Simulation and Model Checking of HMGB1 Signaling Pathway in Cancer

Haijun Gong[1*], Paolo Zuliani[1], Qinsi Wang[1], Anvesh Komuravelli[1], James Faeder[2], Michael Lotze[2], Edmund Clarke[1]

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

[2]University of Pittsburgh, Pittsburgh, PA 15260

*Email: haijung@cs.cmu.edu

Abstract

High-mobility group box-1 (HMGB1) protein has been recently shown to be associated with the cell proliferation of various types of cancers. The protein HMGB1 can activate a number of signaling networks – p53, NFkB, Ras and Rb pathways – which control many physiological processes of the cell. We propose a rule-based model and a Boolean model of the HMGB1-p53-NFkB-Ras-Rb network to investigate how these couplings influence proliferation and apoptosis of cancer cells. The rule-based model was implemented using the BioNetGen language which can simulate both ordinary differential equations and Gillespie's stochastic simulation algorithm. Then, we analyzed and verified qualitative properties of the model by means of simulations and statistical model checking. For Boolean network model, Symbolic Model Checking (SMV) is applied to query and verify some temporal logic properties of HMGB1 model.

Our simulations show that, if HMGB1 is overexpressed, the expression level of oncoproteins CyclinD/E, which regulate cell proliferation, are upregulated, while tumor suppressor proteins which regulate cell apoptosis, such as p53, are repressed. The discrete, stochastic simulations show that HMGB1-activated receptors can generate sustained oscillations of irregular amplitude for the p53, MDM2 and NFkB proteins. Moreover, the models predict that mutation or overexpression of RAS, ARF, p21 and IkB kinase could influence the cancer cell's fate -- apoptosis or survival – through the crosstalk of different pathways. Our work shows that computational modeling and model checking can be effectively combined in the study of biological signaling pathways.

HMGB1 Signaling Pathway [1,2]

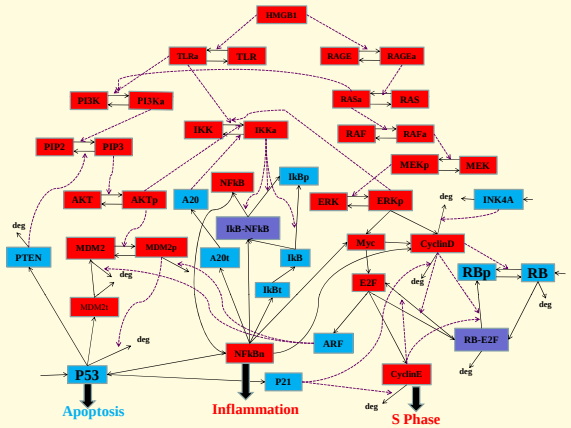


Figure 1. Schematic view of HMGB1 signal transduction. Blue nodes represent tumor suppressor proteins, red nodes represents oncoproteins/lipids. Solid lines with arrows denote protein transcription, degradation or changes of molecular species; dashed line with arrows denote activation processes.

Rule-based Simulation Method

We formulated a reaction model corresponding to the reactions illustrated in Fig.1 in the form of rules specified in the BioNetGen language, The ordinary differential equation (ODE) method and stochastic simulation algorithm (SSA) are used to simulate the model. Example ODE and BioNetGen rules:

$$\frac{d}{dt}MDM2_p(t) = k_1 AKT_p(t) MDM2(t) - d_1 MDM_p(t) - d_2 MDM2_p(t) - d_3 ARF(t) MDM2_p(t)$$

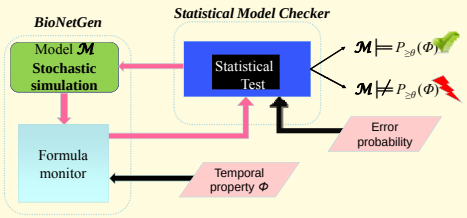
•MDM2 phosphorylation:	MDM2(a~U) + AKTp → MDM2(a~p) + AKTp	k1
•MDM2$_p$ dephosphorylation:	MDM2(a~p) → MDM2(a~U)	d1
•MDM2$_p$ degradation:	MDM2(a~p) → Trash()	d2
•MDM2$_p$ degradation :	MDM2(a~p) + ARF → ARF	d3

Linear Temporal Logic & Computation Tree Logic

LTL formula describes the properties of an infinite sequence of states. LTL *temporal* operators describe the properties of a path: **X**p – p holds in the *second* state of the path; **F**p – p holds at some state in the *future* on the path; **G**p – p holds *globally (always)* at every state on the path; p**U**q – p holds *until* q holds. A **CTL** temporal logic formula is constructed from *path* quantifiers and temporal operators. Two path quantifiers describe the branching structure in the computation tree: **A** – for *all* paths, and **E** – there *exists* a path.

Statistical Model Checking

Statistical Model Checking of biochemical models: $\mathcal{M} \models P_{\geq\theta}(\Phi)$?



Require: *Property* $P_{\geq\theta}(\Phi)$, *Threshold* $T \geq 1$, *Prior* g
$n := 0$ {*number of traces drawn so far*}
$x := 0$ {*number of traces satisfying* Φ *so far*}
repeat
  $\sigma :=$ draw a sample trace from BioNetGen (iid)
  $n := n + 1$
  if $\sigma \models \Phi$ **then**
    $x := x + 1$
  endif
  $\mathcal{B} := BayesFactor(n, x, \theta, g) = \frac{P(X|H_0)}{P(X|H_1)}$
until ($\mathcal{B} > T \vee \mathcal{B} < 1/T$)
if ($\mathcal{B} > T$) **then**
  return "H_0 *accepted*"
else
  return "H_0 *rejected*"
endif

Figure 2. Flow chart and Pseudocode for Statistical Model Checking

Simulation Results [1,2]

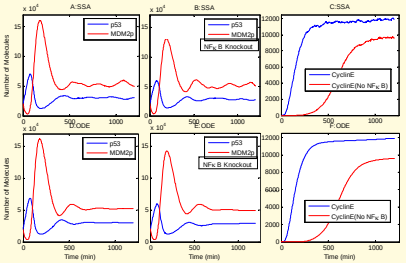


Figure 3. Baseline simulation results for the SSA (A-C) and ODE (D-F) models.

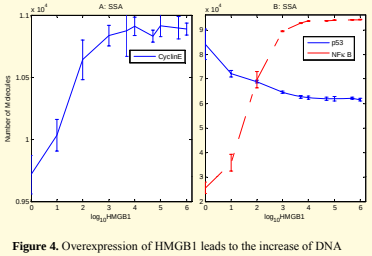


Figure 4. Overexpression of HMGB1 leads to the increase of DNA replication proteins Cyclin E and NFkB, decrease of p53.

Verification of Rule-based HMGB1 Model [1,2]

Property 1: *PI3K will be activated in order of minutes after HMGB1 binds to RAGE.*

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

Table 1 validates that half of PI3K will be activated within 20 minutes with different values of HMGB1.

Table 1

Property 1: $Pr_{\geq 0.9}[\mathbf{F}^{20}(PI3K_a/PI3K_{tot} > 0.5)]$				
HMGB1	# of Samples	# of Successes	Result	Time (s)
10^3	9	0	False	6.49
9×10^3	380	315	False	285.16
10^5	22	22	True	16.39

Property 2: *NFkB's expression level oscillates due to the negative feedback loops activated by IkB and A20 which are NFkB's transcription targets.*

$$Pr_{\geq 0.9}[\mathbf{F}(\frac{NFkB_n}{NFkB_{tot}} \geqslant 0.65 \;\&\; \mathbf{F}(\frac{NFkB_n}{NFkB_{tot}} \leqslant 0.20 \;\&\; \mathbf{F}(\frac{NFkB_n}{NFkB_{tot}} \geqslant 0.20 \;\&\; \mathbf{F}(\frac{NFkB_n}{NFkB_{tot}} \leqslant 0.20))))]$$

Table 2

HMGB1	t(min)	# of Samples	# of Success	Result	Time (s)
10^2	45	13	1	False	76.77
10^2	60	22	22	True	111.76
10^2	75	104	98	True	728.65
10^5	30	4	0	False	5.76
10^5	60	4	0	False	31.98

Property 3: *A large proportion of PI3K, RAS and IKK will be activated and stay at a high expression level all the time when HMGB1 is overexpressed:*

$$Pr_{\geq 0.9}[\mathbf{F}\mathbf{G}^{180}(PI3K_a/PI3K_{tot} > 0.9 \;\&\; RAS_p/RAS_{tot} > 0.8 \;\&\; IKK_a/IKK_{tot} > 0.6)]$$

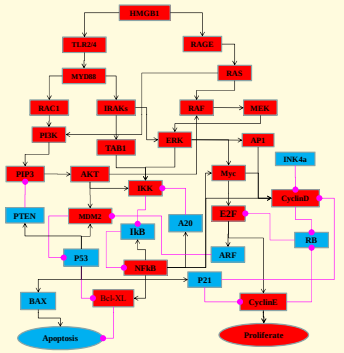
Property 4: *The overexpression of IKK will promote the translocation of NFkB into nucleus, inducing the transcription of Cyclin E:* $Pr_{\geq 0.9}[\mathbf{F}^{300}\mathbf{G}^{300}(CyclinE >= 10,000)]$.

Table 3 shows the verifications of the property 3 and 4 with various values of *t* and IKK expression level.

Table 3

Property 3					Property 4				
t(min)	Samples	Success	Result	Time (s)	IKK	Samples	Success	Result	Time (s)
90	9	0	False	21.27	10^3	22	22	True	547.52
110	38	37	True	362.19	2×10^4	9	2	False	55.86
120	22	22	True	214.38	10^2	4	0	False	16.89

Boolean Network Model [4]



A Boolean network is composed of a graph G and a Boolean transfer function for each node. The state of each node could be either ON (1) or OFF (0) at any time step.

The Boolean transfer function describes the transformation of the state of node from time t to t + 1, which depends on its current state and that of its parents, which can be parental activators or parental inhibitors, that is,

$$x_i(t+1) = \left(x_i(t) \vee \bigvee_{Pa} x_{Pa}^{a}(t)\right) \wedge \neg\left(\bigvee_{Pa} x_{Pa}^{in}(t)\right)$$

Symbolic Model Checking (SMV)

```
MODULE MAIN
VAR
        HMGB1: boolean;
        PI3K, PIP3, PTEN, ... : boolean;
ASSIGN
        init(PI3K):={0,1}; init(PTEN):={0,1};
        init(PIP3):={0,1}; init(INK4a):=0; ...
        next(PIP3):=  //update rule for PIP3
          case
            PI3K & !PTEN: 1;
                  PTEN: 0;
                    1: PIP3;
          ... ...     //update rules for other variables
SPEC  AG(RAS → AF(CyclinE)); // property verification
```

SMV code can be divided into three parts: ***variable declarations*** ("boolean");

initialization of the states for each variable with **init**;

Implementation – updating the state of each node in the state transition diagram with **next**

The verification of CTL properties is encoded using the **SPEC** statement.

SMV Applications [4]

Verification of five types of properties *occurrence, consequence, steady states, oscillation, and pathway.*

Property 1: If HMGB1 is overexpressed, cancer cell will **necessarily** activate *Proliferate* at some time in the future. The CTL property is verified: **AF**(*Proliferate*) : *True*

Property 2: If RAS is continuously activated, the cell will eventually satisfy (!Apoptosis & Proliferate) **AG**(*(RAS)* → **AF**(*!Apoptosis & Proliferate*))

Property 3: Are the states (!P53 & !Apoptosis & CyclinE & Proliferate) steady? **AF**[**AG**(*!P53 & !Apoptosis & CyclinE & Proliferate*)]

Property 4: The release of HMGB1 will oscillations of NF*k*B's expression level in the nucleus. **AG**((*!NFκB* → **AF**(*NFκB*)) & (*NFκB* → **AF**(*!NFκB*)))

Property 5: Is NF*k*B's activation a necessary checkpoint that the cancer cell should go through before it reaches (!Apoptosis & Proliferate)? **!E**[(*!NFκB*) **U** (*!Apoptosis & Proliferate*)]

Conclusions

1. Overexpression of HMGB1 will promote the expression level of cell cycle regulatory protein Cyclin E and NFkB, inhibit the pro-apoptotic protein p53.
2. Statistical Model Checking and Symbolic Model Checking techniques can be effectively combined in the signaling pathway and verify some important temporal properties.

Acknowledgement

This work was supported by a grant from the U.S. national Science Foundation's Expeditions in Computing Program (award ID 0926181).

References

[1] H. Gong, P. Zuliani, A. Komuravelli, J. Faeder, E. Clarke, Analysis and Verification of the HMGB1 Signaling Pathway, *BMC Bioinformatics 2010*

[2] H. Gong, P. Zuliani, A. Komuravelli, J. Faeder, E. Clarke, Computational Modeling and Verification of Signaling Pathways in Cancer, *Proceedings of Algebraic and Numeric Biology 2010*

[3] P. Zuliani, A. Platzer, E. Clarke, Bayesian statistical model checking with application to Simulink/Stateflow verification, *HSCC 2010*

[4] H. Gong, Q. Wang, P. Zuliani, J. Faeder, M. Lotze, E. Clarke, Symbolic Model Checking of Signaling Pathways in Pancreatic Cancer, *Proceeding of International Conference on Bioinformatics and Computational Biology 2011*