

On Coupling Models using Model-Checking: Effects of Irinotecan Injections on the Mammalian Cell Cycle

Elisabetta De Maria, François Fages, and Sylvain Soliman

Project-team Contraintes, INRIA Paris-Rocquencourt, France

Abstract. In systems biology, the number of models of cellular processes increases rapidly, but re-using models in different contexts or for different questions remains a challenging issue. In this paper, we show how the validation of a coupled model and the optimization of its parameters with respect to biological properties formalized in temporal logics, can be done automatically by model-checking. More specifically, we illustrate this approach with the coupling of existing models of the mammalian cell cycle, the p53-based DNA-damage repair network, and irinotecan metabolism, with respect to the biological properties of this anticancer drug.

1 Introduction

In systems biology, the number of models of cellular processes increases rapidly. To date, most of the effort has been devoted to building models and making them freely available, through the design of standard exchange formats, such as for instance the Systems Markup Language SBML [26], the making of model repositories, such as for instance <http://biomodels.net/>, the making of biological ontologies to establish the links between molecular synonyms, species, units, etc., and the development of modeling tools, such as Cell Designer, BIOCHAM [7], BioNetGen [4], Pathway Logic [15], Bio-ambients [35], etc. Despite these efforts however, re-using models in different contexts or for different questions remains a challenging issue. In practice, most of the models are developed, refined, simplified or coupled with respect to other models by hand with no direct support from the tools to re-use models in a systematic way using a specification of the expected behavior.

In this paper, we show how the validation of a coupled model and the optimization of its parameters with respect to biological properties formalized in temporal logics, can be done automatically by model-checking. More specifically, we illustrate this approach with the coupling of existing models of the mammalian cell cycle, the p53-based DNA-damage repair network, and irinotecan metabolism, with respect to the biological properties of the latter one.

Irinotecan is an anti-carcinogenic inhibitor of topoisomerase-1 which started to be used in clinical treatments approximately twenty years ago. It shows significant efficacy against a variety of solid tumors, including lung, colorectal, and

cervical cancers. Scientists are currently trying to optimize the irinotecan therapy in order to understand how to limit its toxicity and to increase its efficacy. In this context, it is crucial to comprehend how the presence of this medication influences cellular proliferation. In this work we present a model-checking approach to the problem. There are in the literature many models of the mammalian cell-cycle, a few ones of the cell's DNA-damage repair pathways, and recently some preliminary models of irinotecan intracellular pharmacodynamics. However these modules need to be assembled in a coherent way to provide meaningful answers. After reformulating the existing models in the rule-based language of the biochemical abstract machine Biocham [19,7], we assemble them into a coupled model under the constraint of satisfying some relevant biological properties, formalized in temporal logic and automatically checked by model-checking.

Model-checking is the process of algorithmically verifying whether a given structure is a model for a given temporal logic formula [13]. In literature, there are various applications of model-checking techniques to biology. In [8,15], temporal logic was introduced as a query language for biochemical networks and for validating boolean models of biological processes by model-checking techniques. Some experimental results were obtained on Kohn's map of the mammalian cell cycle control (800 reaction rules, 500 variables) using the symbolic model-checker NuSMV, and on a small ordinary differential equation (ODE) model using the constraint-based model checker DMC. This approach to verifying biological processes has pushed the development of model-checking techniques for quantitative properties, and continuous, stochastic or hybrid models.

For (non-linear) ODE models, numerical integration techniques provide numerical traces on which linear time temporal logic with numerical constraints can be evaluated by model-checking. Simpathica [2] and the biochemical abstract machine Biocham [6] are two examples of computational tools integrating such model-checkers for quantitative models. This approach has been generalized to temporal logic constraint solving in [18], allowing for efficient kinetic parameter optimization [37] and robustness analyses [36] w.r.t. quantitative temporal logic properties.

In [25], Heath et al. apply the probabilistic model-checker PRISM to the study of a complex biological system, namely, the Fibroblast Growth Factor (FGF) signalling pathway. In [12], Clarke et al. apply statistical model-checking on a stochastic model of a T-cell receptor.

In [3] Batt et al. develop a modeling framework based on differential equations to analyze genetic regulatory networks with parameter uncertainty. The values of uncertain parameters are given in terms of intervals and dynamical properties of the networks are expressed in temporal logic. Model-checking techniques are then exploited to prove that, for every possible parameter value, the modeled systems satisfy the expected properties and to find valid subsets of a given set of parameter values (such an approach is exploited in RoVerGeNe, a tool for robust verification of gene networks). In [33], Piazza et al. propose semi-algebraic hybrid systems as a natural framework for modeling biochemical

networks, taking advantage of the decidability of the model-checking problem for TCTL (Timed Computation Tree Logic) over this large class of systems.

In this paper, we focus on the use of model-checking for integrating biological models. To compose the selected models, we assume a finite set of hypotheses concerning the structure of the “linking reaction rules”, and we search for kinetic parameter values that will make the composite model interact in a proper way. For this, the biological properties of the coupled model are formalized in temporal logic with numerical constraints, and model-checking with parameter optimization techniques are used to find parameter values for the new kinetic rules so that the expected properties are satisfied by the resulting model.

The paper is organized as follows. Section 2 provides the needed biological background on the mammalian cell cycle, on the tumor-suppressor protein p53, and on irinotecan. Section 3 describes Biocham models and explains how such models can be queried in temporal logic. Section 4 presents our coupled model on the effects of irinotecan and Section 5 describes the biological properties of the model that were automatically checked. Finally, in Section 6 we outline some ongoing developments of this work. All the models used are available in BIOCHAM format at http://contraintes.inria.fr/supplementary_material/CMSB09/.

2 Mammalian Cell Cycle, DNA-damage Repair and Irinotecan

2.1 Mammalian Cell Division Cycle

Cells reproduce by duplicating their contents and then dividing in two. To produce a pair of genetically identical daughter cells, the DNA must be faithfully replicated, and the replicated chromosomes must be segregated into two separate cells. The duration of the cell cycle varies greatly from one cell type to another; in a mammalian cell it lasts about 24 hours. The cycle is traditionally divided into the following four distinct phases [1]: the **G1-phase**, that is the temporal gap between the completion of mitosis and the beginning of DNA synthesis, the **S-phase (synthesis)**, that is the period of DNA replication, the **G2-phase**, that is the temporal gap between the end of DNA synthesis and the beginning of mitosis, and the **M-phase (mitosis)**, when replicated DNA molecules are finally separated in two daughter cells.

The cell cycle is regulated by different checkpoints, that are moments when the cell progression is stopped to verify the state of the cell and, if needed, to repair it before damaged DNA is transmitted to progeny cells. DNA damaging agents trigger checkpoints that produce arrest in G1 and G2 stages of the cell cycle. Cells can also arrest in S, which amounts to a prolonged S phase with slowed DNA synthesis. Arrest in G1 allows repair before DNA replication, whereas arrest in G2 allows repair before chromosome separation in mitosis.

The proper alternation between synthesis and mitosis is coordinated by a complicated network that regulates the activity of a family of key proteins. These

proteins are composed of two subunits: a catalytic subunit, the cyclin-dependent kinase, *cdk* for short, and a regulatory subunit, a *cyclin*. A cdk has to associate with a cyclin partner to form a dimer and has to be appropriately phosphorylated in order to be active. The progression through cell cycle is orchestrated by the rise and fall of the Cdk/cyclin dimers.

In this work we refer to the model of mammalian cell division proposed by Novák and Tyson in [30] where the authors present both a set of Ordinary Differential Equations (ODE) and a process diagram to represent the molecular network regulating the mammalian cell cycle. The model comprises 18 differential equations and 4 steady-state relations.

2.2 Protein p53

This subsection is devoted to the description of protein p53, a tumor suppressor protein which is activated in reply to DNA damage. In normal conditions, the concentration of p53 in the nucleus of a cell is feeble: its level is controlled by another protein, Mdm2. These two proteins present a loop of negative regulation. In fact, p53 activates the transcription of Mdm2 while the latter accelerates the degradation of the former. DNA damage increases the degradation rate of Mdm2 so that the control of this protein on p53 becomes weaker and p53 can exercise its functions. This protein is responsible for the activation of many mechanisms: in an indirect way, it stops the DNA synthesis process, it activates the production of proteins charged with DNA reparation, and it can lead to apoptosis, that is, cell death.

When DNA is damaged, Mdm2 loosens its influence on p53 and it is possible to observe some oscillations of p53 and Mdm2 concentrations. The answer to a stronger damage is a bigger number of oscillations. Oscillations have a very regular period. In literature, several models have been proposed to model the oscillatory behaviour of proteins p53 and Mdm2. The most interesting models are undoubtedly the ones proposed by Chickermane et al. [9], by Ciliberto et al. [10], and by Geva-Zatorsky et al. [20]. In this work we build upon the one described in [10], that consists of 6 differential equations.

2.3 Irinotecan

Camptothecins are substances that can be extracted from the Chinese tree “*Camptotheca acuminata* Decne” and are mainly used for the treatment of digestive cancers. Their anticancerogenic properties have been discovered at the end of the Fifties but the first clinical tests have been interrupted owing to heavy effects due to the toxicity of the substances. In the Eighties researchers discovered that camptothecins are inhibitors of topoisomerase-1 (Top1 for short), essential enzyme for DNA synthesis. Afterwards, they started to focus on some semi-synthetic derivative of water-soluble camptothecins, such as irinotecan and topotecan. Irinotecan is pro-medicine and must be transformed in its active metabolite, SN38, to be effectively cytotoxic. In fact the anticancerogenic activity of irinotecan (CPT11) is approximately 100 times less effective than the one

of SN38. The activation is due to carboxylesterase, an enzyme mainly located in the liver, in the intestine, and in the tumoral tissues. SN38 is then detoxified through glucorono-conjugation: this realizes uridine diphosphate glucuronosyl transferase 1A1.

Mechanisms through which irinotecan damages the cell are very complex and have not completely been explained yet. It is sure that DNA lesions appear after the inhibition of Top1 by SN38. Top1 is a protein which is present in all living organisms and which checks DNA replication and transcription. It intervenes to modify the DNA winding degree, acting on one strand. More precisely, Top1 links itself to extremity 3' of DNA forming a transitory cleavage complex and cuts a DNA strand, that in such a way is able to unroll. Then such a complex dissociates and a new ligature comes up. In normal conditions, the connection process is favored with respect to the cleavage one. The target of irinotecan, and above all of its active metabolite SN38, is the complex Top1-DNA. SN38 links to the complex through a covalent bond, preventing in such a way from the ligature of the DNA strand. As clearly written in the title of [34], SN38 acts like a "foot in the door": it keeps opened the DNA strand to which Top1 is linked as to prevent a door from closing. These complexes are still reversible and do not cause DNA lesions. However, they favor them: some lesions can rise as a consequence of the possible collisions with the transcription complexes or with the replication fork. This induces the arrest of the cell cycle. In this case we speak of irreversible complexes. Lesions due to the inhibition of Top1 are therefore consecutive to the stages of the cell metabolism. It means that irinotecan injections must be repeated and abundant in order to be effective. Besides irinotecan is more effective during the DNA replication phase [31,38]. Furthermore, the inhibition of the DNA synthesis takes rapidly place (in a few minutes) and lasts several hours.

Defence answers of cells subjected to irinotecan injections are multiple and vary according to the drug dose. The administration of a very light dose suffices to slow down the S phase of the cell cycle and to delay the G2-M transition. If the dose is more substantial, the lag time in the S phase is much more significant and the cell cycle arrest in the G2-M transition can last more than sixty hours or even be permanent. In this latter case, some genes responsible for the cell cycle arrest (as an example, p21) and involved in the aptototic pathway are over-expressed. These genes are activated by p53, and this suggests the intervention of the protein in reply to a DNA damage due to the dissociation of Top1 from DNA [38].

In this work we refer to a pharmacokinetics/pharmacodinamics (PK/PD) model of irinotecan developed by Dimitrio [14], that takes aim at representing the action of the drug on the body (pharmacodinamic) and the action of the body on the drug (pharmacokinetic), and thus the drug metabolism and its transformations. Such a model is made up of 8 differential equations.

3 The BIOCHAM Abstract Machine

In the last years, one of the main challenges of computational system biology became the creation of powerful simulation, analysis, and reasoning tools for biologists to decipher existing data, devise new experiments, and thus understand functional properties of genomes, proteomes, cells, organs, and organisms. Many of the goals of this emerging discipline have been investigated in the logical setting of temporal logics and have been partly achieved in this approach in the Biochemical Abstract Machine (*BIOCHAM*) [19,7], a formal modeling environment for network biology developed at INRIA-Rocquencourt since 2002.

In this section we briefly recall *BIOCHAM* syntax, its continuous differential semantics and show how a temporal logic with numerical constraints is used to formalize the biological properties of the models and automatically check their satisfaction.

3.1 Syntax and Differential Semantics of *BIOCHAM* Reaction Models

Following SBML and *BIOCHAM* conventions, a model of a biochemical system is formally a set of reaction rules of the form $e \text{ for } S \Rightarrow S'$, where S is a *solution*, that is, a set of molecules given with their stoichiometric coefficient, S' is the transformed solution, and e is a kinetic expression involving the concentrations of molecules. The reaction rules represent biomolecular interactions between chemical or biochemical compounds, ranging from small molecules to proteins and genes.

Reaction rules transform one formal solution into another one. The following abbreviations are used: $A \text{ } \underset{[C]}{\Rightarrow} \text{ } B$ for the catalyzed reaction $A+C \Rightarrow C+B$, and $A \Leftrightarrow B$ for the reversible reaction equivalent to the two symmetrical reactions $A \Rightarrow B$ and $B \Rightarrow A$. The constant $_$ represents the empty solution. It is used for instance in protein *degradation rules*, such as $A \Rightarrow _$, and in *synthesis rules*, such as $_ \underset{[G]}{\Rightarrow} A$ for the synthesis of A by (activated gene) catalyst G . The other main rule schemas are (*de*)*complexation rules*, such as $A + B \Rightarrow A\text{-}B$ for the complexation of A and B , (*de*)*phosphorylation rules*, such as $A \underset{[B]}{\Rightarrow} A\tilde{\{p\}}$ for the phosphorylation of A catalyzed by kinase B , and *transport rules*, such as $A::\text{nucleus} \Rightarrow A::\text{cytoplasm}$ for the transport of A from the nucleus to the cytoplasm.

Reactions can be given kinetic expressions. For instance, $k*[A]*[B]$ for $A\underset{[B]}{\Rightarrow}A\tilde{\{p\}}$ specifies a mass action law kinetics with parameter k for the reaction. Classical kinetics expressions are the following ones:

- the **mass action law kinetics** $k * \prod_{i=1}^n x_i^{l_i}$, which refers to a reaction with n reactants x_i , where l_i is the stoichiometric coefficient of x_i as a reactant;

- the **Michaelis-Menten kinetics** $V_m * x_s / (K_m + x_s)$, which states for an enzymatic reaction of the form $x_s = [x_e] \Rightarrow x_p$, where¹ $V_m = k * (x_e + x_e * x_s / K_m)$;
- the **Hill’s kinetics** $V_m * x_s^n / (K_m^n + x_s^n)$, of which the Michaelis-Menten kinetics is a special case for $n = 1$.

Kinetic expressions can be written either explicitly, allowing any kinetics, or using shortcuts such as **MA(k)** for a Mass Action law with parameter k , or **MM(V_m,K_m)** for a Michaelian kinetics.

A set of reaction rules $\{e_i \text{ for } S_i \Rightarrow S'_i\}_{i=1,\dots,n}$ over molecular concentration variables $\{x_1, \dots, x_m\}$ can be interpreted under different semantics. In this paper we refer to the traditional *differential semantics*, that interprets the rules by the following system of Ordinary Differential Equations (ODE):

$$dx_k/dt = \sum_{i=1}^n r_i(x_k) * e_i - \sum_{j=1}^n l_j(x_k) * e_j,$$

where $r_i(x_k)$ (resp., l_i) is the stoichiometric coefficient of x_k in the right (resp., left) member of rule i . Given a set of reaction rules, BIOCHAM allows to obtain a simulation of the model by solving the set of corresponding differential equations.

3.2 Querying BIOCHAM Models in Temporal Logic

Temporal logics and model-checking algorithms have been proved to be useful to respectively express biological properties of complex biochemical systems and automatically verify their satisfaction [17]. Having a formal language not only for describing models, (i.e., transition systems based on process calculi [35,32], rules [15], Petri nets [22], ODEs [39], etc.), but also for formalizing the biological properties of the system known from biological experiments under various conditions, opens a whole avenue of research for designing automated reasoning tools inspired from circuit and program verification to help the modeler.

In this paper, we use a version of *linear time* logic LTL with numerical constraints, named *Constraint-LTL* [6]. Constraint-LTL formulae are formed over first-order atomic formulae with equality, inequality and arithmetic operators ranging over real values of concentrations and of their derivatives, using the logical connectives and the usual *temporal operators* of LTL: in particular the “always in the future” operator **G**, and the “sometimes in the future” operator **F**, the next time operator **X**, and the binary operator until **U**.

For instance, $F([A] > 10)$ expresses that the concentration of A eventually gets above the threshold value 10 and $G([A] + [B] < [C])$ states that the concentration of C is always greater than the sum of the concentrations of A and B . Oscillation properties, abbreviated as *oscil(M, K)*, are defined as a change of sign

¹ $x_e * x_s / K_m$ is the concentration of the enzyme-substrate complex, supposed constant in the Michaelian approximation, and $x_e + x_e * x_s / K_m$ is thus the total amount of enzyme.

of the derivative of M at least K times: $F((d[M]/dt > 0) \& F((d[M]/dt < 0) \& F((d[M]/dt > 0) \dots)))$. The abbreviated formula $oscil(M, K, V)$ adds the constraint that the maximum concentration of M must be above the threshold V in at least K oscillations.

LTL formulae are interpreted in linear Kripke structures which represent either an experimental data time series or a simulation trace, both completed with loops on terminal states. Given the system of ordinary differential equations (ODE) corresponding to a reaction model, under the hypothesis that the initial state is completely defined, a discrete simulation trace is easily obtained by means of numerical integration methods (such as Runge-Kutta or Rosenbrock method for stiff systems). Since constraints refer not only to concentrations, but also to their derivatives, traces of the form

$$(< t_0, x_0, dx_0/dt, d^2x_0/dt^2 >, < t_1, x_1, dx_1/dt, d^2x_1/dt^2 >, \dots)$$

are considered, where at each time point t_i , the trace associates the concentration values x_i to the variables, and the values of their first and second derivatives dx_i/dt and d^2x_i/dt^2 . It is worth noting that in adaptive step size integration methods of ODE systems, the step size $t_{i+1} - t_i$ is not constant and is determined through an estimation of the error made by the discretization.

Observe that the notion of *next state* refers to the state of the following time point in a discretized trace, and thus does not necessarily imply a real time neighborhood. The rationale is that the numerical trace contains enough relevant points, and in particular those where the derivatives change abruptly, to correctly evaluate temporal logic formulae. An innovative feature of BIOCHAM is that it places at the user's disposal a procedure `learn_parameters` for finding parameter values such that a given LTL formula is satisfied [37]. This search procedure actually replicates and automates part of what the modeler currently does by hand: trying different parameter values, between bounds that are thought reasonable, or computed by other methods such as bifurcation diagrams, in order to obtain behaviors in accordance with the experimental knowledge.

4 Coupling the Models of the Mammalian Cell Cycle, p53-based DNA-damage Repair System, and Irinotecan Pharmacodynamics

4.1 Three Different Modules

The first step of our approach to the investigation of the influence of irinotecan on the mammalian cell cycle consists in the encoding of the selected models of irinotecan [14], p53/Mdm2 [10], and mammalian cell cycle [30] in the BIOCHAM rule-based language.

SBML versions of the irinotecan and p53/Mdm2 modules being available, they were imported in BIOCHAM. The renaming of the variable representing DNA-damage was the only modification necessary in this precise case. More

generally it would be necessary to rely on existing databases and ontologies to match corresponding entities in different models.

For the other model, we looked in parallel at the corresponding set of ordinary differential equations and at the available diagrammatic notation to obtain a set of BIOCHAM reaction rules. Since ODEs can be automatically extracted back from the reactions, and displayed with the BIOCHAM command `show_kinetics`, it was possible to check that the obtained model was indeed coherent with the original one.

4.2 A Diagrammatic Coupled Model

In order to assemble the sub-models to get the coupled model, we reviewed the literature about known links between the different building blocks.

The expected behaviour of the cell is graphically depicted in Figure 1 and can be described as follows. Injections of irinotecan (CPT11) induce DNA damage. In reply to this, the cell reacts by activating protein p53, which blocks the cell cycle at a checkpoint. This arrest aims at repairing critical damage before DNA replications occurs, thereby avoiding the propagation of genetic lesions to progeny cells. Thus, while the cell cycle is arrested, the protein p53 will activate the DNA-damage repair mechanisms. If it is possible for the cell to recover, the cell cycle will be restarted; otherwise, if the damage is too extensive, the cell will undergo apoptosis.

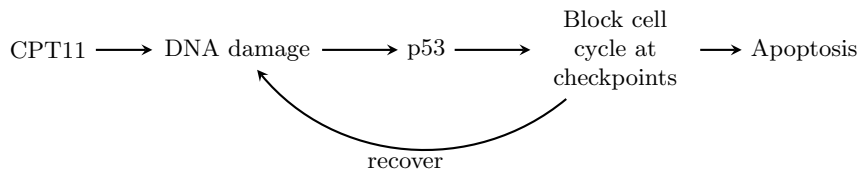


Fig. 1. Expected behaviour of the coupled model.

The strategy we chose to draw the three models together is illustrated in Figure 2. As remarked in Section 2, in literature we found evidence of the fact that, if irinotecan is injected in a cell during the S phase of the cell cycle, then more DNA damage will be caused with respect to the other phases of the cell cycle [31,38]. Keeping this fact in mind, and considering that the dimer characterizing S phase is CycA/Cdk2 (CycA for short), we inserted in cell cycle model a rule stating that a high concentration of CycA determines a high concentration of Top1, an enzyme that, as previously explained, contributes to cause DNA single-strand breaks in presence of irinotecan. In this way we linked the cell cycle model to the irinotecan one.

The link between the irinotecan model and the p53/Mdm2 one is given by DNA damage. In fact, irinotecan injections cause DNA damage, which in turn

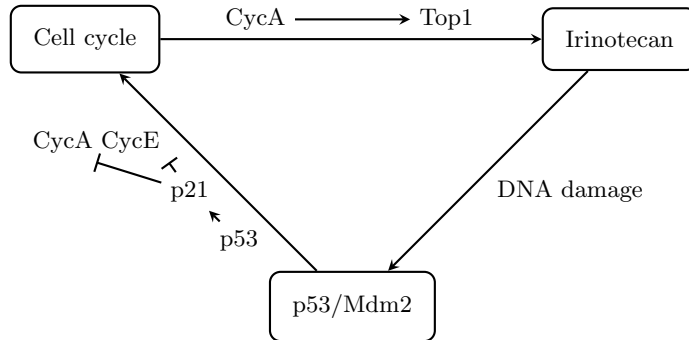


Fig. 2. Our coupled model.

triggers the activity of protein p53. In order to link the p53/Mdm2 and cell cycle models, we inserted in the p53/Mdm2 model a rule which fixes that p53 activates p21, and two further rules imposing that p21 inhibits CycA and CycE, respectively.

It is worth noting that we also investigated the possibility to abstract the previous *expanded* rules by letting p53 directly inhibit CycA and CycE. In the following, we will refer to this last version of the link as to the *contracted* one.

Finally, in order to define precisely the kinetics underlying the links and to validate the coupled model, some biological knowledge was formalized as temporal logic specifications allowing us to use model-checking and parameter optimization techniques.

5 Specifying and Validating the Links through Model-Checking

In this section, we will show how the integration of model-checking and parameter learning techniques allowed us to both specify and validate our linking rules. In both cases, we took advantage of constraint-LTL formulae to query numerical simulations in a much more flexible way than by doing curve fitting. Our Kripke structures are constituted of simulation traces over a time window of 100 hours, obtained by numerical integration (using Rosenbrock’s implicit method for stiff systems), extended with the first and second derivatives of the system’s variables.

As a matter of fact, both at specification and validation time, first we considered two models at a time, and finally we dealt with the complete model made up by the three ones. In the following, we provide examples of biological properties that helped us in defining the kinetics underlying the links: for each property, first of all we express it through the natural language, then we formalize it in constraint-LTL, and finally we present the results of the corresponding model-checking query.

We encoded the link between the cell cycle and irinotecan models (see Figure 2) by means of the following kinetics rule:

MA(top1bis) for $_=[CycA]=>TOP1$.

Then, we used the procedure `learn_parameters` to find out the minimum value for `top1bis` such that property F1, that correlates the concentration values of CycA and Top1, holds.

F1: Whenever CycA gets over 1, there exists a future state where Top1 is greater than 1.

LTL: $G(([CycA] > 1) \rightarrow F([TOP1] > 1))$.

Results: The minimum value for `top1bis` such that F1 is satisfied turned out to be 0.45.

The following Biocham rules encode the link between the p53/Mdm2 and cell cycle models (see Figure 2):

MA(k5321) for $_=[p53]=>p21$.

MA(kA21) for $CycA=[p21]=>_$.

MA(kA21) for $CycE=[p21]=>_$.

As for the contracted version, the encoding is the following one:

MA(kA53) for $CycA=[p53]=>_$.

MA(kA53) for $CycE=[p53]=>_$.

Again, we took advantage of the procedure `learn_parameters` to find suitable parameter values for `k5321` and `kA21` (`kA53` in the second case) so that property F2, that expresses the CycA oscillating behaviour exhibited in the cell cycle model, is conserved in the coupled model in case there are no irinotecan injection.

F2: Within a time interval of 100 time units, CycA is greater than 2 in at least 8 oscillations.

LTL: $oscil([CycA], 8, 2)$.

Results: We found out that suitable parameter values are $k5321 = 0.18$, $kA21 = 0.08$, and $kA53 = 0.25$. Property F2 also turned out to be true when there are injections but the p53/Mdm2 model is not taken into account. In fact, as expected, even if DNA damage occurs, when protein p53 does not act, the cell cycle is not affected, and thus CycA exhibits a regular oscillating behaviour.

On the other hand, when the p53/Mdm2 model is added, in case of repeated injections (and thus of sustained DNA-damage) the oscillations of CycA should be affected.

F3: When there is sustained DNA damage (after an initial period), the amplitude of CycA decreases before 70 time units and then stays low.

LTL: $F((Time < 13) \wedge G([DNAdam] > 0.5)) \rightarrow F((Time < 70) \wedge G([CycA] < 1.2))$.

Results: With the expanded version of the links the amplitude of oscillations gradually decreases, satisfying the property. With the contracted one oscillations are very irregular, as graphically depicted in Figure 3.

The next property regards the DNA repairing power of the cell.

F4: After an irinotecan injection is performed, DNA damage is able to go under the threshold of 0.1 before the next injection is done.

LTL: $G(([CPT11] > 9.45) \vee ((([CPT11] \leq 9.45)U([DNAdam] < 0.1))))$.

Results: Before testing the property, we decided to parameterize the lapse of time between consecutive irinotecan injections. Then we took advantage of the procedure `learn_parameters` to find the minimum k such that, if one injection is performed every k hours, then property F4 is true. We found out that the minimum k multiple of 12 which makes F4 true is 36. Thus, one injection every 36 hours should be performed in order to allow DNA damage to be recovered before the next injection. Then we tried to see what it happens if, at each injection, we double the irinotecan dose. In this case, one injection every 48 hours should be done.

The next property requires the oscillating trend of proteins p53 and Mdm2 to stop before a new injection is performed.

F5: When an injection is performed, p53 and Mdm2 are in a steady state, that is, their derivatives approach 0.

LTL: $G((CPT11 > 9.45) \rightarrow ((d[p53] \leq 0.05) \wedge (d[p53] \geq -0.05) \wedge (d[Mdm2 :: n] \leq 0.05) \wedge (d[Mdm2 :: n] \geq -0.05)))$.

Results: As for the previous specification, we parameterized the lapse of time between consecutive irinotecan injections and we used the procedure `learn_parameters`. The the minimum k multiple of 12 which makes F5 true is 48.

The next property deals with DNA damage.

F6: If p53 is not functional, DNA damage is an increasing function.

LTL: $G([p53] \leq 0.1) \rightarrow G(d([DNA_{dam}])/dt \geq 0)$.

Results: As matter of fact, the premise of this property only holds when the p53/Mdm2 model is not taken into account, thus DNA damage is not recovered and continuously increases until apoptosis is reached.

Finally, the last two properties concern the oscillating trend of proteins p53 and Mdm2 caused by irinotecan injections.

F7: An irinotecan injection causes at least one oscillation of proteins p53 and Mdm2.

LTL: $G((CPT11 > 9.45) \rightarrow F(oscil([p53], 1) \wedge F(oscil([Mdm2], 1))))$.

F8: p53 oscillations are alternated by Mdm2 ones.

LTL: $G(oscil([p53], 1) \rightarrow X(\neg oscil([p53], 1)U(oscil([Mdm2 :: n], 1))))$.

Results: Both F7 and F8 are satisfied in the coupled model.

Such a model-checking approach turned out to be very efficient: the execution times were lower than one second in almost all cases ². Furthermore, it proved to be effective, allowing us to express relevant biological properties of the model (and concentration values that make specifications true) that could not be easily encoded as curve fitting problems for instance. There is a common recognition of the potentialities of model-checking in bio-informatics and our contribution goes in this direction, showing its utility to compose and validate biological models.

In Figure 3 there is a 100 hours-simulation of the complete model, where the contracted version of the link between the p53/Mdm2 and cell cycle models is considered and one injection every 36 hours is performed. The plot puts in evidence how DNA damage increases after every injection. The oscillating trend

² All the experiments have been run on a Centrino Duo 2.00 GHz Windows machine.

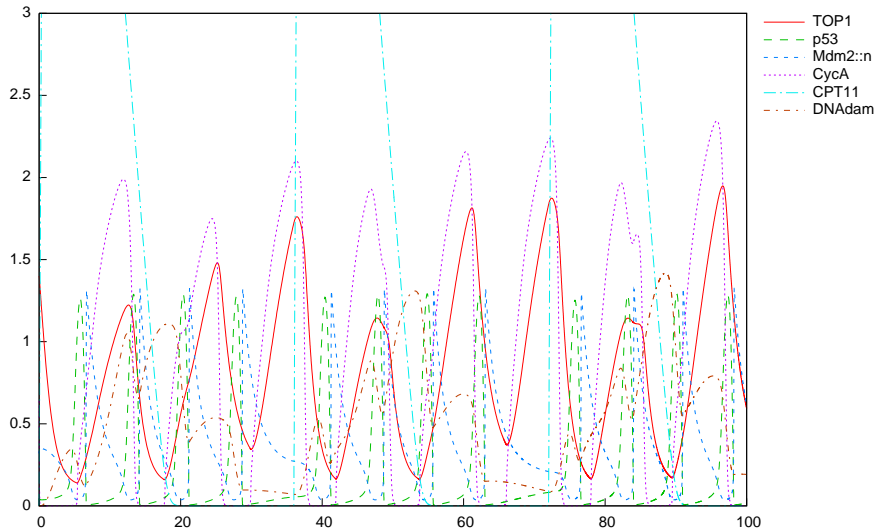


Fig. 3. Zoomed simulation plot of the resulting model with injections every 36 hours.

of proteins p53 and Mdm2 is well highlighted. Furthermore, it is possible to notice the irregular behaviour assumed by CycA after irinotecan injections and the dependence of Top1 from CycA.

6 Conclusion and Perspectives

In this paper we presented a model-checking approach to the investigation of the influence of irinotecan on the mammalian cell cycle. We coupled in the rule-based language of BIOCHAM three models of respectively the mammalian cell cycle, p53-based DNA-damage repair, and irinotecan intracellular PK/PD, using BIOCHAM's procedure for optimizing parameters with respect to biological properties formalized in temporal logic. Model-checking techniques proved to be particularly suitable for studying the irinotecan metabolic pathways. It let us get a better understanding of the drug influence on the mammalian cell cycle and infer some properties to be exploited in the drug therapy, such as optimal injection times and doses.

In order to study how irinotecan interferes with tumor cell proliferation, a further component should be taken into account, that is, the *circadian clock*, which regulates the synchronous progression of cells through each stage of the cell cycle, determining the daily time windows during which cells can traverse certain phases of the cell cycle [28]. The circadian synchronization of cell cycle progression, which characterizes healthy normal tissues, is often altered in cellular tissues affected by malignant tumors [29]. Roughly speaking, it means that,

unlike safe cells, different tumor cells at the same time can be in different phases of the cell cycle.

As a matter of fact, the behaviour of a single cell is not sensibly affected by malignant tumors, but the presence of a disease is generally detected by observing groups of at least one hundred cells. At the level of single cells, a slowing down of biochemical reactions is the only observed phenomenon in presence of tumor. An interesting extension of the work so far presented would consist in adding to our coupled model a new “circadian clock module” to synchronize cells and simulating the resulting model on a group of about one hundred cells. The three already present modules would remain almost unchanged, apart from a speed reduction of the reactions inserted in Figure 2 to link them.

Acknowledgements

This work is partly supported by EU FP6 STREP project TEMPO on cancer chronotherapies. We acknowledge fruitful discussions with the partners of this project.

References

1. B. Alberts, J. Alexander, L. Julian, R. Martin, and R. Keith. *Molecular biology of the cell*. Garland editor, 2008.
2. M. Antoniotti, A. Policriti, N. Ugel, and B. Mishra. Model building and model checking for biochemical processes. *Cell Biochem Biophys*, 38(3):271-86, 2003.
3. G. Batt, C. Belta, and R. Weiss. Temporal Logic Analysis of Gene Networks under Parameter Uncertainty. *IEEE Transactions of Automatic Control*, 53:215-229, 2008.
4. M.L. Blinov, J.R. Faeder, B. Goldstein, and W.S. Hlavacek. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289-3291, 2004. Applications note.
5. M. Calder, V. Vyshemirsky, D. Gilbert, and R. Orton. Analysis of signalling pathways using the continuous time Markov chains. In C.Priami and G.Plotkin,(eds.). *Transactions on Computational Systems Biology VI*, vol. 4220 of LNBI, pp. 44-67. Springer, Heidelberg, 2006.
6. L. Calzone, N. Chabrier-Rivier, F. Fages, and S. Soliman. Machine learning biochemical networks from temporal logic properties. In G.Plotkin, (ed.), *Transactions on Computational Systems Biology VI*, vol. 4220 of LNBI, pp. 68-94. Springer-Verlag, 2006. CMSB’05 Special Issue.
7. L. Calzone, F. Fages, and S. Soliman. BIOCHAM: An Environment for Modeling Biological Systems and Formalizing Experimental Knowledge. *Bioinformatics* 22: 1805-1807, 2006.
8. N. Chabrier and F. Fages. Symbolic model checking of biochemical networks. In *Proc. of the 1st Workshop on Computational Methods in Systems Biology (CMSB)*, C.Priami (ed.), vol. 2602 of LNCS, pp. 149-162, Springer, 2003.
9. V. Chickermane, A. Ray, H.M. Sauro, and A. Nadim. A model for p53 Dynamics Triggered by DNA damage. *Siam Journal on Applied Dynamical Systems*, 6:1, pp.61-78, 2007.

10. A. Ciliberto, B. Novak, and J.J. Tyson. Steady States on Oscillations in the p53/Mdm2 Network. *Cell Cycle*, 4:3, pp.488-493, 2005.
11. A. Cimatti, E. Clarke, F. Giunghiglia, E. Giunghiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchinella. Nusmv 2: An opensource tool for symbolic model checking. In *Proc. of the International Conference on Computer-Aided Verification (CAV' 02)*, vol. 2404 of LNCS, pp. 27-31, 2002.
12. E.M. Clarke, J.R. Faeder, C.J. Langmead, L. Harris, S.K. Jha, and A. Legay. Statistical Model Checking in *Biolab*: Applications to the Automated Analysis of T-Cell Receptor Signaling Pathway. In M.Heiner and A.M.Uhrmacher (eds.), *CMSB'08: Proc. of the sixth international conference on Computational Methods in Systems Biology*, vol. 5307 of LNBI, pp. 231-250. Springer-Verlag, 2008.
13. E.M. Clarke, O. Grumberg, and D.A. Peled. Model Checking. Cambridge, Mass.: MIT Press, 1999.
14. L. Dimitrio. Irinotecan: Modelling intracellular pharmacokinetics and pharmacodynamics. M2 master thesis (in French, English summary), University Pierre-et-Marie-Curie and INRIA internal report, June 2007.
15. S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, and M.K. Sönmez. Pathway logic: Symbolic analysis of biological signaling. In *Proc. of the seventh Pacific Symposium on Biocomputing*, pp. 400-412, 2002.
16. E.A. Emerson. Temporal and modal logic. In *the Handbook of Theoretical Computer Science*, Volume B (chapter 16), pp. 995-1072, J. van Leeuwen (ed.), Elsevier Science Publisher, 1990.
17. F. Fages. Temporal logic constraints in the biochemical abstract machine biocham (invited talk). In P.M.Hill, (ed.), *LOPSTR 2005*, vol. 3901 of LNCS, Springer, Heidelberg, 2006.
18. F. Fages and A. Rizk. From Model-Checking to Temporal Logic Constraint Solving. In *CP'09: Proc. of the fifteenth International Conference on Principles and Practice of Constraint Programming*. To appear in LNCS, Springer-Verlag, 2009.
19. F. Fages, S. Soliman, and N. Chabrier-Rivier. Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM. *Journal of Biological Physics and Chemistry* 4(2), pp.64-73, 2004.
20. N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liton, P. Polak, G. Lahav, and U. Alon. Oscillations and variability in the p53 system. *Molecular System Biology*, 2006.0033, 2006.
21. M.A. Gibson, and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry* 104, 1876- 1889, 2000.
22. D. Gilbert, M. Heiner, and S. Lehrack. A unifying framework for modelling and analysing biochemical pathways using Petri nets. In M.Calder and S.Gilmore, (eds.), *CMSB'07: Proc. of the fifth international conference on Computational Methods in Systems Biology*, vol. 4695 of LNBI, Springer, Heidelberg, 2007.
23. D.T. Gillespie. General method for numerically simulating stochastic time evolution of coupled chemical-reactions. *Journal of Computational Physics* 22, 403-434, 1976.
24. H. Hansson and B. Jonsson. A logic for reasoning about time and reliability. *Formal Aspects of Computing* 6, 512-535, 1994.
25. J. Heath, M. Kwiatkowska, G. Norman, D. Parker and O. Tymchyshyn. Probabilistic model checking of complex biological pathways. In C.Priami (ed.), *Proc. of the 4th Workshop on Computational Methods in Systems Biology (CMSB)*, vol. 4210 of Lecture Notes in Bioinformatics, pp. 32-47, Springer, 2006.

26. M. Hucka, et al. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19:524-531, 2003.
27. M.Z. Kwiatkowska, G. Norman, and D. Parker. Prism 2.0: A tool for probabilistic model checking. In *First International Conference on Quantitative Evaluation of Systems* (QEST 2004), pp. 322-323, IEEE Computer Society, Los Alamitos, 2004.
28. T. Matsuo, S. Yamaguchi, S. Mitsui, A. Emi, F. Shimoda, and H. Okamura. Control mechanism of the circadian clock for timing of cell division in vivo. *Science*, 302:255-259, 2003.
29. M.C. Mormont, and F. Levi. Circadian system alterations during cancer processes. *International Journal of Cancer*, 70:241-247, 1997.
30. B. Novák and J.J. Tyson. A model for restriction point control of the mammalian cell cycle. *Journal of Theoretical Biology* 230, pp.563-579, 2004.
31. S. Ohdo, T. Makinosumi, T. Ishizaki, E. Yukawa, S. Higuchi, S. Nakano, and N. Ogawa. Cell Cycle-Dependent Chronotoxicity of Irinotecan Hydrochloride in Mice. *Journal of Pharmacology and Experimental Therapeutics* 283(3):1383-1388, 1997.
32. A. Phillips and L. Cardelli. A correct abstract machine for the stochastic pi-calculus. In *Proc. of Concurrent Models in Molecular Biology* (Bioconcur'04), affiliated with CONCUR'04, 2004.
33. C. Piazza, M. Antoniotti, V. Mysore, A. Policriti, F. Winkler, and B. Mishra. Algorithmic Algebraic Model Checking I: Challenges from Systems Biology. In *Proc. of the 17th International Conference on Computer Aided Verification (CAV)*, vol. 3576 of LNCS, pp. 5-19, Springer, 2005.
34. Y. Pommier. Camptothecins and Topoisomerase I: A Foot in the Door. Targeting the genome beyond Topoisomerase I with camptothecins and Novel Anti-cancer drugs: importance of DNA Replication, Repair and Cell Cycle Checkpoints. Preprint NIH (NCI), <http://discover.nci.nih.gov/pommier/Pommier.Top1.pdf>, 2004.
35. A. Regev, W. Silverman, and E.Y. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Proc. of the sixth Pacific Symposium of Biocomputing*, pp. 459-470, 2001.
36. A. Rizk, G. Batt, F. Fages, and S. Soliman. A general computational method for robustness analysis with applications to synthetic gene networks. *Bioinformatics*, July 2009. To appear.
37. A. Rizk, G. Batt, F. Fages, and S. Soliman. On a Continuous Degree of Satisfaction of Temporal Logic Formulae with Applications to Systems Biology. In M.Heiner and A.M.Uhrmacher (eds.), *CMSB'08: Proc. of the sixth international conference on Computational Methods in Systems Biology*, vol. 5307 of LNBI, pp.251-268, Springer, 2008.
38. Y. Zhou, F.G. Gwadry, W.C. Reinhold, L.D. Miller, L.H. Smith, U. Scherf, E.T. Liu, K.W. kohn, Y. Pommier, and J.N. Weinstein. Transcriptional Regulation of Mitotic Genes by Camptothecin-induced DNA Damage : Microarray Analysis of Doseand Time-dependent Effects. *Cancer Research* 62, 1668-1695, 2002.
39. D. Zwillinger. Handbook of Differential Equations (3rd edition). Academic Press, Boston, 1997.