Bull Math Biol (2011) 73: 1271–1291
DOI 10.1007/s11538-010-9558-8

ORIGINAL ARTICLE

A Mathematical Model for Timing the Release from Sequestration and the Resultant Brownian Migration of SeqA Clusters in *E. coli*

Donald A. Drew · Gretchen A. Koch · Shannon Hitchcock · Jill Kowalski · Ronak Talati · Vera Valakh

Received: 12 November 2009 / Accepted: 4 June 2010 / Published online: 17 July 2010

Abstract DNA replication in *Escherichia coli* is initiated by DnaA binding to *oriC*, the replication origin. During the process of assembly of the replication factory, the DnaA is released back into the cytoplasm, where it is competent to reinitiate replication. Premature reinitiation is prevented by binding SeqA to newly formed GATC sites near the replication origin. Resolution of the resulting SeqA cluster is one aspect of timing for reinitiation. A Markov model accounting for the competition between SeqA binding and methylation for one or several GATC sites relates the timing to reaction rates, and consequently to the concentrations of SeqA and methylase. A model is proposed for segregation, the motion of the two daughter DNAs into opposite poles of the cell before septation. This model assumes that the binding of SeqA and its subsequent clustering results in loops from both daughter nucleoids attached to the SeqA cluster at the GATC sites. As desequestration occurs, the cluster is divided in two, one associated with each daughter. As the loops of DNA uncoil, the two subclusters migrate apart due to the Brownian ratchet effect of the DNA loop.

Keywords Bacterial cell division · *E. coli* · Sequestration · Brownian migration · SeqA · DNA methylation

1 Introduction

Cell replication begins with replication of its DNA, followed by the partitioning of the daughter strands into separate parts of the cell. These processes must be carefully

D.A. Drew (✉) · S. Hitchcock · J. Kowalski · R. Talati · V. Valakh
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA
e-mail: drewd@rpi.edu

G.A. Koch
Department of Mathematics, Goucher College, Baltimore, MD 21204, USA

regulated by the cell. A description of the many subprocesses involved in bacterial cell division is given in (Bramhill 1997). The key events in bacterial reproduction are replication, segregation, and septation. Further, during replication, the cell must prevent premature reinitiation, using a process called sequestration.

The process is illustrated most succinctly by the reproduction of *E. coli*. Here, the basic processes of DNA replication from initiation through to septation, with timing and spatial location control, occur in a system that is well studied and documented, with a well-documented genome and proteome. In this paper, we offer mathematical models for some division processes, some as documented in the literature, and others postulated in this paper. Specifically, we propose and analyze a model for bundling the DNA chain by preferential binding and the subsequent release of SeqA, followed by a kinematic chain straightening by a rectified Brownian diffusion.

In order to motivate the mathematical models that we propose and analyze, we next give a discussion of biological aspects of bacterial cell division that we feel to have relevance to the models. The modeling sections can be read without reading the remainder of this section.

A mature *E. coli* bacterium has a loop of DNA, located centrally between the two poles in the tubular shaped cell. Replication begins when a cluster of DnaA molecules binds to a neighborhood of the chromosomal locus *oriC*, and recruits other proteins that assemble into two DNA polymerase Replication Factories (Bramhill 1997; Bramhill and Kornberg 1988; Baker and Kornberg 1988) that replicate the DNA sequentially outward from the replication origin in opposite directions along the DNA, melting and assembling the double strands into two daughter strands. The availability of the initiator protein, DnaA, affects the time of initiation in the cell cycle (Boye et al. 1988, 1996). Shortly after the DnaA-modified chromosome recruits the proteins for the Replication Factory, the DnaA proteins are released from the site and exist as viable proteins in the cytoplasm, able to participate in reinitiation. Even though reinitiation is possible, it is not always desirable. Secondary initiations of the newly replicated origins are controlled by a process called sequestration (Boye et al. 2000; Nievera et al. 2006). When the chromosome is sequestered, reinitiations do not occur for substantial periods during a cell cycle (Skarstad et al. 1986; Campbell and Kleckner 1990; Boye et al. 2000; Taghbalout et al. 2000; Norris et al. 2000; Waldminghaus and Skarstad 2009). Sequestration is accomplished by the interaction of the protein SeqA with specific sites on the DNA coded with GATC. During this eclipse period, the adenines in these GATC sites are methylated on only one strand, presumably the mature strand. Newly replicated DNA is sequestered by binding the SeqA protein to hemimethylated GATC sites in the neighborhood of the replication origin (*oriC*). This binding prevents remethylation of the N6-adenine residue (Campbell and Kleckner 1990), and sterically inhibits binding of DnaA for reinitiation (Boye et al. 2000; Torheim and Skarstad 1999; Waldminghaus and Skarstad 2009). During later times in the cell cycle, GATC sites are methylated on both strands by Dam methylase. In the presence of SeqA, the eclipse period is highly influenced by the level of Dam methylase present in the cell (Freiesleben et al. 2000; Nielsen and Løbner-Olesen 2008). Furthermore, in vitro, Dam methylase exhibits lower affinity for *oriC* than SeqA and is unable to displace any SeqA proteins already bound to hemimethylated *oriC* (Kang et al. 1999).

The process of breaking the bond between SeqA and DNA is important in timing the reinitiation. The binding of SeqA to the GATC sequence in the DNA is reversible, since it consists of hydrogen "bonds" and the entire process is thermally active. The binding and unbinding of SeqA allows the competitive binding of the methyl group, which effectively ends the possibility of rebinding SeqA. In addition, it has been noted that the remethylation process occurs at different rates for the different strands of the DNA (Stancheva et al. 1999). This asymmetry may well be important in segregation.

DNA assembly proceeds whereby the two existing "old" strands are each assembled into new strands by adding the second strand of new nucleotides to each of the old. This process proceeds from the 3' to the 5' ("leading") direction on one-half, and from the 5' to the 3' ("lagging") direction on the other. The assembly process differs between the two strands, with a relatively direct assembly of nucleotides into the new strand on the leading half, but assembly of nucleotides into short pieces of DNA called Okazaki fragments, which then are incorporated with the lagging strand. It is interesting that a GATC site on the leading strand corresponds to the sequence CTAG on the lagging strand, which when read from the 3' to 5' direction, is seen as a GATC sequence. It is not known whether the different assembly processes lead to different competition for methylation versus binding of SeqA. However, as observed in Stancheva et al. (1999), the rate of remethylation of the lagging strand is approximately half as fast as the leading strand. Since SeqA is implicated in the delay of remethylation by DAM methylase, it is reasonable to assume that a greater amount of sequestration by SeqA (and consequent DNA looping) must be present in the lagging strand when compared to the leading strand post replication (Fig. 2). This causes an asymmetry between the tethering of loops in the lagging and leading strands of the replicated DNA.

As the DNA is being replicated, it moves in the cell, eventually winding up with the two daughter nucleoids in opposite halves of the cell, with the replication origins *oriC* toward the poles, and *ter* toward the midplane. This particular motion, called segregation, has not been adequately explained. Jacob et al. (1963) proposed that replicating chromosomes attach to the cell wall and are carried by the insertion of structural material. Jun and Mulder (2006) argue that cell wall growth is not fast enough to account for the speed of segregation, on the order of 0.1 μm/sec. When looped sections of the leading strand are unbound from SeqA, its loops are free to stretch out, leading to a tract of SeqA, one of the *oriC* sites, and the Replication Apparatus moving apart toward cellular poles. This motion is consistent with the motions observed by Webb et al. (1997). In the next paragraphs, we summarize possible propulsion mechanisms.

The Replication Apparatus attaches nucleotides to the leading strand of the DNA, resulting in pulling the DNA chain through the Replication Apparatus and pushing out the replicated strands. This potential motive force must be counteracted by tethering while the Replication Apparatus remains stationary at midcell, and could be (partially) responsible for the forces that cause the rapid migration of the replicated DNA to the quarter positions before septation. We propose a model for the timing of segregation and the subsequent rapid migration of the Replication Apparatus, the daughter oriCs and two SeqA tracts based on the DNA looping and clustering action of SeqA and the de-sequestration by an increase in DAM Methylase.

One proposed mechanism for propulsion is the existence of a propulsion protein. There is speculation (Kruse et al. 2006) that the actin homolog MreB RNA polymerase (RNAP) provide a force for segregation of the daughter DNA molecules. MreB is important for formation of the cytoskeleton in *E. coli*, and the observation that MreB and RNAP interact suggests that MreB may be important for chromosome segregation. However, Formstone and Errington (2005) report that deletion of mreB shows that MreB is important for control of cell width and that virtually normal growth was possible in the mreB null mutant.

Jun and Mulder (2006) hypothesize a propulsion mechanism that assumes that the intertwined strands act as a "entropic spring" that favors the state of nonintertwined strands, and random motions progress toward this state. The model that they study assumes that the spring relaxes in a manner akin to "reptation," by strands moving in and out of channels in the tangle of the chain. They imply that their entropic spring model does not lead to sufficiently fast relaxation. Our proposed mechanism is also an "entropic spring," but behaves quite differently from the disfavored intertwining model, assuming that the relaxation to the "straight" state occurs due to the relative differences in the relative number of possible "straighter" states and "more looped" states of the chain.

It is observed in vivo that tagged SeqA protein undergoes a rapid migration at a certain time during the cell replication cycle (Onogi et al. 1999; Hiraga 2000; Gordon et al. 2004). Immediately after initiation, a relatively stable cluster of SeqA proteins is associated with the newly replicated DNA, and remains at midcell for about 40 minutes (Hiraga et al. 2000). After that, the SeqA cluster divides into two clusters that then migrate to opposite cell quarter positions in preparation for septation. The migration is rapid at approximately 0.1 μm/min. The mechanism for the observed rapid SeqA-DNA migration is as yet unknown (Onogi et al. 1999; Webb et al. 1997; Waldminghaus and Skarstad 2009).

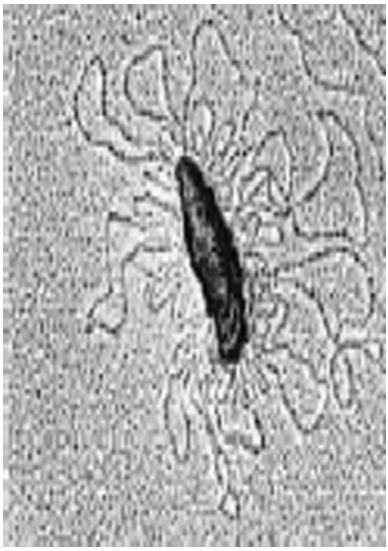


Fig. 1 Looping of intervening DNA by SeqA. Multiple DNA loops between tight SeqA clusters observed in *E. coli* (Skarstad et al. 2000)

Sequestration of DNA by SeqA has been shown to result in formation of intervening loops of DNA (Brendler et al. 2000). In addition, it is known that SeqA exists as a cluster, or multimer, to generate the observed delay of remethylation (Hiraga 2000). Binding of *oriC* plasmid to SeqA in vitro shows visible loops as observed by electron microscopy; see Fig. 1 (Skarstad et al. 2000). Although the purpose of the observed looping is in question, loops of DNA hinder the migration of replications structures from the cell midplane. The observed migration of the replication structures along with SeqA tracts could result from a combination of the uncoiling of the loops and

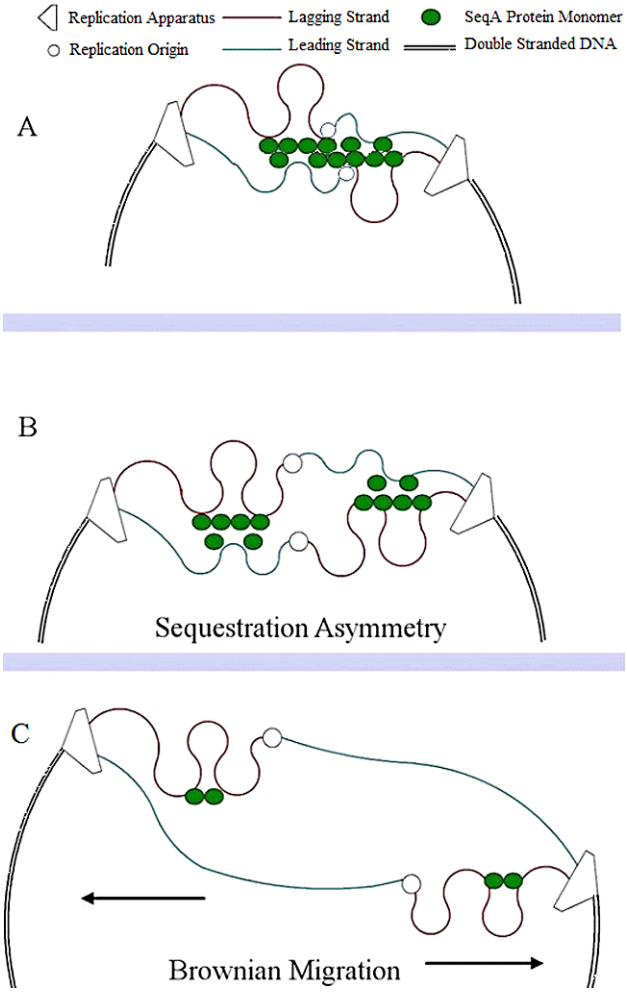


Fig. 2 Brownian migration. **a** A cluster of SeqA molecules with DNA loops attached. **b** Asymmetrical sequestration of lagging vs. leading strands in newly replicated DNA leads to different levels of looping in which SeqA dislodges at the same rate. **c** Less looped structures such those of the leading strand begin to unloop and straighten. As a result, Brownian motions in the DNA loops participate in the net polar migration

the action of the Replication Apparatus as it pulls in DNA and pushes out the replicated DNA strands (Fig. 2). Loop uncoiling is a result of a change in the structure of the SeqA clusters that is triggered by an increase in methylation of the GATC sites.

Post replication, the SeqA supercoiled structure relaxes. It seems likely that this is induced by the known competition between remethylation (DAM methylase) and sequestration (SeqA). The onset of competition can be attributed to an increase in DAM methylase production. The observed migration of the DNA-SeqA structures to the quarter positions in the cell occurs rapidly after its onset.

Fan et al. (2007) postulate that the transport of DNA could be caused by the uncoiling of supercoiled structures in the DNA. They cite the model of entropy driven process offered by Jun and Mulder (2006). The analysis offered here is also in this spirit. We shall discuss a model whereby a chain in the form of a loop uncoils when released from the spatially intermittent SeqA binding, moving the remaining bound SeqA and the DNA itself away from the site of the original cluster. The essence of this motion is that the loops represent low entropy states, which are formed by the capture of the loops by SeqA as they are formed. When the loops are released by the SeqA clusters, they migrate toward a higher entropy state with a less coiled structure. The thermal energy in the cytoplasm, the molecules in the SeqA cluster, and the DNA itself seeks equilibration in the entire structure. The subunits of the loops undergo random motions that result in an uncoiling of the loops, moving the DNA-SeqA structures through a process which we call Brownian migration. Figure 2 shows the conceptual model for this process.

Since lagging and leading strands contain different amounts of SeqA, and DAM Methylase dislodges it at a constant rate, then unlooping of the leading strand would occur first since it contains less SeqA followed by the lagging strand. During DNA unlooping, the Replication Apparatus, along with what remains of the SeqA cluster, is pushed in the direction of unlooping (Fig. 2).

In the following sections, we present models for the timing and dynamics of SeqA-related processes that suggest that segregation could be due to the asymmetry of the SeqA clusters bound to daughter DNA strands, and the subsequent release and rectified motion of the SeqA clusters when the loops of DNA straighten.

2 Mathematical Model for Sequestration

The purpose of this paper is to introduce and study mathematical models of the processes of desequestration and loop uncoiling. The modeling is directed at ascertaining whether the proposed mechanisms can act in the manner suggested in a reasonable time frame. Moreover, the models can suggest features of the processes which might be further verified by experimental observation.

2.1 Sequestration

We first present models for the rate at which several sequestered sites become sequentially unsequestered in terms of the rates of binding and unbinding of SeqA and methyl and the rate of production of binding sites. These models describe different

features of the sequestration process, and allow the comparison of desequestration rates under an assumption about whether desequestration of sequestered sites occurs sequentially or independently.

We start with a model of one site, having three states: sequestered (denoted by S), unbound (U), and methylated (M). Treating the mechanism as a Markov process gives the probability distribution of the times spent in each state. We use this model to get a sense of the time taken to methylate a site in terms of the rates of binding and unbinding SeqA and binding methyl. We then present a model to calculate the probability distributions for a series of methylation reactions for several sites. The model shows that the probability distribution function for the time taken to methylate the ith site in the series is the convolution of the probability distribution functions of the times to methylate any one site, which is the result of the single site model.

In order to model the way that a DNA receptor site becomes sequestered, unsequestered, and methylated, two Markov chain models were utilized. The first model is a three-state model, which describes the probability that a given sequestration site is sequestered (i.e., bound to SeqA), unbound, or (fully) methylated. This model describes how several independent sites would de-sequester. However, it is also possible that the sequestered sites may de-sequester in order. Then, in order to study this process of sequential desequestration, we examine an infinite chain model. The motivation for sequential desequestration is that the cluster of SeqA molecules sterically hinder access to the GATC-SeqA sites so that they must desequester in order.

Desequestration: Independent Site Model This model describes the probability of seeing a site in the states corresponding to having no SeqA protein or methyl group attached, having a SeqA protein attached, or having a methyl attached. We can view the probability as giving the fraction of sites in these various states in a uniform population of cells. Let $s_0(t)$ be the probability that the site is sequestered at time t, $u_0(t)$ the probability that it is unbound, and $m_0(t)$ the probability that the site is methylated.

The Markov equations are

$$\frac{du_0}{dt} = -(k_{\text{on}} + k_\mu)u_0 + k_{\text{off}}s_0$$
$$\frac{ds_0}{dt} = k_{\text{on}}u_0 - k_{\text{off}}s_0$$
$$\frac{dm_0}{dt} = k_\mu u_0$$

The initial conditions for this system are $u_0(0) = 0$, $s_0(0) = 1$, $m_0(0) = 0$, reflecting the starting condition that the site is sequestered. Note that by adding the three equations, integrating, and using the initial conditions, we find that $u_0 + s_0 + m_0 = 1$.

The variable m_0 does not appear in the first two equations. Thus, we can solve the 2×2 linear system for $s_2(t)$ and $u_0(t)$, and subsequently find $m_0(t)$ by quadrature.

We shall nondimensionalize the system with a time scale given by the rate of unbinding of SeqA. That is, we define a dimensionless time variable τ by $\tau = k_{\text{off}}t$.

The system of equations becomes

$$\frac{du_0}{d\tau} = -(K_{\text{eq}} + K_{\text{eq}}K_c)u_0 + s_0$$
$$\frac{ds_0}{d\tau} = K_{\text{eq}}u_0 - s_0$$

where we introduce the sequestration equilibrium constant, $K_{\text{eq}} = k_{\text{on}}/k_{\text{off}}$ and the competition constant, $K_{\text{co}} = k_\mu/k_{\text{on}}$. The sequestration equilibrium constant is the probability that a site is sequestered in equilibrium in the absence of the methylation apparatus. The competition constant is the ratio of the rates of methylation to sequestration. If this fraction is bigger than one, methylation has an advantage, while sequestration has an advantage if it is less than one.

Results for Independent Site Model The solution of this system is given by

$$u_0(\tau) = \frac{1}{\sqrt{D}}\left[\exp\left(-\frac{1}{2}\left(K - \sqrt{D}\right)\tau\right) - \exp\left(-\frac{1}{2}\left(K + \sqrt{D}\right)\tau\right)\right]$$
$$s_0(\tau) = \frac{1}{2\sqrt{D}}\left[a\exp\left(-\frac{1}{2}\left(K - \sqrt{D}\right)\tau\right) + b\exp\left(-\frac{1}{2}\left(K + \sqrt{D}\right)\tau\right)\right]$$

where $K = 1 + K_{\text{eq}} + K_{\text{eq}}K_{\text{co}}$ and $D = 1 + 2K_{\text{eq}} - 2K_{\text{eq}}K_{\text{co}} + K_{\text{eq}}^2 + 2K_{\text{eq}}^2K_{\text{co}} + K_{\text{eq}}^2K_{\text{co}}^2$, $a = -(1 - K_{\text{eq}} - K_{\text{eq}}K_{\text{co}}) + \sqrt{D}$ and $b = 1 - K_{\text{eq}} - K_{\text{eq}}K_{\text{co}} + \sqrt{D}$.

The solutions for $K_{\text{eq}} = 1$ and $K_{\text{co}} = 1$ are shown in Fig. 3, along with $m_0(\tau) = 1 - u_0(\tau) - s_0(\tau)$, which is the probability that the site becomes methylated by time τ.

The time taken to become methylated with probability 50% is a measure of how the parameters affect the time that the DNA spends sequestered. This time is plotted versus K_{eq} and K_{co} in Fig. 4.

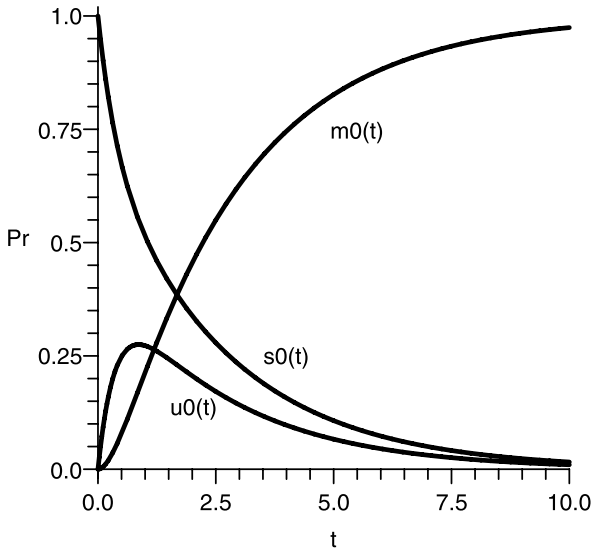


Fig. 3 Independent site probabilities of sequestered ($s_0(t)$), unbound ($u_0(t)$), and methylated ($m_0(t)$) for $K_{\text{eq}} = 1$, $K_{\text{co}} = 1$

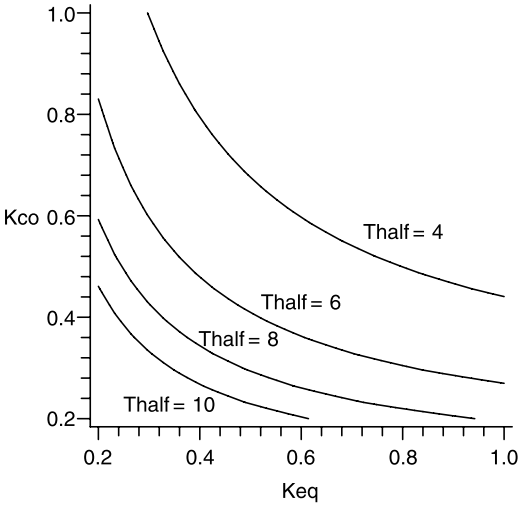


Fig. 4 Dimensionless de-sequestration time versus rate constants

Sequential Desequestration This model describes the probability of observing a sequence of sites, each in the states corresponding to having no SeqA protein or methyl group attached, having a SeqA protein attached, or having a methyl attached. Let $s_i(t)$ be the probability that site i is sequestered at time t, $u_i(t)$ the probability that it is unbound, and $m_i(t)$ the probability that the site is methylated. The model assumes that the SeqA structure sterically obscures the SeqA binding sites except for the one that has just attached a SeqA to the most recently replicated GATC site. The desequestration starts at the "first" site (taken to be the last site sequestered by the sequestration process), and when that has happened, the next site is available to bind and unbind the SeqA molecule or the methylase. This process continues to the next site, and so on. The system of differential equations for the probabilities are

$$\begin{aligned}
\frac{du_1}{dt} &= -(k_{\text{on}} + k_\mu)u_1 + k_{\text{off}}s_1 \\
\frac{ds_1}{dt} &= k_{\text{on}}u_1 - k_{\text{off}}s_1 \\
\frac{du_2}{dt} &= -(k_{\text{on}} + k_\mu)u_2 + k_{\text{off}}s_2 \\
\frac{ds_2}{dt} &= k_{\text{on}}u_2 - k_{\text{off}}s_2 + k_\mu u_1 \\
&\vdots \\
\frac{du_i}{dt} &= -(k_{\text{on}} + k_\mu)u_i + k_{\text{off}}s_i \\
\frac{ds_i}{dt} &= k_{\text{on}}u_i - k_{\text{off}}s_i + k_\mu u_{i-1}
\end{aligned}$$

with initial conditions

$$\begin{aligned} s_1(0) &= 1 \\ u_1(0) &= 0 \\ s_2(0) &= 0 \\ u_2(0) &= 0 \\ &\vdots \\ s_i(0) &= 0 \\ u_i(0) &= 0 \end{aligned}$$

This system reflects a similar set of dynamics for the single site model, except that the rate of methylation for the first (and subsequent) sites provide a "source" of the next site in the sequestered state. Moreover, the initial conditions reflect the fact that the system starts with the first site sequestered, and all other possible states having probability zero. Note that the equations for the first site are identical to the equations solved above for the Simple Model.

Results for Multiple Site Model We nondimensionalize the equations in the same way as the Single-Site Model. The solution for the first site ($i = 1$) is as given above. The solutions for subsequent sites ($i = 2, 3, \ldots$) can be related to $u_{i-1}(t)$ by solving the inhomogeneous system with homogeneous initial conditions. The solutions can be written as

$$\begin{bmatrix} u_i(\tau) \\ s_i(\tau) \end{bmatrix} = K_{\text{eq}} K_{\text{co}} \int_0^\tau u_{i-1}(s) \begin{bmatrix} Q_1(\tau - s) \\ Q_2(\tau - s) \end{bmatrix} ds$$

where, remarkably, $Q_1(\xi) = u_1(\xi)$ and $Q_2(\xi) = s_1(\xi)$. To see the reason for this, let us change variables in the convolution integral to get

$$u_i(\tau) = K_{\text{eq}} K_{\text{co}} \int_0^\tau u_{i-1}(\tau - \sigma) u_1(\sigma)\, d\sigma$$

Then we can interpret $K_{\text{eq}} K_{\text{co}} u_1(\sigma)$ as a "transfer function", representing the rate that a site that is sequestered at time $\sigma = 0$ makes the transition from unsequestered to methylated at time σ. Since $u_{i-1}(\tau - \sigma)$ is the probability that site $i - 1$ is unsequestered at time $\tau - \sigma$, then summing (i.e., integrating) over all possible times gives the probability that site i is unsequestered at time τ. A similar argument applies for the probability that site i is sequestered.

The results for u_i and m_i, $i = 1, 2, 3, 4$ are plotted in Figs. 5 and 6, found by numerically solving the equations for $i = 1, 2, 3, 4$.

2.2 Interpretation

The results for the various rates studied appears to point to a rule of thumb for the timing of subsequent desequestrations of a strand. To see this, consider the probability that site i is methylated by time t. This can be derived by recalling that the rate

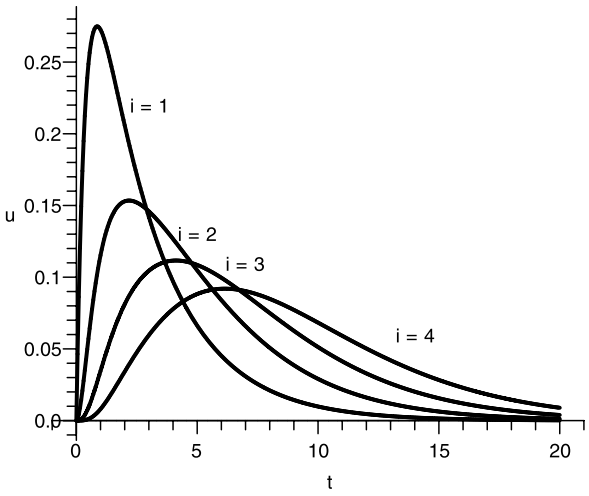


Fig. 5 Multiple site probability. Probability that site i is unbound at time t, for $K_{\text{eq}} = 1$, $K_{\text{co}} = 1$

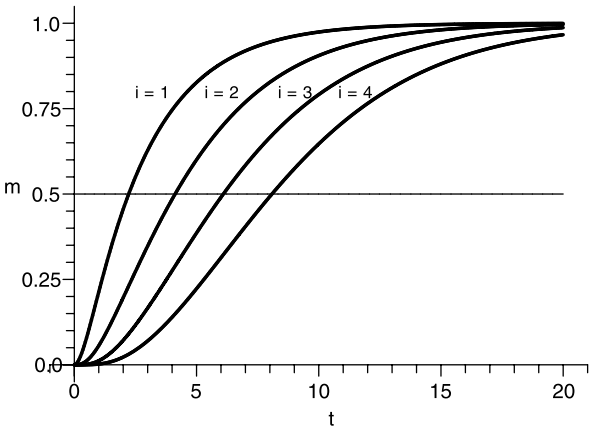


Fig. 6 Multiple site probability. Probability that site i is methylated at time t, again for $K_{\text{eq}} = 1$, $K_{\text{co}} = 1$

at which site i becomes methylated is $dm_i/dt = K_{\text{eq}} K_{\text{co}} u_i$. In Fig. 6, the probability level $m = \frac{1}{2}$ is also shown. The time taken for the first site to become methylated with probability $\frac{1}{2}$ is approximately 2.3. The time taken for the second site to become methylated with probability $\frac{1}{2}$ is approximately 4.2, about 1.9 time units after the first site reaches that level. The third site reaches probability $\frac{1}{2}$ at time 6.1, and the fourth at time 8.0. Thus, after the first site reaches probability $\frac{1}{2}$, the time intervals for subsequent site methylation probabilities to reach $\frac{1}{2}$ are the same, and correctly scaled by the time for the first site to reach probability $\frac{1}{2}$. This result suggests that observations of the dynamics of SeqA as bound to DNA or unbound gives very different results, roughly that sequential desequstration takes about n times as long as independent desequestration, where n is the number of sites to be desequestered.

This suggests that there is a significant difference in the desequestration rates between sequential and parallel desequestration. If we assume that the individual

protein binding and unbinding rates are approximately the same, the sequential desequestration takes many times longer, since the sequential desequestration time is approximately n times the single site rate, where n is the number if sites that must be desequestered.

The rates for the various subprocesses in the desequestration model are largely unknown, and it is also not known if the SeqA binding has an efficient catalyst and is called the diffusion limit and is about 10^8 to 10^9 $M^{-1}\,s^{-1}$. At this point, every collision of the enzyme with its substrate will result in catalysis, and the rate of product formation is not limited by the reaction rate but by the diffusion rate.

3 Mathematical Model for Segregation

At the end of the replication process, the daughter DNA must wind up in the proper location for septation to result in viable daughter cells. The process by which the DNA is transported is unknown. Here, we present a model that allows the directed migration of two SeqA sub-clusters, each with a Replication Factory tethered to it distally from the center of the original SeqA cluster. We assume that the original SeqA cluster (Fig. 2) breaks in two relatively equal pieces by the process of methylation. The subclusters are tethered to each other through two daughter substrands of DNA, which consists of several loops that have been recently released from the SeqA cluster.

The mechanism that we postulate for segregation is Brownian migration, suggested in Fig. 2C. In this mechanism, the thermal motions of the SeqA clusters are rectified by the presence of the DNA loops, which begin in a state of several loops, a condition of low entropy (highly ordered state). If the clusters move thermally apart, the entropy in the loops increases. If they move toward each other, the entropy in the loops may decrease. The model we propose and analyze here is a semiempirical random walk on integer separations between the clusters. The random walk is assumed to occur between the two SeqA clusters connected by a chain of length L.

The random walk of the two SeqA clusters is modulated by the presence of a chain of finite length connecting the clusters. We assume that the clusters and the chain are both thermally active, i.e., subject to small random fluctuations driven by Brownian motions in the cytoplasm. The motions sample from the ensemble of possible motions, and the SeqA cluster motions interact with each other through the chain.

The basic premise that we work from is that the thermal fluctuations result in minimal motions. Specifically, we shall consider one dimension of the motion of the clusters—either together or apart, and assume that the chains consist of discrete nodes and links, with the node-to-node length fixed but with some flexibility. We further restrict the chain model so that the flexibility is manifested in discrete angle changes in two dimensions. While we do not expect this model to have detailed physical analogy to reality, we expect to demonstrate the concept through this set of assumptions.

Chain Model In order to model the behavior of DNA loops, a two-dimensional discrete mathematical model of chains was formulated. Consider a model for the evolution of a discrete two-dimensional chain having links of unit length and allowing

each link to differ from the one before it by a turn to the right or left of some fixed amount $\Delta\theta$, or make no turn. With this paradigm, a loop is a segment of chain having more in a gradual overall turn in the same direction.

The evolution of the chain corresponds to Brownian (thermal) fluctuations in the system. For these fluctuations, we shall assume that they change the configuration of the chain in a neighborhood of the application of some fluctuation "force." The fluctuations that we consider are changes in the turns in the chain. We constrain these fluctuations so that they (i) do not result in any turns that are not 0 or $\pm\Delta\theta$; and (ii) do not move more than one node at a time.

A chain is then specified by a turn vector, t_j, $j = 1, \ldots, N$, where $t_j \in \{-1, 0, 1\}$, each element representing the change in direction from the current orientation that the links will turn at node j. A turn of -1 represents a turn to the right, 0 is straight, and 1 is a turn to the left of angle $\Delta\theta$. We implement the generation of random chains in MATLAB© by generating a random sequence $\{t_j\}$ and represent the chain on the complex plane according to the formula $z_j = z_{j-1} + \exp(i\,\Delta\theta \sum_{k=0}^{j} t_k)$. Without loss of generality, the initial node is set to $(0, 0)$ with the following link going "straight" out to the right to $(1, 0)$. The end-to-end distance given as a function of its length is $|z_N|$.

Figure 7 shows several such chains with length $N = 20$ and $\Delta\theta = \pi/5$. Through this method, one can generate chains with specific structures (i.e., large loops) based on a pre-determined turn vector or random chains based on randomly produced turn vectors. Furthermore, evolution of a chain requires an algorithm to change the turn vector t_j^{τ}, $j = 1, \ldots, N$ representing the chain at some instant τ to a new turn vector $t_j^{\tau+1}$, $j = 1, \ldots, N$ at the "next" instant $\tau + 1$. The algorithm for chain evolution

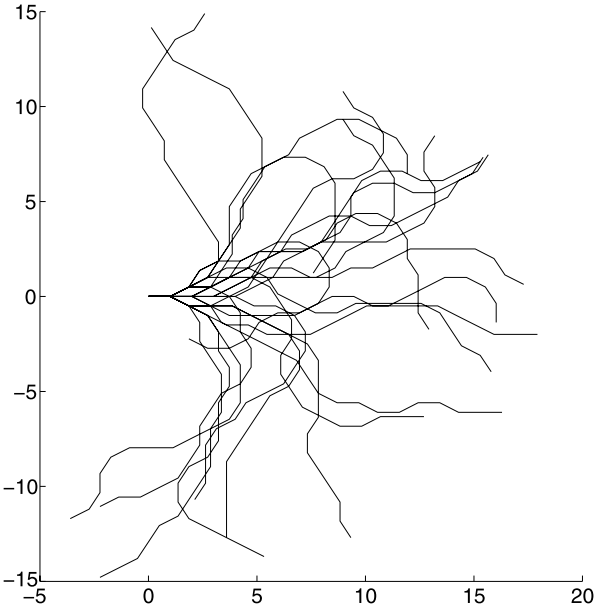


Fig. 7 Several random chains

should be fast and efficient, and should change the chain in a physically and biologically relevant manner.

A set of rules reflecting small movements of the individual chain units governs chain evolution; these small movements correspond to random fluctuations in the chain such as those induced by Brownian fluctuations. The movements allowed by the rules minimize the quantity $\sum_{j=0}^{N} \|z_j^{\tau+1} - z_j^{\tau}\|^2$ while not allowing the angle between the nodes to exceed the angle allowed in generating the chain. Thus, the angle between nodes should not have any value other than 0 or $\pm\Delta\theta$.

In the middle of the chain, $2 \le j \le N-2$, all the individual turns are compared in triads to a known list of patterns that are allowed to evolve. If these triads are members of the allowed list, the index is added to a list of eligible turns. The middle-of-the-chain turns that may evolve are shown in Fig. 8. The nodes at the end of the chain obey different evolution rules, shown in Fig. 9. If the last two turns are members of the end-of-the-chain turns shown, they are added to the list of eligible turns. Note that the terminal end node is free to move in any direction, as befits an unfettered link. To evolve the chain, a node is chosen at random from the list of eligible nodes, and the corresponding evolution is performed, thus updating the turn vector.

Note that the evolution specified by Fig. 8 moves only one node. To see this, assume that node j^* is selected from the eligible nodes. Then, the turn vector $t_j^{\tau+1} = t_j^{\tau}$ is unchanged for $j = 1, \ldots, j^* - 1$ and for $j = j^* + 1, \ldots, N-1$. For $j^* - 1 \le j \le j^* + 1$, the nodes change as shown in Fig. 8. It is evident that $\sum_{j=j^*-1}^{j^*+1} t_j^{\tau+1} = \sum_{j=j^*-1}^{j^*+1} t_j^{\tau}$. Next consider the positions $z_j^{\tau} = z_{j-1}^{\tau} + \exp(i\Delta\theta \sum_{k=0}^{j} t_k^{\tau})$ and $z_j^{\tau+1} = z_{j-1}^{\tau+1} + \exp(i\Delta\theta \sum_{k=0}^{j} t_k^{\tau+1})$. Since the sums in the exponent are identical for $j = 1, \ldots, j^* - 1$, it is clear that $z_j^{\tau+1} = z_j^{\tau}$. Specifically, $z_{j^*-1}^{\tau+1} = z_{j^*-1}^{\tau}$. From Fig. 8, then $z_{j^*+1}^{\tau+1} = z_{j^*+1}^{\tau}$. Finally, observe that for $j = j^* + 1, \ldots, N-1$ $\sum_{k=1}^{j} t_k^{\tau+1} = \sum_{k=1}^{j^*-1} t_k^{\tau+1} + \sum_{k=j*-1}^{j*+1} t_k^{\tau+1} + \sum_{k=j^*+1}^{j} t_k^{\tau+1}$. The first and last sums on the right

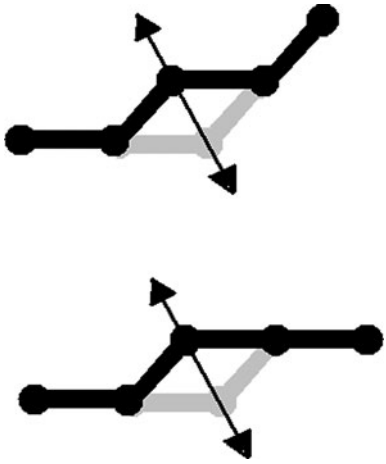
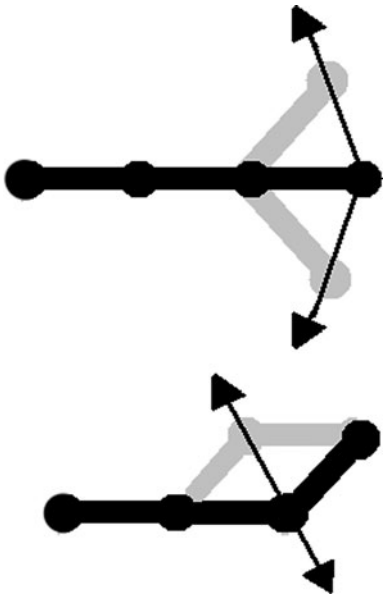


Fig. 8 Allowable motions involving one node and maintaining angles of 0, $\pm\Delta\theta$. Note that left–right and up–down symmetries are also allowed

Fig. 9 Allowable end motions



do not change, so that $\sum_{k=1}^{j^*-1} t_k^{\tau+1} = \sum_{k=1}^{j^*-1} t_k^{\tau}$ and $\sum_{k=j^*+1}^{j} t_k^{\tau+1} = \sum_{k=j^*+1}^{j} t_k^{\tau}$. Since the middle sum does not change, $z_j^{\tau+1} = z_j^{\tau}$ for $j = j^* + 1, \ldots, N-1$.

Given this method of evolution in conjunction with the possible turn evolutions shown in Figs. 8 and 9, only the two types of end evolutions result in a change of the end-to-end distance of the chain.

After implementing this algorithm in MATLAB©, it became desirable to optimize it for efficiency. In order to do this, a method of maximum likelihood sampling was implemented. Since moving end nodes are most efficient in changing end-to-end distance, the likelihood measure that was implemented was expressed as

$$p_j = \frac{e^{\lambda j/N}}{\sum e^{\lambda j/N}}$$

where p_j is the probability of picking node j from the list of eligible nodes, λ is a parameter which emphasizes the end of the chain, and the summation in the denominator is over the eligible nodes. The probability of a node being chosen is dependent of its position in the chain and the lambda value specified. If $\lambda > 0$, the closer the node is to the free (terminal) end of the chain, the more likely it is to be picked. If $\lambda = 0$, then all nodes are equally likely to be picked. The lambda values used range from 0.05 to 0.5, with 0.5 being a high value that chooses only the end nodes regularly. The program then proceeds to evolve the chain in the same manner as described above.

Using the above algorithm speeds up the calculations, but it is difficult to know which value of λ to use. One way to make the simulation more dynamic is to use a process called annealing, named after the annealing process used in metallurgy. In metallurgy, the temperatures to which a metal is subject vary; initially, the metal is

super-heated and then it is cooled slowly. This method improves ductility in the metal and lessens the failure potential. In the case of chain evolution, annealing is used to propagate major changes to the end of the chain inward to the middle of the chain. This is accomplished by changing the λvalue within each execution of the code. The code initially starts out with a high lambda value (e.g. 0.5) to evolve the end nodes and then switches to a smaller value (e.g., 0.05) to evolve the internal nodes. This process is repeated multiple times.

Brownian Migration Let the probability that the distance between the two SeqA clusters is equal to δ at time t be $P(\delta, t)$. We assume that in order to arrive at separation δ, $0 \le \delta \le L$, at time $t + \Delta t$, the clusters could have only been in three possible states at time t, either separation $\delta - \Delta\delta$, δ, or $\delta + \Delta\delta$, and increased, decreased, or stayed the same. Thus,

$$P(\delta, t + \Delta t) = P(\delta - \Delta\delta, t)\tau_{\text{up}}(\delta - \Delta\delta) + P(\delta, t)\tau_0(\delta) + P(\delta + \Delta\delta, t)\tau_{\text{down}}(\delta + \Delta\delta)$$

where $\tau_{\text{up}}(\delta)$, $\tau_0(\delta)$, and $\tau_{\text{down}}(\delta)$ are the probabilities of transition from separation d to $\delta - \Delta\delta$, δ, and $\delta + \Delta\delta$, respectively, in unit time. Note that $\tau_{\text{up}}(\delta) + \tau_0(\delta) + \tau_{\text{down}}(\delta) = 1$.

In terms of $\tau_{\text{up}}(\delta)$ and $\tau_{\text{down}}(\delta)$, we have

$$\begin{aligned} P(\delta, t + \Delta t) = {} & P(\delta, t) + P(\delta - \Delta\delta, t)\tau_{\text{up}}(\delta - \Delta\delta) \\ & - P(\delta, t)\big(\tau_{\text{up}}(\delta) + \tau_{\text{down}}(\delta)\big) \\ & + P(\delta + \Delta\delta, t)\tau_{\text{down}}(\delta + \Delta\delta) \end{aligned}$$

Brownian motion mimics the thermal bombardments and small random-like fluctuations of the cytoplasm and the DNA itself. Brownian simulations on DNA for this purpose is validated by experimental results such as those of Cunha et al. (2005), Woldringh (2002) and Woldringh and Nanninga (2006) in which tagged segments of DNA in *E. coli* cells show a Gaussian distribution of lateral motion over time indicating Brownian dynamics.

In order to determine the transition probabilities $\tau_{\text{up}}(\delta)$, $\tau_0(\delta)$, and $\tau_{\text{down}}(\delta)$, we use the ideas of chain evolution discussed above in the following way: We first generate a large number of random chains of a given length. Each of these chains has an end-to-end length sampled at random from all possible end-to-end lengths. Then we consider all the possible end-to-end length *changes* that could be produced by a motion of either end. We associate these length changes with the chain's end-to-end length, generating frequencies $f_{\text{up}}(\delta)$ for moving up, and $f_{\text{down}}(\delta)$ for moving down simply by counting all the end-to-end length changes that are greater than $1/2$ (for *up*) or less than $-1/2$ (for *down*) and dividing by the total number of length changes. This produces numerically empirical transition probabilities $\tau_{\text{up}}(\delta)$ and $\tau_{\text{down}}(\delta)$. Typical transitional probabilities as a function of the separation of the ends are shown in Fig. 10. The evolution of the probability density function from an initial spike is shown in Fig. 11. The average end-to-end distance, defined by $\langle\delta\rangle(t) = \int_0^L \delta P(\delta, t)\, d\delta$ is shown in Fig. 12 as a function of time for several different chain lengths. These average lengths evolve to equilibrium values, which are

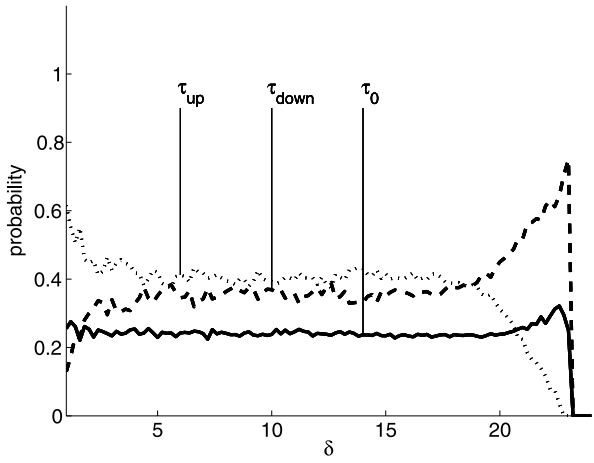


Fig. 10 Transition probabilities for $L = 25$

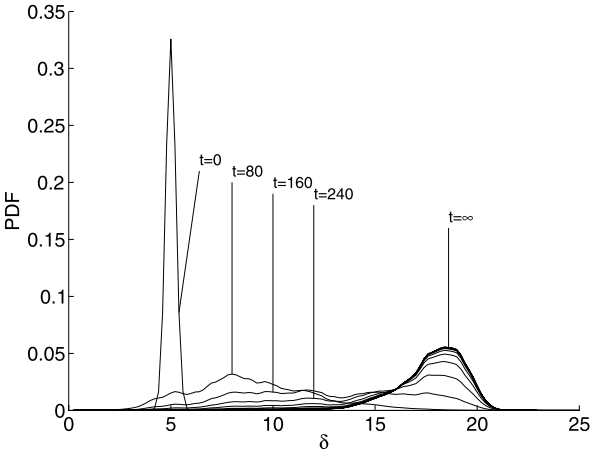


Fig. 11 Evolution from a separation of 5 units to equilibrium

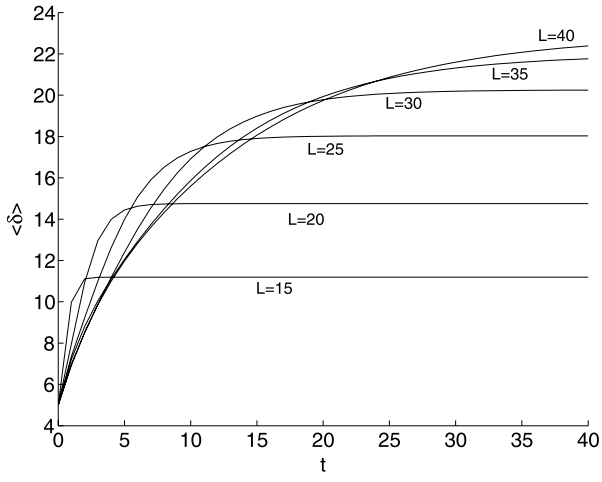


Fig. 12 Average separation of ends versus time for several different chain lengths

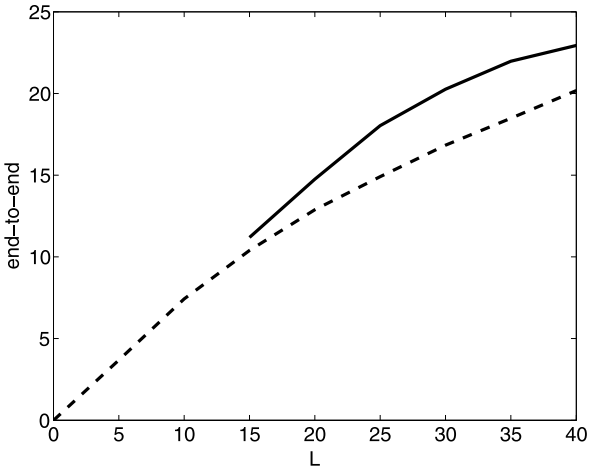


Fig. 13 Equilibrium end-to-end distance versus length (*solid curve*). Also shown is the mean end-to-end distance over all possible chains (*dashed curve*)

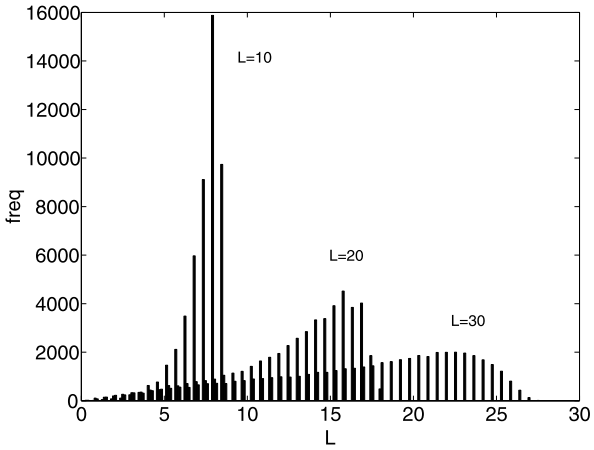


Fig. 14 Histogram of end-to-end distances for three different chain lengths

plotted versus chain length in Fig. 13. Also shown in Fig. 13 is the average end-to-end distance as a function of length, computed by generating 1000 chains, recording the end-to-end distance of each, and averaging. Figure 14 shows the histogram for the end-to-end distance for a chain of length X, along with the predicted probability density function for the separation (see Fig. 11).

3.1 Interpretation

For chain evolution, as discussed in Sect. 3, not all links are in motion at any instant. For the estimate of how fast this part of the system runs, we assume that the units that are moving are base pairs. Loops created by attaching SeqA to GATC sites would be on the order of 20 base pairs apart; consequently, we assume that there are about 20 links of the chain modeling each loop. Note that this is substantially smaller than the "persistence length" of DNA, taken to be on the order of 150 base pairs (bp) (about 50 nm) (Smith et al. 1996). There are about 3000 base pairs per μm and the

uncoiled DNA loop is about 1500 μm in length, giving a link length (bp separation) of $\delta = 0.3$ nm. Taking $k_B T \approx 2 \times 10^{-21}$ J and assuming only few (2 or 3) degrees of freedom gives a thermal velocity of 10 m/sec, and a time scale of $\Delta t = l/v \approx 5$ nsec, much smaller than a reasonable fluctuation time for the chain. Using this paradigm, however, brings into question the assumption of the number of degrees of freedom. A chain of rigid links may have two degrees of freedom per node. Note that at the scale of individual atoms, there are many degrees of freedom corresponding to bond elongations and rotations, as well as atomic motions. Thus, a molecular chain will have many more than two degrees of freedom. Since the model does not need every link to move at every conceivable time step, the energy available for vibrations according to the Boltzmann assumption may be spread through the scales of the molecules, from vibrations of individual atoms to vibrations of substructures, only one of these correspond to chain link motions of the sort that are postulated in Sect. 3. Another approach to time scaling is through observations of GFP tagged proteins attached to DNA chains. If these motions are interpreted as diffusion[1] the observed diffusivity is on the order of $D \approx 10^{-2}$ μm^2/sec. Using this value and a length of $\delta = 0.3$ nm gives a time step of $\Delta t = 10$ μsec.

Since we estimate from the chain model in Sect. 3 that it takes a few thousand time steps for a small (20 links) loop to straighten when the chain is allowed to evolve one node at a time. Thus, a short loop where several nodes could be in motion more or less simultaneously would straighten in about a second.

For Brownian migration, the case shown in Fig. 11 takes several hundred time units for the system to approach equilibrium, with a separation of about 18 units for a chain of length 25. The loop is acting as a rectifier of the motion, so that the space unit may be more like a Brownian step for the SeqA cluster. Since these clusters are relatively large, the Brownian steps will be slower than the chain time steps. Consequently, the assumption of near equilibrium for the chain that is assumed in Sect. 3 is valid. While it is not clear that a base pair represents a valid unit for this calculation, if we assume so, and if we use a diffusivity for protein clusters of several (say, 10), the cluster would have about twice the radius of a single SeqA protein. If we estimate the motion parameters by assuming the SeqA clusters are spherical with a diameter of 100 nm, the force needed to move the sphere at velocity U is $F = 6\pi\mu a U$, where μ is the viscosity and a is the sphere radius. Using $\mu = 10^{-2}$ Poise, $a = 10^{-7}$ m and $U = 10^{-5}$ m/sec, the force is on the order of $F = 0.2$ picoNewtons.

4 Conclusion

The models presented in this paper show the feasibility of two roles for SeqA. First, it is known that binding SeqA to GATC sites prevents reinitiation of DNA replication. We present a series of models that argue that the process of subsequent release of SeqA acts as a timing mechanism for reinitiation. In this model, an upswing in activity of Dam methylase could control competition between SeqA binding and methylation.

[1] Tagged SeqA-GFP in *E. coli* DNA show Brownian motion that is interpreted as "self-diffusion" on the order of 0.12 (μm)^2/sec (Cunha et al. 2005; Robertson and Smith 2007).

The timing is further affected by whether the desequestration occurs sequentially or simultaneously.

We further modeled the possibility that as DNA attached to SeqA unloops, Brownian motion causes the migration of the two daughter loops towards opposite cellular poles. Considering the time scale, the Brownian migration toward the poles is a reasonable event if "unlooping" of DNA after SeqA dislodges occurs with Brownian steps (0.1 μm/step) at 1 second increments in which it takes 60 seconds to cause a lateral motion of 1 μm. Using the model presented above, Brownian migration is quite capable of participating in the initial migration of replicated DNA toward opposing cellular poles. In the context of this model, the role of the cytoskeletal protein MreB is not clear. We postulate that the Replication Factory could interact with the cytoskeleton structure for guidance while the chain modulated by SeqA binding and release rectifies the Brownian motion of these factories through the proposed ratchet mechanism.

Acknowledgements This work is a partially the result of an Undergraduate Research Project (SH, JK, RT, and VV). The support of the National Science Foundation Division of Mathematical Sciences and Division of Biological Sciences through Grant DMS 0214585 and a Supplement to support Undergraduate Research in Biology and Mathematics is appreciated.

References

Baker, T. A., & Kornberg, A. (1988). Transcriptional activation of initiation of replication from the *E. coli* chromosomal origin: an RNA-DNA hybrid near *oriC*. *Cell*, *55*(1), 113–123.

Boye, E., Løbner-Olesen, A., & Skarstad, K. (1988). Timing of chromosomal replication in Escherichia coli. *Biochim. Biophys. Acta*, *951*(2–3), 359–364.

Boye, E., Stokke, T., Kleckner, N., & Skarstad, K. (1996). Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. *Proc. Natl. Acad. Sci. USA*, *93*(22), 12206–12211.

Boye, E., Løbner-Olesen, A., & Skarstad, K. (2000). Limiting DNA replication to once and only once. *EMBO Rep.*, *1*(6), 479–483.

Bramhill, D. (1997). Bacterial cell division. *Annu. Rev. Cell Dev. Biol.*, *13*, 395–424.

Bramhill, D., & Kornberg, A. (1988). A model for initiation at origins of DNA-replication. *Cell*, *54*(7), 915–918.

Brendler, T., et al. (2000). A case for sliding SeqA tracts at anchored replication forks during Escherichia coli chromosome replication and segregation. *EMBO J.*, *19*(22), 6249–6258.

Campbell, J. L., & Kleckner, N. (1990). *E. coli oriC* and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell*, *62*(5), 967–979.

Cunha, S., Woldringh, C. L., & Odijk, T. (2005). Restricted diffusion of DNA segments within the isolated Escherichia coli nucleoid. *J. Struct. Biol.*, *150*(2), 226–232.

Fan, J., Tuncay, K., & Ortoleva, P. J. (2007). Chromosome segregation in Escherichia coli division: a free energy-driven string model. *Comput. Biol. Chem.*, *31*(4), 257–264.

Formstone, A., & Errington, J. (2005). A magnesium-dependent mreB null mutant: implications for the role of mreB in Bacillus subtilis. *Mol. Microbiol.*, *55*(6), 1646–1657.

Freiesleben, U., Krekling, M. A., Hansen, F. G., & Løbner-Olesen, A. (2000). The eclipse period of Escherichia coli. *EMBO J.*, *19*(22), 6240–6248.

Gordon, S., Rech, J., Lane, D., & Wright, A. (2004). Kinetics of plasmid segregation in Escherichia coli. *Mol. Microbiol.*, *51*(2), 461–469.

Hiraga, S. (2000). Dynamic localization of bacterial and plasmid chromosomes. *Annu. Rev. Genet.*, *34*, 21–59.

Hiraga, S., et al. (2000). Bidirectional migration of SeqA-bound hemimethylated DNA clusters and pairing of *oriC* copies in Escherichia coli. *Genes Cells*, *5*(5), 327–341.

Jacob, F., Brenner, S., & Cuzin, F. (1963). On regulation of DNA replication in bacteria. *Cold Spring Harbor Symp. Quant. Biol.*, *23*, 329–348.

Jun, S., & Mulder, B. (2006). Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome. *Proc. Natl. Acad. Sci. USA*, *103*, 12388–12393.

Kang, S., Lee, H., Han, J. S., & Hwang, D. S. (1999). Interaction of SeqA and dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication. *J. Biol. Chem.*, *274*(17), 11463–11468.

Kruse, T., Blagoev, B., Løbner-Olesen, A., Wachi, M., Sasaki, K., Iwai, N., Mann, M., & Gerdes, K. (2006). Actin homolog MreB and RNA polymerase interact and are both required for chromosome segregation in Escherichia coli. *Genes Dev.*, *20*(1), 113–124.

Nielsen, O., & Løbner-Olesen, A. (2008). Once in a lifetime: strategies for preventing re-replication in prokaryotic and eukaryotic cells. *EMBO Rep.*, *9*(2), 151–156.

Nievera, C., Torgue, J. J., Grimwade, J. E., & Leonard, A. C. (2006). SeqA blocking of DnaA-*oriC* interactions ensures staged assembly of the *E. coli* pre-RC. *Mol. Cell*, *24*(4), 581–592.

Norris, V., Fralick, J., & Danchin, A. (2000). A SeqA hyperstructure and its interactions direct the replication and sequestration of DNA. *Mol. Microbiol.*, *37*(4), 696–702.

Onogi, T., et al. (1999). The assembly and migration of SeqA-Gfp fusion in living cells of Escherichia coli. *Mol. Microbiol.*, *31*(6), 1775–1782.

Robertson, R. M., & Smith, D. E. (2007). Self-diffusion of entangled linear and circular DNA molecules: dependence on length and concentration. *Macromolecules*, *40*(9), 3373–3377.

Skarstad, K., Boye, E., & Steen, H. B. (1986). Timing of initiation of chromosome replication in individual Escherichia coli cells. *EMBO J.*, *5*(7), 1711–1717. Erratum in: *EMBO J., 5*(11), 3074 (1986).

Skarstad, K., Lueder, G., Lurz, R., Speck, C., & Messer, W. (2000). The Escherichia coli SeqA protein binds specifically and co-operatively to two sites in hemimethylated and fully methylated *oriC*. *Mol. Microbiol.*, *36*(6), 1319–1326.

Smith, S. B., Cui, Y., & Bustamante, C. (1996). Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, *271*(5250), 795–799.

Stancheva, I., Koller, T., & Sogo, J. M. (1999). Asymmetry of Dam remethylation on the leading and lagging arms of plasmid replicative intermediates. *EMBO J.*, *18*(22), 6542–6551.

Taghbalout, A., Landoulsi, A., Kern, R., Yamazoe, M., Hiraga, S., Holland, B., Kohiyama, M., & Malki, A. (2000). Competition between the replication initiator DnaA and the sequestration factor SeqA for binding to the hemimethylated chromosomal origin of *E. coli* in vitro. *Genes Cells*, *5*(11), 873–884.

Torheim, N. K., & Skarstad, K. (1999). Escherichia coli SeqA protein affects DNA topology and inhibits open complex formation at *oriC*. *EMBO J.*, *18*(17), 4882–4888.

Waldminghaus, T., & Skarstad, K. (2009). The Escherichia coli SeqA protein. *Plasmid*. doi:10.1016/j.plasmid.2009.02.004.

Webb, C. D., et al. (1997). Bipolar localization of the replication origin regions of chromosomes in vegetative and sporulating cells of B. subtilis. *Cell*, *88*, 667–674.

Woldringh, C. L. (2002). The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Mol. Microbiol.*, *45*, 17–29.

Woldringh, C. L., & Nanninga, N. (2006). Structural and physical aspects of bacterial chromosome segregation. *J. Struct. Biol.*, *156*(2), 273–283.