Replication timing and its emergence from stochastic processes

John Bechhoefer[1] and Nicholas Rhind[2]

[1] Department of Physics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada
[2] Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

The temporal organization of DNA replication has puzzled cell biologists since before the mechanism of replication was understood. The realization that replication timing correlates with important features, such as transcription, chromatin structure and genome evolution, and is misregulated in cancer and aging has only deepened the fascination. Many ideas about replication timing have been proposed, but most have been short on mechanistic detail. However, recent work has begun to elucidate basic principles of replication timing. In particular, mathematical modeling of replication kinetics in several systems has shown that the reproducible replication timing patterns seen in population studies can be explained by stochastic origin firing at the single-cell level. This work suggests that replication timing need not be controlled by a hierarchical mechanism that imposes replication timing from a central regulator, but instead results from simple rules that affect individual origins.

Replication origins: correlated or independent?

The duplication of the genome of a cell by DNA replication is an essential step in the cell cycle. In bacteria, the overall situation is straightforward, in that DNA replication initiates at a single, well-defined location in the genome (e.g. the *oriC* site in *Escherichia coli*) and terminates at a second, well-defined region (*ter* in *E. coli*) [1]. Eukaryotic organisms, with 10–1000 times more DNA and with 10–100 times slower replication forks, depend on the firing of multiple origins of replication along the DNA. These origins are defined by a two-step process [2]. Licensing, the first step, occurs in G1 phase, when the origin recognition complex (ORC) binds to chromatin and, with the aid of Cdc6 and Cdt1, loads onto the DNA head-to-head pairs of the barrel-shaped heterohexameric MCM complex, the catalytic core of the replicative helicase [3,4]. Each pair of MCM complexes is a potential origin of DNA replication. Initiation (or origin firing), the second step, occurs in S phase, when a pair of MCMs is activated via a complex process involving numerous proteins, including recruitment of Sld2, Sld3, the GINS complex and Cdc45, as well as the phosphorylation of various components by the CDK and DDK replication kinases [5]. The regulation of the spatial binding of the ORC and the temporal activation

of MCMs largely determines the kinetics of replication during S phase, which is referred to as the replication program.

The question of how replication programs are regulated is an active, and sometimes controversial, field. Although the specific mechanisms that control timing are still obscure, recent work has revealed basic principles that appear to apply to eukaryotic replication in general. In particular, mathematical modeling of genome-wide replication timing data shows that replication timing can be explained by stochastic mechanisms. The significance of this conclusion is that it explains the regulation of replication timing in terms of simple rules that affect the individual probabilities of origin firing. In such models, replication timing is controlled by changing the firing rate of individual origins, instead of by directly regulating the time at which origins fire. Although this distinction may seem semantic, it is important because it recasts black-box mechanisms of global replication timing in terms of biochemically plausible effects on individual origins.

Over the past decade, two views about replication timing mechanisms have been developed. In the first, origin firing is a stochastic event that is (largely) independent of the replication state of neighboring origins. In particular, it has been postulated that there is an initiation function $I(x,t)$ that describes the rate of initiation, per time and per length of unreplicated DNA, of a site x along the genome at time t after the beginning of S phase [6,7] (Figure 1; Box 1). This type of origin firing can manifest in at least three different ways, depending on the experimental model considered. In species such as budding yeast, in which replication initiates at well-defined loci, the function $I(x,t)$ forms a discrete spike at the replication origin [8] (Figure 1a). At the other end of the spectrum, amphibian embryos lack origin specificity, and DNA replication can initiate anywhere along the genome [6]. In an intermediate case, mammalian somatic cells can display clusters of origins or broad initiation zones that are not homogeneously distributed throughout the genome [9–11] (Figure 1b). Each of these three cases is discussed in detail below. We refer to the hypothesis of a locally determined initiation rate as the independent origin hypothesis because it is distinguished by the feature that origins fire independently from the firing of neighboring origins. The attraction of the independent origin hypothesis is its simplicity: one does not need to postulate biological mechanisms that would cause correlated initiations. The potential weakness

---
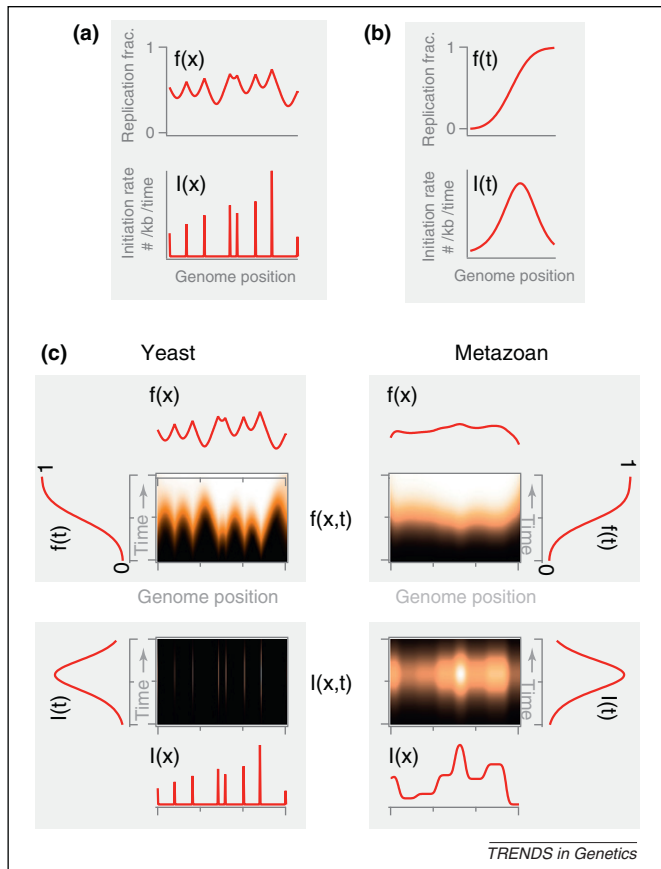
Figure 1. Replication fractions and initiation rates. **(a,b)** The relation between replication fractions *f* and initiation rates *I*, as illustrated for budding yeast. (a) Spatially resolved data, averaged over an asynchronous cell population. (b) Time course data, averaged over the genome. **(c)** Illustration of typical replication timing data for budding yeast (left) and a metazoan organism (right). Top-left image shows the replication fraction *f(x,t)*, as it might be inferred from a microarray timing experiment with several time points of data from synchronized cell populations. Black represents low-replication levels and white represents high-replication levels. Averaging the replication fraction over the genome gives the curve *f(t)*, depicted to the left of the *f(x,t)* image, which goes from 0 to 1. Averaging the replication fraction over time, as in an experiment on asynchronous cell populations, gives the curve above the *f(x,t)* image. The bottom-right group shows the inferred *I(x,t)* image, as well as the averaged curves *I(t)* and *I(x)*. Note that, in budding yeast, replication origins are well localized, as indicated by the spikes in the function *I(x)*. [When viewed or printed at low resolution, not all spikes in *I(x,t)* may be visible.] The right-hand groups illustrate similar concepts for a typical metazoan organism. The main difference is that origins are not well localized, so that the function *I(x)* has broad features, representing zones where initiations are more or less likely to occur.

of the independent origin hypothesis is that, if too simple, it may fail to describe experiments accurately or that implausible coincidences of parameters may be required to fit the data.

In a second scenario, the initiation of an origin, although still stochastic, is linked to the state of the genome in its vicinity. For example, observations of origin clustering [12–14] have led several authors to hypothesize that the presence of a replication fork can increase the firing rate of nearby origins, for instance, the 'next-in-line' model [15] and the 'domino-cascade' model [16,17]. We refer to this second scenario in general as the correlated origin hypothesis.

Previously, there was considerable debate as to whether replication was stochastic and whether origins are independent. At present, it is generally accepted that all models

of replication are stochastic at the level of molecular interactions. It is important to note that stochastic models do not require that origins all fire with the same probability, nor is stochastic firing incompatible with late firing origin [18]. However, there is evidence in some cases for correlation in origin initiation activity. As a result, the current picture is an intermediate one that mixes both stochastic elements and mechanisms for correlations in origin initiation [15,19]. Still, differences remain concerning what is essential and what is incidental in the above picture and what kind of underlying mechanisms are likely to be important in controlling the replication program. In this review, we argue that, for the simpler cases such as unicellular yeast and for the embryonic cells of some multicellular animals, recent experiments and modeling efforts have shown that much of the available replication data may be understood in terms of the simpler independent origin hypothesis and that correlations probably play a minor role in the replication program. Replication in the somatic cells of metazoan organisms is more complex, and we outline recent efforts in this area.

Replication in yeast
The past few years have marked a turning point in the understanding of replication in yeast. First came a series of high-resolution combing and microarray experiments (Box 2). For example, high-resolution timing data of synchronized populations of wild type and *clb5Δ Saccharomyces cerevisiae* show clear average timing patterns [20]. Their measurements, as mentioned in Box 2, amounted to measurements of $f(x,t)$, with spatial information resolved to a few kilobases and temporal information resolved to 5 min. At around the same time, DNA combing studies in budding and fission yeast showed that initiations at the single-molecule scale are stochastic, with different sets of origins chosen in each cell cycle [21,22]. Indeed, in budding yeast, it is now clear that there are as many as 700 potential origin sites, of which only approximately 200 are used in any given cycle.

In parallel work, the rate of origin firing in budding and fission yeast was shown to be regulated by competition for limiting activators, such as the Cdc45 initiation factor and the DDK initiation kinase [23–26]. Competition for limiting activators provides an explanation for why origin firing is less efficient than might be possible. The stochastic interaction between origins and diffusible activators also provides a mechanism for stochastic firing of origins.

The stochasticity of individual origins turns out to be an important effect. In contrast to earlier models, in which the firing of specific origins was envisaged to be limited to narrow windows of S phase, it is now clear that the width of the firing-time distribution for an individual origin can be a substantial fraction of S phase. Indeed, models that fail to incorporate the width of the timing distribution fail to reproduce many of the experimental details adequately [27]. By contrast, stochastic models that take into account the width of the firing-time distribution can successfully fit the microarray data [8,28,29]. Several notable insights and results come from these analyses: first, it is possible to generate models with independent initiation scenarios [initiation rate $I(x,t)$ and constant fork velocity v] that lead to good fits of the data. This result shows that the independent origin hypothesis suffices to explain microarray data

Box 1. *f* and *I*: mathematical functions that describe replication kinetics

DNA replication kinetics can be described using two related but distinct mathematical functions: the replication fraction *f* and the initiation rate *I*. The first, *f*, is a complete description of replication kinetics and can be directly determined from experimental data (Box 2). The second, *I*, only describes the kinetics of origin initiation and cannot be directly measured; it must be inferred from *f*. However, if fork rates are assumed to be nearly constant, as is frequently done in models of replication kinetics, then *I* is sufficient to completely determine *f*. Both *f* and *I* can be defined for every spatial point (*x*) in the genome and every time point (*t*) in S phase, to give *f*(*x,t*) and *I*(*x,t*) (Figure 1c, main text).

It is often useful to consider the time-averaged functions, *f*(*x*) and *I*(*x*) (Figure 1a, main text). *f*(*x*) can be thought of as the average replication time of each point in the genome and is generally measured on asynchronous populations of cells. It is closely related to the median replication time t_{rep} at a site that is inferred from time course data on synchronized cell populations. The peaks in *f*(*x*) represent the origins, and taller peaks indicate origins that fire, on average, earlier in S phase. *I*(*x*) represents the average initiation rate of each point in the genome. In yeast, where origins are well defined, *I*(*x*) = 0 for most of the genome and forms spikes over the origins, with taller spikes reflecting a higher average probability of origin firing (Figure 1a,c). In metazoans, origins appear to be more diffuse, and thus so is *I*(*x*) (Figure 1c). It is important to realize that the height of the peaks in *I*(*x*) (e.g. the average firing probability of an origin) cannot be directly inferred from the height of the peaks in *f*(*x*), because *f*(*x*) convolves both passive replication and active firing of each origin; *I*(*x*) can only be extracted by mathematical modeling of *f*(*x*).

It can also be useful to consider the spatially averaged functions, *f*(*t*) and *I*(*t*) (Figure 1b, main text). The replication fraction *f*(*t*) is generally sigmoidal, as cells go from unreplicated in G1 to replicated in G2. The exact shape of the sigmoid depends on the details of the replication program, such as of the distribution of origins and the shape of *I*(*t*). As discussed in the main text, *I*(*t*) has been proposed to generally increase for most of S phase and then decline in late S phase.

on replication timing in yeast. Second, the intrinsic parameters characterizing each origin have values that are independent of their neighbors, again suggesting that the initiation of each origin is an independent stochastic event [8]. Studies in fission yeast have also led to the conclusion that local initiation models suffice to explain the available experimental data [30,31]. However, several biologically different scenarios can lead to similar overall timing patterns [32], and more complicated mechanisms, such as *trans*-acting regulators of origin activity and chromosome structure, can affect origin timing [33,34]. Clearly, further iterations of modeling and experiment will be needed to come to a final picture.

Replication in embryos

Embryonic cells in metazoans represent an interesting intermediate case of complexity. On the one hand, they have the full amount of DNA of somatic cells. On the other hand, they undergo a rapid, simplified cell cycle that is largely transcriptionally silent, which removes one major source of complication in the replication of somatic cells. *In vitro* studies of *Xenopus* cell-free extracts have been especially detailed and fruitful [13,35–37] and have led to associated modeling efforts [6,7,19,38,39]. The replication program in *Xenopus* embryos is relatively simple and much faster than in somatic cells. In particular, there are no fixed

origin sites, presumably because the lack of transcription and more uniform chromatin structure allows the ORC to load MCM anywhere in the genome [40,41]. Although variation in initiation rates and, hence, replication timing does occur at the megabase scale [37], modeling efforts to date have focused on understanding the temporal variation of the initiation rate, *I*(*t*), averaged over the genome. The main conclusion is that the initiation rate increases over most of S phase, before decreasing to zero near the end of it. This variation of initiation rates over S phase is significant because it leads to a relatively narrow distribution of lengths for S phase, which, because of the stochasticity of origin placement and initiation time, varies with each cell cycle [42]. In embryos, it is particularly important that there be little variation in genome duplication time, as the cell cycle lacks checkpoints that can delay the start of mitosis if replication is not complete. In *Xenopus* embryos, for example, the typical S phase duration is 20 min and that of mitosis is 5 min, all within a 25-min cell cycle [43]. Thus, variations of S phase of more than 5 min can be lethal for a cell. Such variations are proposed to be suppressed by the increasing nature of initiation rate *I*(*t*). It has even been postulated that the increasing form of *I*(*t*) is a universal characteristic of eukaryotic replication [44]. Preliminary assessment of replication data from *S. cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens* supports this scenario, although better data and more extensive analysis are required. The initial increase of initiation rate *I*(*t*) has been attributed to competition for a limiting factor required for replication fork function [35,38] or origin firing (e.g. the DDK replication kinase [23]), whereas the decrease of *I*(*t*) at the end of S phase has been variously attributed to a fork-dependent control mechanism [38] or to increasing diffusion search times for the limiting factor to find its target [39].

Replication in metazoan somatic cells

The replication of DNA in metazoan germline and somatic cells is more complicated than in embryonic cells. Replication in somatic cells can take up to 100 times longer than in embryonic cells [45], and this increase in replication time is not spread equally across the genome. Instead, different regions of the genome replicate at characteristic times during the elongated S phase, and the replication timing of a locus correlates with several other important chromosomal characteristics. The best-established correlation is between late replication and constitutive heterochromatin, the repetitive, transcriptionally inactive regions of the genome that remain condensed throughout the cell cycle [46]. Conversely, gene-rich, transcriptionally active regions of the genome tend to replicate earlier in S phase [47]. The correlation between the transcriptional activity of individual genes and their replication timing is not strong [48]. However, when averaged over large groups of neighboring genes, transcriptional activity correlates well with replication timing [49,50]. An even more remarkable correlation is seen between chromosome interaction maps and replication timing [51,52]. The contiguous regions of the genome that replicate with similar timing are referred to as replication domains. The correlations between the average transcriptional activity, chromatin interactions and

Box 2. Experimental techniques for analyzing DNA replication timing

The recent gains in our understanding of replication timing are built on experimental advances that have greatly increased the quality and quantity of data available. Defined patterns of DNA replication were first observed in fiber autoradiography studies of tritiated thymidine incorporation in bacterial and mammalian cells [12,79]. By *in vivo* pulse-labeling cells with tritiated thymidine and then stretching the labeled DNA on a photosensitive film, it was possible to map replication patterns (which regions have replicated and which have not) at a given time. A significant technical improvement was the substitution of fluorescently labeled thymine analogs, such as BrdU, that could be observed using an optical microscope [80,81]. Molecular combing, which stretches DNA more controllably, improved the latter technique by allowing one to more reliably associate positions on an image of a stretched fiber with genomic positions and by simplifying the identification of individual fibers taken anonymously from the genome [82,83] or with the genome location identified [54,84]. In parallel with fiber-based techniques, live-cell imaging has also yielded much valuable information. Although the size of origins and even their separations are well below the resolution of conventional light microscopy, clever techniques can yield spatial and temporal information. For example, specific sites can be labeled with fusion proteins whose intensity doubles after replication, an event that can readily be observed [85]. In the future, 'live' single-molecule studies based on flow and optical or magnetic tweezers [86], nano-engineered capillaries [87,88] and other molecular-scale structures may lead to even greater insights, especially into local mechanisms at the fork and initiation sites.

A second set of techniques provides information about the fraction of cells in a population that has replicated at a particular location x and time t. This fraction of replicated cells can be described by the function $f(x,t)$, if replication kinetics throughout S phase are measured, or simply as $f(x)$, if measurements are performed on asynchronous cell populations (Figure 1, main text). Such measurements originally used microarrays [89,90], with one approach based on local changes in copy number during replication. In a population of unreplicated cells, a baseline intensity is measured at each locus [$f(x) = 0$]. After all cells have replicated, the measured intensity at each locus should be double [$f(x) = 1$]. During replication, intermediate levels of replication are detected as intermediate intensity levels [$0 < f(x) < 1$]. For example, if half of the cells in the population have replicated at a location x, then $f(x) = 0.5$. More recently, direct sequencing to determine local DNA copy number has given similar information with fewer artifacts [91,92]. Initial studies used multiple time points in cultures of synchronized cells to directly measure $f(x,t)$ [89,93], and this approach is still the state of the art in yeast [20,64]. However, comparable results can be derived by sorting asynchronous cells of any type into G1 and S populations [90].

the replication timing of replication domains has led to qualitative models in which the chromosome accessibility of a domain affects its replication timing [53].

Although replication domains replicate with reproducible timing, origin firing within domains is heterogeneous because of stochastic origin firing [10,54,55]. As in yeast, origin firing in metazoans appears to be regulated by limiting activators. Mammalian Cdc45 is substoichiometric, relative to OCR and MCM, and increasing Cdc45 levels increases the rate of origin firing [56]. Moreover, modulating the levels of the CDK replication kinase affects the efficiency of origin firing [57–59]. An additional reason for the heterogeneity of origin firing in metazoans is that metazoan origins are not well-defined loci; at least in some cases, MCM seems to be loaded heterogeneously throughout a region [60–62], which can be thought of as a

cluster of many inefficient origins or as a diffuse initiation zone.

Mechanisms for timing

Although replication timing appears to be uniform and well coordinated at the population level, this average behavior hides heterogeneous replication kinetics in individual cells. This apparent conflict between heterogeneity at the single-cell level and organization at the population level is resolved by observing that the average of the heterogeneous single-cell data recapitulates the results from ensemble studies [22]. This observation has led to models in which the average replication time of a locus is a function of the firing probability of individual origins, regardless of whether those probabilities are independent or coordinated (Box 3). Such models predict a correlation between the probability and timing of origin firing, a correlation seen in budding yeast [8,28]. Furthermore, recent budding yeast studies have shown that, in most cases in which the length of S phase is significantly increased, the relative timing program is maintained [63,64]; that is, the overall ordering of replication timing of different regions is preserved, even as the scale of timing is altered. Such a result would be expected if S phase length changes because the initiation rates have been altered globally (Naama Barkai, personal communication). As discussed above, initiation rates are thought to be regulated by competition among origins for limiting activation factors. One recently proposed model makes the case both theoretically and experimentally that the limiting factor is a protein associated with active replication forks [65]. The Cdc45 protein, which is required to activate the MCM helicase complex, is one such candidate [56]. Alternatively, factors such as DDK, which phosphorylates and activates MCM, have been seen to be rate limiting in fission yeast [23].

The competition for limiting activators explains why origins fire stochastically but not why some origins fire with higher probability than others. One obvious explanation for differing probabilities of origin firing is the effect of chromatin structure on the accessibility of origins to initiation factors [53]. In the context of competition between origins for limiting activators, it is natural to imagine that chromatin structure affects that competition, allowing euchromatic origins greater access to activators and so higher firing probabilities. This possibility fits well with the strong correlation observed between heterochromatin and late replication [46]. Another possibility that we have recently proposed is based on the observation that multiple MCMs are loaded at each origin [8,60]. In this model, each MCM loaded has a low probability of firing; however, because multiple MCMs are loaded at each origin, origins that have more MCMs loaded will have a higher aggregate firing probability. Thus, the probability of origin firing is set in part by the number of MCMs loaded at a given origin site. The probability of origin firing can then be subsequently altered by chromatin context. For example, a recent study has shown that Rif1, which affects telomere chromatin structure, also binds to chromosome arms and alters origin initiation rates at these sites, perhaps by altering the loading of the Cdc45 that is required for MCM helicase activation [33].

A scenario comprising stochastically firing origins with different firing probabilities naturally leads to a reproducible replication-timing program [66]. Origins with high firing probabilities will be more likely to fire in early S phase and so will have early average replication times. In general, low-probability origins would be unlikely to fire efficiently even in late S phase. However, if the firing rate, $I(t)$, increases during S phase, as described above, even low-probability origins, if not passively replicated, will have a high probability of firing late in S phase, leading to efficient replication of late-replicating regions [18]. Here, we distinguish between $I(t)$, which describes the timing program, and the underlying biological mechanisms, which try to explain why $I(t)$ has an observed form. This description of origin timing applies not only to the individual origins of simpler genomes, such as budding yeast, but also to the complicated replication domains of metazoan genomes. In the latter case, euchromatic replication domains of high-probability origins reproducibly replicate earlier than do domains of lower-probability origins, but heterochromatic domains, which harbor the lowest-probability origins, nonetheless replicate efficiently in late S phase. Thus, the order in which various domains of metazoan genomes replicate may be a secondary consequence of the effect of

their chromatin structure on the firing probabilities of their origins. This possibility is consistent with the strong correlation between chromatin interactions and replication timing [52].

Correlated origin initiations

Although much of observed replication timing can be explained in terms of a picture of independent initiations, there is also evidence for correlations in initiation. For example, DNA fiber studies observe clusters of nearby origins that initiated at approximately the same time [12,13]. One plausible mechanism for origin clustering is that the polymerases and other proteins responsible for replication are localized within the nucleus in small foci known as replication factories [67,68]. As a consequence, if the DNA is tethered to a location in the cell nucleus while replicating, it may loop around and find another set of replication machinery in the same factory. Such looping could increase the likelihood of origin firing of origins located approximately 10 kb from an active fork and decrease origin firing for closer origins [19].

Another line of argument suggesting the possibility of correlated initiation lies in an observation of small biases in the DNA base sequence near certain regions. It has been shown that if a region of the genome is repeatedly replicated by a polymerase on the leading strand, mutations will eventually lead to strand compositional asymmetries (an excess of G over C and T over A) [69]. Indeed, a large proportion of known origins for *H. sapiens* have been found by looking for signatures of compositional skew [70]. Early replicating regions are then marked by an abrupt jump in the local skew. Because adjacent early replicating regions are separated by approximately 1 Mb and because the average distance between origins is approximately 100 kb, there must be multiple initiations between each early region. To explain the observation that the compositional skew varies linearly between compositional discontinuities associated with origins, it was postulated that a wave of correlated initiations occurs, which leads to a 'domino' [16,17,71] or 'next-in-line' model [15]. It is not clear whether a looping mechanism [19] can explain such effects, whether some more complicated form of coupling between initiation and fork progression is required, or whether the difference in chromatin structure between early- and late-replicating regions can account for these observations. Such a possibility would avoid the need to invoke coordinated origin firing. In support of this idea, a recent single-molecule replication kinetics analysis of the mouse *Igh* locus is consistent with a stochastic model that lacks any origin coordination [11] (Paolo Norio, personal communication).

In addition to temporal ordering of origin initiation, some models include spatial correlations in the positioning of origins. Recently, it was proposed that the clustering of initiated origins observed in *Xenopus* embryos and, to a lesser extent in yeast origins, may speed up the overall completion of S phase [72]. A shorter S phase is particularly helpful in *Xenopus* embryos, as it prevents the mitotic catastrophe discussed above. Clustering several inefficient origins together can lead to a group that is collectively efficient in that one or the other of the origins is likely to

fire early. Although the periodic distribution of such groups of origins would be an efficient way to replicate the genome, mechanisms that could achieve this global order are not clear, at present.

Concluding remarks

The hypothesis that replication is largely controlled by the local rate of initiation has received wide support from recent experiments and analyses. Models based on local replication rates $I(x,t)$ have successfully described the replication process in budding and fission yeast, in *Xeno-pus* embryos and in the *Igh* locus of mouse pro-B cells [6,8,11,28,30,38] (Paolo Norio, personal communication). A limiting factor in this work is that each of the above analyses involved a long-term collaboration between experimental biologists and modeling laboratories (the latter from a variety of fields, including physics, engineering and computer science). To broaden the use of quantitative analyses of replication and to analyze the growing number of data sets, it is important that the software and analysis procedures be usable by non-specialists. The recent derivation of 'inversion' formulas (A. Baker, PhD thesis, ENS de Lyon, France, 2011) that give $I(x,t)$ directly from data on the local average replication fraction $f(x,t)$ obtainable from microarray or deep sequencing studies on synchronized cell populations are a first step in that direction.

A second research direction is a more precise understanding of the relation between the replication program, as described above, and the effects of DNA damage, with its concomitant activation of DNA repair mechanisms. For example, one consequence of damage that stalls replication forks is the activation of additional origins, which now have more time to initiate [73,74], an effect that is straightforward to simulate [75] and model analytically [76]. The modeling of fork stalls predicts that there is a critical density of stalled forks (approximately one per replicon), above which there is a global delay in S phase and below which the effects are minor and localized. Interestingly, this threshold density matches the observed stall densities in fragile zones and in cells with activated oncogenes [76]. However, DNA damage can also induce checkpoints that inhibit subsequent origin firing [77], complicating the overall effect of DNA damage on replication timing. A related topic is the interrelation between mutation rates and events in S phase. Although formal models to handle such situations are beginning to be developed [69], more work is needed to understand observations, such as the link between mutation rate and S phase timing [78].

Although the independent origin hypothesis is attractive in its simplicity and so far remarkably successful in its application, there is evidence for correlated initiations in somatic metazoan cells. Some of the correlation is explainable as straightforward consequences of the physical constraints of clustering polymerases. In such a view, the primary method of controlling timing in S phase remains the local modulation of overall initiation rates, and the correlations in the initiation of neighboring origins are produced by the geometrical effects of loops induced by replication factories. Whether such mechanisms suffice or whether a more complicated control mechanism is at play is at present unclear. Time will tell.

References

1 Baker, T.A. and Wickner, S.H. (1992) Genetics and enzymology of DNA replication in *Escherichia coli*. *Annu. Rev. Genet.* 26, 447–477
2 Masai, H. *et al.* (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu. Rev. Biochem.* 79, 89–130
3 Remus, D. *et al.* (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139, 719–730
4 Evrin, C. *et al.* (2009) A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20240–20245
5 Labib, K. (2010) How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev.* 24, 1208–1219
6 Herrick, J. *et al.* (2002) Kinetic model of DNA replication in eukaryotic organisms. *J. Mol. Biol.* 320, 741–750
7 Jun, S. and Bechhoefer, J. (2005) Nucleation and growth in one dimension. II. Application to DNA replication kinetics. *Phys. Rev. E* 71, 011909
8 Yang, S.C. *et al.* (2010) Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol. Syst. Biol.* 6, 404
9 Hamlin, J.L. *et al.* (2008) A revisionist replicon model for higher eukaryotic genomes. *J. Cell. Biochem.* 105, 321–329
10 Norio, P. *et al.* (2005) Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol. Cell* 20, 575–587
11 Gauthier, M.G. *et al.* (2012) Modeling inhomogeneous DNA replication kinetics. *PLoS ONE* 7, e32053
12 Huberman, J.A. and Riggs, A.D. (1968) On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* 32, 327–341
13 Blow, J.J. *et al.* (2001) Replication origins in *Xenopus* egg extract are 5–15 kilobases apart and are activated in clusters that fire at different times. *J. Cell Biol.* 152, 15–25
14 Pasero, P. *et al.* (2002) Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. *Genes Dev.* 16, 2479–2484
15 Shaw, A. *et al.* (2010) S-phase progression in mammalian cells: modelling the influence of nuclear organization. *Chromosome Res.* 18, 163–178
16 Audit, B. *et al.* (2009) Open chromatin encoded in DNA sequence is the signature of 'master' replication origins in human cells. *Nucleic Acids Res.* 37, 6064–6075
17 Guilbaud, G. *et al.* (2011) Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput. Biol.* 7, e1002322
18 Rhind, N. *et al.* (2010) Reconciling stochastic origin firing with defined replication timing. *Chromosome Res.* 18, 35–43
19 Jun, S. *et al.* (2004) Persistence length of chromatin determines origin spacing in *Xenopus* early-embryo DNA replication: quantitative comparisons between theory and experiment. *Cell Cycle* 3, 223–229
20 McCune, H.J. *et al.* (2008) The temporal program of chromosome replication: genomewide replication in *clb5Δ Saccharomyces cerevisiae*. *Genetics* 180, 1833–1847
21 Patel, P.K. *et al.* (2006) DNA replication origins fire stochastically in fission yeast. *Mol. Biol. Cell* 17, 308–316
22 Czajkowsky, D.M. *et al.* (2008) DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *J. Mol. Biol.* 375, 12–19
23 Patel, P.K. *et al.* (2008) The Hsk1(Cdc7) replication kinase regulates origin efficiency. *Mol. Biol. Cell* 19, 5550–5558
24 Mantiero, D. *et al.* (2011) Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J.* 30, 4805–4814
25 Wu, P.Y. and Nurse, P. (2009) Establishing the program of origin firing during S phase in fission yeast. *Cell* 136, 852–864
26 Tanaka, S. *et al.* (2011) Origin association of sld3, sld7, and cdc45 proteins is a key step for determination of origin-firing timing. *Curr. Biol.* 21, 2055–2063

27 Spiesser, T.W. *et al.* (2009) A model for the spatiotemporal organization of DNA replication in *Saccharomyces cerevisiae*. *Mol. Genet. Genomics* 282, 25–35

28 de Moura, A.P. *et al.* (2010) Mathematical modelling of whole chromosome replication. *Nucleic Acids Res.* 38, 5623–5633

29 Luo, H. *et al.* (2010) Genome-wide estimation of firing efficiencies of origins of DNA replication from time-course copy number variation data. *BMC Bioinform.* 11, 247

30 Lygeros, J. *et al.* (2008) Stochastic hybrid modeling of DNA replication across a complete genome. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12295–12300

31 Koutroumpas, K. and Lygeros, J. (2011) Modeling and analysis of DNA replication. *Automatica* 47, 1156–1164

32 Raghuraman, M.K. and Brewer, B.J. (2010) Molecular analysis of the replication program in unicellular model organisms. *Chromosome Res.* 18, 19–34

33 Hayano, M. *et al.* (2011) Mrc1 marks early-firing origins and coordinates timing and efficiency of initiation in fission yeast. *Mol. Cell. Biol.* 31, 2380–2391

34 Knott, S.R. *et al.* (2012) Forkhead transcription factors establish origin timing and long-range clustering in *S. cerevisiae*. *Cell* 148, 99–111

35 Herrick, J. *et al.* (2000) Replication fork density increases during DNA synthesis in *X. laevis* egg extracts. *J. Mol. Biol.* 300, 1133–1142

36 Lucas, I. *et al.* (2000) Mechanisms ensuring rapid and complete DNA replication despite random initiation in *Xenopus* early embryos. *J. Mol. Biol.* 296, 769–786

37 Labit, H. *et al.* (2008) DNA replication timing is deterministic at the level of chromosomal domains but stochastic at the level of replicons in *Xenopus* egg extracts. *Nucleic Acids Res.* 36, 5623–5634

38 Goldar, A. *et al.* (2008) A dynamic stochastic model for DNA replication initiation in early embryos. *PLoS ONE* 3, e2919

39 Gauthier, M.G. and Bechhoefer, J. (2009) Control of DNA replication by anomalous reaction–diffusion kinetics. *Phys. Rev.Lett.* 102, 158104

40 Harland, R.M. and Laskey, R.A. (1980) Regulated replication of DNA microinjected into eggs of *Xenopus laevis*. *Cell* 21, 761–771

41 Hyrien, O. and 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

42 Yang, S.C. and Bechhoefer, J. (2008) How *Xenopus laevis* embryos replicate reliably: investigating the random-completion problem. *Phys. Rev. E* 78, 041917

43 Graham, C.F. (1966) The regulation of DNA synthesis and mitosis in multinucleate frog eggs. *J. Cell Sci.* 1, 363–374

44 Goldar, A. *et al.* (2009) Universal temporal profile of replication origin activation in eukaryotes. *PLoS ONE* 4, e5899

45 Blumenthal, A.B. *et al.* (1974) The units of DNA replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* 38, 205–223

46 Lima-de-Faria, A. and Jaworska, H. (1968) Late DNA synthesis in heterochromatin. *Nature* 217, 138–142

47 Gilbert, N. *et al.* (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* 118, 555–566

48 Schwaiger, M. and Schubeler, D. (2006) A question of timing: emerging links between transcription and replication. *Curr. Opin. Genet. Dev.* 16, 177–183

49 MacAlpine, D.M. *et al.* (2004) Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.* 18, 3094–3105

50 Hiratani, I. *et al.* (2009) Replication timing and transcriptional control: beyond cause and effect: part II. *Curr. Opin. Genet. Dev.* 19, 142–149

51 Lieberman-Aiden, E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293

52 Ryba, T. *et al.* (2010) Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* 20, 761–770

53 Hayashi, M.T. and Masukata, H. (2011) Regulation of DNA replication by chromatin structures: accessibility and recruitment. *Chromosoma* 120, 39–46

54 Lebofsky, R. *et al.* (2006) DNA replication origin interference increases the spacing between initiation events in human cells. *Mol. Biol. Cell* 17, 5337–5345

55 Cayrou, C. *et al.* (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* 21, 1438–1449

56 Wong, P.G. *et al.* (2011) Cdc45 limits replicon usage from a low density of preRCs in mammalian cells. *PLoS ONE* 6, e17533

57 Krasinska, L. *et al.* (2008) Cdk1 and Cdk2 activity levels determine the efficiency of replication origin firing in *Xenopus*. *EMBO J.* 27, 758–769

58 Katsuno, Y. *et al.* (2009) Cyclin A-Cdk1 regulates the origin firing program in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3184–3189

59 Thomson, A.M. *et al.* (2010) Replication factory activation can be decoupled from the replication timing program by modulating Cdk levels. *J. Cell Biol.* 188, 209–221

60 Edwards, M.C. *et al.* (2002) MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in *Xenopus* egg extracts. *J. Biol. Chem.* 277, 33049–33057

61 Dijkwel, P.A. *et al.* (2002) Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies. *Mol. Cell. Biol.* 22, 3053–3065

62 Harvey, K.J. and Newport, J. (2003) CpG methylation of DNA restricts prereplication complex assembly in *Xenopus* egg extracts. *Mol. Cell. Biol.* 23, 6769–6779

63 Koren, A. *et al.* (2010) MRC1-dependent scaling of the budding yeast DNA replication timing program. *Genome Res.* 20, 781–790

64 Alvino, G.M. *et al.* (2007) Replication in hydroxyurea: it's a matter of time. *Mol. Cell. Biol.* 27, 6396–6406

65 Ma, E. *et al.* (2012) Do replication forks control late origin firing in *Saccharomyces cerevisiae*? *Nucleic Acids Res.* 40, 2010–2019

66 Rhind, N. (2006) DNA replication timing: random thoughts about origin firing. *Nat. Cell Biol.* 8, 1313–1316

67 Hozak, P. and Cook, P.R. (1994) Replication factories. *Trends Cell Biol.* 4, 48–52

68 Baddeley, D. *et al.* (2010) Measurement of replication structures at the nanometer scale using super-resolution light microscopy. *Nucleic Acids Res.* 38, e8

69 Chen, C.L. *et al.* (2011) Replication-associated mutational asymmetry in the human genome. *Mol. Biol. Evol.* 28, 2327–2337

70 Touchon, M. *et al.* (2005) Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9836–9841

71 Chagin, V.O. *et al.* (2010) Organization of DNA replication. *Cold Spring Harb. Perspect. Biol.* 2, a000737

72 Karschau, J. *et al.* (2012) Optimal placement of origins for DNA replication. *Phys. Rev. Lett.* 108, 058101

73 Ge, X.Q. *et al.* (2007) Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev.* 21, 3331–3341

74 Blow, J.J. *et al.* (2011) How dormant origins promote complete genome replication. *Trends Biochem. Sci.* 36, 405–414

75 Blow, J.J. and Ge, X.Q. (2009) A model for DNA replication showing how dormant origins safeguard against replication fork failure. *EMBO Rep.* 10, 406–412

76 Gauthier, M.G. *et al.* (2010) Defects and DNA replication. *Phys. Rev. Lett.* 104, 218104

77 Sancar, A. *et al.* (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* 73, 39–85

78 Herrick, J. (2011) Genetic variation and DNA replication timing, or why is there late replicating DNA? *Evolution* 65, 3031–3047

79 Cairns, J. (1963) The bacterial chromosome and its manner of replication as seen by autoradiography. *J. Mol. Biol.* 6, 208–213

80 Gratzner, H.G. (1982) Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: a new reagent for detection of DNA replication. *Science* 218, 474–475

81 Jackson, D.A. and Pombo, A. (1998) Replicon 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

82 Bensimon, A. *et al.* (1994) Alignment and sensitive detection of DNA by a moving interface. *Science* 265, 2096–2098

83 Michalet, X. *et al.* (1997) Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* 277, 1518–2123

84 Norio, P. and Schildkraut, C.L. (2001) Visualization of DNA replication on individual Epstein–Barr virus episomes. *Science* 294, 2361–2364

85 Kitamura, E. *et al.* (2006) Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell* 125, 1297–1308

86 van Oijen, A.M. and Loparo, J.J. (2010) Single-molecule studies of the replisome. *Annu. Rev. Biophys.* 39, 429–448

87 Riehn, R. *et al.* (2005) Restriction mapping in nanofluidic devices. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10012–10016

88 Sidorova, J.M. *et al.* (2009) Microfluidic-assisted analysis of replicating DNA molecules. *Nat. Protoc.* 4, 849–861

89 Raghuraman, M.K. *et al.* (2001) Replication dynamics of the yeast genome. *Science* 294, 115–121

90 Woodfine, K. *et al.* (2004) Replication timing of the human genome. *Hum. Mol. Genet.* 13, 191–202

91 Desprat, R. *et al.* (2009) Predictable dynamic program of timing of DNA replication in human cells. *Genome Res.* 19, 2288–2299

92 Chen, C.L. *et al.* (2010) Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.* 20, 447–457

93 Yabuki, N. *et al.* (2002) Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* 7, 781–789

94 Retkute, R. *et al.* (2011) Dynamics of DNA replication in yeast. *Phys. Rev. Lett.* 107, 068103

95 Jun, S. *et al.* (2005) Nucleation and growth in one dimension. I. The generalized Kolmogorov–Johnson–Mehl–Avrami model. *Phys. Rev. E* 71, 011908

96 Yang, S.C. *et al.* (2009) Computational methods to study kinetics of DNA replication. *Methods Mol. Biol.* 521, 555–573

97 Brummer, A. *et al.* (2010) Mathematical modelling of DNA replication reveals a trade-off between coherence of origin activation and robustness against rereplication. *PLoS Comput. Biol.* 6, e1000783