Equilibrium analysis of the efficiency of an autonomous molecular computer

John A. Rose*
Institute of Physics, The University of Tokyo, Tokyo 153-8902, Japan

Russell J. Deaton†
Department of Computer Science and Computer Engineering, The University of Arkansas, Fayetteville, Arkansas 72701

Masami Hagiya‡
Department of Computer Science, The University of Tokyo, Tokyo 113-0033, Japan

Akira Suyama§
Institute of Physics, The University of Tokyo, Tokyo 153-8902, Japan

(Received 1 November 2000; revised manuscript received 25 September 2001; published 25 January 2002)

In the whiplash polymerase chain reaction (WPCR), autonomous molecular computation is implemented *in vitro* by the recursive, self-directed polymerase extension of a mixture of DNA hairpins. Although computational efficiency is known to be reduced by a tendency for DNAs to self-inhibit by backhybridization, both the magnitude of this effect and its dependence on the reaction conditions have remained open questions. In this paper, the impact of backhybridization on WPCR efficiency is addressed by modeling the recursive extension of each strand as a Markov chain. The extension efficiency per effective polymerase-DNA encounter is then estimated within the framework of a statistical thermodynamic model. Model predictions are shown to provide close agreement with the premature halting of computation reported in a recent *in vitro* WPCR implementation, a particularly significant result, given that backhybridization had been discounted as the dominant error process. The scaling behavior further indicates completion times to be sufficiently long to render WPCR-based massive parallelism infeasible. A modified architecture, PNA-mediated WPCR (PWPCR) is then proposed in which the occupancy of backhybridized hairpins is reduced by targeted PNA_2/DNA triplex formation. The efficiency of PWPCR is discussed using a modified form of the model developed for WPCR. Predictions indicate the PWPCR efficiency is sufficient to allow the implementation of autonomous molecular computation on a massive scale.

DOI: 10.1103/PhysRevE.65.021910 PACS number(s): 87.15.Aa, 89.20.Ff, 87.14.Gg, 07.10.Cm

I. INTRODUCTION

The whiplash polymerase chain reaction (WPCR) was proposed in 1997 by Hagiya, *et al.* [1], as an alternative to methods of DNA computing based on a *generate and search* strategy, such as those of Adleman [2] and Lipton [3]. In *generate and search*, a combinatorial library of DNA molecules is generated, each of which encodes a candidate solution to the computational problem of interest. This library is then screened for a molecule that encodes a correct solution, by the parallel application of a sequence of standard operations from biotechnology. While such a strategy efficiently exploits the data parallelism inherent in molecular computing, the uniform application of a single set of biological steps to all strands restricts implementation to computations that execute a single set of instructions on a single input set.

In WPCR molecular computation is implemented by the recursive polymerase extension of a set of encoded single-stranded DNA (ssDNA) molecules. As each member of the set of extension processes executed by a given ssDNA is directed by the strand's base sequence, complex data structures and/or multistep computational procedures may be encoded into each strand. As a result, WPCR has the potential to implement the parallel execution of multiple programs with different inputs, a significant improvement over the Adleman-Lipton paradigm.

The feasibility of executing single [1] and successive [4] state transitions by using WPCR has been established experimentally. A recent attempt to implement a longer path, however, reported failure after only a few transitions [5]. Although WPCR is known to be susceptible to backhybridization, a systematic form of self-inhibition [1], it was assumed in Ref. [5] that problems due to backhybridization had been overcome by the application of an experimentally optimized thermal protocol, and the observed efficiency problem was attributed to other error processes.

The primary focus of this paper is the development of a theoretical model of the impact of backhybridization on WPCR efficiency. This model is then used to establish backhybridization as the principal factor responsible for the computational failure reported in Ref. [5], and to support the rational redesign of WPCR to reduce the impact of backhy-

*Electronic address: johnrose@genta.c.u-tokyo.ac.jp; URL:http://hagi.is.s.u-tokyo.ac.jp/johnrose
†Electronic address: rdeaton@uark.edu; URL:http://csce.uark.edu/~rdeaton/
‡Electronic address: hagiya@is.s.u-tokyo.ac.jp; URL:http://hagi.is.s.u-tokyo.ac.jp/
§Electronic address: suyama@dna.c.u-tokyo.ac.jp; URL:http://talent.c.u-tokyo.ac.jp/suyama/

bridization. The organization is as follows. In Sec. II, the WPCR architecture is described, and backhybridization is discussed conceptually. In Sec. III, the length distribution of extended strands, as a function of polymerization time and reaction conditions is examined by modeling the recursive extension of each strand as a Markov chain. The extension efficiency per polymerase-DNA encounter is then discussed using a statistical thermodynamic model of DNA hybridization. Model predictions are shown to provide close agreement with experimentally observed behavior [5]. Based on the scaling behavior predicted by the model, completion times are long enough to render WPCR-based massive parallelism infeasible. In Sec. IV, a modified architecture, PNA-mediated WPCR (PWPCR) is proposed in which backhybridization is inhibited by targeted PNA_2/DNA triplex formation. The efficiency of PWPCR is discussed using a modified form of the model developed for WPCR. Model predictions indicate that targeted triplex formation is accompanied by a large increase in efficiency, which is sufficient to support the implementation of autonomous molecular computation on a massive scale.

II. THE WHIPLASH POLYMERASE CHAIN REACTION

A. The basic architecture

In WPCR [1], state transitions are implemented by the template-directed, recursive polymerase extension of a mixture of DNA hairpins. For clarity, this process is described in the context of the WPCR implementation of the simple three-state, two-step computational path, $a \rightarrow b \rightarrow c$. Prior to the beginning of the computation, a ssDNA molecule is encoded with three distinctive regions, as shown in Fig. 1 (i). The *transition rule region* encodes the computation's set of transition rules. Each rule has the form of a simple state transition from computational state x to state y; and is encoded into the DNA as a *rule block* (denoted $x \rightarrow y$). If X denotes the DNA codeword for state x, and $\overline{X}$ denotes its Watson-Crick reverse complement, then the rule block for transition $x \rightarrow y$ is formed by the adjacent pair of DNA codewords, $5'$-$\overline{YX}$-$3'$. The completed transition rule region is formed by the catenation of all rule blocks, punctuated by short *stop* sequences. In Fig. 1, the transition rule region contains rule blocks, $a \rightarrow b$ and $b \rightarrow c$. Notice that the path encoded by a strand is independent of the physical ordering of the blocks. The $3'$ most codeword forms the *head*, which encodes the strand's initial state. For the computation shown in Fig. 1, this corresponds to the DNA codeword *A*. The final segment is the *spacer*, which carries no information but guarantees that planned hybridizations occur freely.

Each state transition is implemented by a three-step process of hybridization, polymerase extension, and denaturation, as shown in Fig. 1(ii), for transitions $a \rightarrow b$ and $b \rightarrow c$. The transition process is initiated by the spontaneous hybridization of the $3'$ head with a complementary codeword in the transition rule region. The state transition encoded by the hybridized rule block is then executed by polymerase extension of the head (horizontal arrow), which appends the DNA codeword for the new state to the strand's $3'$ end. Extension is terminated automatically when the polymerase encounters an adjacent *stop* sequence. This *polymerization stop* operation may be implemented by a length 3 polyadenine stop sequence, combined with the absence of dTTP ($2'$ deoxythymidine $5'$ triphosphate) in the reaction mixture. After denaturation, the extended strand is ready for a new round of hybridization and extension, and so on. At completion, a record of the computed path is encoded at the strand's $3'$ end, as a string of catenated DNA codewords. In Fig. 1(iii), this corresponds to the codeword string, $5'$-*ABC*-$3'$.

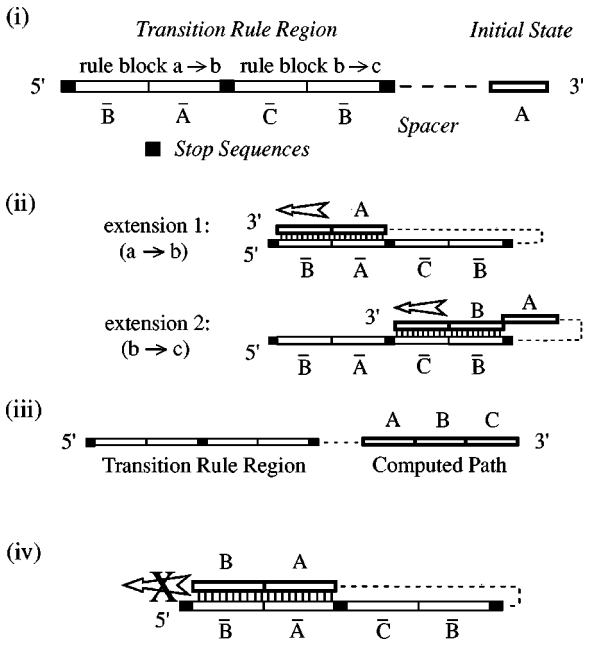

FIG. 1. WPCR implementation of a short computational path. (i) DNA single-strand encoding a WPCR program for executing the path, $a \rightarrow b \rightarrow c$. (ii) Each state transition begins with hybridization of the $3'$ head with a complementary DNA codeword in the transition rule region (e.g., codewords *A* and $\overline{A}$ in extension 1). Execution of the state transition is then implemented by the polymerase extension of the head (horizontal arrow) to the adjacent stop sequence (see text). Hydrogen bonded regions are indicated by vertical dashes. (iii) At completion, the strand's $3'$ end (sequence *ABC*) encodes the path, $a \rightarrow b \rightarrow c$. (iv) Execution of the second extension ($b \rightarrow c$), however, is inhibited by occupancy of the hairpin generated by the first extension [panel (ii), $a \rightarrow b$], which is both unextendable and much more energetically favorable than the hairpin required for extension 2. This effect is known as *backhybridization*.

As each strand may be encoded with a distinct set of transition rules and a distinct initial state, parallel computation on a massive scale is achievable, in principle, by the production of a combinatorial mixture of DNA strands, each of which is encoded to execute a single path from the computational graph of interest, followed by the iterative application of a thermal program appropriate for enforcing the parallel hybridization, extension, and denaturation of all encoded species. Theoretically, this process is computationally equivalent to the *in vitro*, parallel implementation of a set of finite state machines [1]. As a result, the computational power of WPCR is limited to recognition on the class of

regular languages [6]. Taken in combination with the ability to generate an arbitrary pool of initially encoded WPCR molecules (e.g., using the annealing biostep of Adleman's algorithm [2]), however, WPCR is capable of solving instances of a variety of problems in NP-complete [1,4].

B. Backhybridization

The feasibility of using WPCR to implement both single [1] and multiple [4,5] state transitions has been demonstrated in the laboratory. A serious barrier that confronts the scaling of the WPCR process, as well as its practical implementation for small model systems, however, is a systematic tendency for strands to participate in a powerful form of self-inhibition known as *backhybridization* [1,4,5]. As a concrete example consider again the WPCR implementation of the two-step path in Fig. 1. As shown in Fig. 1(ii), the second extension process requires the hybridization of codeword B with codeword $\overline{B}$ of rule block, $b \rightarrow c$. The resulting hairpin structure may then be extended by DNA polymerase (horizontal arrow). The efficiency of this process, however, is compromised by the ability of the DNA strand to form the alternative backhybridized hairpin shown in Fig. 1(iv), which is both unextendable and much more energetically favorable than the planned, extendable hairpin. For a longer WPCR implementation, the number of distinct backhybridized configurations increases as computation continues. For a strand undergoing the rth extension, a total of r alternative hairpin structures will be accessible, only one of which is extendable by DNA polymerase. Occupancy of the $r-1$ backhybridized structures serves to reduce the concentration of ssDNAs available to participate in computation.

Experimental investigation of WPCR has focused on tuning the thermal program to increase the efficiency of recursive extension [4,5]. In the original WPCR protocol, the recommended thermal program was the iterated application of a three-step thermal cycle: (1) a hybridization period (rapid cooling on ice), (2) a polymerization period (68 °C), and (3) a denaturation period (90 °C) [1]. In [4], both the explicit annealing and denaturation phases were eliminated, in favor of a two-step thermal cycle consisting of a brief incubation period at 65 °C, followed by a longer polymerization period at 80 °C, the experimentally determined optimum temperature. This thermal program, which was intended to establish concurrent annealing, polymerization, and denaturation has been experimentally verified to improve efficiency [4,5]. Recently, the effectiveness of this thermal protocol at implementing a single eight-step path was experimentally tested [5]. Although premature failure was reported after only five extensions, the authors in Ref. [5] assumed that difficulties due to backhybridization had been overcome by the thermal program, and attributed the premature failure to other error processes.

III. THE EFFICIENCY OF WHIPLASH PCR

A. The distribution of molecular lengths

In order to provide a theoretical framework for evaluating the impact of backhybridization on overall WPCR efficiency, the recursive extension of each strand may be modeled as an independent Markov chain [7]. Although a general analysis would also require an assessment of the effect of intermolecular interaction, in WPCR all DNA strands are anchored to a solid support, so that strand-strand interaction may be neglected. Let the *extension state* r of a WPCR-encoded strand be defined to equal the number of times the molecule has been successfully extended, plus 1. For a q-step WPCR implementation, extending strands may occupy a total of $q+1$ extension states, ranging from $r=1$ (completely unextended) to $r=q+1$ (fully extended). Note that for extension states, $r<q+1$, the r value of a strand also enumerates the extension process currently being executed by the strand.

Let ϵ_r denote the probability that an effective polymerase encounter with a strand in extension state r will find the strand in the extendable configuration. Here, the term ''effective'' refers to an encounter during which the orientations of the colliding molecules allow the formation of an enzyme-substrate complex. For simplicity, the use of a uniform codeword length shorter than the mean polymerase processivity is assumed, so that any effective polymerase-strand encounter results in the all-or-none, codeword-length extension of the encountered strand. In this case, a strand will increment its extension state during each effective encounter by either 1 or 0, with probabilities ϵ_r and $1-\epsilon_r$, respectively. For strands that reach the final absorbing state $q+1$, no further extension is possible (i.e., $\epsilon_{q+1}=0$). According to the Chapman-Kolmogorov equation [7], the single-step transition matrix for this process is given by

$$T=\begin{bmatrix} 1-\epsilon_1 & \epsilon_1 & \cdots & 0 & 0 \\ 0 & 1-\epsilon_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\epsilon_q & \epsilon_q \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{1}$$

For a WPCR implementation composed of N_o initially identical strands, the application of a single polymerization period, which implements an average of $\overline{N}_e$ effective encounters per strand at constant reaction temperature is modeled by the $\overline{N}_e$-step polymerization matrix, $\mathbf{E}=\mathbf{T}^{\overline{N}_e}$. The overall state occupancies after $\overline{N}_e$ effective encounters are then given by the elements of the row vector,

$$[N_1 N_2 \cdots N_{q+1}]=[N_o 0 \cdots 0]\cdot\mathbf{E}, \tag{2}$$

where $N_{r\geq 1}$ denotes the total number of strands occupying state r, and all N_o strands are assumed to begin in the fully unextended state. If a more complicated thermal program, which consists of several polymerization periods of diverse temperature and duration is applied, the extension process may be modeled by (1) estimating an $\overline{N}_e$ value for each subcycle, (2) constructing a transition matrix of the form provided by Eq. (1) for each subcycle, according to the reaction conditions employed, and (3) applying the resulting set of transition matrices, in order, to the initial state occupancy vector.

B. The efficiency per polymerase-DNA encounter

The quantity, ϵ_r may be estimated within the framework of a statistical thermodynamic model. Consider an ensemble, S_r of identical WPCR molecules, each of which occupies extension state, r (i.e., has been extended $r-1$ times). Assuming an all-or-none model of duplex formation, members of S_r will be distributed amongst a total of $r+1$ configurations: an unfolded ssDNA species, a polymerase extendable hairpin species, and a set of $r-1$ unextendable hairpin species, each of which is a backhybridized artifact of a previous extension. If the zero free-energy state is defined to correspond to that of the fully unstacked, unbonded ssDNA molecule, then the statistical weight of a simple hairpin configuration, which consists of a terminal loop of n unpaired bases and a lone duplex island of length j paired bases is estimated by

$$K=\frac{\sigma M(n)Z_j}{(n+1)^{1.5}}. \tag{3}$$

Here, Z_j is the statistical weight of stacking in the duplex island, σ is the cooperativity parameter [8], and $M(n)$ accounts for significant deviations from the Jacobson-Stockmayer 1.5 power law, observed for hairpins with very small loops due to chain stiffness and steric hindrance [8]. As each member of the set of hairpins that form in a WPCR implementation will contain a large terminal loop (i.e., $n \geqslant l_s$, the spacer length in bases), the formation of small loops is explicitly prevented. Effects due to chain stiffness and loop steric hindrance are therefore neglected $[M(n)=1]$.

Ensuring the uniformity of the various extension reactions of an implementation is critical to efficiency [4]. WPCR codewords are therefore typically selected to have uniform *GC* content, which results in an approximately equal Gibbs free energy of stacking for each codeword with its Watson-Crick complement [5]. In this case, the statistical weight of a duplex island of length j may be estimated by $Z_j=s^{j-1}$ [9], where s is the statistical weight of stacking for the average base pair doublet of the implementation.

Let the τ_r denote the characteristic relaxation time for a WPCR-encoded strand in state r. For simplicity, attention is restricted to WPCR implementations in which $\tau_r \ll \overline{t_e}$, the mean time between successive effective encounters with DNA polymerase, for a typical ssDNA in the mixture. In this case, ϵ_r corresponds to the equilibrium fraction of strands in S_r which occupy the extendable configuration, which is estimated by the ratio of the statistical weight of the extendable hairpin to the sum of the statistical weights of all structures. Constructing this ratio with the particular values $j=l$ (the uniform codeword length) and $j=2l$ for the single planned and $r-1$ backhybridized hairpin configurations, respectively yields

$$\epsilon_r=\left[1+\gamma_r s^l+\frac{(n_r+1)^{1.5}}{\sigma s^{l-1}}\right]^{-1} \tag{4}$$

Here, the parameter

$$\gamma_r=\sum_{i=1}^{r-1}\left[\frac{n_r+1}{n_i+1}\right]^{1.5}, \tag{5}$$

expresses the impact of variations in loop length between competing hairpin structures, as a function of extension state r, where n_r is the number of loop nucleotides of the extendable configuration, and each n_i is the number of loop nucleotides in the hairpin structure extended during previous extension process, i.

C. The mean efficiency of the *q*-step WPCR encoding

The evaluation of the set, $\{\epsilon_r;r=1,...,q\}$ for a specific WPCR-encoded DNA strand requires knowledge of the reaction conditions, and the static (i.e., specifics of the DNA encoding, and the number of transition rules, q) and dynamic (i.e., γ_r and n_r) characteristics of the process encoded by the strand. Here, the term dynamic refers to properties that vary with r. Determination of the precise r dependence of the dynamic characteristics requires knowledge of the ordered set of loop sizes, $\{n_r;r=1,...,q\}$ of the strand's q planned configurations, which is determined by the specific ordering of the encoded transition rules. The low order polynomial dependence of ϵ_r upon loop size, however, suggests that variations resulting from permutations in the rule ordering will be small. For this reason, the impact on efficiency due to the specific transition rule ordering is neglected. ϵ_r is then taken to be adequately approximated by $\overline{\epsilon_r}$, the efficiency per effective polymerase encounter for a strand with the ''mean'' q-step WPCR transition-rule encoding, in which the loop length of the extendable configuration of each extension state r is equal to $\overline{n_r}$, the mean length for that state averaged over all possible transition rule orderings.

In order to estimate the set, $\{\overline{n_r}\}$ let the location of the transition rule implemented by a strand in extension state r be modeled as a discrete random variable, distributed uniformly over the set of rules encoded on the strand. For a q-step path, with a uniform codeword length of l bases, the mean loop length of an extendable WPCR-encoded hairpin in state, $r=1$ is given by the average over the loop lengths of q extendable hybridized configurations, or $\overline{n_1}=(1/q)\Sigma_{h=0}^{q-1}(l_s+2lh)=l_s+l(q-1)$, where l_s is the spacer length, and the small contribution due to stop sequences has been neglected. Since each extension adds l bases to the spacing between the head and the transition rule region, the mean loop size of the extendable hairpin for a strand in extension state r is given by $\overline{n_r}=\overline{n_1}+l(r-1)\approx(q+r)l$, where the use of a short spacer sequence ($ql \gg l_s$) has been assumed. Insertion into Eq. (5), followed by application of the Euler-MacLaurin summation formula [10], retaining only the leading integral yields the approximation

$$\overline{\gamma_r}\approx 2(1-\delta_{1r})\left[\frac{(q+r)^{1.5}}{\sqrt{q}}-(q+r)\right] \tag{6}$$

Here, δ_{lr} is the Krönecker delta, and the factor $(1-\delta_{lr})$ must be included in order to properly model the absence of backhybridized structures during the first extension process. $\overline{\gamma_r}$

has limiting behavior $\overline{\gamma_r} \approx r$ for $1 < r \ll q$, and $\overline{\gamma_r} \approx 1.66r$ for $r = q$. If the minor deviation from $1.66r$ observed for small $r > 1$ is neglected, insertion of $\overline{\gamma_r}$ and n_r into Eq. (4) yields the estimate,

$$\overline{\epsilon_r} \approx \left\{ 1 + 1.66(1-\delta_{1r}) r s^l + \frac{[l(q+r)]^{1.5}}{\sigma s^{l-1}} \right\}^{-1}. \quad (7)$$

Although the single-path implementation is conceptually interesting, the primary appeal of WPCR lies in the potential to implement, in parallel a vast number of distinct computational paths. In addition to estimating the efficiency per effective polymerase encounter for a single-path WPCR implementation, Eq. (7) also has limited applicability to parallel WPCR. In particular, $\overline{\epsilon_r}$ estimates the mean efficiency per effective encounter for each strand, for the special case of parallel WPCR in which the various encoded species differ only in transition rule region structure.

D. The mean effective polymerase-DNA encounter rate

An estimate of the mean number of effective DNA polymerase encounters per DNA strand that occur in a single incubation period, in the limit of saturating substrate concentration, may be constructed by means of stochastic model of polymerase-ssDNA interaction. Consider an N_o strand WPCR implementation, in which a total of N_u units of *Thermus aquaticus* (*Taq*) DNA polymerase (1 unit corresponds to the synthesis of 10 nmol of product in 30 min, using an excess of activated salmon sperm DNA as substrate [11]) are utilized to implement recursive extension. For convenience, let the quantity v_t be defined as the number of completed oligonucleotide-length extensions performed per second by 1 unit of *Taq* DNA polymerase under optimal conditions, given excess dsDNA substrate (both target and primer) and the absence of competition from unextendable dsDNA substrate. *Taq* DNA polymerase is characterized as both fast and highly processive [11]. It is therefore assumed that (1) the polymerase-substrate dissociation time is large compared to both the time required for an oligonucleotide-length extension and the mean free time between encounters, and (2) each polymerase-ssDNA encounter results in the all-or-none, codeword-length extension of the strand. In this case, the total number of effective polymerase-DNA encounters in a single incubation period of length, Δt_p, is invariant to the extendability of the DNA substrate, and may be estimated by the product, $N_{\rm enc} = N_u v_t \Delta t_p$. If $N_{\rm enc}$ is assumed to be distributed uniformly amongst the entire set of strands, then the mean number of effective encounters per ssDNA, per polymerization period is estimated by

$$\overline{N_e} = \frac{N_{\rm enc}}{N_o} = \frac{N_u v_t \Delta t_p}{N_o}, \quad (8)$$

and the mean effective polymerase-ssDNA encounter rate is given by

$$\overline{r_e} = \frac{\overline{N_e}}{\Delta t_p} = \frac{N_u v_t}{N_o}. \quad (9)$$

E. Comparison with experiment

The *in vitro* WPCR implementation of an eight-step path was recently attempted [5]. The experimental protocol consisted of $N_o \approx 1.2 \times 10^{13}$ streptavidin bead-immobilized DNA strands and five units of *Taq* DNA polymerase, in a total reaction volume of 400 μL. DNA codewords were each 15 bases in length, with a uniform *GC* composition of 8/15, and the DNA spacer length was $l_s = 21$ bases. Each strand's transition rule region contained a rule block for implementing each of the eight transition rules. Prior to the computation, a head sequence encoding the first two states of the eight-step path was appended to each strand, using a novel ''input'' PCR technique. As a result, each strand initiated the computation at the second internal computational state, and in the second extension state ($r = 2$). The applied thermal program consisted of 15 iterations of a thermal cycle composed of (1) a 30 s incubation at 64 °C, (2) an increase to 80 °C in 60 s, (3) 300 s of polymerization at 80 °C, and (4) a decrease to 64 °C in 120 s. Constant pH = 7.0 and ionic strength, $I = 0.205M$ (estimated from the experimental conditions, $[K^+] = 0.05M$, $[Mg^{2+}] = 1.5\,mM$, using the methodology in Ref. [12]) were maintained throughout the experiment.

In Ref. [5], a total of seven replicates, denoted $\{R_r; r = 2,...,8\}$ of the above WPCR implementation were performed, where R_r was designed to independently assess the ability of the protocol to implement the rth transition. In particular, for each replicate R_r, a primer was designed to append, by polymerase extension, a short, nonextending ''output'' sequence to the $3'$ end of an r-fold extended strand. The corresponding primer was added to each R_r after a total of $2(r-1)$ thermal cycles. Following completion of 15 thermal cycles, the performance of each replicate was evaluated by means of a secondary, ''output'' PCR amplification. For R_r, this process essentially accomplished the PCR amplification of the set of r-fold extended strands. Results were visualized by gel electrophoresis. Output PCR therefore evaluated the success of each round on an all-or-none basis: the production of even a tiny fraction of r-fold extended product in replicate R_r was interpreted as success for the rth extension. In the output gel, bright bands were observed at the mobilities characteristic of molecules having undergone a total of 1–5 extensions (including the first, which was implemented by input PCR). Numerous very faint bands, observed at various mobilities, are assumed to be indicative of error extension during WPCR and output PCR, rather than correctly extended product. Although premature failure occurred after only five transitions, the authors assumed that problems due to backhybridization had been overcome by the optimized thermal program, and attributed this failure to error hybridization during the output. A comparison between the predictions of the current model of efficiency and the experimentally observed behavior in Ref. [5] provides an opportunity to evaluate the theoretical validity of this assumption.

The implementation in Ref. [5] of the first extension by input PCR was modeled by assigning the value, $\epsilon_1 = 1$. For extensions, $r > 1$, each ϵ_r value was approximated by the mean value, $\overline{\epsilon_r}$ provided by Eq. (7). The consensus value of

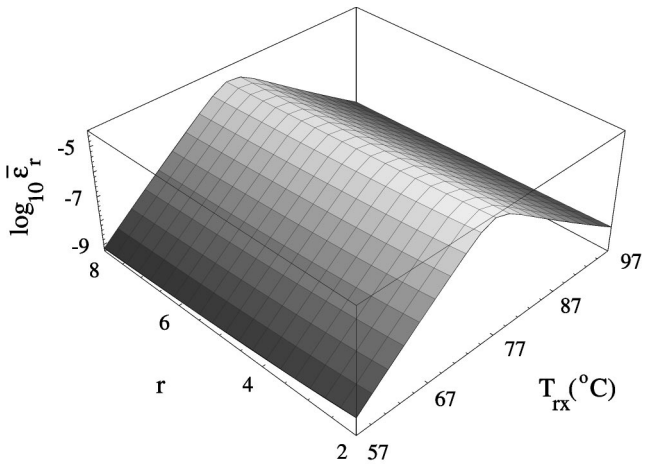

FIG. 2. An estimate of the mean extension efficiency per effective polymerase encounter, $\overline{\epsilon_r}$ for the WPCR implementation in Ref. [5] as a function of reaction temperature T_{rx} and extension process r.

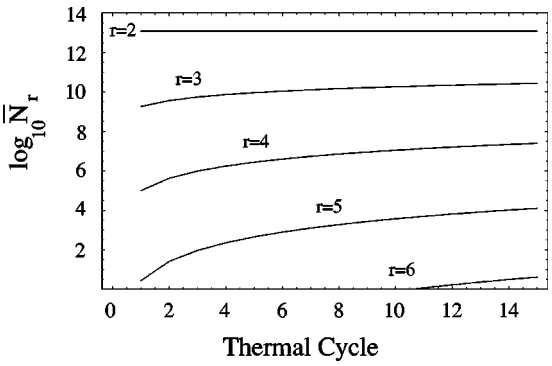

FIG. 3. The overall efficiency of Komiya *et al.* [5]. An estimate of the mean number of DNA strands $\overline{N_r}$ predicted to have undergone from 1 extension ($r=2$) to 5 extensions ($r=6$), after the completion of 1–15 thermal cycles, in the eight-step WPCR implementation reported in Ref. [5]. The total strand number was approximately 1.2×10^{13} strands.

the cooperativity parameter, $\sigma\approx4.5\times10^{-5}$ was adopted [9]. The free energy of duplex formation, ΔG^o, of each member of the codeword set applied in Ref. [5], with its Watson-Crick complement, was estimated using the nearest-neighbor model of Ref. [13]. Free energies were verified to approximately satisfy the assumption of codeword energetic uniformity. Following adjustment for experimental ionic strength by using the methodology described in Ref. [13], codeword energies were averaged, and divided by $l-1$ to obtain the mean doublet Gibbs free energy of stacking $\overline{\Delta G^o_{nn}}$ The corresponding mean doublet statistical weight, $s=-\overline{\Delta G^o_{nn}}/RT_{rx}$ was then combined with Eq. (7) to estimate $\overline{\epsilon_r}$ as a function of r and T_{rx}. As shown in Fig. 2, for the WPCR implementation in Ref. [5], ϵ_r for $r>1$ is predicted to reach a maximum value of 3.2×10^{-5}, at $r=2$ and $T_{rx}\approx77\,^\circ$C, which is in good agreement with the experimentally determined optimum of 80 °C. This optimal T_{rx} is predicted to vary by less than 1 °C over the entire range of extension states, $r=2,...,8$. Optimal extension efficiency is predicted to decrease slowly with increasing r, to a minimum of 1.2×10^{-5} for strands in extension state $r=8$, due to the generation of additional backhybridized structures.

An estimate of the mean number of DNA polymerase encounters per strand per round, $\overline{N_e}$, for each component of the four-stage thermal cycle of Ref. [5] was estimated using Eq. (8). v_t was estimated by dividing the bulk rate of addition of bases defined to equal 1 unit of enzyme [3.35×10^{12} bases/(unit s), under optimal conditions, using excess dsDNA substrate [11]] and the mean number of bases added per polymerase-substrate encounter (i.e., processivity). Based on the manufacturer's estimate, a mean processivity of 50 bases per effective encounter was assumed [14], which yielded the estimate, $v_t\approx6.70\times10^{10}$ effective encounters/(unit s). Application of Eq. (8) yielded $\overline{N_e}\approx8.4$ effective encounters per strand per 300 s period at 80 °C, and $\overline{N_e}\approx1.7$ effective encounters per strand per 60 s period at 64 °C. The polymerization process during each ramp was modeled as a discrete set of brief polymerization periods of equal length, where the interval size was selected so that each period implemented an average of one effective polymerase encounter per strand. All $\overline{N_e}$ and $\overline{\epsilon_r}$ estimates were then applied in combination with the Markov chain model of Sec. III A, to estimate the number of strands in Ref. [5] having undergone each of from 1 (denoted by $\overline{N_2}$) to 8 (denoted by $\overline{N_9}$) extensions, as a function of thermal cycle. Here, the overscore denotes the use of the set $\{\overline{\epsilon_r}\}$ in the single-step transition matrix [Eq. (1)].

As illustrated in Fig. 3, the production of small fractions of molecules that have successfully undergone from 1 to 4 extensions is predicted during the first thermal cycle (i.e., $\overline{N_2}$ through $\overline{N_5}$ each exceed unity following one thermal cycle). The production of longer molecules, however, is delayed until the 11th thermal cycle, when the appearance of fivefold extended molecules is predicted ($\overline{N_6}>1$). The production of sixfold extended molecules is not predicted during the entire course of the experiment. This predicted performance is in excellent agreement with the experimentally observed behavior reported in Ref. [5] (i.e., failure after a total of five transitions), a result that lends strong support to the view that backhybridization was responsible for the reported premature failure, rather than some other, unmodeled process.

F. The feasibility of WPCR

Strategies to increase WPCR efficiency may be classed into two categories: (1) attempts to decrease the fractional occupancy of backhybridized structures, $1-\epsilon_r$ and (2) attempts to increase the mean total number of effective polymerase encounters per strand, $\overline{N_{\rm tot}}$. As the generation of backhybridized hairpins is a natural consequence of extension, the elimination of these structures by encoding or simple redesign appears to be impossible. The impact of backhybridization on $\overline{\epsilon_r}$ is determined by the magnitude of $1.66rs^l$ [cf., Eq. (7)]. Although this l dependence suggests that efficiency may be enhanced by reducing l, the use of primers shorter than 16 bases is not generally recommended

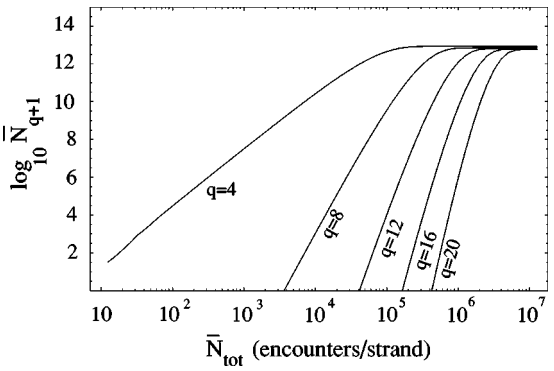

FIG. 4. The scaling of WPCR efficiency. An estimate of the mean number of fully extended DNA strands, $\overline{N}_{q+1}$ produced as a function of the mean total number of effective polymerase encounters/strand, $\overline{N}_{\mathrm{tot}}$ for WPCR implementations of computational paths of length $q=4$, 8, 12, 16, and 20. In each case, DNA codewords are energetically equivalent to those used in Ref. [5], and the application of the estimated optimal T_{rx} has been assumed.

for PCR, due to an increased tendency to dissociate [15]. The 15 base codewords used in Ref. [5] are, therefore, likely to be of optimal length. This trend is general: any simple strategy, that seeks to enhance efficiency by decreasing the stability of backhybridized structures (i.e., decreasing s or l) also decreases the stability of desired structures.

The most straightforward method to increase $\overline{N}_{\mathrm{tot}}$ is the application of additional polymerization time. As shown in Fig. 4, the WPCR implementation in Ref. [5] (at $T_{rx}=77\,^{\circ}\mathrm{C}$) is predicted to reach *minimal completion* ($\overline{N}_9\approx 1$) after roughly $\overline{N}_{\mathrm{tot}}=2.9\times 10^3$ effective encounters per strand ($q=8$ curve, x intercept). At the estimated encounter rate of 1.68 effective encounters/(strand min) predicted for conditions of Ref. [5] this requires a total of 28.4 h of polymerization. As shown in Fig. 4, the attainment of minimal completion is predicted to become much more difficult as q is increased. For a $q=20$ WPCR implementation (at $T_{rx}=77\,^{\circ}\mathrm{C}$), minimal completion requires a total of roughly 3.4×10^5 effective encounters/strand, or 142 days at 1.68 effective encounters/(strand min).

A related strategy is to decrease the mean time between polymerase-DNA encounters, $\overline{t}_e=\overline{r}_e^{-1}$ [cf. Eq. (9)]. This may be accomplished by (1) applying excess polymerase, N_u, (2) decreasing the number strands, N_o, and/or (3) using an polymerase with a higher v_t value for oligonucleotide-length substrate (i.e., a lower processivity). For instance, application of an excess of *Taq* DNA polymerase ($N_u=50$ units, which yields $\overline{t}_e=3.6\,\mathrm{s}$) is predicted to decrease the time for minimal completion of a $q=20$ WPCR implementation to 14.2 days. Although further decreases in $\overline{t}_e$ may be accomplished straightforwardly, a substantial reduction in N_o is inconsistent with the goal of massive parallelism. Furthermore, the use of a very large excess of polymerase is known to increase the rate of nonspecific extension. In any case, the success of decreases in $\overline{t}_e$ to enhance overall efficiency is limited kinetically by the condition for equilibrium, $\overline{t}_e \gg \tau_r$. Although a detailed treatment of WPCR kinetics is beyond the scope of this work, a rough estimate of $\tau_{r>1}$ may be obtained by modeling the WPCR reaction in terms of a simplified r-state model of hairpin kinetics, in which strands in extension state $r>1$ are distributed amongst a set composed of $r-1$ identical, but distinguishable backhybridized hairpin structures, and the random coil (occupancy of the extendable hairpin is neglected). For this purpose, the forward (k_r^+) and reverse (k_r^-) rate constants of hairpin formation for a strand in extension state r were modeled by the mean values, averaged over all $r-1$ backhybridized hairpins accessible to the strand. This model yields the relaxation time, $\tau_r=[(r-1)k_r^+ + k_r^-]^{-1}$. Following [16], k^+ values for individual hairpins may be modeled by the empirical expression, $k^+ \approx A(n+l)^{-2.6}$, where the constant, $A\approx 6.7\times 10^7$ has been estimated by using the reported specific value in Ref. [16] of $k^+=1.4\times 10^4$ for a loop of $n=21$ thymine residues. The mean WPCR loop lengths for extension processes $r>1$, for an $l=15$, $q=20$ WPCR implementation range from $\overline{n}_2 \approx 330$ to $\overline{n}_{20}\approx 600$ bases. Corresponding relaxation times for $T_{rx}=77\,^{\circ}\mathrm{C}$ range from $\tau_2=1.9\,\mathrm{ms}$ to $\tau_{20}=2.9\,\mathrm{ms}$, which is roughly three orders of magnitude less than the mean encounter time of $\overline{t}_e=3.6\,\mathrm{s}$, obtained using $N_u=50$ units of *Taq* DNA polymerase and $N_o=1.2\times 10^{13}$ DNA strands. Substantial additional decreases in $\overline{t}_e$ are, therefore, expected to rapidly become ineffective.

An examination of Fig. 4 indicates that the application of WPCR to massively parallel computation is substantially less feasible. For the WPCR implementation in Ref. [5], adjusted to the optimal T_{rx} of 77 °C, a total of roughly 6.0×10^5 encounters per strand is predicted to be required for completion of 50% of the $N_o=1.2\times 10^{13}$ strands. This corresponds to 248 days and 24.8 days of polymerization, using $N_u=5$ units, and $N_u=50$ units of *Taq* DNA polymerase, respectively. For a $q=20$ implementation, attaining a 50% completion rate requires a total of roughly 10^7 encounters per strand, or 417 days of polymerization, using $N_u=50$ units of *Taq* DNA polymerase.

IV. PNA-MEDIATED WPCR

A. The *P* loop

An alternative approach is to redesign the WPCR architecture to enable the specific inhibition of backhybridized structures by targeted PNA_2/DNA triplex formation. The ability of polyamine nucleic acid strands (PNAs) to bind to complementary ssDNA with extremely high affinity and sequence specificity is well characterized [17]. For homopyrimidine PNA strands, binding to a complementary ssDNA target sequence occurs with stoichiometry 2 PNA:1 DNA, indicating the formation of a PNA_2/DNA triplex. Under appropriate reaction conditions, rapid formation of the triplex structure occurs even if the target DNA single strand is involved in hybridization with a complementary DNA strand (i.e., by strand invasion). In this case, triplex formation results in the extrusion of the target-complementary DNA strand, forming a "*P* loop" [18].

Formation of the PNA_2/DNA triplex is a two-step process [19]. In the first step, sequence-specific invasion of the

target region of the DNA double strand is accomplished by a single PNA molecule. The result of this reversible nucleation phase is the formation of an antiparallel, Watson-Crick base paired PNA/DNA duplex (By convention, PNAs are oriented from N to C terminal). The second phase of triplex formation is a virtually irreversible ''locking'' phase, in which a second PNA strand complexes into the major groove of the PNA/DNA duplex by means of Hoogsteen base pairing [20]. This Hoogsteen strand adopts a parallel orientation relative to the target DNA strand. The extraordinary stability of this triplex structure is due to (1) the absence of the strong electrostatic repulsion that exists between the sugar phosphate backbones of a DNA duplex or triplex (PNA is electrostatically neutral), and (2) the presence of a strong hydrogen bonding interaction between the phosphate oxygen atoms of the target DNA backbone and the amide nitrogen bonds of the Hoogsteen strand [21]. Conditions that promote the unwinding of the DNA are, however, required for the initial nucleation step [18]. At constant temperature, this is accomplished by the application of low salt concentration ($[Na^+] \leq 40mM$). Alternatively, unwinding may be accomplished thermally. Once formed, the locked triplex remains stable, even under conditions that inhibit the initial nucleation phase (e.g., high $[Na^+]$ and low T_{rx}) [18].

Achieving a high saturation of target sequences is critical to the effectiveness of antisense and antigene strategies based on PNA_2/DNA formation. The use of ''bis-PNA,'' in which the two PNA strands involved in *P* loop formation are connected by means of a flexible linker has been verified to be a highly effective means of enhancing the rate of strand invasion [22]. The use of cationic pseudoisocytosine (*J*)-containing PNA sequences has also been suggested [23]. The combined application of these two techniques, under conditions of low salt ($20mM[Na^+]$) and excess PNA, was reported to result in the rapid, virtually irreversible attainment of a near unity fractional saturation of dsDNA target [23]. For the reported pseudo-first-order rate constant of 2.33 min^{-1} obtained using 1.0n*M* DNA target and 1.0μM bis-PNA, at 37 °C and 20m*M* $[Na^+]$ conditions, fractional saturations of 0.9, 0.99, and 0.999 are predicted in roughly 60, 120, and 180 s, respectively.

B. PNA-mediated WPCR

In order to enable the uniform inhibition of backhybridized hairpins by targeted PNA_2/DNA triplex formation, the basic structure of the WPCR transition rule block may be modified so that a ssDNA target sequence for bis-PNA binding is synthesized between the initial and newly polymerized state codewords, during each state transition. In particular, separation of the source and target codeword pair of each of the *q* rule blocks of an encoded strand by the ''target generator'' sequence $5'$-$T_4CT_2CT_2$-$3'$ results in the separation of successively polymerized state codewords by $5'$-$A_2GA_2GA_4$-$3'$, the target for the cationic bis-PNA molecule described in Ref. [23]. Computation may then be implemented by subjecting the mixture to a total of *q* polymerization periods, each of which (with the exception of the final period) is followed by a round of bis-PNA treatment.

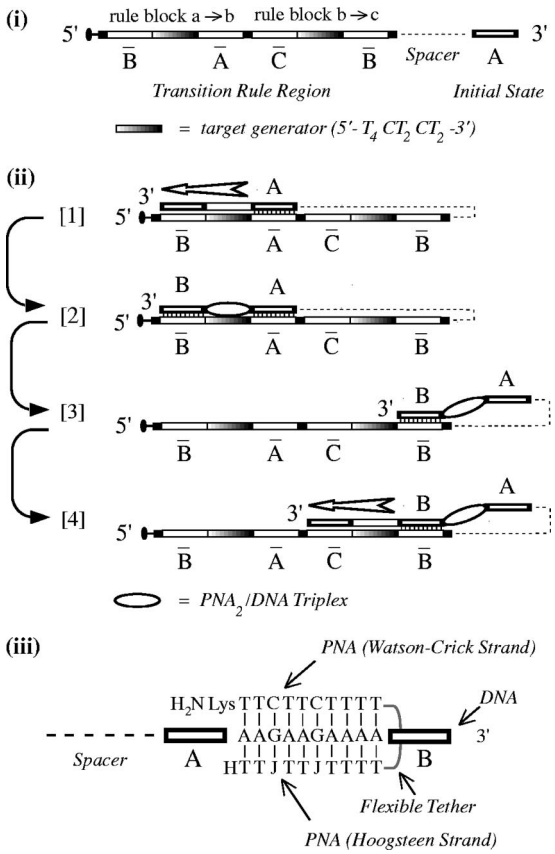

FIG. 5. PNA-mediated WPCR. (i) PWPCR encoding for computing the path $a \rightarrow b \rightarrow c$. This encoding scheme differs from WPCR, in that each transition rule block contains a ''target generator'' sequence, $5'$-$T_4CT_2CT_2$-$3'$. (ii) Each state transition (horizontal arrow) then produces the ''target'' sequence, $5'$-$A_2GA_2GA_4$-$3'$ between the source and target codewords. This process is illustrated in panel 1 for the transition, $a \rightarrow b$. Addition of bis-PNA then results in the formation of a PNA_2/DNA triplex (shown in panel 2 as an ellipse) at the target sequence. This triplex region destabilizes the extended hairpin structure (panel 2), increasing the occupancy of the extendable hairpin (panel 3) required for the next transition (panel 4). (iii) PNA_2/DNA triplex structure.

Following completion of all *q* cycles. bis-PNA clamps may be removed *en masse* by a high temperature denaturing wash. This modified computational architecture, which is referred to as PNA-mediated whiplash PCR (PWPCR), is illustrated in Fig. 5 for the implementation of the three-state, two-step path, $a \rightarrow b \rightarrow c$. As shown in Fig. 5(ii) panel 1, the first state transition ($a \rightarrow b$), which is implemented by polymerase extension, produces target sequence $5'$-$A_2GA_2GA_4$-$3'$ between state codewords *A* and *B*. A round of bis-PNA treatment, which is implemented by subjecting the mixture to a 180 s, low $[Na^+]$, excess bis-PNA wash then results in the effective saturation of instances of the target sequence with *P*-loop forming bis-PNA [Fig. 5(ii), panel 2]. Formation of this triplex structure, shown in detail in Fig. 5(iii), in turn destabilizes the backhybridized hairpin, and increases the occupancy of the extendable hairpin [Fig. 5(ii), panel 3], which is required for the execution of the next transition, $b \rightarrow c$ [Fig. 5(ii), panel 4].

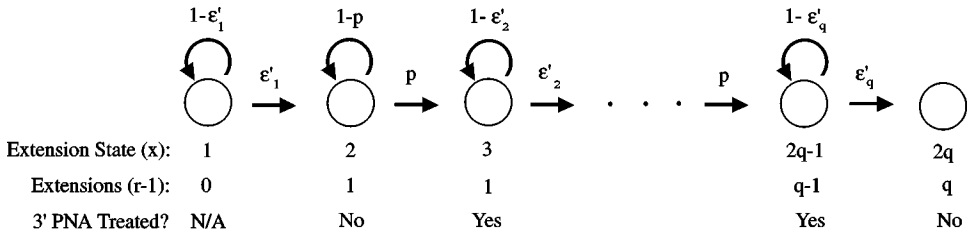

FIG. 6. Markov chain model of the PWPCR protocol. Each strand progresses in single-step increments through an ordered set of $2q$ states, ranging from $x=1$ (fully unextended) to $x=2q$ (fully extended). Each ϵ_r' value denotes the probability of executing the rth extension, during a single effective polymerase encounter. p denotes the probability of PNA_2/DNA triplex formation at the $3'$ most target site during a PNA treatment. Each effective polymerase encounter is modeled by applying the accompanying state transition matrix to the state occupancy row vector (see text), under the condition $p=0$. Each bis-PNA treatment phase is modeled by application of the state transition matrix, under the condition $\epsilon_r'=0$, $\forall r$.

C. The efficiency of PWPCR

In order to provide a framework for estimating the impact of bis-PNA targeting on the overall efficiency of the *in vitro* implementation of a q-step computational path, the process of recursive extension of each PWPCR-encoded strand may be modeled as a Markov chain. As in WPCR, strands are assumed to be anchored to a solid substrate, so that the potential for strand-strand interaction may be neglected. In addition, following each successful extension, a PWPCR strand is assumed to require a round of PNA treatment prior to further extension, due to the very high stability of a PNA-untreated, backhybridized structure (cf. Sec. IV D). As a result, in addition to the $q+1$ states necessary to model a set of strands distributed over $q+1$ distinctive lengths (cf., Sec. III C), an additional set of $q-1$ intermediate states is required to model the q bis-PNA treatments, for a total of $2q$ extension states. Let the extension state of a PWPCR-encoded strand be denoted by x. Note that the PNA treatment of the fully extended state is neglected.

Another complication is the need to model each bis-PNA treatment separately from each polymerization period. In particular, these two protocols cannot be performed concurrently, due to the low ionic strength required to achieve a rapid, high fractional saturation of target regions with bis-PNA. In order to facilitate the modeling of the overall process of strand extension with a single Markov chain, two types of probabilities are adopted: ϵ_r', which represents the probability that a strand will execute the rth extension process during a single polymerase encounter (note that in PWPCR, r does not denote the strand's extension state), and p, which represents the probability that a strand's $3'$ most target sequence will become PNA treated during a round of bis-PNA treatment. A single polymerase encounter is then modeled, for each strand in the mixture, by executing a single transition with $p=0$, while a single round of PNA treatment is modeled by executing a single transition with $\epsilon_r'=0,\forall r$.

The resulting Markov chain model is illustrated in Fig. 6. Each strand progresses, in single-step increments, through an ordered set of $2q$ states, ranging from $x=1$ (fully unextended) to $x=2q$ (fully extended). The subset of even states corresponds to the set of *unextendable* DNA hairpins, where a strand in state $x=2r$ has undergone a total of r previous extensions, but has not been subjected to a subsequent round of PNA treatment. During a round of bis-PNA treatment, PNA targeting of this strand's $3'$ target site occurs with probability p, an event that is modeled as a transition to the next odd state. Similarly, the subset of odd states corresponds to the set of *extendable* DNA hairpins, where a strand in state $x=2r-1$ has undergone a total of $r-1$ previous extensions, and the strand has also been subjected to a subsequent round of PNA treatment. During a single polymerase encounter, this strand may execute the rth extension with probability ϵ_r', a process that is modeled as a transition to the next even state. The single-step transition matrix has the form

$$\mathbf{T}=\begin{bmatrix} 1-\epsilon_1' & \epsilon_1' & 0 & \cdots & 0 & 0 \\ 0 & 1-p & p & \cdots & 0 & 0 \\ 0 & 0 & 1-\epsilon_2' & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-\epsilon_q' & \epsilon_q' \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \quad (10)$$

where $\mathbf{T}=\mathbf{T}(\{\epsilon_r'\},p)$. A single round of polymerization, which implements an average of $\overline{N}_e$ effective encounters/strand, is modeled by the $\overline{N}_e$-step polymerization matrix, $\mathbf{E}=\mathbf{T}(\{\epsilon_r'\},0)^{\overline{N}_e}$. Similarly, a single round of bis-PNA treatment is modeled by the single-step transition matrix, $\mathbf{P}=\mathbf{T}(\{0\},p)$. The overall state occupancies for a PWPCR protocol that consists of a set of q extension rounds, each of which is composed of a single round of polymerization followed by a single bis-PNA treatment, are then given by the elements of the row vector

$$[N_1 N_2 \cdots N_{2q}]=[N_o 0 \cdots 0]\cdot(\mathbf{E}\cdot\mathbf{P})^q, \quad (11)$$

where N_o denotes the total strand number and $N_{x\geqslant 1}$ denotes the total number of strands occupying state x.

D. The efficiency per polymerase-DNA encounter

The effect of PNA_2/DNA triplex formation on the per encounter polymerization extension efficiency, ϵ_r', may be estimated by means of a statistical thermodynamic model. Due to the experimentally reported compactness of the P loop, the presence of a triplex region immediately adjacent to

the head sequence is assumed to have a negligible impact on the ability of the head to hybridize with a complementary sequence in the transition rule region. As a result, the statistical weight K_r of the extendable configuration anticipated for round r is constructed as described in Sec. III B. Estimating the effect of the triplex region on the statistical weight of backhybridized structures requires more care. For simplicity, an all-or-none model of duplex formation is assumed. In this case, each extension facilitates the formation of three distinct backhybridized structures: (1) an extended backhybridized hairpin, which is composed of a pair of length l duplex islands punctuated by a P loop [e.g., Fig. 5(ii), panel 2], and (2,3) two shorter backhybridized hairpins, each of which results from formation of only one of the two duplex islands associated with the full-length structure. The total statistical weight of backhybridization due to extension process r is equal to the sum of the statistical weights of these three structures,

$$K_r = \sigma s^{l-1}\left[\frac{1+\sigma s^{l-1} Z_p}{\left(n_r'+1\right)^{1.5}} + \frac{1}{\left(n_r'+\frac{10l}{3}+1\right)^{1.5}}\right], \quad (12)$$

where a target sequence of length $l_p=\frac{2}{3}l$ has been assumed, n_r' denotes the terminal loop length of a strand in extension process r's planned configuration, and Z_p is the statistical weight of the P loop region. It should be carefully noted that this quantity is the statistical weight of the P loop region, *given* the presence of an irreversible PNA_2/DNA triplex, and as such does not include the energetics of PNA_2/DNA triplex formation. Rather, the PNA_2/DNA triplex is assumed to be present, and hence carries a statistical weight of 1. For simplicity, the decrease in the Jacobson-Stockmayer loop closure probability anticipated due to the increased local chain stiffness in the short triplex region is neglected.

A precise estimate of the energetics of interaction between the postinvasion P loop components (the triplex and extruded single strand) is absent from the literature. The observation of a distinctive eye structure for the P loop [18], however, suggests the absence of stabilizing interactions between the extruded single-stranded target-complementary strand and the PNA_2/DNA triplex. Molecular mechanics calculations also indicate that stabilizing interaction between the two P loop components is minimal [24]. In Ref. [25] the electrostatic interaction between the positive charges of a cationic bis-PNA molecule and the phosphates of the adjacent DNA backbones was also reported to be negligible. Z_p is therefore assumed to be entirely entropic in origin, and is modeled as a stiff chain with excluded volume. The general form of the Jacobson-Stockmayer equation for the statistical weight of a closed loop of m links is $Z=(m+d)^{\alpha}$, where $\alpha=-1.7$ for an internal loop with excluded volume [8], and d is an empirical parameter that accounts for chain stiffness [26]. The local value of d is roughly equal to the order of magnitude of the persistence length of the chain [26]. In the absence of specific information regarding the persistence length of a PNA_2/DNA triplex, the decrease in probability of loop closure associated with the stiffnesses of the ssDNA and triplex regions is neglected, and the value $d=0$ is adopted. It should be noted that this value of d corresponds to the assumption of the worst case behavior, from the standpoint of PWPCR efficiency. Since an internal loop formed by k broken base pairs contains a total of $2k+2$ links between bases, for the P loop region, $m=2+4l/3$, and $Z_p=(2+4l/3)^{-1.7}$.

The equilibrium fraction of extendable structures is estimated by the ratio of the statistical weight of the extendable structure with all structures,

$$\epsilon_r' = \left\{1+\gamma_r''+\gamma_r'\left[1+\frac{\sigma s^{l-1}}{\left(2+\frac{4}{3}l\right)^{1.7}}\right]+\frac{(n_r'+1)^{1.5}}{\sigma s^{l-1}}\right\}^{-1}. \quad (13)$$

For a target sequence length of $l_p=\frac{2}{3}l$, each member of the trio of backhybridized structures generated in the rth round of extension will contain a terminal loop of length either n_r' or $n_r''=n_r'+10l/3$ bases. γ_r' and γ_r'' are defined using n_r' and n_r'', respectively, in exact analogy with γ_r [Eq. (5)]. Explicit evaluation of ϵ_r' for a given strand requires knowledge of the specific encoding and reaction conditions. As in the case of WPCR, the low order polynomial dependence of ϵ_r' on loop size suggests that variations in ϵ_r' resulting from permutations in transition rule order will be small. The impact of specific transition rule ordering is, therefore, again neglected, and the efficiency per polymerase encounter of a PWPCR strand is considered to be well approximated by the value that results from the mean ordering $\overline{\epsilon_r'}$. Calculation of the loop quantities $(\overline{n_r'},\overline{\gamma_r'},\overline{\gamma_r''})$ that characterize the mean ordering for a PWPCR-encoded strand with characteristics, q, l, and s proceeds in exact analogy with the calculation of the quantities $\overline{n_r},\overline{\gamma_r}$, discussed for the WPCR architecture in Sec. III C. For a target sequence length of $2l/3$ nucleotides, $\overline{n_r'}\approx 2l(r+2q/3)$, and $r\leqslant\overline{\gamma_r'}\approx\overline{\gamma_r''}\leqslant 1.94r$, for $1<r\leqslant q$. Adopting the maximum limiting behavior of $\overline{\gamma_r'}$ yields the final expression

$$\overline{\epsilon_r'}\approx\left\{1+1.94r(1-\delta_{1r})\left[2+\frac{\sigma s^{l-1}}{\left(2+\frac{4}{3}l\right)^{1.7}}\right]+\frac{\left[l(2r+\frac{4}{3}q)\right]^{1.5}}{\sigma s^{l-1}}\right\}^{-1} \quad (14)$$

for the extension efficiency per polymerase-ssDNA encounter for the rth extension of a PNA-treated PWPCR strand with static characteristics q, l, and s. Here the Krönecker delta δ_{1r} accounts for the absence of backhybridized structures during the first extension.

E. The parallelization of PWPCR

The ultimate aim of both WPCR and PWPCR is the parallel, *in vitro* implementation of a massive number of distinct computational paths. The ability of either architecture to

implement parallel computations may be assessed as follows. Consider the simple parallelization of a PWPCR implementation, in which the total copy number N_o has been parsed into P distinct species, each of which encodes a distinct computational path and is present with equal copy number, $N_c = N_o/P$. The maximum parallelism obtainable by this implementation is equal to $P_{\max}=N_o/N_c$. However, this is practically obtained only when N_c is sufficiently large to ensure the full extension of at least one copy per path.

The use of codeword sets with uniform length and energetics is assumed. In the absence of bimolecular interaction, the completion of each strand is an independent process. Therefore, the mean probability of failure per path (i.e., for all N_c copies) is given by $p_f=(1-\chi)^{N_c}$, where $\chi \equiv N_{2q}/N_o$ is the fraction of fully extended strands at the completion of the implementation. As the mean number of failed paths in a parallel implementation is equal to the product $p_f P$, the condition for completion of a parallel PWPCR implementation may be expressed by the threshold relation $p_f P<1$. Combining these expressions, maximum parallelism for a PWPCR implementation that consists of N_o strands, each of which completes with a mean probability of χ, is predicted when N_c is chosen such that

$$(1-\chi)^{N_c}=\frac{N_c}{N_o}. \tag{15}$$

For convenience, the value of N_c that satisfies this equation will be denoted $N_c(\mathrm{opt})$. Although Eq. (15) may not, in general, be used to produce a closed-form expression for $N_c(\mathrm{opt})$, an estimate may be obtained numerically for any specific $\{\chi,N_o\}$ pair of interest. The accompanying maximum parallelism is then estimated by $P_{\max}=N_o/N_c(\mathrm{opt})$.

F. Model predictions

A comparison of Eqs. (7) and (14) indicates that the most significant effect of targeted triplex formation on the efficiency per encounter is the destabilization of the full length backhybridized configuration by a factor of σ. The impact of this effect is best illustrated by concrete application. For this purpose, ϵ_r' for each of the last seven steps of a PWPCR implementation of the eight-step computational path described in Ref. [5] has been estimated using the mean value over all transition rule orderings, $\overline{\epsilon}_r'$, provided by Eq. (14). For consistency, codeword energetics and buffer conditions for the extension reaction were selected that are identical to those in Ref. [5]. In addition, the first extension process of each strand was assumed to be implemented by input PCR [5]. Results are illustrated in Fig. 7. As shown, the optimal T_{rx} for extension is predicted to be roughly 60 °C, for strands in extension states $r=2,...,8$, with a variation of less than 1 °C. At this temperature, the PWPCR extension process is predicted to proceed with an efficiency per effective encounter that varies from $\overline{\epsilon}_2'\approx 0.04$ $(r=2)$ to $\overline{\epsilon}_8'\approx 0.01$ $(r=8)$, an improvement of roughly three orders of magnitude over the equivalent WPCR implementation (cf., Fig. 2).

The overall efficiency of the PWPCR protocol, in terms of $\overline{N}_{2q}$, the estimated total number of fully extended strands

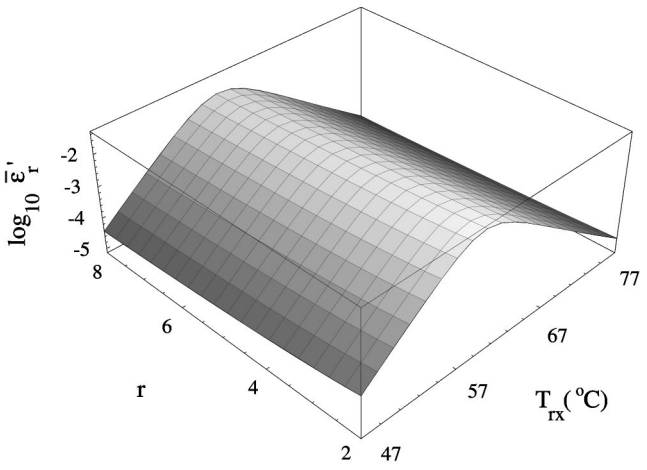

FIG. 7. An estimate of the mean extension efficiency per effective polymerase encounter $\overline{\epsilon}_r'$ for the PWPCR implementation of an eight-step path, as a function of reaction temperature T_{rx} and extension process r. Codeword energetics and buffer conditions are assumed to be identical to those in Ref. [5].

produced after q experimental cycles of polymerization and bis-PNA treatment, has been estimated using the Markov chain model of Sec. IV C. Here, the overscore denotes the use of the set of mean values, $\{\overline{\epsilon}_r'\}$ in the transition matrix [Eq. (10)]. Codeword energetics, *Taq* polymerization buffer conditions, and total strand number ($N_o=1.2\times 10^{13}$ strands) were taken to be identical to those applied in Ref. [5]. The total number of effective polymerase-DNA encounters/(strand cycle) was estimated using Eq. (8). For consistency, the implementation of the first extension ($r=1$) by input PCR was assumed. This was modeled by assigning an efficiency of unity for the first extension ($\epsilon_1'=1$). The application of a near-optimal polymerization temperature of $T_{rx}\approx 60\,^{\circ}\mathrm{C}$ was assumed in the polymerization phase of each PWPCR cycle. Each round of bis-PNA treatment was assumed to be sufficient to attain a near-unity fractional saturation of target sequences in solution (i.e., exposure of the anchored strands for 180 s to a 1.0 μM bis-PNA solution, at 20mM [Na^+], and room temperature). Results are illustrated in Fig. 8, as a function of the number of polymerase encounters implemented per cycle, $\overline{N}_e$ for PWPCR implementations of length $q=8$, 16, 20, 32, and 64 steps.

As shown by Fig. 8, the targeted triplex formation characteristic of PWPCR is predicted to be accompanied by a large increase in overall strand completion efficiency [cf., Fig. 4, where for comparison purposes, $\overline{N}_e$ values for PWPCR should be converted to total effective encounter number by $\overline{N}_{\mathrm{tot}}=(q-1)\overline{N}_e$]. For instance, the PWPCR implementation of the eight-step path reported in Ref. [5], at an effective encounter rate of $\overline{N}_e\approx 18$ encounters/(strand cycle) (for a total of $\overline{N}_{\mathrm{tot}}=126$ encounters/strand, which is identical to the value estimated in Ref. [5]), is predicted to complete roughly 1.2×10^9 of the initial 1.2×10^{13} encoded strands, for a per-strand completion rate of $\chi\approx 1.0\times 10^{-4}$. For comparison purposes, a WPCR implementation of this computation, at the optimal T_{rx} is predicted to require roughly $\overline{N}_{\mathrm{tot}}=3$

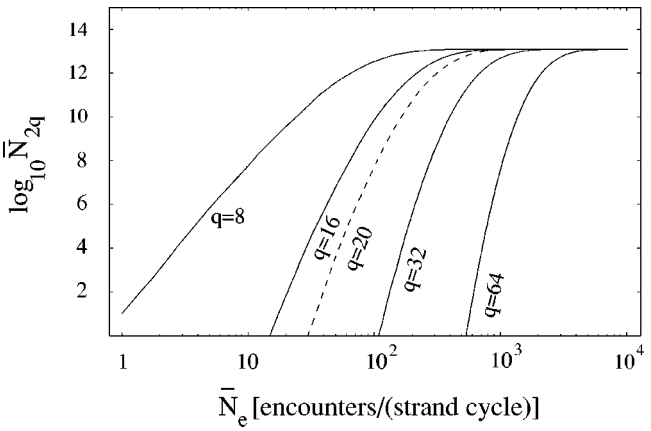

FIG. 8. The efficiency of PWPCR. An estimate of the mean number of fully extended DNA strands $\bar{N}_{2q}$ produced as a function of the mean number of effective polymerase encounters per strand per cycle, $\bar{N}_e$, for PWPCR implementations of paths of length $q=8$, 16, 20, 32, and 64. Codeword energetics and buffer conditions are assumed to be identical to those in Ref. [5], and the application of the estimated optimum T_{rx} has been assumed.

$\times 10^3$ encounters/strand. The maximum parallelism $P_{\max}$ achievable by this $q=8$ PWPCR implementation may be estimated by solving Eq. (15) for $N_c(\text{opt})$, and applying the simple relationship, $P_{\max}=N_o/N_c(\text{opt})$. This procedure predicts that a maximum parallelism of $P_{\max}\approx 7.5\times 10^7$ distinct paths is obtained using a copy number of $N_c(\text{opt})\approx 1.6\times 10^5$, which corresponds to the parallel implementation of roughly $N_{\text{ops}}=qP_{\max}\approx 10^9$ distinct computational operations.

An examination of the scaling behavior of PWPCR efficiency indicates that a substantial improvement may be obtained by a realistic set of modifications to the extension protocol. For instance, according to Fig. 8 (dashed curve), the application of a realistic set of extension reaction conditions [$T_{rx}=60\,^\circ$C; $N_u=27$ units and $\Delta t_p=60$ min, yielding $\bar{N}_e\approx 542.7$ effective encounters/(strand cycle), for a mean total of $\bar{N}_{\text{tot}}\approx 10^4$ effective encounters per strand over all 19 total hours of polymerization] allows the PWPCR implementation of a computation of length $q=20$ rounds, with a per strand completion efficiency of $\chi\approx 0.50$. Using Eq. (15), a maximum parallelism of $P_{\max}=N_o/N_c\approx 3.2\times 10^{11}$ distinct computational paths is predicted to be obtained by this $q=20$ PWPCR implementation at a copy number of roughly $N_c(\text{opt})\approx 38$ copies per path. This value corresponds to the parallel implementation of roughly $N_{\text{ops}}=qP_{\max}\approx 6.4\times 10^{12}$ distinct computational operations, in a reaction volume of 400 μL.

G. Practical considerations

The current analysis is based on an equilibrium model of hairpin formation. Model predictions therefore apply strictly to experimental conditions at which $\bar{t}_e \gg \tau_r$. The potential for kinetic trapping due to unplanned intramolecular interaction complicates an estimation of τ_r, and is a particularly critical issue at reaction temperatures much less than the melting temperature (T_m) of backhybridized hairpins. However, the reaction temperatures suggested for WPCR and PWPCR (77 and 60 °C, respectively) are both approximately equal to the estimated melting temperatures of the backhybridized hairpins in the respective architectures. Furthermore, hairpins used are typically encoded to minimize unplanned sequence similarity. These two strategies were regarded to be sufficient to guarantee the negligibility of kinetic trapping due to unplanned sequence similarity for the implementations investigated. Relaxation times for PWPCR were then estimated by application of the simple model of hairpin kinetics discussed in Sec. III F. Mean hairpin loop sizes for a PWPCR implementation of a 20-step path range from 430 bases during the first extension process, to 1000 bases, at completion. Estimated relaxation times ranged from $\tau_2=14$ ms to $\tau_{20}=9.6$ ms. As the mean time between effective encounters for PWPCR ranged from $\bar{t}_e=35.7$ s ($\bar{N}_e=8.4$ encounters in 300 s) to $\bar{t}_e=6.7$ s ($\bar{N}_e=542.7$ encounters in 60 min), the assumption of equilibrium is expected to be valid for the range of reaction conditions considered.

Ensuring the consistency of the cationic bis-PNA targeting and polymerization processes (i.e., so that PNAs do not fall off at polymerization temperatures) is also critical. Cytosine-bearing, cationic bis-PNAs of the same length as those used in the text (i.e., complementary to a ten-nucleotide target sequence) have been reported to melt at approximately 85 °C at 0.1 M Na^+, with a very narrow melting transition [22]. As this T_m value is 25 °C higher than the predicted optimal PWPCR polymerization temperature of 60 °C, the suggested PWPCR protocol is regarded to adequately avoid significant melting of the PNA_2/DNA complexes formed during the bis-PNA treatment.

Although preventing a high occupancy of backhybridized structures is critical to achieving efficient computation, a variety of secondary factors also exist, which are likely to have an effect on overall PWPCR performance in practice. For instance, estimated DNA-enzyme encounter rates assume the optimal performance of the DNA polymerase employed. One advantage of employing a thermostable enzyme such as *Taq* DNA polymerase is the reported maintenance of enzyme activity with long exposure to denaturing temperatures. Based on the manufacturer's estimates of the temperature dependence of enzyme half-life ($t_{1/2}=10$ min and $t_{1/2}>130$ min at 97.5 and 92.5 °C, respectively [14]), the application of the polymerization protocol suggested for PWPCR is unlikely to result in a significant decrease in enzyme performance. However, the use of *Taq* DNA polymerase also carries with it the penalty of a modestly reduced activity at the reaction temperature suggested above (a 50% reduction at 60 °C [15]). In practice it is therefore likely that a modest increase in enzyme concentration, over that suggested above (or equivalently an increased polymerization time) will be required in order to obtain the predicted performance.

Another practical issue of importance is the potential for reduced extension efficiency due to the presence of suboptimal, but marginally stable unplanned hairpin structures. As noted above, WPCR strands are typically encoded specifically to minimize the potential for unplanned secondary structure. Recent statistical thermodynamic simulations of

DNA hybridization fidelity suggest that a very low occupancy of error configurations ($<10^{-6}$) can be assured for oligonucleotide-length annealing, by the application of a variety of encoding strategies [27]. It is therefore likely that the primary effect of error hybridization will be upon the fidelity, rather than the efficiency, of the PWPCR computational process, which is a related but separate issue.

V. CONCLUSION

In this paper, the impact of backhybridization on the extension efficiency of a DNA-based autonomous molecular computer, whiplash PCR was discussed within the framework of a Markov chain model of recursive polymerase extension. The extension probability per effective polymerase-DNA encounter was estimated for each strand length by using a statistical thermodynamic model of DNA duplex formation. The direct application of this model was demonstrated to provide close agreement with recent experimental observations, which report the premature failure of an eight-step, single-path WPCR implementation. The scaling behavior of the overall efficiency of WPCR was then discussed, and predicted times to completion were shown to be long enough to render WPCR-based massive parallelism infeasible. In an effort to enhance the efficiency of computation by reducing the impact of backhybridization, a modified architecture, PWPCR, was introduced that enables the specific inhibition of backhybridized structures through targeted PNA_2/DNA triplex formation. Application of this protocol under realistic reaction conditions is predicted to dramatically increase both the per encounter and overall computational efficiency, and to allow the practical implementation of massively parallel, autonomous computation.

ACKNOWLEDGMENTS

The authors are grateful to A. Nishikawa of the Osaka Electro-Communication Junior College, M. Arita of the Tokyo Electrotechnical Laboratory, and M. Miller and K. Komiya of the University of Tokyo, for critical reviews and helpful corrections of the manuscript. The authors also wish to thank M. Takano and K. Hamaguchi of the University of Tokyo for providing assistance during the preparation of figures. Financial support was provided by the Japan Society for the Promotion of Science (JSPS) ''Research for the Future'' Program (JSPS-RFTF 96I00101), and by a JSPS Grant-in-Aid.

[1] M. Hagiya *et al.*, in *DNA Based Computers III*, edited by H. Rubin and D. Wood (American Mathematical Society, Providence, RI, 1999), p. 57.
[2] L. Adleman, Science **266**, 1021 (1994).
[3] R. Lipton, Science **268**, 542 (1995).
[4] K. Sakamoto *et al.*, Biosystems **52**, 81 (1999).
[5] K. Komiya *et al.*, in Ref. [28], p. 19.
[6] D. Bovet and P. Crescenzi, *Introduction to the Theory of Complexity* (Prentice-Hall, Herfordshire, 1994).
[7] S. Ross, *Introduction to Probability Models*, 7th ed. (Academic, San Diego, 2000).
[8] A. Benight *et al.*, J. Mol. Biol. **200**, 377 (1988).
[9] R. Wartell and A. Benight, Phys. Rep. **126**, 67 (1985).
[10] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), p. 16.
[11] M. Innis *et al.*, Proc. Natl. Acad. Sci. U.S.A. **85**, 9436 (1988).
[12] J. G. Wetmur, in *DNA Based Computers III*, edited by H. Rubin and D. Wood (American Mathematical Society Providence, RI, 1999), p. 1.
[13] J. SantaLucia, Jr., Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[14] http://www.pebio.com/pc/catalog/pg9.html
[15] J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989), Vol. 2, pp. 5.50, 14.15.
[16] G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. U.S.A. **95**, 8602 (1998).
[17] A. Lomakin and M. D. Frank-Kamenetskii, J. Mol. Biol. **276**, 57 (1998).
[18] D. Cherny *et al.*, Proc. Natl. Acad. Sci. U.S.A. **90**, 1667 (1993).
[19] S. Kim *et al.*, J. Am. Chem. Soc. **115**, 6477 (1993).
[20] V. Demidov *et al.*, Proc. Natl. Acad. Sci. U.S.A. **92**, 2637 (1995).
[21] L. Betts *et al.*, Science **270**, 1838 (1995).
[22] M. Griffith *et al.*, J. Am. Chem. Soc. **117**, 831 (1995).
[23] H. Kuhn *et al.*, Nucleic Acids Res. **26**, 582 (1998).
[24] O. Almarsson *et al.*, Proc. Natl. Acad. Sci. U.S.A. **90**, 7518 (1983).
[25] Y. Kosagonov *et al.*, Biochemistry **39**, 11742 (2000).
[26] M. Fixman and J. Friere, Biopolymers **16**, 2693 (1977).
[27] J. Rose and R. Deaton, in Ref. [28], p. 231.
[28] *DNA Computing*, edited by A. Condon and G. Rozenberg (Springer-Verlag, Berlin, 2001); available by anonymous ftp at http://link.springer-ny.com/link/service/series/0558/tocs/t2054.htm#toc2054.