Mathematical Model for the LacZYA promoter activity

Here we give a detailed derivation of the model for the *lacZYA* promoter activity in *Setty et. al. PNAS 2003*.

In order to find a suitable formula for the promoter activity of the *lacZYA* operon we constructed a mathematical model which incorporate the various chemical reactions between the inducers, the transcription factors(TF) and the promoter sites in the *lacZYA* system. Each TF is activated (or deactivated) by the binding of a specific inducer. Each site may be free or occupied by one or more proteins. In turn, the specific proteins which binds to the DNA site determines the promoter activity i.e. amount of mRNA produced per unit time.

The present model is not meant to be a detailed account of the full biochemistry of the Lac system, rather a toy model that captures the essential behavior. For example, multi-operator DNA loops in the system may introduce additional effects. Furthermore interaction between CRP and LacI can be added to the model, however it was found that this does not improve the fit to the experimental data, and so was neglected.

Typo

The formula for promoter activity in *Setty et. al. PNAS 2003* has $2b$ instead of b. This does not affect any of the paper's conclusions, and is derived below. We thank A. Levchenko for pointing this out.

Notations

The inducers in the model are cyclic Adenine Mono Phosphate (**cAMP**) and Isopropyl β-D-thiogalactoside (**IPTG**). Both are designated by their chemical names.

The TFs cAMP receptor protein (**CRP**) and **LacI** are designated by **C** and **R**, respectively. A star marks the inactive form of the TF. The RNA polymerase (RNAP) is designated by **P**.

A free site is designated by **S**. **S** concatenated by a letter indicates a site occupied by a protein. For example, **SC** indicates a site occupied by CRP, **SPC** indicates a site occupied by RNAP as well as CRP, etc.

Assumptions

- The concentrations of the inducers (cAMP and IPTG) are much higher than those of the TFs (CRP and LacI, respectively). Only a small portion of the total amount of the inducers actually binds to the TFs.
- The number of *lacZYA* promoter sites comprise a small portion of the total number of promoter sites of the bacteria. Since the RNAP is responsible

for transcription of every TFs, the conclusion must be that the amount of RNAP which binds to the *lacZYA* promoter sites comprise a small portion of the total RNAP. We also assume that the free RNAP level remains constant during the reactions.

- The transcription of CRP is both activated and repressed by the active state of the protein. This mechanism of regulation suggest that the total concentration of CRP remains constant. The LacI is constituently expressed and there is no known mechanism which regulates its transcription. We assume that its total concentration also remains constant.
- The reactions between the inducers, the TFs, and the DNA occur on a much shorter time scale than do other process in the cell, so it is safe to assume that detailed balance is quickly obtained.

Molecular reactions in the model

The reactions in the model can be divided into two levels:

- *TF-inducer reactions* which includes the reaction between **CRP** and its inducer **cAMP**, and **LacI** and its inducer **IPTG**.
- *Protein-DNA reactions* which includes the protein-DNA reactions between active **TFs**, **RNAP** and the *lacZYA* promoter site.

TF-inducer reactions

Two different TF's, each with its own inducer, participate in the model.

The first, **CRP** serves as an activator for the transcription of the *lacZYA* operon. It is activated when its inducer (cAMP) binds to it. When the complex CRP-cAMP binds to the DNA it rises the binding affinity of the RNAP to the DNA site and stabiles the reaction. The result is an elevated production of mRNA per unit time.

The second TF, the **LacI** repressor binds to the DNA at a site overlapping the binding site of the σ^{70} unit of the RNAP. Thus it physically suppress the binding of the RNAP to the DNA, and so transcription is not initialized.

LacI is deactivated by IPTG: once IPTG binds to LacI a conformational change is induced in the LacI structure; due to this change the LacI-IPTG complex is unable to bind to the DNA, the σ^{70} site is free to accept the RNAP, and transcription is initialized.

The corresponding chemical reactions between the TF's and the inducer are:

$$\begin{aligned} [C] + [cAMP] &\overset{K_{cAMP}}{\rightleftharpoons} [C^*], \\ [R^*] + [IPTG] &\overset{K_{IPTG}}{\rightleftharpoons} [R], \end{aligned}$$

with the following detailed balance equations:

$$\begin{aligned}
[C] &= [C^*] \cdot [cAMP] \cdot K_{cAMP}, \\
[C_T] &= [C] + [C^*], \\
[R] &= [R^*] \cdot [IPTG] \cdot K_{LacI}, \\
[R_T] &= [R] + [R^*].
\end{aligned}$$

Solving for $[C]$ and $[R]$ we find

$$\begin{aligned}
[C] &= [C_T] \frac{X^n}{1+X^n} = [C_T]\mathcal{A}, \\
[R] &= [R_T] \frac{1}{1+Y^m} = [R_T]\mathcal{R},
\end{aligned}$$

where $X = [cAMP]/K_{-cAMP}$ and $Y[IPTG]/K_{-IPTG}$ designate cAMP and IPTG concentrations in dimensionless units respectively and n and m are Hill coefficient.

Reactions at the protein-DNA level

The *lacZYA* site includes binding sites for CRP, LacI and RNAP. The binding site for σ^{70} unit of the RNAP and LacI overlaps, so that it is impossible for both to bind the site at the same time (fig1). Keeping this in mind the chemical reactions in the protein-TF level obey the following equations:

$$\begin{aligned}
\partial_t[S] &= [SC]K_{-c} + [SP]K_{-p} + [SR]K_{-r} - [S]([C]K_c + [P]K_p + [R]K_r), \\
\partial_t[SP] &= [S][P]K_p + [SPC]K_{-pc} - [SP]([C]K_{pc} + K_{-p}), \\
\partial_t[SC] &= [S][C]K_c + [SPC]K_{-cp} - [SC]([P]K_{cp} + K_{-c}), \\
\partial_t[SR] &= [S][R]K_r - [SR]K_{-r}, \\
\partial_t[SPC] &= [SC][P]K_{cp} + [SP][C]K_{pc} - [SPC](K_{-cp} + K_{-pc}).
\end{aligned}$$

Assuming Quasi-steady state and detailed balance [1]:

$$\frac{K_c\, K_{cp}}{K_{-c}\, K_{-cp}} = \frac{K_p\, K_{pc}}{K_{-p}\, K_{-pc}}$$

the kinetic equations reduce to the following algebraic equations:

$$\begin{aligned}
[SR] &= [S][R]\, \frac{K_r}{K_{-r}}, \\
[SP] &= [S][P]\, \frac{K_p}{K_{-p}}, \\
[SC] &= [S][C]\, \frac{K_c}{K_{-c}}, \\
[SPC] &= [S][P][C]\, \frac{K_p\, K_{pc}}{K_{-p}\, K_{-pc}} \\
[S] &+ [SR] + [SC] + [SP] + [SPC] = 1.
\end{aligned} \tag{1}$$

[1] Note that this relation can be derived by based on analysis of the thermodynamical box in the model (see fig1)

where the last equation describes conservation of mass. Define now an active state as a state in which transcription occurs (that is RNAP has bounded to the promotor and has started to transcribe). The first type of active state is when RNAP is bounded to the site as well as active CRP (**SPC** in our notation). We identify a second type of active state in which only RNAP binds to the site (**SP** in our notation). The promoter activity in the SPC state is higher than in the SP state, since as discussed above, the binding of active CRP to the site activates the transcription. Accordingly we define the total promoter activity

$$f = \alpha[SP] + \beta[SPC] + \gamma([S] + [SC] + [SR]).$$

γ is the basal level of transcription without the influence of the inducers. The fact that $\gamma \neq 0$ (as was found by detailed analysis of the data) can be understood as kind of leakage from the promotor. α and β are parameters which gauge the contribution of each active state to the promotor activity. As was revealed in previous studies, the CRP may have two possible roles in the activation of the RNAP: the first is to stabilize the RNAP binding, while the second is to increase the rate by which the RNAP undergoes a conformation change and initiate transcription. The first of these two possible roles is described by the ratio between a and b whereas the second by the ratio between α and β.

Solving the system of linear equations above (1) and substituting the solution into the definition f yields the final result:

$$f = \frac{\alpha\, a + \ \beta b\, d\, \mathcal{A} + \gamma\, (c\, \mathcal{R} + d\, \mathcal{A} + 1)}{1 + a + (b+1) d\, \mathcal{A} + c\, \mathcal{R}}$$

where

$a = [P]K_p$	RNA-polymerase binding to a free site.
$b = [P]K_pK_{pc}/K_c$	RNA-polymerase binding to a site with CRP-cAMP.
$c = [R_T]K_r$	LacI binding to a free site.
$d = [C_T]K_c$	CRP-cAMP binding to a free site.

The promoter activity can also be expressed as:

$$f = V_1 \frac{1 + V_2\mathcal{A} + V_3\mathcal{R}}{1 + V_4\mathcal{A} + V_5\mathcal{R}},$$

where $V_1, \ldots, V_5$ are combinations of the biochemical parameters:

$$\begin{aligned}
V_1 &= (\alpha\, a + \gamma)/(1 + a), \\
V_2 &= d\, (\beta\, b + \gamma)/(\alpha\, a + \gamma), \\
V_3 &= \gamma\, c/(\alpha\, a + \gamma), \\
V_4 &= d\, (\, b + 1)/(1 + a), \\
V_5 &= c/(1 + a).
\end{aligned}$$

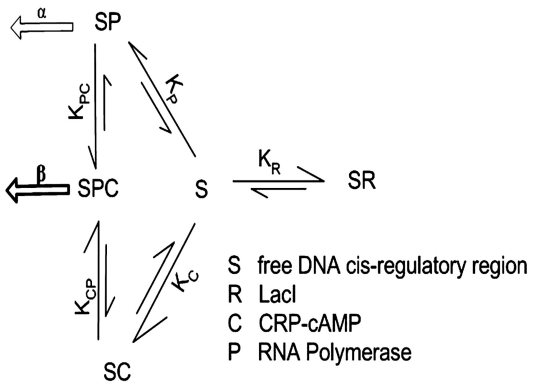

Figure 1: Simple mathematical model of the *lac* region, protein-DNA level reactions. Transcription occurs from states where RNA polymerase is bound (large open arrows). The activity of the promoter in these states are α and β. An additional parameter γ describes transcription from other states (leakiness). The K parameters are the equilibrium disassociation constants of the various reactions.