An *in silico* mathematical model of the initiation of DNA replication

Rohan D. Gidvani[a1], Brendan J. McConkey[a], Bernard P. Duncker[a] and Brian P. Ingalls[a,b]

[a]Department of Biology, University of Waterloo, 200 University Ave. W, Waterloo, Canada; [b]Department of Applied Mathematics, University of Waterloo, 200 University Ave. W, Waterloo, Canada

ABSTRACT

Proper eukaryotic cell proliferation depends upon DNA replication, a closely regulated process mediated by the actions of a multitude of factors. The initiation of replication is regulated by the heterohexameric Origin Recognition Complex (ORC). At origins of replication, ORC recruits and/or associates with protein factors such as Cdt1, Cdc6, the MCM2-7 complex, Cdc45 and the Dbf4-Cdc7 kinase. The mechanisms controlling these associations are well documented, allowing the development of a mathematical model that allows us to explore the network's behaviour. Using budding yeast as a model organism, we have developed an ordinary differential equation (ODE)-based model of the protein-protein interaction network regulating replication initiation. Precise quantification of protein factors at various timepoints is critical to calibration of the model parameters. To this end, we have made use of genetic manipulations and quantitative protein expression analysis. Using chromatin extracts from synchronized cell cultures, we were able to monitor the fluctuation of a number of the aforementioned proteins. This information was used to infer qualities of the protein network and to calibrate a predictive mathematical model of the process of DNA replication initiation, which can be integrated into existing models of the entire budding yeast cell cycle.

Keywords: DNA replication; origin recognition complex; pre-replicative complex; mathematical modeling, systems biology, cell cycle

1 INTRODUCTION

The machinery of the eukaryotic cell cycle has been extensively dissected and described from simple to complex organisms. Cell proliferation hinges on the ability to replicate the genome with high fidelity, segregate the chromosomes equally and finally divide the cell, resulting in two genetically identical copies. Equally important are the monitoring modules that oversee these pathways and that intervene under unfavourable conditions, such as DNA damage, and trigger the ensuing repair mechanisms. These steps have been extensively characterized and the cycle organized into an approximate pathway of sequential events.

In eukaryotes, a functionally-conserved heterohexameric protein complex – ORC (Origin Recognition Complex) acts as a selector for origins of DNA replication. ORC then serves as a scaffold for the association of a number of additional replication factors, which collectively form the pre-replicative complex (Pre-RC). The protein encoded by the *CDC6* gene is also essential and is required for initiation via its role in loading the heterohexameric MCM (minichromosome maintenance) complex onto origin DNA. The six subunits, Mcm2-7, when formed into an active complex, collectively act as the replicative helicase. Formed in the cytoplasm, the MCM complex is co-transported to the nucleus with Cdt1 and is recruited to the ORC- and Cdc6-bound DNA (reviewed in [1]). This is promoted by the direct interaction between Orc6 and Cdt1 [2]. These steps culminate in the loading of the MCM rings onto DNA, whereupon they unwind the double helix bidirectionally and provide

[1] Corresponding author. Department of Biology, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1. Tel: + 1 519 888 4567 x36551; Email: rdgidvan@uwaterloo.ca

access for the DNA polymerases. This tight loading is dependent on a stepwise ATP-hydrolysis dependent mechanism involving Cdc6 and ORC [3].

Pre-RCs are set up at about 300 of the approximately 500 consensus binding sequences, to define potential origins. Firing of a particular origin is dependent upon the association of another group of proteins resulting in the formation of the Pre-IC (Pre-Initiation Complex). Cdc45, the GINS complex, Sld2, Sld3 and Dpb11 must be recruited to a licensed origin, with Cdc45 being a limiting factor. Crucial to this step is the phosphorylation of a number of these proteins by the cyclin-dependent kinase, Cdk1. This is the common name generally referring to the combination of the protein kinase Cdc28 with a cyclin (in the case of DNA replication, cyclin-B5, the gene product of *CLB5*) thus providing a controlling input required for passage through the cell cycle. Finally a second kinase is required primarily to phosphorylate various MCM subunits, activating the ring and triggering initiation. This protein, Cdc7 is also controlled by a limiting regulator: Dbf4. Together they form a kinase complex commonly referred to as Dbf4-dependent kinase (DDK). Dbf4 expression is constitutive, but its degradation is controlled by the anaphase-promoting complex or APC. These mechanisms are reviewed in [1].

In order to maintain genomic stability and prevent over- or under-replication of the genome, the cell has evolved mechanisms to ensure that replication occurs exactly once per cell cycle. This is paramount to avoiding loss of genome integrity and/or cell viability [4]. The inhibitory effects of CDKs (promoted by the abundance of Clb5) on Pre-RC components in *S. cerevisiae* are well documented, and ultimately manifest as deactivation, degradation or nuclear export of these factors. Thus the cell cycle exhibits a dual-state behaviour. When Clb5 levels and consequently CDK activity is high, Pre-RCs cannot be established. Once an origin fires in a DDK- and CDK-dependent manner, a new Pre-RC cannot be established until the next cell cycle due to the inhibitory effects of various CDKs, whose activities peak at S phase and remain high until the end of mitosis.

The complex yet elegant network of cell cycle proteins is sophisticated enough to warrant an attempt at modeling, both because many of its key steps are known and because of the inherent difficulty in intuitively determining individual protein behaviours and interactions under varying biological circumstances. Not only does our model seek to elucidate the fundamentals of DNA replication initiation in yeast, but it also strives to attain predictive power. Given that many of the replication factors as well as the processes that oversee their functions are highly conserved from budding yeast to humans, the model has the potential to be extrapolated to attend to biomedical questions. As defects in the cell cycle and particularly DNA can give rise to human cancers, a predictive model of the cell cycle is invaluable in designing targeted cancer therapies and in determining potential side effects. An added level of rigor is provided by the relationship between our model and that of the whole yeast cell cycle [5] using levels of key cell cycle determinants as specified by the latter.

2 METHODS AND RESULTS

2.1 Building a Kinetic Model

In deconstructing a biological system, a network must be created to represent the key proteins involved and the salient interactions. Passage through replication can be thought to begin at the point of ORC binding DNA. This represents our first “species”, which is a relevant unit in the pathway whose abundance changes over time. As described, multiple proteins bind sequentially and exert their well-defined function. The network diagram is shown in Figure 1. This allows us to intuitively separate and examine the various steps. By measuring the abundance of a given protein at a particular point in the cycle, we can ascertain the levels of the various species of which it is a component. In addition, by defining the enzymatic reactions linking the multiple species in terms of parameters and protein concentrations, we can monitor flux through the pathway.

2.2 *In silico* Modeling

Our network is justified by information regarding the qualitative nature of the individual reactions. Our goal was to convert this knowledge into a consensus picture of the molecular reactions as defined by a set of nonlinear, ordinary differential equations (ODE). A simulated annealing algorithm was used to calibrate the models behaviour by minimizing the least-squares-error in comparison with experimental data.

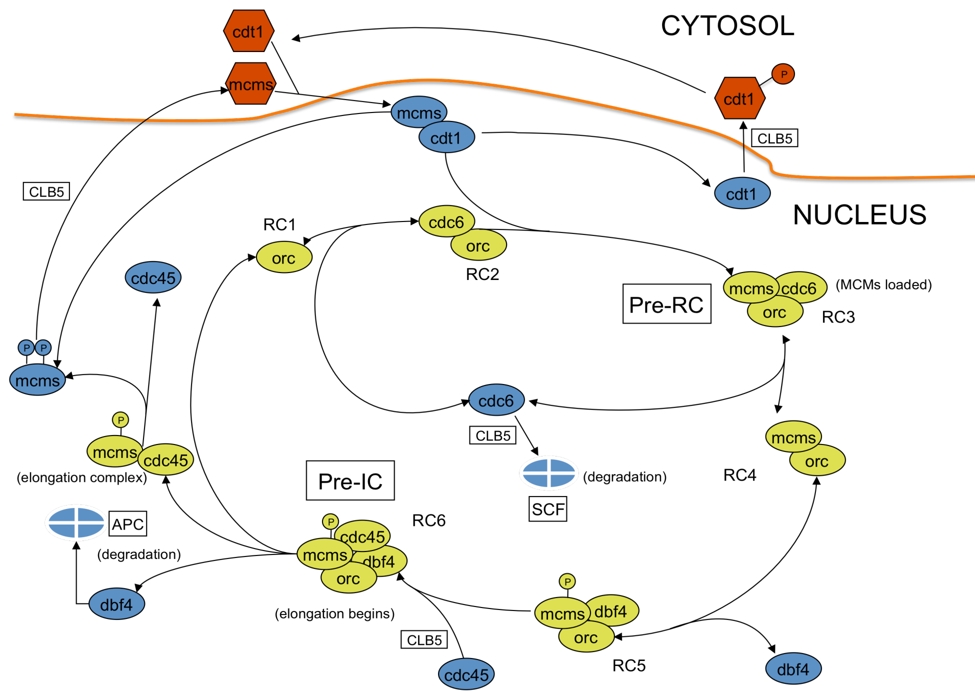

Figure 1. A consensus picture of the network describing the initiation of DNA replication. Reactions are modeled with Mass Action kinetics. The model consists of 11 independent state variables and involves 19 parameters. The levels of APC and CLB5 (representative of CDK activity) are taken as time-varying inputs, with values according to the cell cycle model of Chen et al. [5]. The ODEs are arrived at using the reaction rates shown in Table 1.

2.3 Accumulation of *in vivo data*

In order to determine concentrations of individual proteins implicated in our consensus model we implemented the following methods:

Logarithmically growing asynchronous yeast cultures of a number of strains were arrested in late G1 phase using α-factor. For Cdc6 and Cdc45, myc-epitope tags were incorporated into the open reading frames of these genes and the resultant fusion protein abundances were assayed in separate trials, but by the same method. Cells were released from the G1 block synchronously into the cell cycle. This was confirmed by fluorescence activated cell sorting (FACS). Samples were taken at specific time intervals and were processed by chromatin fractionation to separate proteins bound to chromatin from those that were not. Samples were analyzed by Western blotting, using antibodies directed to the myc-tag or to the protein itself in the case of Mcm2, Dbf4 and Orc2. The amount of a particular protein bound to chromatin (making the assumption that this represents Pre-RC/Pre-IC inclusion) was measured as was the amount unbound. Information about Cdt1 behaviour was obtained from published data [6]. For each protein, at least three trials were performed.

Protein concentrations were determined first by densitometry of Western blots followed by normalization to the number of molecules/cell determined by GFP-tagging experiments described in [8].

2.4 Model Fitting

We used Western blot analysis to measure the abundance of Cdc6, Cdc45, Mcm2 and Dbf4 at eight time points along the cell cycle. This data was compared with the model output and a simulated annealing algorithm was run to minimize the associated least square error associated (normalized by the experimental variance). Figure 2 illustrates the fluctuation of species that comprise the network over time. Figure 3 shows representative examples of the resulting best-fit behaviour.

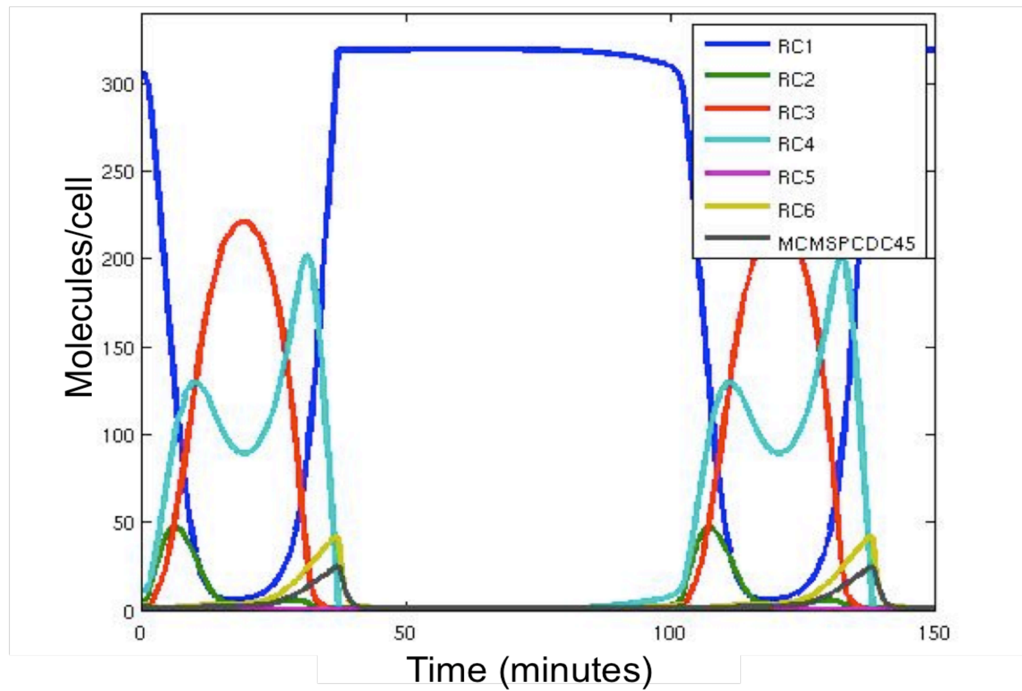

Figure 2. Abundance of network species over time in a 100 minute cell cycle.

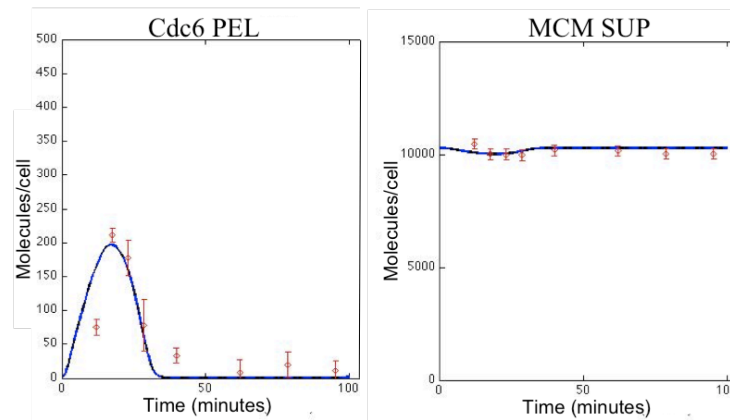

Figure 3. Model simulations of given protein abundances in the indicated cellular fraction (solid line) fit to data points (dots) using a best-fit parameter set. Error bars represent trial-to-trial variation in the data set. The same units used in Figure 1 apply for the respective axes. PEL (pellet) denotes chromatin-bound protein whereas SUP (supernatant) refers to non-chromatin bound protein.

Table 1. Reaction rates. State variables corresponding to the species in Figure 1 are indicated in uppercase.

Description	**Rate reaction**
Expression of CDC6:	k_1
Degradation of CDC6:	k_2.CLB5.CDC6
Expression of MCMs:	k_3
Degradation of MCMs:	k_4.MCM SUP
Expression of DBF4:	k_5
Degradation of DBF4:	k_6.DBF4.APC
Expression of CDC45:	k_7
Degradation of CDC45:	k_8.CDC45
Formation of the Pre-RC	
Association of ORC and CDC6:	k_9.RC1.CDC6

Association and nuclear import of MCMS and CDT1:	k_10.MCM SUP.CDT1 SUP
Loading of MCMS by CDT1:	k_11.RC2.MCM.CDT1
Nuclear export of CDT1:	k_12.CLB5.CDT1
Dissociation/re-association of nuclear MCM-CDT1 complex:	k_13.MCM.CDT1 - k_14.CDT1.MCM PEL
Formation of the initiation complex	
Dissociation/re-association of CDC6 from the pre-RC:	k_14.RC3 - k_15.CDC6.RC4
Association/dissassociaton of DBF4 and the pre-RC:	k_16.RC4.DBF4 - k_17.RC5
Association of CDC45 and the pre-RC:	k_18.RC5.CDC45.CLB5

3 DISCUSSION

Our consensus model has produced a high level of matching with the experimental data, suggesting that it represents a good estimation of the network behaviour. In addition to validation against wild-type protein levels, we also compared the model's behaviour to reports of system behaviour under various perturbations (gene knockdowns, shutoffs, over-expressions). Based on the observation that Dbf4 is degraded by the APC and that this process is one of the redundant mechanisms that prevents re-firing of origins, we simulated Dbf4 as being refractory to degradation and observed an increase in the flux through the fork initiation portion of the cycle. Tanaka and Diffley [6] observed co-transport of Cdt1 and MCMs into the nucleus and we mimicked their experiment which abolished transport of either by preventing the other from entering the nucleus. The abundance of the ORC-Cdc6-MCM complex was reduced drastically in simulations. Finally, given that a major driving force of the network is the initial Cdc6-DNA-ORC binding event, increasing its abundance would be expected to have a profound effect on origin firing. Notwithstanding, re-replication is prevented by degradation of Cdc6 by CDKs. There is a reciprocal relationship in that Cdc6 inhibits CDKs themselves (reviewed in [7]). By increasing the amount of Cdc6 by lowering CDK levels (which additionally maintain a high level of MCMs in the nucleus) in our model we were able to produce a cycle in which origins were firing continuously (i.e. re-replication) indicated by the presence of Pre-RCs and replication forks simultaneously. In addition to results from the literature, we are also undertaking our own *in vivo* perturbation experiments to provide further validation of the model. This will allow greater predictive power in using the model to investigate wild-type and disrupted cell-cycle behaviour.

4 REFERENCES

1. Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333-374 (2002).

2. Chen, S., de Vries, M. A. & Bell, S. P. Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev.* **21**, 2897-2907 (2007).

3. Randell, J. C., Bowers, J. L., Rodriguez, H. K. & Bell, S. P. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol. Cell* **21**, 29-39 (2006).

4. Green, B. M. & Li, J. J. Loss of rereplication control in Saccharomyces cerevisiae results in extensive DNA damage. *Mol. Biol. Cell* **16**, 421-432 (2005).

5. Chen, K. C. *et al*. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* **11**, 369-391 (2000).

6. Tanaka, S. & Diffley, J. F. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat. Cell Biol.* **4**, 198-207 (2002).

7. Honey, S. & Futcher, B. Roles of the CDK phosphorylation sites of yeast Cdc6 in chromatin binding and rereplication. *Mol. Biol. Cell* **18**, 1324-1336 (2007).

8. Huh, W. K. *et al*. Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691 (2003).