

# Model Simplification of Signal Transduction Pathway Networks via a Hybrid Inference Strategy<sup>\*</sup>

Jianfang Jia<sup>\*</sup> Hong Yue<sup>\*\*</sup>

<sup>\*</sup> North University of China, Taiyuan 030051, China (e-mail: jiajf2002@163.com).

<sup>\*\*</sup> University of Strathclyde, Glasgow G1 1QE, UK (e-mail: hong.yue@eee.strath.ac.uk).

---

**Abstract:** A full-scale mathematical model of cellular networks normally involves a large number of variables and parameters. How to effectively develop manageable and reliable models is crucial for effective computation, analysis and design of such systems. The aim of model simplification is to eliminate parts of a model that are unimportant for the properties of interest. In this work, a model reduction strategy via hybrid inference is proposed for signal pathway networks. It integrates multiple techniques including conservation analysis, local sensitivity analysis, principal component analysis and flux analysis to identify the reactions and variables that can be considered to be eliminated from the full-scale model. Using an I $\kappa$ B-NF- $\kappa$ B signalling pathway model as an example, simulation analysis demonstrates that the simplified model quantitatively predicts the dynamic behaviours of the network.

---

## 1. INTRODUCTION

Cell signal transduction networks are composed of a number of signal transduction pathways through the cross talking between pathways. The complexity of signalling pathways is not directly understandable, and mathematical modelling provides a way for comprehensive analysis of such systems (Heinrich and Schuster, 1996; Asthagiri and Lauffenburger, 2000). With the development of molecular biology and high throughput experimental techniques, a large number of data sets have been obtained which makes it possible to study cell signalling transduction networks quantitatively (Bhalla, 2002; Neves and Iyengar, 2002). It is then expected that mathematical models of signal transduction networks can be properly developed, and the dynamic behaviours of complex signalling systems can be analyzed based on manageable and reliable models. However, since the inner structure and functions of signalling transduction networks are, in most cases, not fully understood, the mathematical models developed according to the mechanism analysis, experimental data and some hypothesis, normally consist of a large number of variables and kinetics parameters which are not suitable for theoretical and computational analysis. On the other hand, the biochemical network contains strong nonlinearity, temporal behaviour, and the number of the parameters increases dramatically with the increase of the model complexity.

Functional NF- $\kappa$ B exists in dimeric form that is composed of combinations of five proteins, cRel, RelA, RelB, p50 and p52 (Hoffmann and Baltimore, 2006). Several versions of the computational models have been published in recent years for this pathway. An early model (Hoffmann et al., 2002), which measured the behaviour of cell populations

of the TNF $\alpha$ -induced NF- $\kappa$ B signal pathway, can be described by 24 state variables (components with distinct nuclear and cytoplasmic localizations) and 64 reactions (Ihekwaba et al., 2004). Using this model, the impact of the I $\kappa$ B $\alpha$ -mediated negative feedback on oscillations of NF- $\kappa$ B activity can be observed and analyzed. Imaging studies by microscopy on single cell real-time data provided further evidence to the oscillation behavior in this system (Nelson et al., 2004). Further investigations revealed that I $\kappa$ B $\epsilon$  also mediates functional negative feedback on NF- $\kappa$ B activity but with a time delay and it dampens the I $\kappa$ B $\alpha$ -induced oscillations (Kearns et al., 2006). It is also discovered that different inflammatory stimuli induce distinct IKK profiles (Werner et al., 2005). In the extended models based on these new discoveries, several degradation reactions were added, which leads to 70 reactions in the model. When NF- $\kappa$ B inducible mRNA syntheses of I $\kappa$ B $\epsilon$  and  $-\beta$  were added, the model consists of 72 reactions in total (Werner et al., 2005; Kearns et al., 2006). In the work of (Lipniacki et al., 2004), the protein A20 was also considered as an inhibitor of IKK, which introduced several new variables in the model. Some other modifications can be found in literature when certain extra issues need to be addressed (Covert et al., 2005; Cheong et al., 2006; O'Dea et al., 2007). The literature study shows that for such a complex cell network, a number of models are presented according to different understanding of the reaction mechanisms. It is therefore essential to study how to achieve a simple yet reliable 'core' model without losing basic dynamic properties.

Various approaches have been developed for simplifying biological systems such as lumping methods, timescale-based techniques, model reduction based on sensitivity analysis, and balanced truncation method, etc. To name but a few examples, a framework for model reduction of

---

<sup>\*</sup> This work was supported in part by the National Natural Science Foundation of China #30770560.

Wnt/ $\beta$ -catenin pathway was proposed based on the different time scales of biochemical reaction processes (Kruger and Heinrich, 2004). A heuristic approach was developed combining simulation studies and system analysis, and applied to the model of EGF receptor signalling. The complex networks was decomposed into modules with low retroactivity in order to gain the reduced model through the input/output analysis (Conzelmann et al., 2004). When combined with flux analysis, parametric sensitivity analysis can also allow model reduction, which is used to the EGF mediated signalling (Liu et al., 2005). For complex metabolism models, sensitivity analysis based on parameter tuning importance and principle component analysis was used for model reduction (Degenring et al., 2004). In recent years, to implement the simplification of biochemical models, some new approaches have been developed from the viewpoint of systems engineering and control theory. A systematic framework was presented with a step-by-step bottom-up methodology to guide the development of modular reduced-order signalling pathway components (Maurya et al., 2006). A domain-oriented method has been proposed for reducing non-linear kinetic models of signal transduction networks, in which a large class of signalling networks was modelled by means of macro-states instead of micro-states (Conzelmann et al., 2006).

It is recognized that model simplification is not only important for effective quantitative comprehension of systems, but may also provide new insights into signal pathways such as identifying model robustness. Due to the complexity of cellular networks, most of the current methods were proposed targeting certain type of networks. No single method is superior in all cases as the efficiency of model simplification is related to both model complexity and the objective of the simplification (Dano et al., 2006). In this work, we intend to develop a systematic approach to simplify signal pathway models. This method integrates conservation analysis, local sensitivity analysis (LSA), principal component analysis (PCA) and flux analysis of signal pathway models. The outline of the paper is organized as follows. Section 1 gives the research background and motivation. Section 2 describes the hybrid inference method. This approach is demonstrated through a case study on an I $\kappa$ B-NF- $\kappa$ B signalling pathway model in Section 3. Conclusions are given in Section 4.

## 2. HYBRID INFERENCE METHOD

### 2.1 Conservation Analysis

Consider a biochemical model consisting of  $n$  chemical species and  $m$  reactions. The network can be modelled by the following ordinary differential equation (ODE)

$$dX(t)/dt = Nv(X(t), p, t), \quad X(0) = X_0. \quad (1)$$

The state vector  $X \in \mathbb{R}^n$  is composed of the concentrations of each reaction species. The constant  $r$ -vector  $p$  is composed of the parameters of interest in the model.  $r$  is normally taken to be the same as  $m$  in mass-action reaction systems. The function  $v$  describes the rate of each reaction as a function of species concentrations and parameter values.  $N \in \mathbb{R}^{n \times m}$  is the stoichiometry matrix of the networks.

To consider structural conservations in the network, linear dependencies inherent in the state variables need to be identified before performing further analysis. Denoting  $k$  as the row rank of the stoichiometry matrix  $N$ , if  $k = n$ , then all the state variables are independent. If  $k < n$ , then the matrix  $N$  can be written as

$$N = LN_R \quad (2)$$

where  $N_R \in \mathbb{R}^{k \times m}$  is the reduced stoichiometry matrix composed of the  $k$  linearly independent rows. The link matrix  $L \in \mathbb{R}^{n \times k}$  has the form of  $L = \begin{bmatrix} I_k \\ L_0 \end{bmatrix}$ , where  $I_k$  denotes the  $k$ -dimensional identity matrix, and  $L_0 \in \mathbb{R}^{(n-k) \times k}$ .

Since a structural conservation allows one species concentration to be determined as a function of the others, the states vector  $X$  can be decomposed into an independent species vector  $\tilde{X}$  and a dependent species vector  $\hat{X}$  (Ingalls and Sauro, 2003). Denoting  $X = \begin{bmatrix} \tilde{X} \\ \hat{X} \end{bmatrix}$ , where  $\tilde{X} \in \mathbb{R}^k$  and  $\hat{X} \in \mathbb{R}^{n-k}$ , (1) is reformulated as

$$\frac{d}{dt} \begin{bmatrix} \tilde{X}(t) \\ \hat{X}(t) \end{bmatrix} = \begin{bmatrix} I_k \\ L_0 \end{bmatrix} N_R v(X(t), p, t). \quad (3)$$

It follows that

$$d\hat{X}(t)/dt = L_0 N_R v(X(t), p, t) = L_0 d\tilde{X}(t)/dt. \quad (4)$$

Integrating both sides of (4) for all  $t \geq 0$ , the dependent species vector  $\hat{X}$  can be represented as

$$\begin{aligned} \hat{X}(t) &= L_0 \tilde{X}(t) - L_0 \tilde{X}(0) + \hat{X}(0) \\ &= L_0 \tilde{X}(t) + T \end{aligned} \quad (5)$$

where  $T = \hat{X}(0) - L_0 \tilde{X}(0)$  is an  $(n - k)$ -dimensional constant vector that is determined by the initial conditions of states. The algebraic equation (AE) (5) shows that the concentration trajectory of a dependent species can be represented by independent species concentrations.

As a consequence of the conservation analysis, the system model (1) is decomposed into two parts: the ODEs of the independent variables and the AEs of the dependent variables

$$d\tilde{X}(t)/dt = N_R v(\tilde{X}(t), \hat{X}(t), p, t) \quad (6)$$

$$\hat{X}(t) = \tilde{X}(t) + T \quad (7)$$

### 2.2 Local Sensitivity Analysis

Sensitivity analysis is used to analyze the variation of the outputs of a system with respect to the changes in parameters, inputs or initial conditions. Parametric sensitivities describe the importance of model parameters to the measurement variables, and it plays an important role in data analysis, parameter estimation, model validation and experimental design, etc. Based on the model decomposition from conservation analysis, in this section, LSA is performed to analyze the independent model of the biochemical networks.

According to the original system model (1), the derivatives of the absolute sensitivity function  $S = \partial X / \partial p$  is represented as

$$\frac{d}{dt} S(t) = \frac{d}{dt} \frac{\partial X(t)}{\partial p} = N \left( \frac{\partial v(t)}{\partial X(t)} \frac{\partial X(t)}{\partial p} + \frac{\partial v(t)}{\partial p} \right). \quad (8)$$

In the following, time  $t$  in bracket is omitted for representation convenience. When focusing on the independent variables, only the sensitivity function of  $\tilde{X}$  with respect to  $p$  will be considered, i.e.

$$\begin{aligned} \frac{d}{dt} \tilde{S} &= \frac{d}{dt} \frac{\partial \tilde{X}}{\partial p} \\ &= N_R \left[ \frac{\partial v}{\partial \tilde{X}} \frac{\partial \tilde{X}}{\partial p} + \frac{\partial v}{\partial \tilde{X}} \frac{\partial}{\partial p} (L_0 \tilde{X} + T) + \frac{\partial v}{\partial p} \right] \\ &= N_R \left[ \left( \frac{\partial v}{\partial \tilde{X}} + \frac{\partial v}{\partial \tilde{X}} L_0 \right) \frac{\partial \tilde{X}}{\partial p} + \frac{\partial v}{\partial \tilde{X}} \frac{\partial T}{\partial p} + \frac{\partial v}{\partial p} \right] \\ &= N_R \left[ \frac{\partial v}{\partial \tilde{X}} L \frac{\partial \tilde{X}}{\partial p} + \frac{\partial v}{\partial \tilde{X}} \frac{\partial T}{\partial p} + \frac{\partial v}{\partial p} \right]. \end{aligned} \quad (9)$$

Since  $T$  is a constant,  $\partial T / \partial p = 0$ . For the  $j$ -th parameter,  $\tilde{S}_j = \partial \tilde{X} / \partial p_j = [\tilde{s}_{1,j} \ \tilde{s}_{2,j} \ \cdots \ \tilde{s}_{k,j}]^T$  is a column sensitivity vector, then from (9) we have

$$\frac{d\tilde{S}_j}{dt} = N_R \frac{\partial v}{\partial \tilde{X}} L \tilde{S}_j + N_R \frac{\partial v}{\partial p_j}. \quad (10)$$

The initial conditions of  $\tilde{s}_{i,j}$  can be represented as  $\tilde{s}_{i,j}(0) = \delta(p_j - \tilde{x}_i(0))$ , where  $\delta$  is the Kronecker delta function,  $\tilde{x}_i(0)$  is the initial value of the  $i$ -th independent state variable. By solving (6) and (10) in parallel, the sensitivity column vector  $\tilde{S}_j$  and the independent state vector  $\tilde{X}$  can be obtained simultaneously.

To allow direct comparison of responses at different species and across different parameters, the scaled sensitivity is used for further analysis

$$\bar{S}_j = \frac{\partial \tilde{X} / \tilde{X}}{\partial p_j / p_j} = \frac{\partial \ln \tilde{X}}{\partial \ln p_j}. \quad (11)$$

For a system with  $r$  parameters in consideration, the scaled sensitivity matrix is formed as  $\bar{S} = [\bar{s}_1, \dots, \bar{s}_r]$ .

### 2.3 Principal Component Analysis

Principal component analysis is a classical statistical method, which has been widely used in data analysis and dimensionality reduction. PCA involves a mathematical procedure that transforms a number of correlated variables into a number of uncorrelated variables called principal components. The aim of PCA is to reduce data dimension by performing a covariance analysis between factors.

In order to study the effect of the parameter variations on the state variables, the product matrix  $\bar{S}^T \bar{S}$  is diagonalized using its eigenvalues and eigenvectors as follows:

$$\bar{S}^T \bar{S} = U \Lambda U^T \quad (12)$$

where  $U = [u_{j,i}] \in \mathbb{R}^{r \times r}$  denotes the eigenvectors matrix. The diagonal matrix  $\Lambda$  is composed of the individual eigenvalues  $\lambda_i$  of the matrix  $\bar{S}^T \bar{S}$ , i.e.,  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$ .

In such an analysis, the eigenvalue provides an absolute measure of the significance of certain parts of the biological system that is composed of strongly coupled reactions. Each eigenvector is a linear combination of reactions, and the relative magnitude of the elements of each eigenvector measures the relative importance of each reaction for the corresponding eigenvalue. Thus, the following PCA parameter  $e_j$  can be used to measure the importance of the  $j$ -th reaction (Liu et al., 2005):

$$e_j = \sum_{l=1}^r \lambda_l u_{j,l} / \sum_{l=1}^r \lambda_l \quad (13)$$

Since PCA quantitatively weights the parametric effects on the network features, reactions can be classified into 'essential' and 'non-essential' sets based on the PCA parameter. The threshold value for this classification is chosen to be

$$\hat{e} = \alpha \cdot \max \{|e_j|, j = 1, \dots, r\}. \quad (14)$$

where  $\alpha$  ( $0 \leq \alpha \leq 1$ ) is termed the critical reduction factor. For the  $j$ -th reaction, if  $|e_j| \geq \hat{e}$ , it is classified as 'essential', otherwise 'non-essential'. When  $\alpha = 0$ , all the reactions are taken to be 'essential' and  $\alpha = 1$  takes all reactions to be 'non-essential'. The size of 'essential' and 'non-essential' reactions can thus be adjusted by changing the value of  $\alpha$ . It should be noted that not all 'non-essential' reactions can be eliminated from the model. Removal of those 'non-essential' reactions that have relative high flux in the biological systems may exert serious influence on the model output. Therefore, flux analysis is undertaken to evaluate the flux measure of each reaction in the model.

### 2.4 Flux Analysis

Flux analysis is an established methodology that allows the quantification of intracellular fluxes. The average reaction flux of the  $j$ -th reaction is represented as

$$\bar{f}_j = \int_{t_0}^{t_p} |v_j| dt \quad (15)$$

where  $t_0$  and  $t_p$  are the initial and terminal time for process simulation. Supposing that the set of the average reaction flux of all 'essential' reactions is  $\{\bar{f}_{\text{essential}}\}$ , the threshold flux value can be calculated as

$$\hat{f} = \beta \cdot \min \{\bar{f}_{\text{essential}}\} \quad (16)$$

where  $\beta$  ( $0 \leq \beta \leq 1$ ) is termed the critical flux factor.

In this simplification method, only those reactions that have both low sensitivities ('non-essential') and low flux measures can be removed from the ODE model. For the 'non-essential' reactions, if its average reaction flux is smaller than the critical flux factor  $\hat{f}$ , it can be considered to be eliminated from the biological model due to its relative little influence on the system output. If, on the contrary, the average reaction flux is larger than  $\hat{f}$ , then the reaction should be retained in the model even though it's identified as 'non-essential'.

The implementation of this hybrid inference strategy is summarized as follows.

(1) Decompose the system model into ODEs for independent variables and AEs for dependent variables through conservation analysis. The dimension of the dynamic model is reduced if there are structural conservations in the network.

(2) Perform LSA for the model with independent variables, which shows the relative importance of the impact of each reaction on the system output.

(3) By selecting the appropriate critical reaction factor  $\alpha$ , classify the reactions into 'essential' and 'non-essential' groups through PCA of the LSA results.

(4) Implement flux analysis to all reactions and set the threshold flux value from the 'essential' reactions. Among those 'non-essential' reactions, further check if any of them has an average reaction flux that is lower than the threshold value. If this is the case, the reaction can be eliminated from the model.

### 3. CASE STUDY: MODEL SIMPLIFICATION OF A SIGNAL PATHWAY MODEL

The  $I\kappa B$ -NF- $\kappa B$  model for study has 24 time-varying reaction species and 64 reactions (Hoffmann et al., 2002). The 24 states are defined as follows: IKK ( $x_1$ ), IKKI $\kappa B\alpha$  ( $x_2$ ), IKKI $\kappa B\beta$  ( $x_3$ ), IKKI $\kappa B\varepsilon$  ( $x_4$ ), IKKI $\kappa B\alpha$ -NF- $\kappa B$  ( $x_5$ ), IKKI $\kappa B\beta$ -NF- $\kappa B$  ( $x_6$ ), IKKI $\kappa B\varepsilon$ -NF- $\kappa B$  ( $x_7$ ), NF- $\kappa B$  ( $x_8$ ), NF- $\kappa B_n$  ( $x_9$ ), I $\kappa B\alpha$  ( $x_{10}$ ), I $\kappa B\beta$  ( $x_{11}$ ), I $\kappa B\varepsilon$  ( $x_{12}$ ), I $\kappa B\alpha_n$  ( $x_{13}$ ), I $\kappa B\beta_n$  ( $x_{14}$ ), I $\kappa B\varepsilon_n$  ( $x_{15}$ ), I $\kappa B\alpha$ -t ( $x_{16}$ ), I $\kappa B\beta$ -t ( $x_{17}$ ), I $\kappa B\varepsilon$ -t ( $x_{18}$ ), I $\kappa B\alpha$ -NF- $\kappa B$  ( $x_{19}$ ), I $\kappa B\beta$ -NF- $\kappa B$  ( $x_{20}$ ), I $\kappa B\varepsilon$ -NF- $\kappa B$  ( $x_{21}$ ), I $\kappa B\alpha_n$ -NF- $\kappa B_n$  ( $x_{22}$ ), I $\kappa B\beta_n$ -NF- $\kappa B_n$  ( $x_{23}$ ), I $\kappa B\varepsilon_n$ -NF- $\kappa B_n$  ( $x_{24}$ ). Subscript  $n$  means the protein is in the nucleus.

For this model, the row rank of the stoichiometry matrix  $N$  is calculated to be 23. Then, the matrix  $N$  can be written as  $N = LN_R$ ,  $L \in \mathbb{R}^{24 \times 23}$ ,  $N_R \in \mathbb{R}^{23 \times 64}$ . As a result, the states vector  $X$  is decomposed into the independent vector  $\tilde{X} = [x_1 \ x_2 \ \dots \ x_{23}]^T$  and the dependent vector  $\hat{X} = x_{24}$ . The system model can be rewritten in the form of (6) for the independent variables, and

$$\hat{X}(t) = x_{24}(t) = L_0 \tilde{X} + T = -x_5 - x_6 - x_7 - x_8 - x_9 - x_{19} - x_{20} - x_{21} - x_{22} - x_{23} + T \quad (17)$$

for the dependent variable.

In the next stage, LSA was undertaken for the independent states model. Here the concentration of the NF- $\kappa B$  in the nucleus ( $x_9$ ) was selected as the output of interest because it is an important protein in interpreting the behaviour of this signal pathway network (Hoffmann et al., 2002). The initial conditions of the states were taken from the equilibrium with  $x_1 = 0.1\mu M$  as an activation input (IKK). The initial conditions of the dynamic sensitivities are set to be zeros. The simulation time is taken to be 400 minutes and the sample frequency is 1 per minute during calculation. Once the sensitivity column vector  $\tilde{S}_9$  is obtained, taking principle component analysis by (12) and (13), the PCA parameters  $e_j$  ( $j = 1, 2, \dots, 64$ ) can be obtained. The results of the sensitivity rankings and the PCA parameters are omitted here due to page limit.

To separate the 64 reactions into 'essential' and 'non-essential' groups, the critical reduction factor  $\alpha$  was taken

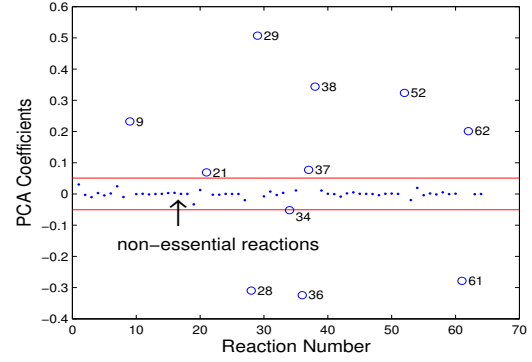


Fig. 1. Classify reactions by PCA: circle means 'essential', dot means 'non-essential'

to be 0.1. The results are shown in Fig. 1, in which there are 11 'essential' reactions and 53 'non-essential' reactions.

In the next flux analysis, the average reaction flux of all reactions were calculated by (15). Taking the critical flux factor  $\beta$  to be 0.1, the threshold flux value was calculated by (16). It is identified that out of the 53 'non-essential' reactions, the following 15 reactions have flux values less than the threshold value:

- $v_{10}$ : IKKI $\kappa B\beta$  + NF- $\kappa B \rightarrow$  IKKI $\kappa B\beta$ -NF- $\kappa B$
- $v_{11}$ : IKKI $\kappa B\beta$ -NF- $\kappa B \rightarrow$  NF- $\kappa B$  + IKKI $\kappa B\beta$
- $v_{13}$ : IKKI $\kappa B\varepsilon$  + NF- $\kappa B \rightarrow$  IKKI $\kappa B\varepsilon$ -NF- $\kappa B$
- $v_{14}$ : IKKI $\kappa B\varepsilon$ -NF- $\kappa B \rightarrow$  IKKI $\kappa B\varepsilon$  + NF- $\kappa B$
- $v_{17}$ : I $\kappa B\beta$ -NF- $\kappa B \rightarrow$  NF- $\kappa B$
- $v_{18}$ : I $\kappa B\varepsilon$ -NF- $\kappa B \rightarrow$  NF- $\kappa B$
- $v_{24}$ : I $\kappa B\beta_n$ -NF- $\kappa B_n \rightarrow$  I $\kappa B\beta_n$  + NF- $\kappa B_n$
- $v_{26}$ : I $\kappa B\varepsilon_n$ -NF- $\kappa B_n \rightarrow$  I $\kappa B\varepsilon_n$  + NF- $\kappa B_n$
- $v_{30}$ : source  $\rightarrow$  I $\kappa B\beta$ -t
- $v_{31}$ : I $\kappa B\beta$ -t  $\rightarrow$  sink
- $v_{32}$ : source  $\rightarrow$  I $\kappa B\varepsilon$ -t
- $v_{33}$ : I $\kappa B\varepsilon$ -t  $\rightarrow$  sink
- $v_{45}$ : I $\kappa B\beta_n \rightarrow$  I $\kappa B\beta$ (Export)
- $v_{47}$ : IKKI $\kappa B\varepsilon \rightarrow$  IKK + I $\kappa B\varepsilon$
- $v_{51}$ : I $\kappa B\varepsilon_n \rightarrow$  I $\kappa B\varepsilon$ (Export)

All these reactions relate to I $\kappa B\beta$  and I $\kappa B\varepsilon$ . This agrees with the biophysical mechanisms in the model since it focuses on the negative feedback mediated by I $\kappa B\alpha$ . By removing the above 15 reactions, 2 states,  $x_{17}$  (I $\kappa B\beta$ -t) and  $x_{18}$  (I $\kappa B\varepsilon$ -t), were consequently eliminated from the model. The simplified model therefore includes 49 reactions and 22 states. The time courses of several important variables of the original model and the reduced model are compared in Fig. 2. It can be seen that the simplified model matches well with the complete model. In all the species relating to I $\kappa B\alpha$ , the difference can hardly be noticed via visual inspection, while only in the concentrations of IKKI $\kappa B\beta$  and IKKI $\kappa B\varepsilon$ , the small difference can be observed. This result shows that by eliminating certain reactions through the proposed hybrid inference algorithm, the major characteristics of the NF- $\kappa B$  model are well-preserved.

In this hybrid inference simplification strategy, the factor  $\alpha$  corresponds to the PCA classification based on sensitivity analysis, and  $\beta$  corresponds to the selection based on flux analysis. From the threshold definitions in (14) and

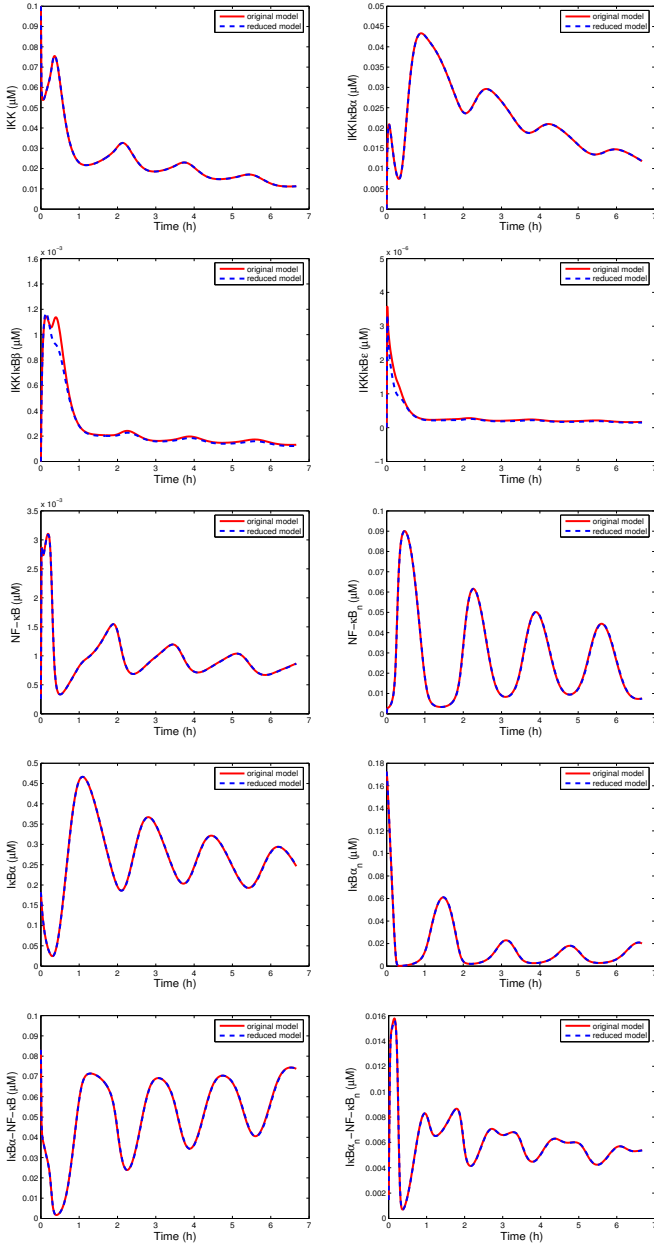


Fig. 2. Selected concentration profiles in the full-scale and simplified models

(16), both  $\alpha$  and  $\beta$  are set to the  $[0,1]$  interval. Within this range, increasing  $\alpha$  will increase the number of ‘non-essential’ reactions. Under the same value of  $\alpha$ , when  $\beta$  is increased, the threshold range to remove low-flux reactions is enlarged and the number of the reduced reactions is likely to be increased. The effects of  $\alpha$  and  $\beta$  on the model reduction can be illustrated by the change of  $\text{NF-}\kappa\text{B}_n$  in Fig. 3. Some numerical calculation results are listed in Table 1, in which NER and NNER are the number of ‘essential’ and ‘non-essential’ reactions, NRR stands for the number of reduced reactions, NRV is the number of reduced variables, and  $Res$  is the residual error calculated by

$$Res = \int_{t_0}^{t_p} (x_9^{orig}(t) - x_9^{redu}(t))^2 dt \quad (18)$$

where  $x_9^{orig}(t)$  and  $x_9^{redu}(t)$  denote the original and the reduced model output at sample time  $t$ .

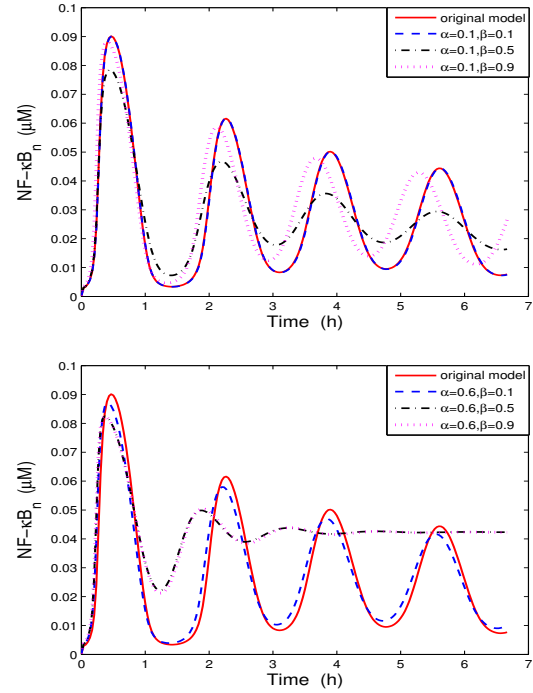


Fig. 3. Oscillations of  $\text{NF-}\kappa\text{B}$  in the nucleus with different values of  $\alpha$  and  $\beta$

Table 1. Model reduction results

$\alpha$	$\beta$	NER	NNER	NRR	NRV	$Res$
0.1	0.1	11	53	15	2	$2.1638e-5$
0.1	0.5	11	53	32	5	$4.3761e-4$
0.1	0.9	11	53	40	10	$5.7319e-4$
0.6	0.1	5	59	37	8	$2.1627e-4$
0.6	0.5	5	59	47	13	$1.1077e-3$
0.6	0.9	5	59	49	13	$1.1012e-3$

It can be seen from the simulation analysis that both  $\alpha$  and  $\beta$  have impacts on the model reduction results, which means that each analysis effort in the hybrid inference strategy plays its role in assuring the quality of model simplification. It should be noted that in the current method, the selection of  $\alpha$  and  $\beta$  is still heuristic.

#### 4. CONCLUSIONS

A hybrid inference method has been proposed to simplify signal transduction pathway models that include a large number of reactions and reaction species. When applying this algorithm to an  $\text{I}\kappa\text{B-NF-}\kappa\text{B}$  signal pathway model, it is found that when the critical reaction factor  $\alpha$  and the critical flux factor  $\beta$  are properly designed, the temporal output of the simplified model highly agrees with that of the complete model. To simplify a complex biological system model, a good understanding of the biophysical mechanisms is crucial, and sometimes, may be more efficient. Taking this signal pathway as an example, without using such an algorithm, a reduced model constructed from physiological analysis contains only a few variables but can still retain important characteristics of the complex model (Krishna et al., 2006). However, a systematic method will

always be helpful for model reduction not only in dealing with systems in a systematic way, but also in uncovering properties that may be ignored by experimental studies.

Cellular networks are well recognized to have inherent robustness. For the  $\kappa$ B-NF- $\kappa$ B signal pathway model, previous sensitivity analyses show that out of all the parameters (more than 60), only a small number of them (less than 10) have significant impacts on the oscillation behaviour of the NF- $\kappa$ B<sub>n</sub> concentration (Ihekweba et al., 2004; Yue et al., 2006, 2007), which suggests that this system is robust to the variation of a large number of parameters (Jia et al., 2007). The work in this paper further shows that the system is also very robust in model structure. The degree of robustness should be measured with respect to maintenance of steady state and this can be quantitatively investigated by system theory in the future work.

## REFERENCES

- A.R. Asthagiri, and D.A. Lauffenburger. Bioengineering models of cell signaling. *Annual Review of Biomedical Engineering*, 2:31–53, 2000.
- U.S. Bhalla. Biochemical signaling networks decode temporal patterns of synaptic input. *J. Computational Neuroscience*, 13:49–62, 2002.
- R. Cheong, A. Bergmann, S.L. Werner, J. Regal, A. Hoffmann, and A. Levchenko. Transient  $\kappa$ B kinase activity mediates temporal NF- $\kappa$ B dynamics in response to a wide range of tumor necrosis factor- $\alpha$  doses. *J. Biological Chemistry*, 281:2945–2950, 2006.
- H. Conzelmann, J. Saez-Rodriguez, T. Sauter, E. Bullinger, F. Allgower, and E.D. Gilles. Reduction of mathematical models of signal transduction networks: simulation-based approach applied to EGF receptor signalling. *IEE Proc. Systems Biology*, 1:159–169, 2004.
- H. Conzelmann, J. Saez-Rodriguez, T. Sauter, B.N. Kholodenko, and E.D. Gilles. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics*, 7:34, 2006.
- M.W. Covert, T.H. Leung, J.E. Gaston, and D. Baltimore. Achieving stability of lipopolysaccharide-induced NF- $\kappa$ B activation. *Science*, 309:1854–1857, 2005.
- S. Dano, M.F. Madsen, H. Schmidt, and G. Cedersund. Reduction of a biochemical model with preservation of its basic dynamic properties. *FEBS J.*, 273:4862–4877, 2006.
- D. Degenring, C. Froemel, G. Dikta, and R. Takors. Sensitivity analysis for the reduction of complex metabolism models. *J. Process Control*, 14:729–745, 2004.
- D.A. Fell. *Understanding the control of metabolism*. Portland Press, London, 1996.
- R. Heinrich, and S. Schuster. *The Regulation of Cellular Systems*. Springer, New York, 1996.
- A. Hoffmann, and D. Baltimore. Circuitry of nuclear factor  $\kappa$ B signaling. *Immunological Reviews*, 210:171–186, 2006.
- A. Hoffmann, A. Levchenko, M.L. Scott, and D. Baltimore. The  $\kappa$ B-NF- $\kappa$ B signaling module: temporal control and selective gene activation. *Science*, 298:1241–1245, 2002.
- A.E.C. Ihekweba, D.S. Broomhead, R.L. Grimley, N. Benson, M.R.H. White, and D.B. Kell. Sensitivity analysis of parameters controlling oscillatory signalling in the NF- $\kappa$ B pathway: the roles of IKK and  $\kappa$ B $\alpha$ . *IEE Proc. Systems Biology*, 1:93–103, 2004.
- B.P. Ingalls, and H. M. Sauro. Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories. *J. Theoretical Biology*, 222:23–36, 2003.
- J.F. Jia, T.Y. Liu, H. Yue and H. Wang. Robustness analysis of signaling transduction networks based on Monte-Carlo method. *Computer Simulation*, 24: 295–299, 2007.
- J.D. Kearns, S. Basak, S.L. Werner, C.S. Huang, and A. Hoffmann.  $\kappa$ B $\epsilon$  provides negative feedback to control NF- $\kappa$ B oscillations, signaling dynamics, and inflammatory gene expression. *J. Cell Biology*, 173:659–664, 2006.
- S. Krishna, M.H. Jensen, and K. Sneppen. Minimal model of spiky oscillations in NF- $\kappa$ B signaling. *PNAS*, 103: 10840–10845, 2006.
- R. Kruger, and R. Heinrich. Model reduction and analysis of robustness for the Wnt/ $\beta$ -catenin signal transduction pathways. *Genome Informatics*, 15:138–148, 2004.
- T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon, and M. Kimmel. Mathematical model of NF- $\kappa$ B regulatory module. *J. Theoretical Biology*, 228:195–215, 2004.
- G. Liu, M.T. Swihart, and S. Neelamegham. Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling. *Bioinformatics*, 21:1194–1202, 2005.
- M.R. Mauryaa, S. Katara, P.R. Patkar, A.E. Rundell, and V. Venkatasubramanian. A systematic framework for the design of reduced-order models for signal transduction pathways from a control theoretic perspective. *Computers & Chemical Engineering*, 30:437–452, 2006.
- D.E. Nelson, A.E.C. Ihekweba, M. Elliott, J.R. Johnson, C.A. Gibney, B.E. Foreman, G. Nelson, V. See, C.A. Horton, D.G. Spiller, S.W. Edwards, H.P. McDowell, J.F. Unitt, E. Sullivan, R.L. Grimley, N. Benson, D.S. Broomhead, D.B. Kell, and M.R.H. White. Oscillations in NF- $\kappa$ B signaling control the dynamics of gene expression. *Science*, 306:704–708, 2004.
- S.R. Neves, and R. Iyengar. Modeling of signaling networks. *Bioessays*, 24:1110–1117, 2002.
- E.L. O’Dea, D. Barken, R.Q. Peralta, K.T. Tran, S.L. Werner, J.D. Kearns, A. Levchenko, and A. Hoffmann. A homeostatic model of  $\kappa$ B metabolism to control constitutive NF- $\kappa$ B activity. *Molecular Systems Biology*, 3:111, 2007.
- S.L. Werner, D. Barken, and A. Hoffmann. Stimulus specificity of gene expression programs determined by temporal control of ikk activity. *Science*, 309:1857–1861, 2007.
- M.C. Wildermuth. Metabolic control analysis: biological applications and insights. *Genome Biology*, 1:1031.1–1031.6, 2000.
- H. Yue, M. Brown, J. Knowles, H. Wang, D.S. Broomhead and D.B. Kell. Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF- $\kappa$ B signaling pathway. *Molecular Biosystems*, 2:640–649, 2006.
- H. Yue, M. Brown, Y. Wang and D.B. Kell. Sensitivity analysis of an oscillatory signal transduction pathway. *Proc. 2nd Foundations of Systems Biology in Engineering*, Stuttgart, Germany: 99–104, 2007.