Stability and time response constraints on the sharpness of the λ phage biological switch.

V. Bérubé

Department of Physics, Massachusetts Institute of Technology

A simple model of interacting transcription factors (TFs) and RNA polymerases (RNAp) at thermodynamic equilibrium was used to study the properties of cooperative repression on gene activity. The cI_2 repression of the P_r promoter, which is part of the operon that dictates the lysogeny-lysis behavior of the λ phage, was studied and its sigmoidal dependence on the parameters of the system determined. The system's response to an increase in cI_2 concentration was found to be far from optimal sensitivity. Stability studies demonstrate that the phage seems to operate in a region which is a compromise between good stability to noise and strong switch-like behavior. A tentative explanation of the operating range of concentration of repressor in the λ phage, based on time response to an external signal, is also given.

Introduction

Experimental techniques like fluorescence resonance energy transfer (FRET) have shown increasing evidence that multiple functions in eukaryotes and prokaryotes such as gene activity regulation are performed by cooperative action of several proteins [1]. Both theoretical [2, 3] and experimental work, using one or two-hybrid systems [4], have been done in the past years to shed light on the interaction mechanisms between transcription factors (TF) on the cis-regulatory region of a promoter. Famous examples of cooperating systems studied in the literature include the cooperative co-activation of MelAB transcription by CRP and MelR in E. coli [5], as well as cooperative co-repression by cl_2 of the P_r promoter that regulates the lysogeny-lysis switch in λ phage [6]. While it is known that cooperativity between TFs enhances sensitivity (gain, control coefficient) and can create a switch-like behavior in the activity of a gene, the exact thermodynamic representation of the mechanisms underlying these reactions is still elusive. Modeling such systems to predict the correct behavior is often made difficult by the lack of information on the numerous parameters involved. So far, attempts at modeling the TFs interactions have been limited to simple models that nevertheless provide insight on the underlying physics.

An interesting thermodynamic approach was recently proposed by Bintu *et al.* [7, 8]. Their model, based on the different energies involved in interaction between the TF's and their specific and non-specific binding sites on DNA, was able to reproduce the observed behavior of the fold change[1] in the repression of lac promoter by lacI as measured by Oehler *et al*. [9]. In the case of co-repression of P_r by cI_2 in the lambda phage, the cell is clearly working away from the most sensitive region regarding its response to cI_2 concentration. Even if it reproduces the right behavior, their model does not explain why the organism would have evolved to operate in the given regime. The following work proposes plausible arguments explaining why it may be advantageous for real biological systems to work off the optimal domain.

Model

Our model is based on the work of Bintu *et al.* [7, 8] and the reader may refer to that paper for a thorough explanation of the derivation of the partition function. In this model, the gene expression is measured by the probability of RNAp to be bound to the promoter. Even if the assumption that the system is in equilibrium represents a potentially big approximation, we hope that such a model can teach us a lot about the basic physics of TF interactions. We consider a strand of DNA divided in N sites of n base pairs. For E. coli, the minimal length (l_{min}) of a TF must be ~12bps to avoid false positive in site recognition [10]. For a 4,5Mbps genome, if the number of binding units times l_{min} is small compared to the genome size, we approximately have $N = 4.5*10^6$ non-specific

[1] The fold change is defined as the ratio of the occupation probability of the promoter by RNAp (gene expression) in the presence and absence of the TFs involved in the reaction.

binding sites. We assume that the DNA acts as a reservoir so that the TF's and RNAp are bound to the DNA at all

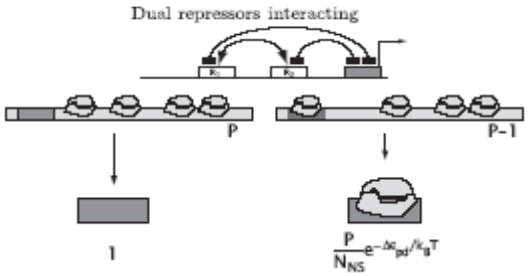

Figure 1[2] TOP: Schematic of cooperative repression by two TFs. BOTTOM: Representation of the "DNA as a reservoir for RNAp" model. The gray site represents the specific site of DNA on P_r and the statistical weight of the state is given on the right.

time. Furthermore, we consider the case of cooperative regulation where two different TFs suppress the binding of RNAp when bound to their specific binding sites on the promoter. For cooperative co-repression of a given gene by TFs A and B, the partition function leads to the following expression for the probability of RNAp in the limit where the number of non-specific site N is large compare to the number of A, B and RNAp (P).

Equation 1

$$P_{bound} = \frac{1}{1+\frac{N}{PF_{reg}}\exp(\frac{\Delta E_p}{k_B T})}$$

$$F_{reg} = \frac{1}{1+a+b+ab\exp(-\frac{E_{ab}}{k_B T})}$$

F_{reg} here is the so called regulation factor that modulates the effective number of P in the cell. For F_{reg} smaller than one we have repression while F_{reg} bigger than one represents cooperation. The effective number of TF A and B (noted a and b) are function of the difference between specific and non-specific binding energy on DNA. E_{ij} is the cooperation energy between i and j and ΔE_p is the difference of energy between specific and non-specific binding of RNAp on DNA.

Equation 2

$$x = \frac{X}{N}\exp(-\frac{E_{xs} - E_{xns}}{k_B T})$$

Defining an equivalent dissociation constant for A and B as $K_x = [X]/x$ where [X] is the concentration of a given TF and x is given by Eq.2, we can rewrite Eq.1 as

Equation 3

$$F_{reg} = \frac{1}{1+\frac{[A]}{K_A}+\frac{[B]}{K_B}+\frac{[A][B]w}{K_A K_B}}$$

$$w = \exp(\frac{-E_{ab}}{k_B T})$$

w being the interaction term representing the cooperation between the TFs.

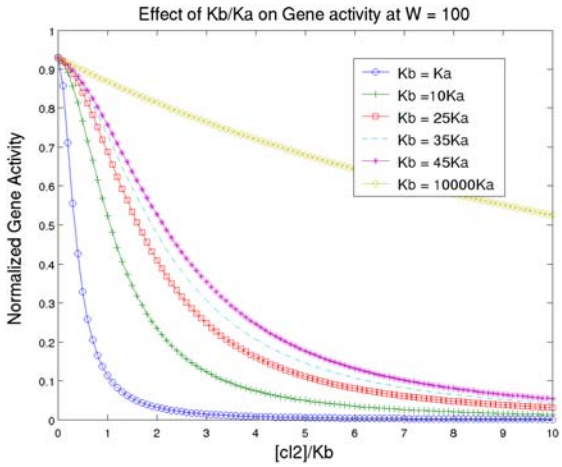

Figure 2 Genetic activity dependence on K_B/K_A. For a value of w corresponding to a weak interaction energy of $-4{,}6k_BT$(w =100), the system can still show great sensitivity if the 2 binding sites have the same specific energy. The -□- represent the biological regime.

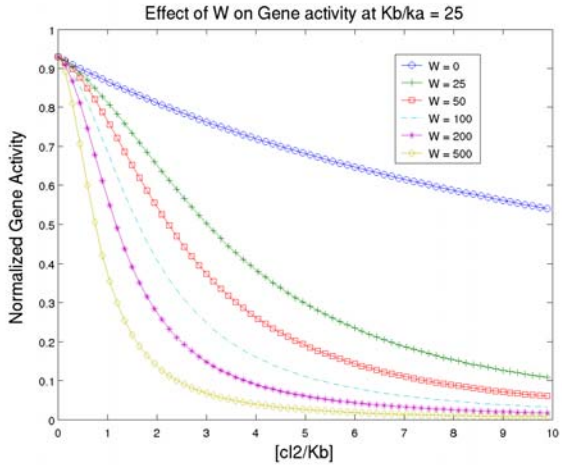

Figure 3 Genetic activity dependence on cooperative energy between the two copies of the repressor. As expected, interaction energy is essential to obtaining a sigmoidal response. As in figure 3, we see that the phage operates far from the optimal sigmoidal response region. The _._ give the biological regime.

Results

Using the above description, we modeled the behavior of cooperative co-repression by cI_2 on P_r by

[2] Figures reproduced from [7,8] with the author's authorization.

fixing [A] = [B] = [cI_2] and analyzed the effect of noise on the system as well as its time response. In vitro experiments by Koblan *et al.*[11] have shown that K_B/K_A ~

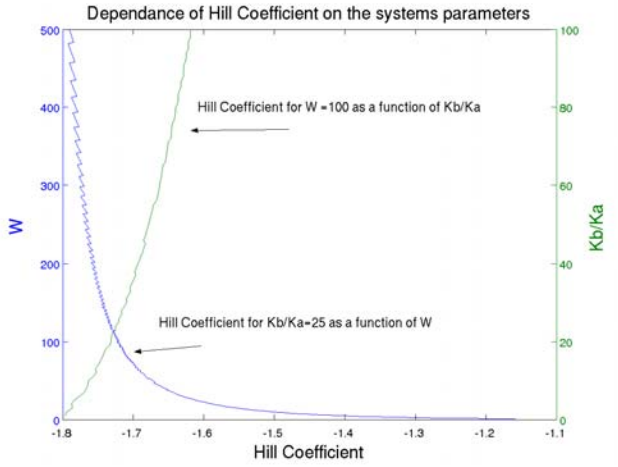

Figure 4 The Hill coefficient (negative here because it is cooperative repression) expressed here on x axis is shown as a function of w and K_B-/K_A. The intersection shows the biological regime. Even lower values could be obtained for simultaneous high w and low K_B/K_A.

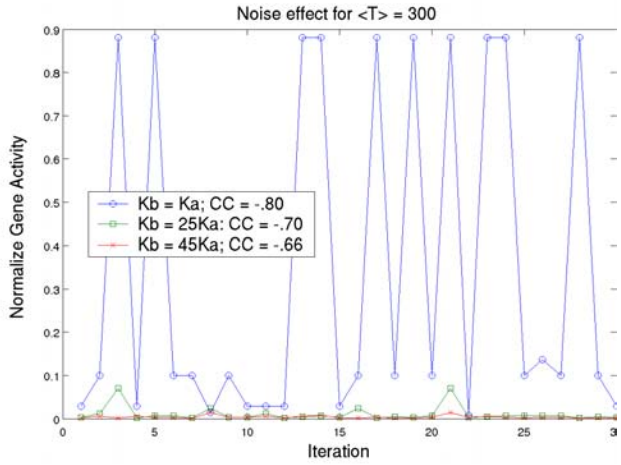

Figure 5 Systems with higher sensitivity (gain, Hill coefficient) are also more sensitive to noise. For K_B/K_A =1, the system loses its stability and may result in the phage adopting different behaviors for the same input signal. The correlation coefficients for the different cases are also indicated on the figure [12].

25 and w ~ 100. ΔEp was set to -9k_BT [13]. This value is approximate since the numerous specific sites for RNAp make it difficult to define its non-specific energy. Since the exact in vivo values of K_A and K_B are still unknown, the gene activity[3] was plotted as a function of [cI_2]/K_B for different values of the ratio K_B/K_A and w. Fig.2 and Fig.3 show that high w and low K_B/K_A significantly increase the sigmoidal response of the system. The corresponding values of the Hill coefficients are plotted in Fig.4. It shows that a much better response to cI_2 signal would be obtained if the 2 specific sites for cI_2 had the same affinity instead of having a significantly weaker binding energy (~3.25k_BT according to Eq. 2) for one of the two sites. By estimating the free energy cost of a broken hydrogen bound [14], the energy lost to a mismatch protein to nucleic acid bound on the specific site can be estimated to be of the order of a few k_BT. This shows that a single mutation on the cognate region would suffice to bring the system in the much more sensitive region where K_B/K_A~1. Since the two TFs are the same in that case, it is clear that a ratio of K_B/K_A ~1 would be possible if the O_{R1} and O_{R2} sites were identical and one may ask himself why the λ phage did not evolve to optimize its response to the repressor.

Consequently noise was introduced to study our model in a biological context. For a fixed value of w =100, the consequence of Gaussian noise (Eq. 4) due to local fluctuation in the number of cI_2 molecules was studied for different regimes driven by the values of K_B/K_A. To see if noise would result in a mixed response as to weather the phage would adopt the lysogeny or the lysis mode, we studied the consequence of noise around the transition toward fully closed switch (stable lysogeny): P_{bound} = 0.1.

Equation 4

$$P([cI_2]) = \frac{1}{\sqrt{2\pi\langle[cI_2]\rangle^2}} \exp\left(-\frac{([cI_2]-\langle[cI_2]\rangle)^2}{2\langle[cI_2]\rangle^2}\right)$$

The variance was chosen to equal the mean which was in turn fixed as the value of [cI_2][4] at P_{bound} = 0.1. Fig.5 shows the effect of the noise and clearly demonstrates that too sharp a switch is unstable because a small variation in the repressor concentration can bring the system in a totally different regime. This goes against the experimental results showing that in rich medium, the lysogeny state exhibits great stability (spontaneous lysis does not exceed $2*10^{-9}$/generation in wild type [15]). More importantly, cI_2 is present in a finite quantity in the wild type cell (60 to 120 dimmers when the environment is rich in nutrients[5]) [13, 15]. In a mutant cell where K_B/K_A ~1, we infer from Eq. 3 that a transition from 1 to 5 copies of cI_2 would drive P_{bound} from 0.1 to 0.9 and rendered the system extremely volatile and almost impossible to switch off. The ideal system thus appears to be a tradeoff between energy efficiency (low TFs synthesis), stability and strong sigmoidal behavior.

Another reason for which the phage would compromise on sigmoidal response could be to gain on activation speed. Starting from a lysis state where *cro*

[3] The gene activity in this model is assumed to be equivalent to the probability of finding the RNAp bound on the promoter.

[4] When using Equation 4, negative concentrations of $[cI_2]$ where not allowed in the simulations.

[5] It is interesting to note that an order of magnitude $E_{specific} - E_{nonSpecific}$ for cI_2 on O_{R1} and O_{R2} can be estimated from this number and equation 3. We get it to be around 13-15k_BT while experiments [13] found that they were around 19-23 k_BT.

dominates, the phage initiates lysogeny by increasing the concentration of cI that then dimerizes in cI_2 whose binding to the promoter represses transcription. In the two state model of TF specific site recognition proposed by von Hippel *et al.* [10, 13, 16, 17, 18] and investigated by Slutsky *et al.* [19], the mean time needed for a TF to find its specific site is inversely proportional to the number 'n' of such TF and independent of the binding energy if we assume this energy to be low enough to ensure folding upon passage of the TF on the cognate site (Eq. 5). A cell that would operate in a regime with a high number of TFs with lower specific affinity then a system with higher specific affinity but less TF would then experience a faster response to a signal even if the equilibrium properties of the two systems are the same for RNAp occupancy. Assuming that the phage would operate in the region where it produces just enough cI molecules to ensure stability of the switch, it would take 5-10 times longer in average for cI_2 to find its site and stop transcription of *cro* for $K_B/K_A \sim 1$ than for the $K_B/K_A \sim 25$.

Equation 5

$$t^{opt} \propto \frac{N}{2n}\sqrt{\frac{\pi\tau_{3D}}{D_{1D}}}$$

Conclusion

The lysogeny state of the λ phage is known for its great stability and it is not surprising that the phage operates in biological conditions that maximize this stability and provide good response time while keeping a sharp switch-like dependence on cI_2. Even if our model gives valuable insight on the simple system studied here, the complete mechanism driving the λ phage behavior involves many more proteins affecting the concentration of cI_2 and its capacity to bind specifically. Inasmuch, assuming that the system is at equilibrium with all of the elements bound on DNA at all times is convenient yet far from reality. We thus propose using a model based on a grand canonical ensemble where the number of TFs and RNAp would not be fixed but driven by an effective chemical potential. This potential would depend on the current state in the cell since the concentration of other proteins will affect binding probability of cI_2 on DNA, dimerization of cI into cI_2 and even its degradation.

Despite some discrepancies, extensive data now exists on the different energies [13] (specific and non-specific) involved in the genetic switch as well as time dependant measures of concentration of the key proteins regulating the switch. Such data could be used to generate a simplified dynamical model in which specific and non specific TFs as well as RNAp diffuse on a chain with possibility of hopping to remote sites. The possibility of other specific sites or kinetic traps for the different proteins involved could also be considered by the introduction of a random potential. Keeping track of gene transcription and introducing degradation could eventually lead to a real in-silico phage that could be studied dynamically and provide a general framework that could be applied easily to other systems.

References

[1] Alberts, B. (1998). *Cell* 92, 291-294

[2] Ben-Naim *J. Chem. Phys.* April 22, 1998, Volume 108, Issue 16, pp. 6937-6946

[3] Ben-Naim, *J. Chem. Phys*. 107, 10242 (1997)

[4] Hu, J., M. G. Kornacker, and A. Hochschild. 2000. *Methods* 20:80-94

[5] Wade JT, Belyaeva TA, Hyde EI, Busby *SJ. EMBO J.* 2001 Dec 17;20(24):7160-7

[6] Ian B. Dodd, Keith E. Shearwin Alison J. Perkins, Tom Burr, Ann Hochschild and J. Barry Egan *GENES & DEVELOPMENT* 18:344-354, 2004.

[7] Bintu L, Buchler N, Garcia H, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R: *arXiv:q-bio.MN/0412010*

[8] Bintu L, Buchler N, Garcia H, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R: *arXiv:q-bio.MN/0412011*

[9] Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Muller-Hill B. *EMBO J.* 1994 Jul 15;13(14):3348-55

[10] Winter, R.B. and P.H. von Hippel. *Biochemistry* 20, 6948-6960 (1981)

[11] OR. Koblan KS, Ackers GK. *Biochemistry*. 1992 Jan 14;31(1):57-65

[12] Juan M. Pedraza and Alexander van Oudenaarden, *Science*, Vol 307, Issue 5717, 1965-1969 , 25 March 2005

[13] KCl.Koblan KS, Ackers GK. *Biochemistry*. 1991 Aug 6;30(31):7817-21

[14] von Hippel, P.H. and O.G. Berg. *Proc. Natl. Acad. Sci*. USA 83, 1608-1612 (1986)

[15] Erik Aurell, Stanley Brown, Johan Johanson, Kim Sneppen. *cond-mat/0010286*

[16] Berg, O.G., Winter, R.B. and P.H. von Hippel. *Biochemistry* 20, 6926-6948 (1981)

[17] Winter, R.B., Berg, O.G. and P.H. von Hippel. *Biochemistry* 20, 6961-6977 (1981)

[18] von Hippel, P.H., W,A, Rees, K. Rippe and K.S. Wilson *Biophysical Chemistry* (W.F. Harrington Memorial Issue), 59, 231-246 (1996)

[19] Michael Slutsky, Leonid A. Mirny *Biophysical Journal* 87:4021-4035 (2004)