doi:10.1016/S0022-2836(02)00522-3 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 320, 741–750

JMB

Kinetic Model of DNA Replication in Eukaryotic Organisms

John Herrick[1], Suckjoon Jun[2], John Bechhoefer[2]* and Aaron Bensimon[1]*

[1]*Unité de Stabilité des Génomes Département Structure et Dynamique des Génomes Institut Pasteur 25-28 rue du Dr Roux 75724 Paris Cedex 15, France*

[2]*Department of Physics Simon Fraser University Burnaby, BC V5A 1S6, Canada*

**Corresponding authors*

We formulate a kinetic model of DNA replication that quantitatively describes recent results on DNA replication in the *in vitro* system of *Xenopus laevis* prior to the mid-blastula transition. The model describes well a large amount of different data within a simple theoretical framework. This allows one, for the first time, to determine the parameters governing the DNA replication program in a eukaryote on a genome-wide basis. In particular, we have determined the frequency of origin activation in time and space during the cell cycle. Although we focus on a specific stage of development, this model can easily be adapted to describe replication in many other organisms, including budding yeast.

Keywords: DNA replication; replicon; S-phase; *Xenopus laevis*; molecular combing

Introduction

Although the organization of the genome for DNA replication varies considerably from species to species, the duplication of most eukaryotic genomes shares a number of common features:

(1) DNA is organized into a sequential series of replication units, or replicons, each of which contains a single origin of replication.[1,2]
(2) Each origin is activated not more than once during the cell-division cycle.
(3) DNA synthesis propagates at replication forks bidirectionally from each origin.[3]
(4) DNA synthesis stops when two newly replicated regions of DNA meet.

Understanding how these parameters are coordinated during the replication of the genome is essential for elucidating the mechanism by which S-phase is regulated in eukaryotic cells. Here, we formulate a stochastic model based on these observations that yields a mathematical description of the process of DNA replication and provides a convenient way to use the full statistics gathered in any particular replication experiment. It allows one to deduce accurate values for the parameters that regulate DNA replication in the *Xenopus laevis* replication system, and it can be generalized to describe replication in any other eukaryotic system. This type of model has also been shown to apply for the case of RecA polymerizing on a single molecule of DNA.[4] The model, as described in Methods below, turns out to be formally equivalent to a well-known stochastic description of the kinetics of crystal growth, which allows us to draw on a number of previously derived results and, perhaps equally important, suggests a vocabulary that we find useful and intuitive for understanding the process of replication.

Since the kinetics of DNA replication in any cell system depends on two fundamental quantities, replication fork velocity and initiation frequency, one of the principal goals of this kind of analysis is to derive accurate values for these quantities, including any variation, during the course of S-phase. As replicon size and the duration of S-phase depend on the values of these parameters, this information is indispensable for understanding the mechanisms regulating S-phase in any given cell system.[5–12]

Abbreviations used: KJMA, Kolmogorov–Johnson–Mehl–Avrami; ORC, origin recognition complex; XORC, *Xenopus* origin recognition complex; ARS, autonomous replication sequence element; POR, potential origin of replication; pre-RC, pre-replication complex.

E-mail addresses of the corresponding authors: johnb@sfu.ca; abensim@pasteur.fr

Results

Summary of the *X. laevis* replication experiment

Here, we describe recent experimental results obtained on the kinetics of DNA replication in the

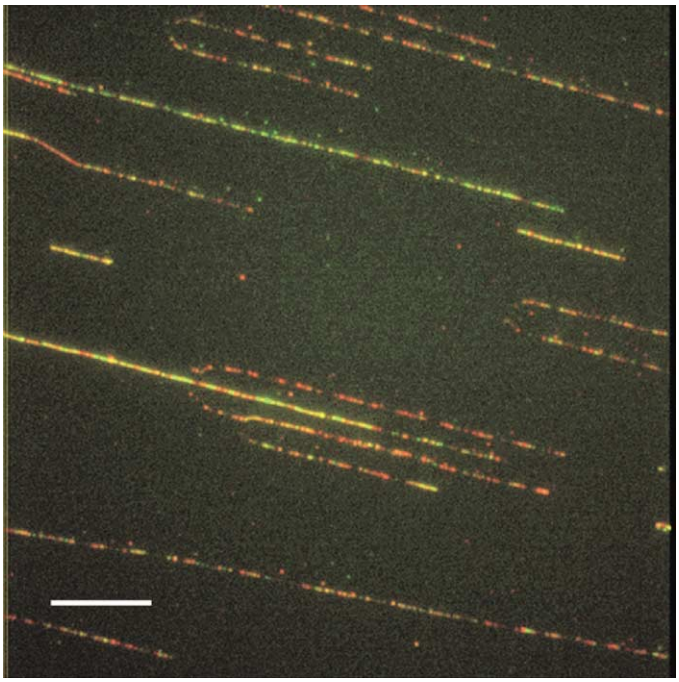

Figure 1. A fluorescence micrograph (bar represents 20 μm). Early replicating sequences labeled with biotin-dUTP are visualized using red fluorescing antibodies (Texas Red). Later replicating sequences are in addition labeled with dig-dUTP and visualized using green (FITC) fluorescing antibodies.

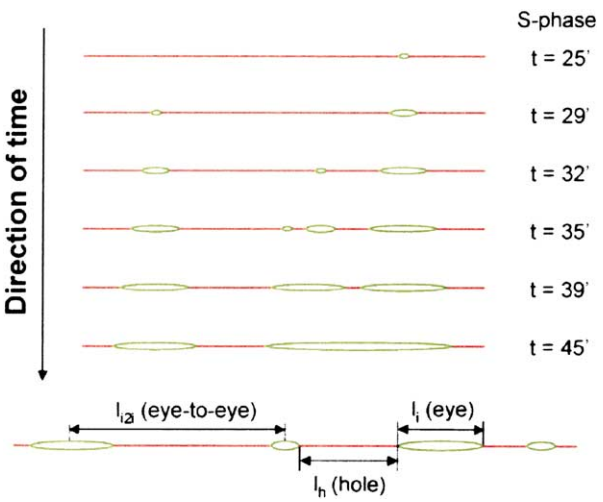

Figure 2. Schematic representation of labeled and combed DNA molecules. Since replication initiates at multiple dispersed sites throughout the genome, the DNA can be differentially labeled, so that each linearized molecule contains alternating subregions stained with either one or both dyes. The bubbles correspond to sequences synthesized in the presence of a single dye (red). The green segments correspond to those sequences that were synthesized after the second dye (green) was added. The result is an unambiguous distinction between eyes and holes (earlier and later replicating sequences) along the linearized molecules. Replication is assumed to have begun at the midpoints of the bubble sequences and to have preceded bidirectionally from the site where DNA synthesis was initiated. Measurements between the centers of adjacent eyes provide information about replicon sizes (eye-to-eye distances). The fraction of the molecule already replicated by a given time, $f(\tau)$, is determined by summing the lengths of the bubbles and dividing that by the total length of the respective molecule.

well-characterized *X. laevis* cell-free system.[13,14] One of the main goals of this paper will be to show that using the theoretical approach described below, one can extract more information, and more reliably, than before from such experiments.

In the *Xenopus* replication experiments, fragments of DNA that have completed one cycle of replication are stretched out on a glass surface using molecular combing.[15–17] Typical two-color epifluorescence images of the combed DNA are shown in Figure 1. The DNA that has replicated prior to some chosen time t is labeled with a single fluorescent dye, while DNA that replicated after that time is labeled with two dyes. The result is a series of samples, each of which corresponds to a different time t during S-phase. Using an optical microscope, one can directly measure eye, hole, and eye-to-eye lengths at that time. We can thus monitor the evolution of genome duplication from time point to time point, as DNA synthesis advances (see Figure 2).

Cell-free extracts of eggs from *X. laevis* support the major transitions of the eukaryotic cell cycle, including complete chromosome replication under normal cell-cycle control and offer the opportunity to study the way that DNA replication is coordinated within the cell cycle. In the experiment, cell extract was added at $t = 2$ minutes, and S-phase began 15–20 minutes later. DNA replication was monitored by incorporating two different fluorescent dyes into the newly synthesized DNA. The first dye was added before the cell enters S-phase in order to label the entire genome. The second dye was added at successive time points $t = 25$, 29, 32, 35, 39, and 45 minutes, in order to label the later replicating DNA. DNA taken from each time point was combed, and measurements were made

on replicated and unreplicated regions. The experimental details are described elsewhere,[13] but the approach is similar to DNA fiber autoradiography, a method that has been in use for the last 30 years.[18,19] Indeed, the same approach has recently been adapted to study the regulatory parameters of DNA replication in HeLa cells.[20] Molecular combing, however, has the advantage that a large amount of DNA may be extended and aligned on a glass slide which ensures significantly better statistics (over several thousand measurements corresponding to several hundred genomes per coverslip). Indeed, the molecular combing experiments provide, for the first time, easy access to the quantities of data necessary for testing models such as the one advanced here.

Generalization of the model to account for specific features of the *X. laevis* experiment

The experimental results obtained on the kinetics of DNA replication in the *in vitro* cell-free system of *X. laevis*[13,14] were analyzed using the kinetic model developed below. In formulating that model, we found that we had to take into account explicitly a number of observations that are peculiar to the particular experiment analyzed:

(1) One goal of the experiment is to measure the initiation function $I(\tau)$, which is the probability of initiating an origin at time τ, per unit length of unreplicated DNA. The simplest assumptions, in terms of our model, would be that either I is peaked at or near $\tau = 0$ (all origins initiated at the beginning of S-phase) or $I(\tau) =$ constant (origins initiated at constant rate throughout S-phase). However, neither assumption turns out to be consistent with the data analyzed here; thus, we formulated our model to allow for arbitrary initiation patterns and deduced an estimate for $I(\tau)$ directly from the data. We note that initiation is believed to occur synchronously during the first half of S-phase in *Drosophila melanogaster* early embryos.[10,21] Initiation in the myxomycete *Physarum polycephalum*, on the other hand, occurs in a very broad temporal window, suggesting that initiation occurs continuously throughout S-phase.[5] Finally, recent observations suggest that in *X. laevis* early embryos nucleation may occur with increasing frequency as DNA synthesis advances.[13,14] By choosing an appropriate form for $I(\tau)$, one can account for any of these scenarios. Below, we show how measured quantities may, using the model, be inverted to provide an estimate for $I(\tau)$.

(2) The basic form of the model assumes implicitly that the DNA analyzed began replication at $\tau = 0$, but this may not be so, for two reasons:
   (i) In the experimental protocols, the DNA analyzed comes from approximately 20,000 independently replicating nuclei. Before each genome can replicate, its nuclear membrane must form, along with, presumably, the replication factories. This process takes 15–20 minutes.[22–24] Because the exact amount of time can vary from cell to cell, the DNA analyzed at time t in the laboratory may have started replicating over a relatively wide range of times.
   (ii) In eukaryotic organisms, origin activation may be distributed in a programmed manner throughout the length of S-phase, and, as a consequence, each origin is turned on at a specific time (early and late).[25]

In the current experiment, the lack of information about the locations of the measured DNA segments along the genome means that we cannot distinguish between asynchrony due to reasons (i) or (ii). We can however account for their combined effects by introducing a starting-time distribution $\phi(t')$, which is the probability, for whatever reason, that a given piece of analyzed DNA began replicating at time t' in the laboratory. Using our model, we can directly extract the starting time distribution from the data. The models described above assumed that statistics could be calculated on infinitely long segments of DNA. In the experimental approach, the combed DNA is broken down into relatively short segments (100 kb, typically). Although it is difficult to account for this effect analytically, we wrote a Monte-Carlo simulation that can mimic such "finite-size" effects. The experiments are all analyzed using an epifluorescence microscope to visualize the fluorescent tracks of combed DNA on glass slides. The spatial resolution (≈ 0.3 μm) means that smaller signals will not be detectable. Thus, two replicated segments separated by an unreplicated region of size < 0.3 μm will be falsely assumed to be one longer replicated segment. We accounted for this in the Monte-Carlo simulations by calculating statistics on a coarse lattice whose size equaled the optical resolution, while the simulation itself takes place on a finer lattice.

Application of the kinetic model to the analysis of DNA replication in *X. laevis*

Using the generalizations discussed above, we analyzed recent results obtained on DNA replication in the *X. laevis* cell-free system. DNA taken from each time point was combed, and measurements were made on replicated and unreplicated regions. Statistics from each time point were then compiled into six histograms (one for each time point) of the distribution $\rho(f, t)$ of replicated fractions f at time t (Figure 3).

One can immediately see from Figure 3 the need to account for the spread in starting times. If all the segments of DNA that were analyzed had started

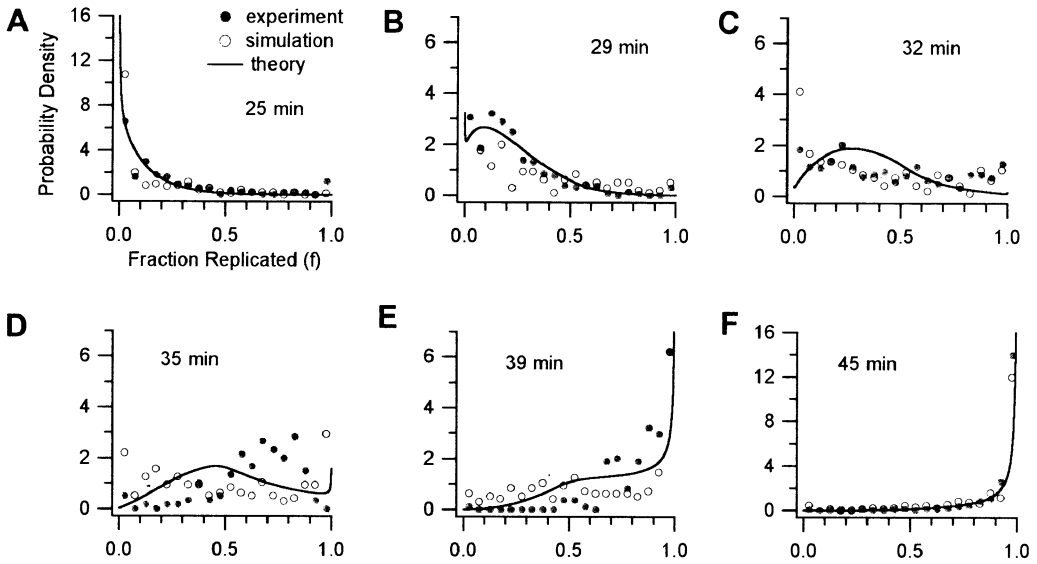

Figure 3. $\rho(f, t)$ distributions for the six time points. The curves show the probability that a molecule at a given time point (a–f) has undergone a certain amount of replication before the second dye was added. The filled circles represent the experimental data. The results of the Monte-Carlo simulation are shown in open circles; analytical curves are the global fitting.

replicating at the same time, then the distributions would have been concentrated over a very small range of f. But, as one can see in Figure 3(c), some segments of DNA (within the same time point) have already finished replicating ($f = 1$) before others have even started ($f = 0$). This spread is far larger than would be expected on account of the finite length of the segments analyzed.

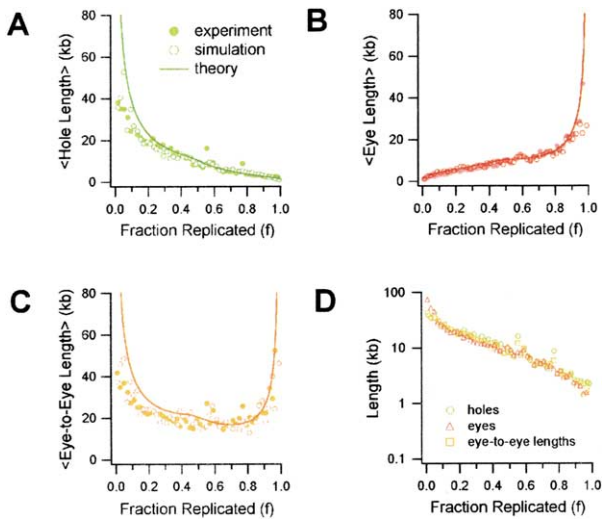

Figure 4. Mean quantities *versus* replication fraction: (a) average hole size $\ell_h(f)$; (b) average eye size $\ell_i(f)$; (c) average eye-to-eye size $\ell_{i2i}(f)$. Filled circles are data; open circles are from the Monte-Carlo simulation; the continuous curve is a least-squares fit, based on a two-segment $I(\tau)$; (d) curves in (a)–(c) collapsed onto a single plot, confirming mean-field hypothesis. (The discrepancies near $f = 0$ and 1 reflect measurement errors. Very small eyes or holes may be missed because of limited optical resolution; very large eyes or holes may be eliminated because of finite segment sizes.)

Because of the need to account for the spread in starting times, it is simpler to begin by sorting data by the replicated fraction f of the measured segment. We thus assume that all segments with a similar fraction f are at roughly the same point in S-phase, an assumption that we can check by partitioning the data into subsets and redoing our measurements on the subsets. In Figure 4(a)–(c), we plot the mean values ℓ_h, ℓ_i, and ℓ_{i2i} against f. We then find $f(\tau)$, $I(\tau)$, and the cumulative distribution of lengths between activated origins of replication, $I_{tot}(\tau)$ (see Figure 5).

The direct inversion for $I(\tau)$ (Figure 5(b)) shows several surprising features: first, origin activation takes place throughout S-phase and with increasing probability (measured relative to the amount of unreplicated DNA), as recently inferred from a cruder analysis of data from the same system using plasmid DNA.[14] Second, about halfway through S-phase, there is a marked increase in initiation rate, an observation that, if confirmed, would have biological significance. It is not known what might cause a sudden increase (break point) in initiation frequency halfway through S-phase. The increase could reflect a change in chromatin structure that may occur after a given fraction of the genome has undergone replication. This in turn may increase the number of potential origins as DNA synthesis advances.[26]

The smooth curves in Figure 4(a)–(c) are fits based on the model, using an $I(\tau)$ that has two linearly increasing regions, with arbitrary slopes and "break point" (three free parameters). The fits

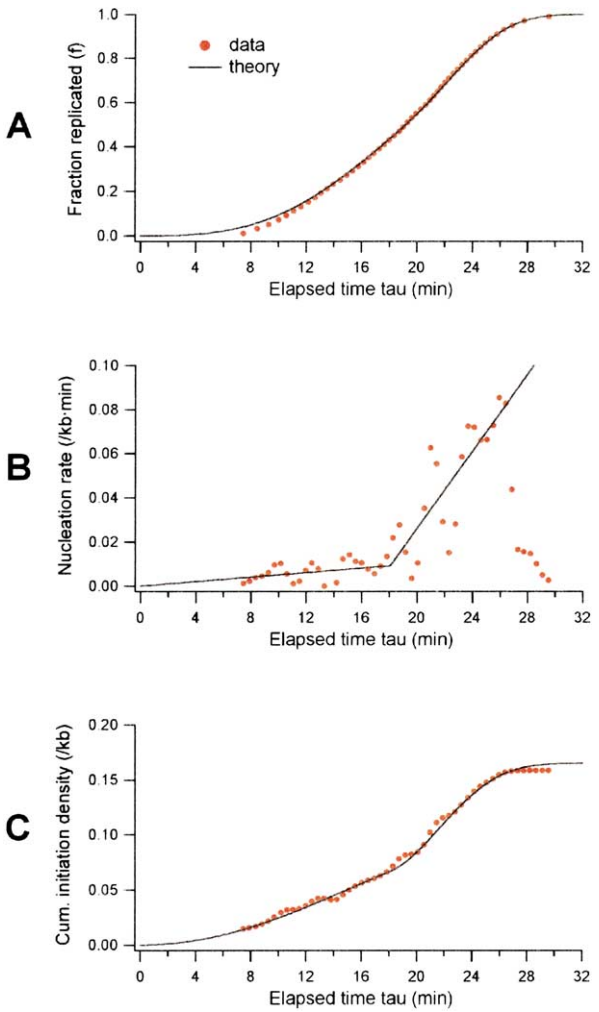

Figure 5. (a) Fraction of replication completed, $f(\tau)$. Red points are derived from the measurements of mean hole, eye, and eye-to-eye lengths. Black curve is an analytic fit (see below). (b) Initiation rate $I(\tau)$. The large statistical scatter arises because the data points are obtained by taking two numerical derivatives of the $f(\tau)$ points in (a). (c) Integrated origin separation, $I_{tot}(\tau)$, which gives the average distance between all origins activated up to time τ. In (a)–(c), the black curves are from fits that assume that $I(\tau)$ has two linear regimes of different slopes. The form we chose for $I(\tau)$ was the simplest analytic form consistent with the data in (b). The parameters for the least-squares fits (slopes I_1 and I_2, break point τ_1) are obtained from a global fit to the eight data sets in Figures 3(a)–(f) and 4(a) and (b), i.e. $\rho(f)$ from six time points, $\ell_h(f)$ and $\ell_i(f)$.

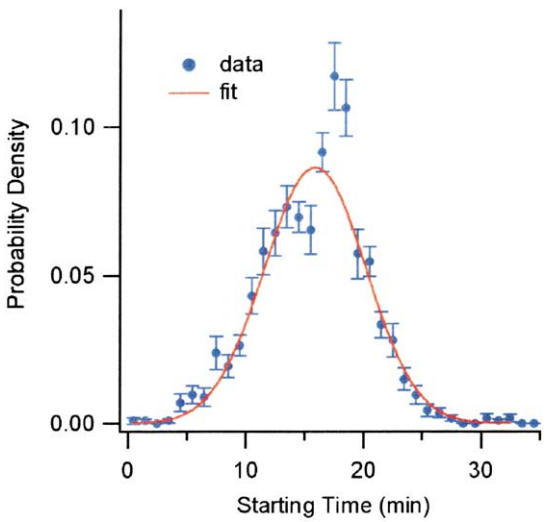

Figure 6. Starting-time distribution $\phi(t)$. Continuous curve is a least-squares fit to a Gaussian distribution.

are quite good, except where the finite size of the combed DNA fragments becomes relevant. For example, when mean hole, eye, and eye-to-eye lengths exceed about 10% of the mean fragment size, larger segments in the distribution for $\ell_h(f)$, etc., are excluded and the averages are biased down. We confirmed this with the Monte-Carlo simulations, the results of which are overlaid on the experimental data. The finite fragment size in the simulation matches that of the experiment, leading to the same downward bias. In Figure 5, we overlay the fits on the experimental data. We emphasize that we obtain $I(\tau)$ directly from the data, with no fit parameters, apart from an overall scaling of the time axis. The analytical form is just a model that summarizes the main features of the origin-initiation rate we determine *via* our model, from the experimental data. The important result is $I(\tau)$.

From the maximum of $I_{tot}(\tau)$, we find a mean spacing between activated origins of 6.3($\pm$0.3) kb, which is much smaller than the minimum mean eye-to-eye separation 14.4($\pm$1.5) kb. In our model, the two quantities differ if initiation takes place throughout S-phase, as coalescence of replicated regions leads to fewer domains, and hence fewer inferred origins (see the note below equation (5)). The mean eye-to-eye separation is of particular interest because its inverse is just the domain density (number of active domains per length), which can be used to estimate the number of active replication forks at each moment during S-phase. For example, the saturation value of I_{tot} corresponds to the maximum number (about 480,000/genome) of active origins of replication. Since there are about 400 replication foci/cell nucleus, this would indicate a partitioning of approximately 1200 origins (or, equivalently, about 7.5 Mb) per replication focus.[22,27]

The distribution of f values in the $\rho(f, t)$ plots can be used to deduce the starting-time distribution ($\phi(t')$), along with the fork velocity v (Figure 6). The spread in starting times ϕ is consistent with a Gaussian distribution, with a mean of 15.9($\pm$0.6) minutes and a standard deviation of 6.1($\pm$0.6) minutes. For the fork velocity, we find $v = 615(\pm 35)$ bases/minutes, in excellent agreement with previous estimates.[28,29] As with the f data, we extracted $\phi(t)$ and v from a global fit to data from all six time points.

Discussion

Initiation throughout S-phase

The view that we are led to here, of random initiation events occurring continuously during the replication of *Xenopus* sperm chromatin in egg extracts, is in striking contrast to what has until recently been the accepted view of a regular periodic organization of replication origins throughout the genome.[8,9,30,31] For a discussion of experiments that raise doubts on such a view, see Berezney *et al.*[32] The application of our model to the results of Herrick *et al.* indicates that the kinetics of DNA replication in the *X. laevis in vitro* system closely resembles that of genome duplication in early embryos. Specifically, we find that the time required to duplicate the genome *in vitro* agrees well with what is observed *in vivo*. In addition, the model yields accurate values for replicon sizes and replication fork velocities that confirm previous observations.[7,28] Though replication *in vitro* may differ biologically from what occurs *in vivo*, the results nevertheless demonstrate that the kinetics remains essentially the same. Of course, the specific finding of an increasing rate of initiation invites a biological interpretation involving a kind of autocatalysis, whereby the replication process itself leads to the release of a factor whose concentration determines the rate of initiation. This will be explored in future work.

Directions for future experiments in *X. laevis*

One effect that we did not include in our analysis is a variable fork velocity. For example, v might decrease as forks coalesce or as replication factor becomes limiting toward the end of S-phase.[5,22–24] Such effects, if present, are too small to see in the data analyzed here.

Another important question is to separate the effects of any intrinsic distribution due to early and late-replicating regions of the genome of a single cell from the extrinsic distribution caused by having many cells in the experiment. One approach would be to isolate and comb the DNA from a single cell. Although difficult, such an experiment is technically feasible. The latter problem could be resolved by *in situ* fluorescence observations of the chosen cell.

Applications to other systems

One can entertain many further applications of the basic model discussed above, which can be generalized, if need be. For example, Blumenthal *et al.* interpreted their results on replication in *D. melanogaster* for $\rho_{i2i}(\ell, f)$ to imply periodically spaced origins in the genome[21] (see their Figure 7). It is difficult to judge whether their peaks are real or statistical happenstance, but if the conclusion is indeed that the origins in that system are arranged periodically, the kinetics model could be generalized in a straightforward way (introducing an $I(x, \tau)$ that was periodic in x).

Very recently, detailed data on the replication of budding yeast (*Saccharomyces cerevisiae*) have become available.[33] The data provide information on the locations of origins and the timings of their initiation during S-phase. These data support the view of origin initiation throughout S-phase. Unlike replication in *Xenopus* prior to the mid-blastula transition, origins in budding yeast are associated with highly conserved sequence elements (autonomous replication sequence elements, or ARSs). Raghuraman *et al.*[33] also give the first estimates of the distribution of fork velocities during replication. Although broad, the distribution is apparently stationary, and there is no correlation between velocities and the time in S-phase when the forks are initiated. The model developed here could be generalized in a straightforward way to the case of budding yeast. Knowing the sequence of the genome and hence the location of potential origins means that the initiation function would be an explicit function of position x along the genome, with peaks of varying heights at each potential origin. The advantage of the kind of modeling advanced here would be the opportunity to derive quantities such as the replication fraction as a function of time in S-phase. Raghuraman *et al.* fit their data for this "timing curve" to an arbitrarily chosen sigmoidal function (see their supplementary data, Section II-5). Such modeling will make it easier to find meaningful biological explanations of the programming of S-phase evolution.

The origin-spacing problem

One outstanding issue in DNA replication in eukaryotes is the observation that the replication origins cannot be too far apart, as this would prevent the genome from being replicated completely within the length of a single S-phase.[34] One solution that has been proposed is that there is an excess of pre-replication complexes (pre-RCs) of highly conserved proteins, which assemble at ORC-bound DNA sites before the cell enters S-phase (e.g. Lucas *et al.*,[14] and references therein). In this case, the position of each potential origin of replication (POR) can be distributed randomly, with a statistically insignificant probability of having large gaps between PORs. The problem with this solution is that the average POR spacing must be much smaller (less than 1–2 kb) than the reported values of *Xenopus* origin recognition complex (XORC) spacing of 7–16 kb.[6,35]

A second proposed solution to the origin-spacing problem is to invoke correlations in POR spacings. In other words, instead of assuming a purely random pre-RC distribution, one imposes constraints that force a partial periodicity on the POR spacing, so that most of the origins are spaced 5–15 kb apart (Blow *et al.*,[36] and references therein). This suppresses the formation of large gaps but

raises other issues. First, it requires an unknown mechanism to achieve this periodicity of POR spacing. Second, it assumes implicitly that most of the PORs fire during S-phase, to prevent the 30 kb gap that could arise from an origin's failure to initiate, which is not obvious at all. Third, if origins initiate throughout S-phase, then there needs to be some kind of correlation that forces the more widely spaced origin groups to initiate early enough in S-phase to complete replication in the required time.

Implicitly, our model adopts language consistent with the first solution, but it is straightforward to consider the correlations assumed in the second solution. The presence of significant correlations in PORs would not invalidate the results presented here, which pertain to mean quantities (e.g. Figure 4); however, it would change their interpretation and could change biological models that one might try to make to explain the observed kinetic parameters we extract using the KJMA model. We plan to investigate these questions, along with the effect of origin efficiency on DNA replication kinetics, in future work.

Conclusion

Here, we have introduced a class of theoretical models for describing replication kinetics that is inspired by well-known models of crystal growth kinetics. The model allows us to extract the rate of initiation of new origins, a quantity whose time dependence has not previously been measured. With remarkably few parameters, the model fits quantitatively the most detailed existing experiment on replication in *Xenopus*. It reproduces known results (for example, the fork velocity) and provides the first reliable description of the temporal organization of replication initiation in a higher eukaryote. Perhaps most important, the model can be generalized in a straightforward way to describe replication and extract relevant parameters in essentially any organism.

Methods

Mathematical analogy between crystal growth and the kinetics of DNA replication

In this section, we describe how certain features of the mathematics describing crystal growth may be mapped onto a model describing the kinetics of DNA replication. We emphasize that the analogy is a formal one: the underlying processes are completely different. However, by mapping our problem onto one that has been long studied in a different context, we can take over a number of results that have already been derived, and we can develop useful intuitions about how to look at experimental results about DNA replication.

In the 1930s, several scientists independently derived a stochastic model that described the kinetics of crystal growth.[37–41] The "Kolmogorov–Johnson–Mehl–Avrami" (KJMA) model has since been widely used by metallurgists and other scientists to analyze thermodynamic phase transformations.[42]

In the KJMA model, freezing kinetics result from three simultaneous processes:

(1) nucleation, which leads to discrete solid domains;
(2) growth of the domain;
(3) coalescence, which occurs when two expanding domains merge.

Each of these processes has an analog in DNA replication in higher eukaryotes, and more specifically embryos:

(1) The activation of an origin of replication is analogous to the nucleation of the solid domains during crystal growth.
(2) Symmetric bidirectional DNA synthesis initiated (nucleated) at the origin corresponds to solid-domain growth.
(3) Coalescence in crystal growth is analogous to multiple dispersed sites of replicating DNA (replication fork) that advance from opposite directions until they merge.

Simple version of the KJMA model for DNA replication

In the simplest form of the KJMA model, solids nucleate anywhere in the liquid, with equal probability for all spatial locations ("homogeneous nucleation"), although it is straightforward to describe nucleation at pre-specified sites ("heterogeneous nucleation"), which would correspond to a case where replication origins are specified by fixed genetic sites along the genome. Once a solid domain has been nucleated, it grows out as a sphere at constant velocity v. When two solid domains impinge, growth ceases at the point of contact, while continuing elsewhere. KJMA used elementary methods to calculate quantities such as $f(\tau)$, the fraction of the volume that has crystallized by time (τ). Much later, more sophisticated methods were developed to describe the detailed statistics of domain sizes and spacings.[43,44]

DNA replication, of course, corresponds to one-dimensional crystal growth; the shape in three dimensions of the one-dimensional DNA strand does not directly affect the kinetics modeling. (In the model, replication is one-dimensional along the DNA. The configuration of DNA in three dimensions is not directly relevant to the model but can enter indirectly *via* the nucleation function $I(x, \tau)$. For example, if, for steric reasons, certain regions of the DNA are inaccessible to replication factories, those regions would have a lower (or even zero) value of I.) The one-dimensional version of the KJMA model assumes that domains grow out at velocity v, assumed to remain constant. The nucleation rate $I(x, \tau) = I_0$ is defined to be the probability of domain formation per unit length of unreplicated DNA per unit time, at the position x and time τ. Following the analogy to the one-dimensional KJMA model, we can calculate the kinetics of DNA replication during S-phase. This requires determining the fraction of the genome $f(\tau)$ that has already been replicated at any given moment during S-phase. One finds:

$$f(\tau) = 1 - e^{-I_0 v \tau^2} \quad (1)$$

which defines a sigmoidal curve. (Equation (1) assumes

an infinite genome length. The relative importance of the finite size of chromosomes is set by the ratio (fork velocity × duration of S-phase)/chromosome length.[45] In the case of the experiment analyzed here, this ratio is ≈ 10 bases/seconds × 1000 seconds/10^7 bases/chromosome $\approx 10^{-3}$, which we neglect.)

A more complete description of replication kinetics requires detailed analysis of different statistical quantities, including measurements made on replicated regions (eyes), unreplicated regions (holes), and eye-to-eye sizes (the eye-to-eye size is defined as the length between the center of one eye and the center of a neighboring eye). The probability distributions may be expressed as functions either of time τ or replicated fraction f. For example, the distribution of holes of size ℓ at time τ, $\rho_h(\ell, \tau)$ can be derived by a simple extension of the argument leading to equation (1):

$$\rho_h(\ell, \tau) = I_0 \tau \, e^{-I_0 \tau \ell} \tag{2}$$

From equation (2), the mean size of holes at time τ is

$$\ell_h(\tau) = \frac{1}{I_0 \tau} \tag{3}$$

Determining the probability distributions of replicated lengths (eye sizes) is complicated because a given replicated length may come from a single origin or it may result from the merger of two or more replicated regions. Thus, one must calculate in effect an infinite number of probabilities; by contrast, holes of a given length arise in only one way.[44] One can nonetheless derive a simple expression for $\ell_i(\tau)$, the mean replicated length at time τ, from a "mean-field" hypothesis:[46] the probability distribution of a given replicated length is assumed to be independent of the actual size of its neighbor. One can show that this mean-field hypothesis must always be true in one-dimensional growth problems, but not necessarily in the ordinary three-dimensional setting of crystal growth. In particular, if $I(\tau)$ depends on space, one expects correlations to be important. Using the mean-field hypothesis, we find:

$$\ell_i(\tau) = \ell_h(\tau) \frac{f}{1-f} = \frac{e^{I_0 v \tau^2} - 1}{I_0 \tau} \tag{4}$$

and

$$\ell_{i2i}(\tau) = \ell_i(\tau) + \ell_h(\tau) = \frac{\ell_h(\tau)}{1-f} = \frac{e^{I_0 v \tau^2}}{I_0 \tau} \tag{5}$$

The minimum average eye-to-eye size, obtained by differentiating equation (5), is $\ell^*_{i2i} = \sqrt{2e}\sqrt{v/I_0}$. These expressions for $\ell_i(\tau)$ and $\ell_{i2i}(\tau)$ allow one to collapse the experimental observations of ℓ_h, ℓ_i, and ℓ_{i2i} (the mean eye-to-eye separation) onto a single curve (see Figure 4(d)).

Finally, we can calculate the average distance between origins of replication that were activated at different times during the replication process, which is just the inverse of I_{tot}, the time-integrated nucleation probability per unit length:

$$\ell_0 \equiv I_{tot}^{-1} = \frac{2}{\sqrt{\pi}} \sqrt{\frac{v}{I_0}} \tag{6}$$

The last expression shows that, as might have been guessed by dimensional analysis of the model parameters (I_0 and v), the basic length scale in the model is set by $\ell^* \equiv \sqrt{v/I_0}$. Note that because initiation in the model is occurring throughout S-phase, the minimum eye-to-eye distance ℓ_{i2i_min} is not the same as the average separation between origins, ℓ_0. For this simple case, $\ell_{i2i\_min}/\ell_0 = \sqrt{e\pi/2} \approx 2.1$.

Generalizations of the KJMA model

On the basis of the specific results of the *Xenopus* experiments discussed above, we generalized the simple version of the KJMA model in several ways.

The first generalization is to allow for arbitrary $I(\tau)$. Equation (1) then becomes:

$$f(\tau) = 1 - e^{-g(\tau)} \qquad \text{with} \qquad g(\tau) = 2v \int_0^\tau I(\tau')(\tau - \tau') d\tau' \tag{7}$$

and, similarly, equation (3) becomes:

$$\ell_h(\tau) = \left[\int_0^\tau I(\tau') d\tau' \right]^{-1} \tag{8}$$

The other mean lengths, $\ell_i(\tau)$ and $\ell_{i2i}(\tau)$, continue to be related to $\ell_h(\tau)$ by the general expressions given in equations (4) and (5). In the experiment, one measures ℓ_h, ℓ_i, and ℓ_{i2i} as functions of both τ and f. (Because of the start-time ambiguity, the f data are easier to interpret.) The goal is to invert this data to find $I(\tau)$. Using equations (7) and (8), we find:

$$\tau(f) = \frac{1}{2v} \int_0^f \ell_{i2i}(f') df' = \frac{1}{2v} \int_0^f \frac{\ell_h(f')}{1-f'} df' \tag{9}$$

Because $\tau(f)$ increases monotonically, one can numerically invert it to find $f(\tau)$. From $f(\tau)$, one can derive all quantities of interest, including $I(\tau)$.

The starting time distribution $\phi(t)$ can be deduced looking at each molecular fragment, measuring its replication fraction f, and extrapolating back to a starting time using the experimentally determined $f(\tau)$ curve. (Fragments that are fully replicated ($f = 1$) are excluded.) The starting times are then binned to give $\phi(t)$ directly.

Monte-Carlo simulations

We wrote a Monte-Carlo simulation using the programming language of Igor Pro (wave-metrics) to test various experimental effects that were difficult to model analytically. These included the effects of finite sampling of DNA fragments (on average, 190 molecules per time point), the finite optical resolution of the scanned images, and, most important, the effect of the finite size of the combed DNA fragments. The size of each molecular fragment in the simulation was drawn randomly from an estimate of the actual size distribution of the experimental data. This distribution was approximately log-normal, with an average length of 102 kb and a standard deviation of 75 kb.

In the simulations, we consider each DNA molecular fragment as a one-dimensional lattice, and each lattice site is updated with a time-step $\Delta t = 0.2$ minute. An origin is initiated (lattice site changed from 0 to 1) with a probability determined by the initiation rate $I(\tau)$. Once an origin has been initiated, replication forks grow bidirectionally at a constant rate v. The natural size of lattice then would be $v\Delta t$, which is 123 bp for the measured fork velocity $v = 615$ bp/minute and chosen time-step Δt. The lattice scale is then roughly the size of origin recognition complex proteins. We sample the simulation results at the same time points as the actual experiments ($t = 25$, 29, 32, 35, 39, 45 minutes). Each sampled molecule is cut at a random site to simulate the

combing process. The lattice is then "coarse grained" by averaging over approximately four pixels. The coarse lattice length scale is then 0.24 μm, which roughly corresponds to that of the scanned optical images. Finally, the coarse-grained fragments were analyzed to compile statistics concerning replicon sizes, eye-to-eye sizes, etc., that were directly compared to experimental data.

In a first version of the simulation, the lattice was directly simulated using a vector with one element for each lattice site. In a more refined version of the simulation, we noted only the position of the replication forks, which greatly increased the speed of the simulations.

We also used the simulation to test a previous algorithm for extracting $I(f)$, the initiation rate as a function of overall replication fraction. The previous algorithm[13,47] looked for small replicated regions and extrapolated back to an assumed initiation time. We tested this algorithm using our Monte-Carlo analysis and found significant bias in the inferred $I(f)$, while the algorithms we introduce here showed no such bias.

Parameter extraction from data

We extracted data from both the real experiments and the Monte-Carlo simulations by a global least-squares fit that took into account simultaneously the different data collected (i.e. the different curves in Figures 3 and 4). As discussed above, we fit a two-segment straight line to the $I(\tau)$ curve extracted directly from the data for analytic simplicity. Assuming this form for $I(\tau)$, we derive explicit formulae for the curves in Figures 3 and 4.

The finite size of the molecular fragments studied ($102(\pm 75)$ kb) causes systematic deviation from the "infinite-length" formulae we can calculate. Such deviations could be detected using the Monte-Carlo simulations by comparing the extracted values of parameters with those input. The deviations show themselves mainly in two settings: first, whenever the mean length of holes, eyes, or eye-to-eye distances approaches the mean segment length, the observed mean lengths will be systematically too small because the larger end of the experimental distributions is cut off by the finite fragment length. We dealt with this complication by restricting our fit to areas where the mean length being measured is less than 10% of the mean fragment size. The second complication is that the inferred fork velocity is systematically reduced (by about 5% for the fragment size in the experiments analyzed here). We measured this bias using the Monte-Carlo simulations and then corrected the "raw" fork velocity that is given by our least-squares fits.

One further subtle point in a global fit is the relative weighting to be given to the data in the $\rho(f)$ curves (Figure 3) relative to the data in the mean-value curves (Figure 4). We estimated the weights using the bootstrap method.[48] In a similar spirit, we used repeated Monte-Carlo simulations to estimate statistical errors in experimentally extracted quantities.

Acknowledgments

We thank M. Wortis and B. -Y. Ha for helpful comments and insights. This work was supported by grants from the Fondation de France, NSERC, and NIH.

References

1. Hand, R. (1978). Eukaryotic DNA: organization of the genome for replication. *Cell*, **15**, 317–325.
2. Friedman, K. L., Brewer, B. J. & Fangman, W. L. (1997). Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells*, **2**, 667–678.
3. Cairns, J. (1963). The chromosome of *E. coli*. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 43–46.
4. Shivashankar, G. V., Feingold, M., Krichevsky, O. & Libchaber, A. (1999). RecA polymerization on double-stranded DNA by using single-molecule manipulation: the role of ATP hydrolysis. *Proc. Natl Acad. Sci. USA*, **96**, 7916–7921.
5. Pierron, G. & Benard, M. (1996). DNA replication in physarum. In *DNA Replication in Eukaryotic Cells* (DePamphilis, M., ed.), pp. 933–946, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
6. Walter, J. & Newport, J. W. (1997). Regulation of replicon size in *Xenopus* egg extracts. *Science*, **275**, 993–995.
7. Hyrien, O. & Mechali, M. (1993). Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of *Xenopus* early embryos. *EMBO J.* **12**, 4511–4520.
8. Coverley, D. & Laskey, R. A. (1994). Regulation of eukaryotic DNA replication. *Annu. Rev. Biochem.* **63**, 745–776.
9. Blow, J. J. & Chong, J. P. (1996). DNA replication in *Xenopus*. In *DNA Replication in Eukaryotic Cells*, pp. 971–982, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. Shinomiya, T. & Ina, S. (1991). Analysis of chromosomal replicons in early embryos of *Drosophila melanogaster* by two-dimensional gel electrophoresis. *Nucl. Acids Res.* **19**, 3935–3941.
11. Brewer, B. J. & Fangman, W. L. (1993). Initiation at closely spaced replication origins in a yeast chromosome. *Science*, **262**, 1728–1731.
12. Gomez, M. & Antequera, F. (1999). Organization of DNA replication origins in the fission yeast genome. *EMBO J.* **18**, 5683–5690.
13. Herrick, J., Stanislawski, P., Hyrien, O. & Bensimon, A. (2000). Replication fork density increases during DNA synthesis in *X. laevis* egg extracts. *J. Mol. Biol.* **300**, 1133–1142.
14. Lucas, I., Chevrier-Miller, M., Sogo, J. M. & Hyrien, O. (2000). Mechanisms ensuring rapid and complete DNA replication despite random initiation in *Xenopus* early embryos. *J. Mol. Biol.* **296**, 769–786.
15. Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F. & Bensimon, D. (1994). Alignment and sensitive detection of DNA by a moving interface. *Science*, **265**, 2096–2098.
16. Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Shurra, C., Hornigold, N. *et al.* (1997). Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science*, **277**, 1518–1523.
17. Herrick, J., Michalet, X., Conti, C., Shurra, C. & Bensimon, A. (2000). Quantifying single gene copy number by measuring fluorescent probe lengths on combed genomic DNA. *Proc. Natl Acad. Sci. USA*, **97**, 222–227.
18. Huberman, J. A. & Riggs, A. D. (1966). Autoradiography of chromosomal DNA fibers from Chinese hamster cells. *Proc. Natl Acad. Sci. USA*, **55**, 599–606.

19. Jasny, B. R. & Tamm, I. (1979). Temporal organization of replication in DNA fibers of mammalian cells. *J. Cell Biol.* **81**, 692–697.
20. Jackson, D. A. & Pombo, A. (1998). Replication clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140**, 1285–1295.
21. Blumenthal, A. B., Kriegstein, H. J. & Hogness, D. S. (1974). The units of DNA replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* **38**, 205–223.
22. Blow, J. J. & Laskey, R. A. (1986). Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of *Xenopus* eggs. *Cell*, **47**, 577–587.
23. Blow, J. J. & Watson, J. V. (1987). Nuclei act as independent and integrated units of replication in a *Xenopus* cell-free DNA replication system. *EMBO J.* **6**, 1997–2002.
24. Wu, J. R., Yu, G. & Gilbert, D. M. (1997). Origin-specific initiation of mammalian nuclear DNA replication in a *Xenopus* cell-free system. *Methods*, **13**, 313–324.
25. Simon, I., Tenzen, T., Reubinoff, B. E., Hillman, D., McCarrey, J. R. & Cedar, H. (1999). Asynchronous replication of imprinted genes is established in the gametes and maintained during development. *Nature*, **401**, 929–932.
26. Pasero, P. & Schwob, E. (2000). Think global, act local—how to regulate S phase from individual replication origins. *Curr. Opin. Genet. Dev.* **10**, 178–186.
27. Mills, A. D., Blow, J. J., White, J. G., Amos, W. B., Wilcock, D. & Laskey, R. A. (1989). Replication occurs at discrete foci spaced throughout nuclei replicating *in vitro*. *J. Cell Sci.* **94**, 471–477.
28. Mahbubani, H. M., Paull, T., Elder, J. K. & Blow, J. J. (1992). DNA replication initiates at multiple sites on plasmid DNA in *Xenopus* egg extracts. *Nucl. Acids Res.* **20**, 1457–1462.
29. Lu, Z. H., Sittman, D. B., Romanowski, P. & Leno, G. H. (1998). Histone H1 reduces the frequency of initiation in *Xenopus* egg extract by limiting the assembly of prereplication complexes on sperm chromatin. *Mol. Biol. Cell*, **9**, 1163–1176.
30. Buongiorno-Nardelli, M., Michelli, G., Carri, M. T. & Marilley, M. (1982). A relationship between replicon size and supercoiled loop domains in the eukaryotic genome. *Nature*, **298**, 100–102.
31. Laskey, R. A. (1985). Chromosome replication in early development of *Xenopus laevis*. *J. Embryol. Expt. Morphol.* **89**, 285–296.
32. Berezney, R., Dubey, D. D. & Huberman, J. A. (2000). Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma*, **108**, 471–484.
33. Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodlicka, L., Conway, A. *et al.* (2001). Replication dynamics of the yeast genome. *Science*, **294**, 115–121.
34. Gilbert, D. M. (2001). Making sense of eukaryotic DNA replication origins. *Science*, **294**, 96–100.
35. Rowles, A., Chong, J. P., Brown, L., Howell, M., Evan, G. I. & Blow, J. J. (1996). Interaction between the origin recognition complex and the replication licensing system in *Xenopus*. *Cell*, **87**, 287–296.
36. Blow, J. J., Gillespie, P. J., Francis, D. & Jackson, D. A. (2001). Replication origins in *Xenopus* egg extract are 515 kilobases apart and are activated in clusters that fire at different times. *J. Cell Biol.* **152**, 15–25.
37. Kolmogorov, A. N. (1937). On the statistical theory of crystallization in metals. *Izv. Akad. Nauk SSSR, Ser. Fiz.* **1**, 355–359.
38. Johnson, W. A. & Mehl, P. A. (1939). Reaction kinetics in processes of nucleation and growth. *Trans. AIMME*, 416–442. Discussion, pp. 442–458.
39. Avrami, M. (1939). Kinetics of phase change. I. General theory. *J. Chem. Phys.* **7**, 1103–1112.
40. Avrami, M. (1940). Kinetics of phase change. II. Transformation–time relations for random distribution of nuclei. *J. Chem. Phys.* **8**, 212–224.
41. Avrami, M. (1941). Kinetics of phase change. III. Granulation, phase change, and microstructure. *J. Chem. Phys.* **9**, 177–184.
42. Christian, J. W. (1981). *The Theory of Phase Transformations in Metals and Alloys, Part I: Equilibrium and General Kinetic Theory*, Pergamon Press, New York.
43. Sekimoto, K. (1991). Evolution of the domain structure during the nucleation-and-growth process with non-conserved order parameter. *Int. J. Mod. Phys. B*, **5**, 1843–1869.
44. Ben-Naim, E. & Krapivsky, P. L. (1996). Nucleation and growth in one dimension. *Phys. Rev. E*, **54**, 3562–3568.
45. Cahn, J. W. (1996). Johnson–Mehl–Avrami kinetics on a finite growing domain with time and position dependent nucleation and growth rates. *Mater. Res. Soc. Symp. Proc.* **398**, 425–438.
46. Plischke, M. & Bergersen, B. (1994). *Equilibrium Statistical Physics*, 2nd edit., Chapt. 3, World Scientific, Singapore.
47. Marheineke, K. & Hyrien, O. (2001). Aphidicolin triggers a block to replication origin firing in *Xenopus* egg extracts. *J. Biol. Chem.* **276**, 17092–17100.
48. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edit., Chapt. 15, Cambridge University Press, Cambridge.

Edited by M. Yaniv

(Received 14 December 2001; received in revised form 23 April 2002; accepted 13 May 2002)