MODEL REDUCTION AND MODEL COMPARISON FOR NFκB SIGNALING

O. Radulescu [*] **A. Zinovyev** [**] **A. Lilienbaum** [***]

[*] *IRISA (INRIA, CNRS UMR 6074) and IRMAR (U.Rennes 1, CNRS UMR 6025), Campus de Beaulieu, 35042 Rennes, France*
[**] *Institut Curie, Bioinformatique, 26 rue d'Ulm, Paris F-75248, France, and Institute of Computational Modeling SB RAS, Krasnoyarsk, Akademgorodok, Russia, 660036*
[***] *Cytosquelette et Développement (CNRS URA 2115), Faculté de Médecine Pitié-Salpêtrière, 105, boulevard de l'Hôpital, 75634 Paris cedex 13, France*

Abstract: We introduce a methodology allowing to reduce and to compare systems biology models. This is based on several reduction tools. The first tool is a combination of Clarke's graphical technique and idempotent algebra. The second tool is the Karhunen-Loève expansion, providing a linear embedding for the invariant manifold. The nonlinear dimension of the invariant manifold is estimated by a third method. We also introduce a novel, more realistic model for NFκB signaling. This model is reduced and compared to existing models.

Keywords: Model reduction; Systems biology; NFκB signaling network

1. INTRODUCTION

Model reduction techniques are used to produce small, but still accurate models, from larger ones. In systems biology, model reduction methods were applied to signal transduction models (Conzelmann *et al.*, 2004) and to clocks (Indic *et al.*, 2006).

There is a need for general, eventually automatized, reduction methods. We may distinguish among three classes of model reduction techniques. *Trajectory based techniques* use the integration of the dynamical equations and look for a small number of reduced variables. The empirical orthogonal eigenfunctions (EOF), also called Proper Orthogonal Decomposition (POD), or Karhunen-Loève expansion (KL) method, consists in finding a low dimension linear space, containing (or sufficiently closed to) the trajectories. *Singular perturbations techniques* eliminate fast variables whose dynamics is slaved by the slower variables. This is equivalent to finding a lower dimensional, invariant manifold, containing the dynamics. Invariant manifolds methods have been applied to chemical kinetics (Gorban *et al.*, 2004). *Graph contraction* methods such as Clarke's (Clarke, 1992) replace the reactions mechanism by a simpler mechanism in which some intermediate species are absent.

There are many reasons for simplifying systems biology models. Models contain unnecessary complexity which conceals design principles, and renders analysis difficult. Sensitivity studies, critical parameters and regulation loops identification become easier for reduced models. To find common

[1] this work was supported by ACI IMPBio and by EC contract ESBIC-D (LSHG-CT-2005-518192)

patterns (model comparison) models should be simplified to a common level of complexity. In this paper we show how models of NF-κB signaling system can be simplified.

The transcription factor NF-κB has been discovered 20 years ago, and is still in the spotlights. NF-κB is involved in a wide diversity of domains such as the immune and inflammatory responses, cell survival and apoptosis, cellular stress and neurodegenerative diseases, cancer and development. NF-κB is sequestered in the cytoplasm by an inactivating protein, IκBα. When cells receive the appropriate signals, IκBα is phosphorylated by a kinase complex IKK, which triggers its ubiquitinylation, and which targets the molecule for degradation by the proteasomal complex. NF-κB that is bound to IκBα is released in the cytosol, then actively transported to the nucleus where it activates its target genes. Understanding such a complex biological system requires modeling. In the last years, several mathematical models of NF-κB have been published. The first model describes a single NF-κB molecule, which binds to IκBα, IκBβ and IκBϵ. This work lead to the demonstration of oscillation in NF-κB activity, confirmed by experimental data (Hoffmann and al., 2002), oscillation which is made possible by the feed-back loop of IκBα activated by NF-κB and inhibiting it. The model set by (Lipniacki *et al.*, 2004) takes the same approach, and adds to the complexity by modeling in addition the A20 molecule whose production is enhanced upon NF-κB stimulation, and which negatively regulates IKK activity. A third model analyzed the critical parameters necessary for maintaining oscillations, with given amplitude and frequency (Ihekwaba and al., 2004).

With the goal to better render the complexity of the NF-κB system, and to fit more to reality, we propose a fourth model with more complex description that takes into account transcription, translation and degradation of NF-κB. We also describe the production of the different NF-κB subunits, p50 which is cleaved from the p105 molecule, and the combinatorial possibilities between p50 and p65, which leads to the formation of p50.p65, p50.p50 and p65.p65. These dimers of NF-κB are characterized by different affinities for DNA sites, and associate differentially to IκBα, β and ϵ, generating thus 9 species with different proportion and characteristics upon signalling and degradation.

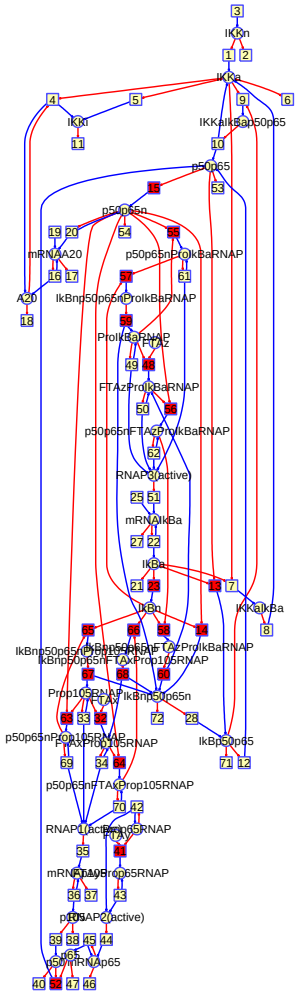

Fig. 1. Model $\mathcal{M}(39, 67, 88)$ in oriented bipartite graph representation: squares represent reactions, circles represent species.

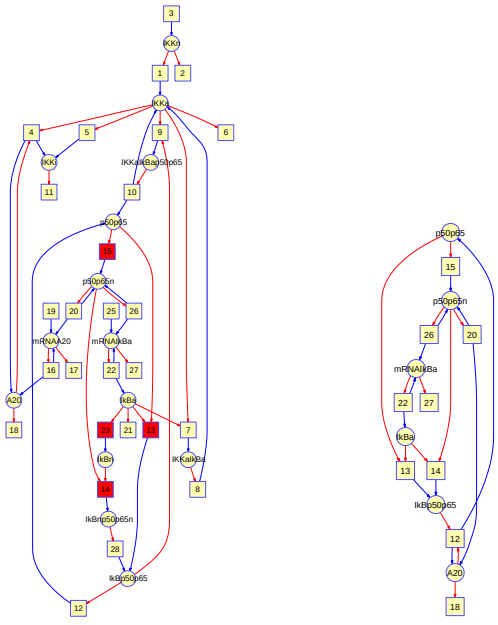

Fig. 2. Models $\mathcal{M}(14, 27, 36)$ and $\mathcal{M}(6, 9, 13)$

2. METHODS

2.1 Models

We consider chemical kinetics models of the form:

$$\frac{dX}{dt} = \mathbf{SR}(X) = \sum_{i=1}^{m} S^i R_i(X) \qquad (1)$$

where $X \in \mathbb{R}^n$ is the vector of concentrations of different species, $\mathbf{S}$ is the stoichiometric matrix. Each column of the stoichiometric matrix S^i corresponds to an elementary reaction.

Like in (Lipniacki *et al.*, 2004) we use a two-compartment approximation to cope with translocation between nucleus and cytoplasm. This amounts to considering reactions with non-integer stoichiometry: nuclear influx reactions have the stoichiometry $(\ldots, -1, \ldots, k_v, \ldots)$ where the real number k_v is the ratio between the volumes of the cytoplasm and the nucleus.

2.2 Clarke's reduction method

Let I be the set of intermediate species, that will be eliminated in the reduction procedure. $\mathcal{R}^{(I)}$ is the set of reactions that either produce or consume some species in I. $S^{(I)}$ is the restriction of the stoichiometry matrix to I. All other species, different from I and involved in reactions $\mathcal{R}^{(I)}$ are called terminal. A reaction route is any integer coefficient vector γ that satisfies:

$$S^{(I)}\gamma = 0,\ \gamma_i \geq 0,\ \text{if}\, R_i \in \mathcal{R}^{(I)} \text{is irreversible} \qquad (2)$$

A submechanism is made of all reactions in the reaction route (reactions of indices i such that $\gamma_i \neq 0$). A submechanism that does not contain a smaller submechanism is called simple. Simple mechanisms are analogue to elementary modes from flux balance analysis (Gagneur and Klamt, 2004).

The simplification method replaces $\mathcal{R}^{(I)}$ by its simple submechanisms. The rates of the submechanisms are obtained from the quasistationarity of the intermediates (at imposed terminal concentrations). The resulting rates are functions of the concentrations of the terminal species only. By this procedure the steady states of the simplified model can be perfectly imbedded in the steady states of the original model. Although it presents some dangers (it eliminates some time delays that can affect the stability of limit cycles), the method works well in networks with hierarchical time scales. In such networks a few variables dictate a simple dimensional dynamics. Other variables are either much more rapid (and are slaved by the simple dynamics) or are much slower (and modulate the simple dynamics).

2.3 Limit simplifications and idempotent algebras

Even supposing that elementary reactions follow mass action law kinetics, the reduced rates are complicated functions of the concentrations. More seriously, the decrease of the number of parameters is rather limited in Clarke's approach. A much more efficient reduction method is to use idempotent algebras in order to simplify rate functions. This consists in keeping only the dominant terms in the expressions of the rates. Polynomials become monomials or sums of few monomials. We call the models thus obtained limit simplifications.

A simple example of limit simplification is the min funnel. The min funnel combines two molecules to form a complex. The molecules are produced with rates R_1, R_2 and are degraded with constants k_{deg1},k_{deg2}. The complex formation is a reversible reaction with constants k_c, k'_c. Ignoring losses in the funnel $min(R_1, R_2) >> \frac{k_{deg1}k_{deg2}}{k_c}$ and considering that the complex C does no accumulate in huge amounts at the funnel exit (the order of $k'_c C$ is at most the order of R_1,R_2) we obtain the following limit simplification for the production rate of the complex: $R_C \approx min(R_1, R_2)$.

2.4 PCA (Karhunen-Loève expansion) for "flat" dimension

The dynamics of models that we present in this paper is close to a limit cycle, embedded in the multidimensional space of species concentrations. However, the support of the dynamics may be a low-dimensional linear manifold. One efficient method allowing to detect the low-dimensional linear manifold is the Principal Component Analysis (PCA) also known as Karhunen-Loève expansion.

PCA allows to calculate a new orthonormal basis in the concentration space such that the vectors of this new basis are ordered with respect to the species variation measured in the corresponding directions. New basis vectors are eigenvectors of the covariance matrix calculated for discretized system trajectory (represented in the phase space as a finite number of "snapshots"). The explained variation is quantified by the distribution of the corresponding eigen-values. The simplest method of giving "reduced" description consists in taking first k eigen-vectors such that the cumulative variation explained (relative sum of the first k eigen values) would exceed a certain threshold (for example, 95% of the total variance).

2.5 Estimating the "nonlinear" dimension

The approach that we use is based on a local version of PCA, with several ideas taken from

(Hundley and Kirby, 2003; Solis, 1999). It consists of two steps: sampling the invariant manifold and estimating the (local) dimension of the manifold.

We estimated the local dimension of a manifold M in point $c \in M$ where the manifold is sampled in a finite number m of vectors $X_i, i = 1..m$. Following (Hundley and Kirby, 2003), one can rely on the fact that in some vicinity of c the manifold M can be approximated by a linear expansion. The effective dimension of this expansion can be estimated using local variant of PCA. For an ϵ-ball around c, $B_\epsilon(c) = \{X_i : \|X_i - c\| < \epsilon\}$, we calculate the principal components for $X_i \in B_\epsilon$. The functions of the corresponding eigenvalues $\lambda_k(\epsilon)$ on the value of ϵ are called *singular value curves* (SVC). It can be shown that in the case of k-dimensional manifold with uniform sampling in the limit of many samples one should have exactly k SVCs of the form $\lambda(\epsilon) = \frac{1}{k+2}\epsilon^2$. All other $\lambda(\epsilon)$ functions should have different scalings (proportional to ϵ^4, ϵ^6, etc.) This result is difficult to use directly because in practice it is not possible to provide uniform and sufficiently dense sampling, and the sampling is often spoiled with noise. We found that a better choice is to estimate the local dimension of the invariant manifold as a number of linearly scaling at $\epsilon = 0$ functions $\sigma(\epsilon) = \sqrt{\lambda(\epsilon)}$, with the largest slopes separated from the others by a "spectral gap".

3. RESULTS

3.1 Hierarchical reduction of the model

As an illustration, we reduce a simpler version of our model (that employs only one member of the IκB family, namely the IκBα, which is the most important) towards the model proposed by Lipniacky (Lipniacki *et al.*, 2004). After that, we continue the reduction procedure towards even simpler models, thus obtaining an hierarchy of models. The complexity of a model is quantified by a triplet of positive integers (n, m, p) (numbers of species, reactions, and parameters). In the unreduced model, elementary reactions follow mass action law kinetics. Thus, each reversible reaction involves two kinetic parameters and each irreversible reaction only one. Another parameter is kv, the cytoplasm to nucleus volume ratio. Thus, the unreduced model is $\mathcal{M}(39, 67, 88)$ (there are 39 species, 67 reactions, among which 20 are reversible). Lipniacky's model is $\mathcal{M}(14, 27, 29)$ (all reactions are irreversible, but the conserved total NF-κB stands for a new parameter).

The reduction from $\mathcal{M}(39, 67, 88)$ to $\mathcal{M}(14, 27, 29)$ is performed in 5 steps. The first three steps correspond to limit simplifications of the mechanisms producing the proteins $p50$, $p65$, and the mRNA for $I\kappa B\alpha$. Thus, the reactions R_{41-46}, R_{32-39}, R_{63-70}, R_{48-51}, R_{55-62} are replaced by the simple submechanisms $\varnothing \to$ p65, $\varnothing \to$ p50, $\varnothing \to$ mRNAIkBa, with the rates:

$$R_{45} \approx \frac{k_{43}k_{45}}{k_{46}}[Prop50], \quad (3)$$

$$R_{39} \approx \frac{k_{36}k_{34}[Prop105]}{k_{37}(1 + \frac{k_{64}x_7}{k_{68}})}, \quad (4)$$

$$R_{26} \approx [ProIkB]\frac{k_{50} + \frac{k'_{58}x_7}{k'_{56}k'_{58} + k_{60}k_{58}x_{11}}}{1 + \frac{k_{56}}{k_{60}}x_7} \quad (5)$$

where $x_7 = [p50p65n]$, $x_{11} = [I\kappa Bn]$.

The fourth step is a min funnel simplification of the production of $p50p65$. R_{39}, R_{45} are replaced by $R_{52} \approx min(R_{39}, R_{45}) = R_{39}$.

In the fifth step, we use two time scale averaging to replace the non-conservative model by a conservative model. Without the reactions R_{52}, R_{53}, R_{54}, R_{71}, R_{72} that produce and consume $p50p65$, the total amount of $p50p65$ (free or complexed with other species) would be conserved. Considering that the reactions R_{53}, R_{54}, R_{71}, R_{72} have the same constant k, the total amount X of $p50p65$ satisfies the equation :

$$\frac{dX}{dt} = -kX + R_{52} \quad (6)$$

R_{52} depends on $[p50p65n]$ and weakly on $[IkBn]$. During the accumulation of p50p65, the dynamics has two time scales, with the slow timescale (about 10 h) controlled by Eq.6 and the rapid timescale controlled by the oscillations of the model (period 1-2 h). An averaging argument allows us to estimate the asymptotic amount of $p50p65$: $X_\infty =< R_{52} > /k$, where the average is over a period of the oscillations for sustained oscillations or the value at steady state for damped oscillations.

Thus, the fifth reduction step consists in eliminating the reactions R_{52}, R_{53}, R_{54}, R_{71}, R_{72} and using as initial condition for the total amount of $p50p60$ the asymptotic value X_∞ that becomes a parameter of the reduced model. We obtain the model $\mathcal{M}(14, 27, 36)$ that has the same species and reactions as Lipniacky's model $\mathcal{M}(14, 27, 29)$, but slightly more parameters. The difference in the number of parameters comes from the more complex expressions of the mRNA IκB transcription rate R_{26} (this rate is a function of x_7, x_{11} and four parameters in $\mathcal{M}(14, 27, 36)$ as in Eq.5 and is simply $R_{26} = k_{26}x_7$ in $\mathcal{M}(14, 27, 29)$) and from the fact that in $\mathcal{M}(14, 27, 36)$ four reactions are reversible.

Although the reversibility of some reactions is not really an important difference, the two ex-

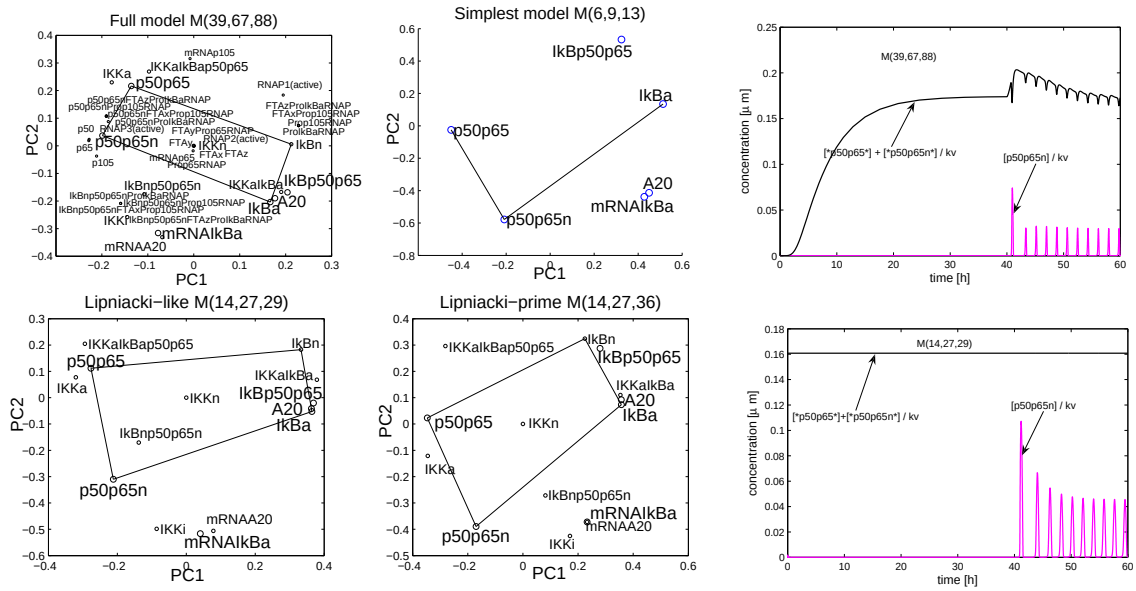

Fig. 3. Visualization of components of first two principal vectors calculated in concentration space. Grouped species oscillate "in-phase" on the limiting trajectory. Signal was applied at t=40h.

pressions for R_{26} are sufficiently different to distinguish the two models. To compare the dynamics we have chosen a parameter $k_{26} = \frac{k'_{58}P_3}{(k'_{56}k'_{58}+k_{60}k_{58}<x_{11}>)(1+\frac{k_{56}}{k_{60}}<x_7>)}$ for which we expect comparable rates R_{26} for average values of x_7,x_{11}, in the two models.

Three more reduction steps transform $\mathcal{M}(14,27,29)$ into $\mathcal{M}(6,9,13)$. The kinase transformations reactions R_{1-11} are replaced by two submechanisms IkBa $\rightarrow \varnothing$, IkBp50p65 $\rightarrow$ p50p65, of rates:

$$R_{21} \approx [k_{21} + \frac{k_3 k_7}{k_5 + k_4 x_8}]x_{20} \quad (7)$$

$$R_{12} \approx \frac{k_3 k_9}{k_5 + k_4 x_8} x_{13} \quad (8)$$

where $x_{10} = [I\kappa B]$, $x_8 = [A20]$, $x_{13} = [I\kappa Bp50p65]$.

A20 production can be simplified to $\varnothing \rightarrow$ A20, with the rate:

$$R_{20} = \frac{k_{16}k_{20}x_7}{k_{17}} \quad (9)$$

The nuclear complex formation reactions R_{14}, R_{23}, R_{28} are replaced by $I\kappa B\alpha + k_v p50p65n \rightarrow I\kappa Bp50p65$, with the rate:

$$R_{14} = \frac{k_{23}k_{14}}{k_v k_{24} + k_{14}x_7} x_{10}x_7 \quad (10)$$

Sensitivity analysis shows that reactions R_{21}, R_{25} have little influence on the oscillating behavior of the model. The same is true about the reversibility of R_{13-14}, R_{15}. This leads us to the simplest model $\mathcal{M}(6,9,13)$.

3.2 Model comparison and dynamical dimension

Model	80%	95%	99%	99.9%	100%
$\mathcal{M}(39,67,88)$	2	4	6	8	15
$\mathcal{M}(14,27,36)$	3	4	6	7	11
$\mathcal{M}(14,27,29)$	2	3	4	5	9
$\mathcal{M}(6,9,13)$	2	3	4	4	5

Table 1. Dimensions of linear manifolds embedding the limiting trajectories for various "explained" variance threshold.

All reduced models have oscillating behavior. Although sustained oscillations were not observed in biological experiments, the NF-κB system undeniably functions close to a Hopf bifurcation. One can pass from damped to sustained oscillations by changing X_∞ (lowering it produces damping). We used the limit cycle ideal behavior to compare models. We found little qualitative differences among reduced models (Fig.3). Using the PCA analysis we conclude that the effective dimension of linear manifolds containing the trajectories is low and does not change much by model reduction (Table1). The dynamics of the most complex model is not more dimensional than the reduced model $\mathcal{M}(14,27,36)$ at 99%. The same is true if we compare $\mathcal{M}(14,27,29)$ and $\mathcal{M}(6,9,13)$.

However, one simplification of the system (linearization of the transcription rate R_{26} of IκB mRNA as function of nuclear p50p65, i.e. transition from $\mathcal{M}(14,27,36)$ to $\mathcal{M}(14,27,29)$) gives drastic decrease of linear dimension. The reason for this should be looked for in the distribution of timescales for the two models. Although the period is similar, the duration of the peaks is quite different. In our model the inhibitor x_{11} relatively quickly turns down the transcription rate (see Eq.5) and reduces the duration of the

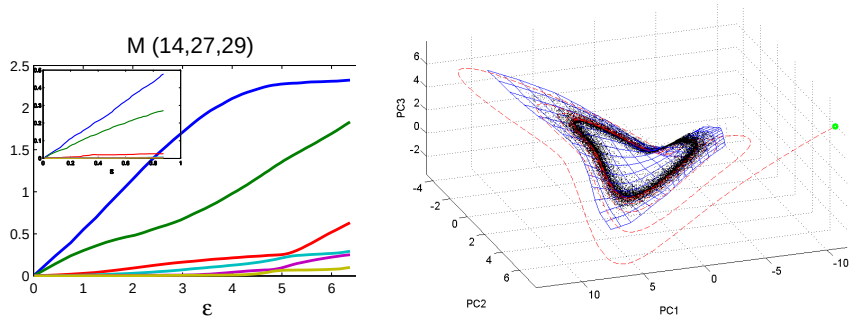

Fig. 4. Singular value curves (SVC) $\sigma_i(\epsilon)$ and invariant manifold for $\mathcal{M}(14, 27, 29)$; manifold sampling (dots) and one particular system trajectory (dashed line) are shown. Notice two dominating SVCs, separated by a spectral gap from the others: an indication of local existence of 2D invariant manifold.

peaks. In Lipniacky's model this effect is not taken into account and the duration of the peaks is longer (Fig.3). Hence, several rapid modes that can be reduced without loss for this model (kinase complexification and nuclear complex formation), can not be reduced for our model.

The nonlinear analysis confirms these findings. From Fig. 4 it follows that for $\mathcal{M}(14, 27, 29)$ there exists a well defined nonlinear two-dimensional invariant manifold. One can estimate the local dynamics dimension to be 3 for $\mathcal{M}(14, 27, 36)$ and about 5 for $\mathcal{M}(39, 67, 88)$. To illustrate the existence of a two-dimensional invariant manifold, we present a sampling produced for the model $\mathcal{M}(14, 27, 29)$ on Fig. 4, where the distribution of dynamical "snapshots" is approximated by the *elastic principal manifold* (Gorban and Zinovyev, 2005).

4. CONCLUSIONS

Our approach generated an hierarchy of models of various structural complexity. The model $\mathcal{M}(14, 27, 36)$ obtained by reduction from our most complex model $\mathcal{M}(39, 67, 88)$ has precisely the same set of molecular species and biochemical reactions as the existing $\mathcal{M}(14, 27, 29)$ (Lipniacki *et al.*, 2004). We spotted the differences between these models and identified their origin. Using the dimension of the invariant manifold as a criterion, we showed that our model has higher dynamical complexity. We also showed that $\mathcal{M}(14, 27, 29)$ permits a series of simplifications. Although these do not lead to the simplest dynamical dimension (which is 2) they allow to sort biochemical regulations according to their effects on dynamics. For our model we can mention the importance of IκB-induced detachment of NF-κB from DNA (reaction R_{58}, R_{60}). As seen in Eq.5, this effect is the main modulator of mRNA IκB transcription rate. By eliminating R_{58}, the limit simplification changes and the oscillation period doubles. These reasonings are crucial for designing experiments to test the validity of different models and the relative roles of various regulations.

REFERENCES

Clarke, B.L. (1992). General method for simplifying chemical networks while preserving overall stoichiometry in reduced mechanisms. *J.Phys.Chem.* **97**, 4066–4071.

Conzelmann, H., J. Saez-Rodriguez, T. Sauter, E. Bullinger, F. Allgwer and E.D. Gilles (2004). Reduction of mathematical models of signal transduction networks: simulation-based approach applied to egf receptor signalling. *Syst.Biol.* **1**, 159–169.

Gagneur, J. and S. Klamt (2004). Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* **5**, 175.

Gorban, A.N. and A. Zinovyev (2005). Elastic principal graphs and manifolds and their practical applications. *Computing* **75**, 359–379.

Gorban, A.N., I.V. Karlin and A.Yu. Zinovyev (2004). Invariant grids for reaction kinetics. *Physica A* **333**, 106–154.

Hoffmann, A. and al. (2002). The iκb-nf-κb signaling module: temporal control and selective gene activation. *Science* **298**, 1241–1245.

Hundley, D. and M. Kirby (2003). Estimation of topological dimension. In: *Proceedings of the Third SIAM International Conference on Data Mining* (D. Barbara and C. Kamath, Eds.). pp. 194–202.

Ihekwaba, A.E.C. and al. (2004). Sensitivity analysis of parameters controlling oscillatory signalling in the nf-κb pathway: the roles of ikk and iκbα. *Syst. Biol.* **1**, 93–102.

Indic, P., K. Gurdziel, R.E. Kronauer and E.B. Klerman (2006). Development of a two-dimension manifold to represent high dimension mathematical models of the intracellular mammalian clock. *J.Biol.Rhythms* **21**, 222–232.

Lipniacki, T., P. Paszek, A.R. Brasier, B. Luxon and M. Kimmel (2004). Mathematical model of nf-κb regulatory module. *J.Theor.Biol.* **228**, 195–215.

Solis, F.J. (1999). Dimension and local structures of attracting manifolds of smooth dynamical systems. *Appl. Math. Comp.* **100**, 169–175.