ELSEVIER

Biophysical Chemistry 106 (2003) 173–178

Biophysical Chemistry

www.elsevier.com/locate/bpc

A theory on the origin of cooperativity in DNA renaturation kinetics

R. Murugan*

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India

Received 29 May 2003; received in revised form 27 June 2003; accepted 29 June 2003

Abstract

Theoretical proof for the existence of cooperativity (experimental evidences have already been reported in our earlier work: Biochem. Biophys. Res. Commun., 293 (2002) 870) in the phenomenon of DNA renaturation has been obtained using a generalized birth–death master equation approach. Results have shown that no special searching mechanism (i.e. a facilitated mechanism) was necessary for complementary DNA strands to find their correct-contacts but searching by a 'random jump' method was enough to explain not only the observed cooperativity phenomenon but also the magnitude of correct-contact-forming rate constant. Impact of cooperativity phenomenon on length of primers and their specificity in polymerase chain reaction has also been discussed.

Keywords: DNA renaturation; Cooperativity; Polymerase chain reaction; Master equation; Random jump

1. Introduction

The renaturation kinetics of complementary strands of DNA is an important aspect in molecular biology, especially in C_0t analysis [1–3], polymerase chain reaction (PCR) and denaturing high-performance liquid chromatography [4]. Several mathematical models on kinetics of renaturation of DNA in solution have already been developed and experimentally tested [5–8]. Conventionally, the renaturation of ssDNA was described by a second order kinetics scheme as follows:

$$2[\mathrm{ssDNA}] \xrightarrow{k_1} [\mathrm{dsDNA}]$$

Here k_1 (mol^{-1} s^{-1}) is bimolecular rate constant, [ssDNA] is concentration (mol l^{-1}) of single-stranded DNA and [dsDNA] is concentration (mol l^{-1}) of double stranded DNA. The corresponding differential (rate) equation can be written as follows:

$$\frac{\mathrm{d}[\mathrm{dsDNA}]}{\mathrm{d}t} = k_1\big(a_0 - [\mathrm{dsDNA}]\big)^2 \qquad (1)$$

The formal solution of Eq. (1) for initial condition at $t=0$, $[\mathrm{ssDNA}]=a_0$ can be given as

$$[\mathrm{ssDNA}] = \frac{1}{1+k_1a_0t},\ [\mathrm{dsDNA}] = \frac{k_1a_0t^2}{1+k_1a_0t} \qquad (2)$$

$C_0t_{1/2}$ is an important parameter obtained from

---

*Tel.: +91-22-215-2971; fax: +91-22-215-2181/2110.
E-mail address: muruga@tifr.res.in (R. Murugan).

0301-4622/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0301-4622(03)00190-X

renaturation studies, where $C_0 t_{1/2}=1/k_1$, $C_0=a_0$ (i.e. initial concentration of ssDNA in mol l^{-1} and $t_{1/2}=1/k_1 C_0$ is the time (t) to obtain [ssDNA] = $C_0/2$), which is used to find the 'complexity' (a term used to denote the unique number of base-pairs contained in a genome and its repetitive nature) of given genome. But recent studies [9–11] have shown that renaturation was not just a simple second order process but contained a correct-contact-forming phase (which is bimolecular and fast), followed by an intramolecular zipping phase (which is unimolecular, cooperative and slow) and the correct order parameter to describe the renaturation process was neither [dsDNA] nor [ssDNA] but the 'number of mol-bases' in ssDNA or dsDNA. According to cooperative model Eq. (1) modifies to

$$\frac{\mathrm{d}n_{\mathrm{DS}}}{\mathrm{d}t}=n_c k_2(n_m-n_{\mathrm{DS}})\beta n_{\mathrm{DS}} \quad (3)$$

where n_{DS} denotes the number of mol-bases in dsDNA form and n_m is total number of mol-bases of DNA, k_2 is intramolecular (uni-molecular) zipping rate (s^{-1}) and β (mol-base^{-1}) is the cooperativity index and n_c is the (here $n_c=k_1 a_0 t^2/(1+k_1 a_0 t)$ where k_1 is the rate constant for the formation of correct-contacts) mol-number of correct-contacts attained at time t which is just equal to [dsDNA] as given in Eq. (2). It was also shown that the theoretical $C_0 t$ values obtained from cooperative models fairly agreed well with the experimental $C_0 t$ values whereas non-cooperative models under-estimated the $C_0 t$. Though it was argued that the cooperativity arose due to the memory effect of already zipped form of DNA (i.e. n_{DS}), there was no phenomenological proof for this. Moreover, it is still unclear how ssDNA molecules exactly find their complementary position in solution conditions. A DNA molecule having N number of base pairs has a rate constant of finding correct-contact by a random search as $k_1=2w/N^2$ (here it is $\text{mol-correct-contact}^{-1}\ \text{s}^{-1}$) where w is searching rate (or simply the collision rate) and the multiplication factor 2 is due to the fact that simultaneously both strands search for correct-contacts on each other. For example, in case of linearized pBR32 plasmid DNA, $N=4632$ bp [12] and assuming a diffusion controlled limit (maximum) of $w\sim 10^8$, we obtain $k_1\sim 10$ ($\text{mol-correct-contact}^{-1}\ \text{s}^{-1}$) which is $\sim 10^6$ times lower than the observed value of k_1 [9] and thus indicates the existence of an 'efficient' search process. In this article we have derived a theoretical proof for the existence of cooperativity in renaturation of DNA and using that we have also obtained the correct expression for k_1.

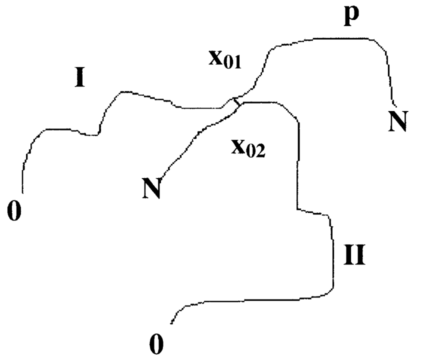

Fig. 1. Here x_{02}th base position of strand (II) is searching for correct-contact 'p' (an absorbing boundary at strand I) starting from x_{01}th position of strand (I) at $t=0$. Here 0 (a reflecting boundary) and N are the helical ends.

2. Theory

As stated in the earlier paragraph, DNA renaturation has two distinct steps, viz. correct-contact-forming and zipping steps. According to our theory, the correct-contact-forming step itself contains two microscopic (which are of course strongly coupled) steps, namely, a contact-forming phase (i.e. a non-specific and diffusion controlled) followed by a searching phase (i.e. search for a correct-contact). The overall correct-contact-forming rate can be derived as follows: let us assume ssDNA with a length of N bases undergoing renaturation (as in Fig. 1) where the first contact happened between x_{01}th base of first strand (I) and x_{02}th base of second strand (II) (where $0<x_{01}<N$ and $0<x_{02}<N$ and they are of course complementary in nature) and second strand search for correct-contact on first strand (for time being let us neglect the helical ends) by a 'random jump'

with a maximum jump size of 'k' and the correct-contact lies at 'p'th position of strand I (i.e. $0 < x_{01} < p < N$). The corresponding transition probabilities for the jump size of 'k' (which include all possible jumps with jump-sizes less than k) for an infinitesimal time Δt can be given as follows:

$$\sum_{i=1}^{k} P(x \to x-i,\ t) = \sum_{i=1}^{k} w_i \Delta t$$

$$\sum_{i=1}^{k} P(x \to x+i,\ t) = \sum_{i=1}^{k} w_{-i} \Delta t \tag{4}$$

$$\sum_{i=1}^{k} P(x \to x,\ t) = \sum_{i=1}^{k} \{1-(w_i+w_{-i})\} \Delta t$$

Now the rate of change in probability of finding the position of x_{02} on strand I can be described by a generalized birth–death master equation as follows:

$$\begin{aligned}\partial_t P(x,\ t) &= \lim_{\Delta t \to 0} \frac{P(x,\ t+\Delta t)-P(x,\ t)}{\Delta t} \\ &= \sum_{i=1}^{k} (w_i P(x-i,\ t)+w_{-i} P(x+i,\ t) \\ &\quad -(w_i+w_{-i}) P(x,\ t))\end{aligned} \tag{5}$$

Here, w_i is transition rate of x_{02} towards 'p' and w_{-i} is away from 'p' with a jump size of i, k is the maximum jump size and $P(x, t)$ is the probability of finding x_{02} at x in time t. We are summing over all the values of i due to the fact that a jump size of k includes all the possibilities within k. Eq. (5) can be solved by the method of generating functions to obtain $P(x, t)$, which can be defined as follows:

$$G(s,\ t) = \sum_{x=0}^{x_0} s^x P(x,\ t) \tag{6}$$

Now by substituting Eq. (6) into Eq. (5), we can obtain the following differential equation for the generating function $G(s, t)$.

$$\begin{aligned}&\partial_t G(s,\ t) \\ &\quad = G(s,\ t) \sum_{i=1}^{k} \left\{ \frac{w_i s^{2i} + w_{-i} - (w_i+w_{-i}) s^i}{s^i} \right\}\end{aligned} \tag{7}$$

Using the initial condition as $P(x_{01}, 0) = 1$ which is due to the fact that at time $t = 0$, the probability of finding x_{02} at x_{01}th position was unity (therefore $G(s,\ 0) = s^{x_{01}}$) and the solution of Eq. (7) for aforementioned initial condition can be given as

$$\begin{aligned}&G(s,\ t) \\ &\quad = s^{x_{01}} \exp\left(t \sum_{i=1}^{k} \left\{ \frac{w_i s^{2i} + w_{-i} - (w_i+w_{-i}) s^i}{s^i} \right\} \right)\end{aligned} \tag{8}$$

In order to get the probability distribution function $P(x, t)$, one has to expand Eq. (8) in to a series of powers of s and find the limit as $s \to 1$ (from the definition of generating function). Since $G(s, t)$ has an essential singularity at $s = 0$, expanding $G(s, t)$ in a Laurent series and setting $s = 1$, we can obtain the expression for the probability distribution $P(x, t)$ as (though it is not expressible in terms of elementary functions)

$$\begin{aligned}P(x,\ t) = \lim_{s \to 1} &\left\{ \frac{1}{2\pi i} \sum_{n=0}^{x} s^n \int_{C_1} \frac{G(s',\ t)}{s'^{n+1}} \mathrm{d}s \right. \\ &\left. + \frac{1}{2\pi i} \sum_{n=-x}^{-1} s^n \int_{C_2} \frac{G(s',\ t)}{s'^{n+1}} \mathrm{d}s \right\}\end{aligned} \tag{9}$$

Since we are interested in the time evolution of mean (this is the expectation value of x at time t) and variance (this is the expected mean square deviation of x at time t) of the position of x_{02} on first strand that can be calculated as follows:

$$\langle x \rangle = \lim_{s \to 1} \partial_s G(s,\ t) = x_{01} + t \sum_{i=1}^{k} i(w_i - w_{-i}) \tag{10}$$

$$\begin{aligned}\mathrm{Var}\{\mathrm{x}\} &= \lim_{s \to 1} \partial_s^2 G(s,\ t) + \langle x \rangle - \langle x \rangle^2 \\ &= t^2 \Big\{ \sum_{i=1}^{k} i^2 (w_i - w_{-i})^2 + \sum_{i<j,j=1}^{k} i \\ &\quad \times (w_i - w_{-i}) j (w_j - w_{-j}) \Big\} \\ &\quad + t \Big\{ 2x_{01} \sum_{i=1}^{k} i(w_i - w_{-i}) + 2 \sum_{i=1}^{k-1} i \\ &\quad \times (i-1)(w_{-i} + w_{i+1}) + k(k+1) w_{-k} \Big\}\end{aligned} \tag{11}$$

Eq. (10) clearly shows that as time evolves, the mean position of x_{02} on the first strand (x) will change linearly with time either left or right to x_{01} depend on the transition rates (w_i and w_{-i}) and Eq. (11) shows that as the time evolves the variance in the position (x) of x_{02} on the first strand will change quadratically with time (this indirectly indicates a spread in the probability distribution $P(x, t)$ as time evolve). The mean first passage time (mfpt) taken by x_{02} to reach 'p' (i.e. to form a correct-contact) starting from any x can be calculated by simplifying Eq. (4) into the following Fokker–Plank equation [13]:

$$\partial_t P(x, t) = -\partial_x[\alpha_1 P(x, t)] + \frac{1}{2}\partial_x^2[\alpha_2 P(x, t)] \tag{12}$$

where $\alpha_1 = \sum_{i=1}^{k} i(w_i - w_{-i})$ is the drift vector and $\alpha_2 = \sum_{i=1}^{k} i^2(w_i + w_{-i})$ is the diffusion matrix. Now defining $\beta(z) = \exp\{\int_0^z (2\alpha_1/\alpha_2)\mathrm{d}z\}$, the mfpt for a jump size of k can be given as

$$T(x, k) = 2\int_x^p \beta(y)^{-1}\mathrm{d}y \int_0^y \beta(z)\alpha_2^{-1}\mathrm{d}z \tag{13}$$

Here '0'th position on strand I (i.e. helical end of the ssDNA strand) is reflecting boundary (i.e. it cannot jump beyond that) and 'p'th position is absorbing boundary (i.e. once x_{02} meet the 'p', it can form the correct-contact and thus can start zipping). Performing the integration of Eq. (13) we obtain

$$T(x, k) = \frac{\alpha_2}{4\alpha_1^2}\left\{\exp\left(\frac{4\alpha_1 p}{\alpha_2}\right) - \exp\left(\frac{2\alpha_1 p}{\alpha_2}\right) - \exp\left(\frac{4\alpha_1 x}{\alpha_2}\right) + \exp\left(\frac{2\alpha_1 x}{\alpha_2}\right)\right\} \tag{14}$$

Eq. (14) gives the mean time taken by x_{02} to reach the correct-contact (p) on its complementary strand starting from a position (x) with a random jump size of k. The problem can be simplified when one of the following conditions holds:

Case I: $w_i = w_{-i}$, $w_i \neq w_j$

When the forward and reverse transition rates for 'i'th jump size are equal, Eq. (7) simplifies to the following form.

$$\partial_t G(s, t) = G(s, t)\sum_{i=1}^{k}\left\{w_i\frac{s^{2i}+1-2s^i}{s^i}\right\} \tag{15}$$

Integration of Eq. (15) yields

$$G(s, t) = s^{x_{01}}\exp\left(\sum_{i=1}^{k}\left\{w_i\frac{s^{2i}+1-2s^i}{s^i}\right\}\right) \tag{16}$$

As in the earlier analysis, the time evolution of mean and variance of position of x_{02} on its complementary strand I (x) can be given as follows:

$$\langle x\rangle = \lim_{s\to 1}\partial_s G(s, t) = x_{01} \tag{17}$$

$$\begin{aligned}\mathrm{Var}\{\mathrm{x}\} &= \lim_{s\to 1}\partial_s^2 G(s, t) + \langle x\rangle - \langle x\rangle^2 \\ &= 2t\sum_{i=1}^{k} w_i i^2\end{aligned} \tag{18}$$

From Eqs. (17) and (18) we can easily conclude that though the mean (given by Eq. (17)) does not evolve with time, the probability distribution function $P(x, t)$ is flattened by the linear evolution of variance. Now from Eq. (13), the mfpt for a jump size k can be given as (here $\alpha_1 = 0$ and $\alpha_2 = 2\sum_{i=1}^{k} i^2 w_i$) follows:

$$T(x, k) = \frac{1}{\alpha_2}(p^2 - x^2) \tag{19}$$

Case II: $w_i = w_{-i}$, $w_i = w_j = w$

When all the transition rates are equal, Eq. (7) reduces to a simple form as follows:

$$\begin{aligned}\partial_t G(s, t) &= wG(s, t)\sum_{i=1}^{k}\left\{\frac{s^{2i}+1-2s^i}{s^i}\right\} \\ &= wG(s, t)\left\{\frac{(1+2k)(1-s) - s^{-k} - s^{k+1}}{s-1}\right\}\end{aligned} \tag{20}$$

The solution of Eq. (20) for the same initial condition as in case I can be given as

$$G(s, t) = s^{x_{01}}\exp\left(wt\sum_{i=1}^{k}\left\{\frac{s^{2i}+1-2s^i}{s^i}\right\}\right) \tag{21}$$

Now the mean and variance of x becomes

$$\langle x\rangle = \lim_{s\to 1}\partial_s G(s,\ t) = x_{01} \tag{22}$$

$$\mathrm{Var}\{\mathrm{x}\} = \lim_{s\to 1}\partial_s^2 G(s,\ t) + \langle x\rangle - \langle x\rangle^2 = 2wt\sum_{i=1}^{k} i^2 \tag{23}$$

Finally, the mfpt for the jump size of k can be given (from Eq. (13) where $\alpha_1 = 0$ and $\alpha_2 = 2w\sum_{i=1}^{k} i^2$) as follows:

$$T(x,\ k) = \frac{1}{\alpha_2}(p^2 - x^2) = \frac{3(p^2 - x^2)}{k(k+1)(2k+1)w} \tag{24}$$

Now from Eq. (24) the minimum (i.e. starting the search from position $x = 0$ and taking $p = N$) value of rate constant for the formation of correct-contact can be easily given as

$$r_{K\min} = \frac{1}{T(0,\ k)} = \frac{2w}{p^2}\frac{k(k+1)(2k+1)}{6} \tag{25}$$

One also should note that

$$\lim_{p\to N} r_{K\min} = \frac{2w}{N^2}\frac{k(k+1)(2k+1)}{6} = k_1\frac{k(k+1)(2k+1)}{6} \tag{26}$$

And it is easy to verify that $\lim_{k\to 1} r_{K\ \min} = k_1$ and when k is sufficiently large,

$$\lim_{k\to p} r_{K\min} \cong 2wp/3 \tag{27}$$

Thus, $r_{K\ \min} = \beta k$, where $2w/3 = \beta$. Moreover, the variation of $r_{K\ \min}$ with N and k has been shown in Fig. 2 (here $w = 3$ used for simplicity), which clearly indicated that at sufficiently large values of N and k, correct-contact-forming rate was almost a constant quantity. Since the so-called jump (with a size of k bp) can happen only when the complementary ssDNA strands are in contact (i.e. in the dsDNA form), the maximum possible jump-size at any time t is $k = n_{DS}$. According to the law of mass action the rate of formation of n_{DS} is directly proportional to number of correct-contacts (i.e. n_c times $r_{K\ \min}$) and number of mol-bases in ssDNA form (i.e. $n_m - n_{DS}$) at time t. Therefore, using the relation $r_{K\ \min} = \beta k$, the final version of differential rate equation can be written as

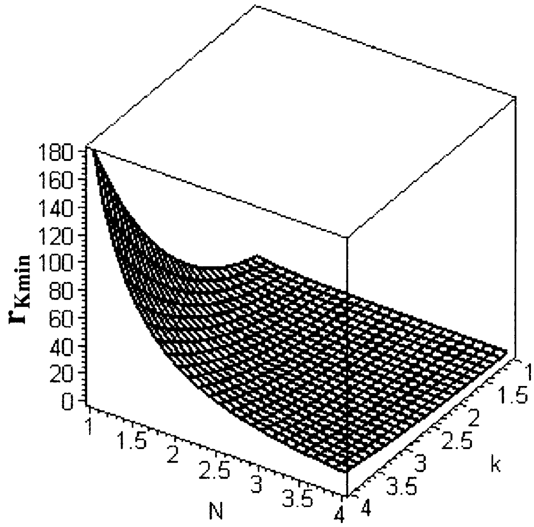

Fig. 2. Shows the variation of rate of formation of correct-contacts ($r_{K\ \min}$) with a maximum jump size (k) and length of ssDNA (in base pairs). Here $w = 3$ is used for simplicity.

$$\frac{dn_{DS}}{dt} \alpha\, n_c r_{K\min}(n_m - n_{DS})\, \alpha\, n_c k\beta(n_m - n_{DS}) \tag{28}$$

$$\frac{dn_{DS}}{dt} = n_c k_2(n_m - n_{DS})\beta n_{DS} \tag{29}$$

Surprisingly, Eq. (29) is simply the differential rate equation derived earlier for cooperative renaturation [9], where k_2 is intramolecular zipping rate (s^{-1}). Thus, we can conclude that cooperativity could possibly be the result of searching for correct-contacts by random jumps.

3. Results and discussions

Using Eq. (26) we can easily solve the paradox arose in the renaturation of linearized pBR322 plasmid. As we have already seen, the theoretical rate constant for finding the correct-contact of

$k_1 \sim 10$ (mol-correct-contact^{-1} s^{-1}) was $\sim 10^6$ times lower than the experimental value of k_1 which suggested that a minimum jump size of $k \sim 100$ bp (i.e. starting from x_{01}, after one searching move, x_{02} can be found in the range of $x_{01} \pm 100$ bp with equal probability and assuming $w \sim 10^8$ collisions s^{-1}, and $N = 4632$ bp) was necessary. The correct-contact-forming step is especially important in case of PCR. Here a primer of few base pairs (let us say z bp) in length has to find its complementary sequence on template DNA (with a length of N bp) by random search before starting to anneal. Now the expression for rate of formation of correct-contacts between template DNA and primer can be derived as follows: since here the maximum number of possible base-pairs in dsDNA form (thus the jump size) is just z bp, the minimum correct-contact-forming rate becomes

$$r_{K\min} = \frac{2w}{N^2} \frac{z(z+1)(2z+1)}{6} \tag{30}$$

which clearly indicates that $r_{K\ \min} \propto z^3/N^2$ and, therefore, in a PCR reaction for a particular fragment of a template DNA, the forward and reverse primers should be equal in length which otherwise lead to some sort of asymmetry in amplification and it is obvious that higher primer length leads to higher specificity (due to the higher rate of formation of correct-contact with the template) which otherwise require a longer annealing times to attain the same specificity.

Acknowledgments

This work was supported by TIFR endowment fund for carrier development. The author thanks Professor S. Mazumdar for his constant encouragement and help.

References

[1] R.J. Britten, D.E. Graham, B.R. Neufeld, Analysis of repeating DNA sequences by reassociation, Methods Enzymol. 29 (1974) 363–405.

[2] D.G. Peterson, W.R. Pearson, S.M. Stack, Characterization of tomato genome using in vitro and in situ DNA reassociation, Genome 41 (1998) 346–356.

[3] R.B. Goldberg, From C_0t curves to Genomics. How gene cloning established new concepts in plant biology, Plant Physiol. 125 (2001) 4–8.

[4] E. Schaeffeler, T. Lang, U.M. Zanger, M. Eichelbaum, M. Schwab, High-throughput genotyping of thiopurine *S*-methyltransferase by denaturing HPLC, Clin. Chem. 47 (2001) 548–555.

[5] Anshelevich, V.V. Vologodskii, J. Biomol. Struct. Dyn. 4 (1986) 251–262.

[6] A. Sayonara, A. Wada, Biopolymers 23 (1984) 409.

[7] G.S. Manning, Q. Rev. Biophys. 11 (1978) 179–246.

[8] A.V. Vologradskii, A.V. Lukashin, M.D. Frank-Kamenetskii, Biopolymers 23 (1984) 39.

[9] R. Murugan, Revised theory on DNA renaturation kinetics and its experimental verification, Biochem. Biophys. Res. Commun. 293 (2002) 870–873.

[10] R. Murugan, Sample size autocorrelation analysis of kinetic data: resolving reaction path heterogeneity, J. Chem. Phys. 117 (2002) 6372–6377.

[11] R. Murugan, A stochastic model on DNA renaturation kinetics, Biophys. Chem., 104 (2003) 536–541.

[12] J.G. Sutcliffe, Cold Spring Harbor Symp. Quant. Biol. 43 (1979) 77–90.

[13] Gardiner, in: (Ed.), Handbook of Stochastic Methods, Springer, Berlin, 1983.