ELSEVIER

BioSystems 52 (1999) 81–91

BioSystems

www.elsevier.com/locate/biosystems

State transitions by molecules

Kensaku Sakamoto [a], Daisuke Kiga [a], Ken Komiya [a], Hidetaka Gouzu [a], Shigeyuki Yokoyama [a], Shuji Ikeda [b], Hiroshi Sugiyama [b], Masami Hagiya [c,*]

[a] *Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Hongo 7-3-1 Bukyo-ku, Tokyo, 113-0033 Japan*

[b] *Institute for Medical and Dental Engineering, Tokyo Medical and Dental University, 2-3-10 Kanda Surugadai, Chiyoda-ku, Tokyo, 101-0062 Japan*

[c] *Department of Information Science, Graduate School of Science, University of Tokyo, Hongo 7-3-1 Bukyo-ku, Tokyo, 113-0033 Japan*

Abstract

In our previous paper, we described a method by which a state machine is implemented by a single-stranded DNA molecule whose 3′-end sequence encodes the current state of the machine. Successive state transitions are performed in such a way that the current state is annealed onto an appropriate portion of DNA encoding the transition table of the state machine and the next state is copied to the 3′-end by extension with polymerase. In this paper, we first show that combined with parallel overlap assembly, a single series of successive transitions can solve NP-complete problems. This means that the number of necessary laboratory steps is independent from the problem size. We then report the results of two experiments concerning the implementation of our method. One is on isothermal reactions which greatly increase the efficiency of state transitions compared with reactions controlled by thermal cycles. The other is on the use of unnatural bases for avoiding out-of-frame annealing. The latter result can also be applied to many DNA-based computing paradigms.

Keywords: Successive state transition; Hairpin structure; Whiplash PCR; DNA-Based computing; Isothermal reaction and reaction with thermal cycles

1. Introduction

In our previous paper (Hagiya et al., 1997), we described a method for implementing state machines by DNA molecules. In our method, which we now call successive localized polymerization[1], the transition table of a state machine is implemented by a single-stranded DNA molecule of the form

$$\text{stopper} - state'_1 - state_1 - \text{stopper} - state'_2 - state_2 - \cdots - \text{stopper} - state'_n - state_n,$$

* Corresponding author. Tel.: +81-3-38124177; fax: +81-3-38181073.

E-mail address: hagiya@is.s.u-tokyo.ac.jp (M. Hagiya)

[1] Winfree proposes to use the phrase whiplash PCR, coined by Adleman, for denoting intramolecular reactions employing polymerization stop. The phrase *successive localized polymerization* is intended to cover intermolecular reactions as well.

0303-2647/99/$ - see front matter
PII: S0303-2647(99)00035-0

where in each pair ($state'_i$–$state_i$) of states, $state_i$ denotes the state before a transition, and $state'_i$ the state after the transition. Each state is represented by an appropriate number of bases, called a state sequence. We currently use 15-mers in our experiments. To the left of each pair is inserted a sequence called a stopper sequence, designated as stopper in figures. Each sequence

$$\text{stopper} - state'_i - state_i$$

enables the transition from $state_i$ to $state'_i$.

State transitions are performed as in Fig. 1. The sequence at the 3′-end represents the current state of the state machine. It is the Watson–Crick complement of some state in the transition table. Assume that it is the complement of $state_i$. Under appropriate conditions, it is annealed to $state_i$ in the transition table by forming a hairpin structure (Fig. 1a). By DNA polymerase, the 3′-end is then extended and the complement of $state'_i$ becomes the new current state (Fig. 1b). The stopper sequence to the left of $state'_i$ stops the extension right after the complement of $state'_i$ is attached to the 3′-end. The implementation of stopper sequences will be described in the next paragraph. Assume that $state'_i$ is identical to $state_j$. After denaturation, the new current state is annealed to $state_j$ (Fig. 1c). The 3′-end is again extended by DNA polymerase and the complement of $state'_j$ becomes the next state. This process of state transitions can be repeated in

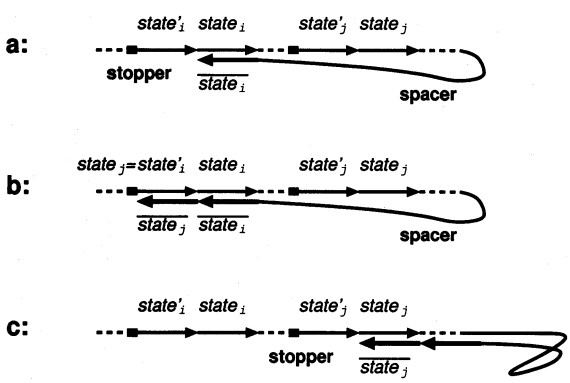

Fig. 1. State transitions. The current state is annealed onto the transition table by forming a hairpin structure (a). The current state is then extended and the next state is copied from the transition table (b). After denaturation, the new current state is annealed to another part of the transition table to enable the next transition (c).

a single tube by a simple thermal program consisting of thermal cycles for denaturation, annealing, and polymerization.

The polymerization buffer lacks one kind of base out of the common four. Assume that T is missing in the polymerization buffer. Then any repetition of A can serve as a stopper sequence because extension should stop when A is encountered. Of course, we cannot use A in our encoding of states, i.e. the representation of each state should consist of only T, G, and C. We currently use a triplet AAA as a stopper sequence in our experiments.

In this paper, we first argue that our method can greatly enhance the power of Adleman and Lipton's approach to DNA-computing (Adleman, 1994; Lipton, 1995). For solving instances of NP-complete problems, they first generate the space of candidate solutions in a tube, where each candidate is represented by a DNA molecule. Hybridization and ligation are employed for the generation of the candidate space; recently, the technique of parallel overlap assembly is also used (Ouyang et al., 1997). The candidate space is then explored by a number of laboratory steps that together implement the condition for a candidate to be a real solution.

Our method can be employed in this second step of exploring the candidate space and extracting the real solutions. NP-complete problems are solved by a single series of successive state transitions as described above. Since a series of state transitions can be considered as one big-step, this means that the number of laboratory steps needed to explore the candidate space is constant, i.e. $O(1)$. We will see that by adopting appropriate encoding, many NP-complete problems can be efficiently solved by the combination of parallel overlap assembly and a single series of state transitions, possibly with a fixed number applications of electrophoresis, PCR and affinity separation.

In his recent paper (Winfree, 1998), Winfree gives a framework for implementing GOTO programs by employing our method, and discusses how to solve NP-complete problems in his framework.

The molecular computing group at University of Memphis also reports their work on implementing state machines by DNA molecules (Garzon et al., 1997). In their methods, however, transition tables are encoded in input sequences

and therefore a single tube can only implement one state machine. In our method, on the other hand, more than one state machine can be executed in parallel in a single test tube.

As for the implementation of state transitions by polymerization, the work by Guarnieri et al. (1996) which makes DNA add, is related to ours, because polymerization appends a new state and a new digit to the current state of addition. There are a number of differences between their work and ours. Although they call their method horizontal chain reaction, it is not at all obvious how to perform successive reactions in a single test tube. This is a theme of the experimental part of this paper. While they only consider intermolecular reactions, we are basically concerned with intramolecular ones. Finally, since we use the technique of polymerization stop, all the templates for polymerization can be placed on one DNA molecule.

In the experimental section of this report, we describe improvement of the thermal program for two successive transitions, and use of unnatural bases for guiding appropriate priming of polymerization.

A difficulty with the thermal program reported previously (Hagiya et al., 1997) was that this program can only inefficiently implement the successive transitions. We inferred that this difficulty comes from the state machine taking two intramolecularly annealed forms after the first transition. One form takes place just after this transition, and in this form, one pair of state sequences are annealed onto the transition table, and no further extension of the DNA can occur (Fig. 1b). In the other form, only the 'current state' sequence is annealed onto the transition table, and primes polymerization for the second transition (Fig. 1c). At the annealing step, formations of these two types of the hairpins compete against each other. Naturally, the hairpin with the longer stem is formed preferentially. This means that there is little chance for the second transition to occur. In order to solve this difficulty, we performed the successive transitions in an isothermal reaction. At a fixed temperature, an equilibrium is reached between these two hairpin forms. Since the 'shorter-stem' hairpin can undergo further reaction, addition of polymerase leads to the decrease in the concentration of this hairpin, which pushes the DNA towards its formation. Thus, conversion of the 'longer-stem' hairpin to the less stable 'shorter-stem' hairpin can be achieved in the isothermal reaction.

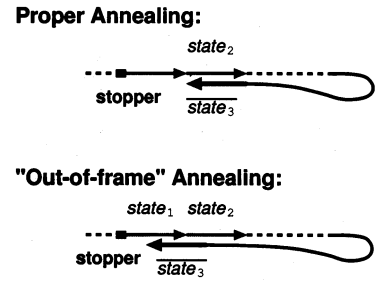

Fig. 2. Proper annealing and out-of-frame annealing. The upper panel shows proper annealing where the current state is properly annealed onto a transition table. The lower panel shows out-of-frame annealing where flanking parts of two juxtaposed sequences constitute the sequence complementary to the current state.

The second experimental part is concerning a technique by which the priming of polymerization can be properly guided. Any state sequence is not allowed to form base pairs with some residues from a state sequence and those from another state sequence simultaneously (Fig. 2); such 'out-of-frame' annealing makes the implementation meaningless. This 'out-of-frame' annealing will occur if flanking parts of two juxtaposed sequences constitute the sequence complementary to a third state sequence. In our scheme of transition, annealing of a state sequence onto the transition table primes polymerization. In order to prevent inappropriate priming caused by 'out-of-frame' annealing, we introduced isoguanosines (iG) at both side of a state sequence and 5-methylisocytidines (iC) at the corresponding positions in the transition table. These unnatural nucleotides are specific to each other, and do not form a base pair with any of the four common nucleotides, except that the minor enol form of isoguanosine can form a base pair with T (Switzer et al., 1989). If the transition table has no T residues, the state sequence has a matched base pair at its 3′-end and thus can prime polymerization only in the case of proper annealing.

2. Solving NP-complete problems

We first prepare a set S of states. Each state that appears in state transitions is taken from this set. A pair $\langle s', s \rangle$ of states before and after a transition is called a state-pair. We assume that a stopper sequence is always inserted to the left of a state-pair when a state-pair is represented in a DNA molecule.

$$\text{stopper} - s' - s$$

A finite set $\{\langle s'_1, s_1\rangle, \langle s'_2, s_2\rangle, \cdots, \langle s'_k, s_k\rangle\}$ of state-pairs is then called a fragment. A fragment is represented by the concatenation of state-pairs in the fragment.

$$\text{stopper} - s'_1 - s_1 - \text{stopper} - s'_2 - s_2 - \cdots$$

$$- \text{stopper} - s'_k - s_k$$

The order among state-pairs is irrelevant. A fragment can be empty; it is called an empty fragment.

Let F_i^+ and F_i^- be fragments for $1 \leqslant i \leqslant n$. By parallel overlap assembly, it is possible to generate a random library of sequences of the form

$$b_0 - F_1^{s_1} - b_1 - F_2^{s_2} - b_2 - \cdots - b_{n-1} - F_n^{s_n} - b_n,$$

where s_i is either $+$ or $-$, and b_i is a bridge sequence for connecting fragments $F_i^{s_i}$ and $F_{i+1}^{s_{i+1}}$. Parallel overlap assembly is accomplished by thermal cycles on the mixture of $b_{i-1} - F_i^{s_i} - b_i$ for odd i and the Watson–Crick complement of $b_{i-1} - F_i^{s_i} - b_i$ for even i. Ignoring bridge sequences, the resulting sequences can be considered as transition tables.

2.1. CNF-SAT

The satisfiability problem of a conjunctive normal form is solved as follows. Let $C_1 \wedge C_2 \wedge \cdots \wedge C_k$ be a given conjunctive normal form which contains variables $x_1, x_2, \cdots, x_n$. Let the state set S be $\{1, 2, \cdots, k, k+1\}$. For $1 \leqslant i \leqslant n$, let F_i^+ be the fragment consisting of all state-pairs $\langle j+1, j \rangle$ such that the literal x_i occurs in clause C_j. And let F_i^- be the fragment consisting of all state-pairs $\langle j+1, j \rangle$ such that the negation of x_i occurs in clause C_j. For example, consider the following conjunctive normal form.

$$(x_1 \vee x_2 \vee \neg x_5) \wedge (\neg x_3 \vee x_4 \vee x_5) \wedge (\neg x_1 \vee x_3 \vee x_6)$$

$$\wedge (x_2 \vee \neg x_4 \vee \neg x_6)$$

There are four clauses:

$$C_1 = x_1 \vee x_2 \vee \neg x_5$$

$$C_2 = \neg x_3 \vee x_4 \vee x_5$$

$$C_3 = \neg x_1 \vee x_3 \vee x_6$$

$$C_4 = x_2 \vee \neg x_4 \vee \neg x_6,$$

and six variables: $x_1, x_2, \cdots, x_6$. Since the literal x_1 appears in C_1, $F_1^+ = \{\langle 2, 1\rangle\}$. On the other hand, since the literal $\neg x_1$ appears in C_3, $F_1^- = \{\langle 4, 3\rangle\}$. As for x_2, since x_2 appears in C_1 and C_4, $F_2^+ = \{\langle 2, 1\rangle, \langle 5, 4\rangle\}$. On the other hand, since $\neg x_2$ appears in no clause, F_2^- is the empty fragment.

By parallel overlap assembly, we generate sequences of the form

$$b_0 - F_1^{s_1} - b_1 - F_2^{s_2} - b_2 - \cdots - b_{n-1} - F_n^{s_n} - b_n.$$

We then add state 1 as the initial state by attaching the Watson–Crick complement of the representation of state 1 with a spacer sequence of an appropriate length.

$$b_0 - F_1^{s_1} - b_1 - F_2^{s_2} - b_2 - \cdots - b_{n-1} - F_n^{s_n} - b_n$$

$$- \text{spacer} - \overline{1}$$

Notice that since the state-pair $\langle j+1, j \rangle$ may appear in several fragments $F_i^{\pm}$, the final DNA sequence may contain several copies of $\langle j+1, j \rangle$. In fact, it appears once in every fragment corresponding to a literal in clause C_j. Therefore, it is easy to see that those sequences that allow state transitions from state 1 to state $k+1$ correspond to those that satisfy the given conjunctive normal form. The transition from state j to state $j+1$ is possible in a sequence if and only if the clause C_j is true in the sequence, where occurrence of the fragment F_i^+ in the sequence means to set x_i to true while F_i^- means false. For example, the sequence

$$b_0 - F_1^- - b_1 - F_2^+ - b_2 - F_3^- - b_3 - F_4^+ - b_4 - F_5^+$$

$$- b_5 - F_6^+ - b_6$$

in the above example allows transitions from state 1 to state 5. In this sequence, x_1 is false, x_2 is true, x_3 is false, etc.

Those sequences that allow transitions from state 1 to state $k+1$ can be extracted, for instance, by PCR with b_0 and state $k+1$ as primers, or by affinity separation with state $k+1$.

2.2. Vertex cover

Let $G = \langle V, E \rangle$ be an undirected graph, where

$$V = \{v_1, v_2, \ldots, v_n\}$$

is the set of vertices and

$$E = \{e_1, e_2, \ldots, e_k\}$$

is the set of edges. A vertex cover is a subset C of V such that each edge in E is incident to some vertex in C. The vertex cover problem is that of finding the minimum size of a vertex cover.

Let the state set S be $\{1, 2, \cdots, k, k+1\}$. For $1 \leqslant i \leqslant n$, let F_i^+ be the fragment consisting of all state-pairs $\langle j+1, j \rangle$ such that v_i is incident to e_j. By inserting a noncoding sequence of an appropriate length, we make the representation of F_i^+ have the same length for each i. On the other hand, F_i^- is defined to be the empty fragment for $1 \leqslant i \leqslant n$. We then add state 1 as the initial state to each sequence generated by parallel overlap assembly.

The state-pair $\langle j+1, j \rangle$ appears either twice, once, or not at all, depending upon whether both, one, or neither of its incident vertices is in the vertex set defined by the DNA. Therefore, if a sequence allows state transitions from state 1 to state $k+1$, it corresponds to a vertex cover that contains vertex v_i if and only if F_i^+ is in the sequence. By electrophoresis, it is possible to measure the minimum length of such sequences, which gives the minimum size of vertex covers.

2.3. Direct sum cover

Let U be a finite set and $\mathscr{F} = \{X_1, X_2, \cdots, X_n\}$ be a family of subsets of U. A direct sum cover is a subfamily $\mathscr{C}$ of $\mathscr{F}$ such that any pair of subsets in $\mathscr{C}$ is disjoint and the union of all the subsets in $\mathscr{C}$ is identical to U. The direct sum cover problem is that of finding a direct sum cover.

For simplicity, we assume $U = \{1, 2, \cdots, k\}$. Let the state set S be $\{1, 2, \cdots, k, k+1\}$. For $1 \leqslant i \leqslant n$, let F_i^+ be the fragment consisting of all state-pairs $\langle j+1, j \rangle$ such that j is in subset X_i. F_i^- is defined to be the empty fragment. We then add state 1 as the initial state to each sequence generated by parallel overlap assembly.

If a sequence allows state transitions from state 1 to state $k+1$, it corresponds to a cover of U that contains subset X_i if and only if F_i^+ is in the sequence. Since the length of F_i^+ is proportional to the size of X_i, we know the length of the sequence that corresponds to a direct sum cover. Therefore, by electrophoresis, it is possible to check whether a direct sum cover exists or not.

2.4. Hamiltonian path

Needless to say, it is possible to solve the directed hamiltonian path problem (DHPP) as in the last subsections by generating paths in a directed graph instead of generating sequences by simple parallel overlap assembly. This means to randomly generate paths in a directed graph by Adleman's method (Adleman, 1994).

Let $G = (V, E)$ be a directed graph, where

$$V = \{v_1, v_2, \cdots, v_k\}$$

Let the state set S be $\{1, 2, \cdots, k, k+1\}$. The representation of vertex v_j is assumed to contain the state-pair $\langle j+1, j \rangle$. Then a path containing all the vertices allows transitions from state 1 to state $k+1$. By electrophoresis, it is possible to check whether there exists a path whose length is the same as that of Hamiltonian paths.

In his recent paper (Winfree, 1998), Winfree gives a method for solving the DHPP in his framework of GOTO programs. His method, though it requires a preprocessing step, is more efficient than ours because it only generates paths whose length is equal to that of a Hamiltonian path.

3. Experiments

3.1. Materials and methods

The sequences of the oligodeoxyribonucleotides used in this paper are listed in Fig. 3. These

Tran1	CCAAATCCCGTTCTGGGCCTGCACGATCTAGGAAAGTGGTTTGCGCTCGTTCCCGTTCTGGGCCTTTTCCTAGATCGTGC
iC1	CCC AAG CTT GAG TGC GCA CGA TCT AGG iCTC CCG TTC TGG GCC TiCAAA
iC2	CCC AAG CTT GAG TGC GCA CGA TCT AGG iCiCTC CCG TTC TGG GCC TiCiC AAA
C1	CCC AAG CTT GAG TGC GCA CGA TCT AGG CTC CCG TTC TGG GCC TCAAA
RiG1	CCC GAA TTC GAG GAA AGC GCA CGA iGVV VVV VVV VVV VVV ViG (V=A, C, G)
RiG2	CCC GAA TTC GAG GAA AGC GCA CGA iGiGVV VVV VVV VVV VVV ViGiG (V=A, C, G)
RG1	CCC GAA TTC GAG GAA AGC GCA CGA GVV VVV VVV VVV VVV VG (V=A, C, G)
primer1	CCC AAG CTT GAG TGC GCA CG
primer2	CCC GAA TTC GAG GAA AGC GC

Fig. 3. Oligomers used in the experiments. Oligomer Tran1 is used in the first experiment on isothermal transitions. Other oligomers are used in the second experiment on the use of unnatural bases for avoiding mispairing.

oligomers, except for the iG/iC-containing oligomers, were commercially synthesized by Amersham Pharmacia biotech Co., Ltd. (Tokyo). Oligomers iC1, iC2, RiG1, and RiG2 were chemically synthesized (Sugiyama et al., 1996). Two successive transitions were performed in a reaction mixture containing oligomer Tran1 (3 μM), 2 mM $MgCl_2$, dATP, dGTP, and dCTP (250 μM each), 2 units of AmpliTaq DNA polymerase (Perkin–Elmer), and the PCR buffer provided by the supplier. Thermal cycler 480 (Perkin–Elmer) was used for incubation of the reaction mixture, and for PCR described below.

Similar conditions were employed for the model reactions of a single transition, except for the addition of dTTP, and use of RiG1, RiG2 or RG1 (1.6 μg) and iC1, iC2, or C1 (1.4 g), instead of Tran1. These model reactions were performed in a 25 μl scale at 65°C for 5 min, and an aliquot (1 μl) was subjected to amplification by PCR with primers 1 and 2. The thermal program for PCR comprises 30 cycles of the denaturation at 94°C for 0.5 min, the annealing at 65°C for 0.5 min, and the polymerization at 72°C for 0.5 min.

3.2. Results

3.2.1. Improvement of the thermal program for two successive transitions

In the previous paper, we employed a thermal cycle in order to perform successive transitions with oligomer Tran1, which is of the form ini-s2-s1-s3-s2-$\overline{\mathbf{s1}}$ and consists of subsequences defined in Fig. 4. The expected extension is shown in Fig. 5. Each thermal cycle comprises three steps of denaturation (90°C, 1 min), annealing (on ice, 1 min, and then 40°C, 30 s), and polymerization (68°C, 30 s). It was found that even after the fourth

s1	GCACGATCTAGGAAA
s2	TCCCGTTCTGGGCCT
s3	GTGGTTTGCGCTCGT
ini	CCAAA

Fig. 4. DNA sequences comprising Tran1. Oligomer Tran1 is of the form ini-s2-s1-s3-s2-$\overline{\mathbf{s1}}$, where $\overline{\mathbf{s1}}$ denotes the Watson–Crick complement of s1.

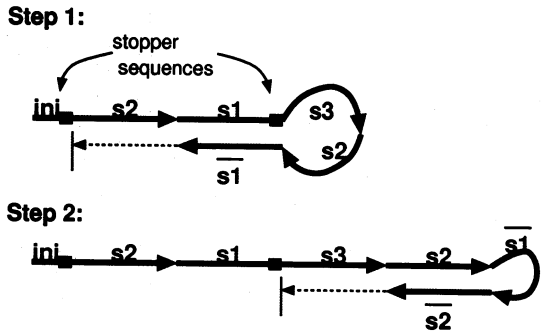

Fig. 5. Expected extension of Tran1. Oligomer Tran1 was designed so that two successive transitions are possible.

cycle, a significant population of Tran1 did not complete the successive transitions yet. On the other hand, even after the first cycle, a fractional amount of Tran1 completed the two successive transitions. Note here that these Tran1 oligomers underwent extension twice, although the reaction mixture was exposed to the temperature for polymerization only once. We inferred that at this polymerization step, not only the first extension was completed, but the denaturation of the extended DNA and then the second extension also occurred. In order to test this possibility, the reaction mixture was exposed first to the denaturation (90°C, 1 min) and annealing (on ice, 1 min) steps and then to various temperatures for 5 min (Fig. 6).

Fig. 6 indicates that significant amounts of the product of the successive transitions were obtained

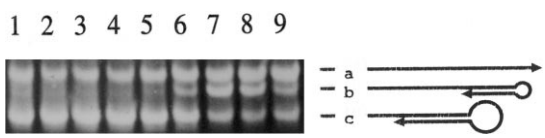

Fig. 6. Two successive transitions performed under various temperatures. The products of the transition reactions were analyzed on a denaturing polyacrylamide gel. The reaction temperature was, 56°C (lane 1); 60°C (lane 2); 64°C (lane 3); 68°C (lane 4); 72°C (lane 5); 76°C (lane 6); 80°C (lane 7); 84°C (lane 8); and 88°C (lane 9). The bands corresponding to Tran1 (a); the product of the first transition (c); and that of the second transition (b) were assigned as described previously (Hagiya et al., 1997). The products of the first and second transitions moved faster than the Tran1 oligomer on the gel, suggesting that they formed hairpin structures (Hagiya et al., 1997). Also in DNA sequence analyses, formation of secondary structures by ssDNA in urea-containing gels has been observed (Deininger 1983).

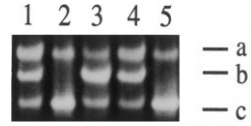

Fig. 7. Two successive transitions performed at 64°C and/or 80°C. The products of the transition reactions were analyzed on a denaturing polyacrylamide gel. The reaction mixture was incubated, after the denaturation and annealing steps, at 80°C for 5 min (lane 1); at 64°C for 5 min (lane 2); at 64°C for 5 min six times, each time followed by the incubation at 80°C for 5 min (lane 3); at 80°C for 30 min (lane 4), and at 64°C for 30 min (lane 5). The bands a, b, and c correspond to Tran1 (a); the product of the first transition (c); and that of the second transition (b).

at 76, 80, 84 and 88°C; the largest yield was obtained at 80°C. The optimum temperatures for the first and second transitions appear to differ from each other. By judging from the least amount of Tran1 that did not undergo the first transition yet, the optimum temperature for the first one was 64°C (Fig. 6, lane 3). Similarly, judging from the least amount of Tran1 that underwent the first transition but not the second one yet, the optimum temperature for the second one was 80°C (Fig. 6, lane 7). Then, we performed the reaction at 64°C and/or 80°C.

Incubation at 64°C for 5 min was found to be enough for almost all the Tran1 oligomers to complete the first transition, although the second transition did not occur at all (Fig. 7, lane 2). It took 30 min at 80°C for a major population of the DNA to complete the two successive transitions (Fig. 7, lane 4). Incubation at 64°C for 5 min and then at 80°C for 5 min was repeated six times; the yield was found to be the maximum (Fig. 7, lane 3). Furthermore, the incubation at 64°C for 5 min and then at 80°C for 5 min was done only once, following the denaturation at 94°C for 1 min the yield was almost the same as that from the six-times repeat (data not shown). Thus, we found this thermal program (94°C, 1 min; 64°C, 5 min; 80°C, 5 min) to be remarkably efficient for carrying out the two successive transitions.

3.2.2. Introduction of unnatural bases at the boundaries of the state sequences

In order to examine the availability of unnatural bases for guiding proper priming by the state

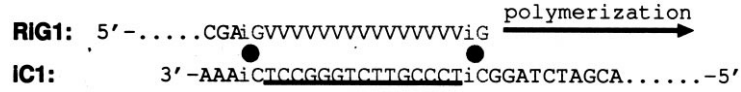

Fig. 8. Annealing between RiG1 and iC1 guided by the iC–iG base pairs. The 3′-regions of each oligomer are shown. The defined state sequence on the iC1 strand is underlined. V represents A, G, or C.

sequences, we designed a model reaction of a single 'intermolecular' transition. In this model reaction, an iC-containing oligodeoxyribonucleotide, iC1, and an iG-containing oligomer, RiG1, were annealed onto each other, priming the extension of RiG1 with *Taq* DNA polymerase by use of iC1 as a template (Fig. 8). Oligomer iC1 with a defined state sequence of 15 nucleotides, encoded with C, G, and T, has iC residues at both ends of this sequence; these iC residues frame the state sequence. Similarly, a 15-nucleotides random sequence in RiG1, comprising only A, C, and G, is framed by iG residues at both ends of this random region. Annealing between the defined sequence and the random region was to be guided by base pairings between these 'framing' iG and iC residues. The extended RiG1 was then amplified by PCR (primer1 and primer2). Furthermore, similar experiments were performed with iC2 and RiG2, which contain doublets iCiC and iGiG in place of singlets iC and iG, respectively, and with C1 and RG1, which contain C and G in place of iC and iG, respectively.

The expected length of PCR product in the case of annealing properly guided by iC–iG/C–G base pairs was 68-mer. The reaction with iC1 and RiG1, after amplification by PCR, yielded a major band of the corresponding length, together with minor bands of the significantly shorter lengths (Fig. 9, lanes 1–2). These minor bands appeared even when iC1 was not included in the reaction mixture (data not shown), and were supposed to be produced due to annealing of primer1 onto the random region RiG. Similar minor bands were detected also in the other two reactions (Fig. 9, lanes 3–6). The reaction with iC2 and RiG2 yielded doublet bands, one of which was of the expected length (Fig. 9, lanes 3–4). The yield of these bands were much lower as compared with that of the product with RiG1 and iC1. The third model reaction included RG1 and C1 instead of iG/iC-containing oligomers, and produced a broad band together with some minor bands (Fig. 9, lanes 5–6). These products were subjected to sequence analysis.

The result of sequence analysis is summarized in Fig. 10. All of these sequences were derived from the RiG1/RG1 strand rather than the iC1/C1 strand, because none of them have the same sequence as the defined state sequence. According to the sequence design for RiG1/RG1, its defined sequence, 5′-···ACGA-3′, should be followed by the sequences (designated as RDS) derived from the two framing bases and the 15-mer random sequence, and the 3′-framing base. In the sequences obtained with RiG1 and iC1 (1–1 to 1–8), iG were found to be replaced by A, with two exceptions where the nucleotides corresponding to the 3′-framing iG residues are missing (sequences 1–1 and 1–2). The 3′-halves of the listed RDS sequences were found to reflect the iC1/C1 sequences, which indicates 7–8 base pairs are enough for priming the polymerization under

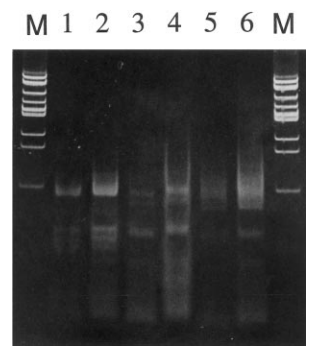

Fig. 9. Amplified products from the single-transition model reactions with RiG1 and iC1 (lanes 1–2), and with RiG2 and iC2 (lanes 3–4), and with RG1 and C1 (lanes 5–6). An aliquot of each PCR mixture was loaded on the gel (0.5 μl for lanes 1 and 5; 2 μl for lane 3), together with its 4-fold volume in the next lane (lanes 2, 4, and 6). Some known marker was loaded (lanes M).

1-1	AGCGCACGA	AAGCGGGAGGACGGGA	CCTAGATCGTGCGCACTCAA	(14)
1-2	AGCGCACGA	AACGGGGAACGGGA	CCTAGATCGTGCGCACTCAA	(12)
1-3	AGCGCACGA	AAGGCACGGAGCGGGAA	CCTAGATCGTGCGCACTCAA	(15)
1-4	AGCGCACGA	AGGCGCCGGAACGGAA	CCTAGATCGTGCGCACTCAA	(14)
1-5	AGCGCACGA	AGGCCAAGCACGGGAA	CCTAGATCGTGCGCACTCAA	(14)
1-6	AGCGCACGA	AAGGACCAAAGCGGGAA	CCTAGATCGTGCGCACTCAA	(15)
1-7	AGCGCACGA	AGCCCCCAGGACGGGAA	CCTAGATCGTGCGCACTCAA	(15)
1-8	AGCGCACGA	AGGGACCACAGCGGGAA	CCTAGATCGTGCGCACTCAA	(15)
3-1	AGCGCACGA	GGGGGCAGAACGGAAG	CCTAGATCGTGCGCACTCAA	(14)
3-2	AGCGCACGA	GCGAGCAAGAACGGGAG	CCTAGATCGTGCGCACTCAA	(15)
3-3	AGCGCACGA	GGCAAGGGCCGAACGGGAG	CCTAGATCGTGCGCACTCAA	(17)
3-4	AGCGCACGA	GGGGGGCAGAACGGGAG	CCTAGATCGTGCGCACTCAA	(15)
3-5	AGCGCACGA	GCCGGGCGCCAGAACGGGAG	CCTAGATCGTGCGCACTCAA	(18)
3-6	AGCGCACGA	GGAAGGCCGGAACGGGGG	CCTAGATCGTGCGCACTCAA	(16)
3-7	AGCGCACGA	GAGGCGCAGGACGGGAG	CCTAGATCGTGCGCACTCAA	(15)
3-8	AGCGCACGA	GCGGGGACAGACCGGGAG	CCTAGATCGTGCGCACTCAA	(16)

Fig. 10. Sequences of the products from the single-transition model reactions with oligomers RiG1 and iC1 (1–1 to 1–8), and with oligomers RG1 and C1 (3–1 to 3–8). The RDS sequences are shown in the middle. The residues in the RDS that have their counterparts in iC1/C1 are underlined. The numbers of the residues in the RDS is shown in the parentheses.

the employed conditions. Some sequences have the RDS of more or less than 17 nucleotides. The short length for the RDSs of sequences 1–1, 1–2, 1–4, 1–5, and 3-1 may be due to deletion during PCR. In contrast, the lengths of 16, 17, and 18 nucleotides for the RDS were found only for the sequences obtained with RG1 and C1. It is likely that in these cases, the 3′-framing G of RG1 formed a base pair with the residues one to three bases before the 5′-framing C of C1.

3.3. Discussion

In the present study, we improved the thermal program for the two successive transitions, and thus, a remarkably efficient implementation of the two successive transitions was achieved. Isothermal reaction is the key feature of the new thermal program, which consists of only two incubation steps at 64 and 80°C. The temperature of 64°C supports both the annealing and polymerization for the first transition, while the denaturation, annealing, and polymerization for the second transition occur at 80°C. Thus, this thermal program is different from those including thermal cycles, such as PCR, where in each cycle, denaturation, annealing, and polymerization are carried out at their specific temperatures.

For application of the state transitions, we assume that the state machine DNA forms a hairpin, and transitions occur in an intramolecular manner rather than intermolecular ones. After the first and second transitions, the extended Tran1 DNAs were supposed to take a hairpin form, because they moved faster in the polyacrylamide gel than the Tran1 of the initial length. Nevertheless, this observation did not immediately show that these transitions occurred as intermolecular reactions. Our plan to guarantee independent implementation for each state machine DNA molecule is described in Section 4.

In the second experiment, out of three model reactions of intermolecular transition, only the reaction with RiG1 and iC1 yielded a single major product with no confusing by-products. The quality of each oligomer was much the same (data not shown), and therefore the broad band produced with RG1 and C1 was not due to incomplete purification of the oligomers. Note here that any DNA extends from its 3′-end after annealing onto a template. Therefore, base pairing by the 3′-end residue of the DNA is necessary for starting the extension. Since iG at the 3′-end of RiG1 can form a base pair only with the framing iC residues in iC1, only the proper annealing can prime the extension, for the set of RiG1 and iC1. Thus, the use of iG and iC as the framing bases is especially effective for the case of implementation of the state transitions.

The sequence analysis revealed that only the 7–8 residues at the 3′-end of RiG1 formed base pairs with iC1. Therefore, state sequences of 7–8 residues are probably enough for priming the extension of the oligomer under the employed conditions. The substitution of iG to A indicates that PCR amplification of the extended RiG1 was due to misincorporation of T in respond to iG. This explains the lower yield in the amplification for the extended RiG2, which contains two iGiG doublets. Out of the obtained sequences, no obvi-

ous case of inappropriate priming by 'out-of-frame' annealing was found for RiG1 and iC1, while half of the sequences for RG1 and C1 were supposed to derive from inappropriate priming.

The availability of unnatural bases as demonstrated here may contribute to various implementations with DNA molecules in the following three points. The first one is avoiding inappropriate priming caused by 'out-of-frame' annealing. Although careful sequence design also may be useful for the same purpose, this necessarily imposes some restrictions on selection of sequence. The second is increasing the number of the nucleotide bases available for encoding states, variables, and so on. The availability of unnatural bases has been proposed and demonstrated in this type of application (Piccirilli et al., 1990; Bain et al., 1992). The third is preventing secondary-structure and/or dimer formations by single-stranded DNA. Imagine an ssDNA molecule consisting of A, C, and iG. No proper base pair will not be formed among these three nucleotides. The feasibility of this application of iG and iC is a subject of further study.

4. Concluding remark

In this paper, we first showed that NP-complete problems can be solved by a single series of successive transitions, combined with parallel overlap assembly and some other operations.

However, more time should be spent on successive transitions as more transitions are required. In the first experiment, we reported that isothermal reactions greatly increased the efficiency of state transitions compared with reactions controlled by thermal cycles.

In the second experiment, we examined the use of unnatural bases for avoiding out-of-frame annealing. This result is expected to be applied to many DNA-based computing paradigms.

We did not touch on the issue of enforcing intramolecular reactions in this paper. In order for a DNA molecule to behave as an independent state machine, it should not interact with other DNA molecules. We are currently employing streptavidin-coated magnetic beads on which biotin-inserted DNA molecules are bound to avoid intermolecular reactions. Results of this experiment will be reported in the near future.

Acknowledgements

We deeply thank James Wetmur for his suggestion on isothermal reactions and Erik Winfree for his comments on the earlier draft of the paper. We also thank Akio Nishikawa for carefully reading the final version of the paper. This work is supported by the Japan Society for the Promotion of Science 'Research for the Future' Program (JSPS-RFTF 96IOO101).

References

Adleman, M., 1994. Molecular computation of solution to combinatorial problems. Science 266, 1021–1024.

Bain, J.D., Switzer, C., Chamberlin, A.R., Benner, S.A., 1992. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. Nature 356, 537–539.

Deininger, P.L., 1983. Approaches to rapid DNA sequence analysis. Anal. Biochem. 135, 247–263.

Garzon, M., Gao, Y., Rose, J.A., Murphy, R.C., Deaton, R., Franceschetti, D.R., Stevens Jr, S.E., 1997. In vitro implementation of finite-state machines, Proceedings of the Workshop on Implementing Automata (WIA'97), Springer LNCS, in press.

Guarnieri, F., Fliss, M., Bancroft, C., 1996. Making DNA add. Science 273, 220–223.

Hagiya, M., Arita, M., Kiga, D., Sakamoto, K., Yokoyama, S., Towards parallel evaluation and learning of boolean μ-formulas with molecules, Preliminary Proceedings of the Third DIMACS Workshop on DNA Based Computers, June 23–25, 1997, University of Pennsylvania, pp. 105–114.

Lipton, R.J., 1995. DNA solution of hard computational problems. Science 268, 542–545.

Ouyang, Q., Kaplan, P.D., Liu, S., Libchaber, A., 1997. DNA solution of the maximal clique problems. Science 278, 446–449.

Piccirilli, J.A., Krauch, T., Moroney, S.E., Benner, S.A., 1990. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. Nature 343, 33–37.

Sugiyama, H., Ikeda, S., Saito, I., 1996. Remarkably stable parallel-stranded oligonucleotides containing 5-methylisocytosine and isoguanine. J. Am. Chem. Soc. 118, 9994–9995.

Switzer, C., Moroney, S.E., Benner, S.A., 1989. Enzymatic incorporation of a new base pair into DNA and RNA. J. Am. Chem. Soc. 111, 8322–8383.

Winfree, E. Whiplash PCR for $O(1)$ computing, Fourth DIMACS Workshop on DNA Based Computers, 1998, submitted for publication.