Mathematical modelling of eukaryotic DNA replication

Olivier Hyrien · Arach Goldar

Published online: 20 November 2009
© Springer Science + Business Media B.V. 2009

Abstract Eukaryotic DNA replication is a complex process. Replication starts at thousand origins that are activated at different times in S phase and terminates when converging replication forks meet. Potential origins are much more abundant than actually fire within a given S phase. The choice of replication origins and their time of activation is never exactly the same in any two cells. Individual origins show different efficiencies and different firing time probability distributions, conferring stochasticity to the DNA replication process. High-throughput microarray and sequencing techniques are providing increasingly huge datasets on the population-averaged spatiotemporal patterns of DNA replication in several organisms. On the other hand, single-molecule replication mapping techniques such as DNA combing provide unique information about cell-to-cell variability in DNA replication patterns. Mathematical modelling is required to fully comprehend

the complexity of the chromosome replication process and to correctly interpret these data. Mathematical analysis and computer simulations have been recently used to model and interpret genome-wide replication data in the yeast *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, in *Xenopus* egg extracts and in mammalian cells. These works reveal how stochasticity in origin usage confers robustness and reliability to the DNA replication process.

Keywords Replication origins · Replication fork progression · Genome stability · 1D nucleation and growth processes · Stochastic models

Abbreviations

ATM	Ataxia-telangiectasia mutated
ATR	ATM and Rad3 related
AT-rich	Adenine- and thymine-rich
Cdc6	Cell division cycle (factor) 6
Cdc45	Cell division cycle (factor) 45
Cdk2	Cell division kinase 2
Cdt1	cdc10-Dependent transcript 1
Clb5	Cyclin B5
Clb6	Cyclin B6
Dbf4	Dumbell-forming (factor) 4
Drf1	dbf4-Related factor 1
HU	Hydroxyurea
KJMA	Kolmogorov–Johnson–Mehl–Avrami (model for nucleation-and-growth)
MCM2-7	Minichromosome maintenance (proteins) 2 to 7

Responsible Editor: Marie-Noëlle Prioleau and Dean Jackson.

O. Hyrien (✉)
Ecole Normale Supérieure, UMR CNRS 8541,
46 rue d'Ulm,
75005 Paris, France
e-mail: hyrien@biologie.ens.fr

A. Goldar
Commissariat à l'Energie Atomique (CEA),
iBiTec-S,
91191 Gif-sur-Yvette, France
e-mail: arach.goldar@cea.fr

ORC	Origin recognition complex
1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional

Introduction

The duplication of eukaryotic chromosomes starts at multiple origins that are spread at irregular intervals and fire at different times in S phase (Huberman and Riggs 1968; Machida et al. 2005; Aladjem 2007). A large degree of flexibility in origin usage is observed between different tissues, developmental stages or even between individual cells of the same type (McKnight and Miller 1977; Hyrien et al. 1995; Norio et al. 2005; Czajkowsky et al. 2008). This suggests that potential origins are more abundant than are used in a single S phase, and that origin choice is partly stochastic.

A molecular basis for origin redundancy has been suggested by biochemical studies of the DNA replication "licensing" system (Blow and Dutta 2005). Chromatin is licensed for replication in late mitosis and G1 by the loading of multiple complexes of the MCM2-7 proteins in a process that depends on the origin recognition complex (ORC), Cdc6 and Cdt1 proteins. Cyclin- and Drf1/Dbf4-dependent protein kinases then phosphorylate MCM2-7 and other replication factors in S phase, resulting in the assembly of replication forks at origins. MCM2-7 provides helicase activity in front of the forks and are displaced from replicated DNA. Once S phase has started, a variety of mechanisms prevent MCM2-7 from rebinding chromatin until past mitosis, so that DNA is licensed and replicated only once per cell cycle. Importantly, 10–20 times more MCM2-7 complexes are loaded onto chromatin than ORC and actual initiation events (Burkhart et al. 1995; Lei et al. 1996; Rowles et al. 1996). In some organisms, MCM2-7 complexes appear to spread at varying distances from their ORC-bound loading sites, and it has been suggested that each of them is competent for initiation, although only a small fraction of them are actually used during a given S phase (Lucas et al. 2000; Edwards et al. 2002; Harvey and Newport 2003). Thus, MCM2-7 chromatin loading can be artificially reduced to a significant extent without

compromising genome duplication kinetics (Mahbubani et al. 1997; Ying and Gautier 2005; Woodward et al. 2006). Nevertheless, excess MCM2-7 complexes become critical for genome stability in the presence of replication inhibitors, as they provide back-up origins than can rescue stalled forks (Woodward et al. 2006; Ge et al. 2007). Further reduction of MCM2-7 loading can compromise genome stability during normal S phase (Ibarra et al. 2008).

Genome duplication must be completed prior to cell division. This necessitates an appropriate distribution of initiation events and safeguard mechanisms that protect against replication fork failure (Hyrien et al. 2003; Willis and Rhind 2009). The number of activated origins per genome per cell cycle ranges from a few hundreds in yeasts to hundreds of thousands in frog early embryos. Individual origins have such a broad range of efficiencies (see Nomenclature for definitions) and firing times that the history of initiation events is never the same in any two cells.

Furthermore, replication fork velocity, as estimated by the slopes of replication timing profiles in yeast (Raghuraman et al. 2001) and by DNA combing in mammalian cells (Conti et al. 2007), appears to vary over a 10-fold range when travelling along different parts of the genome. However, the vast majority are within ± 2 -fold of the mean. The extremes of the range are quite rare and may be artefactual. Velocities may be overestimated using DNA combing if closely spaced replicons fire and merge during a single labelling pulse, and in population-averaged replication timing profiles, slopes no longer reflect fork velocities when the considered region can be replicated in opposite directions from alternative origins.

High-throughput techniques are providing increasingly huge datasets on the spatiotemporal patterns of DNA replication in several organisms, both in the form of population-averaged profiles and single-molecule data, which provide unique information about cell-to-cell variability. Mathematical modelling is required to fully comprehend the complexity of the chromosome replication process and to correctly interpret these data. Mathematical modelling has been used to address several important questions. Is stochastic origin usage compatible with a reliable replication end time? What is the time-dependent rate of origin usage and how is it controlled? Can stochastic origin usage offer protection against replication fork failure? How should we take stochasticity

into account to interpret whole-genome replication timing profiles?

The random completion problem: modelling DNA replication in *Xenopus*

Life without specific origins

Early embryos of the frog *Xenopus laevis* provide an extreme case of stochasticity in DNA replication (Hyrien et al. 2003). Early embryonic S phase is very brief (20 min) although the fork progression rate is homogeneously slow ($v=0.5$ kb/min at 20°C) across the genome. Rapid replication is ensured by a close spacing of initiation events. Quantitative 2D gel analysis of replication intermediates in *Xenopus* eggs, egg extracts and early embryos have shown that replication initiates at random with respect to DNA sequence and at mean ~ 10 -kb intervals (Hyrien and Méchali 1992, 1993; Mahbubani et al. 1992; Lucas et al. 2000). However, a random distribution of origins with a mean 10-kb spacing would result in $e^{-20/10}=13.5\%$ of interorigin distances >20 kb, the maximum length of DNA that can be replicated by a pair of forks within 20 min. Nevertheless, early embryonic blastomeres reliably complete replication within 20 min.

Two potential solutions have been proposed to explain this paradox, known as the random completion problem (Hyrien et al. 2003). The "regular spacing" model assumes that origins are located at regular (not random) intervals, albeit with no regard to specific DNA sequences, and are activated with a close to 100% efficiency (Hyrien and Méchali 1993; Blow et al. 2001). The "origin redundancy" model instead assumes that potential origins are in excess and that the probability of origin firing increases as replication progresses so that additional initiations ensure the completion of unreplicated gaps before the end of S phase (Lucas et al. 2000). Electron microscopy, DNA combing and other DNA fibre techniques have been used to study the distribution of replication "eyes" (or bubbles; these terms come from the appearance of replicated segments on electron micrographs) on single DNA molecules replicating in egg extracts (Herrick et al. 2000; Lucas et al. 2000; Blow et al. 2001; Marheineke and Hyrien 2001). These studies revealed that (1) initiation occurs

throughout S phase; (2) eye-to-eye distances are not regularly distributed; (3) the ease with which potential origins fire, deduced from the abundance of small eyes at various stages of replication, increases as replication progresses. These data favoured the origin redundancy/increasing firing probability model. The way origin firing probability changes during S phase defines the replication timing programme of the early *Xenopus* embryo. Once the initiations have been set, the rest of the process is deterministic.

Application of the KJMA model to DNA replication kinetics

Bechhoefer and colleagues have adapted theories of phase transition kinetics to extract an analytical expression for the time-dependent changes in origin activation probability from the DNA fibre data (Herrick et al. 2002; Jun and Bechhoefer 2005; Jun et al. 2005; Zhang and Bechhoefer 2006). Some phase transition phenomena, e.g. growth of a crystal from a liquid, result from three simultaneous processes: nucleation of solid domains, growth of existing domains, and coalescence (merge of expanding domains), which are formally analogous to replication initiation, elongation and termination. In the 1930s, Kolmogorov (1937), Johnson and Mehl (1939), and Avrami (1939, 1940, 1941) independently derived a stochastic model (the KJMA model) that could describe the fraction of volume $f(t)$ of a liquid that has crystallised by time t , which experimentally is a sigmoidal curve. In the model's simplest form, nucleation occurs within the liquid at a constant rate in space and time and growth occurs at a constant speed. In the 1980s, Sekimoto (1984a, b, 1991) showed that if growth occurs in only one dimension (as in DNA replication) the analysis can be pushed further to describe the statistics and time evolution of solid domain sizes and spacing (replication eyes, gaps and eye-to-eye distances). Bechhoefer and colleagues extended this analysis to the case of an arbitrary initiation rate $I(t)$, defined as the spatially averaged number of initiations per unit time per unit length of unreplicated DNA. They developed a numerical inversion procedure to infer $I(t)$ from v , the observed velocity of replication forks and $f(t)$, the replicated fraction at time t .

$I(t)$ is assumed to be spatially homogeneous, which is justified if potential origins are abundant and

independent, but it needs not be constant in time. Note that $I(t)$ is a different concept from origin efficiency. Origin efficiency is the overall frequency at which an origin fires within a population of cells, whereas $I(t)$ measures the ease with which unused origins fire at different times in S phase.

The extracted $I(t)$ was found to markedly increase halfway through S phase then decrease sharply (Herrick et al. 2002; Zhang and Bechhoefer 2006). Initially, the decreasing part of $I(t)$ was not taken in consideration because the numerical inversion procedure tends to amplify noise and the possibility of systematic errors at the end of S phase could not be discarded. When an analytical expression of $I(t)$ that fits the increasing part of the data was incorporated into Monte-Carlo simulations of a random replication process, the simulations could account for the mean eye lengths, gap lengths and eye-to-eye distances observed at different times in S phase. It was concluded that this kinetic model, starting from the two fundamental parameters $I(t)$ and v , adequately describes the progress of DNA replication, independently of the underlying biochemical mechanisms.

Relating replication initiation rate to replication end time distributions

Firing origins more rapidly as S phase progresses offers a plausible mechanism to accelerate the completion of unreplicated gaps. But what is the robustness of this solution? Using concepts from extreme-value statistics, the distribution of replication end times for an arbitrary $I(t)$ was mathematically calculated (Bechhoefer and Marshall 2007; Yang and Bechhoefer 2008). Adjusting the amplitude of $I(t)$ so as to fit the mode of the distribution to the data, it was found that initiating all origins at the beginning of S phase would lead to the broadest possible dispersion of end times around the mode. In contrast, a constant rate of initiation narrowed this distribution, a linearly increasing $I(t)$ narrowed it more, and a quadratically increasing $I(t)$, which best fit the experimental $I(t)$, narrowed it even more, ensuring a more reproducible length of S phase.

Recently, the $I(t)$ profile was determined without relying on a numerical inversion procedure but by counting individual initiation events on combed DNA fibres at different $f(t)$ (Goldar et al. 2008). This analysis showed that the decreasing part of $I(t)$, which

was not initially taken into account, is not a statistical artefact but a real feature of chromosome replication. The inclusion of the decreasing part of $I(t)$ did not dramatically alter the distribution of replication end times, suggesting that increasing origin usage around mid-S phase rather than at its end can also ensure a reliable replication end-time despite random initiation (Yang and Bechhoefer 2008). Interestingly, this shape of $I(t)$ was found to be conserved among distant eukaryotes, suggesting that it serves more general purposes than resolving the random completion problem (Goldar et al. 2009; see below).

Explaining the observed rate of origin usage

Assuming that potential origins are in excess and independent, the temporal programme of DNA replication is defined solely by $I(t)$ and v . All other observables (fork density, replicated fraction, distributions of eye, gap and eye-to-eye sizes) follow from these two fundamental parameters. Two models to explain the observed shape of $I(t)$ in terms of protein-DNA interactions have been proposed.

Goldar et al. (2008) have performed simulations wherein the encounter of a limiting, recyclable replication fork factor with potential origins results in productive initiation with probability P . When forks meet, the factor is released and made available again for initiation. The total number of factor molecules N_T and the probability P were either held constant or allowed to change with time. When N_T and P were held constant, the calculated $I(t)$ profile could not match the data, even after extensive parameter testing. However, an excellent match was obtained when two conditions were combined: N_T linearly increased with time, and P increased with fork density in a self-limiting manner. By itself, the first condition allowed a match to the increasing but not the decreasing part of the data. The second condition alone resulted in $I(t)$ profiles that increased then decreased but could not match the data.

The requirement for an increasing N_T can be related to the observation that several replication factors are actively concentrated in the nucleus during replication in egg extracts (Walter et al. 1998). An interpretation of the self-limiting dependency of initiation on fork density is that forks may facilitate nearby initiation until their density exceeds some threshold. Forks may stimulate initiation because

Cdc45, a stable fork component, can recruit Cdk2, a kinase involved in origin activation, at replication foci (Alexandrow and Hamlin 2005). The limitation of fork-stimulated initiation at high fork density may be related to the observation that ATM/ATR checkpoints negatively regulate initiation through a feedback mechanism originating from exposure of single-stranded DNA at replication forks (Marheineke and Hyrien 2004; Shechter et al. 2004).

Gauthier and Bechhoefer (2009) have proposed a different model to explain the observed shape of $I(t)$, based on anomalous diffusion theories. In their model, the initiation rate is set by the time needed for initiation factors to find (by diffusion) and then activate (reaction) potential origins. The search time t_s for one initiator molecule to find one potential origin was calculated according to previous theories in which a searcher particle diffuses in the 3D space until it nonspecifically binds to the DNA, then perform a 1D search along the DNA, then dissociates and starts another search cycle until it finds its target (Berg et al. 1981). The model assumes no spatial correlations between consecutive nonspecific bindings. The search time t_s was expressed as a function of: the duration of one search cycle; the average 1D-search distance; the density of potential origins; the fork velocity; the genome length; the replicated fraction; and the diffusion exponent α , which defines how the mean-square displacement of a particle scales with time ($\langle x^2 \rangle \sim t^\alpha$). Assuming a constant activation time t_r , the number of initiations per time per length of unreplicated DNA for N_s searchers was then calculated. Two conditions were required to fit this expression to combing data. First, the number of searchers, $N_s(t)$, had to increase with time, as proposed by Goldar et al. (2008). Second, the decreasing part of the data could be explained if $\alpha < 1$, i.e. if the interaction between DNA and searchers is sub-diffusive. It was concluded that initiation is reaction-limited during most of S phase but becomes suppressed at the end of S phase because the diffusion-limited search time dominates.

In conclusion, both models require that the amount of a limiting replication factor increases during S phase, e.g. by nuclear import, to explain the increasing part of $I(t)$. The diffusion model can naturally account for both the increasing and the decreasing part of $I(t)$ without explicitly tying initiation rate to fork density. However, the fork-density dependent

model can more easily incorporate biological observations such as the downregulation of initiation by ATM/ATR when fork density increases above some threshold.

Is DNA replication completely random in space and time in *Xenopus* egg extracts?

Some experimental data do not entirely agree with a purely random model for *Xenopus* DNA replication. First, initiation shows a ~ 2 -fold preference for AT-rich DNA sequences (Labit et al. 2008; Stanojic et al. 2008). Second, in *Xenopus* as in other eukaryotes, active replication forks at any given time in S phase are not randomly distributed in the nucleus but are concentrated in a few hundreds of intranuclear sites that are known as replication foci and are believed to represent large (~ 1 Mb) chromosomal segments that replicate at specific times (Nakamura et al. 1986; Berezney et al. 2000). Labit et al. (2008) labelled DNA replicated at the onset of two successive S phases in two different colours. When doubly labelled nuclei were visualised, a strong coincidence of both labels was observed at the level of replication foci, showing that chromosomal segments comprising hundreds of replicons are replicated in a non-random, reproducible temporal sequence. When the DNA from similarly labelled nuclei was purified and combed; however, the two labels did not coincide on the combed fibres; therefore, within each domain origins are activated in a random temporal order. Although the determinants of the large-scale, deterministic temporal order of replication are unknown, its existence has implications for the interpretation of $I(t)$ profiles. Calculations of $I(t)$ profiles assumed that DNA fibres of comparable $f(t)$ have similar replication starting times, which is not entirely exact. The time origin in the $I(t)$ profiles more exactly reflects each fibre's starting time than the start of S phase.

A third inconsistency with a purely random replication model is the existence of a small but significant size correlation between adjacent replication eyes, suggesting that neighbouring initiations tend to occur synchronously (Blow et al. 2001; Jun et al. 2004; Marheineke and Hyrien 2004). Furthermore, the distributions of eye-to-eye distances are slightly peakier than predicted by the KJMA model, suggesting spatial correlations (Jun et al. 2004). A model in which loop formation between a potential origin and a

fork stimulates initiation was proposed to explain these spatiotemporal correlations (Jun et al. 2004). In a semi-flexible polymer, loops that are too small cost too much bending energy, while loops that are too large cost too much entropy because they reduce the number of conformations that the chain can explore. Balancing these effects gives an optimal loop size that depends on the fibre's stiffness, usually expressed as its persistence length. Jun et al. (2004) estimated this parameter by fitting a theoretical loop size probability distribution to the eye-to-eye distance distribution, then incorporated the calculated persistence length in simulations in which each origin had a different probability of initiation depending on how far it was from a left and right approaching forks (note that this model implicitly ties the initiation rate to the local fork density and enforces synchrony of neighbouring origins). The results predicted an optimal loop size of ~11 kb and an eye-to-eye distance distribution that better agreed with the data than the standard 1D KJMA model, suggesting that mechanical properties of the chromatin fibre may contribute to regularising the distribution of initiation events.

Yang and Bechhoefer (2008) investigated whether the experimentally observed degree of spatial ordering was sufficient to make the replication end-time distribution deviate from the random case. The genome was modelled as a lattice of potential origins whose spacing and efficiency were continuously increased, allowing a continuous interpolation from complete randomness to perfect periodicity while maintaining the total initiation probability constant. The resulting end-time distribution was unaltered as long as the lattice spacing remained <6.5 kb, the average inter-origin spacing in the random case. Spacing origins further apart shifted the end-time distribution to longer times but did not alter its width. As many observed eye-to-eye distances are <6.5 kb, any underlying regularity in origin placement must be too weak to be pertinent to the random completion problem.

In summary, a combination of biochemical experiments, mathematical modelling and computer simulations has shown how stochastic location and activation of origins ensures the fast and reliable replication needed for rapid development of early *Xenopus* embryos. These works provide a solid foundation to formulate more complex models appropriate to cells where origin location is less random

and fork velocity is more variable (Yang et al. 2009). The fact that a large genome can be faithfully duplicated while initiating origins randomly raises the question why initiation is random neither in space nor in time in many other organisms or in later developmental stages. Non-randomness probably results from the need to coordinate replication and transcription. This constraint is absent in early *Xenopus* embryos whose genome is transcriptionally silent until the mid-blastula stage.

Modelling DNA replication in *Schizosaccharomyces pombe*

Origin islands in a sea of genes

Replication origins in *S. pombe* were first identified using autonomous plasmid replication assays (Maundrell et al. 1988) and 2D gel electrophoresis of replication-fork-containing restriction fragments (Zhu et al. 1992; Dubey et al. 1994). Origins are 500–1,000 bp long, are relatively inefficient (30% on average) and display no consensus sequence except for an extreme AT richness, which has been used bioinformatically to identify 384 potential origins (Segurado et al. 2003). Later microarray studies confirmed these predictions and mapped nearly all potential origins in *S. pombe* (Feng et al. 2006; Heinricher et al. 2006; Eshaghi et al. 2007; Hayashi et al. 2007). All origins map to intergenic regions, and their sequences properties are similar to those of inter-genes in general. The relative ability of an inter-gene to serve as an origin does not depend on its precise sequence but only on its length and AT content, suggesting that origins active in any given S phase are recruited stochastically among these intergenic regions (Dai et al. 2005). DNA combing experiments have shown that (1) eye-to-eye distances are well fit by an exponential distribution; (2) there is no significant correlation between firing of neighbouring origins, inconsistent with coordinated regulation of origin clusters and (3) there is no correlation between origin firings in sequential S phases (Patel et al. 2006). All these results are consistent with stochastic firing of origins, although in contrast to *Xenopus* embryos, origins do not occur at random sequences and display a broad range of efficiencies.

Defining origin efficiency

Origin efficiencies have been estimated by several methods. 2D gel electrophoresis of replication intermediates reveals the fraction of replication intermediates of a specific restriction fragment that contain either a replication bubble (initiation) or a single fork (passive replication). These studies first suggested that *S. pombe* origins fire in a variable and often small fraction of the cell cycles (Dubey et al. 1994). Combing DNA molecules labelled with BrdU in the presence of hydroxyurea (HU), an agent which slows down replication forks and confines DNA synthesis to the vicinity of origins, showed that origins fired early on from 8% or less to 59% of fibres (mean 33%) (Patel et al. 2006). Heichinger et al. (2006) used microarrays to identify 904 potential origins and defined their efficiencies as the relative increase in DNA content in cells synchronously released from a late G2 block into S phase in the presence of HU. Only a few origins were >50% efficient and most were <10% efficient, consistent with the combing study. Plotting efficiency as a function of replication time (determined in the absence of HU) revealed that early origins tend to be more efficient. A caveat in these experiments is that cells were exposed to HU for a time that only allowed limited DNA synthesis, so that late-firing but efficient origins may have gone unnoticed. In an independent microarray study, Eshaghi et al. (2007) have defined efficiency as the rate of copy number increase in cells released from an HU block and reported that efficiently replicating origins tend to fire late, an opposite conclusion to Heichinger et al. However, the method of Eshaghi et al. does not discriminate between initiation and passive replication by incoming forks from neighbouring origins. More experiments are required to determine the fraction of cells in which each potential origin fires in unperturbed cells, and to determine the time-dependent rate of initiation of individual origins.

Stochastic hybrid model of DNA replication in *S. pombe*

Lygeros et al. (2008) have proposed a model for DNA replication in *S. pombe* where initiations events are characterised by uncertainty in time and space. Each origin was characterised by an intrinsic "propensity" to fire, which denotes the instantaneous probability of

firing of an as yet unfired origin (see Nomenclature). This concept is similar to $I(t)$ except that in *S. pombe*, unlike in *Xenopus*, the probability of firing is not spatially homogeneous. Initially, the firing propensity of each origin, λ_i , was assumed to be constant in time, so that in the absence of passive replication, the probability that an origin has fired by time t is $P(t) = 1 - e^{-\lambda_i t}$. This equation implies that strong origins fire early and weak origins fire late. To estimate the parameter λ_i for each origin, $P(t)$ was equated with the origin efficiencies determined by Heichinger et al. (2006), setting $t=20$ min as the normal duration of S phase. Fork velocity was set constant at 3 kb/min. The model was used to simulate the DNA combing experiment of Patel et al. (2006). A close agreement was observed between the simulated and observed inter-bubble distances. However, the predicted replication end time was broadly distributed with a much longer mean (67 min) than observed (20 min), although the number of fired origins was close to observed. Increasing the firing propensity of all origins decreased the predicted duration of S phase but also increased the number of fired origins to unrealistic values. Increasing fork velocity (to 6 kb/min) or introducing additional inefficient origins decreased S phase length to 45 min, a still too long time. The model was then modified to let the firing propensity of unfired origins increase during S phase, keeping the total firing propensity constant through S phase. This constraint is equivalent to postulating that a limiting initiation factor, whose concentration remains constant during S phase, is released following origin firing and binds again to unfired origins. In this scenario the average end time was reduced to 33 min. Similar results (average end time=37 min) were obtained when the limiting factor was released when forks meet or when an origin is passively replicated.

The authors concluded that increasing the firing propensity of individual origins during S phase offers a possible solution to the random completion problem in fission yeast, as in *Xenopus*. They also considered that DNA replication could in fact take longer than usually accepted and extend into what we now define as G2-phase. This possibility can be rejected in *Xenopus* because S phase is immediately followed by mitosis. However, G2 makes the largest part of the fission yeast cell cycle and a prolongation of DNA synthesis across different regions in different cells may have gone unnoticed so far.

The model assumes that the limiting factor released from fired origins binds efficiently to unfired origins and predicts that the mean firing propensity increases most markedly during the last 5 min of S phase (Figure S2B in Lygeros et al. (2008)). However, simulations have shown that recycling is not efficient when the number of unfired origins decreases as S phase progresses, even if the affinity of potential origins for the factor is very high (Figure 2A in Goldar et al. (2008)). Furthermore, the spatially averaged profile of $I(t)$, which is equivalent to the mean firing propensity, was recently extracted (Goldar et al. 2009) from the published *S. pombe* replication timing profiles (Heichinger et al. 2006; Eshaghi et al. 2007). It was found that $I(t)$ increases during the first half of S phase then decreases, in contrast to the predictions of the model. As it seems to us, it would be interesting to match the time-dependent change in mean firing propensity with the experimental $I(t)$ profile and examine if the replication end time distribution becomes more realistic.

Modelling DNA replication in *Saccharomyces cerevisiae*

Replication programme in *S. cerevisiae*

In contrast to other eukaryotes, replication origins in *S. cerevisiae* are strictly determined by DNA sequence. An 11-bp consensus sequence and additional sequences variable in size and location relative to this consensus are required for replication initiation (Newlon and Theis 1993). The efficiencies of most of the origins of chromosome VI have been found by 2D gel and fork direction analysis techniques to be extremely variable, ranging from <5% to >90% (Friedman et al. 1997; Yamashita et al. 1997). The first genome-wide replication timing profile of *S. cerevisiae* was reported by Raghuraman et al. (2001), who identified 332 origins as regions that replicate before their neighbours. Origins show a continuum of replication times. Although an origin's replication time is often assumed to reflect its activation time, this is only true for the most efficient origins. For less efficient origins, replication time is an average of initiation and passive replication times. Fork rates were estimated from the slopes of these peaks to 0.5–

11 kb/min, with a mean of 2.9 kb/min and a median of 2.3 kb/min (as explained in the “Introduction”, the extremes of this range may result from misinterpretation of the profile slopes). Since then, experimental and computational studies (Wyrick et al. 2001; Yabuki et al. 2002; Feng et al. 2006; Nieduszynski et al. 2006; Xu et al. 2006) have identified and mapped a total of 732 origins (Nieduszynski et al. 2007), and there are replication timing data for 454 of them in the study of Raghuraman et al. (2001).

Czajkowski et al. (2008) have used DNA combing to visualise the replication patterns of single chromosome VI molecules. Contrary to what would be expected for a strictly deterministic temporal program, no two molecules exhibited the same pattern, and replication of different regions of the same chromosome occurred independently of each other. Nevertheless, averaging the patterns of all the fibres examined recapitulated the ensemble-averaged patterns obtained from population studies. One extreme interpretation of the single molecule data is that the observed kinetics of replication simply results from stochastic processes acting on origins with different probabilities of activation. Indeed, 2D gel studies suggested that there is a strong—though not absolute—correlation between origin efficiency and firing order (Yamashita et al. 1997).

To investigate whether origin firing time is probabilistic or deterministic, McCune et al. (2008) have compared genome-wide kinetics of replication in the wild-type and in a *clb5Δ* mutant. During normal S phase, origins are activated by Clb5- or Clb6-dependent kinases. Clb6 is rapidly degraded as cells enter S phase whereas Clb5 persists until mitosis (Jackson et al. 2006). In the absence of Clb6, S phase proceeds with no apparent defect whereas in the absence of Clb5 S phase requires 50% more time for completion (Schwob and Nasmyth 1993; Donaldson et al. 1998). McCune et al. (2008) show that regions that replicate early in the wild-type replicate with the same kinetics in the *clb5Δ* mutant whereas regions that replicate late in the wild-type replicate even later in the mutant. The late origins do not seem a preferred substrate for Clb5-dependent kinase since expression of a long-lived version of Clb6 completely rescues replication kinetics. McCune et al. suggest that these results argue against a purely stochastic model of origin firing because the efficiency of only some origins is affected. However, it remains possible that

the most efficient origins have all fired before Clb6 becomes degraded so that only the less efficient late origins are detectably affected. Resolving this issue requires quantitative modelling (see also the review by Rhind et al. in this issue of Chromosome Research).

A deterministic model of *S. cerevisiae* genome replication

The data described above suggest that origins are incompletely efficient and have a wide range of firing time distributions. Neglecting this complexity, Spiesser et al. (2009) have provided a deterministic model for DNA replication in *S. cerevisiae*. They used the length of the chromosomes, the position of 454 origins, their mean initiation time and a fork migration rate assumed to be constant at 3 kb/min to recalculate the replication timing profiles of all 16 chromosomes, assuming that the efficiency of the selected 454 origins was 100%. In spite of these approximations, the recalculated profiles matched the experimental profiles surprisingly well for chromosomes II, V, VII, XI, XIII, XV, XVI. However, significant differences were observed for the other chromosomes, with many sequences replicating later than predicted. Whereas the slope of the recalculated profiles is constant due to the constant fork rate implemented, the experimental curve is smooth with a varying slope, which may reflect a changing fork progression rate or the activation of inefficient origins that were not included in the simulation.

Spiesser et al. (2009) modelled replication in the *clb5Δ* mutant by stopping origin activation at a time point that corresponds to the mean value of origin firing times (27 min, about the midpoint of a normal 60-min S phase). The delayed regions in the calculated profiles exactly matched the in vivo data for chromosomes I to VIII and IX, less exactly for chromosomes IX, X and XIV and very poorly for chromosomes XII, XIII, XV and XVI. A better match may be obtained when origins are each assigned a distribution of firing times and a range of efficiencies, and when Clb6 is modelled to decay over a 5–10 min period as observed in vivo (Jackson et al. 2006) rather than abruptly at 27 min. This will be required to conclusively determine whether other factors than origin efficiencies contribute to their order of activation.

Stochastic models of *S. cerevisiae* chromosome replication

De Moura et al. (C. Nieduszynski, pers. comm.) have recently developed a stochastic model for *S. cerevisiae* chromosome VI replication in which each origin is characterized by its position, its "activation competence" (as estimated from plasmid maintenance assays; (Shirahige et al. 1993; Chang et al. 2008)), and a Gaussian distribution of activation times where the mean is estimated from timed 2D gels (Yamashita et al. 1997) and the standard deviation (σ) is simulated. The fork velocity is held constant (1.4 kb/min, according to Friedman et al. (1997)), and origins are assumed to fire independently of each other. Stochastic simulation of chromosome replication gives the replication time profile of a single cell, and the simulation is repeated a large number of times from which population averages are compared with experimental data, assuming perfect cell synchrony in the population. Note that according to the definitions listed in Nomenclature, plasmid maintenance assays are strictly measurements of potential efficiency, not competence. In practice, however, the difference can perhaps be neglected. Origin competence is a distinct notion from origin efficiency and origin firing probability. Origin efficiency is equal to origin competence times the probability that the origin is not passively replicated, which depends on the positions, competences and activation times of neighbour origins. Note also that by assuming a Gaussian firing time distribution, the model implies that firing probability changes during S phase, to peak at the mean firing time.

Initial simulations assuming a constant σ for all origins reproduced the replication timing curve better than the deterministic model of Spiesser et al. (2009). Varying σ from 0 to 15 min did not dramatically alter the replication timing curves but affected the observed efficiency of each origin to a varying extent. However, no single value of σ resulted in efficiencies agreeing closely with the data. When σ was assumed to be equal to half the activation time for each origin, a better agreement was obtained.

De Moura et al. have stressed that a peak height in a replication timing profile is influenced by both origin activation time and competence as well as by neighbour origins, so that it is not possible to directly interpret a peak height as an origin activation time. To more rigorously interpret replication timing profiles,

the locations of origins were assumed and a genetic algorithm was used to estimate competences, activation times, fork velocity and σ . The estimated origin activation times best agreed with the experimental data of Friedman et al. (1997) when the authors used experimental determination of the proportion of chromosome VI that are replicated at various times in S phase (Alvino et al. 2007), allowing for different values of σ for each origin. The estimated σ values tended to increase with activation time. The estimated competences were not concordant with those measured by plasmid loss experiments (Shirahige et al. 1993; Chang et al. 2008) or in HU-treated checkpoint deficient cells (Feng et al. 2006). However, estimated origin efficiencies were in close agreement with those determined by 2D-gel analyses (Friedman et al. 1997; Yamashita et al. 1997).

Yang et al. (J. Bechhoefer, pers. comm.) have recently developed an analytical model that extends the previous formalism elaborated for *Xenopus* to incorporate variable origin position, variable fork velocities and probabilistic initiation. They used this general formalism describing DNA replication to compute ensemble averages of replication timing profiles and performed fits to the *S. cerevisiae* microarray data of McCune et al. (2008), which includes eight time points from 10 to 45 min into S phase at 5-min resolution. They were able to extract the distribution of firing times for each origin and found that using a constant fork velocity (2 kb/min) gives an equally good fit to the data as using a space- and time-dependent velocity. They also were able to identify origins that were not previously recognised because they do not have apparent peaks on the timing profile but are crucial to the fit. Importantly, the model recapitulates the observed firing-times and initiation rates and the fit reveals that later-firing origins have slower replication kinetics and greater variation in firing times, consistent with lower potential efficiencies. This work conclusively establishes that the reproducible temporal order of replication observed on population averages can emerge as a consequence of stochastic initiations, as proposed by Czajkowsky et al. (2008) (see also the review by Rhind et al. in this issue of Chromosome Research). In the future, this analytical model may be used to extract origin firing-time distributions and fork velocities genome-wide from sufficiently time- and space-resolved datasets for any organism.

Is there an optimal time-dependent rate of origin usage?

Recently, published replication timing profiles of entire genomes or chromosomes have been used to extract the spatially averaged $I(t)$ functions of *S. cerevisiae*, *S. pombe*, *Drosophila melanogaster* and *Homo sapiens* (Goldar et al. 2009). In these profiles, the mean replication time of each locus is plotted against chromosomal position such that efficient origins appear as peaks. The profiles were cut into slices of time and $I(t)$ was calculated for each time point as the ratio between the number of fired origins and the amount of unreplicated DNA. All the $I(t)$ profiles showed a strikingly similar shape to the $I(t)$ extracted from *X. laevis* DNA combing data (Goldar et al. 2008), increasing during the first half of S phase then decreasing before its end. In this work, the mean replication time of only the most efficient origins, rather than the full distribution of origin firing times, was analysed. The result is nevertheless intriguing, given the reported differences in replication initiation control between these organisms. One common feature of all eukaryotes, however, is origin redundancy and its corollary, incomplete efficiency, which confers randomness in origin choice and firing time even to sequence-determined origins. The $I(t)$ function represents the genome-averaged dynamic processes that controls the dynamics of replication fork assembly from soluble factors that are available in limited amounts and its universal shape suggests that it has been selected for optimal use of these factors. Models elaborated to explain the shape of $I(t)$ in *Xenopus* may thus more generally apply.

Evolution probably acts to minimise the rate of cell death due to replication completion failure. This rate has been estimated to $<3 \cdot 10^{-3}$ in *Xenopus* embryos and can be taken as a measure of the reliability of S phase (Yang and Bechhoefer 2008). Although the amplitude of any $I(t)$ could be adjusted to meet a predefined reliability constraint, this would change the maximal number of forks that need to be simultaneously active at one time. Bechhoefer and colleagues have explored the relationship between $I(t)$ and the maximal number of required forks (Bechhoefer and Marshall 2007; Yang and Bechhoefer 2008). They found that the $I(t)$ that minimises the maximum number of required forks to meet a given reliability constraint is the one that maintains a constant number

of forks throughout S phase or, in other words, recycles available fork proteins so efficiently that they work all the time. Confining initiation to the start of S phase, on the other hand, would require an 18-fold higher number of forks. The experimentally observed $I(t)$ requires a maximum number of forks that is only 3-fold higher than the theoretical optimum.

Why does the observed $I(t)$ differ from the theoretical optimum? The calculation of this optimum assumes that replication fork protein availability is constant through S phase, which is probably not the case, and that all the available factors can instantaneously assemble in forks at the start of S phase, and there is no obvious physiological mechanism for such a sharp transition. Maintaining a constant number of forks also implies that the frequency of initiation and the local density of forks increase to infinite values at the end of S phase, which is unrealistic. Replication termination may also require specific factors that can become limiting, and the time-dependent frequency of termination is set by the dynamics of origin activation. It might be advantageous for the cell to spread termination events throughout S phase rather than concentrate them at the end of S phase. If so, the optimum $I(t)$ must be the best compromise between an efficient recycling of rate-limiting initiation, elongation and termination factors. The optimum $I(t)$ defined in this way has not been calculated but it would be interesting to see if it better approximates the experimentally observed $I(t)$.

Modelling the consequences of replication fork failure in metazoan cells

The amount of MCM2-7 complexes loaded onto chromatin is in excess over the origins normally used. In *Xenopus* extracts, inhibiting ATM/ATR kinases with caffeine stimulates initiation ~2-fold, partly by derepressing new origin clusters and partly by enhancing initiations within clusters (Marheineke and Hyrien 2004; Shechter et al. 2004). When aphidicolin is added to slow down forks ~3-fold, caffeine addition can revert replication kinetics to normal without stimulating fork velocity, implying a 3-fold increase in origin firings compared to normal S phase (Woodward et al. 2006). In mammalian tissue culture cells, aphidicolin or HU can stimulate origin firing ~2-fold without the need to abolish checkpoint

activity, and new initiations are preferentially directed towards active clusters (Ge et al. 2007). These experiments suggest that potential origins are at least in 3-fold excess but are normally prevented to fire by checkpoints or by passive replication. The amount of chromatin-bound MCM2-7 can be reduced several fold without impairing normal DNA replication kinetics. In these circumstances, however, fewer origins can be activated following replicative stress and the cells become hypersensitive to stress (Ge et al. 2007; Ibarra et al. 2008). Thus, activation of redundant origins depends on loading a large amount of MCM2-7 proteins and is an important mechanism to protect against replication fork failure and prevent genetic instability.

Can stochastic initiation from redundant origins suffice to explain the levels of origin activation seen in vivo when forks are slowed down? Blow and Ge (2009) modelled a mammalian replicon cluster as a circular 250 kb DNA molecule containing from five to 100 potential origins whose initiation probabilities were constant in time and picked from a random distribution such that an average of five initiations in the 250-kb cluster was maintained. Simulations showed that replication end times were little affected by the trade-off between origin redundancy and efficiency. The increased origin firing obtained when human cells were treated with HU was modelled by reducing fork velocity to 25–33% of normal. As the forks took longer to reach potential origins, some of them underwent initiation in later time steps. In order to reach a mean spacing between fired origins that match the in vivo data, the model required at least 1–3 potential origins for each fired origin, which is well within the ~10-fold excess of MCM2-7 complexes over origins. Thus, the protective effect of redundant origins against fork stalling may not require any special pathway to regulate origin firing.

Forks can stall irreversibly when they encounter DNA lesions. To model how redundant origins can rescue the double stall of two converging forks, clusters were replicated with a fork stall rate of either 5×10^{-6} or 5×10^{-7} per base pair. Increasing the number of potential origins per 250-kb cluster decreased the percentage of double fork stall from 25% with five potential origins to 5% with 25 origins, while the total number of origin firings increased only slightly. Thus, origin redundancy not only solves the random completion problem but also protects against

double-fork stalls while minimising the use of replication fork proteins. As the probability of origin activation was set constant in time, the replication completion time increased significantly with fork stalling. It would be interesting to see how these results are affected when initiation probability increases during S phase.

Conclusion

Stochasticity has emerged as a fundamental component of DNA replication and mathematical modelling is becoming an indispensable tool to correctly interpret population-averaged replication timing profiles and single molecule replication patterns. Further modelling of diverse experimental systems will help elucidate universal and species-specific strategies in eukaryotic DNA replication and reveal molecular mechanisms that underlie position- and time-dependent changes in origin firing probability and ensure replication robustness.

Nomenclature

Origin	A site in the genome where replication can initiate. In some organisms and cell types, such sites may be well defined by <i>cis</i> -acting sequence features; in other organisms or cell types many or even all, sites in the genome may act as origins.
Origin competence	The fraction of cells in a population in which an origin is biochemically competent to fire. A biochemically competent origin may fail to fire and be passively replicated.
Origin firing, Origin activation or Origin initiation (used interchangeably)	The irreversible conversion of a competent origin into bidirectional replication forks.
Origin efficiency	The fraction of cells in which an origin fires during S phase.

Potential origin efficiency

The fraction of cells in which an origin would fire during S phase if not passively replicated. Experimentally, this can be measured in yeasts by plasmid maintenance assays since passive replication is impossible when the plasmid contains a single origin. Note that potential efficiency is always lower than competence because some competent origins may fail to fire (in the absence of passive replication) during a finite length S phase.

Firing probability or Instantaneous firing probability

The probability that an as yet unfired origin will fire during a specific time period. Mathematically, the instantaneous firing probability is described by a probability density function. Lygeros et al. (2008) have used the term origin firing propensity with the same meaning.

Firing rate or Instantaneous firing rate

The number of origins initiated per time per length of unreplicated genome.

Acknowledgements We thank C. Nieduszynski, J. Bechhoefer and N. Rhind for communicating unpublished work (De Moura et al. and Yang et al.) and their review in this volume and for their comments on the manuscript. The O.H. lab is supported by the Association pour la Recherche sur le Cancer, the Ligue Nationale Contre le Cancer and the Fondation pour la Recherche Médicale (équipe labellisée). A.G. is supported by the Commissariat à l'Energie Atomique.

References

- Aladjem MI (2007) Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat Rev Genet* 8:588–600
- Alexandrow MG, Hamlin JL (2005) Chromatin decondensation in S-phase involves recruitment of Cdk2 by Cdc45 and histone H1 phosphorylation. *J Cell Biol* 168:875–886

- Alvino GM, Collingwood D, Murphy JM, Delrow J, Brewer BJ, Raghuraman MK (2007) Replication in hydroxyurea: it's a matter of time. *Mol Cell Biol* 27:6396–6406
- Avrami M (1939) Kinetics of phase change. I. General theory. *J Chem Phys* 7:1103–1112
- Avrami M (1940) Kinetics of phase change. II. Transformation-time relations for random distributions of nuclei. *J Chem Phys* 8:212–224
- Avrami M (1941) Kinetics of phase change. III. Granulation, phase change, and microstructure. *J Chem Phys* 9:177–184
- Bechhoefer J, Marshall B (2007) How *Xenopus laevis* replicates DNA reliably even though its origins of replication are located and initiated stochastically. *Phys Rev Lett* 98:098105
- Berezney R, Dubey DD, Huberman JA (2000) Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* 108:471–484
- Berg OG, Winter RB, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929–6948
- Blow JJ, Dutta A (2005) Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol* 6:476–486
- Blow JJ, Ge XQ (2009) A model for DNA replication showing how dormant origins safeguard against replication fork failure. *EMBO Rep* 10:406–412
- Blow JJ, Gillespie PJ, Francis D, Jackson DA (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* 152:15–26
- Burkhardt R, Schulte D, Hu D, Musahl C, Gohring F, Knippers R (1995) Interactions of human nuclear proteins P1Mcm3 and P1Cdc46. *Eur J Biochem* 228:431–438
- Chang F, Theis JF, Miller J, Nieduszynski CA, Newlon CS, Weinreich M (2008) Analysis of chromosome III replicators reveals an unusual structure for the ARS318 silencer origin and a conserved WTW sequence within the origin recognition complex binding site. *Mol Cell Biol* 28:5071–5081
- Conti C, Sacca B, Herrick J, Lalou C, Pommier Y, Bensimon A (2007) Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol Biol Cell* 18:3059–3067
- Czajkowsky DM, Liu J, Hamlin JL, Shao Z (2008) DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *J Mol Biol* 375:12–19
- Dai J, Chuang RY, Kelly TJ (2005) DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc Natl Acad Sci U S A* 102:337–342
- Donaldson AD, Raghuraman MK, Friedman KL, Cross FR, Brewer BJ, Fangman WL (1998) CLB5-dependent activation of late replication origins in *S. cerevisiae*. *Mol Cell* 2:173–182
- Dubey DD, Zhu J, Carlson DL, Sharma K, Huberman JA (1994) Three ARS elements contribute to the *ura4* replication origin region in the fission yeast, *Schizosaccharomyces pombe*. *Embo J* 13:3638–3647
- Edwards MC, Tutter AV, Cvetic C, Gilbert CH, Prokhorova TA, Walter JC (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
- Eshaghi M, Karuturi RK, Li J, Chu Z, Liu ET, Liu J (2007) Global profiling of DNA replication timing and efficiency reveals that efficient replication/firing occurs late during S-phase in *S. pombe*. *PLoS ONE* 2:e722
- Feng W, Collingwood D, Boeck ME et al (2006) Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* 8:148–155
- Friedman KL, Brewer BJ, Fangman WL (1997) Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2:667–678
- Gauthier MG, Bechhoefer J (2009) Control of DNA replication by anomalous reaction-diffusion kinetics. *Phys Rev Lett* 102:158104
- Ge XQ, Jackson DA, Blow JJ (2007) Dormant origins licensed by excess Mcm2–7 are required for human cells to survive replicative stress. *Genes Dev* 21:3331–3341
- Goldar A, Labit H, Marheineke K, Hyrien O (2008) A dynamic stochastic model for DNA replication initiation in early embryos. *PLoS ONE* 3:e2919
- Goldar A, Marsolier-Kergoat MC, Hyrien O (2009) Universal temporal profile of replication origin activation in eukaryotes. *PLoS One* 4:e5899
- Harvey KJ, Newport J (2003) CpG methylation of DNA restricts prereplication complex assembly in *Xenopus* egg extracts. *Mol Cell Biol* 23:6769–6779
- Hayashi M, Katou Y, Itoh T et al (2007) Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *Embo J* 26:1327–1339
- Heichinger C, Penkett CJ, Bahler J, Nurse P (2006) Genome-wide characterization of fission yeast DNA replication origins. *Embo J* 25:5171–5179
- 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
- Herrick J, Jun S, Bechhoefer J, Bensimon A (2002) Kinetic model of DNA replication in eukaryotic organisms. *J Mol Biol* 320:741–750
- Huberman JA, Riggs AD (1968) On the mechanism of DNA replication in mammalian chromosomes. *J Mol Biol* 32:327–341
- Hyrien O, Méchali M (1992) Plasmid replication in *Xenopus* eggs and egg extracts: a 2D gel electrophoretic analysis. *Nucleic Acids Res* 20:1463–1469
- Hyrien O, Méchali 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
- Hyrien O, Maric C, Méchali M (1995) Transition in specification of embryonic metazoan DNA replication origins. *Science* 270:994–997
- Hyrien O, Marheineke K, Goldar A (2003) Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *BioEssays* 25:116–125
- Ibarra A, Schwob E, Mendez J (2008) Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci U S A* 105:8956–8961
- Jackson LP, Reed SI, Haase SB (2006) Distinct mechanisms control the stability of the related S-phase cyclins Clb5 and Clb6. *Mol Cell Biol* 26:2456–2466

- Johnson WA, Mehl PA (1939) Reaction kinetics in processes of nucleation and growth. *Trans AIMME* 135:416–442
- Jun S, Bechhoefer J (2005) Nucleation and growth in one dimension. II. Application to DNA replication kinetics. *Phys Rev E Stat Nonlin Soft Matter Phys* 71:011909
- Jun S, Herrick J, Bensimon A, Bechhoefer J (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
- Jun S, Zhang H, Bechhoefer J (2005) Nucleation and growth in one dimension. I. The generalized Kolmogorov-Johnson-Mehl-Avrami model. *Phys Rev E Stat Nonlin Soft Matter Phys* 71:011908
- Kolmogorov AN (1937) On the statistical theory of crystallization in metals. *Izv Akad Nauk SSSR, Ser Fiz* 1:355–359
- Labit H, Perewoska I, Germe T, Hyrien O, Marheineke K (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
- Lei M, Kawasaki Y, Tye BK (1996) Physical interactions among Mcm proteins and effects of Mcm dosage on DNA replication in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16:5081–5090
- Lucas I, Chevrier-Miller M, Sogo JM, Hyrien O (2000) Mechanisms ensuring rapid and complete DNA replication despite random initiation in *xenopus* early embryos. *J Mol Biol* 296:769–786
- Lygeros J, Koutroumpas K, Dimopoulos S et al (2008) Stochastic hybrid modeling of DNA replication across a complete genome. *Proc Natl Acad Sci U S A* 105:12295–12300
- Machida YJ, Hamlin JL, Dutta A (2005) Right place, right time, and only once: replication initiation in metazoans. *Cell* 123:13–24
- Mahbubani HM, Paull T, Elder JK, Blow JJ (1992) DNA replication initiates at multiple sites on plasmid DNA in *Xenopus* egg extracts. *Nucleic Acids Res* 20:1457–1462
- Mahbubani HM, Chong JP, Chevalier S, Thommes P, Blow JJ (1997) Cell cycle regulation of the replication licensing system: involvement of a Cdk-dependent inhibitor. *J Cell Biol* 136:125–135
- Marheineke K, Hyrien O (2001) Aphidicolin triggers a block to replication origin firing in *Xenopus* egg extracts. *J Biol Chem* 276:17092–17100
- Marheineke K, Hyrien O (2004) Control of replication origin density and firing time in *Xenopus* egg extracts: role of a caffeine-sensitive, ATR-dependent checkpoint. *J Biol Chem* 279:28071–28081 Epub 22004 Apr 28028
- Maundrell K, Hutchison A, Shall S (1988) Sequence analysis of ARS elements in fission yeast. *Embo J* 7:2203–2209
- McCune HJ, Danielson LS, Alvino GM et al (2008) The Temporal Program of Chromosome Replication: Genome-wide Replication in *clb5*{ Δ } *Saccharomyces cerevisiae*. *Genetics*
- McKnight SL, Miller OL Jr (1977) Electron microscopic analysis of chromatin replication in the cellular blastoderm *Drosophila melanogaster* embryo. *Cell* 12:795–804
- Nakamura H, Morita T, Sato C (1986) Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp Cell Res* 165:291–297
- Newlon CS, Theis JF (1993) The structure and function of yeast ARS elements. *Curr Opin Genet Dev* 3:752–758
- Nieduszynski CA, Knox Y, Donaldson AD (2006) Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev* 20:1874–1879
- Nieduszynski CA, Hiraga S, Ak P, Benham CJ, Donaldson AD (2007) OriDB: a DNA replication origin database. *Nucleic Acids Res* 35:D40–46
- Norio P, Kosiyatrakul S, Yang Q 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
- Patel PK, Arcangioli B, Baker SP, Bensimon A, Rhind N (2006) DNA replication origins fire stochastically in fission yeast. *Mol Biol Cell* 17:308–316
- Raghuraman MK, Winzeler EA, Collingwood D et al (2001) Replication dynamics of the yeast genome. *Science* 294:115–121
- Rowles A, Chong JP, Brown L, Howell M, Evan GI, Blow JJ (1996) Interaction between the origin recognition complex and the replication licensing system in *Xenopus*. *Cell* 87:287–296
- Schwob E, Nasmyth K (1993) CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in *Saccharomyces cerevisiae*. *Genes Dev* 7:1160–1175
- Segurado M, de Luis A, Antequera F (2003) Genome-wide distribution of DNA replication origins at A+T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep* 4:1048–1053
- Sekimoto K (1984a) Kinetics of magnetization switching in a 1-D system - size distribution of unswitched domains. *Physica* 125A:261–269
- Sekimoto K (1984b) Kinetics of magnetization switching in a 1-D system II - Long time behavior of switched domains. *Physica* 128A:132–149
- 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
- Shechter D, Costanzo V, Gautier J (2004) ATR and ATM regulate the timing of DNA replication origin firing. *Nat Cell Biol* 6:648–655 Epub 2004 Jun 2027
- Shirahige K, Iwasaki T, Rashid MB, Ogasawara N, Yoshikawa H (1993) Location and characterization of autonomously replicating sequences from chromosome VI of *Saccharomyces cerevisiae*. *Mol Cell Biol* 13:5043–5056
- Spiesser TW, Klipp E, Barberis M (2009). A model for the spatiotemporal organization of DNA replication in *Saccharomyces cerevisiae*. *Mol Genet Genomics*
- Stanojcic S, Lemaitre JM, Brodolin K, Danis E, Mechali M (2008) In *Xenopus* egg extracts, DNA replication initiates preferentially at or near asymmetric AT sequences. *Mol Cell Biol* 28:5265–5274
- Walter J, Sun L, Newport J (1998) Regulated chromosomal DNA replication in the absence of a nucleus. *Mol Cell* 1:519–529
- Willis N, Rhind N (2009) Regulation of DNA replication by the S-phase DNA damage checkpoint. *Cell Div* 4:13
- Woodward AM, Gohler T, Luciani MG et al (2006) Excess Mcm2–7 license dormant origins of replication that can be used under conditions of replicative stress. *J Cell Biol* 173:673–683
- Wyrick JJ, Aparicio JG, Chen T et al (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*:

- high-resolution mapping of replication origins. *Science* 294:2357–2360
- Xu W, Aparicio JG, Aparicio OM, Tavaré S (2006) Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics* 7:276
- Yabuki N, Terashima H, Kitada K (2002) Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* 7:781–789
- Yamashita M, Hori Y, Shinomiya T et al (1997) The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2:655–665
- Yang SC, Bechhoefer J (2008) How *Xenopus* laevis embryos replicate reliably: Investigating the random-completion problem. *Phys Rev E Stat Nonlin Soft Matter Phys* 78:041917
- Yang SC, Gauthier MG, Bechhoefer J (2009) Computational methods to study kinetics of DNA replication. *Methods Mol Biol* 521:555–573
- Ying CY, Gautier J (2005) The ATPase activity of MCM2–7 is dispensable for pre-RC assembly but is required for DNA unwinding. *Embo J* 24:4334–4344
- Zhang H, Bechhoefer J (2006) Reconstructing DNA replication kinetics from small DNA fragments. *Phys Rev E Stat Nonlin Soft Matter Phys* 73:051903
- Zhu J, Brun C, Kurooka H, Yanagida M, Huberman JA (1992) Identification and characterization of a complex chromosomal replication origin in *Schizosaccharomyces pombe*. *Chromosoma* 102:S7–16