appears only twice, as does x^S. If one of these appearances is with the string y (i.e., one template in which x appears is either xy or yx) then in order to maximize the size of the template set, the remaining appearances must be either with y, y^S or $(y^R)^S$. This results in 4 non-palindromic strings in T. Hence, the best way to construct a template set T containing AAAACCCC is to use 2 of the 3 possible pairs of strings of Equation **2** to construct 4 non-palindromic strings, and then use the remaining pair to obtain 1 more non-palindromic string and 2 palindromic strings. This, together with AAAACCCC, results in 6 non-palindromic strings and 2 palindromic strings, as in the above template set.

We also considered template sets not containing AAAACCCC. It appears that no such template set is better, although a proof of this is tedious. In particular, if a template set contains only templates of the form xy, where one of x and y has 3 A's and 1 C (and the other has 3 C's and 1 A), then there are at most 4 (non-palindromic) strings in T. Finally, we note that the map sets of size 6 and 16 for the palindromic and non-palindromic templates, as described above, were identified using reasoning similar to that used for constructing a good template set.

WORD LABEL SET SELECTION

A second major design consideration in the development of a DNA word strategy is the selection of a set of fixed base word labels to accompany the variable base 8mers described in the previous section. Recall from Equation **1** that a word label consists of two 4mers located on either end of the 16mer DNA word (FFFF....FFFF); these 8 base locations will be used to uniquely identify a particular DNA word in a linked set. An appropriate set of word labels should have the following properties: (i) the 8mer word label sequence should have a G/C content of 50%, (ii) a word label should have many mismatches with the Watson–Crick complement of another word label ('inter-word complements'), (iii) a word label should have many mismatches with the reversal of another word label ('inter-word reversals'), and (iv) a word label sequence should create many mismatches in all of the possible slide match configurations in a word set, where a 'slide match' is defined as the partial hybridization of two DNA words that are not in registry (i.e., arranged so that the bases at the 3′ and 5′ ends of the two DNA words are not aligned).

In a previous paper on single base encoding strategies for DNA words, we have examined possible word labels and the effect of word label structure on slide matches (8). Based on these previous results, we choose that the 16mer words described in this paper all have the format shown in Equation **3**.

$$5'\text{-ABCDvvvvvvvvDCBA-}3' \qquad \mathbf{3}$$

This format has been shown to be very good at reducing slide matches, and also has the advantage that all intra-word reversals are complete mismatches (8bm reversals) in the word label regions. This format also reduces the chance of hairpin structure formation between the two word label 4mers.

The best choice for a set of 4mer sequences of the form ABCD would be ones that differ in all 4 locations (4bm complements) from each other. This set of word labels would have 8 base mismatches between two different words containing the same internal bits (8bm inter-word complements). In Equation **4** we identify a set of four 4mers (Set 1) that are 4bm complements and 2bm reversals of each other:

$$\text{Set 1: \{AACG, TTGC, CGAA, GCTT\}} \qquad \mathbf{4}$$

We will use Set 1 in our initial set of four word labels. In order to test the word label sequences, we have written a simple computer program that calculates the number of matches for a set S of DNA words with its complements and reversals (S^C and S^R) (8). When two DNA molecules are in registry, the match is denoted an 'inherent' match, and when they are allowed to slide past each other it is denoted a 'slide' match. Table 1 shows the results of the program for a set of 108 16mers that use the variable base 8mers described in the previous section and a word label from Set 1 of the form GCTT...TTCG. The first column in the table is the number *m* of correctly matched base pairs that appear in a given duplex (defined as an '*m*-base partial match'), and the subsequent columns are the numbers of inherent, slide and total (= inherent + slide) *m*-base partial matches that occur in the calculation. The 'Inherent' column shows that, as expected, in the 108 DNA word set there are 108 16 base matches (i.e., 108 perfect complementary matches), and no 15, 14 nor 13 base partial matches (i.e., no 1bm, 2bm or 3bm complements or reversals). The numbers for 12 to 8 base partial matches in this column are the numbers of 4bm to 8bm complements reported in the previous section, and the numbers for 4 to 0 base partial matches are the various reversal matches that can be formed. Note that the reversals are well separated from the complements in the inherent matches; this is because the word label GCTT...TTCG is an 8bm reversal of itself.

Table 1. Inherent and slide matches for the 108 set with the word label GCTT........TTCG

Matches	Inherent	Slides	Total
16	108	0	108
15	0	0	0
14	0	0	0
13	0	0	0
12	2140	0	2 140
11	1536	0	1 536
10	4416	54	4 470
9	1536	104	1 640
8	1928	845	2 773
7	0	2 962	2 962
6	0	13 532	13 532
5	0	34 062	34 062
4	1052	70 734	71 786
3	1536	110 918	112 454
2	5952	151 774	157 726
1	1536	153 314	154 850
0	1588	161 541	163 129

The 'Slides' column in Table 1 reveals that the largest *m*-base partial slide match is 10. We have previously defined a quality parameter Q as the largest m-base partial match created by a slide configuration; the lower the Q, the better the arrangement of word labels is at reducing slide matches (8). A Q of 10 demonstrates that the slide matches are not as important as the inherent matches in this word set. Our previous paper examined methods for searching for word labels that minimize Q (8). As the number of words are increased, the Q of the entire set will increase due to the addition of inter-word slides. However, with the proper selection of word labels, the Q of a DNA word set can be kept at or below the Q = 12 level. For example, a Q of 12 was calculated for the word set created from the 108 4bm 8mers and the four word labels of Set 1.

A word label set larger than four can be created by relaxing the mismatch criteria. In Equation **5** we list an additional set of eight 4mers (Set 2) that, when added to Set 1, yield a total set of 12 word labels that are 3bm complements and 2bm reversals of each other:

Set 2: {ACAC, AGGT, TCCA, TGTG, CATC, CTCT, GAGA, GTAG} **5**

The addition of Set 2 provides us with all of the computational power that we will require. For example, imagine that we create a set of DNA words that can store 6 bits of data in the internal base region (corresponding to 64 unique variable 8mers), and use all 12 of the word labels in Sets 1 and 2. By linking together these DNA words we could in principle produce a combinatorial set of $(64)^{12} = 10^{21}$ different DNA molecules (192mers). In our initial studies, we will limit ourselves to just the 4 word label sequences defined by Set 1; as stated in the Introduction, our modest first goal will be to create a combinatorial set of 64K different DNA molecules.

EXPERIMENTAL CONSIDERATIONS

Materials

The chemicals 11-mercaptoundecanoic acid (MUA) (Aldrich), poly(L-lysine) hydrobromide (PL) (Sigma), sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (SSMCC) (Pierce), urea (Bio-Rad Laboratories), and triethanolamine hydrochloride (TEA) (Sigma) were all used as received. Gold substrates were prepared by vapor deposition onto microscope slide covers (No. 2, 18 × 18 mm) that had been silanized with (3-mercaptopropyl)trimethoxysilane (Aldrich) as described previously (9). Millipore filtered water was used for all aqueous solutions and rinsing. All oligonucleotides were synthesized on an ABI DNA synthesizer at the University of Wisconsin Biotechnology Center. Glen Research 5′-Thiol-Modifier C6 and ABI 6-FAM were used for 5′-thiol-modified and 5′-fluorescein-modified oligonucleotides respectively. Prior to purification, thiol-modified oligonucleotides were deprotected as outlined by Glen Research Corp. (10). Before use, each oligonucleotide was purified by reverse-phase binary gradient elution HPLC (Shimadzu SCL-6A). All thiol oligonucleotides were used immediately after purification. Because thiol oligonucleotides slowly oxidize to form disulfide dimers, care must be taken to store free thiol oligonucleotides under an inert atmosphere. All DNA concentrations were verified prior to use with an HP8452A UV-VIS spectrophotometer. The 5′-thiol DNA solutions used in the surface attachment reactions had a DNA concentration of 1 mM in a pH 7, 100 mM triethanolamine (TEA) buffer. DNA hybridization and rinsing employed a pH 7.4 '2× SSPE/0.2% SDS' buffer that consisted of 300 mM NaCl, 20 mM sodium phosphate, 2 mM EDTA and 6.9 mM sodium dodecyl sulfate. Removal of hybridized complementary molecules (referred to as denaturation or 'Unmark') was accomplished by immersing the sample in 8.3 M urea at 37°C for 15 min.

DNA surface attachment chemistry

DNA oligonucleotides were immobilized onto polycrystalline gold thin films via a four step chemical modification depicted in Figure

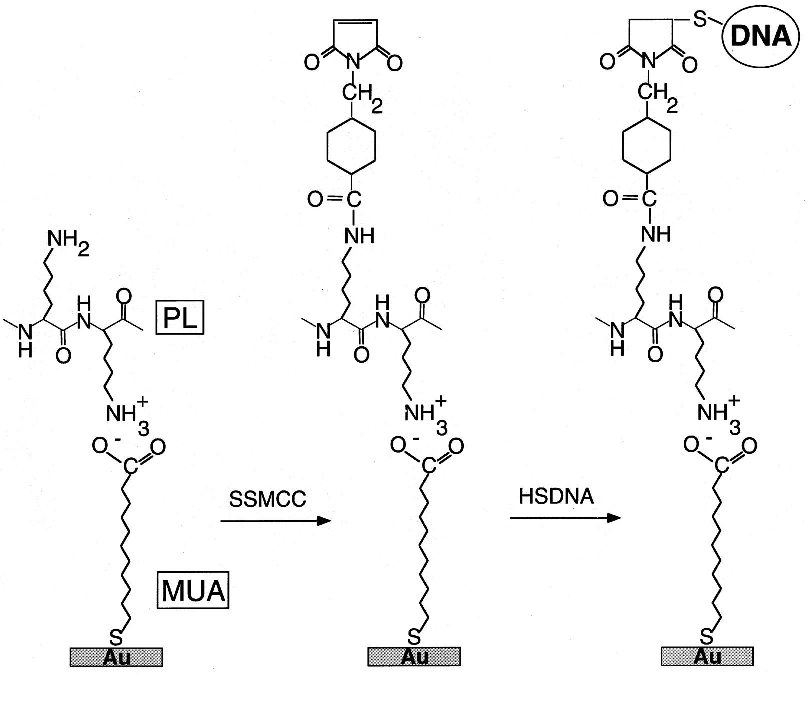

Figure 2. Reaction scheme showing the surface attachment chemistry of HS-DNA onto chemically modified gold films. The gold surface is modified with a monolayer of 11-mercaptoundecanoic acid (MUA) followed by the electrostatic adsorption of a poly-L-lysine monolayer (PL). This amine-terminated surface is then reacted with the bifunctional linker SSMCC, creating a thiol-reactive maleimide surface which is subsequently reacted with single-stranded 5′-thiol modified DNA.

2. All of the chemical modification steps have been thoroughly characterized previously with a combination of polarization modulation FTIR reflection–absorption spectroscopy (PM-FTIRRAS) and surface plasmon resonance (SPR) film thickness measurements (11–13). The first two chemical modification steps were the formation of a MUA alkanethiol self-assembled monolayer on the gold surface, followed by the electrostatic adsorption of a poly-L-lysine (PL) monolayer (12). As shown previously (11), these steps create an amine-terminated gold surface that can then be reacted with the heterobifunctional linker sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (SSMCC). This linker creates a thiol-reactive maleimide terminated surface (depicted in Fig. 2) that can then be reacted with single-stranded 5′-thiol modified DNA by spotting with a 1 mM DNA solution for at least 12 h. After exposure to the DNA solution, the surface was rinsed with water, soaked for at least 1 h in 2× SSPE/0.2% SDS and then subjected to three hybridization/denaturation cycles to remove any non-specifically bound DNA. DNA-modified gold surfaces prepared in this manner were found to be robust and stable, and the attached DNA 'probe molecules' could be cycled through many (>10) hybridization/denaturation steps with minimal degradation. From the PM-FTIRRAS and SPR measurements (13), the DNA probe molecule surface density was estimated to be 5×10^{12} molecules/cm^2.

Surface fluorescence measurements

Surface fluorescence measurements of hybridization adsorption were performed on a Molecular Dynamics FluorImager 575. Hybridization to the attached DNA probe molecules was accomplished by exposure to a 2 µM solution of 5′-fluorescein-labeled 'target' oligonucleotides in 2× SSPE/0.2% SDS. A 20 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 for 30 min, after which the sample was immersed in a beaker of 2× SSPE/0.2% SDS buffer for 10 min. The sample was then placed face down on top of a glass scanner tray with a droplet of 2× SSPE/0.2% SDS buffer between the gold surface and tray and then scanned with the FluorImager. Although a gold surface can in principle quench the fluorescence of an adsorbed monolayer, the target DNA molecules are tethered at a sufficient distance from the substrate (>12 nm) that ample fluorescence signal was observed (14). Washes at 37°C were accomplished by immersing the slide in a preheated beaker of 2× SSPE/0.2% SDS.

Surface exonuclease experiments

The enzymatic destruction of single-stranded oligonucleotides in the presence of hybridized DNA molecules on the gold surface was accomplished by reacting the surface with 20 U of the single-strand-specific enzyme *Escherichia coli* Exonuclease I (Amersham) in a pH 9.5 buffer consisting of 67 mM glycine (Bio-Rad Laboratories), 6.7 mM $MgCl_2$ (New England Biolabs), 10 mM 2-mercaptoethanol (Sigma), 1 M NaCl and 100 µg/ml BSA (New England Biolabs). Enzymatic digestion was allowed to proceed for 3 h at room temperature after which the surface was rinsed with water.

Melting temperature measurements

DNA melting curves were obtained by monitoring the absorbance of DNA solutions at 260 nm as a function of temperature with an HP8452A UV-VIS spectrophotometer equipped with an HP89090A Peltier temperature control accessory. Melting temperatures were measured in pH 7 buffer solutions consisting of 10 mM sodium phosphate, 1 mM EDTA, 1 M NaCl and 2 µM oligonucleotide. A ramp rate of 1°C/min with a hold time of 1 min was used over the range 25–85°C to record the DNA melting curve. The T_m (if observed) was determined as the temperature at which the first derivative of the raw UV absorbance curve was a maximum. T_m data are estimated to be accurate within ±1.5°C.

Table 2. Oligonucleotides used in the four spot fluorescence experiments

Word	5′ HS-$(T)_{15}$GCTT........ TTCG 3′	Complement	3′ CGAA........ AAGC-Fl 5′
W1	TTGGACCA	C1	AACCTGGT
W2	AACCACCA	C2	TTGGTGGT
W3	ATGCAGGA	C3	TACGTCCT
W4	ATCGAGCT	C4	TAGCTCGA

4BM WORD SET HYBRIDIZATION ADSORPTION EXPERIMENTS

Four spot fluorescence measurements

In order to examine the hybridization adsorption behavior of the 4bm DNA word set, a series of fluorescence imaging experiments was performed on a small array of 4 words attached onto a chemically modified gold surface as depicted in Figure 3. The attached words, denoted as W1 through W4 and listed in Table 2, all contain the word label GCTT...TTCG and have internal 8mers that are members of the 108 4bm 8mer set identified in the Variable base (8mer) set generation section. A 15 base poly-T spacer was included at the 5′ end of each surface-bound word in order to facilitate hybridization adsorption by distancing the duplex forming region from the surface (15). The fluorescein-labeled complements of these words are also listed in Table 2 and are

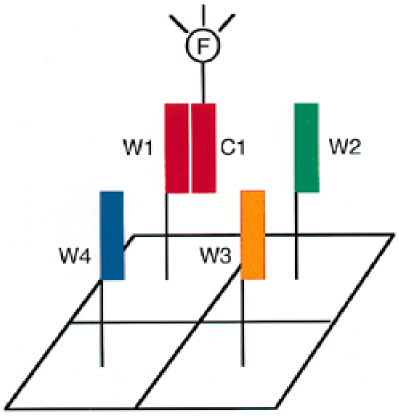

Figure 3. Arrangement of 4 DNA words attached to a chemically modified gold surface used in the four spot fluorescence experiments. The sequences of the words are listed in Table 2.

denoted as C1 through C4; hybridization adsorption to the 4 word set was studied by the sequential exposure of the surface to each of these fluorescent complements. In between each exposure step, the surface was regenerated (i.e., an 'unmark' operation was performed in which all adsorbed complements were removed) by immersing the sample in 8.3 M urea at 37°C for 15 min. Figure 4 shows the results of four successive hybridizations to C1 through C4. The column on the far right shows the internal bits of the surface bound words (top sequence) and the fluorescein-labeled complement (bottom sequence) in each hybridization step; the perfect match duplex is shown in blue and the mismatched bases are shown as underlined, red letters. This particular set of DNA words was chosen so that every mismatch (W*n*–C*m* where $n \neq m$) in the set is a 4bm. For each hybridization step, only the perfect match (W*n*–C*n*) was observed on the surface (within the detection limit of the fluorescence imaging experiment) after washing the surface in a 37°C buffer solution for 5 min. Also shown in Figure 4 is the fluorescence image obtained at room temperature (22°C) prior to washing at 37°C; at room temperature two mismatches appeared (this point is discussed further below), and the perfect match signals were ~4% stronger. The 4% loss in signal from 22 to 37°C did not increase with longer exposure times at the higher temperature. These discrimination results are a significant improvement over our previous experiments using a single-base encoding strategy in which a temperature of 50°C was necessary to effect single base mismatch discrimination, resulting in a loss of ~70–80% of the perfect match (4).

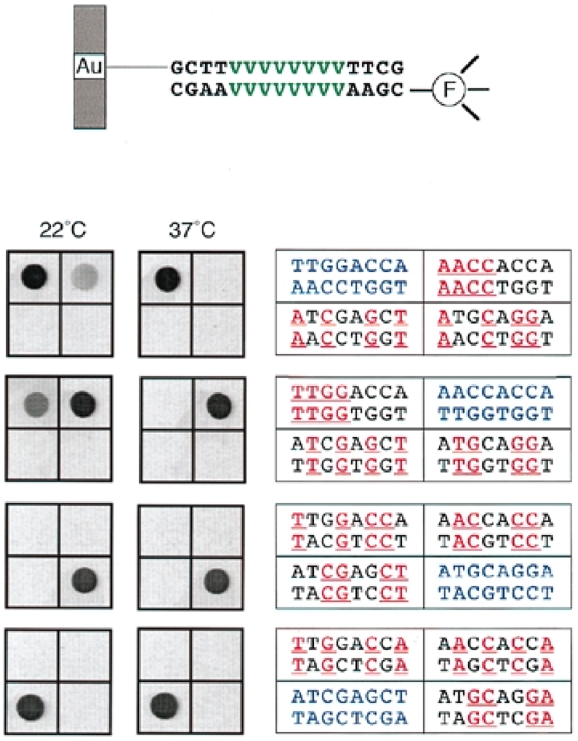

Figure 4. Four spot fluorescence imaging measurements on a representative set of 4 DNA words. Hybridization adsorption to the four word set was studied by the sequential exposure of the surface to each of the fluorescent complements. Between each exposure step, the surface was regenerated by exposure to urea. The far right column shows the internal bits of the surface bound words (top sequence) and the fluorescein-labeled complements (bottom sequence) in each hybridization step; perfect match duplexes are shown in blue, and mismatched bases are shown as red, underlined letters. Note that for each hybridization step, only the perfect match was observed after washing the surface in buffer at 37°C for 5 min.

4bm Melting temperatures and thermodynamic calculations

Our 4bm strategy has been chosen to facilitate discrimination between perfectly matched duplexes and all mismatched duplexes. One measure of a duplex's stability is its melting temperature, T_m. Solution melting curves were measured for each of the duplexes formed between C1 and the words W1–W4, listed in Table 2. As listed in Table 3, a T_m of 71°C was measured for the perfectly matched duplex W1–C1, and a T_m of 40°C was observed for the mismatch W2–C1. This mismatch is one of the two mismatches that was observed at room temperature in the fluorescence measurements in Figure 4; the 30°C decrease in melting temperature from the perfect match is the reason why a high degree of discrimination was achieved by rinsing at 37°C. No melting temperatures were observed for W3–C1 or W4–C1, implying that they are below the starting temperature of the melting curve experiment (25°C).

Table 3. Experimental and predicted T_m[a] for various duplexes

	Internal bases[b]	#bms[c]	#amps[d]	Expt. T_m (°C)	Method 1 -ΔG° (kcal/mol)	Pred. T_m (°C)	Method 2 -ΔG° (kcal/mol)	Pred. T_m (°C)
W1–C1	TTGGACCA AACCTGGT	0	–	70.6	26.9	75.2	24.0	64.5
W2–C1	**AACC**ACCA **AACC**TGGT	4	3	40.0	16.5	55.0	8.4	24.5
W3–C1	**A**TG**C**A**GG**A **A**AC**C**T**GG**T	4	1	<25	12.1	41.0	10.3	30.2
W4–C1	**A**T**C**GA**G**C**T** **A**A**C**CT**G**G**T**	4	0	<25	10.3	33.7	9.8	28.9

[a] In 1 M NaCl and a DNA concentration of 2 µM.

[b] Internal 8 bases of a 16mer (see Table 1); underlines indicate mismatched base pairs.

[c] Number of base mismatches.

[d] Adjacent mismatch pairs.

Table 4. Classification and predicted thermodynamics of 4 base mismatches (4bm) using Method 1

4bm Type[a]	amps[b]	NN Terms[c]	Avg. -ΔG° (kcal/mol)	Avg. -ΔH° (kcal/mol)	Avg. T_m[d] (°C)	#
$\overset{\frown\frown\frown}{\mathrm{XXXX}}\mathrm{oooo}$	3	10	16.2	88.0	51.3	352
$\mathrm{o}\overset{\frown}{\mathrm{XX}}\mathrm{oo}\overset{\frown}{\mathrm{XX}}\mathrm{o}$	2	9	14.3	80.4	45.9	660
$\overset{\frown}{\mathrm{XX}}\mathrm{ooXooX}$	1	8	12.4	72.6	39.3	568
oXoXoXoX	0	7	10.3	64.0	30.9	560
Perfect Match	-	15	25.4	124.4	71.4	108

[a] Internal bits of a 16mer; an'X' represents a mismatched base pair.

[b] Adjacent mismatch pairs.

[c] Nearest neighbour terms.

[d] In 1 M NaCl and a DNA concentration of 2 μM.

Also shown in Table 3 are the results of two simple estimation methods for calculating $\Delta G°$ and T_m for each of the four duplexes. The stability of a particular DNA duplex depends both upon the hydrogen bonding of each base pair and the stacking interactions between nearest neighbors. While estimation of mismatched duplex stability is a very difficult and complex task, Breslauer *et al.* (16) and other researchers (17–19) have developed data sets for estimating the stability of perfectly matched duplexes by summing the contributions of each pair of nearest neighbors in a DNA duplex. We have also used these calculations in our single base mismatch studies (4). Using the set of parameters given by Breslauer *et al.* and an equation suggested by Wetmur (equation 2a from ref. 19) we calculate a $\Delta G°$ of –26.9 kcal/mol, a $\Delta H°$ of –127.1 kcal/mol and a T_m of 75 °C for the perfectly matched duplex W1–C1. Other sets of thermodynamic parameters have been suggested; for example, using the values suggested by Quartin and Wetmur (17) we calculate a $\Delta G°$ of –24.0 kcal/mol, a $\Delta H°$ of –132.1 kcal/mol and a T_m of 64.5 °C for the perfectly matched duplex W1–C1. This calculation includes the contribution of the 5′ dangling end (20), but ignores the influence of the 5′ fluorescein label which would lead to a slightly higher T_m (21). The experimentally observed value for T_m of 71 °C falls in between these two calculations. We have also calculated (using Breslauer's parameters) an average $\Delta G°$ and T_m of –25.4 kcal/mol and 71 °C respectively for the entire set of 108 4bm 16mers of the form GCTT...TTCG using the internal 8mers described in the Variable base (8mer) set generation section. These numbers also match well with the experimentally observed melting temperatures of the perfectly matched duplexes.

While the calculation of perfectly matched DNA duplex stability is well-defined, the treatment of mismatched duplexes such as W2–C1 is not in general straightforward. We have devised a simple modification of the nearest-neighbor pair model in order to estimate the stability of mismatched duplexes (4). This methodology (denoted in Table 3 as 'Method 1') modifies the nearest-neighbor calculation by not including any nearest-neighbor pairs that contain a mismatched base. The results of this calculation are listed in Table 3, and predict melting temperatures of 55, 41 and 34 °C for the mismatch duplexes W2–C1, W3–C1 and W4–C1 respectively. Also listed in Table 3 are the $\Delta G°$ and melting temperatures for these duplexes using a different calculation method (denoted as 'Method 2' in Table 3) for mismatched duplexes suggested by Wetmur (20) that predicts duplex stability based upon the type of mismatch. In this method, the duplex stability is first calculated as if the mismatch had no effect, after which the energetics are corrected by adding an appropriate destabilization factor depending upon the identity of the mismatch. This method predicts a T_m of 21 °C for the W2–C1 duplex; once again, the observed T_m of 40 °C falls in between the two calculations. Whereas Method 2 only considers mismatch identity, Method 1 is only concerned with the number of adjacent mismatch pairs. A more accurate estimation of mismatch hybridization thermodynamics would include consideration of longer range interactions such as the formation of stems and loops.

As seen in Table 3, Method 1 successfully predicts that the W2–C1 duplex is the most stable of the mismatched duplexes. This duplex is also one of only two mismatched duplexes in Figure 4 that has all four of the mismatches adjacent to each other. The other such mismatch, W1–C2, is also the only other mismatched duplex which appears in the 22 °C fluorescence measurements. This suggests, as noted by other researchers (22), that mismatch connectivity plays a significant role in mismatch stability. As mentioned in the Variable base (8mer) set generation section, in the 108 4bm DNA word set there are a total of 2140 4bm complements. We can classify the mismatched duplexes formed from these 4bm complements by the number of adjacent mismatch pairs (amps), which varies from zero to three. Table 4 lists the numbers of each type of 4bm complement and the average thermodynamics calculated by Method 1. For 4bm complements with 3 amps such as W1–C2, an average $\Delta G°$ of –16 kcal/mol is calculated, whereas 4bm complements with 0 amps have a calculated $\Delta G°$ of –10 kcal/mol. It should be noted, however, that no matter how the 4bms are arranged, the 4bm complements are always significantly (>10 kcal/mol) less stable than the perfect matches, as required for the efficient hybridization marking of the DNA word set.

WORD LABEL HYBRIDIZATION ADSORPTION EXPERIMENTS

In a second set of fluorescence imaging experiments, the hybridization behavior of DNA words containing the same internal bases but different word labels was examined. Two 16mers that had the internal 8mer sequence AACCAACC and the word labels AACG...GCAA and GCTT...TTCG were used as a

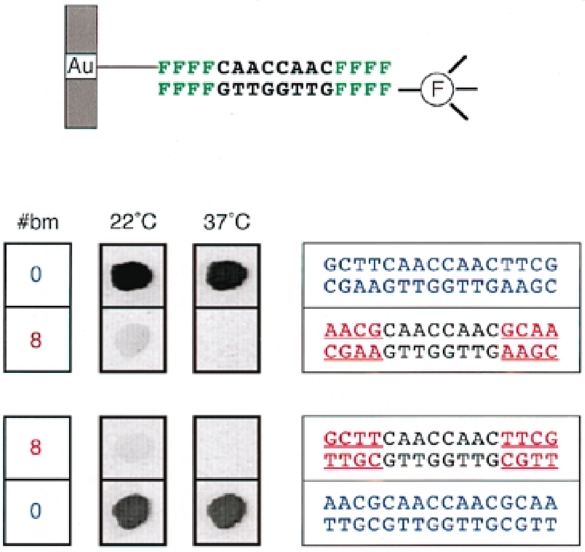

Figure 5. Word label discrimination experiment. Two 16mer words with the same internal 8mer sequence were immobilized onto a chemically modified gold surface and then sequentially exposed to the fluorescein-labeled complement of each word. Mismatched base pairs are shown as red, underlined letters. Discrimination was achieved by rinsing the surface in buffer solution for 5 min at 37°C.

test set (denoted as W5 and W6 respectively). This pair of word labels was chosen from the set of four (Set 1) described in the Word label set selection section, and are 8bm complements with each other. The two DNA words W5 and W6 were immobilized in two different spots on a chemically modified gold surface and then exposed to a hybridization solution containing the fluorescently labeled complement C5. The resulting fluorescence image is shown in Figure 5 along with the sequences for the possible DNA duplexes. As in Figure 4, the perfectly matched duplexes are shown in blue and any mismatched base pairs are shown in red and underlined. As seen in Figure 5, a high level of discrimination between matched and mismatched duplexes was readily achieved by washing the surface in a buffer solution for 5 min at 37°C. The surface was then regenerated (unmarked) and subsequently exposed to a solution containing the fluorescently labeled complement (C6) of the other attached DNA word. Once again, rinsing the surface in a buffer solution for 5 min at 37°C led to virtually complete discrimination. Note that in both of these hybridization adsorption marking experiments, some hybridization adsorption of the mismatched duplexes W5–C6 and W6–C5 was detected at 22°C, even though these duplexes are 8bms. A solution T_m of 45°C was measured for the mismatched duplex W5–C6, and no T_m was observed above 25°C for W6–C5. The most likely explanation for the unusual stability of the 8bm W5–C6 duplex is that the mismatched bases in these molecules are located at the ends of the DNA duplex; mismatches at the ends of an oligomer are known to be less disruptive than mismatches in the middle of the molecule (23–25). If necessary, the word labels could be rearranged in the molecule to include internal base pair locations. However, because the discrimination level observed at 37°C is comparable to that observed in the 4bm word set fluorescence measurements, the word labels should provide sufficient discrimination in any marking operation of linked DNA words.

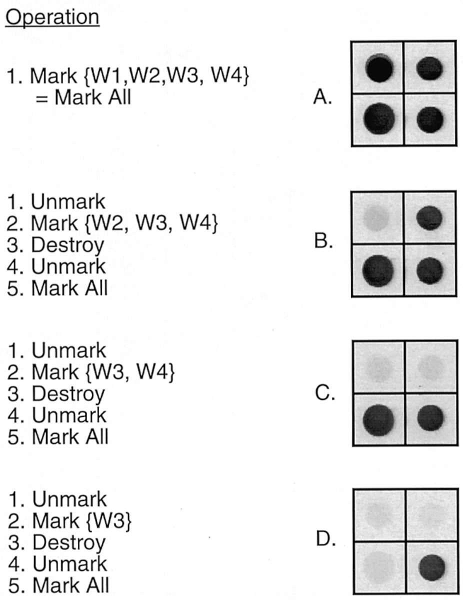

Figure 6. Selective enzymatic destruction experiment. A surface with the 4 DNA words W1–W4 arranged as depicted in Figure 3 was prepared and then exposed to a solution containing the set of fluorescein-labeled complements C1–C4 (i.e. the operation 'Mark All') giving the image shown in (**A**). The surface then underwent the following series of operations to give the image shown in (**B**): 1. Unmark: the removal of all hybridized complements; 2. Mark {W2, W3, W4}: exposure to a solution containing the complements to W2, W3, W4; 3. Destroy: exposure to a solution of the single-strand-specific enzyme Exonuclease I; 4. Unmark; 5. Mark All. This series of operations was repeated two more times to remove W2 and W4 as shown in (**C**) and (**D**) respectively. Exonuclease digestion removed >94% of W1, W2 and W4.

SELECTIVE ENZYMATIC DESTRUCTION EXPERIMENTS

In a final set of fluorescence imaging experiments, the ability to enzymatically destroy unmarked words in the presence of marked words was demonstrated. Our DNA computing strategy requires the selective recognition and enzymatic manipulation of surface-bound DNA molecules. Repeated cycles of the Mark, Destroy and Unmark operations constitute the DNA computation process, permitting subsets of the initial combinatorial space to be eliminated, and leaving the desired solutions as represented by the DNA molecules to the problem of interest. Figure 6 shows the results of a 3 cycle mark and destroy experiment that uses the single-strand-specific enzyme *E.coli* Exonuclease I to remove all single-stranded oligonucleotides from the chemically modified gold surface. A surface with the four DNA words W1 through W4 arranged in the previous 4 spot experimental geometry (see Fig. 3) was prepared, and then exposed to a solution containing the set of all fluorescein-labeled complements C1 through C4 (this operation is denoted as 'Mark{W1, W2, W3, W4}' or 'Mark All'). As expected, the fluorescence image of this surface (Fig. 6A) shows four spots. The surface then underwent the following series of operations:

1. Unmark: the removal of all hybridized complements;
2. Mark{W2, W3, W4}: exposure to a mixture of the complements to W2, W3 and W4;

3. Destroy: exposure to the exonuclease solution;
4. Unmark;
5. Mark All.

The fluorescence image of the surface after this series of operations is shown in Figure 6B, and clearly shows that the DNA word W1 has been removed from the surface. An analysis of the residual fluorescence shows that >94% of W1 has been removed by the exonuclease digestion reaction. This series of operations {Unmark/Mark/Destroy/Unmark/Mark All} was repeated two more times to remove >94% of both W2 and W4; Figure 6C and D clearly show that this enzymatic destruction of single-stranded words can be performed in a repeated fashion.

The intensity of the fluorescence from the remaining word W3 in Figure 6D is decreased from its original value in Figure 6A by ~30%. However, this diminution is not attributed to digestion of the double-stranded DNA by the exonuclease because a similar decrease in fluorescence intensity was observed in a control experiment when the 4 spot surface was exposed three times to the exonuclease buffer solution that did not contain any enzyme. No appreciable loss of fluorescence intensity was observed when the surface was repeatedly exposed to Mark/Unmark cycles, so we attribute the intensity loss to some component of the enzyme buffer. Further experiments are currently in progress to identify the source of this loss.

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we have described a word design strategy for storing and manipulating information in DNA molecules attached to surfaces. A 16 base oligonucleotide is used as the basic word unit; information is stored in the 8 internal variable base locations and the 4 fixed bases on either end serve as a unique word label. A template-map strategy was used to generate a set of 108 8mers that are 4 base mismatches with each other (4bm complements) for use in the variable base region, and sets of 4 and 12 word labels that are 8bm and 6bm complements respectively have been identified. Surface fluorescence experiments on sets of oligonucleotides attached to a chemically modified gold surface have been used to demonstrate that specific DNA molecules in this word set can be identified or ‘marked’ by hybridization adsorption, and that DNA words with the same internal bases but different word labels can also be selectively marked. A combination of simple thermodynamic calculations and melting temperature measurements has been used to help quantify the hybridization selectivity of the word sets. In a final set of preliminary experiments, the enzymatic manipulation of sets of attached DNA words has been demonstrated with the selective enzymatic destruction of unmarked DNA molecules by the single-strand-specific enzyme *E.coli* Exonuclease I. Further characterization and optimization of this and other surface enzyme reactions are currently in progress.

As outlined previously (2), after the mark and destroy operations have been characterized, the next step in the demonstration of how these molecules can be used for surface DNA computations is the creation of larger combinatorial sets using linked DNA word strings on a surface. Slight modifications to the standard procedures for the solid-phase synthesis of oligonucleotides can be used to create combinatorial sets of linked DNA words (5); however, a more complicated mark operation involving multiple surface hybridizations will be required to identify a particular DNA molecule (2). Additional surface fluorescence measurements will be used in the future to demonstrate the multiple marking of DNA word strings attached to chemically modified gold, glass and silicon surfaces.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the National Science Foundation in these studies (Grants CHE-9626607, CCR-9613799 and CCR-9628814).

REFERENCES

1 Adleman, L. M. (1994) *Science*, **226**, 1021–1024.

2 Cai, W., Condon, A. E., Corn, R. M., Glaser, E., Fei, Z., Frutos, T., Guo, Z., Lagally, M. G., Liu, Q., Smith, L. M. and Thiel, A. (1997) *Proceedings of the First Annual International Conference on Computational Molecular Biology (Recomb97).* ACM, pp. 67–74.

3 Liu, Q., Guo, Z., Condon, A. E., Corn, R. M., Lagally, M. G. and Smith, L. M. ‘A Surface-Based Approach to DNA Computation’; (1997) *Proceedings of the American Mathematical Society*, in press.

4 Liu, Q., Frutos, A. G., Thiel, A. J., Corn, R. M. and Smith, L. M. ‘DNA Computing on Surfaces: Encoding Information at the Single Base Level’, submitted for publication.

5 Smith, L. M., Condon, A. E., Frutos, A. G., Lagally, M. G., Liu, Q., Thiel, A. J. and Corn, R. M. ‘A Surface-based Approach to DNA Computation’, submitted for publication.

6 Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. and Davis, R. W. (1996) *Nature Genet.*, **14**, 450–456.

7 Garey, M. R. and Johnson, D. S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York.

8 Gray, J. M., Frutos, A. G., Berman, A. M., Condon, A. E., Lagally, M. G., Smith, L. M. and Corn, R. M. ‘Reducing Errors in DNA Computing by Appropriate Word Design’, in preparation.

9 Frey, B. L., Hanken, D. G. and Corn, R. M. (1993) *Langmuir*, **9**, 1815–1820.

10 Glen Research Corporation (1990) User Guide to DNA Modification and Labelling.

11 Frey, B. L. and Corn, R. M. (1996) *Anal. Chem.*, **68**, 3187–3193.

12 Jordan, C. E., Frey, B. L., Kornguth, S. and Corn, R. M. (1994) *Langmuir*, **10**, 3642–3648.

13 Jordan, C. E., Frutos, A. G., Thiel, A. J. and Corn, R. M. (1997) ‘Surface Plasmon Resonance Imaging Measurements of DNA Hybridization Adsorption and Streptavidin/DNA Multilayer Formation at Chemically Modified Gold Surfaces’, Anal. Chem. in press.

14 Naujok, R. R., Duevel, R. V. and Corn, R. M. (1993) *Langmuir*, **9**, 1771–1774.

15 Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. and Smith, L. M. (1994) *Nucleic Acids Res.*, **22**, 5456–5465.

16 Breslauer, K. J., Frank, R., Blöcker, H. and Marky, L. A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3746–3750.

17 Quartin, R. S. and Wetmur, J. G. (1989) *Biochemistry*, **28**, 1040–1047.

18 SantaLucia, J., Jr, Allawi, H. T. and Senevirante, P. A. (1996) *Biochemistry*, **35**, 3555–3562.

19 Wetmur, J. G. (1991) *Crit. Rev. Biochem. Mol. Biol.*, **26**, 227–259.

20 Wetmur, J. G. (1996) In Myers, R. A. (ed.), *Encyclopedia of Molecular Biology and Molecular Medicine*. VCH Press, New York, Vol. 4, pp. 235–243.

21 Morrison, L. E. and Stols, L. M. (1993) *Biochemistry*, **32**, 3095–3104.

22 Anderson, M. L. M. (1995) In Hames, B. D. and Higgins, S. J. (eds), *Gene Probes 2 A Practical Approach.* IRL Press, Oxford, p. 12.

23 Guo, Z., Liu, Q. and Smith, L. M. (1997) *Nature Biotechnol.*, **15**, 331–335.

24 Gillam, S., Waterman, K. and Smith, M. (1975) *Nucleic Acids Res.*, **2**, 625–634.

25 Persson, B., Stenhag, K., Nilsson, P., Larsson, A., Uhlén, M. and Nygren, P. (1997) *Anal. Biochem.*, **246**, 34–44.