Control of DNA Replication by Anomalous Reaction-Diffusion Kinetics

Michel G. Gauthier and John Bechhoefer

Department of Physics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada
(Received 3 December 2008; published 16 April 2009)

We propose a simple model for the control of DNA replication in which the rate of initiation of replication origins is controlled by protein-DNA interactions. Analyzing recent data from *Xenopus* frog embryos, we find that the initiation rate is reaction limited until nearly the end of replication, when it becomes diffusion limited. Initiation of origins is suppressed when the diffusion-limited search time dominates. To fit the experimental data, we find that the interaction between DNA and the rate-limiting protein must be subdiffusive.

DNA replication occurs at every cell cycle and its reliability is crucial for the survival of daughter cells. Although the process has been studied for decades [1,2], the recent development of molecular-combing techniques has significantly increased the available data on replication kinetics of individual cells [3,4]. Such experiments have given accurate statistics characterizing the size distribution and growth speed of the replication bubbles as a function of time. One of the best-studied cases is that of *Xenopus laevis* frog embryos, where replication is initiated stochastically at multiple locations (origins) along the genome. These origins are distributed nearly at random along the chromosomes and initiate at different times during the synthesis (*S*) phase of the cell cycle [5–7]. Averaged over the genome, the origins are triggered at a rate $I(t)$ (number of initiations per unreplicated length per time) that increases throughout most of *S* phase, before decreasing to zero at the end [8,9].

In previous work, our group introduced a formalism—inspired by the Kolmogorov-Johnson-Mehl-Avrami (KJMA) theory of phase-transition kinetics [10]—that, given $I(t)$, can predict experimental quantities such as the average size of the replicated and nonreplicated domains, the domain densities, and replication fraction [11]. We also showed how to "invert" measured domain-size statistics into an estimate for $I(t)$ [11].

While a generally increasing $I(t)$ helps limit the variability of *S*-phase completion times [12], the biological mechanisms that control the observed $I(t)$ remain unclear. Recently, Goldar *et al.* [9] proposed that the DNA-binding kinetics of a single rate-limiting protein, possibly Cdc45, could lead to the observed $I(t)$. Their model required assumptions—such as a positive correlation between replication-fork density and protein binding—whose biological origins are not always clear, and they ignored any effects due to diffusion.

Here, we show that a simple scenario, where protein-binding kinetics crosses over from reaction to diffusion limited, can explain the observed data, with no need for correlations. Our model leads to analytical equations that

we solve numerically and to predictions that can be tested. Our type of analysis also applies to the work of Goldar *et al.* [9], which relied on Monte Carlo simulations. Interestingly, the observed decrease of $I(t)$ is consistent with the anomalous, subdiffusive motion expected for protein-DNA interactions in the cell nucleus.

The KJMA method.—DNA replication statistics can be analyzed using the KJMA model of nucleation and growth [11]. A key result of this theory is the calculation of the probability $S(t)$ that a given point remains uncovered by the expanding domains as a function of time. As illustrated in Fig. 1(a) for the special case of constant growth rate v, a point X will remain uncovered by the growing domains after a time t if no nucleation has occurred in the shaded triangular spacetime area. That probability is given by

$$S(t) = \lim_{\Delta x, \Delta t \to 0} \prod_{x,t \in \triangle} [1 - I(t)\Delta x \Delta t]$$

$$= \exp\left\{-\iint_{x,t \in \triangle} I(t)dxdt\right\} = e^{-2vh(t)}, \quad (1)$$

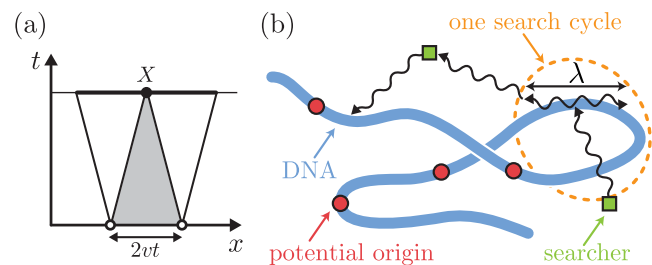where $\triangle$ represents the shaded area in Fig. 1(a) and $2vh(t)$

FIG. 1 (color online). (a) Nucleation and growth spacetime diagram. The point X is not reached by a growing domain at time t only if no nucleation occurred in the shaded area. (b) Schematic representation of the potential origin search process. The searchers (squares) diffuse in the 3D space until they nonspecifically bind to the DNA. They then search along the DNA an average distance λ. If no origin (circles) is found, the searcher dissociates and starts another search cycle. The duration of each cycle (e.g., the dashed circled region) is τ.

is the number of origins expected to fire in $\triangle$. The one-dimensional result derived in Eq. (1) can be used to estimate the fraction $f(t)$ of replicated DNA (for long molecules) as a function of time:

$$f(t) = 1 - S(t) = 1 - e^{-2vh(t)}, \qquad (2)$$

where v is now the average replication-fork velocity and $I(t)$ the replication-initiation rate per time per base pair. Here, we seek to understand the observed form of $I(t)$.

Diffusion-based search.—DNA replication can be divided into two distinct processes: the initiation ("firing") of replication origins and the propagation of replication forks away from the fired origins. Here, we model the control of replication-initiation throughout S phase using a population of proteins ("searchers"). We suppose that the proteins seek "potential origins," protein complexes previously bound to DNA ("licensed") during $G1$ phase. For *Xenopus* embryos, potential origins are assembled in excess prior to S phase [7] and may fire at some point during S phase, generating two replication forks. Alternatively, they may be passively replicated by a replication fork from another origin [13]. In our model, the initiation rate is set by the time to find and then activate potential origins. If the former time scale is slower (faster), the initiation process is diffusion (reaction) limited. The approach we propose can model these two regimes together.

We follow previous work on target searches along DNA in assuming a sequence of three- and one-dimensional explorations [14–16]. The optimization of total search time by choosing the relative amounts of the two types of diffusion has been studied in detail (see [15,17–19] for examples). Here, we focus on whether a reaction-diffusion-based search can explain the observed $I(t)$.

Let an initiation site be found after N_c search cycles, with each cycle consisting of a three-dimensional search of the cell-nucleus volume followed by a one-dimensional search along the DNA that starts from a nonspecific binding site [Fig. 1(b)]. Several experiments indicate that diffusion in the cell cytoplasm and nucleoplasm is subdiffusive [20–23]. This means that the mean-square displacement of a particle in such crowded environment scales sublinearly with time (i.e., $\langle x^2 \rangle \sim t^\alpha$ with $\alpha < 1$). The slowdown of proteins in the cell nucleus can result either from physical constrictions that act as traps or from binding interactions between the diffuser and the cell components. In both cases, long trapping times between successful moves of the diffuser are introduced, and the motion becomes subdiffusive. For these reasons, subdiffusive random walks can be modeled using Markov processes with N steps of constant size whose duration is picked from a continuous time power-law distribution with a long-time tail that goes as $\Psi(t) \sim t^{-(1+\alpha)}$ [24] with $\alpha < 1$. Here, we view each cycle of the 1D/3D diffusion as one such step which has an individual duration drawn from $\Psi(t)$. In such a case, the total time of the walk

is proportional to $N^{1/\alpha}$ [25]. For normal diffusion, the central limit theorem implies that the total time scales linearly with N. Consequently, if the search process depicted in Fig. 1(b) is subdiffusive, the average search time t_s for one searcher to find one potential origin scales as

$$t_s(t) = N_c(t)^{1/\alpha} \tau, \qquad (3)$$

where τ is the duration of one search cycle. Note that the number of cycles, $N_c(t)$, changes because of the changing number of potential origins as replication progresses. Physically, Eq. (3) incorporates the increased likelihood of a long-duration trapping event(s) as the time interval considered is increased. If the typical time of a single cycle is τ, the time for two cycles will typically be longer than 2τ because of the increased chance of a long trapping event. Note that Eq. (3) is similar to the expression proposed in Ref. [14] to model proteins searching for DNA targets.

We now evaluate the average number of search cycles needed to find a target. Let λ be the average one-dimensional length explored during one cycle. Next, we assume that successive cycles explore independent regions of the DNA (no spatial correlation between consecutive nonspecific bindings). We also assume that the binding energies are the same all along the DNA. For one particle searching for one origin, the probability to find a target site after one cycle is $P_c \approx 1 - \exp[-1/L']$, with $L'(t) = L[1 + f(t)]/\lambda$ and where L is the genome length and $f(t)$ is the fraction of replicated DNA as a function of time t since the beginning of replication. The $1 + f(t)$ term accounts for nonspecific binding to both unreplicated and newly replicated DNA. Let the average number of available prelicensed origins—origins that have neither fired nor been passively replicated—at time t be denoted by $N_o(t)$. Assuming a constant density ℓ_0^{-1} of potential origins along unreplicated DNA, we have that $N_o(t) = [1 - f(t)]L/\ell_0$. With these $N_o(t)$ potential origins, $P_c \approx 1 - \exp[-N_o(t)/L']$. The probability of not finding any origin after N_c cycles is given by $(1 - P_c)^{N_c} \approx \exp[-N_c N_o(t)/L']$. Then, the average number of cycles to find one origin, $\bar{N}_c$, is obtained by solving $(1 - P_c)^{\bar{N}_c} \approx e^{-1}$, which gives

$$\bar{N}_c(t) \approx \frac{L'}{N_o(t)} = \frac{[1 + f(t)]\ell_0}{[1 - f(t)]\lambda} = \frac{\ell_0}{\lambda}[2e^{2vh(t)} - 1], \quad (4)$$

where we assumed large L so we could use Eq. (2) for $f(t)$.

Control of DNA replication.—In order to relate Eq. (3) to the replication process, we first assume that the searcher adsorption and desorption rates are the same in both replicated and unreplicated regions of the DNA. In addition, we assume that the diffusion constant is the same all along the DNA chain. Then τ and λ are constant, as well. Combining Eqs. (3) and (4), we get

$$t_s(t) = \bar{N}_c(t)^{1/\alpha} \tau = \left(\frac{\ell_0}{\lambda}\right)^{1/\alpha} [2e^{2vh(t)} - 1]^{1/\alpha} \tau, \quad (5)$$

which corresponds to the time for *one* searcher to find *one* potential origin. We can also allow for a time t_r for an origin to initiate (or activate) after a search particle has found the target site. The average time between two successive initiations from the same searcher is then $(t_s + t_r)$. If t_r is constant throughout time, the initiation function (initiations/time/length-of-unreplicated-DNA) for N_s searchers is

$$I(t) = \frac{N_s(t)}{L[1 - f(t)][t_s(t) + t_r]} = \frac{N_s(t)e^{2vh(t)}}{L[t_s(t) + t_r]}, \quad (6)$$

where, again, Eq. (2) was used to replace $f(t)$. Note that, following Ref. [9], we allow for nuclear import of searchers by letting $N_s = N_s(t)$. (Replication factors accumulate in the nucleus during S phase [26].) If we now assume a constant rate κ_s of nuclear import of searchers, $N_s(t) = \kappa_s t$ [9], we can combine Eqs. (5) and (6) to define a dimensionless initiation function

$$\frac{I(t)}{v/L^2} = \frac{\frac{L}{v}\kappa_s t e^{2vh(t)}}{(\frac{\ell_o}{\lambda})^{1/\alpha}[2e^{2vh(t)} - 1]^{1/\alpha}\tau + t_r}$$
$$= \frac{I_0 t e^{2vh(t)}}{\mathcal{T}_0[2e^{2vh(t)} - 1]^{1/\alpha} + 1} = \frac{I(t)}{\mathcal{T}(t) + 1}, \quad (7)$$

where we used the measured genome size L and fork velocity v to scaled $I(t)$. Equation (7) depends on three free parameters: α, $I_0 = \kappa_s L/vt_r$, and $\mathcal{T}_0 = (\ell_0/\lambda)^{1/\alpha} \times (\tau/t_r)$. The scaled time $\mathcal{T}(t)$ is the ratio of the reaction rate to the search rate, $t_s(t)/t_r$, while $I(t)$ represents the scaled initiation rate in the reaction-limited regime [i.e., $\mathcal{T}(t) \ll 1$]. Since $I(t) = d^2h(t)/dt^2$, Eq. (7) can be integrated to find $h(t)$, with initial conditions $h(0) = \dot{h}(0) = 0$.

Two comments about Eq. (7): First, we assume no lag between successful origin triggering and the beginning of the next search by a protein. Including a lag is straightforward, but the precision of present experimental data does not justify another free parameter. Second, for $t \to \infty$, $I(t) \sim \exp[2vh(t)(1 - 1/\alpha)]$. Since $I \geq 0$, it follows that $h > 0$ and thus $I \to 0$ at long times if $\alpha < 1$. In other words, subdiffusion can explain the observation that $I \to 0$ at long times. If diffusion were ordinary, then one expects $I \to N_s(t)\lambda/2L\tau\ell_0$.

Comparison with experimental data.—The solid line in Fig. 2 presents the fit [27] of the *Xenopus* data [28] to the differential equation (7). The anomalous exponent $\alpha = 0.75 \pm 0.02$, which is consistent with the range of exponents, 0.5–0.85, obtained from fluorescence correlation spectroscopy (FCS) measurement of the diffusion coefficients in eukaryotic cells [20–22]. The value we obtain from the fit is highly correlated with the value we use for fork velocity. However, we can vary v by $\pm 50\%$ and still obtain an exponent within the FCS observed range.
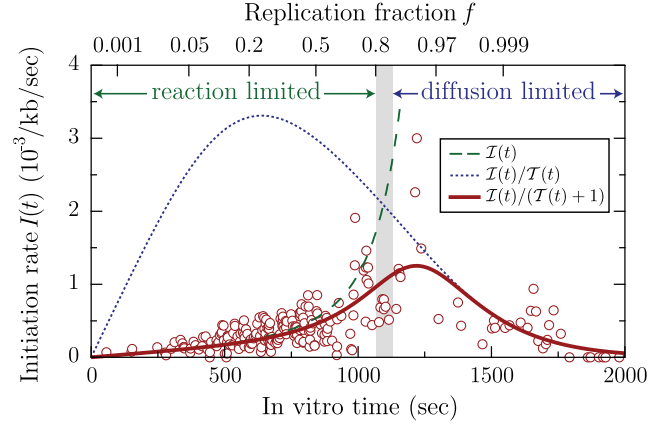

FIG. 2 (color online). *In vitro* origin-initiation data for *Xenopus* embryo replication. Circles denote experimental data (Fig. 7a in Ref. [9]), while the solid curve results from a least-squares fit based on Eq. (7), with $\alpha = 0.75 \pm 0.02$, $I_0 = (3.1 \pm 0.2) \times 10^8$ s^{-1}, and $\mathcal{T}_0 = 0.04 \pm 0.01$. The gray zone denotes the crossover from reaction- to diffusion-limited regimes ($f = 0.83 \pm 0.03$). We fixed the fork velocity at $v = 0.6$ kb/min [35] and genome length at $L = 3.1 \times 10^6$ kb [36].

The fit in Fig. 2 has three free parameters: α, I_0, and $\mathcal{T}_0$. The latter two depend on five physical variables (ℓ_0, t_r, τ, λ, and κ_s). To test whether fit values are reasonable, we need prior estimates of three of those variables. First, potential origins are associated with the hexamer MCM2-7, the helicase that unwinds DNA when an origin fires [7,29]. For *Xenopus* embryo cells, their average separation is estimated to be 0.5–0.8 kb [30] although clustering might increase the effective separation of potential origins to as much as 2 kb [31]. We thus estimate $\ell_0 = 0.5$–2 kb. Second, we estimate $\tau = 0.3$–5 ms and $\lambda = 0.02$–0.09 kb by analogy with transcription factors (TF) that search for DNA binding sites, a similar mechanism [32]. Using the fit values for I_0 and $\mathcal{T}_0$, we estimate $\kappa_s = 0.4$–10 searchers/s and $t_r = 0.3$–10 s.

Figure 2 also shows the border (gray zone) between reaction- and diffusion-limited regimes. Solving $\mathcal{T}(t) = 1$ (reaction time equals search time), we estimated a crossover at $f = 0.83 \pm 0.03$. Thus, most of the replication process occurs during the reaction-limited regime, where searchers spend most of their time bound to a potential origin, waiting to trigger it. In this regime, the number of activated origins $\approx t_{rep}^2\kappa_s/t_r \approx 10^6$ (t_{rep} is the replication time), which is close to the estimate of 5×10^5 origins [5]. At the end of S phase, $t_s(t)$ diverges as potential origins disappear, and $I(t)$ is diffusion limited.

Discussion.—We have presented a new model to describe the replication-initiation rate. Based on the diffusion of proteins that search for potential origins along the genome, our model reproduces the observed $I(t)$ throughout the whole replication process, including a steady increase through most of S phase, followed by a sharp decrease to zero at the end of the S phase. In our model,

P H Y S I C A L R E V I E W L E T T E R S

this transition is caused by the crossover from the reaction- to diffusion-limited regime. Ignoring diffusion, the initiation rate would have the form shown by the dashed line in Fig. 1(b).

We demonstrated that the subdiffusive motion of proteins in the cell nucleus naturally accounts for the observed decrease of $I(t)$ at the end of S phase. Such subdiffusive protein-DNA interactions agree with recent observations of diffusion in the cell. It is not clear, for DNA replication, whether the diffusion of proteins or the DNA itself dominates the search process, and proponents may be found for both views [33]. But subdiffusion is expected in *both* cases: If protein diffusion dominates, the crowded molecular environment of the cell interior leads to trapping by nonspecific-binding sites, which results in subdiffusion [20–23]. If DNA diffusion dominates, the dynamics of monomers within the DNA chain shows anomalous diffusion. In particular, single-stranded DNA shows Rouse dynamics (with $\alpha \approx 1/2$), while double-stranded DNA shows Zimm dynamics (with $\alpha \approx 2/3$) [34]. The dynamics of an interior location of chromatin have yet to be measured, but one anticipates subdiffusive motion.

Our model can be tested experimentally. Changing the density of potential origins or the diffusion coefficient would give easily traceable signature effects on $I(t)$. One can identify the rate-limiting protein by using mutants with altered expression levels of the candidate protein. Fusing the proteins with a fluorescent marker would allow one to check the nuclear-import hypothesis.

Finally, our model is straightforward to generalize as better data become available. The decrease of fork speed v throughout S phase observed *in vitro* [35] can easily be accommodated. The dynamics of the number of searchers can be modified, as well. Equation (7) indicates that a decrease of $N_s(t)$ may also explain the observed decrease of $I(t)$ without the need for subdiffusion. However, a simple degradation mechanism such as $\dot{N}_s(t) = \kappa_s - \gamma N_s(t)$, with γ the degradation rate, implies $N_s \rightarrow \kappa_s/\gamma$. Removing or inactivating searchers at the end of S phase would thus require a further mechanism for actively controlling the import rate κ_s. While nothing yet rules out these more complicated possibilities, the simpler model advanced here accounts for the known experimental data and is readily testable.

---

[1] J. Cairns, J. Mol. Biol. **6**, 208 (1963).

[2] A. B. Blumenthal, H. J. Kriegstein, and D. S. Hogness, Cold Spring Harbor Symposia on Quantitative Biology (1933-) [Proceedings] **38**, 205 (1974).

[3] J. Herrick and A. Bensimon, Chromosome Research : An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology **7**, 409 (1999).

[4] X. Michalet *et al.*, Science **277**, 1518 (1997).

[5] J. Herrick *et al.*, J. Mol. Biol. **320**, 741 (2002).

[6] O. Hyrien and M. Méchali, EMBO J. **12**, 4511 (1993).

[7] J. Walter and J. W. Newport, Science **275**, 993 (1997).

[8] J. Herrick *et al.*, J. Mol. Biol. **300**, 1133 (2000).

[9] A. Goldar *et al.*, PLoS ONE **3**, e2919 (2008).

[10] A. N. Kolmogorov, Bull. Acad. Sci. USSR, Phys. Ser. **1**, 335 (1937).

[11] S. Jun, H. Zhang, and J. Bechhoefer, Phys. Rev. E **71**, 011908 (2005); S. Jun and J. Bechhoefer, Phys. Rev. E **71**, 011909 (2005); H. Zhang and J. Bechhoefer, Phys. Rev. E **73**, 051903 (2006).

[12] J. Bechhoefer and B. Marshall, Phys. Rev. Lett. **98**, 098105 (2007).

[13] J. Herrick and A. Bensimon, Chromosoma **117**, 243 (2008).

[14] A. Cherstvy, A. Kolomeisky, and A. Kornyshev, J. Phys. Chem. B **112**, 4741 (2008).

[15] G. Guigas and M. Weiss, Biophys. J. **94**, 90 (2008).

[16] Z. Wunderlich and L. A. Mirny, Nucleic Acids Res. **36**, 3570 (2008).

[17] K. V. Klenin *et al.*, Phys. Rev. Lett. **96**, 018104 (2006).

[18] S. E. Halford, Nucleic Acids Res. **32**, 3040 (2004).

[19] M. A. Lomholt, T. Ambjörnsson, and R. Metzler, Phys. Rev. Lett. **95**, 260603 (2005).

[20] I. Golding and E. C. Cox, Phys. Rev. Lett. **96**, 098102 (2006).

[21] G. Guigas, C. Kalla, and M. Weiss, Biophys. J. **93**, 316 (2007).

[22] M. Weiss *et al.*, Biophys. J. **87**, 3518 (2004).

[23] J. A. Dix and A. S. Verkman, Annu. Rev. Biophys. **37**, 247 (2008).

[24] H. Scher and E. Montroll, Phys. Rev. B **12**, 2455 (1975).

[25] J. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).

[26] J. Walter, L. Sun, and J. Newport, Mol. Cell **1**, 519 (1998).

[27] We used the software package IGOR PRO (WaveMetrics, Inc). The fit and errors reported here were obtained using the unbinned data and the residuals.

[28] Equation (7) is valid for long genomes, with $L/(v t_{rep}) \gg 1$. Here, $L = 3.07 \times 10^6$ kb, $v = 0.6$ kb/min, and a typical replication time t_{rep} is 40 min, giving $L/(v t_{rep}) = \mathcal{O}(10^5)$.

[29] E. E. Arias and J. C. Walter, Genes Dev. **21**, 497 (2007).

[30] S. C.-H. Yang and J. Bechhoefer, Phys. Rev. E **78**, 041917 (2008).

[31] J. J. Blow (private communication).

[32] J. Elf, G.-W. Li, and X. S. Xie, Science **316**, 1191 (2007).

[33] P. R. Cook, *Principles of Nuclear Structure and Function* (Wiley, New York, 2001), Chap. 3.

[34] R. Shusterman, S. Alon, T. Gavrinyov, and O. Krichevsky, Phys. Rev. Lett. **92**, 048303 (2004).

[35] K. Marheineke and O. Hyrien, J. Biol. Chem. **279**, 28071 (2004).

[36] C. Thiébaud and M. Fischberg, Chromosoma **59**, 253 (1977).