
Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/2492/1/Conrad_Nieduszynski--dynamics_DNA.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see:
http://creativecommons.org/licenses/by/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
Dynamics of DNA Replication in Yeast

Renata Retkute and Conrad A. Nieduszynski
Centre for Genetics and Genomics, University of Nottingham, Nottingham NG7 2UH, United Kingdom

Alessandro de Moura
Institute of Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, AB24 3UE United Kingdom

(Received 15 December 2010; published 4 August 2011)

We present a mathematical model for the spatial dynamics of DNA replication. Using this model we determine the probability distribution for the time at which each chromosomal position is replicated. From this we show, contrary to previous reports, that mean replication time curves cannot be used to directly determine origin parameters. We demonstrate that the stochastic nature of replication dynamics leaves a clear signature in experimentally measured population average data, and we show that the width of the activation time probability distribution can be inferred from this data. Our results compare favorably with experimental measurements in Saccharomyces cerevisiae.

DOI: 10.1103/PhysRevLett.107.068103

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.
We establish that the shape of the average replication time profile has a signature of the stochasticity of the replication process, even though it is a quantity defined by a population average. We derive an analytical expression relating the replication time in the regions between two origins to the standard deviation \( \Delta t \) of the activation time distribution of the origins. This is a valuable result, since few single cell experiments have been able to give direct information about the stochastic properties of the replication dynamics. Our results allow \( \Delta t \) to be obtained from widely available population-wide measurements. We apply this result to measured replication time data, and estimate \( \Delta t \) from the data, obtaining a result in agreement with current estimates in the literature.

In our model we consider a chromosome with \( N \) origins, where each origin \( i \) is defined by the following: its chromosomal position \( x_i \); the probability \( q_i \) that the origin achieves licensing (in a given cell within a population) and is thus capable of activating; and the activation time distribution \( \Pi_i(t) \), which is the probability density of origin \( i \) activating and starting bidirectional replication forks at time \( t \). Since an origin may not be competent in every cell within the population, in general \( q_i < 1 \), and \( \Pi_i(t) \) satisfies \( \int_{-\infty}^{+\infty} \Pi_i(t) dt = q_i \). The fundamental quantity from which all statistical properties of this system can be calculated is the probability density \( P(x, t) \), defined such that \( P(x, t) dt \) is the probability that the replication fork at time \( t \) is replicated between times \( t \) and \( t + dt \). If only origin \( i \) were present, \( P \) would be given by \( P(x, t) = \Pi_i(t - |x - x_i|/v) \), where \( v \) is the fork velocity, which we assume to be a constant.

In the presence of all \( N \) origins, the calculation of \( P(x, t) \) is complicated by the fact that position \( x \) can be replicated by forks originated from any of the origins. Let us assume that position \( x \) is replicated between times \( t \) and \( t + dt \) by a fork from origin \( i \). This requires that (i) origin \( i \) activates at time \( t - |x - x_i|/v \), so that the fork arrives at \( x \) at time \( t \); and (ii) all other origins \( j \neq i \) have either not activated or they have activated, but their forks would arrive at \( x \) later than \( t \). The probability density for event (i) is \( \Pi_i(x, t) = \Pi_i(t - |x - x_i|/v) \), and the probability for event (ii) is \( \Pi_j(x, t) = \prod_{j \neq i} \Pi_j(x, t) \), where \( \Pi_j(x, t) \) is the probability that a fork from origin \( j \) arrives later than \( t \), or fails to activate: \( \Pi_j(x, t) = 1 - \int_{-\infty}^{+\infty} \Pi_j(x, y) dy \), where \( s_i = 1 - q_i \), is the probability of origin \( i \) not being competent. Therefore, the probability density \( P_i(x, t) \) that position \( x \) is replicated by origin \( i \) at time \( t \) is

\[
P_i(x, t) = \Pi_i(x, t) \Pi_i(x, t).
\]

Finally, the probability density that position \( x \) is replicated at time \( t \), irrespective of which origin the fork started from, is

\[
P(x, t) = \sum_{i=1}^{N} P_i(x, t).
\]

One of the most important quantities for comparison with experimental data is the average replication time \( T(x) \) at position \( x \), which is given in terms of \( P \) as

\[
T(x) = \frac{1}{1 - s_1s_2 \cdots s_N} \int_{-\infty}^{+\infty} tP(x, t) dt,
\]

where \( 1 - s_1s_2 \cdots s_N = \int_{-\infty}^{+\infty} P(x, t) dt \) is the probability that at least one of the origins will activate. The average replication time across whole chromosomes \([T(x) \text{ curves}]\) has been measured in a number of organisms. However, caution is required when interpreting \( T(x) \) curves. In some of the biological literature, \( T(x) \) curves are used to directly infer origin parameters [4]. For example, it is widely accepted that the values of \( T \) at \( x_i \) are the average activation times of origins. However, Eq. (3) shows that \( T(x) \) is determined collectively by all origins [8]. This suggests that simple interpretations of \( T(x) \) are not justifiable.

We want to use the general theory presented above to study replication dynamics in a simple setting. From now on we focus on the case of a hypothetical linear chromosome with just two origins. We define the chromosomal coordinates so that one of the origins has position \( x_1 = 0 \); the other origin has position \( x_2 = D \). We assume for simplicity that each origin can activate within a time window \( \Delta t \) with uniform probability; we will argue later that our conclusions are largely independent of the precise shape of the probability distribution. We select origin activation times so that the average activation time of the first origin is 0. The other origin has an average activation time \( \tau \), and we assume without loss of generality that \( \tau \geq 0 \). Thus the activation time distributions are

\[
p_i(t) = \frac{q_i}{\Delta t}, \quad \text{if } t \in \left[ t_i^\Delta - \frac{\Delta t}{2}, t_i^\Delta + \frac{\Delta t}{2} \right],
\]

where \( i = 1, 2 \), \( t_i^\Delta = 0 \), and \( t_i^\Delta = \tau \). \( p_1 \) and \( p_2 \) are set to zero outside the stated intervals.

Using Eqs. (3) and (4), we can write analytical expressions for the probability density \( P(x, t) \) and the average replication time \( T(x) \). From Eq. (4) and (2), \( P(x, t) \) vanishes outside the intervals \( I_1 \) and \( I_2 \) given by

\[
I_1 = \left[ t_i^\Delta - \frac{\Delta t}{2}, \frac{|x_1 - x|}{v} + t_i^\Delta - \frac{\Delta t}{2} \right],
\]

where \( x_1 = 0 \) and \( x_2 = D \). In total there are five scenarios (depending on the relative values of \( \tau \), \( \Delta t \), \( D \), and \( v \)) that differ in the dynamics of how the chromosome is replicated. From here on we will consider just the case where the condition \( \tau + \Delta t < D/v \) is satisfied, since this is the case for many pairs of origins in real chromosomes. This means that the variations in the activation time \( \Delta t \) are small enough that a fork from one origin can only replicate the other origin if that origin is not competent. The expression for \( T(x) \) is then
in fact, the DNA at the origin location will not be replicated for different values of $q_1$. Parameters values: $q_2 = 0.7$, $\tau = 1$, $\Delta t = 4$, $D = 10$, $v = 1$.

A plot of $T(x)$ for different values of $q_1$ is shown in Fig. 1. We see that $T(x)$ has discontinuous derivatives at the origin locations, because the forks originate there. At the origins, the mean replication times are

$$T(x_1) = T(0) = q_2 s_1 (D/v + \tau)/(1 - s_1 s_2);$$

$$T(x_2) = T(D) = (q_1 s_2 D/v + q_2 \tau)/(1 - s_1 s_2).$$

(7)

It is commonly assumed in the replication literature that $T(x)$ has a minimum at an origin, and that the value of this minimum directly gives the average activation time for the origin. However, Eq. (7) shows that this is not the case and, in fact, $T(x_i) \geq t^0_i$; the mean replication time at an origin location is equal to or greater than the origin’s average activation time. Only when an origin has $q_i = 1$ can $T(x_i) = t^0_i$, because if an origin fails to activate in a given cell, the DNA at the origin will not be replicated until a fork from another origin arrives. This means that $T_i$ is higher for origins that are more likely to fail, as seen directly in Fig. 1. Another important conclusion from Eq. (7) is that even when both origins have the same average activation time ($\tau = 0$), generally we have $T(x_1) \neq T(x_2)$. This is again due to the possibility of origins not activating. Therefore, the origin with the lower minimum of $T(x)$ does not necessarily activate earlier than the other origin: minima of $T(x)$ cannot be used to draw conclusions on the relative activation times of the corresponding origins, as previously assumed [4,10]. Equations (1)–(6) show that in general $T(x)$ at any point depends collectively on the parameters of all origins. However, if an origin is highly competent, early activating, and isolated from other origins, $T(x)$ at that origin’s position will be close to the origin’s average activation time.

Equation (6) challenges the assumption that origins are located at minima of $T(x)$. From Eq. (6) the expression for the slope of $T$ near the first origin (for $x > 0$):

$$T'(x) = \frac{q_1 - q_2 s_1}{v(1 - s_1 s_2)}.$$  

(8)

This expression shows that the slope is a function of the competencies $q_i$ of both origins as well as the fork velocity $v$. For the origin at $x = 0$ to be a minimum of $T(x)$, we must have $T' > 0$ for $x > 0$, from which we get the condition $q_1 > q_2 s_1$. This shows that if an origin has low competence compared to its neighbor, it may not be a minimum of $T(x)$, which can be seen in Fig. 1. This phenomenon has been observed in experimental data [4]. Note that if $q_1 > 1/2$, this condition is always satisfied and a minimum is guaranteed for this two-origin system. In addition, Eq. (8) shows that the fork velocity is not given by the slope of $T(x)$, an assumption widely used in the literature [4].

Figure 1 shows that $T(x)$ has sharp corners at origin locations. The reason for this is that in every cell the forks always start at the same locations (the origins), which causes a discontinuous change in the proportion of left-propagating compared to right-propagating replication forks, which in turn causes the discontinuity in the derivative of $T(x)$ at the origins. In contrast, Fig. 1 shows that the local maximum of $T(x)$ between two origins is a smooth curve. The reason is that in different cells in a population forks meet each other and terminate at different locations on the DNA, because of the stochastic variations in activation times. This reasoning suggests that the shape of the maxima of $T$ could be used to infer information about the width $\Delta t$ of the activation time distribution. We expect that sharp maxima should correspond to forks meeting within a narrow time window, and consequently a small value of $\Delta t$; conversely, a broad maximum corresponds to a high $\Delta t$. This can be seen in Fig. 2, where $T(x)$ is plotted for various $\Delta t$.

FIG. 1 (color online). Replication time curves for differing values of competence $q_1$. Parameters values: $q_2 = 0.7$, $\tau = 1$, $\Delta t = 4$, $D = 10$, $v = 1$.

FIG. 2 (color online). Replication time curves for different widths of the activation time window.

068103-3
values of $\Delta t$. In order to investigate this more quantitatively, we use the modulus of the second derivative of $T(x)$ at the maxima to measure how broad the maxima is—low values of $|T''|$ correspond to broad peaks. We now use Eq. (6) to find the relationship between $|T''|$ and the origin parameters. At the maximum of $T(x)$, we find

$$|T''| = \frac{4q_1q_2\sqrt{1 - \frac{1}{q_2} - \frac{1}{q_1}}}{v^2\Delta t(1 - s_1s_2)}.$$  \hspace{1cm} (9)

Thus $|T''|$ is inversely proportional to $\Delta t$. Notice also that $|T''|$ does not depend on $\tau$, which means it is independent of the origins’ average activation times. This expression can be used to calculate $\Delta t$ from an experimental replication time profile $T(x)$, if the origin competencies and the fork velocity are known. This is a very useful result because it allows the determination of a quantity characterizing stochastic properties of the system $\Delta t$ from $T(x)$, which is defined by a population average. This is valuable because experiments to directly measure $\Delta t$ are technically difficult, and there are few results available [5,6]. We note that this does not require assuming that all cells in the population are synchronized, since in each individual cell in an asynchronous population, the statistics of the relative activation times of origins remain unaltered [8].

Equation (9) was obtained using a simple uniform distribution for $p_i(t)$. However, we expect it to be a good approximation for any single-peaked distribution function $p_i(t)$, since Eq. (9) only involves the second moment (the variance) of the distribution, and the replication dynamics are mostly determined by the average activation time and the width of the activation distribution—the first and second moments of $p_i(t)$. To test this assumption we used Eqs. (1)–(6) to numerically compute $T(x)$ for pairs of origins with either a Gaussian or a skewed distribution that lead to sigmoidal cumulative distributions [14]. Choosing parameters such that all these distributions have the same mean and variance, we find that in all cases $T''$ never differs between distributions by more than 10%.

Despite the fact that we have been considering a hypothetical two-origin chromosome, we expect Eq. (6) to be a good approximation for chromosomes with many origins when two neighboring origins are relatively isolated from other origins. To test this, we looked at experimental data [4] for $S.\text{cerevisiae}$ chromosome X, specifically the region containing origins ARS1012/13 and ARS1014 (Fig. 3). The smoothness of the curve—ignoring the fluctuations caused by experimental noise—is direct evidence for stochastic origin activation, in agreement with other results [5,6]. We fitted a parabola through the data points and from this determined the value of $|T''|$. Using Eq. (9) we estimate the value of $\Delta t$ as 10.8 min [19]. This value is in agreement with the limited number of single cell measurements that have been made at other $S.\text{cerevisiae}$ origins [6].

We thank M. Hawkins for valuable discussions. This work has been supported through the Biotechnology and Biological Sciences Research Council (Grants No. BB/E023754/1, No. BB/G001596/1, and No. BB-G010722).

[19] Parameter values used were $q_1 = 0.74$, $q_2 = 0.78$ and $v = 1.9$ kb/min, estimated in [8].