Kinetics of Repeat Propagation in the Microgene Polymerization Reaction

Mark Itsko,[†*] Avinoam Rabinovitch,[‡] and Arieh Zaritsky[†]
[†]Department of Life Sciences and [‡]Department of Physics, Ben-Gurion University of the Negev, Be'er-Sheva 84105, Israel

ABSTRACT Repetitive DNA is a periodic copolymer with the intrinsic property of exponential propagation to longer repeats. Microgene polymerization reaction (MPR) is a model system in which a short nonrepetitive homo-duplex DNA evolves to multiple repetitive products during heat-cool cycles. The mechanism underlying this process involves staggered annealing of complementary DNA strands of variable lengths and polymerase-mediated filling-in of the generated overhangs. MPR is considered here as a process sharing common features with two polymerization types, chain-growth and step-growth, and significant distinctions from both types were highlighted. The involved reaction stages were formulated and a kinetic model was derived and tested experimentally. The model can quantitatively explain MPR propagation and be used as a good approximation for this phenomenon.

INTRODUCTION

DNA is a natural irregular copolymer (1) comprised of four types of basic monomer units, the nucleoside monophosphates, and contains essential genetic information (2). It is built as two complementary strands in a double helical mode that are associated by multiple hydrogen bonds according to the Watson-Crick rules (3). The genetic information passes to successive generations by accurate DNA duplication, accomplished via the template-driven process in which nucleotides are assembled by polymerase on one strand thus reproducing the complementary one (2).

Repetitive DNA (containing multiple oligonucleotide repeats), an example of periodic copolymer (1), is ubiquitous in genomes of eukaryotes (4). It plays an important role in both maintaining chromosome integrity by telomeres (5) and chromosome seggregation by centromeres (6). Expansion of DNA repeats is associated with a variety of human heredity diseases (7) and may reflect the mechanism underlying primordial molecular evolution of primitive DNA sequences into complex genomes (8,9) that are known to include numerous periodicities also encrypted in the encoded proteins (10). The number of repeats in repetitive DNA is prone to expand during replication because its constituent strands slide over each other between the multiple complementary regions (11). Thermodynamically unfavorable structures bulging out from DNA duplex that accompany the strand sliding process can be stabilized by the inner base-pairs which facilitate expansion (12). The genomic repeat-expansion can be simulated in vitro with short repetitive homo-huplexes (HD) and thermophilic DNA polymerases under isothermal conditions (13–16). However, whereas ensemble of various enzymes facilitates repeat-expansion in vivo, high temperature (near the melting temperature of expanding DNA) is the facilitating factor in in vitro repeat-expansion (15). Expansion with nonrepetitive HD was observed in heat-cool cyclers and termed microgene polymerization reaction (MPR) (17). This process was successfully used to produce artificial proteins containing repetitive motifs with useful properties (18–20) or to expose cryptic activities of inactive proteins (21), but a deep understanding of the underlying mechanisms is lacking.

To initiate the process, a nonrepetitive HD is duplicated head-to-tail (17), thus generating a minimum repetitive unit termed initial doublet (ID) (22) that is prone to the succeeding expansion (Fig. 1 *A*). This rare process putatively involves bridging of two molecules of nonrepetitive HD by a third in a manner allowing the DNA polymerase to skip the inter-template gap (22). The ID can be amplified by the original HD before overall expansion starts ((23) and Fig. 1 *B*). The expansion includes staggered annealing of repetitive single strands of varied lengths followed by polymerization that fills in overhangs ((15) and Fig. 1 *C*).

The two MPR stages, initiation and propagation, are likened here for the first time to chain-growth and step-growth types of polymerization, respectively. Both stages of the MPR are formulated; the derived kinetic model is tested experimentally and discussed.

The model

The model considered, not limited by nucleotide concentrations [dNTP], was devised with the following variables and parameters:

- $A_i(N)$ and $B_j(N)$ are the concentrations of forward and reverse complementary DNA strands containing i and j repeats, respectively, at a given MPR cycle N.
- $\Delta A_n(N)$ $[= A_n(N+1) - A_n(N)]$ and ΔB_n $[= B_n(N+1) - B_n(N)]$ are increments in $A_n(N)$ and $B_n(N)$, respectively, after one cycle.
- k_I is the constant rate of MPR initiation.

Submitted September 19, 2008, and accepted for publication October 29, 2008.

*Correspondence: itskom@niehs.nih.gov

Mark Itsko's present Address: Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709.

Editor: Kathleen B. Hall.

0006-3495/09/03/1866/9 $2.00

doi: 10.1016/j.bpj.2008.10.061

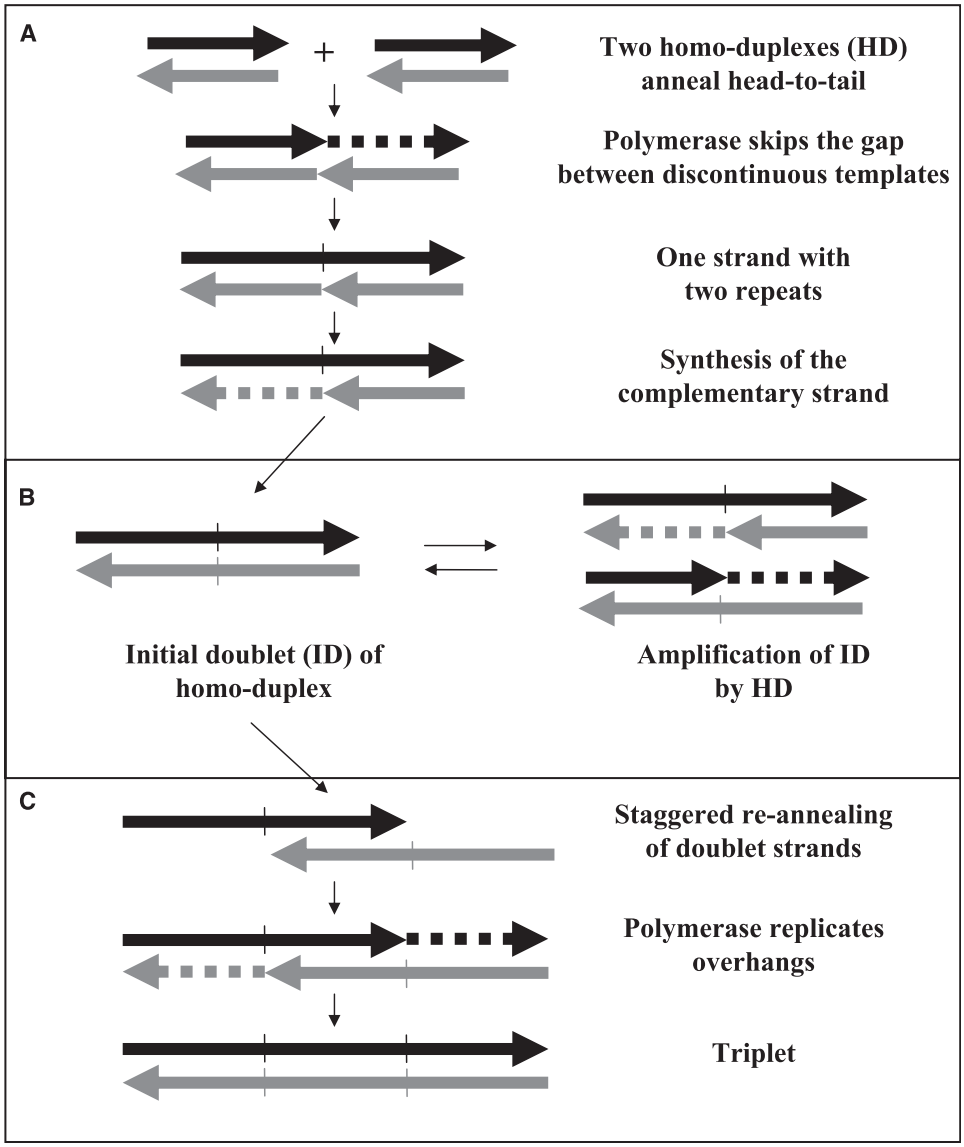


FIGURE 1 A model for MPR. (*A*) Initiation. (*B*) The initiator (ID) and its amplification. (*C*) The propagation of ID in the first heat/cool cycle.

k_{Ampl} is the constant rate of ID amplification by the original HD.

k_{Pr} is the constant rate of MPR repeat propagation per PCR cycle, assumed to be independent of polymer length (measured in repeat units n).

Initiation and amplification

The MPR is initiated when two HD combine to generate an ID (Fig. 1 *A*) by an intricate mechanism (22). The modeled propagation behavior is not sensitive to the exact mechanism of initiation; for the sake of simplicity, a molecularity of 2 was assumed for each of complementary DNA strands. The equation formulating this simplified process (Fig. 1 *A*) is therefore:

$$2A_1 + 2B_1 \xrightarrow{k_I} A_2 + B_2. \quad (1)$$

The initiator (ID) composed of A_2 and B_2 can be amplified (23) by the original HD (composing of A_1 and B_1) (Fig. 1 *B*) according to:

$$A_1 + B_2 \xrightarrow{k_{\mathrm{Ampl}}} A_2 + B_2 \text{ and } A_2 + B_1 \xrightarrow{k_{\mathrm{Ampl}}} A_2 + B_2. \quad (2)$$

Amplification of ID rapidly brings the mass concentration of the initiator to that of the original HD (Appendix A). Assuming that $k_{\mathrm{Ampl}} = k_{\mathrm{Pr}}$, the amplification stage is kinetically included in the following propagation stage.

Propagation (Fig. 1 C)

After generation of the initial doublet, the number of repeats is envisioned as expanding according to

$$A_i + B_j \xrightarrow{k_{\mathrm{Pr}}} A_n + B_n, \quad (3)$$

where $n \leq i + j - 1$ (explained in Appendix B).

Equation 3 describes an irreversible bimolecular process that can be expressed as:

$$\Delta A_n(N) = k_{\mathrm{Pr}} \sum_{i,j} f_{\mathrm{in}}(i,j) A_i(N) B_j(N) - k_{\mathrm{Pr}} A_n(N) \times \sum_j f_{\mathrm{out}}(j) B_j(N) \tag{4}$$

and

$$\Delta B_n(N) = k_{\mathrm{Pr}} \sum_{i,j} f_{\mathrm{in}}(i,j) A_i(N) B_j(N) - k_{\mathrm{Pr}} B_n(N) \times \sum_j f_{\mathrm{out}}(j) A_j(N), \tag{5}$$

where $f_{\mathrm{in}}(i, j)$, and $f_{\mathrm{out}}(j)$ (Appendix B) are the probability functions of the alignment between A_i and B_j that yield inflow and outflow of A_n (or B_n), respectively. Note that in this model both f_{in} and f_{out} do not depend on n. For the sake of simplicity, all possible pairings between A_i and B_j are assumed to have an equal chance to occur. The summation is carried out over all (i, j) pairs that together allow the generation of a product of length n. The following set of difference equations is thus derived for ΔA_n, considering also the doublet formation in the A_2 inflow and A_1 outflow, Eq. 1:

For $n = 1$,

$$\Delta A_1(N) = -k_{\mathrm{Pr}} A_1(N) \sum_{j=2}^{\infty} \frac{(j-1) B_j(N)}{j} - k_l A_1(N)^2 B_1(N)^2. \tag{6}$$

For $n = 2$,

$$\Delta A_2(N) = k_{\mathrm{Pr}} A_1(N) \sum_{i=2}^{\infty} \frac{B_i(N)}{i} - k_{\mathrm{Pr}} A_2(N) \times \sum_{j=2}^{\infty} \frac{(j-1) B_j(N)}{j+1} + k_l A_1(N)^2 B_1(N)^2. \tag{7}$$

For $n > 2$,

$$\Delta A_n(N) = k_{\mathrm{Pr}} \sum_{i=1}^{n-1} A_i(N) \sum_{j=n-i+1}^{\infty} \frac{B_j(N)}{i+j-1} - k_{\mathrm{Pr}} A_n(N) \times \sum_{j=2}^{\infty} \frac{(j-1) B_j(N)}{n+j-1}. \tag{8}$$

A similar set of equations may be formulated for ΔB_n. Since the initial concentrations of the complementary DNA strands dealt with here are identical, $A_1(0) = B_1(0)$, the propagation kinetics of A_n and B_n are identical, hence $B_i = A_i$.

To consider the enzymatic nature of MPR (with a finite quantity of enzyme), the propagation must be specified so that the aligned complex of the complementary strands A_iB_j obtained in the process

$$A_i + B_j \overset{k_1, k_{-1}}{\longleftrightarrow} A_iB_j. \tag{9}$$

associates with the enzyme (polymerase) (Enz) to form the product

$$A_iB_j + \mathrm{Enz} \overset{k_2, k_{-2}}{\longleftrightarrow} A_iB_j\mathrm{Enz} \xrightarrow{k_3} A_n + B_n + \mathrm{Enz}. \tag{10}$$

k_1 and k_{-1} are rate constants of alignment and of melting, respectively; k_2 and k_{-2} are rate constants of association and of dissociation, respectively, and k_3 is the rate constant of the enzyme-driven polymerization (turnover number) for filling in $(n–i)$ repeats.

After applying steady-state kinetics (see Appendix C), MPR rate becomes

$$\Delta A_n(N) = \sum_{i,j} f_{\mathrm{in}}(i,j) G(A_i, B_j) - \sum_i f_{\mathrm{out}}(i) G(A_i, B_n), \tag{11}$$

where

$$G(A_i, B_j) = \frac{k_{\mathrm{Pol}} \mathrm{Enz}_{\mathrm{tot}} A_i(N) B_j(N)}{K_M^{\mathrm{app}} + A_i(N) B_j(N)}, \quad K_M^{\mathrm{app}} = K_M^{\mathrm{app}*} + K_D^{-1} K_M, \quad K_M^{\mathrm{app}*} = \mathrm{Enz}_{\mathrm{tot}} k_3 / k_1, \tag{12}$$

$\mathrm{Enz}_{\mathrm{tot}}$ is the total concentration of enzyme, $K_D = k_1/k_{-1}$ is the equilibrium constant for duplex formation, $K_M = (k_{-2} + k_3)/k_2$ is the Michaelis-Menten constant, and $f_{\mathrm{in}}(i, j)$ and $f_{\mathrm{out}}(j)$ are the terms used in Eqs. 4 and 5. When $K_M^{\mathrm{app}*} \gg A_i(N) B_j(N)$, the kinetics reduces to a simple bimolecular one as in Eq. 3, and k_{Pr} can be expressed by

$$k_{\mathrm{Pr}} = k_{\mathrm{Pol}} \mathrm{Enz}_{\mathrm{tot}} / K_M^{\mathrm{app}*} = k_1 \varepsilon, \tag{13}$$

where ε is defined in Appendix C.

Termination

MPR is never truly terminated, as happens in chain-growth-type polymerization reactions; it is rather finished when the nucleotides are depleted, as in step-growth-type reactions.

Extent of polymerization in MPR

The final polymer length in chain-growth-type reactions can be calculated from the ''kinetic chain length,'' defined as the number of monomer units consumed in the propagation stage per active center produced in the initiation stage (24,25). In the MPR, nucleotide concentration [dNTP] presented in the reaction mixture determines the extent of propagation, whereas that of the initial ID (equal to the nonrepetitive HD $[\mathrm{HD}]_0$) represents active centers (Eq. 2 and Appendix A). Thus, in MPR-produced multiple repetitive DNA, the final length

$$\langle n \rangle = (1/m) \times [\mathrm{dNTP}] / [\mathrm{HD}]_0, \tag{14}$$

where m is the number of nucleotides composing one HD. Equation 14 is concordant with the experimental results presented before (Fig. 2, *A* and *C*, in (22)).

MATERIALS AND METHODS

MPR conditions and length distribution of products

A 55-μl reaction mixture contained 8.8 pmol (160 nM) of EVNA HD (22), 500 μM of each of the dNTPs, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 8 mM $MgSO_4$, 0.1% Triton X-100 and 1 unit of Vent DNA polymerase. The following conditions for T-Gradient Thermoblock cycler (Biometra, Göttingen, Germany) were employed: 10 min at 94°C (for strand separation) and 10 min at 72°C (around the melting temperature T_m of the EVNA HD), then 30 cycles of 94°C for 10 sec and 72°C for 4 min. Length distributions of the products were determined after 14–17 cycles. At the midpoints (2 min) of five additional cycles, samples of 9 μl each were withdrawn, electrophoresed on 0.8% agarose gels with 0.5 μg ml^{-1} of Ethidium Bromide (EtBr), and photographed under UV illumination. The fluorescence intensity of EtBr in the digitized images (reflecting the product-length distributions) was analyzed using National Institutes of Health Object Image software (National Institutes of Health, Bethesda, MD)(http://rsb.info.nih.gov/nih-image/index.html). Each distribution of MPR lengths (in number of repeats n) was normalized (Peak Normalization) in two steps: it was divided by its maximum, and shifted by dividing each n by n_{max} (n at the peak).

Simulation of the MPR kinetics

The set of difference equations (limited to 3000) describing kinetics of MPR was solved numerically using Matlab 7 (MathWorks, Natrick, MA (http://www.mathworks.com)). The rate of exponential MPR propagation $\ln(1+E)$, where E is the amplification efficiency (the fraction of product added per cycle), is determined by the value of the maximum of the 1st derivative of $\ln(X_N)$, where X_N is the total amount of the MPR product after cycle N (assuming a continuous process). The adjustment of parameters is governed by the value of E (between 0.7–0.8) in both End-Point-Detection-PCR (Fig. 2 *C*) and Real-Time-PCR (22) and by the features of Vent polymerase (26), K_M = 0.1 nM, Enz_{tot} = 10 nM and k_3=16.7 s^{-1} for dNTP addition. This last value translates into [16.7/42] = 0.4 repeat sec^{-1} for addition of a 42 bp repetitive unit. Each distribution of MPR lengths (both calculated and experimental) was normalized (Integral Normalization) in two steps: it was divided by its area (integral), and shifted by dividing each n by n_{max} (n at the peak).

RESULTS

Propagation kinetics of the MPR products

To follow DNA propagation, samples were removed from the running MPR mixtures at consecutive cycles and analyzed: DNA products extended in length (Fig. 2, *A* and *B*) and the total amount rose exponentially (Fig. 2 *C*) with E of 0.75 ($e^{0.5575}$-1) to maximal length determined by [HD] (Eq. 14 and (22)). Overlaying Peak-Normalized forms of the length distributions (MPR conditions and length distribution of products) reveals that they are essentially identical (Fig. 2 *D*).

Model's predictions

The model discerns between different behaviors of the exponential expansion in MPR, depending on K_M^{app*}: the latter is either much larger than $\langle A_i(N)B_j(N)\rangle$ (large K_M^{app*}) or not (small K_M^{app*}). In both cases, it predicts a constant rate of total product-propagation (provided an unlimited supply of dNTP), but whereas the accumulation is exponential throughout when $K_M^{app*} \gg \langle A_i(N)B_j(N)\rangle$, it is finally replaced by a stationary phase in the alternative (small K_M^{app*}) case (Fig. 3 *A*): the decrease in rate is due to an ever-increasing

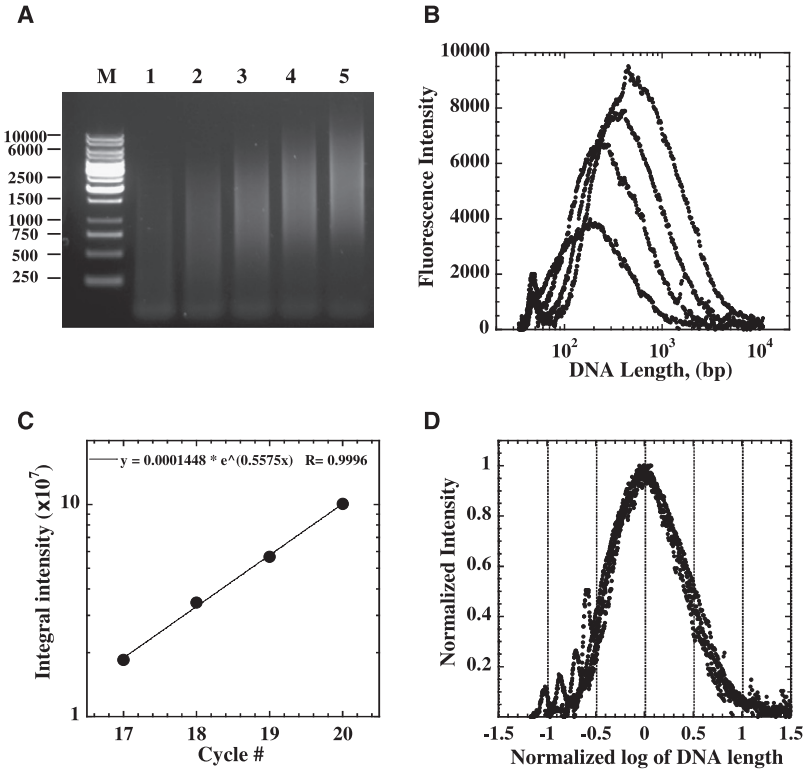


FIGURE 2 Propagation kinetics and distributions of MPR products. (*A*) Consecutive samples run on 0.8% agarose gel: M, DNA size markers; lane 1, cycle 16; lane 2, cycle 17; lane 3, cycle 18; lane 4, cycle 19; lane 5, cycle 20. (*B*) Scanned and digitized samples of cycles 17–20 (from *A*). (*C*) Total (integral of corresponding distributions) fluorescence intensity (DNA) in lanes 2–5 (cycles 17–20; from *A*). (*D*) Juxtaposed, peak normalized (MPR conditions and length distribution of products) distributions (from *B*).

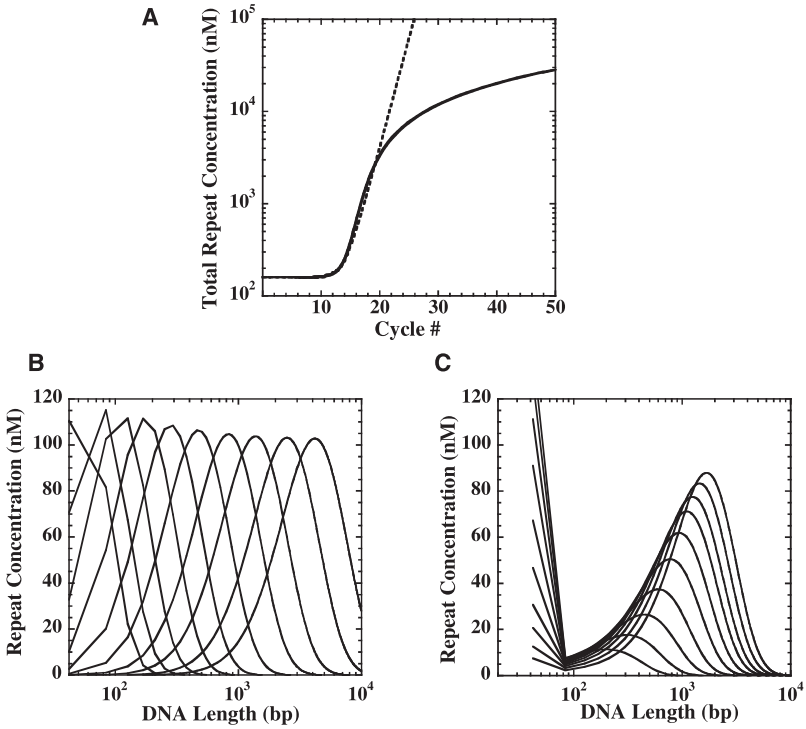


FIGURE 3 Matlab simulations. (*A*) Kinetics of total MPR product: dashed line, large K_M^{app*} case [$k_{Pol} = 96$ repeats cycle^{-1} ($k_3 = 0.4$ repeat sec^{-1}), $K_M^{app*} = 8\times10^4$ nM^2, k_I=1.5×10^{-15} nM^{-3} sec^{-1}; $E = 0.728$]; full line, small K_M^{app*} case [$k_{Pol} = 0.09$ repeat cycle^{-1} ($k_3 = 3.75\times10^{-4}$ repeat sec^{-1}), $K_M^{app*} = 1$ nM^2, $k_I = 10^{-24}$ nM^{-3} sec^{-1}; $E = 0.735$]. Final distributions at cycles 14–23: with large K_M^{app*} (*B*) and small K_M^{app*} (*C*) values.

value of $(K_D)^{-1}$ (reflecting gradual increase in T_m of the extending product) that finally exceeds K_M^{app*} (Eqs. 12 and C13). In addition, the value of the distribution maxima during the exponential phase is almost a constant according to the former case (Fig. 3 *B*), whereas it increases to a constant value at the stationary phase according to the latter (Fig. 3 *C*). The fact that the intensity at the distributions' peak rises (Fig. 2 *B*) is consistent with the small K_M^{app*} case only.

Comparing the experimental results with the model

The experimental results (Fig. 2 *B*) and the model's predictions (Fig. 3 *C*) resemble each other in that both display bimodal distributions. However, the ratios between the areas under the first (HD) peak to that of the main peak in each of the model distributions for early propagation cycles are much higher than those in the experimental ones. This inconsistency is considered in the Discussion Section, but superposition of the two by the Integral Normalization procedure (see Simulation of the MPR kinetics section), yielding Fig. 4 *A*, was performed without the theoretical HD peak.

DISCUSSION

The study presented here is the first attempt, to our knowledge, to describe the empiric phenomenon of MPR in terms of polymer chemistry. The model seems to explain the major quantitative aspects of MPR propagation and can be used to better approximate minor details of this phenomenon. Controlling the extent of MPR propagation via [dNTP] and [HD] allows a rational design of reaction mixtures to obtain higher yields of desirable bioengineered artificial polypeptides.

MPR as a chemical polymerization process

The MPR is kinetically divided into two stages, initiation (enhanced by amplification) and propagation (Fig. 1). The former is slower because the initiator (ID) is generated through an unstable nucleation complex (22,23). However, whereas in common chain-growth polymerization reactions that are limited by unstable intermediates (free radicals or activated precursors in chemical (24) or biological (27–29) macromolecules, respectively), the initiation of MPR (after a transient period) is not rate limiting because ID is stable. The rate-limiting stage here is propagation, the mechanism of which shares common features with step-growth polymerization processes (24,25): each extension step, Eq. 3, occurs between molecules containing any number of repeats, which gradually increases with cycles (Fig. 2). The common step-growth polymerization processes proceeding by condensation is of 2nd-order kinetics resulting in a linear growth of the average polymer molecular weight (25). The MPR, on the other hand, is an autocatalytic process resulting in an exponential growth of the number of repeats per polymer molecule (Fig. 2).

Exponential rise in the average number of repeats per molecule $\langle n\rangle_N$ during propagation is justified as follows. In the extension reaction (Eq. 3), two molecules of lengths n_1 and n_2 (i.e., $n_1 \le n_2$) yield two molecules with lengths in the range between n_1+1 and n_1+n_2-1. Assuming that the population of such repetitive products is uniformly distributed with a common difference of one repeat unit

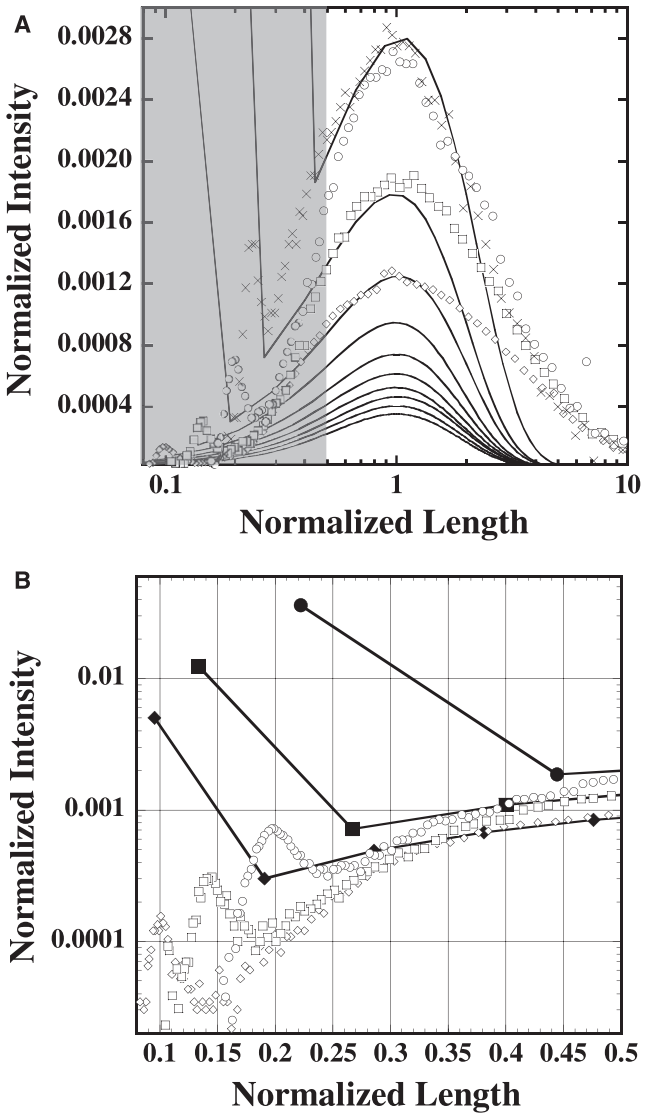


FIGURE 4 Experimental versus model distributions. (*A*) Juxtaposition of the small $K_M^{app^*}$ case model (solid *lines* for cycles 14–23) with four (cycles 17–20) experimental distributions (*crosses*, *open circles*, *open squares*, and *open diamonds*, respectively), each integrally normalized (see Simulation of the MPR kinetics section). The shaded area is expanded in (*B*): The model (*solid symbols* for cycles 14–16) and corresponding experimental (*open symbols*) distributions, excluding cycle No. 17.

(Appendix B), the average length of the product $\langle n\rangle = [(n_1+1)+(n_1+n_2-1)]/2 = n_1 + n_2/2$. Averaging over all possible reactions of this sort yields

$$\langle n\rangle_{N+1} = \langle n_1\rangle_N + \langle n_2\rangle_N/2. \qquad (15)$$

Since $\langle n_1\rangle_N = \langle n_2\rangle_N = \langle n\rangle_N$ (belonging to the same distribution), the average length of the polymer should increase by a factor of 1.5 at each cycle $\langle n\rangle_{N+1} = 3\langle n\rangle_N/2$, corresponding to $E = 0.5$. This propagation scheme, Eq. 3, remains unaffected by assuming that the reacting molecules are duplexes rather than single strands, as is the case in this treatise.

The experimental value of the amplification efficiency E (Fig. 2 *C* and (22)) exceeds 0.5, Eq. 15, implying (see Eq. 3) an overlap between two reacting molecules mainly at their 3′ terminal units rather than uniformly along their whole lengths (as in Appendix B). Unwinding of the reacting duplexes by strand displacement and their intrusion into each other, as does Vent polymerase (26), can explain this high E: shorter overlaps would be favored because stereo hindrance rises with deeper intrusions (see Fig. 6 in (15)). As a consequence, the average length of the product molecules is larger, resulting in $E > 0.5$. Moreover, each 4 min cycle at around the T_m used here is likely to allow more than a single extension step thus further enlarging the average length. Larger E values would thus be anticipated under longer heat-cool MPR cycles.

Adjustment of theoretical reaction-constants

The range of propagation efficiency E obtained here (0.7–0.8) with $k_3 = 0.4$ repeat sec^{-1} (see Simulation of the MPR kinetics section) and $k_{Pol} = 240$ (sec cycle^{-1})$\times k_3 = 96$ repeats cycle^{-1} (Appendix C) requires K_M^{app*} to be ~10^5 nM^2 ($\gg \langle A_i(N)B_j(N)\rangle$; Eqs. 12 and 13), reducing the Michaelis-Menten kinetics, Eqs. 9–11, to a simple bimolecular one, Eq. 3. Thus, the rise of the maximum value of the main peak (Fig. 3 *C*) would be eliminated (Fig. 3 *B*), which is totally inconsistent with the experimental results (Fig. 2 *B*), precluding the validity of this value of k_3 (26). Exclusion of the large K_M^{app*} case (see Model's predictions section) is further supported by a), comparing E values that it predicts at different [HD] with experimental data (Table 1) and b), deriving a low value for k_1 from Eq. 12 (5×10^4 M^{-1} sec^{-1}), far below the diffusion-controlled limit (k_D of between 10^7-10^8 M^{-1} sec^{-1}) derived from the Debye-Smoluchowski equation (30). Since the activation energy involved in hybridization between complementary regions is low, the overall reaction rate during an MPR cycle would be governed by the rate at which DNA molecules diffuse through the medium (25).

The experimental rise in peaks (Fig. 2 *B*) compels (Fig. 3 *C*) the DNA alignment constant k_1 and the enzyme turnover number k_3 to be equal to 3.75×10^6 M^{-1} sec^{-1} and

TABLE 1 Calculated *E* values

nM[†]	Experimental[‡]	large $K_M^{app^*}$[§]	small $K_M^{app^*}$[¶]
320	0.80 ± 0.05	1.92	0.6
160	0.80 ± 0.04	0.71	0.7
80	0.68 ± 0.16	0.31	0.9
40	0.57 ± 0.09	0.14	1.0

[†]Initial concentration of original homo-duplexes ([HD]$_0$).
[‡]Gleaned from (22).
[§]Calculated with the following parameters: $k_{Pol} = 96$ cycle^{-1} ($k_3 = 0.4$ repeat sec^{-1}), $K_M^{app^*} = 8\times10^4$ nM^2, $k_I = 1.5\times10^{-15}$ nM^{-3}sec^{-1}, Enz$_{tot}$ = 10 nM, $K_M = 0.1$ nM.
[¶]Calculated with the following parameters: k_{Pol}=0.09 cycle^{-1} ($k_3 = 3.75\times10^{-4}$ repeat sec^{-1}), $K_M^{app^*} = 1$ nM^2, $k_I = 10^{-24}$ nM^{-3} sec^{-1}, Enz$_{tot}$ = 10 nM, $K_M = 0.1$ nM.

3.75×10^{-4} repeat sec^{-1}, respectively. While k_I is sufficiently close to k_D, the value of k_3 is 10^3-fold lower than that calculated value (0.4 repeat sec^{-1}) from existing data with Vent (26) (see also in the Simulation of the MPR kinetics section). This means that the average number of repeats added per cycle by an enzyme molecule is ~1000 (see Eq. 10).

Similarities and disparities between the model and experimental results

There is a satisfactory resemblance between the model's predictions and the results (Fig. 4) thus supporting it: a), both display bimodal distributions; b), values of all three HD peaks agree; c), three of the experimental length distributions (2nd–4th) overlap the first three of the model at their main peaks. To enhance resolution and comprehension of this issue, the lower end of the integrally normalized distribution curves (*shaded area* in Fig. 4 *A*) was expanded with different scales (Fig. 4 *B*).

Three discrepancies between the derived model and the experimental distributions are evident but can be resolved by testable explanations (Fig. 4): a), the overlap between the first two product-distributions after initiation of MPR is explained by the preceding amplification ((23) and Fig. 1 *B*). The kinetics of this stage that corresponds to early MPR cycles is different than that of the net propagation, depending on the difference between k_{Ampl} and k_{Pr}. Since the product of the first cycle (*crosses* in Fig. 4) includes also the parallel amplification stage, it should be excluded from the comparison between the experimental and the modeled MPR propagation distributions. b), the very low intensities of the HD peaks in the experimental distributions can be explained by the leaching of EtBr from DNA during electrophoresis to the cathode during run on gel (Fig. 2 *B*). Shorter DNA species such as HD are more vulnerable to EtBr leaching and consequently to an underestimate of their real amount. The same reasoning is valid for the shortest repeats in all distributions thereby explaining the disparity at their left hand part. This disparity decreases as the length increases, validating the assertion of the vulnerability of EtBr leaching. c), the experimental distributions are skewed to longer lengths whereas the model predicts negative skewness. This means that for every (A_i, B_j) involved in the extension step (see Eqs. 3 and 10) the generated product is much shorter than the expected maximum ((*i*+*j*-*1*) in Appendix B). This is consistent with lower processivity of the Vent polymerase (26).

APPENDIX A: ASYMPTOTICS

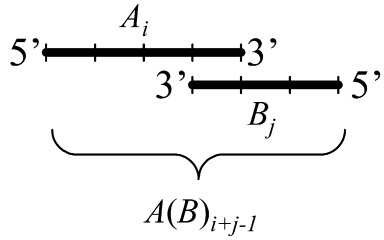
For the sake of simplicity of the consequent evaluation, initiation and amplification stages are formulated without dissecting each HD into the single strands composing it.

Initiation: $2\mathrm{HD} \xrightarrow{k_I} D$; Amplification: $\mathrm{HD} + D \xrightarrow{k_{\mathrm{Ampl}}} 2D$, where HD is the original nonrepetitive HD and D is the head-tail doublet of HD. Thus, the rate of D accumulation is: $d[D]/dt = k_I[\mathrm{HD}]^2 + k_{\mathrm{Ampl}}[\mathrm{HD}][D]$. Expressing [HD] via $[D]$, the equation is rearranged to: $d[D]/dt = k_I([\mathrm{HD}]_0 - [D])^2 + k_{\mathrm{Ampl}}([\mathrm{HD}]_0 - [D])[D]$, where $[\mathrm{HD}]_0$ is the initial concentration of HD. Integration of this equation yields: $[D] = k_I[\mathrm{HD}]_0(e^{k_{\mathrm{Ampl}}[\mathrm{HD}]_0 t} - 1)/(k_{\mathrm{Ampl}} - k_I + k_I e^{k_{\mathrm{Ampl}}[\mathrm{HD}]_0 t})$. Since $k_{\mathrm{Ampl}} \gg k_I$, this expression describes exponential increase in $[D]$ that asymptotically approaches the value of $[\mathrm{HD}]_0$.

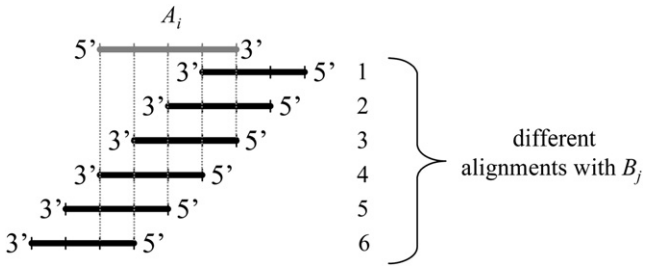
APPENDIX B

Inflow

The length n of the strand generated by reactants A_i and B_i is at most $(i + j - 1)$, since at least one repeat is always hidden in the overlap paired region:



The value $(i + j - 1)$ is also the total number of overlap types between two strands:



Due to the fact that chain growth continues from 3′ termini, only one of the above alignments leads to generation of A_n product of desired length n, whereas alignments numbers 4–6 yield no product at all. From the above alignments, only number 1 generates A_6. Correspondingly, the probability of the specific alignment $(f_{\mathrm{in}}(i, j)) = (i + j - 1)^{-1}$.

Outflow

The outflow of A_n is put into effect in longer products only because the model ignores exonucleolytic activity of DNA polymerase. Since at least one repeat unit in B_j must be 5′-overhanged to allow lengthening of A_n, $(j - 1)$, different alignments of B_j with A_n out of a total $(n + j - 1)$ (see the previous paragraph) lead A_n to extend to a group of longer lengths, $f_{\mathrm{out}}(j) = (j - 1) / (n + j - 1)$.

APPENDIX C: ANALYSIS OF THE KINETIC EQUATIONS

All the processes formulated below occur in the same cycle, hence cycle number index is omitted from concentration symbols and they are put into square brackets.

The mathematical description of the kinetic Eqs 9–11 may be written

$$\frac{d[A_iB_j]}{dt} = k_1[A_i][B_j] - k_{-1}[A_iB_j] - k_2[A_iB_j] \times [\mathrm{Enz}] + k_{-2}[A_iB_j\mathrm{Enz}], \qquad \text{(C1)}$$

$$\frac{d[A_iB_j\mathrm{Enz}]}{dt} = k_2[A_iB_j][\mathrm{Enz}] - k_{-2}[A_iB_j\mathrm{Enz}] - k_3[A_iB_j\mathrm{Enz}], \quad \text{(C2)}$$

and

$$\frac{d[A_{n(i,j)}]}{dt} = k_3[A_iB_j\mathrm{Enz}], \quad \text{(C3)}$$

where $A_{n(i,\ j)}$ is A_n generated from the given i and j and k_3 is the turnover number of the enzyme.

The rate of generation of intermediates such as $[A_iB_j]$ and $[A_iB_j\mathrm{Enz}]$ is taken as close to zero according to the assumption of steady-state kinetics for them (25). Therefore:

$$k_1[A_i][B_j] - k_{-1}[A_iB_j] - k_2[A_iB_j][\mathrm{Enz}] + k_{-2}[A_iB_j\mathrm{Enz}] = 0, \quad \text{(C4)}$$

and

$$k_2[A_iB_j][\mathrm{Enz}] - k_{-2}[A_iB_j\mathrm{Enz}] - k_3[A_iB_j\mathrm{Enz}] = 0. \quad \text{(C5)}$$

From Eq. C5:

$$(k_{-2} + k_3)[A_iB_j\mathrm{Enz}] = k_2[A_iB_j][\mathrm{Enz}], \quad \text{(C6)}$$

and

$$[A_iB_j] = \frac{[A_iB_j\mathrm{Enz}](k_{-2} + k_3)/k_2}{[\mathrm{Enz}]}. \quad \text{(C7)}$$

[Enz] can be expressed as

$$[\mathrm{Enz_{tot}}] - [A_iB_j\mathrm{Enz}], \quad \text{(C8)}$$

so that

$$[A_iB_j] = \frac{[A_iB_j\mathrm{Enz}](k_{-2} + k_3)/k_2}{[\mathrm{Enz_{tot}}] - [A_iB_j\mathrm{Enz}]}. \quad \text{(C9)}$$

Inserting Eqs. C8 and C9 into Eq. C4 yields the following after simplification:

where $k_3[A_iB_j\mathrm{Enz}]^2 - \alpha \times [A_iB_j\mathrm{Enz}] + \beta = 0, \quad \text{(C10)}$

$$\alpha = k_1[A_i][B_j] + k_3[\mathrm{Enz_{tot}}] + \frac{k_{-1}}{k_2}(k_{-2} + k_3) \text{ and}$$
$$\beta = k_1[A_i][B_j][\mathrm{Enz_{tot}}].$$

Solving this equation produces two roots:

$$[A_iB_j\mathrm{Enz}]_{1,2} = 0.5\left(\alpha \pm \sqrt{\alpha^2 - 4\beta k_3}\right)/k_3. \quad \text{(C11)}$$

The only root satisfying the condition $[\mathrm{Enz_{tot}}] - [A_iB_j\mathrm{Enz}] \geq 0$, that the amount of the enzyme joined to DNA should be smaller than its total amount, is

$$[A_iB_j\mathrm{Enz}] = 0.5\left(\alpha - \sqrt{\alpha^2 - 4\beta k_3}\right)/k_3. \quad \text{(C12)}$$

The melting equilibrium constant K_D is concentration-dependent because linear nucleic acid duplexes are formed from two strands. $K_D = 2\theta_{\mathrm{ext}}/\left((1-\theta_{\mathrm{ext}})^2 C_T\right)$, where θ_{ext} is the fraction of strands with at least one intact basepair and C_T is the total concentration of strands forming DNA duplex (31). θ_{ext} at the melting temperature is equal to 0.5, so K_D at this condition is equal to $4/C_T$. Since MPR products are repetitive DNA containing basic primer motive, C_T may be evaluated as the total concentration of MPR repeats that increases with cycles:

$$K_D^{-1} = 0.25\sum_{i=1}^{\infty} i([A_i(N)] + [B_i(N)]). \quad \text{(C13)}$$

Under the conditions of the used [HD] (Table 1) and values of K_M (0.1 nM), k_3 (0.4 repeat sec^{-1}) and [$\mathrm{Enz_{tot}}$] (10 nM) (26),

$$\alpha^2 \gg 4\beta k_3 \text{ or } 4\beta k_3/\alpha^2 \ll 1. \quad \text{(C14)}$$

We assume that this inequality stays true along the process and hence Eq. C12 may be simplified as follows:

$$[A_iB_j\mathrm{Enz}] = 0.5\alpha\left(1 - \sqrt{1 - \frac{4\beta k_3}{\alpha^2}}\right)/k_3 \approx 0.5\alpha \times \left(1 - \left(1 - \frac{2\beta k_3}{\alpha^2}\right)\right)/k_3 = \frac{\beta}{\alpha} \quad \text{(C15)}$$

because $\sqrt{1-x} \approx 1 - 0.5x$ as x approaches to 0.

$$[A_iB_j\mathrm{Enz}] = \frac{[\mathrm{Enz_{tot}}][A_i][B_j]}{\mathrm{Enz_{tot}}k_3/k_1 + K_D^{-1}K_M + [A_i][B_j]} \quad \text{(C16)}$$

so that

$$\frac{dA_{n(i,j)}}{dt} = \frac{k_3[\mathrm{Enz_{tot}}][A_i][B_j]}{\mathrm{Enz_{tot}}k_3/k_1 + K_D^{-1}K_M + [A_i][B_j]}. \quad \text{(C17)}$$

Assuming $t = N\varepsilon$, where N is the number of cycles and ε is the cycle period, yields

$$\Delta A_{n(i,j)} = \frac{k_{\mathrm{Pol}}[\mathrm{Enz_{tot}}][A_i][B_j]}{\mathrm{Enz_{tot}}k_3/k_1 + K_D^{-1}K_M + [A_i][B_j]}, \quad \text{(C18)}$$

where $k_{\mathrm{Pol}} = k_3\varepsilon$, and $\varepsilon = 240$ sec cycle^{-1}.

We thank Eitan Ben-Dov for useful advice and Norbert Visher for assistance in National Institutes of Health Object Image analysis.

This investigation was partially supported by an Eshkol Fellowship for PhD Studentship, Ministry of Science, Jerusalem, Israel (to M.I.).

REFERENCES

1. IUPAC 1996. Glossary of Basic Terms in Polymer Science. *Pure Appl. Chem.* 68:2287–2311.
2. Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, et al. 2002. Molecular Biology of the Cell. 4th ed, Garland Science, New-York, pp. 4–6.
3. Watson, J., and F. Crick. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 171:737–738.
4. Britten, R. J., and D. E. Kohne. 1968. Repeated sequences in DNA. *Science*. 161:529–540.
5. Blackburn, E. H. 1991. Structure and function of telomeres. *Nature*. 350:569–573.
6. Catasti, P., G. Gupta, A. E. Garcia, R. Ratliff, L. Hong, et al. 1994. Unusual structures of the tandem repetitive DNA sequences located at human centromeres. *Biochemistry*. 33:3819–3830.

7. Mirkin, S. M. 2007. Expandable DNA repeats and human disease. *Nature*. 447:932–940.

8. Ohno, S. 1984. Birth of unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence. *Proc. Natl. Acad. Sci. USA*. 81:2421–2425.

9. Ohno, S. 1987. Evolution from primordial oligomeric repeats to modern coding sequences. *J. Mol. Evol.* 25:325–329.

10. Shiba, K., Y. Takahashi, and T. Noda. 2002. On the role of periodism in the origin of proteins. *J. Mol. Biol.* 320:833–840.

11. Harvey, S. C. 1997. Slipped structures in DNA triplet repeat sequences: entropic contributions to genetic instabilities. *Biochemistry*. 36:3047–3049.

12. Kang, S., A. Jaworski, K. Ohshima, and R. D. Wells. 1995. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nat. Genet.* 10:213–218.

13. Ogata, N., and T. Miura. 2000. Elongation of tandem repetitive DNA by the DNA polymerase of the hyperthermophilic archaeon *Thermococcus litoralis* at a hairpin-coil transitional state: a model of amplification of a primordial simple DNA sequence. *Biochemistry*. 39: 13993–14001.

14. Ogata, N., and H. Morino. 2000. Elongation of repetitive DNA by DNA polymerase from a hyperthermophilic bacterium *Thermus thermophilus*. *Nucleic Acids Res*. 28:3999–4004.

15. Tuntiwechapikul, W., and M. Salazar. 2002. Mechanism of in vitro expansion of long DNA repeats: effect of temperature, repeat length, repeat sequence, and DNA polymerases. *Biochemistry*. 41:854–860.

16. Liang, X., K. Jensen, and M. D. Frank-Kamenetskii. 2004. Very efficient template/ primer-independent DNA synthesis by thermophilic DNA polymerase in the presence of a thermophilic restriction endonuclease. *Biochemistry*. 43:13459–13466.

17. Shiba, K., Y. Takahashi, and T. Noda. 1997. Creation of libraries with long ORFs by polymerization of a microgene. *Proc. Natl. Acad. Sci. USA*. 94:3805–3810.

18. Shiba, K. 2004. MolCraft: a hierarchical approach to the synthesis of artificial proteins. *J. Mol. Catal., B Enzym.* 28:145–153.

19. Saito, H., T. Minamisawa, and K. Shiba. 2007. Motif programming: a microgene-based method for creating synthetic proteins containing multiple functional motifs. *Nucleic Acids Res*. 35:10.1093/nar/gkm017.

20. Shiba, K., and T. Minamisawa. 2007. A synthesis approach to understanding repeated peptides conserved in mineralization proteins. *Biomacromolecules*. 8:2659–2664.

21. Itsko, M., and A. Zaritsky. 2007. Exposing cryptic antibacterial activity in Cyt1Ca from *Bacillus thuringiensis israelensis* by genetic manipulations. *FEBS Lett*. 581:1775–1782.

22. Itsko, M., A. Zaritsky, A. Rabinovitch, and E. Ben-Dov. 2008. Initiation of the microgene polymerization reaction with non-repetitive homoduplexes. *Biochem. Biophys. Res. Commun.* 368:606–613.

23. Itsko, M., A. Zaritsky, and A. Rabinovitch. 2008. Thermodynamics of unstable DNA structures from kinetics of the Microgene PCR. *J. Phys. Chem. B*. 112:13149–13156.

24. Flory, P. J. 1953. Principles of Polymer Chemistry. Cornell University Press.

25. Atkins, P. W. 1994. Physical Chemistry, 5th ed. Oxford University Press, pp. 861–959.

26. Kong, H., R. B. Kucera, and W. E. Jack. 1993. Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J. Biol. Chem.* 268:1965–1975.

27. Manor, H., D. Goodman, and G. S. Stent. 1969. RNA chain growth rates in *Escherichia coli*. *J. Mol. Biol.* 39:1–29.

28. Manor, H., M. P. Deutscher, and U. Z. Littauer. 1971. Rates of DNA chain growth in *Escherichia coli*. *J. Mol. Biol.* 61:503–524.

29. Wilhelm, J. M., and R. Haselkorn. 1970. The chain growth rate of lysozyme T4 *in vitro*. *Proc. Natl. Acad. Sci. USA*. 65:388–394.

30. Berg, O. G., R. B. Winter, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*. 20:6929–6948.

31. Owczarzy, R., P. M. Vallone, F. J. Gallo, T. M. Paner, M. J. Lane, et al. 1997. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers*. 44:217–239.