August 21, 2013

Analysis of gene regulatory network modules

PROGRESS REPORT

The Heterodimer Autorepression Loop

Author: Bruno Lannoo

1 Recapitulation

1.1 Biochemical representation

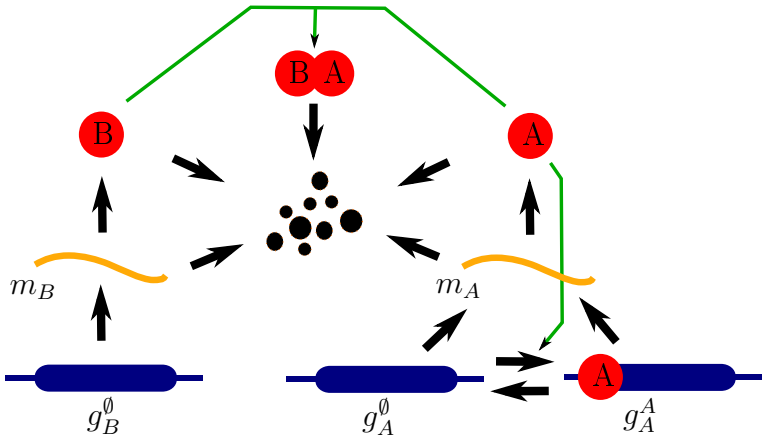


Figure 1: **Full biological representation:** Here you can see the HAL module with all the bio-chemical reactions which define it explicitly denoted. (The black circles denote left over material after degradation)

1.2 list of all the biochemical reactions

Base component	Biochemical reaction	Reactants	$\xrightarrow{\text{rate}}$	Products
Gene A:	$g_a^{\varnothing}$ transcription:	$g_a^{\varnothing}$	$\xrightarrow{\mu_{m_a}^{\varnothing}}$	$g_a^{\varnothing} + m_a$
	m_a degradation:	m_a	$\xrightarrow{\delta_{m_a}}$	$\varnothing$
	m_a translation:	m_a	$\xrightarrow{\mu_A}$	$m_a + A$
	A degradation:	A	$\xrightarrow{\delta_A}$	$\varnothing$
Gene B:	g_b transcription:	g_b	$\xrightarrow{\mu_{m_b}}$	$g_b + m_b$
	m_b degradation:	m_b	$\xrightarrow{\delta_{m_b}}$	$\varnothing$
	m_b translation:	m_b	$\xrightarrow{\mu_B}$	$m_b + B$
	B degradation:	B	$\xrightarrow{\delta_B}$	$\varnothing$
A self repression:	$g_a^{\varnothing}$ repression:	$g_a^{\varnothing} + A$	$\xrightarrow{\alpha}$	g_a^A
	g_a^A deregulation:	g_a^A	$\xrightarrow{\theta}$	$g_a^{\varnothing} + A$
	g_a^A transcription:	g_a^A	$\xrightarrow{\mu_{m_a}^A}$	$g_a^A + m_a$
AB complexation:	AB complexation:	$A + B$	$\xrightarrow{\gamma_{AB}}$	AB
	AB dissociation:	AB	$\xrightarrow{\lambda_{AB}}$	$A + B$
	AB degradation:	AB	$\xrightarrow{\delta_{AB}}$	$\varnothing$

Figure 2: List of all the biochemical reactions which define the HAL module. By convention rates are denoted by: μ for production rates, δ for degradation rates, α for binding rates, θ for unbinding rates, γ for complexation rates and λ for dissociation rates.

1.3 The complete set of differential equations

Translation of the list of biochemical reactions to a set of differential equations leads to the following equations:

$$
\begin{cases}
\frac{d[g_a^{\varnothing}]}{dt} &= \theta[g_A^A] &- \alpha[g_a^{\varnothing}][A] & & & & \\
\frac{d[m_A]}{dt} &= \mu_{m_A}^{\varnothing}[g_a^{\varnothing}] &+ \mu_{m_A}^{A}[g_A^A] &- \delta_{m_A}[m_A] & & & \\
\frac{d[A]}{dt} &= \mu_A[m_A] &- \gamma_{AB}[A][B] &+ \lambda_{AB}[AB] &- \delta_A[A] & & \\
\frac{d[g_b^{\varnothing}]}{dt} &= 0 & & & & & \\
\frac{d[m_B]}{dt} &= \mu_{m_B}[g_b^{\varnothing}] &- \delta_{m_B}[m_B] & & & & \\
\frac{d[B]}{dt} &= \mu_B[m_B] &- \gamma_{AB}[A][B] &+ \lambda_{AB}[AB] &- \delta_B[B] & & \\
\frac{d[g_A^A]}{dt} &= \alpha[g_a^{\varnothing}][A] &- \theta[g_A^A] & & & & \\
\frac{d[AB]}{dt} &= \gamma_{AB}[A][B] &- \lambda_{AB}[AB] &- \delta_{AB}[AB] & & &
\end{cases}
\tag{1}
$$

2 Simplifying the equations

A number of things can be done to simplify the equations without any loss of generality:

- Conservation of gene number $[g_A^A] + [g_a^{\varnothing}] = 1$ gives: $[g_A^A] = 1 - [g_a^{\varnothing}]$
- Solving $\frac{d[g_b^{\varnothing}]}{dt} = 0$ with $[g_b^{\varnothing}](t = 0) = 1$ gives: $[g_b^{\varnothing}] = 1$
- Solving $\frac{d[m_B]}{dt} = \mu_{m_B} - \delta_{m_B}[m_B]$ gives: $[m_B] = \frac{\mu_{m_B}}{\delta_{m_B}}$ for $t \to \infty$
- Elimination of the $[AB]$ equation is possible since it does not feedback into the system.
- Seen no confusion is possible any-more $[g_a^{\varnothing}]$ can be replaced by $[g]$
- Seen no confusion is possible any-more $[m_A]$ can be replaced by $[m]$
- $\mu_B \frac{\mu_{m_B}}{\delta_{m_B}}$ can be replaced by μ_B. Where μ_B is not any-more the production rate of protein B per m_B per unit of time, but is now the production rate of protein B per unit of time given the equillibrium distribution of m_B is maintained.

$$
\begin{cases}
\frac{d[g]}{dt} &= \theta(1-[g]) &- &\alpha[g][A] & & & & & \\
\frac{d[m]}{dt} &= \mu_m^{\varnothing}[g] &+ &\mu_m^{A}(1-[g]) &- &\delta_m[m] & & & \\
\frac{d[A]}{dt} &= \mu_A[m] &- &\gamma_{AB}[A][B] &+ &\lambda_{AB}[AB] &- &\delta_A[A] \\
\frac{d[B]}{dt} &= \mu_B &- &\gamma_{AB}[A][B] &+ &\lambda_{AB}[AB] &- &\delta_B[B] \\
\frac{d[AB]}{dt} &= \gamma_{AB}[A][B] &- &\lambda_{AB}[AB] &- &\delta_{AB}[AB] & & &
\end{cases}
\tag{2}
$$

3 CubeSampling and PCA

Since Sampling of the full parameter space was not biologically relevant and sampling of the full biologically relevant space did not give a clear result, we moved onto sampling of the neighbourhood of a point in the biologically relevant space.

I will deliberately not show hour previous results here (where beta correlates well to the period), since the new analysis comes before those results in the work flow. And while the PCA analysis was inspired by my knowledge of the existence of beta, the analysis is know capable of predicting beta.

The PCA analysis goes as follows:

1. sample N (± 1000) points uniformly in a 12-D hypercube of size S[1] (2) of the parameter space around ${k_{basis}}^2$
2. measure the period for each point and normalize it with ${\delta_m}^3$ (remove the non oscillatory points)
3. optional?: normalize each set of parameters and the set of all the measured periods[4]
4. merge the parameters and the period into one 13^5 by N matrix, which I will call "the data matrix".
5. calculate the covariance[6] of this data matrix
6. Calculate the smallest(???) eigenvalue of the covariance matrix and it's eigenvector, multiply both and call the result n.

Following this process we get:

$$\begin{array}{ccc} n(\theta) = 0.0100 & n(\delta_m) = -0.0207 & n(\mu_m^{\varnothing}) = 0.0102 \\ n(\mu_A) = 0.0102 & n(\mu_B) = -0.0092 & n(\gamma_{AB}) = 0.0003 \\ n(\alpha) = -0.0004 & n(\mu_m^A) = 0.0005 & n(\delta_A) = -0.0004 \\ n(\delta_B) = 0.0001 & n(\delta_{AB}) = -0.0003 & n(\lambda_{AB}) = -0.0004 \end{array}$$

Consider now that $\beta = \frac{\theta \mu_m^{\varnothing} \mu_A}{\mu_B \delta_m^2}$ and consider with $P = (\prod_k k^{n(k)})^c$ with $c = 100$. Can this be a coincidence?

---

[1] if $\overrightarrow{x}$ is a point in an M dimensional space, then an M dimensional hypercube of size S around $\overrightarrow{x}$ is defined **here** as $\{\overrightarrow{y} | \forall i \in 1...M], y(i) \in [x(i)/S, x(i) * S]\}$

[2] $k_{basis} = [\theta = 0.01, \delta_m = 0.05, \mu_m^{\varnothing} = 2, \mu_A = 25, \mu_B = 100, \gamma_{AB} = 1000, \alpha = 0.01, \mu_m^A = 0.001, \delta_A = 0.001, \delta_B = 0.001, \delta_{AB} = 1, \lambda_{AB} = 0.001]$

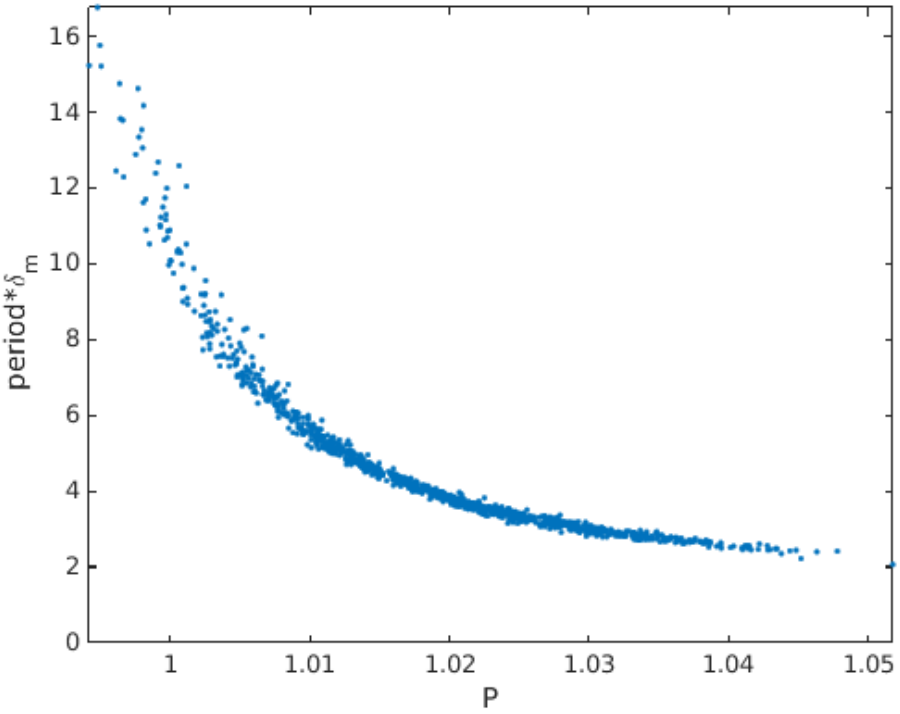
[3] This is necessary, since the analysis only works if the period are measured in a similar scale, but has to be analysed more thoroughly to generalize it to any system.

[4] make their average=0 and std=1

[5] 12 parameters + the period

[6] $cov(A)_{i,j} = \frac{1}{N-1} \sum_{k=1}^{N} (A_{i,k} - < A_{i,*} >)(A_{j,k} - < A_{j,*} >)$ where $< X_{i,*} >$ is the average of the i-th row of matrix X

Even better lets plot the normalized period towards P:



4 Questions to be resolved

From most to least important:

1. Why the smallest eigenvalue? Or is it?
2. Explain more clearly why scale with δ_m and generalize this part of the process
3. Find the name/literature of this analysis[7] (and streamline it to the standards of the field)
4. Understand the literature on non-linear PCA (and use it to perfect this technique)
5. Try with MFL (or other models)
6. Try without the log ($T = c\sum_k n(k)k$ instead of $T = (\prod_k k^{n(k)})^c$)
7. Predict c in $T = (\prod_k k^{n(k)})^c$
8. Try with an arbitrary data point (this one was known to be better then average)
9. Try without simplifying the equations ($n(\mu_B) = -1$ should be replaced by $\{n(\mu_B{}^8) = -1, n(\mu_{m_B}) = -1, n(\delta_{m_B}) = 1\}$
10. Find a standardized mathematical definition of log based hypercubes
11. Check without normalizing parameters and period

[7] I can not believe this has not been done before, this is a combination of 2 trivial extensions of PCA. 1) PCA on log scales 2) PCA on input/output data. However I do not even know the name of each extension separately.

[8] This is not exactly the same μ_B