Note: This is a preprint of a paper being submitted for publication.
Final publication is available from Mary Ann Liebert, Inc., publishers http://dx.doi.org/10.1089/cmb.2016.0065

Computational framework for modelling multiple non-cooperative transcription factor binding and its application to the analysis of NF-κB oscillatory response

Manca Bizjak[†], Nikolaj Zimic[†], Miha Mraz[†]

and Miha Moškon[*†]

[*]To whom correspondence should be addressed.
[†]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

Please cite this paper as:
Bizjak Manca, Zimic Nikolaj, Mraz Miha, and Moškon Miha. Journal of Computational Biology. December 2016, 23(12): 923-933. doi:10.1089/cmb.2016.0065.

Abstract

Recent studies have shown that regulation of many important genes is achieved with multiple transcription factor binding sites with low or no cooperativity. Additionally, non-cooperative binding sites are gaining more and more importance in the field of synthetic biology. Herein we introduce a computational framework that can be applied to dynamical modelling and analysis of gene regulatory networks with multiple non-cooperative transcription factor binding sites. We propose two computational methods to be used within the framework, i.e. average promoter state approximation and expression profiles based modelling. We demonstrate the application of the proposed framework on the analysis of NF-κB oscillatory response analysis. We show that different promoter expression hypotheses in a combination with the number of transcription factor binding sites drastically affect the dynamics of the observed system and should not be ignored in the process of quantitative dynamical modelling, as is usually the case in existent state-of-the-art computational analyses.
Key words: gene regulatory networks, non-cooperative transcription factor binding, quantitative modelling, computational analysis, transcription factor NF-κB.

1. Introduction

Interactions between transcription factors (TFs) at the level of promoter binding sites divide regulatory circuits into two groups with respect to their transcriptional responses. The first group converts concentrations of regulating TFs merely to on/off promoter activity. This behaviour is usually achieved with positive cooperative TF binding, which results in a highly non-linear response.

On the contrary, the second group of regulatory circuits produces more analogue transcriptional response. In this case, promoters without or with very low cooperative regulation between the TF binding sites are involved (Giorgetti *et al.*, 2010). Low cooperativity causes TFs to bind to the promoters gradually, i.e. linearly proportional to their concentrations. Furthermore, promoter activity in this case depends on the exact state of its binding sites (Spitz & Furlong, 2012). Therefore, TFs do not only define the activity of promoters, but may also define the rate of their expression in different manners (Giorgetti *et al.*, 2010; Spitz & Furlong, 2012).

Gradual promoter response on the first hand reflects high regulatory capacity compared to promoters with simple on/off dynamics. Larger clusters of TF binding sites can on the other hand be used to decrease the effects of intrinsic noise (Giorgetti *et al.*, 2010), reduce the crosstalk among different operators and enhance the efficiency of repression (Lebar *et al.*, 2014). Large clusters of TF binding sites can be used when graded, i.e. analogue, transcriptional response is desired (Lorberbaum & Barolo, 2013). However, non-cooperative TF binding sites are not limited to linear behaviour only. Non-linear response can be achieved with the use of different enhancer activity models (Spitz & Furlong, 2012), positive feedback loops and competition for the same DNA operator binding sites between activators and repressors (Lebar *et al.*, 2014). Large numbers of non-cooperarive monomeric TF binding sites thus allow us to design tunable biological systems, which reflect robust behaviour, and can provide higher regulatory capacity with respect to cooperative binding, which typically results in a digital (non-linear) on/off response.

The development of different artificially engineered monomeric TFs such as Zinc fingers (Gommans *et al.*, 2005), TAL effectors (Garg *et al.*, 2012) or CRIPSR/Cas-based repressors (Cong *et al.*, 2013; Qi *et al.*, 2013) has recently enabled us to design biological networks with scalable number of non-cooperative TF binding sites. These platforms have already been applied to the design of boolean logic gates (Gaber *et al.*, 2014) and a toggle switch (Lebar *et al.*, 2014). In these examples multiple non-cooperative TF binding sites are used in order to decrease the intrinsic noise and thus increase the robustness of designed circuits and achieve the desired response with the use of monomeric TFs only. While the analysis of promoters with multiple non-cooperative TF binding sites is important from the perspective of synthetic biology, it may also shed light on the analysis of similar regulatory mechanisms observed in natural systems. For example,

the Epstein-Barr Virus latency switch is regulated by a promoter with 20 non-cooperative TF binding sites (Werner *et al.*, 2007). Moreover, TF NF-κB, controlling inflammation and immunity processes in different eukaryotes (Giorgetti *et al.*, 2010), and TF Msn2, regulating general stress response in yeast (Stewart-Ornstein *et al.*, 2013) also follow similar behaviour.

In this text we present a computational framework that can be used to efficiently model the dynamics of gene regulatory networks (GRNs) with multiple non-cooperative TF binding sites. The framework consists of three levels, i.e. (1) promoter state evaluation, (2) evaluation of transcriptional activity and (3) updating the concentrations. We introduce a methodology that can be used on the first level, namely *average promoter state approximation* (APSA). APSA reduces computational complexity of existing approaches that are commonly used to evaluate the promoter state. We additionally propose a compact way of presenting different promoter states' transcriptional activity, namely *expression profiles based modelling* (EPBM), which can be used on the second level of the modelling framework. We then demonstrate the application of the framework on a case study of NF-κB oscillatory response analysis, where we observe how different numbers of observed binding sites affect the oscillatory dynamics of the system together with different promoter expression hypotheses and other parameters describing the system under study.

The remainder of this text is organized as follows: Section 2 introduces the modelling framework. Its application on the analysis of the NF-κB oscillatory response is described in Section 3. Section 4 discusses the generality of the described approaches, their potentials in practical applications, and provides some guidelines for future work.

2. Computational modelling framework

Quantitative models describing the dynamics of GRNs are usually composed of a set of ordinary differential equations (ODEs) or of a stoichiometric matrix presentation (Le Novere, 2015). Both presentations are derived from a set of chemical reactions describing the changes in the observed system. The size of this set is proportional to the number of observed chemical species, which presents a major obstacle when dealing with more complex systems. For example, let's assume a GRN where promoters have n non-cooperative TF binding sites with m different types of competing TFs. In this case, the number of species that derive from only one promoter equals $(m+1)^n$. Since each promoter state is represented as a separate chemical species, the number of differential equations in a conventional ODE model coincides with the number of different promoter states.

When we are dealing with cooperative binding, these numbers can be drastically reduced with Hill equations (Alon, 2007). These presume that the promoter state is either unbound or fully occupied. While restriction to only two possible promoter states seems reasonable and also significantly lowers the computational load, there are several drawbacks of these models (Weiss, 1997). Moreover, they cannot be applied to non-cooperative binding.

Consequently, accurate quantitative modelling of multiple non-cooperative binding sites is without further presumptions only possible for very small systems, e.g., for a single promoter with only a few binding sites (Murphy *et al.*, 2007). We can usually simplify the model by ignoring the arrangements of bound TFs, which means we only consider their quantities (Bintu *et al.*, 2005a,b; Giorgetti *et al.*, 2010; Sauro, 2012; Werner *et al.*, 2007). This results in $\binom{n+m}{m}$ different promoter states, which still becomes infeasible for numerical simulations of non-trivial scenarios. Further simplifications can be made with the separation of observed reactions into two time scales according to their kinetic rates, i.e. *slow* reactions, that describe the processes of degradation, transcription and translation, and *fast* reactions that describe the binding processes (Hasty *et al.*, 2001). Since fast reactions occur several magnitudes of order faster than slow reactions, we can presume that promoters always reach a quasi-equilibrium state before the changes due to the slow reactions occur (Rao & Arkin, 2003). We can therefore simulate the behaviour of the observed system with the iteration of the following three steps, which are represented as three seperate levels of our computational framework:

1. *Promoter state evaluation*: evaluate current promoter states or their probabilities on the basis of TF concentrations and binding affinities.

2. *Evaluation of transcriptional activity*: evaluate the transcriptional activity of promoters based on their states and their expression profiles.

3. *Updating concentrations*: update concentrations of observed chemical species according to slow reactions, i.e. degradation, transcription and translation.

2.1. Promoter state evaluation

Promoter states are evaluated in dependence on corresponding TF concentrations and their binding/dissociation constants. Two approaches that have already been applied to the modelling of GRNs with multiple TF binding sites are *fractional occupancy* (FO) (Sauro, 2012) and *thermodynamic modelling* (TDM) (Bintu *et al.*, 2005a,b; Giorgetti *et al.*, 2010; Werner *et al.*, 2007).

2.1.1 Fractional occupancy and thermodynamic modelling

Both FO and TDM are based on the estimation of probabilities of different promoter states and can be integrated into proposed computational framework in a straightforward manner. Probabilities of promoter state s_x can be calculated as

$$P(s_x) = \frac{W(s_x)}{\sum_i W(s_i)}, \tag{1}$$

where $W(s_j)$ describes the weight of promoter state j. FO defines the weights as

$$W(s_x) = N_x \cdot K_{A_1,n_1} \cdot [A_1]^{n_1} \cdot K_{A_2,n_2} \cdot [A_2]^{n_2} \cdots K_{A_m,n_m} \cdot [A_m]^{n_m}, \tag{2}$$

where $[A_i]$ is concentration of TF i, n_i number of binding sites occupied by TF i in state s_x and K_{A_i,n_i} the binding constant of n_i TFs A_i to the promoter region. N_x is the number of arrangements that define the observed state and can be expressed as

$$N_x = \frac{n!}{n_1!n_2!\cdots n_m!(n - n_1 - n_2 - \cdots - n_m)!}. \tag{3}$$

TDM is essentially the same method as FO, but uses different data, i.e. binding free energies of each promoter state (Bintu *et al.*, 2005a,b), to evaluate promoter state weights. These are evaluated as

$$W(s_x) = N_x \cdot \mathrm{e}^{-\Delta G_{n_1}/RT} \cdot [A_1]^{n_1} \cdot \mathrm{e}^{-\Delta G_{n_2}/RT} \cdot [A_2]^{n_2} \cdots \mathrm{e}^{-\Delta G_{n_m}/RT} \cdot [A_m]^{n_m}, \tag{4}$$

where ΔG_{n_i} is the free energy change between the n_i bound TFs of type i and the reference promoter state (i.e. fully unbound promoter), T is the absolute system temperature, and R is gas constant. Free energy change of the reference state equals 0 ($\Delta G_0 = 0$). Note that the weight of unbound promoter s_0 equals 1 in both FO and TDM approach.

Described equations presume that no cooperativity among different TFs exists. If there is also no cooperativity among TFs of the same type, it follows that $K_{A_i,n_i} = K_{A_i}^{n_i}$ in FO and $\Delta G_{n_i} = n_i \cdot \Delta G_i$ in TDM, where K_{A_i} is binding constant of a single TF A_i to the promoter, and ΔG_i presents the free energy change between the reference state and the state in which one binding site is occupied by TF A_i. Both approaches presume monomeric TF binding, but can be adjusted to fit

multimeric binding by raising Hill coefficients. Another presumption they make is that all TF binding sites have the same binding affinity for the same TF type on the same promoter. If the latter does not hold, one needs to regard each binding site independently, which again leads to exponential growth of the number of observed chemical species. However, biological plausibility of this presumption has been justified thoroughly in the literature (e.g. see (Bintu *et al.*, 2005a; Giorgetti *et al.*, 2010; Lebar *et al.*, 2014; Werner *et al.*, 2007; Zeiser *et al.*, 2007)).

2.1.2 Average promoter state approximation

Although FO and TDM significantly decrease the computational effort, both still perform a significant number of calculations that are necessary to determine the weights of promoter states in each time step of the simulation. We propose a simplified approach, i.e. *average promoter state approximation* (APSA), which is able to accurately simulate the dynamics of more complex GRNs. APSA reduces the time complexity of promoter state approximation from $O\left(N \cdot \binom{n+m}{m}\right)$ of FO and TDM to $O\left(N \cdot m\right)$ per promoter type, where N is the number of simulation iterations, n the number of binding sites, and m the number of different TF types that bind to these sites (see Supplementary text for time complexity evaluation). Proposed approach evaluates an average promoter state instead of probabilities of each possible state, and can be derived from the FO (see Supplementary text for derivation). When m TF types competitively bind to n binding sites, average number of bound TF A_i can be expressed as

$$\overline{n}_{A_i} = n \frac{K_{A_i}[A_i]}{1 + \sum_{j=1}^{m} K_{A_j}[A_j]}. \tag{5}$$

Note that TF binding sites can be regarded independently when there is no competition between the TFs. Finally, we can describe the average promoter state as a set of average numbers of bound TFs of each type, i.e. $\overline{s} = \left\{\overline{n}_{A_1}, \overline{n}_{A_2}, \ldots, \overline{n}_{A_m}\right\}$.

2.2. Evaluation of transcriptional activity

Once the promoter states or their probabilities are evaluated, we can use them to determine the transcriptional activity of the observed promoter. We propose a compact way of presenting different promoter states' transcriptional activity, namely *expression profiles based modelling* (EPBM), which derives from the matrix presentation. The dimensionality of the EPBM matrix is defined by the number of different TF types that bind to their corresponding binding sites. If m is the number of different TFs, the promoter with n binding sites would require a m-dimensional matrix with $n+1$ being the size of each dimension. For example, a promoter with n binding sites regulated by both activator and repressor yields an expression profiles matrix $M \in \mathbb{R}^{(n+1)\times(n+1)}$, in which each element

corresponds to expression profile of a certain promoter state. Continuing the previous example, M_{ij} represents the expression profile of the promoter state s_{ij} with i bound activators and j bound repressors, which we denote as $\rho(s_{i,j})$. Note that elements that lay bellow the matrix anti-diagonal are not valid in the case of competitive binding, since the sum of bound TFs $i + j$ produces a number larger than the number of available TF binding sites n. When using FO or TDM we can evaluate the transcriptional activity of observed promoter as $\rho = \sum_x \rho(s_x) \cdot P(s_x)$, where $\rho(s_x)$ presents the transcriptional activity of promoter state s_x.

Alternatively, we can describe expression profiles with the m-dimensional interpolation function corresponding to EPBM matrix presentation. It is necessary to have such presentation when using APSA, since it generally yields non-integer numbers corresponding to the average numbers of bound TFs. To determine the promoter activity in this case we have to interpolate values between different promoter states' expression profiles. This can be achieved, e.g., with the weighted sum of all valid nearest promoter states. Another option is to define an m-dimensional interpolation function, i.e. $\rho(\cdot)$, that presents the promoter expression profiles and substitutes the matrix presentation. In both cases we evaluate promoter activity as $\rho = \rho(\bar{s})$.

2.3. Updating concentrations

After promoter activity has been evaluated it can be used to update concentrations of the observed species with respect to slow reactions, i.e. transcription, translation and degradation. This can be achieved with the application of ODE solvers or variations of stochastic simulation algorithm (SSA) (Gillespie, 1977). E.g., when using ODE solvers, the selected mRNA dynamics can be described as

$$\frac{dm}{dt} = \sum_{i=1}^{k} P_i \rho_i - \gamma m, \tag{6}$$

where $P_1, ..., P_k$ and $\rho_1, ..., \rho_k$ denote concentrations and evaluated transcriptional activity of promoters expressing mRNA species m, and γ denotes degradation rate of m. Equations governing the dynamics of mRNA species can be integrated into the system of ODEs together with the description of other slow reactions to obtain a functional model of the observed GRN.

3. Case study: analysis of NF-κB oscillatory response

We demonstrate the application of the proposed computational framework on the analysis of NF-κB oscillatory response. In the light of recent progress, indicating the non-cooperative nature of the NFKBIA gene TF binding sites (Giorgetti *et al.*, 2010), we were interested in how different number of the observed TF binding sites in a combination with different expression hypotheses affect the dynamics of the analysed system.

3.1. Model description

NF-κB is a TF crucial for induction of genes in response to inflammatory stimuli thus regulating inflammation and immune response in mammalian cells (Cheong *et al.*, 2008). NF-κB may in certain conditions exhibit oscillatory behaviour in response to sufficient concentrations of a stimulant TNF-α, tumour necrosis factor alpha. However, it has only recently been shown that many of NF-κB's target genes contain clustered DNA binding sites with negligible cooperativity, resulting in analogue inflammatory response, where graded increase in TNF-α concentration results in a graded increase of nuclear NF-κB concentration (Giorgetti *et al.*, 2010). The inhibitor of NF-κB activity, IκBα, is encoded by NFKBIA gene, one of the most common NF-κB's target genes. NF-κB and IκBα are thus connected via negative feedback loop, a necessary prerequisite for oscillatory response. Production of IκBα is triggered by the presence of IκB kinase (IKK) induced by TNF-α, which phosphorylates IκBα, resulting in its saturated degradation (Basak *et al.*, 2012). Saturated degradation of IκBα induced at a constant level of IKK concentration introduces a time delay, which is a necessary requirement that can lead to sustained oscillations in combination with the core negative feedback loop (Mengel *et al.*, 2010). Note, however, that our model ignores the presence of two other IκB isoforms, namely IκBβ and IκBϵ, otherwise responsible for dampening the oscillations on larger time-scales (Cheong *et al.*, 2008).

We adopted a three-dimensional model presented in Krishna et al. (Krishna *et al.*, 2006) to account for different numbers of TF biding sites. We observed the time course of nuclear NF-κB concentrations, cytoplasmic concentrations of IκBα and its transcript using the following model:

$$\begin{aligned}
\frac{dx}{d\tau} &= A(1-x)\frac{\epsilon}{\epsilon+z} - Bz\frac{x}{\delta+x}, \\
\frac{dy}{d\tau} &= \rho - y, \\
\frac{dz}{d\tau} &= y - C(1-x)\frac{z\epsilon}{\epsilon+z},
\end{aligned} \tag{7}$$

where x, y and z are proportional to NF-κB, IκBα transcript and IκBα concentrations, respectively. Parameters A, B, C, δ and ϵ are derived from basic kinetic rates

(for exact model derivation and parameter values see Supplementary text), and ρ defines promoter activity in dependence on the promoter state using three equally plausible hypotheses, i.e. additive, all-or-none and singular (Giorgetti *et al.*, 2010).

With the additive hypothesis the transcriptional response is proportional to the number of occupied binding sites. With the all-or-none hypothesis promoter is active if all of the n binding sites are occupied. Finally, with the singular hypothesis transcriptional response is triggered by the binding of an arbitrary (>0) number of TFs. Because we are dealing with a single TF, the dimensionality of the EPBM matrix is $(n+1) \times 1$ and has the following forms

$$M = \begin{cases} \left[0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right]; & \textit{if} \text{ additive,} \\ [0, 0, 0, \ldots, 0, 1]; & \textit{if} \text{ all-or-none,} \\ [0, 1, 1, \ldots, 1, 1]; & \textit{if} \text{ singular.} \end{cases} \tag{8}$$

As described in Section 2.2, it is necessary to interpolate these matrices in order to use them when modelling with the APSA method.

3.2. Analysis

We analysed the features of oscillations in terms of their periods, amplitudes and spikiness, a common phenomenon caused by the saturated degradation of a regulator (Mengel *et al.*, 2010). Spikiness was evaluated according to the measure $(\max(x) - \min(x)) / \text{mean}(x)$, where x is the concentration of the observed chemical species (Krishna *et al.*, 2006). For further information regarding the analysis see Section *Performing computational analyses* in Supplementary text.

First, we compared the results of different promoter state evaluation methods, namely FO (which outputs identical results as TDM) and APSA, on our case study with different expression hypotheses on promoters with five TF binding sites (see Figure 1). Note that for one binding site all hypotheses yield equal results. As there is no difference in the results obtained with the additive and all-or-none expression hypotheses and no significant difference in the results obtained with the singular hypothesis, our further analyses were based only on APSA method, due to its computing speed. The reader can refer to additional results in Supplementary text that justify the suitability of APSA with respect to FO.

Next, we analysed the effects of IKK concentrations on the oscillatory dynamics in dependence on the number of modelled TF binding sites (see Figure 2), with the tuning of model parameter C (see Section *Model derivation* in Supplementary text). Results indicate that the number of TF binding sites does not affect the dynamics of the system with the additive hypothesis. However, with the all-or-none and singular hypotheses, the increase in the number of TF binding sites has

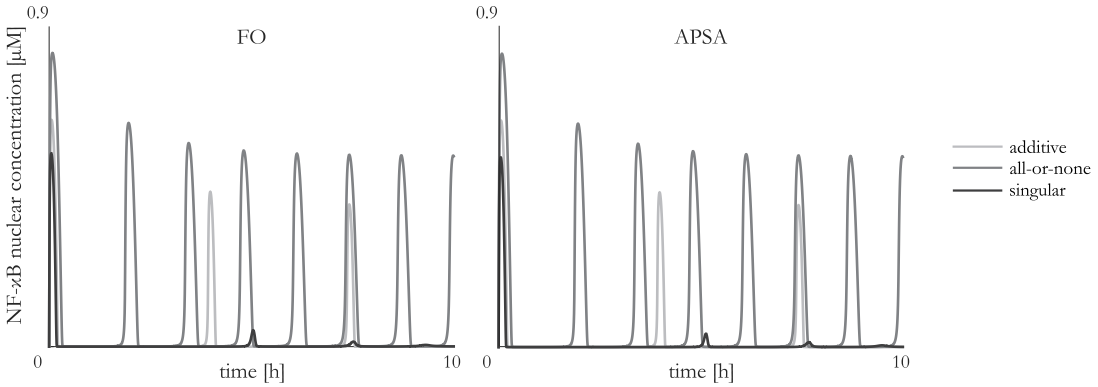

Figure 1: Comparison of results obtained with FO (a) and APSA (b) method. Simulations were performed on a model with nominal parameter values, five TF binding sites and additive, all-or-none and singular expression profile hypotheses.

an opposing effect. While it increases the oscillatory region with the all-or-none hypothesis, it decreases the region with the singular hypothesis. Moreover, all-or-none hypothesis increases the maximal amplitudes of oscillations in the middle of oscillatory region and drastically increases periods of oscillations in the close proximity of the bifurcation point. The values remain more or less pertained in other regions and with the singular hypothesis. All-or-none hypothesis, however, decreases the spikiness of obtained oscillations with at least four binding TF sites and larger IKK concentrations, i.e. far from the bifurcation point. Spikiness is pertained within the remaining oscillatory region with the singular expression hypothesis.

At last, we analysed the oscillatory dynamics around the nominal values of kinetic parameters describing the model with perturbations of two parameters at the same time and again with the variations of the number of TF binding sites. We presumed that IKK concentration is constant during the course of each simulation since we only observed the system response for a limited amount of time. According to IKK degradation rates used in (S. Zambrano, 2014), IKK concentrations do not change significantly within the observed time interval. We perturbed other parameter values by multiplying them with logarithmically spaced values from an interval $[10^{-2}, 10^{2}]$. Results of our analysis are presented in Figures 3, 4 and 5 with the additive, all-or-none and singular expression hypothesis, respectively, and for perturbations of parameters A and B (additional results obtained with FO and results for variations of different parameters are presented in Supplementary text).

Oscillatory regions follow the same trends as when changing the IKK concentrations. Similar dynamics is observed for amplitudes of oscillations with the all-or-none expression hypothesis, which are increased in the middle of oscillatory region. On the other hand, other two hypotheses more or less pertain the maximal oscillation amplitudes with different numbers of TF binding sites. Similar

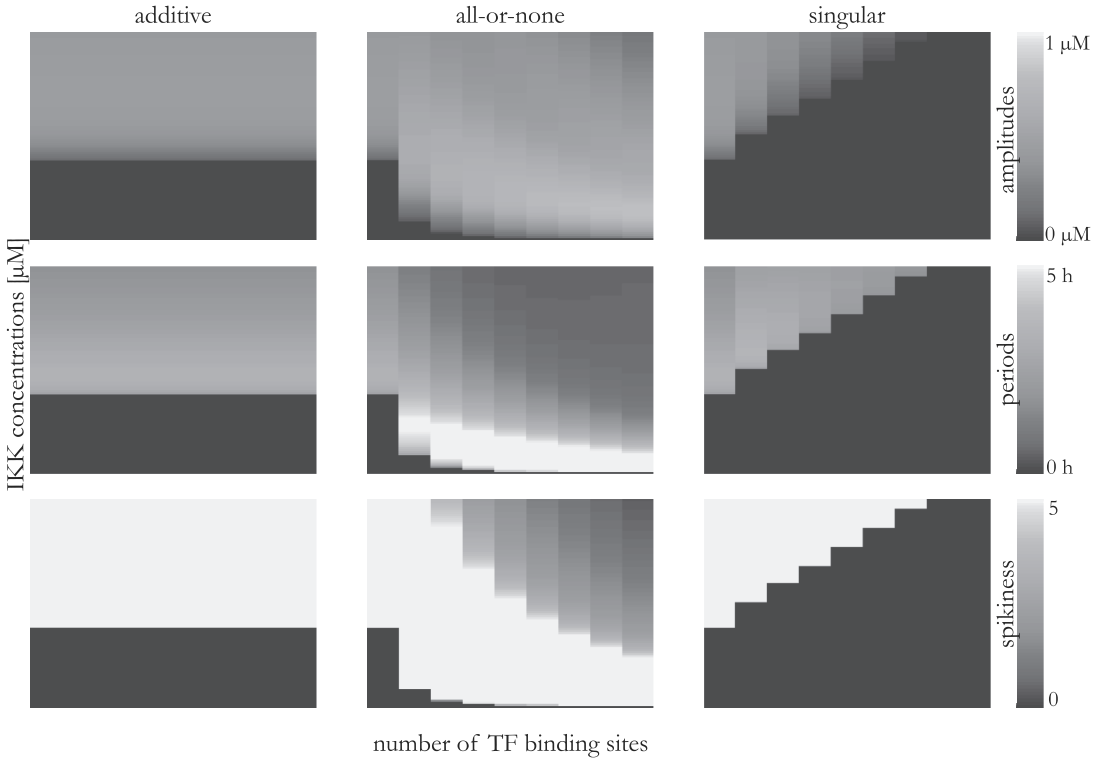

Figure 2: Effects of IKK concentrations (linear range from 0 to 1 μM) in a combination with number of TF binding sites (linear range from 1 to 10) on nuclear NF-κB oscillation *amplitudes* (top row), *periods* (middle row) and *spikiness* (bottom row) with the *additive* (left column), *all-or-none* (middle column) and *singular* (right column) expression hypothesis. Amplitude values are measured in μM, period values in hours. Black colour represents stationary behaviour (no oscillations).

behaviour is reflected by periods of oscillations. Their values are relatively low near the bifurcation point and increase with the distance from it. Higher numbers of TF binding sites increase this distance with the all-or-none hypothesis and decrease it with the singular hypothesis (especially near bifurcations caused by increasing parameter A). While spikiness of oscillations is pertained with the additive and singular hypotheses, it is decreased near bifurcation points with the all-or-none hypothesis.

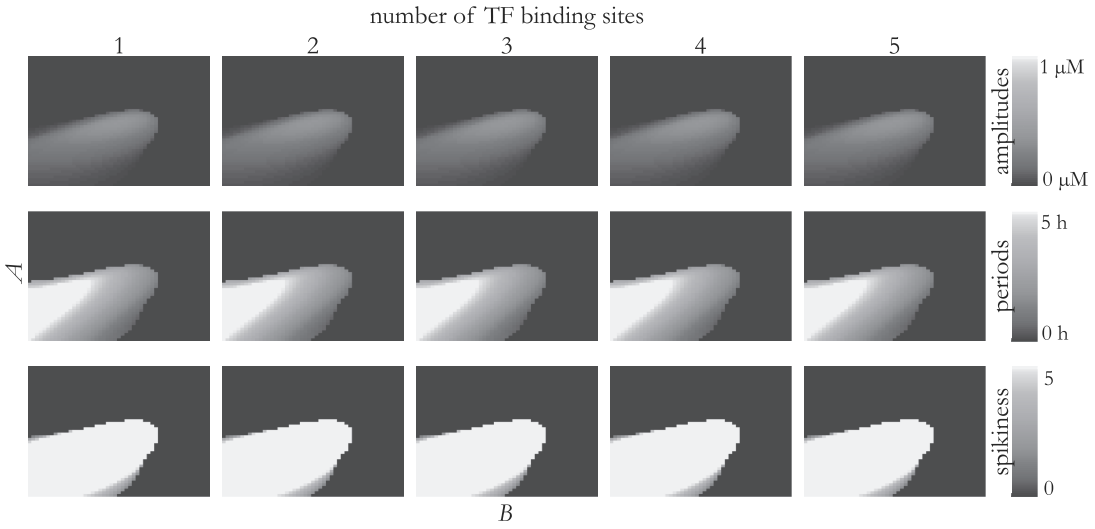

Figure 3: Effects of A (y axis) and B (x axis) parameter perturbations on nuclear NF-κB oscillation *amplitudes* (top row), *periods* (middle row) and *spikiness* (bottom row) for different numbers of TF binding sites ranging from 1 (first column) to 5 (last column) with the *additive* expression hypothesis. Amplitude values are measured in μM, period values in hours. Perturbations were performed by multiplying the parameters with logarithmically spaced values from an interval $[10^{-2}, 10^{2}]$. Black colour represents stationary behaviour (no oscillations).

4. Conclusion

Even though the majority of state-of-the-art computational analyses omit explicit modelling of individual TF binding sites within the cluster, we demonstrated that variations in dynamics caused by different promoter expression hypotheses and different numbers of observed TF binding sites should not be ignored in the exact quantitative modelling of biological systems. We described a computational framework that can be applied to address this problem. We additionally proposed two novel methods within the framework, i.e. APSA to efficiently evaluate promoter states, and EPBM to describe the transcriptional activity of the observed promoters. Introduced framework allows us to evaluate the dynamics of the observed system in dependence on different promoter expression hypotheses, different numbers of observed TF binding sites and different variations of other parameters describing the system in a straightforward manner.

We demonstrated the application of the proposed framework on a case study of NF-κB oscillatory response analysis. Our results indicate that variation of the number of TF binding sites in a combination with different promoter expression hypotheses, i.e. additive, singular and all-or-none, yield substantially different results in essentially the same system. We showed that the number of TF binding sites may indeed play a key role in determining the overall system response. On

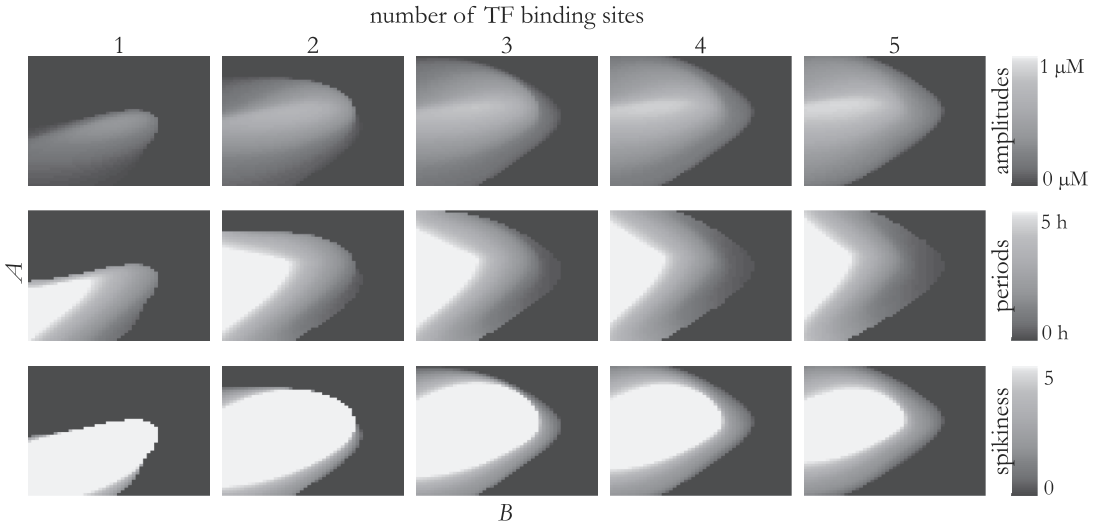

Figure 4: Same as Figure 3 except with the *all-or-none* expression hypothesis.

one hand the number of TF binding sites does not change the dynamics of the system in any way using additive hypothesis. On the other hand, the number of observed TF binding sites considerably changes system's response with the remaining two hypotheses. While the oscillatory ranges decrease with respect to the number of observed TF binding sites with the singular hypothesis, they increase with the all-or-none hypothesis. We reason that in order to perform complete, detailed analyses of regulatory networks or their smaller sub-networks, presence and effects of multiple TF binding sites on the system dynamics should not be omitted.

All of the approaches for evaluating promoter states presented within our computational framework presume that concentrations of regulating TFs are much higher than binding sites concentrations. This presumption is commonly applied to the modelling of GRNs and metabolic networks together with different acknowledged mathematical formalisms, such as Hill equations (Alon, 2007) or Michaelis-Menten kinetics, in which we presume that the substrate concentration is much higher than the enzyme concentration. However, this presumption is not always valid, especially when we are dealing with large clusters of TF binding sites and/or several copies of each promoter regulated by these sites. FO and TDM models are unable to retain the quantitative accuracy in such cases. APSA however, can be extended to cope with such scenarios, but only for small numbers of different TFs, which bind to the same binding site type (see derivation in Supplementary text for one TF per type and (Wang, 1995) for two TFs per type). The cost of this accuracy is additional complexity of equations behind the model. However, these do not considerably increase the computational complexity of simulations.

Finally, we end this discussion with suggestions of further potential applica-

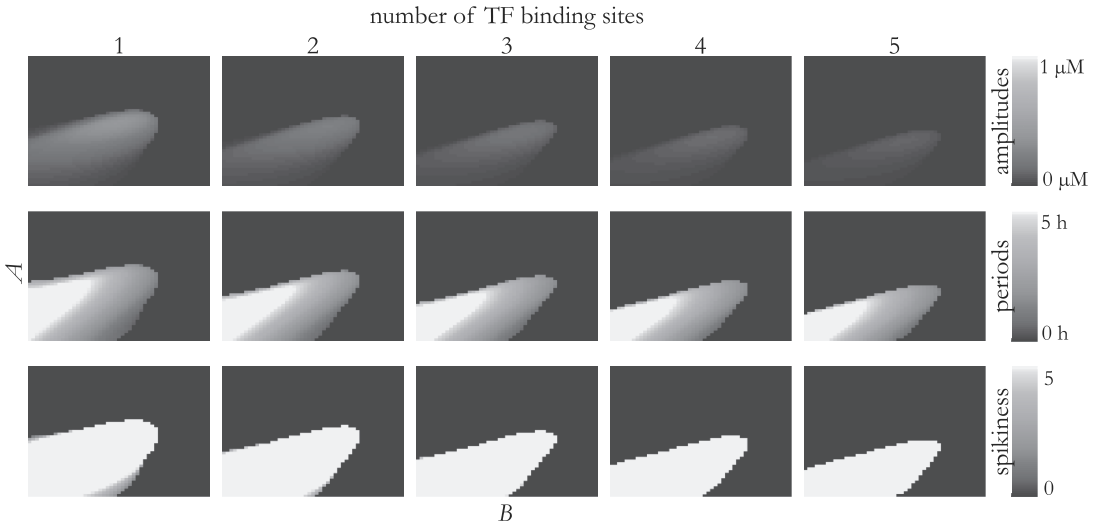

Figure 5: Same as Figure 3 except with the *singular* expression hypothesis.

tions of the presented approaches. Research and analysis of GRNs goes hand in hand with computational modelling, which is sometimes accompanied by easy-to-use computational tools. Such tools (e.g. Copasi (Mendes *et al.*, 2009), TinkerCell (Chandran *et al.*, 2009) or CellDesigner (Funahashi *et al.*, 2006)) already provide a variety of functionality regarding the construction and analysis of GRNs behind intuitive graphical user interfaces with the purpose to speed up and simplify the process of computational analyses end user would like to perform. However, to our knowledge, none of the existing tools provide accurate modelling of the multiple TF binding sites at the present day. Following the recent trends in synthetic biology, which demonstrate progress in the engineering of several different types of monomeric non-cooperative TFs, we anticipate the need for integration of such modelling techniques. Although it is sometimes possible to work around these limitations with additional assumptions about the model itself, this would not be necessary if computational frameworks as the one presented in this text were included as one of the many embedded features.

Acknowledgement

The research was partially supported by the scientific-research programme Pervasive Computing (P2-0359) financed by the Slovenian Research Agency in the years from 2009 to 2017 and by the basic research and application project Designed cellular logic (J1-6740) financed by the Slovenian Research Agency in the years from 2014 to 2017.

Author disclosure statement

The authors declare that no competing financial interests exist.

References

Alon, U. 2007. *An introduction to systems biology*. London, UK: Chapman & Hall/CRC.

Basak, S., Behar, M., & Hoffmann, A. 2012. Lessons from mathematically modeling the NF-kB pathway. *Immunol. Rev.*, **246**, 221–238.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., J., Kondev, Kuhlman, T., & Phillips, R. 2005a. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.*, **15(2)**, 125–35.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., & Phillips, R. 2005b. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, **15(2)**, 116–24.

Chandran, D., Bergmann, F., & Sauro, H. 2009. TinkerCell: modular CAD tool for synthetic biology. *J. Biol. Eng.*, **3**(1), 19.

Cheong, R., Hoffmann, A., & Levchenko, A. 2008. Understanding NF-kB signaling via mathematical modeling. *Mol. Syst. Biol.*, **4**(1).

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**(6121), 819–823.

Funahashi, A., Matsuoka, Y., Jouraku, A., Kitano, H., & Kikuchi, N. 2006. CellDesigner: a modeling tool for biochemical networks. *In: Proceedings of the 38th conference on Winter simulation.*

Gaber, R., Lebar, T., Majerle, A., Šter, B., Dobnikar, A., Benčina, M., & Jerala, R. 2014. Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat. Chem. Biol.*, **10**, 203–208.

Garg, A., Lohmueller, J. J., Silver, P. A., & Armel, T. Z. 2012. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res.*, **40(15)**, 7584–7595.

Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.-US*, **81**, 2340–2361.

Giorgetti, L., Siggers, T., Tiana, G., Caprara, G., Notarbartolo, S., Corona, T., Pasparakis, M., Milani, P., Bulyk, M. L., & Natoli, G. 2010. Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Mol. Cell*, **37**(3), 418 – 428.

Gommans, W. M., Haisma, H. J., & Rots, M. G. 2005. Engineering zinc finger protein transcription factors: the therapeutic relevance of switching endogenous gene expression on or off at command. *J. Mol. Biol.*, **354**(3), 507–519.

Hasty, J., Isaacs, F., Dolnik, M., McMillen, D., & Collins, J. J. 2001. Designer gene networks: towards fundamental cellular control. *Chaos*, **11**, 207–220.

Krishna, S., Jensen, M. H., & Sneppen, K. 2006. Minimal model of spiky oscillations in NF-kB signaling. *P. Natl. A. Sci. USA*, **103**(29), 10840–10845.

Le Novere, N. 2015. Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.*, **16**(3), 146–158.

Lebar, T., Bezeljak, U., Golob, A., Jerala, M., Kadunc, L., Pirš, B., Stražar, M., Vučko, D., Zupančič, U., Benčina, M., Forstnerič, V., Gaber, R., Lonzarič, J., Majerle, A., Oblak, A., Smole, A., & Jerala, R. 2014. A bistable genetic switch based on designable DNA-binding domains. *Nature Commun.*, **5**, 1–13.

Lorberbaum, D. S., & Barolo, S. 2013. Gene regulation: When analog beats digital. *Curr. Biol.*, **23**(23), R1054–R1056.

Mendes, P., Hoops, S., Sahle, S., Gauges, R., Dada, J., & Kummer, U. 2009. Computational modeling of biochemical networks using COPASI. *Method. Mol. Biol.*, **500**, 17–59.

Mengel, B., Hunziker, A., Pedersen, L., Trusina, A., Jensen, M. H., & Krishna, S. 2010. Modeling oscillatory control in NF-kB, p53 and Wnt signaling. *Curr. Opin. Genet. Dev.*, **20**, 656–664.

Murphy, K. F., Balazsi, G., & Collins, J. J. 2007. Combinatorial promoter design for engineering noisy gene expression. *P. Natl. A. Sci. USA*, **104**, 12726–12731.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152(5)**, 1173–1183.

Rao, C. V., & Arkin, A. P. 2003. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *J. Chem. Phys.*, **118**, 4999–5010.

S. Zambrano, M. E. Bianchi, A. Agresti. 2014. A simple model of NF-kB dynamics reproduces experimental observations. *J. Theor. Biol.*, **347**, 44–53.

Sauro, H. M. 2012. *Enzyme kinetics for systems biology*. Seattle, WA: Future Skill Software (Ambrosius Publishing).

Spitz, F., & Furlong, E. E. M. 2012. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.

Stewart-Ornstein, J., Nelson, C., DeRisi, J., Weissman, J. S., & El-Samad, H. 2013. Msn2 coordinates a stoichiometric gene expression program. *Curr. Biol.*, **23**(23), 2336–2345.

Wang, Z.-X. 1995. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.*, 111–114.

Weiss, J. N. 1997. The Hill equation revisited: uses and misuses. *FASEB J.*, **11**, 835–841.

Werner, M., Ernberg, I., Zou, J., Almqvist, J., & Aurell, E. 2007. Epstein-Barr virus latency switch in human B-cells: a physico-chemical model. *BMC Syst. Biol.*, **1**(40).

Zeiser, S., Müller, J., & Liebscher, V. 2007. Modeling the Hes1 oscillator. *J. Comp. Biol.*, **14**, 984–1000.

Address correspondence to:
Miha Moškon
Faculty of Computer and Information Science,
University of Ljubljana,
Večna pot 113, 1000 Ljubljana
Slovenia
E-mail: miha.moskon@fri.uni-lj.si

Computational framework for modelling multiple non-cooperative transcription factor binding and its application to the analysis of NF-κB oscillatory response

Supplementary text

Manca Bizjak, Miha Mraz, Nikolaj Zimic, and Miha Moškon*

[1] Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia

* corresponding author
e-mail: miha.moskon@fri.uni-lj.si

1 From fractional occupancy to average promoter state approximation

Average promoter state approximation (APSA) evaluates promoter's state in dependence on the concentrations of TFs, TF binding sites and their corresponding binding constants. In each simulation step the average promoter state is defined with the average number of each TF type bound to its binding sites. We can describe this state with the use of fractional occupancy (FO) as

$$\overline{n}_A = \sum_{i=0}^{n} i \cdot P(n_A = i), \tag{1}$$

where $\overline{n}_A$ represents the average number of binding sites occupied by TF A, n number of observed TF binding sites and $P(n_A = i)$ probability that i binding sites are occupied by TF A. The probability can be expressed as

$$P(n_A = i) = \frac{W(n_A = i)}{\sum_{j=0}^{n} W(n_A = j)}, \tag{2}$$

where $W(\cdot)$ represents the weight of each promoter state. Extending Equation 1, we obtain

$$\overline{n}_A = \frac{\sum_{i=0}^{n} i \cdot W(n_A = i)}{\sum_{i=0}^{n} W(n_A = i)}. \tag{3}$$

When we are dealing with only one TF type, weights can be expressed as

$$W(n_A = i) = \binom{n}{i} K_A^i [A]^i, \tag{4}$$

where K_A is binding constant between TF A and its binding sites and $[A]$ is TF concentration. Note again that we presume there is no cooperativity among TFs of the same type. For the sake of convenience we rename the product $K_A[A]$ to α. We can now rewrite the Equation 3 as

$$\begin{aligned}
\overline{n}_A &= \frac{\binom{n}{1}\alpha + 2\binom{n}{2}\alpha^2 + \cdots + (n-1)\binom{n}{n-1}\alpha^{n-1} + n\binom{n}{n}\alpha^n}{1 + \binom{n}{1}\alpha + \binom{n}{2}\alpha^2 + \cdots + \binom{n}{n-1}\alpha^{n-1} + \binom{n}{n}\alpha^n} \\
&= \frac{n\alpha + 2\frac{n(n-1)}{2}\alpha^2 + \cdots + (n-1)n\alpha^{n-1} + n\alpha^n}{1 + n\alpha + \frac{n(n-1)}{2}\alpha^2 + \cdots + n\alpha^{n-1} + \alpha^n} \\
&= n\alpha \frac{1 + 2\frac{(n-1)}{2}\alpha + \cdots + (n-1)\alpha^{n-2} + \alpha^{n-1}}{1 + n\alpha + \frac{n(n-1)}{2}\alpha^2 + \cdots + n\alpha^{n-1} + \alpha^n} \frac{1+\alpha}{1+\alpha} \\
&= n\alpha \frac{1 + n\alpha + \frac{n(n-1)}{2}\alpha^2 + \cdots + n\alpha^{n-1} + \alpha^n}{1 + n\alpha + \frac{n(n-1)}{2}\alpha^2 + \cdots + n\alpha^{n-1} + \alpha^n} \frac{1}{1+\alpha} \\
&= \frac{n\alpha}{1+\alpha} \\
&= n\frac{K_A[A]}{1 + K_A[A]}.
\end{aligned} \tag{5}$$

A similar, but drastically more complex derivation can be performed for scenarios with competition of different TF types for the same binding sites, which we do not describe here. Instead, we provide a *Mathematica* script that conducts the derivation [1]. In general, the average number of bound TFs A_i among m types of TFs that competitively bind to n binding sites can be expressed as

$$\overline{n}_{A_i} = n\frac{K_{A_i}[A_i]}{1 + \sum_{j=1}^{m} K_{A_j}[A_j]}. \tag{6}$$

At last, we can define the average promoter state as a set of average numbers of bound TFs of each type:

$$\overline{s} = \{\overline{n}_{A_1}, \overline{n}_{A_2}, \ldots, \overline{n}_{A_m}\}. \tag{7}$$

2 Evaluating the time complexity

We will evaluate and compare the time complexity of FO and APSA approach in terms of the *big O notation*. Note that thermodynamic modelling (TDM)

[1] The code used in this paper is available at http://lrss.fri.uni-lj.si/bio/material/mbs.zip under the Creative Commons Attribution license.

has the same time complexity as fractional occupancy approach and is therefore omitted in the following evaluation.

With FO we must evaluate probabilities of each among $\binom{n+m}{m}$ promoter states in each iteration of the simulation, where n denotes the number of transcription factor (TF) binding sites to which m TFs bind competitively. After evaluation of $\binom{n+m}{m}$ weights, probabilities of promoter states are calculated. In each iteration these probabilities are used in a weighted sum over $\binom{n+m}{m}$ elements to obtain gene expression rate in dependence on the input TFs' concentrations. If we presume that numerical integration of ODEs describing our system is constant (for example, if we use Euler's method), and that the number of simulation steps is fixed, then time complexity of FO is expressed as

$$T(m,n,N) = O\left(N \cdot \binom{n+m}{m}\right) = O\left(N \cdot \frac{(n+m)!}{m! \cdot n!}\right), \tag{8}$$

where N is the number of simulation iterations.

APSA on the other hand only evaluates the average number of promoter bound species for each TF type. If we again presume that numerical integration of ODEs describing our system is constant, then time complexity of APSA is

$$T(m,N) = O\left(N \cdot m\right). \tag{9}$$

In this case, time complexity is drastically reduced compared to FO or TDM approach (see Figure 1 for time complexity of a single iteration).

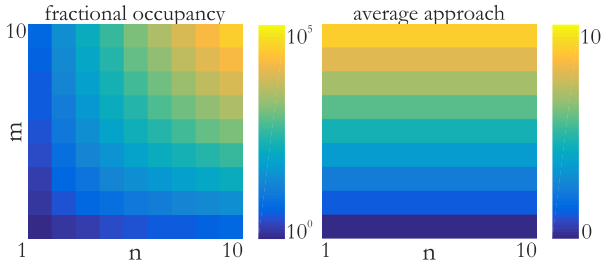

Figure 1: Time complexity of a single simulation iteration. Figures show the time complexity of FO approach (left) and APSA (right) in dependence on the number of observed binding sites (n) and number of different TFs that bind competitively to these sites (m).

On the other hand, we are often presented with a biological system in which parameters m and n are known and usually relatively small constants (case study introduced in the main text presumes $m = 2$ and n between 1 and 10). While it is obvious from Equations 8 and 9 that m and n greatly influence the time complexity (especially in the FO approach), we must not forget that for accurate results, the number of simulation iterations N is usually a much

larger number, which ultimately becomes a limiting factor when running simulations. Figure 2 presents the time complexity of described approaches for a larger number of simulation iterations.

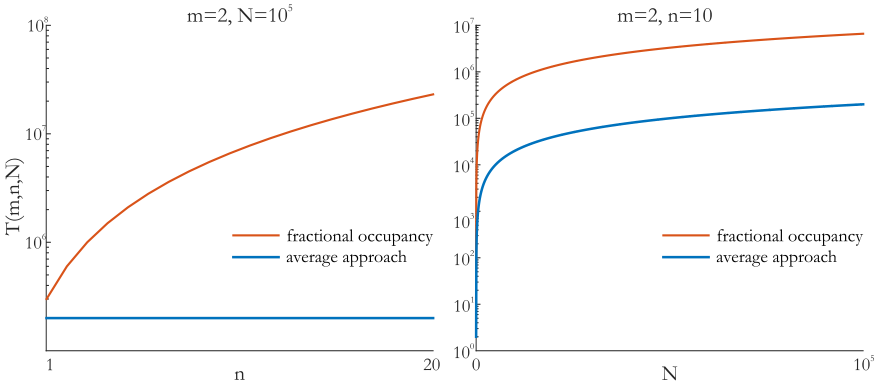

Figure 2: Time complexity for a larger number of simulation iterations. Figures show time complexity of FO approach and APSA in dependence on the number of observed binding sites (n) with two competitive TFs ($m = 2$) in $N = 10^5$ simulation iterations (left), and time complexity of FO approach and APSA in dependence on the number of iterations (N) with two competitive TFs ($m = 2$) for ten binding sites ($n = 10$).

3 NF-κB model derivation

We describe time-course of nuclear concentrations of NF-κB (N_n) and cytoplasmic concentrations of IκBα (I) and its transcript (I_m) with the following system of ordinary differential equations (ODEs):

$$\begin{aligned} \frac{dN_n}{dt} &= k_{Nin}(N_{tot} - N_n)\frac{K_I}{K_I + I} - k_{Iin}\frac{IN_n}{K_N + N_n}, \\ \frac{dIm}{dt} &= P_{tot}k_t\rho - \gamma_m I_m, \\ \frac{dI}{dt} &= k_{tl}I_m - \alpha(N_{tot} - N_n)\frac{I}{K_I + I}, \end{aligned} \tag{10}$$

where P_{tot} is the total number of the NFKBIA gene copies, ρ is predicted transcriptional activity of a promoter, k_t is transcription rate, k_{tl} is translation rate, γ_m is transcript degradation rate, α is degradation rate of IκBα in the IκBα:NF-κB complex caused by IKK ($\alpha = 1.05 \cdot IKK\text{min}^{-1}$, where IKK denotes IKK concentration), k_{Nin} and k_{Iin} are transfer rates of NF-κB and IκBα from cytoplasm to nucleus, respectively, and N_{tot} is the total (nuclear and cytoplasmic)

parameter	value
k_t	$1.03\mu M^{-1}\text{min}^{-1}$
k_{tl}	0.24min^{-1}
γ_m	0.017min^{-1}
k_f	$30\mu M^{-1}\text{min}^{-1}$
k_{fn}	0.03min^{-1}
k_b	0.03min^{-1}
k_{bn}	$30\mu M^{-1}\text{min}^{-1}$
k_{Nin}	0.08min^{-1}
k_{Iin}	0.018min^{-1}
k_{NIout}	0.83min^{-1}
k_{Iout}	0.012min^{-1}
N_{tot}	$1\mu M$
P_{tot}	1
IKK	$0.5\mu M$
K_d	$0.2\mu M$

Table 1: Parameter values used in the basic model derived from [1] and [2].

concentration of NF-κB. Parameters K_I and K_N denote the dissociation constants of cytoplasmic and nuclear IκBα:NF-κB complex, respectively, and can be expressed as $K_I = (k_b + \alpha)/k_f$ and $K_N = (k_{bn} + k_{NIout})/k_{fn}$. Here k_b and k_f denote cytoplasmic IκBα:NF-κB complex on and off rates, respectively, and k_{bn} and k_{fn} nuclear IκBα:NF-κB complex on and off rates, respectively, and k_{NIout} transfer rate of IκBα:NF-κB complex from nucleus to cytoplasm. For parameter values used in our analyses see Table 1. Note that parameter K_d represents TF dissociation constant and is not used directly in the ODEs, but is used to determine promoter activity ρ in each simulation step.

In order to decrease the number of free parameters, we rescale the observed variables and introduce dimensionless time variable τ:

$$\begin{aligned} \tau &= \gamma_m t, \\ x &= \frac{N_n}{N_{tot}}, \\ y &= \frac{\gamma_m I_m}{k_t P_{tot}}, \\ z &= \frac{\gamma_m^2 I}{k_t k_{tl} P_{tot}}. \end{aligned} \tag{11}$$

The above transformations give us the following system of differential equations:

$$\begin{aligned}\frac{dx}{d\tau} &= A(1-x)\frac{\epsilon}{\epsilon+z} - Bz\frac{x}{\delta+x},\\ \frac{dy}{d\tau} &= \rho - y,\\ \frac{dz}{d\tau} &= y - C(1-x)\frac{z\epsilon}{\epsilon+z},\end{aligned} \tag{12}$$

where parameters A, B, C, δ and ϵ are defined as:

$$\begin{aligned}A &= \frac{k_{Nin}}{\gamma_m},\\ B &= \frac{k_{Iin}k_t k_{tl}P_{tot}}{\gamma_m^3 N_{tot}},\\ C &= \frac{\alpha N_{tot}}{\gamma_m K_I},\\ \delta &= \frac{K_N}{N_{tot}},\\ \epsilon &= \frac{K_I\gamma_m^2}{k_t k_{tl}P_{tot}}.\end{aligned} \tag{13}$$

According to Table 1, nominal parameter values are as follows: $A \approx 45.4$, $B \approx 905.7$, $C \approx 1670$, $\delta \approx 0.029$ and $\epsilon \approx 2.16 \cdot 10^{-5}$. We can analyse how dynamics change with respect to the parameter A indirectly, via parameter k_{Nin}. Similarly, we indirectly vary parameters B, C, δ and ϵ via parameters k_{Iin}, k_f, K_N and K_I, respectively.

4 Performing computational analyses

We ran our computational analyses in Matlab. A single simulation was run for every pair of parameter values with numerical integration of ODEs describing the model (see Equations 12). Before numerical integration, promoter state and its activity ρ were determined in each iteration of the ODE solver either with fractional occupancy or with the average promoter state approximation approach.

Simulation results presenting time-course of nuclear NF-κB concentrations were analysed for potential oscillations. In case of sustained oscillations we measured their periods, amplitudes and spikiness. First part of the analyses was identification of peaks in the analysed signal (time-course of species concentrations). We calculated amplitudes as differences between maximal and minimal signal values within one period of a signal. The behaviour was estimated to be oscillatory if amplitudes of the last period in the observed signal were larger than a predefined threshold (note that the treshold value was determined throuh a trial and error process). Spikiness was evaluated using the measure proposed in [1], i.e. $(\max(x) - \min(x))/\text{mean}(x)$, where x is the concentration of the

observed chemical species. Spikiness is displayed when calculated value equals at least 2. We performed simulations for longer time periods, but the signal was analysed only after the stable state was achieved for one oscillation period, which never exceeds 10 hours. Note that our model is able to produce either steady stable state or stable oscillations after initial transient dynamics die out.

5 Additional results

The main text presents the results obtained with variations of the numbers of TF binding sites and parameters A and B on the basis of average promoter state approximation. Here we additionally present the results of variations of the same parameters, but obtained with fractional occupancy in order to observe the differences between the approaches on a wider scale (see Figures 3, 4 and 5). However, we ran simulations for smaller resolution of varied parameter values (25x25 instead of 50x50) due to larger computational complexity of fractional occupancy.

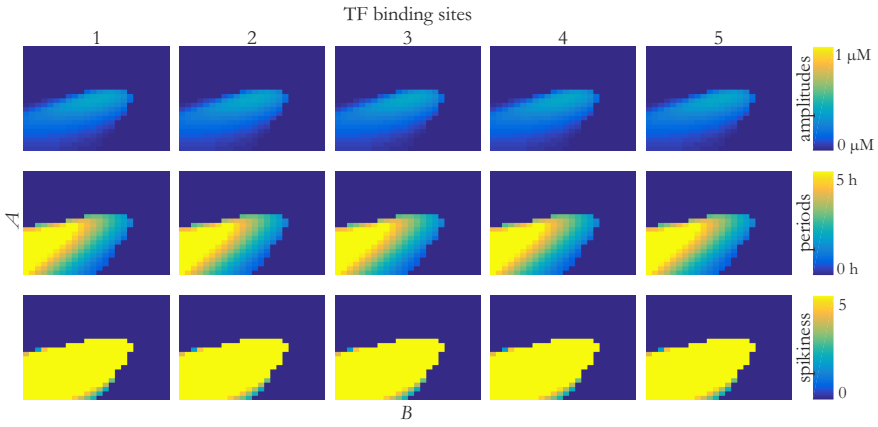

Figure 3: Effects of A (y axis) and B (x axis) parameter perturbations on nuclear NF-κB oscillation *amplitudes* (top row), *periods* (middle row) and *spikiness* (bottom row) for different numbers of TF binding sites ranging from 1 (first column) to 5 (last column) for *additive* expression scenario. Amplitude values are measured in μM, period values in hours. Perturbations were performed by multiplying the parameters with logarithmically spaced values from an interval $[10^{-2}, 10^{2}]$. Results were obtained with fractional occupancy approach.

We additionally analysed effects of variations of parameters B and C on the oscillatory dynamics of observed system obtained with the average promoter state approximation approach (see Figures 6, 7 and 8).

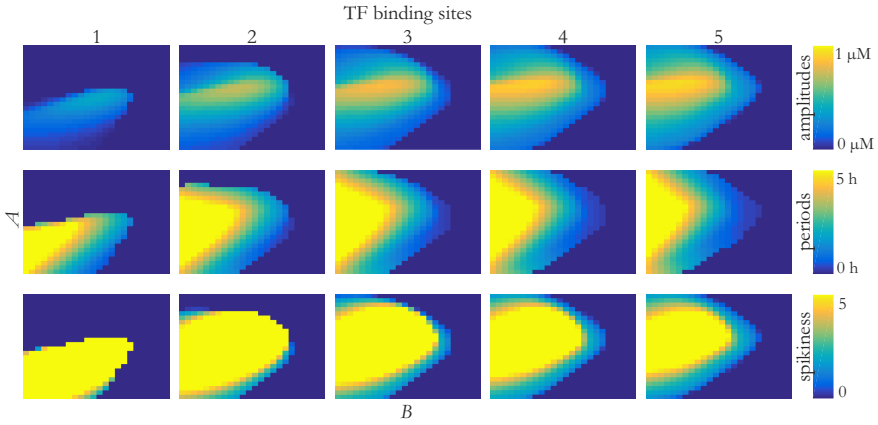

Figure 4: Same as Figure 3, except for *all-or-none* expression scenario.

6 Derivation of exact expression for describing binding of transcription factor to its binding site

Let's presume that $[A]$ denotes TF concentration and $[B]$ its binding site concentration, i.e. the number of TF binding sites per each promoter multiplied by promoter concentration. We can describe binding processes with the reversible chemical reaction:

$$A + B \overset{K_d}{\leftrightarrow} AB,$$

where $K_d = k_{off}/k_{on}$ is the dissociation constant. Conservation of mass requires that

$$[A] + [AB] = [A]_0,$$
$$[B] + [AB] = [B]_0.$$

Steady state concentrations can be expressed with

$$[AB] = \frac{[A][B]}{K_d}.$$

The following equations are derived when no further presumptions are made:

$$[AB] = \frac{([A]_0 - [AB])\,([B]_0 - [AB])}{K_d}$$

$$[AB] = \frac{[A]_0[B]_0 - [AB][B]_0 - [AB][A]_0 + [AB]^2}{K_d}$$

$$0 = \frac{[A]_0[B]_0 - [AB]([B]_0 + [A]_0 + K_d) + [AB]^2}{K_d}$$

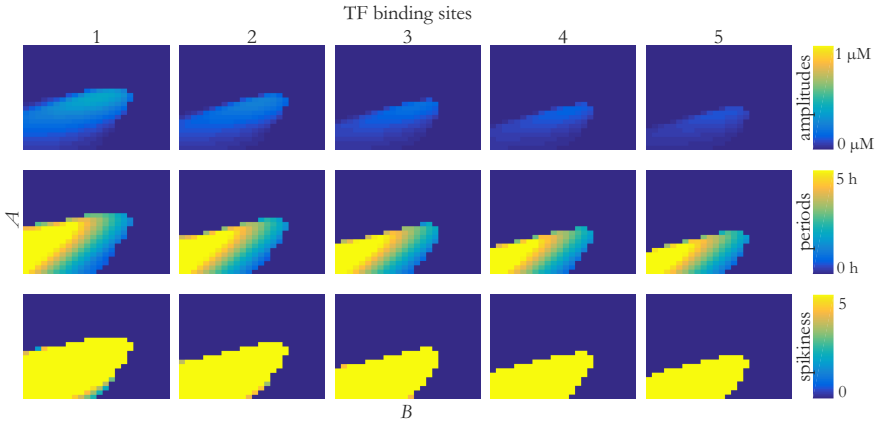

Figure 5: Same as Figure 3, except for *singular* expression scenario.

We can eliminate the scenario in which K_d equals zero and obtain a quadratic equation with solutions

$$[AB]_{1,2} = \frac{[A]_0 + [B]_0 + K_d \pm \sqrt{([B]_0 + [A]_0 + K_d)^2 - 4[A]_0[B]_0}}{2},$$

which define concentrations of occupied TF binding sites and among which only the solution that satisfies conditions $[AB] > 0$, $[AB] < [A]_0$ and $[AB] < [B]_0$ is valid.

Let $[AB]$ denote the valid concentration. The average number of TFs bound to each promoter can be expressed as $n = [AB]/[P]$, where [P] presents the promoter concentration.

7 Description of Matlab and Mathematica code

The code used in this paper is available at http://lrss.fri.uni-lj.si/bio/material/mbs.zip under the Creative Commons Attribution license. Below is its brief documentation.

- 'params.m': saves parameter values to a file 'params.mat', which is used by other files (Matlab).
- 'measureOsc.m': determines if the input signal exhibits oscillatory behaviour and evaluates the features of potential oscillations (frequency, period, amplitude and spikiness) (Matlab).
- 'modelFO.m': fractional occupancy model of NF-κB - IKBα interactions used for basic tests (Matlab).

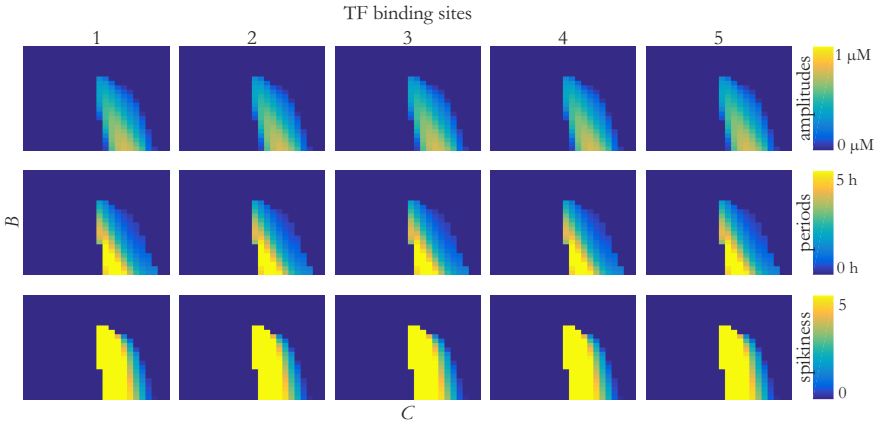

Figure 6: Effects of B (y axis) and C (x axis) parameters perturbations on nuclear NF-κB oscillation *amplitudes* (top row), *periods* (middle row) and *spikiness* (bottom row) for numbers of TF binding sites ranging from 1 (first column) to 5 (last column) for *additive* expression scenario. Amplitude values in μM, period values in hours. Perturbations were performed by multiplying the parameters with logarithmically spaced values from an interval $[10^{-2}, 10^{2}]$.

- 'modelAvgBasic.m': average promoter state approximation model of NF-κB - IKBα interactions used for basic tests (Matlab).
- 'modelAvg.m': average promoter state approximation model of NF-κB - IKBα interactions used by other Matlab files in more complex analyses (Matlab).
- 'simulateForBSRangeIKK.m': analysis of different IKK concentrations given the promoter expression scenario, number of interpolation points for IKK concentrations and number of the observed TF binding sites (Matlab).
- 'simulateForBSRangeParams.m': analysis of different parameter ranges given the promoter expression scenario, number of interpolation points for IKK concentrations and number of the observed TF binding sites (Matlab).
- 'derive_avg.nb': derivation of average state approximation with fractional occupancy approach for one type of TF (Mathematica).
- 'derive_avg_2.nb': derivation of average state approximation with fractional occupancy approach for two types of TF that compete for the same binding sites (Mathematica).
- 'derive_avg_3.nb': derivation of average state approximation with fractional occupancy approach for three types of TF that compete for the

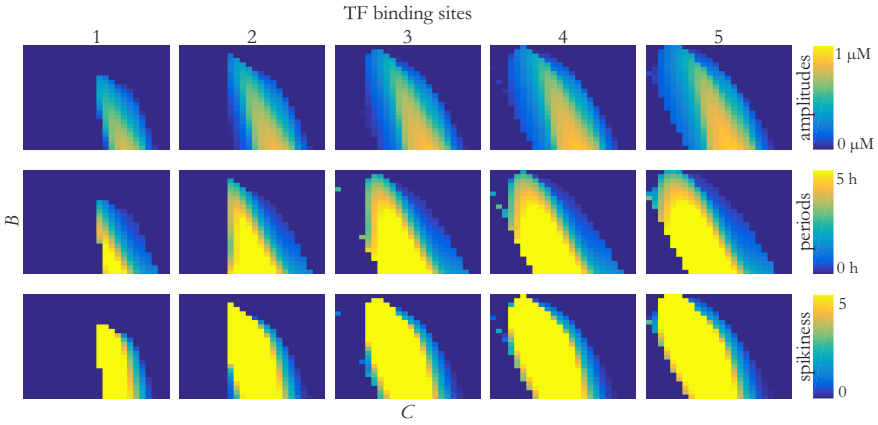

Figure 7: Same as Figure 6, except for *all-or-none* expression scenario.

same binding sites (Mathematica).

References

[1] S. Krishna, M. H. Jensen, and K. Sneppen, "Minimal model of spiky oscillations in NF-kB signaling," *P. Natl. A. Sci. USA*, vol. 103, no. 29, pp. 10 840–10 845, 2006.

[2] L. Giorgetti, T. Siggers, G. Tiana, G. Caprara, S. Notarbartolo, T. Corona, M. Pasparakis, P. Milani, M. L. Bulyk, and G. Natoli, "Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs," *Mol. Cell*, vol. 37, no. 3, pp. 418 – 428, 2010.

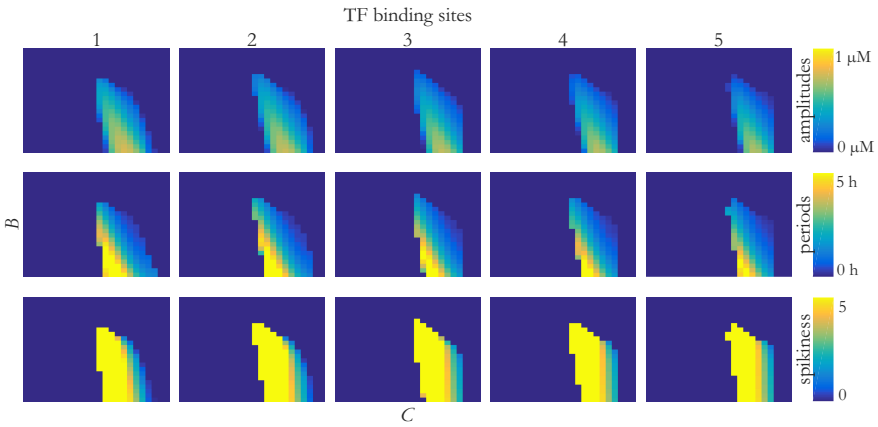

Figure 8: Same as Figure 6, except for *singular* expression scenario.