Analysis and Model Checking of the HMGB1 Signaling Pathway

Haijun Gong,[1] Paolo Zuliani,[1] Anvesh Komuravelli,[1] James R. Faeder,[2] and Edmund M. Clarke[1]

[1]*Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

Email: {haijung,pzuliani,anvesh,emc}@cs.cmu.edu

[2]*Department of Computational Biology, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Email: faeder@pitt.edu

Recent studies have found that overexpression of the High-mobility group box-1 (HMGB1) protein, in conjunction with its receptors for advanced glycation end products (RAGEs) and toll-like receptors (TLRs), is associated with proliferation in cancers of various types including breast and pancreatic. We have developed a rule-based model of crosstalk between HMGB1 signaling and other key cancer signal pathways. The model has been simulated using both ordinary differential equations (ODEs) and discrete stochastic simulation. Our simulations show that, if HMGB1 is overexpressed, then the oncoproteins CyclinD/E, which regulate cell proliferation, are activated or overexpressed, while tumor suppressor proteins which regulate cell apoptosis (programmed cell death), such as p53, are repressed. The discrete, stochastic simulations show that p53 and MDM2 oscillations continue even after 10 hours. This property is not exhibited by the deterministic ODEs simulation. Moreover, the models also predict that mutation of RAS, ARF and P21's could influence the cancer cell's fate – apoptosis or survival – through the crosstalk of different pathways. Finally, we apply an automated verification technique, Statistical Model Checking, to validate formally interesting temporal properties of our model.

Keywords: Signaling Pathway, Cell Cycle, Stochastic Simulation, Statistical Model Checking

I. INTRODUCTION

The cell cycle is strictly regulated and controlled by a complex network of signaling pathways[1], comprised of hundreds of proteins. If some important proteins are mutated or there are defects in the signaling mechanisms, the normal cell growth regulation will break down, possibly leading to the occurrence of cancer in the future. Moreover, a number of extracellular proteins could bind to their receptors and activate signaling pathways to communicate with the nucleus.

The high-mobility group box-1 (HMGB1) protein is a DNA-binding nuclear protein, released actively in response to cytokine stimulation, or passively during cell death, and it is present in almost all eukaryotic cells[2–5]. HMGB1 can activate a series of signaling components, including mitogen-activated protein kinases (MAPKs) and AKT, which play an important role in tumor growth and inflammation, through binding to different surface receptors, such as RAGE and TLR2/4. Some studies have shown that elevated expression of HMGB1 occurs in many tumors[6–9] and accelerates cell-cycle progression. Recent *in vitro* studies with pancreatic cancer cells[10] observed that the targeted knockout or inhibition of HMGB1 and RAGE could increase apoptosis and suppress pancreatic cancer cell growth. This phenomenon has been also observed with lung cancer and other types of cancer cells[7,11].

To the best of the authors' knowledge, no computational model has been proposed to investigate the importance of HMGB1 in tumor proliferation. In this work, we construct a simple model of HMGB1 signal transduction to investigate tumorigenesis, on the basis of known signaling pathways studies[12–17]. Little is known about HMGB1 at the mechanistic level, so our model can provide some insights into the study of HMGB1's roles in tumor proliferation. A series of simulation experiments was conducted to investigate the properties of the HMGB1 pathway. Furthermore, we used Statistical Model Checking to validate formally our pathway model against known experimental results.

FIG. 1: Schematic view of HMGB1 signal transduction. Blue nodes represent tumor suppressor proteins; red nodes represents oncoproteins/lipids. Solid lines with arrows denote protein transcription, degradation or changes of molecular species; dashed lines with arrows denote activation processes.

II. BIOLOGICAL PROBLEM AND MODEL FORMULATION

A. Biological Problem

HMGB1 can activate a series of signaling pathways, with many proteins involved in signal transductions. We seek to understand which signaling pathways are fundamental to describe HMGB1 signal transduction. Moreover, we aim at identifying the pathways and mechanisms that explain recent results linking overexpression of HMGB1 with decrease of apoptosis (and increased cancer cell survival).

Apoptosis and cell proliferation are regulated respectively by the proteins p53 and CyclinE, acting in two different signaling pathways. The protein p53 is one of the most important tumor suppressor proteins: its activation can lead to the cell cycle arrest, DNA repair or apoptosis. Mutations of p53 occur at a frequency of 50% or higher in many different cancer types[18]. CyclinE is a cell cycle regulatory protein which regulates the G1-S phase transition during cell proliferation. Cancer cells often exhibit high expression levels of CyclinE and aberrant CyclinE activity[19]. Many studies have found evidence of crosstalk between the two signaling pathways involving p53 and CyclinE[20]. The crosstalk is regulated by some tumor suppressor proteins including ARF, P21 and FBXW7, which are also frequently mutated in many cancers. How do these proteins and their mutations change the cell's fate – apoptosis or survival – when HMGB1 signaling transduction is activated? We built a simple model to answer these questions.

B. Structure of the Model

The HMGB1 signaling pathway model is illustrated in Fig.1. It includes 31 molecular species (5 tumor suppressor proteins), 59 chemical reactions, and three different signaling pathways activated by HMGB1: the RAS-ERK, Rb-E2F and p53-MDM2 pathways. Since the interaction between HMGB1 and its receptors TLR and RAGE is not clear at the mechanistic level, RAGE is used to represent all the receptors in our model, in order to reduce the number of unknown parameters. We now briefly discuss the three pathways and their crosstalk. We denote activation (or promotion) by →, while inhibition (or repression) is denoted by ⊣.

p53-MDM2 pathway. It is regulated by a negative feedback loop[22]: $PI3K \rightarrow PIP3 \rightarrow AKT \rightarrow MDM2 \dashv p53 \rightarrow MDM2$, and a positive feedback loop: $p53 \rightarrow PTEN \dashv PIP3 \rightarrow AKT \rightarrow MDM2 \dashv p53$. The protein PI3K is activated by the toll-like receptors (TLR2/4) within several minutes upon TLR2/4 activation by HMGB1[23]. In turn, PI3K phosphorylates the phosphatidylinositol 4,5-bisphosphate (PIP2) to phosphatidylinositol (3,4,5)-trisphosphate (PIP3), leading to phosphorylation of AKT. The unphosphorylated oncoprotein MDM2, which is one of p53's transcription targets[24], resides in the cytoplasm, and cannot enter the nucleus until it is phosphorylated by activated AKT. The phosphorylated MDM2 translocates into the nucleus to bind with p53, inhibiting p53's transcription activity, initializing p53 polyubiquitination[25], which targets it for degradation. Also, p53 can regulate the transcription of PTEN[26], a tumor suppressor protein, which can hydrolyze PIP3 to PIP2, thereby inhibiting the activation of AKT and MDM2.

RAS-ERK pathway: $RAS \rightarrow RAF \rightarrow MEK \rightarrow ERK \rightarrow CyclinD$. Activation of RAGE by HMGB1 leads to RAS activation, which in turn activates its effector protein RAF. Activated RAF will phosphorylate the MEK proteins (mitogen-activated protein kinase kinases (MAPKK)), leading to the phosphorylation of ERK1/2 (also called MAPKs). Activated ERK can phosphorylate some transcription factors which activate the expression of the regulatory protein CyclinD and Myc, enabling progression of the cell cycle through the G1 phase. K-RAS, a member of the RAS protein family, is found to be mutated in over 90% of pancreatic cancers[32].

Rb-E2F pathway: $CyclinD \dashv Rb \dashv E2F \rightarrow CyclinE \dashv Rb$. The Rb-E2F pathway regulates the G1-S phase transition in the cell cycle during cell proliferation. E2F is a transcription factor which can activate the transcription of many proteins involved in DNA replication and cell-cycle progression[27]. In quiescent cells, E2F is bound by unphosphorylated Rb – a tumor suppressor protein – forming an Rb-E2F complex which inhibits E2F's transcription activity. E2F will be activated and released when its inhibitor Rb is phosphorylated by some oncoproteins (CyclinD and Myc in Fig.1), leading to the transcription of CyclinE and Cyclin-dependent protein kinases (CDK2) which promotes cell-cycle progression. CyclinE, in turn, continues to inhibit the activity of Rb, leading to a forward positive feedback loop[28–30]. Fig.1 shows that the activity of CyclinD-CDK4/6(only CyclinD is shown in Fig.1) is inhibited by the tumor suppressor protein INK4A, which is inactivated in up to 90% pancreatic cancers[31].

Crosstalk. The three signaling pathways in HMGB1 signal transduction are not independent, the crosstalk between these pathways can influence the cell's fate. As shown in Fig.1, RAS can also activate the PI3K-AKT signaling pathway; the tumor suppressor ARF protein, activated by the overexpressed oncoprotein E2F, can bind to MDM2 to promote its degradation and stabilize p53's expression level, leading to apoptosis. Moreover, experiments[18] have demonstrated that the p53-dependent tumor suppressor proteins p21 and FBXW7 can restrain the activity of CyclinD/CDK4/6 and CyclinE/CDK2 (only p21 is shown in Fig.1 to represent both p21 and FBXW7's contribution). Mutations of RAS, ARF, P21 and FBXW7 have been found in many cancers[31–33]. Our model and simulation will investigate how these mutations affect the cell's fate.

C. Simulation Models

In Fig.1 we give a schematic view of the signaling pathways and proteins involved in our model of the HMGB1 signal transduction. In the model, all substrates are expressed in the number of molecules; protein with a subscript "a" or "p" correspond respectively to active or phosphorylated form of the protein. For example,

- RAGE (RAGE$_a$) - inactive(active) form of HMGB1's receptor

- MDM2 (MDM2$_p$) - unphosphorylated(phosphorylated) MDM2.

We denote by $mdm2$ the mRNA transcript of MDM2. We assume that the total number of active and inactive forms of RAGE, PI3K, PIP, AKT, RAS, RAF, MEK, ERK molecules is constant. For example, $AKT + AKT_p = AKT_{tot}$, $PIP2 + PIP3 = PIP_{tot}$. We sometimes use CD to stand for CyclinD-CDK4/6 complex, CE for CyclinE, and RE for Rb-E2F complex.

TABLE I: Initial values for the model

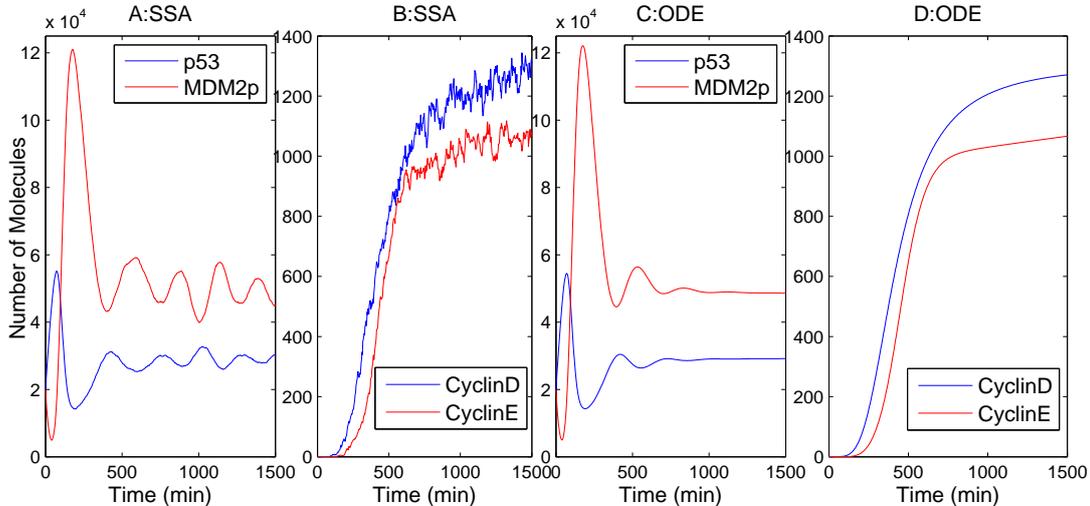| Proteins	$RAGE$	$PI3K$	$PIP2$	AKT	$MDM2$	$MDM2_p$	$P53$	RAS	RAF	MEK	ERK	RE
# of Mol.	10^3	10^5	10^5	10^5	10^4	2×10^4	2×10^4	10^4	10^4	10^4	10^4	10^5

FIG. 2: Number of p53, $MDM2_p$ (A, C), Cyclin D/E (B, D) molecules versus time for baseline simulations with SSA(A-B) and ODE(C-D) models.

The p53-MDM2, Rb-E2F, and RAS-ERK pathways have been investigated individually by many researchers using deterministic ODE methods[12–15,27]. We formulated a reaction model corresponding to the reactions illustrated in Fig.1 in the form of rules specified in the BioNetGen language[34]. We use Hill functions to describe the rate laws governing the transcription of some proteins, including PTEN, MDM2, CyclinD (CD), Myc, E2F and CyclinE (CE), and use mass action rules for other reactions. We use both ODEs and Gillespie's stochastic simulation algorithm (SSA)[35] to simulate the model within BioNetGen[34]. Stochastic simulation is important because when the number of molecules involved in the reactions is small, stochasticity and discretization effects become more prominent[36–38]. A list of ODEs is provided in the Appendix.

III.   SIMULATION RESULTS

We have conducted a series of deterministic and stochastic simulation experiments to study the properties of our HMGB1 signaling model. We first conducted a baseline simulation for four important proteins – p53, MDM2$_p$, CyclinD/E – based on ODE and stochastic simulation models. We set the initial value for the number of HMGB1 molecules to be 10^3; the nonzero initial values for other proteins are given in Table I, and the input parameters are listed in the online supplementary material[55]. The baseline simulations in Fig. 2A demonstrate that the expression levels of p53 and MDM2$_p$ oscillate in the stochastic model even after 10 hours, when the cell enters the S phase. However, no oscillation is observed in the ODE model (Fig.2C) when the cell proceeds to the S phase (we recall that cells usually remain in phase G1 for about 10 hours of the 24 total hours of the cell cycle). The stochastic simulation model is more consistent with Geva-Zatorsky et al.'s experimental result that the oscillations of p53 and MDM2$_p$ expression levels could last more than 72 hours after γ irradiation[39]. Fig.2B and D show that the Cyclin E protein, which regulates the G1-S phase transition in the cell cycle, reaches its maximum at about 10 hours, after which the cell proceeds with DNA replication (S phase).

How does the expression level of HMGB1 and other proteins influence the cell's fate? We varied the levels of HMGB1 and AKT to determine how they affect cell behavior. A number of studies have found that HMGB1 is overexpressed in many cancers, and the overexpression of HMGB1 and its receptors can promote cancer cell proliferation and decrease apoptosis[7,8]. In Fig.3 A-B, we increase the initial values of HMGB1 from 1 to 10^6 and measure p53's maximum expression level in phase G1. We then measure E2F and CyclinD/E's expression levels at 10 hours, which corresponds to the G1-S phase transition point. For the stochastic simulation, the experiment is repeated 10 times per value to compute the mean and standard errors. Fig.3(A,D) demonstrates that the increase of HMGB1 initial value will lead to the decrease of p53's expression level, but when the number of HMGB1 is over 10^5, p53 will not continue to decrease. This is because HMGB1 can also activate and increase the expression level of its downstream protein E2F (Fig.3(A,D)), whose overexpression will activate the transcription of the tumor suppressor protein ARF, which can inhibit MDM2's activity to stabilize p53's level. However, ARF is found to be mutated in up to 80% of pancreatic cancers[31,40]. Fig.3(B,E) shows that the cell cycle regulatory proteins CyclinD/E will increase with the elevated expression of HMGB1. Fig.3(A,B,D,E) explains the experimental discovery that the overexpression of HMGB1 decreases apoptosis and promotes DNA replication and proliferation in cancer cell.

The protein AKT is overexpressed in many cancers[41]. In Fig.3(C,F), we first increase the number of AKT molecules and fix the other proteins' concentration, then measure p53 and $MDM2_p$'s expression levels at 10 hours in phase G1. Fig.3(C,F) shows that with the increase of AKT's expression level, p53 is repressed due to the ubiquitination initiated by the overexpressed $MDM2_p$, which is promoted by the activated and overexpressed AKT protein. The results in Fig.3(C,F) provide a way to inhibit tumor cell proliferation and induce tumor cell apoptosis through the inhibition of protein phosporylation events downstream from AKT kinases in the PI3K/AKT signaling pathway, using some AKT kinase inhibitor (such as the GSK-690693 drug[42]).

K-RAS is a member of the RAS protein family. K-RAS mutation and ARF loss occur in more than 80% of pancreatic cancers[31,40]. The P21 and FBXW7 proteins are also frequently mutated in many cancers[33]. How these mutations influence the HMGB1 signaling transduction, especially ARF and P21, plays an important role in the crosstalk between the p53 and Rb pathways. ARF is able to reroute cells with oncogenic damage to p53-dependent fates through binding to MDM2 and targets its degradation. The p53-dependent tumor suppressor proteins P21 and FBXW7 can inhibit CyclinD/E's activity to prevent the proliferation of cancer cells.

Fig.4 shows how mutation of ARF, P21 and FBXW7, and K-RAS influence tumor suppressor and cell cycle regulatory protein's expression levels at 10 hours. We use the MDM2 degradation rate driven by ARF, d_{ARF} (d'_7 in the ODE model), to describe ARF mutations. Also, we use Cyclin degradation rate driven by P21, d_{P21} (b'_6 in the ODE model), to describe P21 and FBXW7 mutations. Large d_{ARF} and d_{P21} values correspond to small mutation of ARF and P21 respectively, while small d_{ARF} and d_{P21} correspond to large ARF and P21 mutations in the cell.

Fig.4(A,D) show that wild-type ARF (larger d_{ARF}), can decrease the number of $MDM2_p$ molecules and increase p53's expression level to initiate apoptosis even if the cell proceeds to the S phase. Moreover, mutated ARF (smaller d_{ARF}) could not stabilize p53's level and prevent the proliferation of cancer cells if HMGB1 is overexpressed. This could explain the phenomenon that ARF loss exists in over 80% of pancreatic cancers[31]. Fig.4(B,E) demonstrates that CyclinD/E proteins will increase if P21 is mutated (smaller d_{P21}), thereby accelerating cell cycle progression.

K-RAS is mutated in most cancers, especially in pancreatic cancer[32]. Mutated K-RAS can not be deactivated, so it will continuously activate the downstream signaling pathways which promote cell proliferation. Fig.4(C,F) shows that with the increase of RAS deactivation rate d_{RAS} (b_1 in the ODE model), the synthesis of CyclinD/E will be inhibited. The results visualized in Fig.4 provide some ways to inhibit cancer cell proliferation through inhibition or deactivation of the signaling pathways involving RAS, Cyclin, and Cyclin-dependent kinases (CDK). Recently, CDK and RAS inhibitor drugs[43–45] have been developed to inhibit tumor growth.
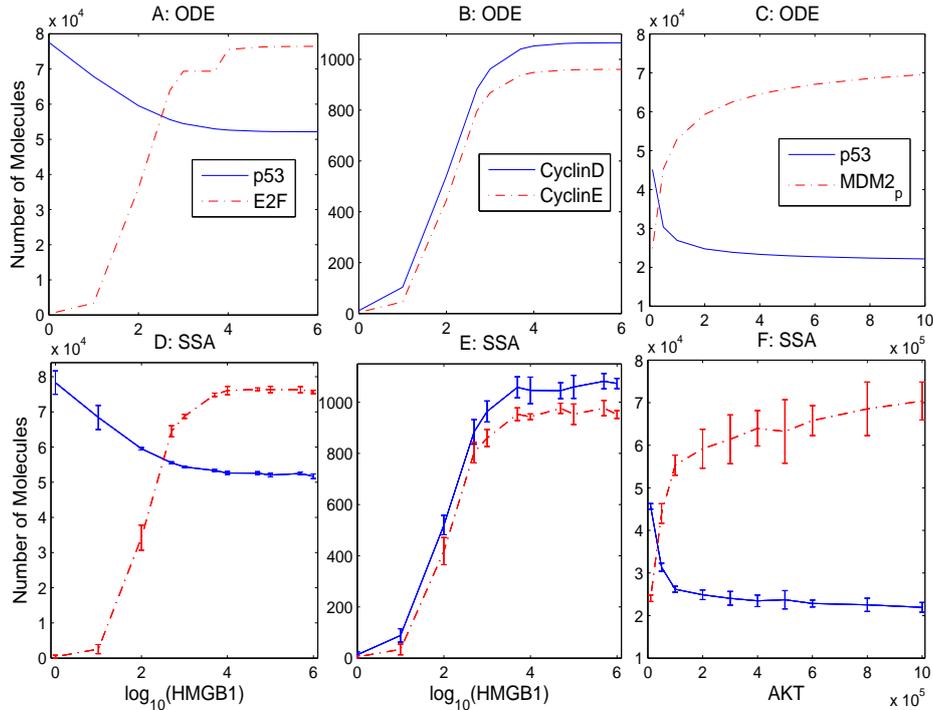
FIG. 3: Overexpression of HMGB1(A-B) leads to the increase of oncoprotein E2F and DNA replication proteins Cyclin D/E, decrease of p53; Overexpression of AKT(C) promote the expression level of $MDM2_p$ and inhibit p53 based on ODE(A-C) and SSA(D-F) models.

IV. MODEL CHECKING

Model Checking[51,52] has emerged as one of the leading techniques for the automated verification and analysis of hardware and software systems. Given a high-level behavior specification, a model checker verifies whether our system (or model) satisfies it. A specification might be satisfied by many different models. Thus, Model Checking is the process of determining whether or not a given system model satisfies (is a model of) a property describing desired behavior of the system. Mathematically, system models take the form of state-transition diagrams, while some version of temporal logic is used to describe the desired properties (specifications) of system executions. A typical property stated in temporal logic is $\mathbf{G}(grant\_req \rightarrow \mathbf{F}\,ack)$, meaning that it is always ($\mathbf{G}$ = globally) true that a grant request eventually ($\mathbf{F}$ = future) triggers an acknowledgment. One important aspect of Model Checking is that it can be performed algorithmically - user intervention is limited to providing a system model and a property to check.

Recently, there has been growing interest in formal verification of stochastic systems, and biological systems in particular[47,50,54], by means of Model Checking techniques. The *Probabilistic* Model Checking problem (PMC) is to decide whether a stochastic model satisfies a temporal logic property with a *probability* greater than or equal to a certain threshold. To express temporal properties we use a logic in which the temporal operators are equipped with *bounds*. For example, the property "CyclinD will always stay below 10 in the next fifty time units" is written as $\mathbf{G}^{50}(CyclinD < 10)$. We now ask whether our stochastic system M satisfies that formula with a probability greater or equal to a fixed threshold (say 0.9), and we write $M \models Pr_{\geqslant 0.9}[\mathbf{G}^{50}(CyclinD < 10)]$. In the next section we formally define the temporal logic used in this work, the Bounded Linear Temporal Logic.
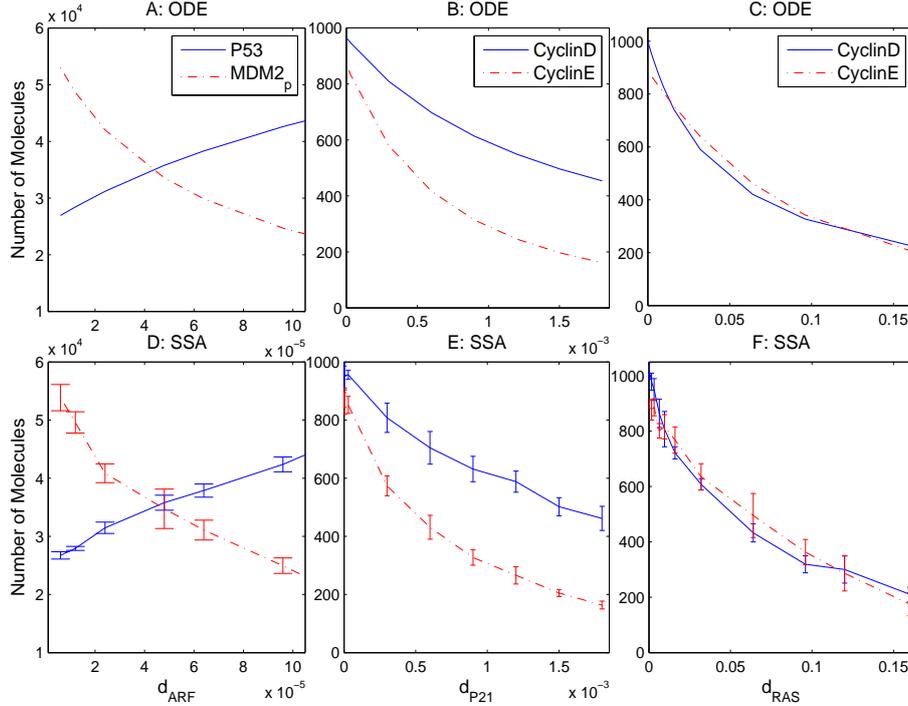
FIG. 4: Mutations of ARF, P21 and RAS affect the cell's fate based on ODE(A-C) and SSA(D-F) models.

A. Bounded Linear Temporal Logic (BLTL)

Let SV be a finite set of real-valued variables, an atomic proposition AP is a boolean predicate of the form $e_1 \sim e_2$, where e_1 and e_2 are arithmethic expressions over variables in SV, and $\sim$ is either $\geq$, $\leq$, or $=$. A BLTL property is built over atomic propositions using Boolean connectives and bounded temporal operators. The syntax of the logic is the following:

$$\phi ::= AP \mid \phi_1 \vee \phi_2 \mid \phi_1 \wedge \phi_2 \mid \neg\phi_1 \mid \phi_1 \mathbf{U}^t \phi_2.$$

The bounded until operator $\phi_1 \mathbf{U}^t \phi_2$ requires that, *within* time t, ϕ_2 will be true and ϕ_1 will hold until then. Bounded versions of the $\mathbf{F}$ and $\mathbf{G}$ operators can be easily defined: $\mathbf{F}^t\phi = true \ \mathbf{U}^t \phi$ requires ϕ to hold true within time t; $\mathbf{G}^t\phi = \neg\mathbf{F}^t\neg\phi$ requires ϕ to hold true up to time t.

The semantics of BLTL is defined with respect to *traces* (or executions) of a system. In our case, a trace will be the output of a simulation of a BioNetGen stochastic model. Formally, a trace is a sequence of time-stamped state transitions of the form $\sigma = (s_0, t_0), (s_1, t_1), ...$, which denotes that the system moved to state s_{i+1} after having sojourned for time t_i in state s_i. The fact that a trace σ satisfies the BLTL property ϕ is written $\sigma \models \phi$. We denote the trace suffix starting at step k by σ^k. We have the following semantics of BLTL:

- $\sigma^k \models AP$ if and only if AP holds true in state s_k;

- $\sigma^k \models \phi_1 \vee \phi_2$ if and only if $\sigma^k \models \phi_1$ or $\sigma^k \models \phi_2$;

- $\sigma^k \models \phi_1 \wedge \phi_2$ if and only if $\sigma^k \models \phi_1$ and $\sigma^k \models \phi_2$;

- $\sigma^k \models \neg\phi_1$ if and only if $\sigma^k \models \phi_1$ does not hold;

- $\sigma^k \models \phi_1 \mathbf{U}^t \phi_2$ if and only if there exits $i \in N$ such that, (a)$\sum_{0 \leq l < i} t_{k+l} \leq t$, (b) $\sigma^{k+i} \models \phi_2$ and (c) for each $0 \leq j < i$, $\sigma^{k+j} \models \phi_1$.

Note that the semantics of BLTL is defined over *infinite* traces, while of course any simulation trace must be finite in length. It can be shown that traces of an appropriate (finite) length are sufficient to decide BLTL properties. The interested reader can find details elsewhere[46].

B. Statistical Model Checking

We briefly explain *Statistical* Model Checking[53], the technique we use for verifying BioNetGen models simulated by Gillespie's algorithm. Statistical Model Checking treats the PMC problem as a statistical inference problem, and solves it by randomized sampling of the traces (simulations) from the model. In particular, the PMC problem is naturally phrased as a hypothesis testing problem, *i.e.*, deciding between two hypotheses – $M \models Pr_{\geqslant \theta}[\phi]$ versus $M \models Pr_{< \theta}[\phi]$. In other words, to determine whether M satisfies ϕ with a probability $p \geqslant \theta$, we test the hypothesis $H_0 : p \geqslant \theta$ against $H_1 : p < \theta$. Sampled traces are model checked individually to determine whether a given property ϕ holds, and the number of satisfying traces is used by a hypothesis testing procedure to decide (approximately) between H_0 and H_1. Note that Statistical Model Checking cannot guarantee a correct answer to the PMC problem. Still, the probability of giving a wrong answer can be arbitrarily bounded by the user.

We introduced a sequential Bayesian hypothesis testing approach and applied to the verification of rule-based models of signaling pathways and other stochastic systems[46,47]. Sequential sampling means that the number of sampled traces is not fixed a priori, but it is instead determined at "run-time", depending on the evidence gathered by the samples seen so far. This often leads to significantly smaller number of sampled traces. The hypothesis test is based on the Bayes Factor, that is, the likelihood ratio of the sampled data with respect to the two hypotheses. Formally, the Bayes Factor of data d and hypotheses H_0 and H_1 is $B = \frac{Pr(d|H_0)}{Pr(d|H_1)}$. Therefore, B can be interpreted as a measure of evidence (given by the data d) in favor of H_0. Now, fix an evidence threshold $T > 1$. Our algorithm iteratively draws independent and identically distributed (iid) sample traces $\sigma_1, \sigma_2, ...$, and checks whether they satisfy ϕ. The algorithm then computes the Bayes Factor B to check if it has obtained conclusive evidence. The algorithm accepts H_0 if $B > T$, and rejects H_0 (accepting H_1) if $B < \frac{1}{T}$. Otherwise ($\frac{1}{T} \leqslant B \leqslant T$) it continues drawing iid samples. It can be shown that when the algorithm terminates, the probability of a wrong answer is bounded above by $\frac{1}{T}$ (details of the algorithm can be found elsewhere[46]).

C. Application to HMGB1 model

We applied Statistical Model Checking (SMC) to verify formally some fundamental properties that our BioNetGen model should satisfy. We test whether the model satisfies a given BLTL property with probability $p \geqslant 0.9$. We set the threshold $T = 1000$ for the verification, so the probability of a wrong answer is smaller than 10^{-3}.

Property 1: p53 is normally expressed at low levels in human cells. We verified the following property

$$Pr_{\geqslant 0.9}[\mathbf{F}^{600}(\mathbf{G}^{900}(P53 < 3.3 \times 10^4))]$$

which informally means that the number of p53 molecules will be less than a threshold value within 10 hours, and it will always stay below this value during the next 900 minutes. SMC accepts this property as true after sampling 22 (satisfying) traces.

Property 2: p53's expression level increases quickly in response to various stresses, including oncoproteins activation. We verified the property

$$Pr_{\geqslant 0.9}[\mathbf{F}^{100}(P53 > 5.3 \times 10^4)]$$

TABLE II: Statistical Model Checking of Property 4 and 5

Property 4: $Pr_{\geqslant 0.9}[\mathbf{F}^{600}(CyclinE > 900)]$				Property 5: $Pr_{\geqslant 0.9}[\mathbf{F}^{600}(CyclinD > 900)]$			
HMGB1	# of Samples	# of Success	Result	d_{RAS}	# of Samples	# of Success	Result
10^2	9	0	False	10^{-6}	22	22	True
10^3	55	16	False	10^{-2}	26	5	False
10^6	22	22	True	10^{-1}	9	0	False

that is, within 100 minutes p53's level will eventually be larger than 5.3×10^4. SMC accepts this property as true, after sampling 38 traces (of which 37 satisfying).

Property 3: PI3K will be activated in order of minutes after HMGB1 binds to RAGE. We verified the following property

$$Pr_{\geqslant 0.9}[\mathbf{F}^{20}(PI3K_a/PI3K_{tot} > 0.5)]$$

which means that half of PI3K will be activated within 20 minutes. If the initial value of HMGB1 is 10^4, this property was accepted as true (22 satisfying traces). But if HMGB1 is set to 10^3, the property was rejected (9 unsatisfying traces).

Property 4: The overexpression of HMGB1 will promote the expression level of Cyclin E before the G1-S phase transition point, thereby facilitating the G1-S phase transition. We verified the property

$$Pr_{\geqslant 0.9}[\mathbf{F}^{600}(CyclinE > 900)]$$

that is, the number of Cyclin E will eventually exceed 900 within 600 minutes (10 hours). We verified this property with various values of HMGB1 and the results are shown in Table II.

Property 5: Mutation in K-RAS leads to continuous activation of downstream pathways and overexpression of CyclinD in the G1 phase. We verified the property

$$Pr_{\geqslant 0.9}[\mathbf{F}^{600}(CyclinD > 900)]$$

with different RAS deactivation rates (d_{RAS}). The results are presented in Table II. Property 4 and 5 show that overexpression of HMGB1 and mutation of RAS (small d_{RAS} value) will accelerate the expression of cell regulatory protein CyclinD/E to promote cell proliferation. However, inhibition of HMGB1 and RAS expression will prevent tumor growth.

Property 6: Within 300 minutes, Cyclin E's expression level is very low until 50% of RAS has been activated. We verified the property

$$Pr_{\geqslant 0.9}[(CyclinE < 10)\mathbf{U}^{300}(RAS_a/RAS_{tot} > 0.5)].$$

SMC accepted this property as true (22 satisfying traces).

V.   DISCUSSION

We presented a reaction network model of the signaling transduction initiated by HMGB1. The model incorporates the contribution from the most important known signaling components of the HMGB1 signal transduction network. The model is expressed in the form of BioNetGen rules, and simulated using ODEs and Gillespie's algorithm under a range of conditions. We used Statistical Model Checking to validate formally and automatically our model with respect to known experimental results.

Our simulations demonstrate a dose-dependent p53 and Cyclin E response curve to increasing HMGB1 stimulus. That is, overexpression of HMGB1 will promote the cell cycle regulatory proteins E2F and CyclinD/E, and inhibit the pro-apoptotic protein p53, leading to increased cancer cell survival and decreased

apoptosis. This is consistent with experimental observations in recent studies on cancer cells. We also investigated the role of different components in the pathway and predict their activity in response to various conditions. We investigated how mutation of the RAS, ARF and P21 proteins influence the fate of cancer cell. In particular, parameter variation showed that the mutated RAS will continuously activate the downstream signal transduction and increase the expression level of Cyclin E, leading to cancer cell proliferation. Mutation or loss of the ARF protein could not inhibit MDM2's activity and stabilize p53's expression level if HMGB1 is overexpressed, resulting in decreased apoptosis. Our model shows that the inhibition (or deactivation) of RAS, Cyclin, and Cyclin-dependent kinases (CDK) could inhibit tumor growth.

Besides the PI3K-AKT and RAS-ERK pathways, HMGB1 can also activate the NFκB signaling pathway[23]. This pathway regulates many pro-apoptotic and anti-apoptotic proteins' transcription[49]. A larger network for HMGB1 signal transduction and more analysis will be conducted in our future work.

A recent study has found that pancreatic tumor cells increase autophagy[10], and release HMGB1 in response to chemotherapy, radiation, and hypoxia, which may promote tumor cell survival[9]. It has been hypothesized that direct inhibition of autophagy may be another way to inhibit tumor growth and enhance the efficacy of cancer therapies[10]. The incorporation of autophagic proteins into the HMGB1 signaling pathway is worth to be considered in future work.

Moreover, understanding of HMGB1 at mechanistic level is still not clear, and reaction rates for some proteins interactions have not been measured by experiments. Although, currently, our model can only qualitatively compare with the experimental behavior, it still provides valuable information about the behavior of HMGB1 signaling transduction in response to different stimuli. Since much research on HMGB1 is on the way, more experimental data, with the help of powerful model checking techniques, will help us to estimate or constrain model parameters and make our models more realistic.

---

1   D. Hanahan, R.A. Weinberg, Cell, 100(1), 57-70 (2000).

2   T. Bonaldi, F. Talamo, P. Scaffidi, D. Ferrera, A. Porto, A. Bachi, A. Rubartelli, EMBO J, 22, 5551-60(2002).

3   H. Wang, O. Bloom, M. Zhang, J. Vishnubhakat, M. Ombrellino, A. Frazier, H. Yang, Science, 285, 248(1999).

4   I.E. Dumitriu, P. Baruah, B. Valentinis, et al J Immunol 174, 7506-15(2005).

5   C. Semino, G. Angelini, A. Poggi, A. Rubartelli Blood 106, 609-16 (2005).

6   L.J. Sparvero, D. Asafu-Adje, R. Kang, D. Tang, N. Amin, et al. J. Transl. Med. 7, 17 (2009).

7   J.E. Ellerman, C.K. Brown, M. de Vera, H.J. Zeh, T. Billiar, et al. Clin. Cancer Res. 13, 2836-48 (2007).

8   M. Lotze, K. Tracey, Nature Reviews 5:331 (2005).

9   J. Vakkila, M. Lotze, Nature Reviews 4,641 (2004).

10   R. Kang, D. Tang, N.E. Schapiro, K.M. Livesey, A. Farkas, P. Loughran, Bierhaus, M.T. Lotze, H.J. Zeh, Cell Death and Differentiation, 17(4), 666-76 (2009).

11   M.L. Brezniceanu, K. Volp, S. Bosser, C. Solbach, P. Lichter, et al. FASEB J, 17, 1295 (2003).

12   K.B. Wee, B.D. Aguda, Biophysical Journal 91, 857-865 (2006).

13   K.B. Wee, U. Surana, R.D. Aguda, PloS One 4(2), e4407 (2009).

14   A. Ciliberto, B. Novak, J. Tyson, Cell Cycle 4 3, 488-493 (2005).

[15] T. Zhang, P. Brazhnik, J. Tyson, Cell Cycle 6 1, 85-e10 (2007)

[16] K. Puszynski, B. Hat, T. Lipniacki, Journal of Theoretical Biology 254, 452-465 (2008).

[17] S. Bottani, B. Grammaticos, Journal of Theoretical Biology 249, 235-245 (2007).

[18] B. Vogelstein, D. Lane, and A.J. Levine, Nature 408, 307-310 (2000)

[19] A. C. Minella , J. E. Grim , M Welcker, B. E. Clurman, Oncogene 26, 6948-6953 (2007)

[20] D.G. Pestov, A Strezoska, and L.F. Lau, Molecular and Cellular Biology, 21 (13),4246-4255 (2001)

[21] C. J. Bakkenist, M.B. Kastan, Nature 421, 499-506 (2003).

[22] S. Larris, A.J. Levine, Oncogene 24, 2899-2908 (2005).

[23] J.R. van Beijnum, W.A. Buurman, A.W. Griffioen, Angiogenesis 11, 91-99(2008).

[24] Y. Barak, T. Juven, R. Haffner, M. Oren, EMBO J, 12, 461-468 (1993).

[25] Y. Haupt, R. Maya, A. Kasaz, M. Oren, Nature 387, 296-299 (1997).

[26] C. Blanco-Aparicio, O. Renner, J.E.M. Leal, A. Carnero, Carcinogenesis 28, 1379-1386 (2007).

[27] G. Yao, T.J. Lee, S. Mori, J. Nevins, L. You, Nature Cell Biology 10, 4 (2008).

[28] C.J. Sherr, F. McCormick, Cancer Cell, 2:103-112 (2002).

[29] R.C. Sears, J.R. Nevins, The Journal of Biological Chemistry 277, 11617-11620 (2002).

[30] J.R. Nevins, Human Molecular Genetics, 10, 699-703 (2001).

[31] N. Bardeesy, R.A. DePinho, Nature Reviews Cancer 2(12):897-909 (2002).

[32] J. Downward, Nature Reviews, 3, 11(2003)

[33] J. Mao, J. Perez-losada, D. Wu, R. DelRosario, R. Tsunematsu, K. I. Nakayama, K. Brown, S. Bryson, A. Balmain, Nature, 9;432(7018), 775-9 (2004)

[34] W.S. Hlavacek, J.R. Faeder, M.L. Blinov, R.G. Posner, M. Hucka, and W. Fontana, Sci. STKE 2006, re6 (2006).

[35] D. T. Gillespie, J. Comp. Phys., 22(4), 403-434, (1976).

[36] H. Gong, H. Sengupta, A. Linstedt, R. Schwartz, Biophysical journal, 95, 1674-1688 (2008).

[37] H. Gong, Y. Guo, A. Linstedt, R. Schwartz, Physical Review E 81(1):011914 (2010).

[38] T. Lipniacki, T. Hat, J.R. Faeder, W.S. Hlavacek, Journal of Theoretical Biology 254, 110-122 (2008).

[39] N. Geva-Zatorsky, N. Rosenfield, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, U. Alon, Mol. Sys. Biol, 0033 (2006).

[40] W.C. Hahn, R.A. Weinberg, Nature Reviews, 2:331 (2002).

[41] D.A. Altomare, H.Q. Wang, K.L. Skele, A. De Rienzo, A.J. Klein-Szanto, A.K. Godwin, J.R. Testa, Oncogene 23 (34), 5853-7, (2004).

[42] N. Rhodes, DA Heerding, DR Duckett, et al. Cancer Research 68, 2366 (2008)

[43] C. McInnes, Drug Discovery Today, 13(19-20), 875-881, (2008).

[44] M. Malumbres, P. Pevarello, M. Barbacid and J.R. Bischoff, Trends in Pharmacological Sciences, 29(1) 16 (2008)

[45] B. Rotblat, M. Ehrlich, R. Haklai, Y. Kloog, Methods Enzymol, 439:467-89 (2008).

[46] P. Zuliani, A. Platzer, E.M. Clarke, HSCC 2010 (to appear)

[47] S. K. Jha, E.M. Clarke, C.J. Langmead, A. Legay, A. Platzer, P. Zuliani, CMSB, 218-234 (2009).

[48] E.M. Clarke, J.R. Faeder, L.A. Harris, C.J. Langmead, A. Legay, S.K. Jha, CMSB, 231-250 (2008).

[49] K. Puszynski, R. Bertolusso, T. Lipniacki, IETDL 3, 356-367 (2009).

[50] C.J. Langmead, CSB, 201-212 (2009).

[51] E.M. Clarke, O. Grumberg, and D.A. Peled, *Model Checking*, MIT Press (1999).

[52] E.M. Clarke, E.A. Emerson, and J. Sifakis, Commun. ACM, 52 (11), 74-84 (2009).

[53] H.L.S. Younes, R.G. Simmons, Inf. and Comp., 204 (9), 1368-1409, (2006).

[54] A. Rizk, G. Batt, F. Fages, and S. Soliman, CMSB, 251-268, (2008).

[55] http://www.cs.cmu.edu/~haijung/research/Supplement.pdf

APPENDIX A: ORDINARY DIFFERENTIAL EQUATIONS

$$\frac{d}{dt}RAGE_a(t) = k_1 RAGE(t)HMGB1 - d_1 RAGE_a(t)$$

$$\frac{d}{dt}PI3K_a(t) = (k_2 RAGE_a(t) + k_2' RAS_a(t))PI3K(t) - d_2 PI3K_a(t)$$

$$\frac{d}{dt}PIP3(t) = k_3 PI3K_a(t)PIP2(t) - d_3 PTEN(t)PIP3(t)$$

$$\frac{d}{dt}AKT_p(t) = k_4 PIP3(t)AKT(t) - d_4 AKT_p(t)$$

$$\frac{d}{dt}PTEN(t) = k_5 P53(t)^3/(K_1^3 + P53(t)^3) - d_5 PTEN(t)$$

$$\frac{d}{dt}mdm2(t) = k_6 P53(t)^3/(K_1^3 + P53(t)^3) - d_6 mdm2(t)$$

$$\frac{d}{dt}MDM2(t) = k_7 mdm2(t) + d_8 MDM2_p(t) - (d_7 + k_8 AKT_p(t) + d_7' ARF(t))MDM2(t)$$

$$\frac{d}{dt}MDM2_p(t) = k_8 AKT_p(t)MDM2(t) - d_8 MDM_p(t) - (d_8' + d_7' ARF(t))MDM2_p(t)$$

$$\frac{d}{dt}P53(t) = k_9 - d_9 P53(t) - d_9' MDM2_p(t)P53(t)$$

$$\frac{d}{dt}P21(t) = k_{10} P53(t)^2/(K_1^2 + P53(t)^2) - d_{10} P21(t)$$

$$\frac{d}{dt}RAS_a(t) = a_1 RAGE_a(t)RAS(t) - b_1 RAS_a(t)$$

$$\frac{d}{dt}RAF_a(t) = a_2 RAS_a(t)RAF(t) - b_2 RAF_a(t)$$

$$\frac{d}{dt}MEK_p(t) = a_3 RAF_a(t)MEK(t) - b_3 MEK_p(t)$$

$$\frac{d}{dt}ERK_p(t) = a_4 MEK_p(t)ERK(t) - b_4 ERKp(t)$$

$$\frac{d}{dt}Myc(t) = a_5 ERK_p(t)/(K_2 + ERK_p(t)) - b_5 Myc(t)$$

$$\frac{d}{dt}CD(t) = \frac{a_6 ERK_p(t)}{K_2 + ERK_p(t)} + \frac{a_6' Myc(t)}{K_2 + Myc(t)} - (b_6 + b_6' P21(t) + b_6'' INK4A(t))CD(t)$$

$$\frac{d}{dt}RE(t) = a_7 RB(t)E2F(t) - b_7 RE(t) - b_7'(CD(t) + CE(t))RE(t)$$

$$\frac{d}{dt}RB_p(t) = (a_8 RB(t) + b_7' RE(t))CD(t) - (b_8 + b_8')RB_p(t) + b_7' CE(t)RE(t)$$

$$\frac{d}{dt}RB(t) = a_9 + b_8 RB_p(t) - a_8 CD(t)RB(t) - (a_7 E2F(t) + b_9)RB(t)$$

$$\frac{d}{dt}E2F(t) = \frac{a_{10} Myc(t)}{K_2 + Myc(t)} + b_7' CD(t)RE(t) - (a_7 RB(t) + b_{10})E2F(t) + b_7' CE(t)RE(t)$$

$$\frac{d}{dt}ARF(t) = a_{11} E2F(t)/(K_3 + E2F(t)) - b_{11} ARF(t)$$

$$\frac{d}{dt}CE(t) = a_{12} E2F(t)/(K_3 + E2F(t)) - b_{12} CE(t) - b_6' CE(t)P21(t)$$

$$\frac{d}{dt}INK4A(t) = a_{13} - b_{13} INK4A(t)$$