Computational approaches for quantitative modeling of gene regulatory networks with multiple transcription factor DNA binding site repeats

Manca Bizjak, Miha Mraz, Nikolaj Zimic, Miha Moškon
Faculty of Computer and Information Science, University of Ljubljana, Slovenia

Problem

Explicit modeling of multiple transcription factor (TF) binding sites is **computationally challenging**, because the number of promoter states we have to regard (N) is exponentially proportional to the number of binding sites ($N = (t+1)^n$, where t is the number of different transcription factors). Computational modeling of such systems presents a challenge for conventional approaches when **no or very low cooperativity** among binding sites exists, since their dynamics cannot be approximated with Hill's equations. However, multiple TF binding sites may **substantially affect the dynamics** of observed systems and should be regarded explicitly. This might be the case in several natural (Epstein-Barr, NF-kB) or engineered biological systems (regulation with synthetic TFs, such as Zinc Fingers, TAL effectors or CRISPR based repressors).

Methods

In order to simplify the system we presume that chemical reactions can be separated in two groups - **slow reactions** (*transcription, translation, degradation*) and **fast reactions** (*binding to and unbinding from promoters*) and that concentrations of binding sites are much lower than concetrations of TFs.

Fractional occupancy

First, we determine the weights of all valid promoter states. **Promoter state** is defined by the number of bound activators and repressors (e.g. S_{21} denotes a state in which 2 activators and 1 repressor are bound to the promoter).

$$W(s_{ij}) = N_{ij}(K_{A,i} \cdot [A]^n \cdot K_{R,j} \cdot [R]^m).$$

N_{ij} is the total number of all arrangements of i activaros and j repressors to n binding sites. Then, we calculate the probability of each promoter state:

$$P(s_{ij}) = \frac{W(s_{ij})}{\sum_{j=1}^{N} W(s_j)}.$$

Transcriptional activity of promoter at a given moment is then expressed as the sum of prodcucts of state probabilities and their expression profiles:

$$\overline{k_{trsc}} = \sum_{j=1}^{N} k_{trsc}(s_j) \cdot P(s_j).$$

This value is then used in the system of ODEs to deretmine changes in the current simulation step.

Thermodynamic modeling

Thermodynamic modeling (TM) is similar to fractional occupancy (FO). However, it relies on the estimation of binding free energies of each promoter state. The procedure is the same as in the fractional occupancy, except the weights are calculated differently:

$$W(s_{ij}) = N_{ij} \cdot e^{a \ln K_A} \cdot [A]^n \cdot e^{r \ln K_R} \cdot [R]^m.$$

FO and TM have already been successfully applied to the modeling of gene regulatory networks with multiple binding sites. We additionally introduce two general and flexible approaches.

Average promoter state approximation

A simplification of FO and TM, which is asymptotically **faster**. Promoter states are not analysed as a whole - each binding site is analysed seperately (whether it is free or occupied with a certain transcription factor), with the probability of its state expressed as

$$P(BS_{t_i}) = \frac{W(BS_{t_i})}{1 + \sum_{j=1}^{t} W(BS_{t_j})}.$$

We then determine the average promoter state with the equations of the form

$$n_{t_i} = n \cdot P(BS_i).$$

Describing promoters' expression

We use matrix presentation to describe transcriptional activity of promoters together with all of the above approaches. Matrix element a_{ij} denotes expression when i activators and j repressors are bound.

repressors

activators

$$\begin{pmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,n} \\ a_{1,0} & \ddots & a_{1,n-1} & 0 \\ \vdots & & & \\ & \cdots & 0 & 0 \\ a_{n-1,0} & a_{n-1,1} & & \\ a_{n,0} & 0 & 0 & 0 \end{pmatrix}$$

Left: matrix presentation when activator and repressor compete for the same binding sites.

The matrix is full in the non-competitive case.

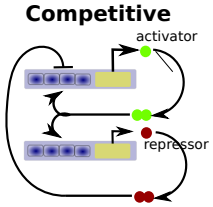
Results

Network design

Amplified negative feedback oscillator

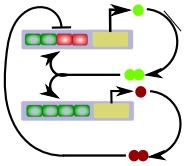
We tested all of the approaches on a simple regulatory network comprising of two genes encoding activator (A) and repressor (R). We studied how the number of modelled binding sites affects the parameter space in which the system exhibits **sustained oscillations** in scenarios where A and R compete for the binding sites at the promoter expressing A, and when there is no competition.

Competitive



A and R compete for the same binding sites at the promoter of gene encoding A.

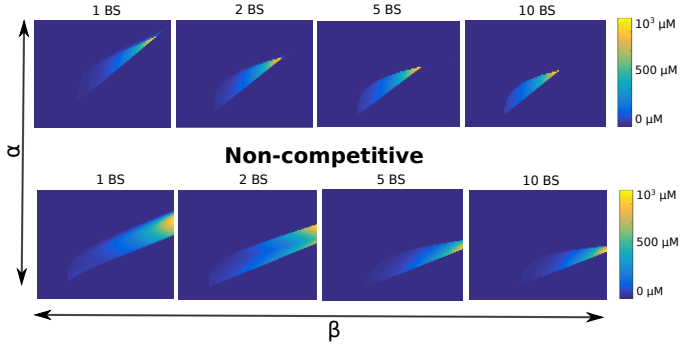
Non-competitive



A and R each bind to seperate binding sites.

Amplitudes of oscillations with respect to different numbers of modelled binding sites

Competitive



Figures above illustrate how the number of modelled BS affects system's dynamics when transcriptional response (promoter activity) is linearly proportional to the number of binding sites. We can see that the size of the oscillatory region decreases in both competitive and non-competitive scenario. All of the presented modelling approaches produced very similar results.

We encourage explicit modeling of noncooperative binding site repeats, since it allows us to obtain quantitative response with higher biological relevance. However, we must adopt one of the presented non-trivial modelling approaches to avoid potential computational infeasibility.

References

[1] Giorgetti, L., Siggers, T., Tiana, G., Caprara, G., Notarbartolo, S., Corona, T., Pasparakis, M., Milani, P.,Bulyk, M.L., Natoli, G.: Noncooperative interactions between transcription factors and clustered DNA bindingsites enable graded transcriptional responses to environmental inputs. Molecular Cell 37(3), 418–428 (2010)
[2] 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.: A bistable genetic switch based on designable DNA-binding domains. Nature Communications (2014)