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
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 neuro-degenerative 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, 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.

2. METHODS

2.1 Models

We consider chemical kinetics models of the form:
\[
\frac{dX}{dt} = \mathbf{SR}(X) = \sum_{i=1}^{m} S^i R_i(X)
\]

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

\[
S^{(\mathcal{I})}\gamma = 0, \quad \gamma_i \geq 0, \text{ if } R_i \in \mathcal{R}^{(\mathcal{I})} \text{ is irreversible}
\]  

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 \( \mathcal{R}^{(\mathcal{I})} \) 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 this 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 polynomial 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 \( k_{\text{deg}1}, k_{\text{deg}2} \). The complex formation is a reversible reaction with constants \( k_v, k_v' \). Ignoring losses in the funnel \( \min(R_1, R_2) \gg \frac{k_{\text{deg}1}k_{\text{deg}2}}{k_v} \) and considering that the complex \( C \) does not accumulate in huge amounts at the funnel exit (the order of \( k_v'C \) is at most the order of \( R_1, R_2) \) we obtain the following limit simplification for the production rate of the complex: \( RC \approx \min(R_1, R_2) \).

### 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, effectively it may be located on 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
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 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 \(k_{v}\), the cytoplasm to nucleus volume ratio. Thus, the unreduced model is \(M(39, 67, 88)\) (there are 39 species, 67 reactions, among which 20 are reversible). Lipniacki’s model is \(M(14, 27, 29)\) (all reactions are irreversible, but the conserved total NF-κB stands for a new parameter). The reduction from \(M(39, 67, 88)\) to \(M(14, 27, 29)\) 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 IκBα. Thus, the reactions \(R_{41–46}, R_{52–39}, R_{60–70}, R_{48–51}, R_{55–62}\) are replaced by the simple submechanisms \(\varepsilon \rightarrow p50, \varepsilon \rightarrow p50, \varepsilon \rightarrow mRNA\), with the rates:

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

\[
R_{59} \approx \frac{k_{38}k_{34}}{k_{37}(1 + \frac{k_{44}x_7}{k_{45}})}, \quad (4)
\]

\[
R_{26} \approx \frac{k_{50} + \frac{k_{56}x_7}{k_{55}x_7 + k_{56}x_7}}{1 + \frac{k_{50}}{k_{57}x_7}} \quad (5)
\]

where \(x_7 = [p50p65n], x_{11} = [IκBn]\).

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

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 \([IκBn]\). 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 \(X_\infty\) that becomes a parameter of the reduced model. We obtain the model \(M(14, 27, 36)\) that has the same species and reactions as Lipniacki’s model \(M(14, 27, 29)\), 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 \(M(14, 27, 36)\) as in Eq.5 and is simply \(R_{26} = k_{26}x_7\) in \(M(14, 27, 29)\)) and from the fact that in \(M(14, 27, 36)\) four reactions are reversible.

Although the reversibility of some reactions is not really an important difference, the two ex-
pressions for \( R_{20} \) are sufficiently different to distinguish the two models. To compare the dynamics we have chosen a parameter \( k_{26} = \frac{k_{14} P_1}{(k_{36} k_{58} + k_{60} k_{58}) x_{11}} \) 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 \( R_{11-11} \) are replaced by two submechanisms \( \text{IkBa} \rightarrow \varnothing, \text{IkBp50p65} \rightarrow p50p65 \), 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 \( x_{10} = [\text{IkB}] \), \( x_8 = [A20] \), \( x_{13} = [\text{IkBp50p65}] \).

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 \( \text{IkBp} + k_6 p50p65n \rightarrow \text{IkBp50p65} \), with the rate:

\[
R_{14} = \frac{k_{23} k_{14}}{k_6 k_{24} + k_{14} x_7} x_{10} x_7 \quad (10)
\]

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

<table>
<thead>
<tr>
<th>Model</th>
<th>80%</th>
<th>95%</th>
<th>99%</th>
<th>99.9%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{M}(39, 67, 88) )</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>( \mathcal{M}(14, 27, 36) )</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>( \mathcal{M}(14, 27, 29) )</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>( \mathcal{M}(6, 9, 13) )</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

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 \( X_{\infty} \) (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 IkB 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
peaks. In Lipniacki’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 in models \( M(6,9,13) \) and \( M(14,27,29) \) there exists a well defined nonlinear two-dimensional invariant manifold, but not in the other models. One can estimate the local dynamics dimension to be 3 for \( M(14,27,36) \) and about 5 for \( M(39,67,88) \). To illustrate the existence of a two-dimensional invariant manifold, we present a sampling produced for the model \( M(14,27,29) \) 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 \( M(14,27,36) \) obtained by reduction from our most complex model \( M(39,67,88) \) has precisely the same set of molecular species and biochemical reactions as the existing \( M(14,27,29) \) (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 \( M(14,27,29) \) 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 IxB-induced detachment of NF-κB from DNA (reaction \( R_{58}, R_{68} \)). As seen in Eq.5, this effect is the main modulator of mRNA IxB 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


