Probabilistic Verification of ER Stress-induced Signaling Pathways

Haijun Gong, Lu Feng
Department of Mathematics and Computer Science, Saint Louis University, St. Louis, MO 63103, USA
Email: hgong2@slu.edu

***Abstract*—Endoplasmic reticulum (ER) is a communication hub for several signaling and secretory pathways that regulate the cell cycle progression. Recent studies revealed that deregulation of ER stress-induced signaling pathways are involved in the pathogenesis of cancer and Alzheimers disease. Computational analysis and verification of these pathways will provide insights into the mechanism linking ER stress with different diseases. Based on the experimental observation and our recent bioinformatics studies, we construct a quantitative model to systematically study the ER stress-induced apoptosis and survival signaling pathways. To overcome the limitations of traditional simulation techniques, a formal verification method, called Probabilistic Model Checking, was introduced and applied to formally analyze the temporal logic properties of the stochastic model, which is written in the PRISM language. Compared with the symbolic model verification method, this PRISM verification technique can not only qualitatively verify the signaling pathway model using sequential probability ratio test (SPRT), but also provide a quantitative estimation of the temporal properties using confidence interval estimation method. The probabilistic verification studies show that, overexpressed ER transmembrane proteins will promote the expression of Cyclin D, Amyloid-β and MDM2, leading to the pathogenesis of cancer and other diseases in the long term, and inhibit the apoptosis. Our work also verified that, varying IKK's expression level will influence the number of NFκB molecules in the nucleus, which explained why IKK inhibitor could inhibit tumorigenesis. The proposed signaling pathway model and PRISM method can help our understanding of the mechanisms that link ER stress with cancer and other diseases.**

Keywords*—ER stress; signaling pathway; cancer; Alzheimer's disease; Probabilistic verification; PRISM; sequential probability ratio test; confidence Interval

I. Introduction

Somatic mutations and deregulation of key signaling pathways [1], [2] are responsible for the pathogenesis of many diseases, including cancer, diabetes and Alzheimers disease (AD). Modern genome sequencing and microarray technique allows biologists to measure the somatic mutations in hundreds of samples from different cancer types and Alzheimer's disease in a quantitative way. Recent studies [3], [4] have identified several genetic signatures and signaling pathways that are frequently altered in the cancer and AD cells. Investigation of the somatic mutations and signaling pathways can help understand the mechanism behind the pathogenesis of cancer and AD, and develop effective strategies for the early diagnosis and treatment of these diseases. Computational biologists have proposed different statistical inference [5], [6], [7], [8], [9], [10], verification algorithms [11], [12], [13], [14], [15], [16] and simulation methods [17], [18], [19] to identify genetic signatures and analyze important regulatory networks. A comprehensive understanding of the signaling networks will help researchers to develop effective multi-gene targeted treatments for the patients. Since the signaling pathway is complex due to a large amount of biochemical reactions, it is not efficient to apply traditional methods to analyze large networks. Our previous work had proposed and successfully applied Statistical Model Checking [20], [21] and Symbolic Model Checking [15], [16] techniques to formally verify some signaling pathways in the cancer cell.

Recent clinical and wet lab experimental studies revealed that, dysfunction of the endoplasmic reticulum (ER) stress signaling is implicated in the pathogenesis of cancer, AD, diabetes and heart disease [22], [23], [24], [25]. ER is a major hub for protein synthesis, folding, sorting and several signaling pathways in the cell [2]. External stimuli, genetic mutations or some pathological conditions, including the hypoxia and viral infection, could induce cellular stress and abnormal accumulation of unfolded proteins in the ER, which can activate the unfolded protein response (UPR) [1]. UPR is a self-protective mechanism which could ameliorate the abnormal accumulation of unfolded proteins. However, prolonged or severe ER stress will activate the apoptosis (cell death) signaling pathway. Investigation of these signaling pathways will provide insights into the mechanism linking ER stress with different diseases.

Three signaling branches initiated by three ER-resident transmembrane proteins have been found to be able to sense the accumulation of unfolded proteins and regulate the cell cycle progression, including the Protein kinase RNA-like ER kinase (PERK), Inositol-requiring enzyme 1 (IRE1) and membrane-bound transcription factor ATF6. Recent experimental work [2] shows that, the pathways regulated by the ER and Golgi apparatus can activate both survival and apoptosis signaling pathways in response to external stress. Computational modeling and analysis are helpful to study the roles of ER stress-induced signaling pathways and secretory pathways in the cell cycle progression and some diseases. Several pathways, including P53, RAS and NFκB, have been well studied using discrete event simulation [18], [19], rule-based stochastic simulation [20], [21] and synchronous symbolic model verification (SMV) [15] methods. In [16], we extend the synchronous SMV and propose an asynchronous model checking method to formally analyze the ER-Golgi-regulated signaling pathways. Using the SMV model checker, we can only qualitatively check some computation tree temporal logic formulas related to the cancer and Alzheimer's disease. In this work, a stochastic simulation model and a probabilistic model verification technique are proposed to quantatively analyze the roles of ER stress-induced signaling pathways in the cell

978-1-4799-5669-2/14/$31.00 ©2014 IEEE

cycle progression, and investigate the molecular mechanisms underlying the pathogenesis of some diseases.

II. ER Stress-induced Signaling Pathways

In our recent work [16], we have summarized some possible signaling pathways induced by the ER stress. In this work, we will briefly reiterate some important pathways that are regulated by the ER membrane-associated transmembrane proteins, including ATF6, PERK and IRE1 sensors. Our objective is to develop a stochastic simulation model and apply probabilistic verification method to investigate the dynamic and temporal behaviors of these pathways in the pathogenesis of cancer and Alzheimer's disease. The crosstalk of these signaling components is depicted in the Fig. 1 modified from [16], where each node represents a regulatory component (protein or mRNA), the solid lines with arrows represent either protein transcription or molecular species changes, dashed lines with arrows represent activation processes. The protein with a subscript "a/p" corresponds to an active or phosphorylated form. For example, IKK (IKK$_a$) is an inactive (active) form of IKK.

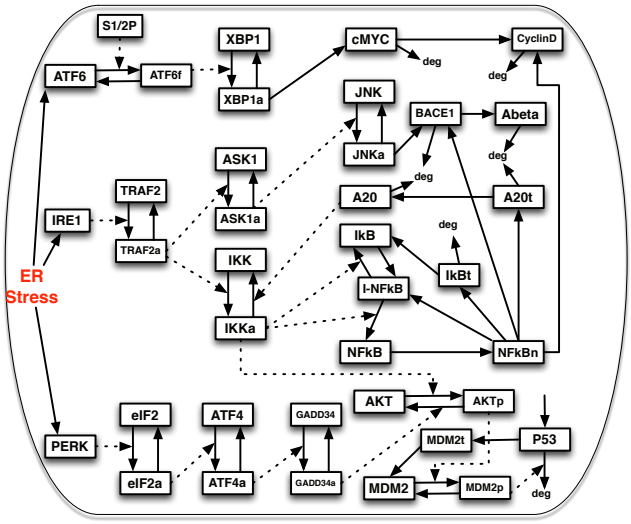

Fig. 1. ER stress-induced signaling pathways: each node represents a regulatory component; solid lines with arrows represent either protein transcription or molecular species changes; dashed lines with arrows represent activation processes.

ATF6 branch: $ATF6 + S1/2P \rightarrow ATF6f \rightarrow XBP1 \rightarrow cMYC \rightarrow CyclinD$. Under stress, with the help of coat protein II (COPII) complex, ER transmembrane protein ATF6 is transported to the Golgi apparatus [1] where it is processed/cleaved by the site 1 protease (S1P) and S2P (S1/2P in Fig.1), releasing the N-terminal cytosolic domain fragment ATF6f, an active form of ATF6. Upon activation, ATF6f enters the nucleus to activate the ER stress response element-dependent gene product X-box-binding protein 1 (XBP1) [1], [26], and upregulate the synthesis of cMYC and Cyclin D which are important in the G1/S phase transition in the cell cycle progression.

PERK branch: $PERK \rightarrow eIF2\alpha \rightarrow ATF4 \rightarrow GADD34 \dashv AKT \rightarrow MDM2 \dashv P53 \rightarrow MDM2t$. In response to ER stress or abnormal accumulation of misfolded proteins, PERK can phosphorylate and activate the α-subunit of eIF2 (eukaryotic translation initiation factor-2), leading to the activation of P53-dependent apoptosis signaling pathway [27] through regulating the downstream tumor suppressor genes including ATF4 (activating transcription factor-4), CHOP and GADD34. In the normal cell, the tumor suppressor P53's transcription activity is inhibited by the phosphorylated MDM2 in the nucleus which can promote the degradation of P53. MDM2's phosphorylation is activated by the oncoprotein AKT. P53 is also a transcription factor for MDM2, leading to a negative feedback loop between P53 and MDM2. In this pathway, P53 and eIF2 are frequently mutated or down-regulated in the cells of cancer and neurodegenerative diseases [28]. PERK branch's activation will lead to a reduction of protein synthesis and reduce protein stress in ER.

IRE1 branch: $IRE1 \rightarrow TRAF2 \rightarrow \{NF\kappa B, JNK\}$ pathways. ER stress can also induce the interaction between IRE1 and TRAF2 (tumor necrosis factor receptor associated factor-2), leading to the activation of NFκB and JNK pathways which are involved in the inflammation, apoptosis, survival and Amyloid-β production [1], [26]. JNK pathway, which is activated by ASK1 (Apoptosis signal-regulating kinase 1), can regulate the synthesis of BACE1 (beta-secretase) and production of Amyloid-β (Aβ) [1], [22], whose abnormal accumulation is one of the major hallmarks in Alzheimer's disease. IRE1 could also activate NFκB pathway which plays an important role in the cell growth, imflammation, and apoptosis: $TRAF2 \rightarrow IKK \dashv I\kappa B \dashv NF\kappa B \rightarrow \{CyclinD, BACE1, A20, I\kappa B\}$. In the resting cell, the tumor suppressor protein IκB binds to NFκB and forms a complex in the cytoplasm to inhibit its transcription activity. Some oncoproteins could activate IKK (IκB kinase), leading to the disassembly of NFκB-IκB complex (I-NFkB in Fig. 1). After NFκB translocates into the nucleus, it will promote the transcription of Cyclin D, BACE1, IκB and A20 [29], [30].

III. Probabilistic Model Verification

A. Linear Temporal Logic

In formal verification, a Kripke structure [15], [31] $M = (S, s_0, R, L)$ represents a finite-state concurrent system, where $s_0 \in S$ is an initial state, R is a transition relation between states S, and L is a labeling function that labels each state S with the set of atomic propositions (AP) true in that state. Given a temporal logic formula ϕ expressing some desired property in a system/model, and a model M with an initial state s_0, Model checking [31] is the process of determining whether or not M satisfies the property ϕ: $M, s_0 \models \phi$. If we are given a stochastic model M with a starting state s_0, a temporal logic formula ϕ, a threshold value $\theta \in [0, 1]$ describing a pre-defined probability, then, the probabilistic model checking (PMC) is to decide whether $M, s_0 \models Pr_{\geq\theta}[\phi]$ is true or not. That is, the property ϕ is true with a probability greater or equal to θ. PMC can also estimate the probability that ϕ satisfies M: $Pr[\phi]$.

In this work, the desired properties are expressed as Linear Temporal Logic (LTL) formulas, which describe properties of an infinite sequence of states. LTL formula is constructed from a set of APs, Boolean logic connectives, and *temporal* operators **X, F, G, U** describing the properties of a path [15], [31]. If M describes a stochastic model, we will use a

time-bounded temporal logic formulas to specify the desired properties. The syntax of the logic is written as

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

Time-bounded **F**, **G**, and **U** operators are defined as: $\mathbf{F}^t\phi$ or $\mathbf{F}(\leq t)[\phi]$ means ϕ holds true within time t; $\mathbf{G}^t\phi$ or $\mathbf{G}(\leq t)[\phi]$ means ϕ holds true globally up to time t; the time-bounded until operator $\phi_1\mathbf{U}^t\phi_2$ or $\phi_1\mathbf{U}(\leq t)\ \phi_2$ means, *within* time t, ϕ_1 will hold until ϕ_2 becomes true. We can also build composite operators using these basic LTL operators, for example, $F^{t_1}G^{t_2}[\phi]$ means, ϕ holds true within time t_1 and will be globally true up to time t_2.

The semantics of time-bounded LTL is defined with respect to the trace of a stochastic system, which is the output of a stochastic simulation. If (s_i, t_i) represents the state and sojourn time at the state s_i, then the trace is a sequence of time-stamped state transitions written as $\sigma = (s_0, t_0), (s_1, t_1), ...$ We use σ^k to denote the trace starting at the step k, then, $\sigma^k \models \phi$ represents the track σ^k satisfying the bounded LTL formula ϕ. The semantics of time-bounded LTL is written as

- $\sigma^k \models AP$ if and only if AP holds true in 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 exists $i \in N$ such that, for each $0 \leq j < i$, $\sigma^{k+j} \models \phi_1$, and if $\sum_{0\leq l<i} t_{k+l} \leq t$, then $\sigma^{k+i} \models \phi_2$.

B. Probabilistic model checker

The probabilistic model checker, PRISM, has been widely used for the modeling and verification of different systems [32], [33]. PRISM supports the verification of three types of probabilistic models, including the discrete-time Markov chains, Markov decision processes and continuous-time Markov chains (CTMCs) models. This work is based on the CTMC process, and the PRISM model file is given an extension .sm.

PRISM Model: ERprism.sm

```
ctmc // Continuous Time Markov Chains Model

// Declaration of constants
      const int N=1000;
      const double k1=0.01; const double d1=0.01;
      ...
module ER
  // Variable declaration and Initialization
      AKTp: [0..N] init 0;
      MDM2: [0..N] init N;  // unphosphorylated MDM2
      MDM2p: [0..N] init N; // phosphorylated MDM2
      ...
  — State transition update
      [ ] AKTp>0 & MDM2>0 //AKTp phosphorylates MDM2
      → k1*AKTp*MDM2: (MDM2p'=MDM2p+1) & (MDM2'=MDM2-1);
      [ ] MDM2p>0 //MDM2p dephosphorylation
      → d1*MDM2p: (MDM2p'=MDM2p-1) & (MDM2'=MDM2+1);
      ...
endmodule
```

Fig. 2. Illustration of the procedure to build a PRISM model of ER stress-induced signaling pathways

The procedure to develop a PRISM model of ER stress-induced signaling pathway is illustrated in Fig. 2. Since the proposed model is a continuous-time stochastic process, the PRISM code starts with the keyword **ctmc**. The values of some constants, for example the constants of transcription rate and (de)phosphorylation rate, are defined in the beginning with the keyword **const**. In the PRISM code, the program can consist of at least one module using the keyword "**module ... endmodule**". In this work, there is only one module "module ER" which contains all the regulatory components and interactions. The variables should be declared and initialized first. For example, "AKTp: [0..N] **init** 0;", which means, AKTp takes an integer value ranging from 0 to N, and its initial value is specified (as 0) with the keyword **init**. The behavior of each module is specified by the "guard" and "updates" taking the form: [] $\mathbf{guard} \rightarrow \mathbf{rates} : \mathbf{updates}$. That is, if the $guard$ (predicate) is true, then, the states in the module will be updated according to the "$updates$" rules with a rate of "$rates$". The $updates$ can have more than one statement in the form of $V' = f(V)$, indicating that the variable V's value is updated to V'. PRISM supports both synchronous and asynchronous update of modules. Each guarded command with an "empty" square bracket [] corresponds to an asynchronous process. Some transitions having the same label (called action) placed in the square brackets in different modules will be updated simultaneously. The signal transduction in this work is a stochastic process and each reaction occurs at different rates, so, the state update of each variable is asynchronous.

PRISM Property Specification: ERtest.csl

```
const double T;
P>= 0.9 [F<= T (CyclinD >= 500)];
P>= 0.9 [G[80, 100] (NFkBn/NFkBt <= 0.35)];
P>= 0.9[F<= T (P53/P53i <= 1 &  (MDM2p/MDM2i >= 10))];
P=? [F<= T (Abeta >= 300)];
P=? [G[80, 100] (NFkBn/NFkBt <= 0.35)];
```

Fig. 3. Some PRISM property specifications of ER stress-induced signaling pathway based on sequential probability ratio test and confidence interval estimation methods

PRISM is a probabilistic model checker which can automatically and formally verify the continuous-time Markov chain models. The continous stochastic logic (CSL), which is a property specification language for CTMC process, will be used to specify the temporal properties in PRISM. Given a property of the form $P_{\geq p}\ [\phi]$, Wald's sequential probability ratio test (SPRT) is applied to verify a property on-the-fly through generating a succession of simulation traces when an answer ("True", "False") can be given with a high confidence. SPRT is dependent on two parameters, δ used as the half-width of an indifference region $[p - \delta, p + \delta]$, and "α/β" which represent the probability of type I/II errors. Fig. 3 illustrates some temporal logic formulas. For example, the property "$P_{>=p}\ [\mathbf{F}_{<=t}\ (CyclinD \geq 500)]$" means, within t minutes, the number of CyclinD molecules in the cell will be greater than 500 with a probability at least p. This property is an assertion, so PRISM will output either "True" or "False", which is similar to our previous statistical model checking method based on the hypothesis testing proposed in [20].

PRISM could also directly evaluate a numerical value of

a bounded property $P_{=?}[\phi]$ using confidence interval (CI) method, which is dependent on two parameters, α (confidence coefficient) and ω (half-width of the interval), or N (number of samples). For example, "$P_{=?}$ $[\mathbf{G}_{[80,100]}\ NFkBp/NFkBt \leq 0.35]$" estimates the probability that at most 35% of NFκB molecules stay in the nucleus in the time interval [80, 100]. Given two temporal logic formulas ϕ_1 and ϕ_2, PRISM could also estimate the conditional probability $P_{=?}(\phi_1|\phi_2) = P_{=?}(\phi_1 \& \phi_2)/P_{=?}(\phi_2)$.

IV. Applications and Results

In this section, we will apply PRISM to analyze the ER stress-induced signaling pathway and investigate how the disruption of these pathways influence the pathogenesis of cancer and neurodegenerative disease. In this model, we assume some regulatory components can take two possible forms, either active (phosphorylated) or inactive (unphosphorylated). All substrates are expressed in the number of molecules; and the subscript "t/i" corresponds to the total/initial number of molecules. We assume the total number of molecules for some substrates (active + inactive forms) is constant. For example, IKK + IKK$_a$ = IKK$_t$ $= 10^4$. The values of most parameters are either abstracted from our previous work [20], [21] or estimated or tuned from the experiments and other literatures. The full PRISM code with parameters is available on [34]. We will first verify whether or not the model satisfies a given time-bounded LTL formula with a probability p. We set $\delta = \alpha = \beta = 0.05$ in the SPRT.

Hypothesis I: The expression level of the oncoprotein Cyclind D and AD-associated protein Amyloid-β will be upregulated due to ER stress.

$$P_{\geq 0.9}(\phi_1) = P_{\geq 0.9}\ [F_{\leq t}\ (CyclinD \geq 500)];$$
$$P_{\geq 0.9}(\phi_1') = P_{\geq 0.9}\ [F_{\leq t}\ (A\beta \geq 300)].$$

Formulas ϕ_1 and ϕ_1' mean that, if PERK, IRE1 and ATF6 are overexpressed (with an initial value 10^3), then, within t minutes, the number of Cyclin D and Amyloid-β molecules will be equal to or greater than a threshold value with a probability of at least 0.9. We checked these properties while varying the value of t. The verification results show that if $t \geq 50$ minutes around, these formulas were true, which explained why the ER stress could promote the pathogenesis of cancer growth and Alzheimer's disease [35], [36]. These properties also confirmed our recent results [16] using symbolic model checker that overexpressed ER transmembrane proteins will finally promote the transcription of Amyloid-β peptides and Cyclin D, inducing cell to reach a state of "Alzheimer = True" and "Cancer=True" on all paths.

Hypothesis II: ER stress can induce apoptosis in the early stage through activating the P53 signaling pathway; however, the tumor suppressor P53's expression is inhibited in the long term.

$$P_{\geq 0.9}(\phi_2) = P_{\geq 0.9}\ \{F_{\leq t}\ (P53/P53i \geq a)\}$$
$$P_{\geq 0.9}(\phi_2') = P_{\geq 0.9}\ \{G_{[20,t_1]}\ (P53/P53i \geq a)\}$$
$$P_{\geq 0.9}(\phi_2'') = P_{\geq 0.9}\ \{F_{\leq t_2}\ (\tfrac{P53}{P53i} \leq b\ \&\ \tfrac{MDM2p}{MDM2i} \geq a)\}$$

Formula ϕ_2 was verified by PRISM: within t (around 20) minutes, P53's expression level will be upregulated to at least $a = 10$ times of its initial value. Formula ϕ_2' was also true when $t_1 = 25$, that is, P53's expression continuously stays at a high level in a short time interval $[20, 25]$, but falsified if t_1 is given a large value. The verification of ϕ_2 and ϕ_2' confirmed our previous SMV results [16] which state that ER stress could promote both apoptosis- and survival-related signal transductions. The verified formula ϕ_2'' claims, prolonged ER stress (within $t_2 = 80$ minutes) will promote the oncoprotein MDM2's phosphorylation ($a = 10$) which can inhibit P53's transcription activity ($b = 1$).

Hypothesis III: In response to the external stimulus or stress, a large amount of NFκB will be released and translocate into the nucleus to promote the transcription of many oncoproteins and tumor suppressors.

$$P_{\geq 0.9}(\phi_3) = P_{\geq 0.9}\ [F_{\leq 30}(NFkBp/NFkBt \geq 0.5)]$$
$$P_{\geq 0.9}(\phi_3') = P_{\geq 0.9}\ [F_{[60,80]}(NFkBp/NFkBt \leq 0.35)]$$
$$P_{\geq 0.9}(\phi_3'') = P_{\geq 0.9}\ [G_{[80,100]}(NFkBp/NFkBt \leq 0.35)]$$

Formula ϕ_3 claims, within 30 minutes, half of $NF\kappa$B molecules will translocate into the nucleus. And, the formula ϕ_3' predicts, within the time interval $[60, 80]$, the percentage of nuclear NFκB will be reduced to a threshold value due to a negative feedback loop regulated by IκB and A20. These two formulas were verified to be true by PRISM. However, the formula ϕ_3'' is falsified, that is, in the long run ([80, 100]), the lower expression level of NFκB in the nucleus is fluctuating instead of stable, indicating the existence of an oscillation phenomena which has been observed in Hoffmann et al.'s experiment [30].

Hypothesis IV: Overexpressed IKK can promote the translocation of NFκB into the nucleus to induce the transcription of Cyclin D which regulates the G1/S phase transition.

We vary the initial value of IKK and check the above formulas in Hypothesis III, the results are presented in Table 1. The verification results confirmed that IKK's verexpression could eventually promote the expression and translocation of NFκB to the nucleus, which plays an important role in the tumorigenesis, inflammation and Alzheimer's disease; while a smaller initial value of IKK could reduce the amount of NFκB molecules in the nucleus. This hypothesis explained why using the IKK inhibitor (e.g., Manumycin A) could inhibit tumorigenesis since NFκB pathway is inhibited, which is consistent with previous studies [11], [15], [16], [21].

Table 1: Verification of Hypothesis IV

IKK(10^3)	2.5	5.0	7.5	10	15	20
$P_{\geq 0.9}(\phi_3)$	False	False	True	True	True	True
$P_{\geq 0.9}(\phi_3')$	True	True	True	True	False	False
$P_{\geq 0.9}(\phi_3'')$	True	True	False	False	False	False

The temporal logic formulas in the Hypothesis I-IV are assertions of some regulatory components' properties. The probability is predefined with $p = 0.9$, and the PRISM model checker will output "True" or "False" only. In many cases, we need directly estimate the probability that a temporal logic property is true. Below, we perform a more accurate estimation of some temporal properties using PRISM.

Estimation 1: What is the probability that the number of Cyclin D and Aβ molecules will be equal to or greater than a threshold value within time t?

$P_{=?}(\phi_1) = P_{=?}\ [F_{\leq t}(CyclinD \geq 500)]$;
$P_{=?}(\phi_1') = P_{=?}\ [F_{\leq t}(A\beta \geq 300)]$.

In Table 2, we present the estimation results varying the time t ($\alpha = \omega$ = 0.05). The increasing probabilities with time indicate that, prolonged ER stress (after around 1 hour) will induce the uncontrolled synthesis and abnormal accumulation of Cyclin D and Amyloid-β molecules. So, targeting the ER stress-induced pathway could possibly find a way to slow the progression of cancer and Alzheimer's disease.

Table 2: Estimation of Hypothesis I						
$time$	30	35	40	45	50	60
$P_{=?}(\phi_1)$	1E-4	0.019	0.31	0.75	0.93	0.99
$P_{=?}(\phi_1')$	0.122	0.984	1.0	1.0	1.0	1.0

Estimation 2: Estimate the probabilities of some properties related to P53's transcription activity and MDM2's phosphorylation due to ER stress in Hypothesis II.

$P_{=?}(\phi_2) = P_{=?}\ \{F_{\leq 20}\ (P53/P53i \geq a)\}$
$P_{=?}(\phi_2') = P_{=?}\ \{G_{[20,25]}\ (P53/P53i \geq a)\}$
$P_{=?}(\phi_2'') = P_{=?}\ \{F_{\leq 80}\ (\frac{P53}{P53i} \leq b\ \&\ \frac{MDM2p}{MDM2i} \geq a)\}$

Table 3 reports the results with different values of a and/or b ($\alpha = \omega$ = 0.05). The estimated probabilities are consistent with the SPRT results, that is, P53's expression level will be upregulated (increase at least 11 times of the initial value) immediately after external stress. However, it will reduce to a lower level (around its initial value) in the long term due to the negative regulation by MDM2 which is upregulated by the ER stress pathways within 80 minutes.

Table 3: Estimation of Hypothesis II						
a	9	10	11	12	13	14
$P_{=?}(\phi_2)$	1.00	0.996	0.97	0.60	0.02	0.0
$P_{=?}(\phi_2')$	0.99	0.93	0.74	0.45	0.02	0.0
b ($a = 10$)	0.5	0.75	1	1.5	2	2.5
$P_{=?}(\phi_2'')$	0.02	0.6	0.92	0.92	0.93	0.93

Estimation 3: Estimate the percentage of NFκB molecules in the nucleus after ER stress, and calculate its probabilities under different conditions in the hypothesis III and IV.

$P_{=?}(\phi_3) = P_{=?}\ [F_{\leq 30}(NFkBp/NFkBt \geq a)]$
$P_{=?}(\phi_3') = P_{=?}\ [F_{[60,T]}(NFkBp/NFkBt \leq 0.35)]$
$P_{=?}(\phi_3'') = P_{=?}\ [G_{[80,100]}(NFkBp/NFkBt \leq 0.35)]$

Table 4 reports the estimated probabilities of the hypothesis III and IV (α = 0.05, N = 100). The results in the first two rows indicate that, the percentage of NFκB in the nucleus hits the peak (a's value is around 0.5$\sim$0.55) within 30 minutes after stress. We can also find the time point when this percentage starts to be lower than a given value through varying T's value (row 3-4), which indicates that, after around 70 minutes, most of NFκB are taken out of the nucleus by its inhibitor IκB and inhibited by A20 due to the negative feedback loops. The last four rows calculate the probabilities of the hypothesis IV with different initial values of IKK (a=0.5). These results confirmed that, overexpressed IKK will promote the translocation of NFκB to the nucleus, that is, $P_{=?}(\phi_3) = 1$. However, if we lower IKK's expression, the number of NFκB molecules in the nucleus will be reduced, that is $P_{=?}(\phi_3')$ and $P_{=?}(\phi_3'')$ will be close to 1. These results are consistent with the qualitative assertations using SPRT method.

Table 4: Estimation of Hypothesis III and IV					
a	0.58	0.56	0.55	0.52	0.5
$P_{=?}(\phi_3)$	0.04	0.52	0.85	1.0	1.0
T	63	65	67	70	75
$P_{=?}(\phi_3')$	0.38	0.61	0.82	0.95	1.0
IKK ($\times 10^3$)	5.0	6.0	7.5	10	15
$P_{=?}(\phi_3)$	0.45	0.99	1.0	1.0	1.0
$P_{=?}(\phi_3')$	1.0	1.0	1.0	1.0	0.88
$P_{=?}(\phi_3'')$	0.99	0.89	0.25	0.0	0.0

Estimation 4: Estimate the conditional probability $P_{=?}(\phi_2''|\phi_2''') = \frac{P_{=?}\ \{F_{\leq 80}\ (\frac{P53}{P53i} \leq b\ \&\ \frac{MDM2p}{MDM2i} \geq 10)\}}{P_{=?}\ \{F_{\leq 80}\ (\frac{MDM2p}{MDM2i} \geq 10)\}}$

Table 5: Conditional Probability						
b	0.5	0.75	1	1.5	2	2.5
$P_{=?}(\phi_2''|\phi_2''')$	0.02	0.63	0.96	0.96	0.97	0.97

Using the results in Table 3 and $P_{=?}(F_{\leq 80}\ (\frac{MDM2p}{MDM2i} \geq 10)) = 0.957$ ($\alpha = \omega = 0.05$), Table 5 estimates the conditional probabilities varying b's initial values.

V. Conclusion

In this work, we developed a stochastic model to describe the ER stress-induced signaling pathway which is implicated in the pathogenesis of a variety of diseases. Then, the probabilistic model checker PRISM, based on sequential probability ratio test and confidence interval estimation methods, is applied to formally analyze some important quantitative properties of the system described by the temporal logic formulas, which are difficult to be addressed using the traditional simulation approaches. Our recent work [16] proposed a discrete value model and applied the asynchronous and synchronous symbolic model verification (SMV) technique to study the roles of ER-Golgi network in the cell cycle progression. Compared with the SMV study, this work could not only qualitatively analyze the the complex biological system, but also provide a "quantitative" estimation of some properties. However, we remark that, since many reactions and parameters in our model are unknown or indirectly estimated or tuned from existing experiments, the current model can only qualitatively compare with the experimental behavior. However, it still provides valuable information about the ER stress-induced signaling pathway in the cell cycle and some diseases.

Probabilistic verification of the temporal logic formulas using hypothesis testing and estimation methods shows that, overexpressed ER transmembrane sensor proteins will promote the expression of Cyclin D, Amyloid-β and MDM2, leading to the pathogenesis of cancer and AD in the long term, and inhibition of the apoptosis. ER stress could upregulate the expression and promote the translocation of NFκB to the nucleus, and its expression level is fluctuating due to the negative feedback loops composed of IKK, IκB, NFκB and A20, which

has been observed in Hoffmann et al's experiment [30]. These results are consistent with our recent SMV studies [16] and previous experimental work. Our work continues to verify that, varying IKK's initial value will influence the number of NFκB molecules in the nucleus, which explained why IKK inhibitor could inhibit tumorigenesis.

The proposed signaling pathway model in this work is composed of three signaling branches initiated by ER-resident transmembrane proteins due to ER stress, and the model construction is based on literature searching. ER and its neighboring compartment Golgi apparatus regulate many signaling and secretory pathways whose deregulation is implicated in several diseases. Since more gene expression data are publicly available, in our future work, we will combine the statistical inference algorithms with the model checking techniques to automatically infer and verify the signaling pathways, comprehensively investigate the roles of ER stress in the pathogenesis of cancer and other diseases.

ACKNOWLEDGMENT

This work was partially supported by HG's new faculty start-up grant and President Research Fund award (230152) from the Saint Louis University.

REFERENCES

[1] H. Kadowaki and H. Nishitoh, "Signaling pathways from the endoplasmic reticulum and their roles in disease," *Genes*, 2013.

[2] D. Wlodkowic, J. Skommer *et al.*, "Er-golgi network - a future target for anti-cancer therapy," *Leuk Res*, 2009.

[3] T. T. Wu, H. Gong, and E. M. Clarke, "A transcriptome analysis by lasso penalized cox regression for pancreatic cancer survival," *Journal of Bioinformatics and Computational Biology*, vol. 9, p. 63, 2011.

[4] H. Gong, T. T. Wu, and E. M. Clarke, "Pathway-gene identification for pancreatic cancer survival via doubly regularized cox regression," *BMC Systems Biology*, vol. 8, 2014.

[5] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, pp. 631–643, 2005.

[6] Y. Luan and H. Li, "Group additive regression models for genomic data analysis," *Biostatistics*, vol. 9, pp. 100–113, 2008.

[7] T. T. Wu, Y. F. Chen, T. Hastie, E. Sobel, and K. Lange, "Genomewide association analysis by lasso penalized logistic regression," *Bioinformatics*, vol. 25, pp. 714–721, 2009.

[8] S. Ma, X. Song, and J. Huang, "Supervised group lasso with applications to microarray data analysis," *BMC Bioinformatics*, vol. 8, pp. 60–76, 2007.

[9] T. T. Wu and S. Wang, "Doubly regularized cox regression for high-dimensional survival data with group structures," *Statisstics and Its Interface*, vol. 6, pp. 175–186, 2013.

[10] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, pp. 49–67, 2006.

[11] H. Gong, Q. Wang, P. Zuliani, M. T. Lotze, J. R. Faeder, and E. M. Clarke, "Symbolic model checking of the signaling pathway in pancreatic cancer," *Proceedings of the International Conference on Bioinformatics and Computational Biology (BICoB)*, 2011.

[12] P. Zuliani, A. Platzer, and E. M. Clarke, "Bayesian statistical model checking with application to simulink/stateflow verification," in *HSCC*, 2010, pp. 243–252.

[13] H. Gong, P. Zuliani, and E. Clarke, "Model checking of a diabetes-cancer model," *3rd International Symposium on Computational Models for Life Sciences*, 2011.

[14] E. M. Clarke, J. R. Faeder, L. A. Harris, C. J. Langmead, A. Legay, and S. K. Jha, "Statistical model checking in biolab: Applications to the automated analysis of t-cell receptor signaling pathway," vol. 5307, pp. 231–250, 2008.

[15] H. Gong, "Analysis of intercellular signal transduction in the tumor microenvironment," *BMC Systems Biology*, vol. 7, p. S5, 2013.

[16] H. Gong and L. Feng, "Computational analysis of the roles of er-golgi network in the cell cycle," *BMC Bioinformatics*, vol. 15, p. S12:S18, 2014.

[17] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comp. Phys.*, vol. 22, no. 4, pp. 403–434, 1976.

[18] H. Gong, H. Sengupta, A. Linstedt, and R. Schwartz, "Simulated de novo assembly of golgi compartments by selective cargo capture during vesicle budding and targeted vesicle fusion," *Biophysical Journal*, vol. 95, pp. 1674–1688, 2008.

[19] H. Gong, Y. Guo, A. Linstedt, and R. Schwartz, "Discrete, continuous, and stochastic models of protein sorting in the golgi apparatus," *Physical Review E*, vol. 81, no. 1, p. 011914, 2010.

[20] H. Gong, P. Zuliani, A. Komuravelli, J. R. Faeder, and E. M. Clarke, "Analysis and verification of the HMGB1 signaling pathway," *BMC Bioinformatics*, vol. 11, no. 7, 2010.

[21] H. Gong, P. Zuliani, A. Komuravelli, J. Faeder, and E. Clarke, "Computational modeling and verification of signaling pathways in cancer," *Proceedings of Algebraic and Numeric Biology, LNCS*, vol. 6479, 2012.

[22] G. Hotamisligil, "Endoplasmic reticulum stress and the inflammatory basis of metabolic disease," *Cell*, vol. 140, 2010.

[23] D. G. Breckenridge, "Regulation of apoptosis by endoplasmic reticulum pathways," *Oncogene*, vol. 22, 2003.

[24] D. T. Rutkowski and R. J. Kaufman, "A trip to the er: coping with stress," *Trends in Cell Biology*, vol. 14, 2004.

[25] C. Pereira, "Crosstalk between endoplasmic reticulum stress and protein misfolding in neurodegenerative diseases," *ISRN Cell Biology*, 2013.

[26] A. Salminen *et al.*, "Er stress in alzheimer's disease: a novel neuronal trigger for inflammation and alzheimer's pathology," *Journal of Neuroinflammation*, vol. 6, 2009.

[27] A. Yagi *et al.*, "Gadd34 induces p53 phosphorylation and p21/waf1 transcription," *J Cell Biochem*, vol. 90(6), pp. 1242–9, 2003.

[28] M. de la Vega C *et al.*, "Possible mechanisms involved in the down-regulation of translation during transient global ischaemia in the rat brain," *Biochem J.*, vol. 357, 2001.

[29] J. R. van Beijnum, W. A. Buurman, and A. W. Griffioen, "Convergence and amplification of toll-like receptor (tlr) and receptor for advanced glycation end products (rage) signaling pathways via high mobility group b1," *Angiogenesis*, vol. 11, pp. 91–99, 2008.

[30] A. Hoffmann, A. Levchenko, M. Scott, and D. Baltimore, "The IκB-NFκB signaling module: Temporal control and selective gene activation," *Science*, vol. 298, pp. 1241–1245, 2002.

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

[32] M. Kwiatkowska, G. Norman, and J. Sproston, "Probabilistic model checking of deadline properties in the ieee1394 firewire root contention protocol," *Formal Aspects of Computing*, vol. 14, pp. 295–318, 2003.

[33] M. Kwiatkowska, G. Norman, and D. Parker, "Probabilistic model checking in practice: Case studies with prism," *ACM SIGMETRICS Performance Evaluation Review*, vol. 32, pp. 16–21, 2005.

[34] Prism code for er stress-induced pathway model. [Online]. Available: http://cs.slu.edu/∼gong/research/ERprism.zip

[35] S. Mokhtar *et al.*, "The beta-amyloid protein of alzheimers disease: Communication breakdown by modifying the neuronal cytoskeleton," *International Journal of Alzheimers Disease*, vol. 2013, 2013.

[36] B. Liu *et al.*, "Endoplasmic reticulum stress promotes amyloid-beta peptides production in rgc-5 cells," *Cell Stress and Chaperones*, 2014.