Automatica 47 (2011) 1156–1164

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Modeling and analysis of DNA replication☆

Konstantinos Koutroumpas, John Lygeros *

Automatic Control Laboratory, Department of Information Technology and Electrical Engineering, ETH Zürich, 8092, Switzerland

ARTICLE INFO

Article history:
Received 1 February 2010
Received in revised form
11 August 2010
Accepted 10 January 2011
Available online 31 March 2011

Keywords:
DNA replication
Modeling
Piecewise deterministic Markov process
Systems biology

ABSTRACT

DNA replication is an important process in the life of a cell. It has to be completed with extreme accuracy in a specific phase of the cell cycle, known as the S phase. Eukaryotic DNA replication is a rather complex and uncertain process. Several mathematical models have been recently proposed in the literature to interpret experimental data from various organisms. A common concern of many of these models is the so-called random gap problem, the observation that eukaryotic DNA replication should last longer than experimental evidence suggests due to its stochastic nature. One of the biological hypotheses proposed for resolving the random gap problem postulates the presence of a limiting factor regulating the rate with which DNA replication initiates. We show how this hypothesis can be captured in the Piecewise Deterministic Markov Process modeling framework. Monte Carlo simulations allow us to analyze the proposed model and compare model predictions with independent experimental data.

1. Introduction

Mathematical modeling of biochemical processes has attracted considerable attention in recent years. Development of Mathematical models that describe gene and protein interactions in a precise and unambiguous manner is recognized as one of the major challenges facing the biology research community today. Such mathematical models allow computer-based simulation and analysis of biochemical processes that can be used for rapid verification or falsification of biological hypotheses, replacing in certain cases costly and time-consuming *in vitro* or *in vivo* experiments. Moreover, *in silico*, *in vitro* and *in vivo* experiments can be used together in a feedback arrangement: mathematical model predictions can assist in the design of *in vitro* and *in vivo* experiments, the results of which can in turn be used to improve the fidelity of the mathematical models.

Here we concentrate on DNA replication, the process of duplication of a cell's genetic material, one of the most fundamental processes in the life of every cell. While the replication process is rather simple in bacteria and viruses, eukaryotic DNA replication is characterized by a higher degree of stochasticity. In eukaryotes replication can initiate at multiple sites along the genome, known as the origins of replication, that are activated with varying efficiencies and at different times. The need to improve our understanding of the process has led to the development of mathematical and computational models (Blow & Ge, 2009; Brummer, Salazar, & Zinzalla, 2010; de Moura, Retkute, & Hawkins, 2010; Gauthier & Bechhoefer, 2009; Gauthier, Herrick, & Bechhoefer, 2010; Goldar, Labit, Marheineke, & Hyrien, 2008; Goldar, Marsolier-Kergoat, & Hyrien, 2009; Herrick, Jun, Bechhoefer, & Bensimon, 2002; Lygeros et al., 2008; Spiesser, Klipp, & Barberis, 2009; Yang, Rhind, & Bechhoefer, 2010) (see Hyrien & Goldar, 2009, Rhind, Yang, & Bechhoefer, 2009, Herrick, 2010 for overviews) that capture the mechanisms of replication for different organisms. In previous work (Kouretas, Koutroumpas, Lygeros, Lygerou, & Lygeros, 2006) authors and co-workers had developed a mathematical model for DNA replication based on the Piecewise Deterministic Markov Process framework. Model predictions (Lygeros et al., 2008), in accordance with experimental studies Patel, Arcangioli, Baker, Bensimon, and Rhind (2006), suggest that stochastic origin firing can result in a duration of replication considerably longer than previously believed, due to randomly generated large inter-origin gaps (random gap problem). This in turn suggests the existence of an (as of yet unidentified) biological mechanism for alleviating this problem.

One of the biological hypotheses proposed for this purpose postulates the presence of a limiting factor that increases the activation probability of unreplicated origins of replication during the replication process. In Lygeros et al. (2008) this hypothesis was tested *in silico* and results indicated that the proposed firing propensity redistribution mechanism offers a possible solution to the random gap problem.

In this paper we revisit the firing propensity redistribution hypothesis focusing on two possible alternatives for the redistribution mechanism. We show how the previously developed model

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Graziano Chesi under the direction of Guest Editor Francis J. Doyle III.

* Corresponding author. Tel.: +41 44 6328970; fax: +41 44 6321211.
E-mail addresses: koutroumpas@control.ee.ethz.ch (K. Koutroumpas), lygeros@control.ee.ethz.ch (J. Lygeros).

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.02.007

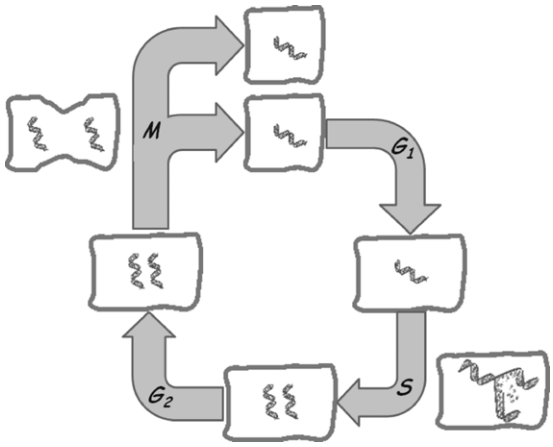

Fig. 1. The cell cycle.

(Kouretas et al., 2006) can be extended to encode these alternatives in the framework of the Piecewise Deterministic Markov Process. Model formalization allows us to extend standard simulation algorithms to analyze the models *in silico* and lends theoretical support to the results. Monte Carlo simulations are used to perform predictions and statistically analyze the properties of the two alternatives. The simulation study in Lygeros et al. (2008) showed that firing propensity redistribution may indeed provide an explanation for the random gap problem. To strengthen this conclusion we show here that the firing propensity redistribution models match better the experimental data than the model without redistribution and are also compatible with other experimental observations. Simulation results show that firing propensity redistribution may lead to the appearance of temporal regulation of origin firing without the need to assume a dedicated mechanism for this purpose (Barberis, Spiesser, & Klipp, 2010; Yang et al., 2010). This allows us to draw novel conclusions about the mechanisms that may regulate temporal origin firing (or the lack thereof).

The rest of the paper is organized as follows. We start with a short biological description of the DNA replication and the random gap problem, Section 2. Section 3.1 outlines the stochastic hybrid model for the DNA replication process. In Section 3.2 we give a brief overview of the PDMP modeling framework that is used later for the DNA replication model, Section 3.3. Then, simulation results, followed by a discussion, are summarized in Section 4 and concluding remarks are drawn in Section 5.

2. Biological background

2.1. DNA replication in the Cell Cycle

Cell reproduction occurs by an ordered sequence of events, known as the cell cycle. During a cycle, all of the cell's components are duplicated and then distributed into two distinct cells, the "daughter cells". The cell cycle comprises four phases: G_1, a cell growth (gap) phase; S (synthesis) phase, when DNA is replicated; G_2, a second gap phase; and an M (mitosis) phase, in which cell division takes place (Morgan, 2007) (Fig. 1). Given that daughter cells have to be genetically identical, cells must duplicate their DNA with extraordinary accuracy before cell division. Incorrect replication of even a small part of the genome would disrupt proper segregation of the genetic material to the two daughter cells during mitosis, leading to genomic instability. DNA re-replication, a situation where a cell makes more copies of some of its genetic material, is associated with cancer genesis.

The process of DNA replication begins with initiator proteins that bind and unwind the double-stranded DNA, breaking the hydrogen bonds between the bases. The positions at which the DNA helix is first opened are called replication origins. In bacteria replication begins at a single origin and continues bidirectionally. In eukaryotes, because of the large size of most eukaryotic chromosomes, replication initiates from multiple points, allowing the replication of the whole genome in a timely manner. For most eukaryotes active origin selection is not deterministic (Dai, Chuang, & Kelly, 2005; Heichinger, Penkett, Bahler, & Nurse, 2006; Patel et al., 2006; Rhind et al., 2009): a specific origin will fire in some but not all cell cycles. Moreover, even if an origin does fire, the time when it will do so is still uncertain and differs from cell to cell. When an origin fires the pre-replicative complex is disassembled, prohibiting origin re-firing, and DNA is opened creating two Y-shaped DNA structures, called replication forks. At the replication fork a multi-enzyme complex synthesizes the DNA of both new daughter strands. Replication forks are formed in pairs and create a replication bubble as they move in opposite directions away from a common origin, stopping when they meet a replication fork moving on the opposite direction (fork conversion) or when they reach a chromosome end.

2.2. The random gap problem

Recent single cell experiments (Patel et al., 2006) have shown that random selection and activation of origins of replication results in an exponential distribution of distances between origins. This suggest that with high probability some large distances between active origins would be obtained in every cell cycle. Since the completion time of DNA replication is determined by the largest of these gaps and the limited replication speed, the total replication time would be very long. The observation that because of the randomly distributed origin activation DNA replication should last much longer than what conventional experimental wisdom would suggest (Hyrien, Marheineke, & Goldar, 2003; Laskey, 1985; Lucas, Chevrier-Miller, Sogo, & Hyrien, 2000; Lygeros et al., 2008) is known as the random completion or random gap problem. The mismatch between modeling predictions and experimental evidence exposed by the random gap problem is very important from a biological point of view as it suggests the existence of an (as of yet unknown) mechanism that ensures the replication of DNA in a timely manner despite the stochastic activation of origins.

Different biological mechanisms have been proposed to explain the random gap problem. According to one of them DNA replication may last longer than believed and extend into the G_2 phase (Lygeros et al., 2008). Such late DNA replication may have gone experimentally undetected to date due to the inherent limitations of experimental procedures and the fact that the stretches of late replicating DNA lie at different locations in each cell cycle. An alternative explanation is that there is a mechanism that evenly distributes origin firing across the genome. For instance, origins within specific clusters could be selected to fire, or active origins could suppress their neighbors by lateral inhibition (Mesner, Li, Dijkwel, & Hamlin, 2003; Shechter & Gautier, 2005). In such a case, the distribution of inter-bubble distances, distances between the replication bubbles, would be non exponential (Patel et al., 2006), contrary to experimental results based on single cell techniques.

An alternative solution is that origins fire stochastically, but their firing efficiency increases during S-phase (Rhind et al., 2009). This assumes that origins in unreplicated regions will fire with increasing probability as S-phase proceeds (Goldar et al., 2009; Herrick et al., 2002; Lucas et al., 2000). Increase of the firing probability can be brought about by several mechanisms, see Rhind et al. (2009) for a recent review. One of these mechanisms assumes