Kinetics of *cis*-regulatory domain binding by multiple transcription factors

Christopher Zopf
Department of Chemical Engineering, Massachusetts Institute of Technology, Boston, MA 02139

For some time researchers were puzzled by the seemingly impossibly fast rate at which transcription factors found their target sequence. Diffusion limitations were found to be overcome by hopping and sliding along the DNA molecule. The kinetics of multiple target sites in a *cis*-regulatory domain would also include these effects, but an equilibrium state evolves as the transcription factors hover in the vicinity of the domain while their neighbors migrate to the target. The time to bind N transcription factors simultaneously is modeled and calculated numerically. It is seen that cooperative binding has little effect on time to binding as this is dominated by the migration to the site, but increases the overall probability that all N transcription factors will bind their targets simultaneously.

I. Introduction

The transcription of many genes is regulated dynamically by the binding of effector proteins (transcription factors, TF) to the *cis*-regulatory region of the promoter; however, in 1970 it was found that the regulatory region could be bound by the TF 1000 times faster than diffusion through the cytosol should allow [1]. This paradox was resolved by considering several mechanisms for target site location by the TF in addition to diffusion through the cytosol. These included association-dissociation events with the DNA on the macro scale and the local scale ("hopping"), intersegment transfer and "sliding" along the DNA [2]. The development of this model in terms of the binding energy landscape has shown it capable of predicting experimentally observed binding rates which cannot be explained by 3D diffusion alone [3]. These models consider the search time for a single TF molecule to locate a single target sequence within the genome. While considerations have been made for multiple target sites [4] and multiple TF competing for the same target site [5], these authors have only considered a *cis*-regulatory domain in which the promoter is regulated by a single bound TF.

In the present discussion, the hope is to develop a model which can predict the kinetics of a *cis*-regulatory domain in which transcription is induced when N TF are bound to N sites. As the previous work was interested in the time for a single TF to find its target, dissociation from the target site could only occur after the time of interest. In a scheme with N TF, the possibility of the TF dissociating before all N are bound must be considered. The probability that a TF has encountered the *cis*-regulatory domain will be determined by the distribution of first-pass time (FPT) using the 1D/3D diffusion model. It is not necessary to account for which site is bound because the number of 1D-3D search rounds is very large. Diffusion to a particular site within the vicinity of the targets will be negligible compared to finding the regulatory region within the entire genome [3]. While the proximity of several target sites will not effect the FPT, it may create a funnel in the binding energy landscape, which will cause the TF to rebind quickly in the region despite its attempted dissociations [3,6]. This will lead to a pseudo-equilibrium state in the vicinity of the target sites. By analogy to protein-DNA binding in thermodynamic models of gene expression, the equilibrium probability of a TF bound to the DNA can be developed for independently acting TFs and TFs which weakly interact [7]. The overall probability that N TFs are bound at a point in time will include a dead-time while the TFs locate the targets and a weighting of probabilities based on the binding energy of the TF to the target sequence and to each other (Figure 1).

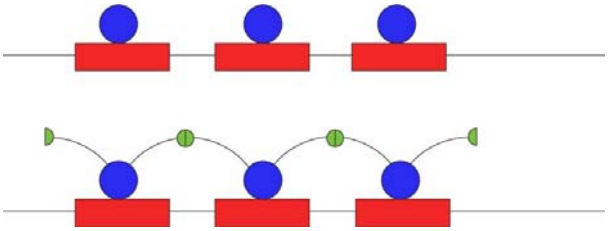

Figure 1. (*top*) TFs binding to the *cis*-regulatory region independently. (*bottom*) TFs binding to the *cis*-regulatory domain cooperatively.

II. Model

The model will describe the kinetics of multiple TF binding to their target sites after their introduction to the system at time $t = 0$. The TF will undergo many rounds of 1D/3D diffusion to locate the target sequence with some distribution of encounter times. Then a pseudo-equilibrium state occurs as the TF is "trapped" in the vicinity of the targets due to a deep energy funnel. Each TF then binds to a target site with probability based on the binding energy of the target

sequence and any cooperative interaction between neighboring TFs.

The probability a target is occupied at time t is given by the product of the probability of a TF encountering the site by time t and the equilibrium probability it is bound:

$$p_1(t) = (1 - S(t)) \cdot \phi(t) \tag{1}$$

where $S(t)$ is the survival probability of the unbound target site. Solokov *et al* give the survival probability to be

$$S(t) = e^{-2v_0\sqrt{\frac{D_{1d}t}{\pi}}} \tag{2}$$

where v_0 = TF concentration on DNA ≈ N_{TF}/(M*3.4 Å), and D_{1d} is the diffusion coefficient for sliding along the DNA. Slutsky and Mirny develop an expression for D_{1d} as a function of the roughness of the binding energy landscape, σ:

$$D_{1d} = \frac{e^{-\frac{7}{4}\beta^2\sigma^2}}{2\tau_0}\sqrt{1+\frac{1}{2}\beta^2\sigma^2} \tag{3}$$

where τ_0 is the reciprocal of the effective attempt frequency for hopping to a neighboring site. Adapting the equilibrium binding probability for a protein to DNA derived by Bintu *et al*,

$$\varphi(t) = \frac{1}{1+\frac{1}{\gamma(1-S(t))}e^{\beta(U_b - E_{ns})}} \tag{4}$$

where γ = $N_{TF}/N_{nonspecific}$ ≈ target density in *cis*-regulatory domain at equilibrium, U_b is the binding energy at the target site, and E_{ns} is the non-specific binding energy. Substituting the expression for optimized E_{ns} determined by Slutsky *et al*,

$$E_{ns} = \frac{1}{\beta}\left[\ln\left(\frac{\tau_{3d}}{\tau_0}\right) - \frac{1}{2}\beta^2\sigma^2\right] \tag{5}$$

into (4) yields

$$\varphi(t) = \frac{1}{1+\frac{\tau_0}{\tau_{3d}\gamma(1-S(t))}e^{\frac{1}{2}\beta^2\sigma^2+\beta U_b}} \tag{6}$$

where τ_{3d} is the time constant for a round of 3D diffusion. Substituting (2) and (6) into (1) gives the expression for the probability a target is occupied at time t to be

$$p_1(t) = \frac{1-e^{-2v_0\sqrt{\frac{D_{1d}t}{\pi}}}}{1+\frac{\tau_0}{\tau_{3d}\gamma\left(1-e^{-2v_0\sqrt{\frac{D_{1d}t}{\pi}}}\right)}e^{\frac{1}{2}\beta^2\sigma^2+\beta U_b}} \tag{7}$$

The probability that n sites are occupied at time t is given by the product of the probability for the individual sites weighted by the number of unoccupied sites

$$\begin{aligned} P(n,t) &= \prod_{i=1}^{n} p_1(t)(N-(i-1)) \\ &= (p_1(t))^n \frac{N!}{(N-n)!} \end{aligned} \tag{8}$$

Therefore, the overall probability of N TFs being bound simultaneously at time t is given by

$$P_N(t) = (p_1(t))^N N! \tag{9}$$

Normalizing this for $\varphi(t)$ = 1 such that $P_N(\infty)$ = 1 removes the N! coefficient

$$P_N(t) = (p_1(t))^N \tag{10}$$

The above expression for $p_1(t)$ was developed for independently acting TF. Considering cooperative binding between neighboring TF will add a stabilizing energy and increase the probability that a TF is bound at any time t [7]. This is represented here by adding a term to the exponential in the denominator of (6), effectively deepening the binding energy well at the target site. The resulting probability of occupancy for a single target site i is then

$$p_{1,i}(t) = \frac{1-e^{-2v_0\sqrt{\frac{D_{1d}t}{\pi}}}}{1+\frac{\tau_0}{\tau_{3d}\gamma\left(1-e^{-2v_0\sqrt{\frac{D_{1d}t}{\pi}}}\right)}e^{\frac{1}{2}\beta^2\sigma^2+\beta(U_b+U_c p_{1,i-1}(t))}}$$

(11)

The probability of occupancy is dependent on the binding of the previous TF. The probability that all N sites are occupied becomes

$$P_{N,c}(t) = \prod_{i=1}^{N} p_{1,i}(t)$$

$$= \prod_{i=1}^{N} \frac{1 - e^{-2v_0\sqrt{\frac{D_{1d}t}{\pi}}}}{1 + \frac{\tau_0}{\tau_{3d}\gamma\left(1 - e^{-2v_0\sqrt{\frac{D_{1d}t}{\pi}}}\right)} e^{\frac{1}{2}\beta^2\sigma^2 + \beta(U_b + U_c p_{1,i-1}(t))}}$$

(12)

III. Results and Discussion

The model was first applied to the purR TF in *Bacillus subtilis*. The specific binding energy and roughness were obtained from the weight matrix available in the online Database of Transcriptional Regulation in *Bacillus subtilis* [8] using the specific binding sites and randomly generated sites, respectively. This resulted in U_b = -8.85 k_BT and σ = 3.74 k_BT. Substituting into equation (3) (estimating $\tau_0 \sim 10^{-8}$ s as in [3]) results in D_{1d} = 0.003 cm^2/s. This is far greater than the maximum for 1D diffusion estimated in [3] and possibly the result of competing forces for fast diffusion and stable binding also observed in [3]. Therefore, the rest of the modeling was performed using the estimate $D_{1d} \sim 10^{-7}$ cm^2/s as in [3].

To investigate the qualitative behavior of the model, the following parameters were used for the purR repressor in *Escherichia coli* [3]: $\tau_0 = 10^{-8}$s, $\tau_{3d} = 10^{-3}$s, γ = 0.5 (estimated from purr binding site), σ = 6.5 β, U_b= -15 β (for stability), and v_0 = 100 TF/3.4e-2 cm. The resulting probability of occupancy is shown in Figure 2 for 3 target sites. Assuming an interaction energy U_c = -5 k_BT, the probability of occupancy with cooperative binding is also shown. The figure shows the expected behavior for the various distributions. $S(t)$ for a single TF should increase to one as t goes to infinity. When the possibility of dissociation is considered in $p_1(t)$, the probability of the TF being bound at each time point is slightly decreased. Increasing the number of target sites from 1 to 3 shows a significant increase in the dead-time while the TFs migrate to the target region. This is expected as the more objects need to randomly find a location, the longer it should take for them all to arrive. The probability as t goes to infinity is also decreased due to multiple TFs having to be bound in equilibrium simultaneously. When cooperative binding is considered, the dead-time remains the same, but the probability of binding becomes higher as the TFs stabilize each other. For the interaction energy chosen, the probability matches that for a single transcription factor as t goes to infinity. This is unexpected, but is logical for large interaction energies. If several TFs are bound and one attempts dissociation, it will be even less likely to break free of the energy well.

The probability density of the time to the event where all N TF are bound, $\psi(t)$ is found by taking the derivative of $P_N(t)$ since it is a cumulative distribution. This is done numerically for N = 10 TF, and the results are shown in Figure 3 for independent and cooperative binding. The mean time to "first pass" can be evaluated numerically by integrating $t\,\psi(t)$ from 0 to infinity:

t_1 = 2.2679 s
t_{10} =13.2585 s
$t_{10,c}$ =13.1999 s

It is again seen that cooperative binding does not affect the time to binding significantly. If the difference between the specific and nonspecific binding energies was smaller, it is more likely the cooperative binding would play a greater role in securing the binding complex faster.

IV. Conclusions

This model may be applicable to regulatory schemes in which multiple transcription factors bind either independently or cooperatively, but due to a lack of data it could not be quantitatively validated. The qualitative behavior was as expected, giving promise to the applicability of this model. It could also easily be generalized for non-identical TFs by assigning individual binding energies to each $p_1(t)$. One could imagine an *in vitro* experiment could be performed to obtain binding rates making use of FRET to determine when TFs are interacting. There is much room for exploration with these regulatory schemes.

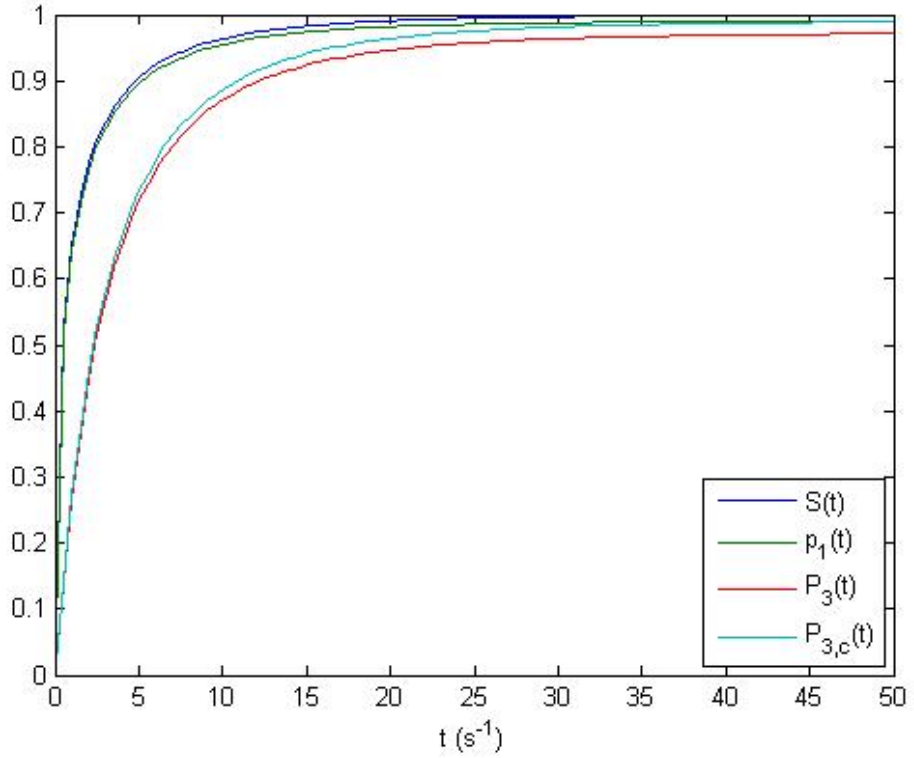

Figure 2. Probabilities of target site occupancy after time *t*.

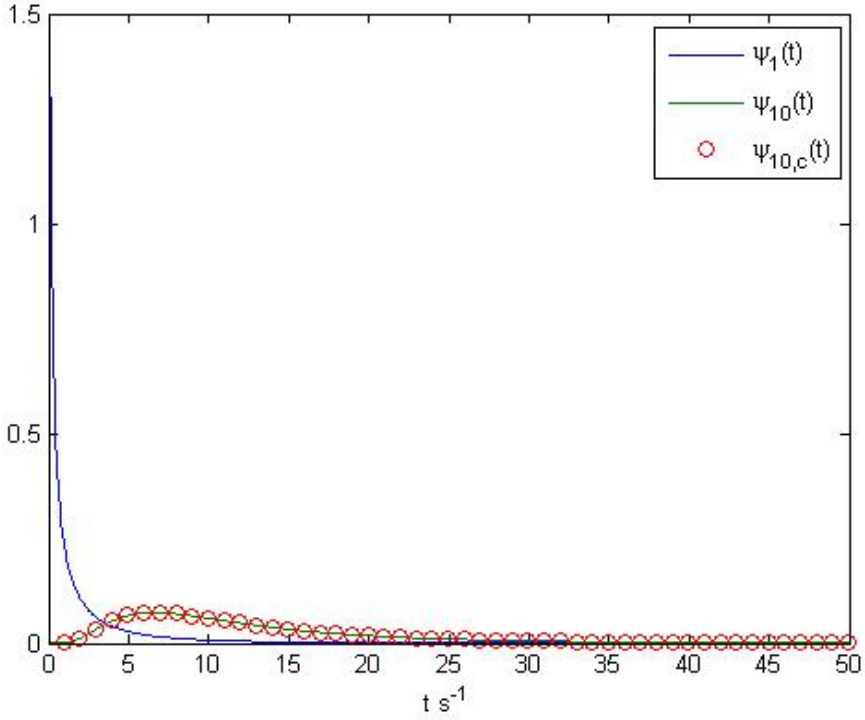

Figure 3. "First passage" time density

[1] Riggs, A. D., Bourgeois, S., and Cohn, M. **(1970)** ***J. Mol. Biol.*** **53**,401-417.
[2] Berg, O. G., R. B. Winter, and P. H. von Hippel. (1981) ***Biochemistry***, **20,** 6929–6948.
[3] Slutsky, M., Mirny, L. (2004) ***Biophysical J.*** **87,** 4021-4035.
[4] Gerland, U. *et al.* (2002) ***PNAS,*** **99,** 12015-12020.
[5] Sokolov, I., *et al.* (2005) ***Biophysical J.*** **89,** 895-902.
[6] van Zon, J., *et al.* (2006) ***Biophysical J.*** **91,** 4350-4367.
[7] Bintu, L. *et al*. (2005) ***Curr Op Gen Dev.*** **15**, 116-124.
[8] Makita, Y., *et al*. "DBTBS: database of transcriptional regulation in Bacillus subtilis and its contribution to comparative genomics Nucleic Acids Res.", 32,D75-77 (2004). Available at <http://dbtbs.hgc.jp/>.