When activators repress and repressors activate: a qualitative analysis of the Shea-Ackers model

Tomas Gedeon[1,2] Konstantin Mischaikow[3,4]
Kate Patterson[1] Eliane Traldi[3]

September 10, 2007

[1] Department of Mathematics, Montana State University, Bozeman, MT 59715, USA

[2] Center for Computational Biology, Montana State University, Bozeman, MT 59715, USA

[3] Department of Mathematics, Rutgers, The State University of New Jersey, 110 Frelinghuysen Rd,Piscataway, NJ 08854-8019, USA

[4] BioMaPS Institute, Rutgers, The State University of New Jersey, 110 Frelinghuysen Rd,Piscataway, NJ 08854-8019, USA

Abstract

The concept of activation in transcriptional regulation is based on the assumption that product mRNA increases monotonically as a function of regulator concentration. We analyze the Shea-Ackers model of transcription and find this assumption to be correct only for the simplest of promoters. We define a new regulatory constant that is a nonlinear combination of association and transcription initiation constants characterizing activation and repression for more complicated promoters. Our results can guide the synthesis of new promoters and lead to a deeper understanding of the constraints guiding the natural promoters evolution.

1 Introduction

Synthetic biology suggests the possibility of developing organisms with different functional abilities that may provide solutions to a wide variety of fundamental problems ranging from medicine to renewable energy. Producing such organisms may require a deep understanding of existing as well as novel signal transduction/gene regulatory network designs. Recent work has shown the feasibility of complete genome transplantation [1], thus, in theory, completely original networks could be employed. In practice synthetic circuits have already been constructed [2, 3]. However, for many aspects ranging from the construction of the individual components to the design of the architecture of the networks themselves, much remains to be understood.

On the network level Alon [4] provides a compelling framework for understanding the design principles of biological circuits as it relates local models for transcriptional regulation and network design to phenomenological function of the system as a whole. The local model is in accordance with the concept of regulated recruitment [5], wherein the rate of transcription of mRNA is determined by the local structure of the DNA and concentrations of regulatory proteins, often referred to as activators and repressors. As the names suggest activators enhance and repressors decrease the rate of transcription. For the most part, Hill functions are used in [4] to model the transcription rate: $f(r) = ar^n/(b + r^n)$ for activators and $f(r) = a/(b + r^n)$ for repressors with $n \geq 1$. Observe that these are monotone functions of the regulatory protein r.

The assumption of monotone regulatory interaction is widespread. The most common representation of a regulatory network is a graph with vertices corresponding to the chemical species or genes and edges corresponding to reactions. Each reaction is usually labelled with a positive or a negative sign corresponding to up- or down-regulation. Considerable effort has been spent deducing dynamics and function from such representations of a network [4, 6]. The theory of motifs is a result of such activity.

Assuming all chemical reactions on the regulatory region involving the regulatory proteins and RNA polymerase (RNAP) equilibrate on a much faster time scale than transcription, Shea and Ackers [7] construct a nonlinear model for the rate of transcription. Since the first time scale is on the order of milliseconds (bacteria) to seconds (eukaryotes) and the other on the order of minutes [4], this is a reasonable assumption in both bacterial and eukaryotic cells. The Shea-Ackers model provides a broadly accepted

quantitative framework [8] and has been experimentally validated for a variety of gene networks [9, 10, 11, 12, 7]. It should also be noted that since the Hill function is derived from the assumption of equilibrium binding of one transcription factor to the promoter, the Shea-Ackers nonlinearity is a generalization of the Hill function that naturally allows for multiple binding factors.

There are many ways in which transcription initiation is controlled and very likely more ways will be discovered in the future, but (at least in prokaryotes) there are three main steps in this process. The first is the binding of RNA polymerase to the DNA (characterized by the association constant K_B that is directly related to the binding energy of RNA polymerase to DNA). The second is isomerization of the closed RNAP-DNA complex to an open complex (characterized by a rate constant k_f), and finally successful clearance of the promoter by RNA polymerase (characterized by the constant k_{clear}). In the Shea-Ackers framework the last two processes are modelled as a *transcription initiation rate* and are lumped into one constant k. Both the binding energy of RNAP and the transcription initiation rate are controlled by the transcription factors. Thus to each control state s, which is a particular configuration of regulatory proteins and RNAP bound to the DNA, there is an association constant $K_B(s)$ and an initiation rate constant $k(s)$.

Within the context of the model of regulated recruitment the set of control states, the association constants and the initiation rate constants are the fine levers by which the cell controls transcription. These can be measured. Of course, what is of interest to systems biologists is the effect of particular regulators on transcription. Transcription factors are activators if an increase in their concentration leads to an increase in the rate of transcription and they are repressors if an increase in their concentration leads to a decrease in the rate of transcription. The Shea-Ackers transcription rate function is a sufficiently well established quantitative model of these interactions to justify a mathematical investigation of its behavior as a function of association constants K_B, and transcription initiation rates k.

Observe that the above definition of an activator or repressor is equivalent to an assumption of monotonicity with respect to the concentration of the regulatory protein. While this is true for Hill functions, we show in this paper that it need not be the case for the Shea-Ackers function. While this should not be a surprise to biologists - in low concentrations the regulatory protein CI_2 in the phage lambda switch is an activator for the cI gene, but at high concentrations it becomes a repressor - the theoretical extent to which

non-monotonicity may occur has not, to the best of our knowledge, been made clear.

The mathematical implications of non-monotone reaction functions can be significant. As an example, the global dynamics of cyclic feedback systems with arbitrarily many components with monotone reaction functions exhibits very simple dynamics; asymptotically one can have only equilibria or periodic orbits [13]. However, if the reaction functions are not monotone, then one can have chaotic dynamics [14].

In principle, the lack of monotonicity of the Shea-Ackers function could have an equally significant impact on the conclusions expressed in [4] concerning the design principles of biological circuits. In reality, it is quite possible that the biologically constrained parameters prevent this non-monotonicity. Understanding and design of transcriptional regulation require the ability to easily identify the appropriate constraints on the set of states, their association constants, and their initiation rate constants. With this in mind we introduce what we refer to as the *regulatory constant*, ρ, which is a nonlinear combination of various association and initiation rate constants, that reduces the determination of regions of monotonicity to linear equations. If r is a regulatory protein with regulatory constant ρ_r, then in the absence of any other regulatory proteins $\rho_r > 1$ implies that r is an activator and $\rho_r < 1$ implies that r is a repressor.

An outline of this paper is as follows. In Section 2 we review the Shea-Ackers model and illustrate it in the context of the *trp* and *lac* operons of E. coli. In Section 3 we introduce various concepts and notation. We begin our analysis of the Shea-Ackers model by showing that in this model the transcription rate is entirely controlled by the association and initiation rate constants (see Theorem 3.6). Section 4 contains the main results of this paper. In Section 4.1 we examine the case of a regulatory region with a single binding site and a single regulatory protein. Though the Shea-Ackers function is more general than a Hill function, monotonicity is still preserved. We also derive a relationship between the association constant and the initiation rate constant for the regulatory protein that determines whether the protein is an activator or a repressor (see Figure 5).

This relation leads to the definition of the regulatory constant ρ_r. Since it is well known [15] that multiple regulatory proteins can bind at the same site, in Section 4.2 we consider the case of a regulatory region with a single binding site but multiple regulatory proteins. Formulas which exactly determine when proteins will be activators or repressors as a function of their regula-

tory constants are presented. A complete classification for the case of two regulatory proteins is given in Theorem 4.6 and Corollary 4.7. Section 4.3 examines the unequal impact of the association constant K_B and transcription initiation rate constant k on the Shea-Ackers function. In Section 4.4 we extend the results of Section 4.2 to a generic gene with one regulatory protein that has two possible binding sites. We define a *regulatory constant for a pair* ρ_{12} and again we are able to determine if the regulatory protein r is an activator or a repressor using values of the regulatory constants ρ_1, ρ_2, and ρ_{12}. We apply our results to a phage λ model. Finally in Section 4.5 we calculate the Shea-Ackers function for a gene with two regulators and two binding sites where one of the regulator binding sites overlaps the RNAP binding site. Imposing further restrictions we recover a model of the *lac* operon.

2 The Shea-Ackers Model

The model for regulated recruitment begins with the concept of regulatory proteins binding in various configurations and at different sites in the regulatory region of a particular gene. To capture this we consider a collection of control states. The simplest state is the *empty state* which occurs when no regulatory proteins and no RNA polymerase is bound to the regulatory region. We denote this by $s_\emptyset$. The set of possible non-empty states is denoted by $\mathcal{S}$. For the purposes of this paper a state in $\mathcal{S}$ is typically determined by the configurations of the regulatory proteins for that particular gene, $\{r_1, \ldots, r_n\}$, and the presence or absence of RNAP, though in principle other control factors could be included. The simplest non-empty states consist of those for which a single regulatory protein or a single RNAP is bound to the DNA. These states are called *elementary states* and denoted by $\mathcal{E} \subset \mathcal{S}$. Within the context of the model of regulated recruitment we can use the elementary states to describe the minimal information associated with any non-empty state $s \in \mathcal{S}$. This leads to the following definition.

Definition 2.1 A *decomposition* of the state $s \in \mathcal{S}$ is the list of elementary states $\{s_i \mid i = 1, \ldots, I\} \subset \mathcal{E}$ which indicates whether regulatory proteins and/or RNAP are bound to the DNA when the state s occurs. In an abuse of notation we will often write $s = \{s_i \mid i = 1, \ldots, I\}$.

RNAP plays an essential role in that without its presence transcription cannot occur. We use $[\cdot]$ to denote concentration, and P represents RNA polymerase. Although the concentration of RNAP is a variable, [P], in order to keep our focus on the effects of regulatory proteins, it will be treated as a constant. The elementary state where only RNAP is bound to the DNA is denoted by s_P. Let $\mathcal{S}_0 \subset \mathcal{S}$ be the set of states which do not have RNAP bound to the promoter.

Under the assumption that the binding of RNAP and proteins r_i to the DNA is sufficiently more rapid than the transcription process Ackers et. al. [9] define the probability of the occurrence of the control state s to be

$$\mathbb{P}(s) = \mathbb{P}(s)([P], [r_1], \ldots, [r_m]) = \frac{K_B(s)[P]^{\alpha_s}[r_1]^{\alpha_s^1}[r_2]^{\alpha_s^2}...[r_m]^{\alpha_s^m}}{Z},$$

where

$$K_B(s) := e^{-\frac{E_s}{RT}} \tag{1}$$

and the partition function Z is given by

$$Z([P], [r_1], \ldots, [r_m]) = 1 + \sum_{s \in \mathcal{S}} K_B(s)[P]^{\alpha_s}[r_1]^{\alpha_s^1}[r_2]^{\alpha_s^2}...[r_m]^{\alpha_s^m}. \tag{2}$$

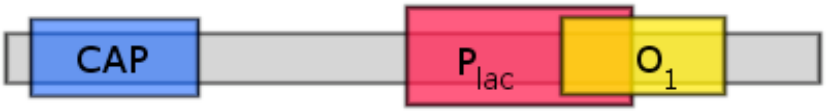
In this formula E_s denotes the energy associated to the state $s \in \mathcal{S}$ under the normalization that $E_{s_\emptyset} = 0$. The exponents α_s^i indicate the number of r_i molecules bound to the regulatory region in state s and similarly, α_s denotes the number of RNAP molecules bound to the regulatory region in state s. As is standard, T is the temperature and R is the universal gas constant [16].

Let $k(s)$ be the rate of transcription initiation of the binding state s. In particular, if $\alpha_s = 0$, i.e. $s \in \mathcal{S}_0$, then it is assumed that $k(s) = 0$. Under these assumptions the Shea-Ackers transcription rate function [7] of the gene in question is

$$f([P], [r_1], \ldots, [r_m]) = \sum_{s \in \mathcal{S}} k(s)\mathbb{P}(s). \tag{3}$$

Remark 2.2 From (3) it should be clear that to describe transcription regulation of a gene within the context of the Shea-Ackers function it is sufficient to know the set of states $\mathcal{S}$ and for each state s to know the association constant $K_B(s)$ and the transcription initiation rate $k(s)$. Because of its frequent use we define $K_P := K_B(s_P)$ and $k_P := k(s_P)$.

A



B

Figure 1: Regulatory regions of *trp* and *lac* operon in E. Coli. (A) The *trp* operator O (binding site for TrpR) overlaps the *trp* promoter P_{trp} (RNA polymerase binding site). When TrpR is bound to the DNA transcription cannot occur. (B) The *lac* operator O_1 (binding site for lacI) overlaps the *lac* promoter P_{lac}. When lacI is bound to the DNA, transcription cannot occur. The CAP binding site is upstream of the *lac* promoter and when bound by the CAP-cAMP complex it enhances the probability of RNAP binding to the promoter.

Example 2.3 The *trp* operon of E. coli is regulated by the TrpR repressor protein (see Figure 1.A). When tryptophan is present, it binds the TrpR repressor inducing conformational change in that protein and enabling it to bind the *trp* operator. This binding prevents transcription, since the operator overlaps with the RNA polymerase binding site. When tryptophan is limiting, the TrpR repressor is free of its corepressor (tryptophan) and cannot bind to the operator, allowing RNAP to bind the promoter and start transcription [17].[1]

[1]The *trp* operon is also subject to transcription attenuation and feedback inhibition [17]. Within the context of this model we do not consider these types of regulation.

For this genetic regulatory region, there are three states, $s_\emptyset$, s_P, and the elementary state s_r in which TrpR is bound to the DNA. Let $[r]$ denote the concentration of TrpR. To simplify the notation let $K_r := K_B(s_r)$. The partition function is given by $Z([r],[P]) = 1 + K_r[r] + K_P[P]$. Since s_P is the only regulatory state leading to transcription, the Shea-Ackers function for the *trp* operon is

$$f([r],[P]) = \frac{k_P K_P[P]}{1 + K_r[r] + K_P[P]}. \tag{4}$$

This function can be viewed as a generalization of the Hill function for protein binding, with the addition of RNA polymerase. When $[P]$ is assumed to be constant f becomes a Hill function for a repressor with $a = k_P K_P[P]$ and $b = 1 + K_P[P]$.

Example 2.4 The transcription of the E. coli *lac* operon is controlled by lacI and CAP-cAMP complex. The lacI binding region (operator O_1) overlaps with the RNA polymerase binding site and the CAP-cAMP complex binding site is located upstream of the promoter [15] (see Figure 1.B). In the absence of the inducer allolactose, lacI will bind to the operator and prevent binding of RNAP which prevents transcription. In the presence of allolactose, the repressor can no longer bind to the operator because of the interaction of the inducer with the repressor. Only in the absence of adequate glucose supply will the CAP-cAMP complex bind to the DNA, which is required for RNAP to effectively bind the promoter. Therefore transcription will occur only in the presence of allolactose and absence of glucose, that is, only when the CAP-cAMP complex is bound to the DNA and lacI is not bound. The set of possible regulatory states are $s_\emptyset$, s_P, s_c, s_r, $s_{cP} = \{s_c, s_P\}$, and $s_{cr} = \{s_c, s_r\}$. The elementary states s_c and s_r correspond to binding of the CAP-cAMP complex to the DNA and the repressor to the DNA, respectively. To simplify notation let us denote $k(s_*)$ and $K_B(s_*)$ by k_* and K_*, respectively. The Shea-Ackers function is given by

$$f([c],[r],[P]) = \frac{[P]}{Z}(k_P K_P + k_{cP} K_{cP}[c]), \tag{5}$$

where

$$Z([c],[r],[P]) = 1 + K_c[c] + K_r[r] + K_{cr}[c][r] + K_P[P] + K_{cP}[c][P]).$$

3 Activators and Repressors

The simplest control design would involve each regulatory protein acting as either an "activator" or a "repressor." Heuristically, increasing the presence of an activator should result in a higher expression of the gene, while increasing the presence of a repressor should lead to a lower level of gene expression. To tease out the intricacies in the Shea-Ackers function, we begin our analysis with a generic regulatory region. We use this to extract a binding dependence constant and a normalized transcription initiation constant. These constants define control within the Shea-Ackers function. This is the macroscopic characterization of the activator and the repressor. With this in mind we make the following definitions.

Definition 3.1 A regulatory protein r is a *phenomenological activator* for a gene if the transcription rate of this gene always increases with the concentration of r, that is, $\frac{\partial f}{\partial [r]} > 0$ for all $[r] \geq 0$. Conversely, r is a *phenomenological repressor* for the gene in question if the transcription rate of this gene always decreases with the concentration of r, that is, $\frac{\partial f}{\partial [r]} < 0$ for all $[r] \geq 0$.

Example 3.2 Consider the *trp* operon as in Example 2.3. The Shea-Ackers function is given by (4). Differentiation gives

$$\frac{\partial f}{\partial [r]}([r],[P]) = -\frac{k_P K_r K_P [P]}{Z^2} < 0,$$

and hence the regulator r is a phenomenological repressor.

Example 3.3 Consider the *tox* gene regulation by DtxR protein. The diphtheria toxin is composed of two subunits that are synthesized from the *tox* gene. The regulatory region for the *tox* gene consists of an operator overlapping with the RNA polymerase binding site [17] (see Figure 2). The DtxR protein binds to the operator for the *tox* gene only in the presence of ferrous ions (Fe^{2+}) and prevents transcription. The *tox* gene is turned on when there is a low level of free iron. Since the regulatory region for the *tox* gene has the same configuration of Example 3.2, the Shea-Ackers function will be given by (4) with r representing the DtxR protein. It follows from Example 3.2 that the regulator DtxR is a phenomenological repressor.

An argument following the lines of Examples 3.2 and 3.3 leads to the following result.

Figure 2: The *tox* operon regulatory region: The *tox* operator O (binding site for DtxR) overlaps with the *tox* promoter P_{tox}. When DtxR is bound to the DNA transcription cannot occur.

Proposition 3.4 *Consider a gene with regulators r_i all of whose binding sites overlap with the RNAP binding site, and whose binding excludes binding of RNAP. Then the regulators r_i are phenomenological repressors.*

As the name suggests, we use the adjective phenomenological to indicate the most directly observable relationship between the concentration of a regulatory protein and the production of the associated mRNA of the gene. However, given a set of possible states, the Shea-Ackers function has two free parameters, the association constant, $K_B(s)$, and the transcription initiation rate constant, $k(s)$. In principle these quantities can be determined by experiment. Thus, it makes sense to try to understand the phenomenological properties of regulatory proteins in terms of their association constants and transcription initiation rates. We begin with the following definition which is justified by Theorem 3.6.

Definition 3.5 The decomposition $\{s_i \mid i = 1, \ldots, I\}$ of a state s in terms of its elementary states is *independent* if

$$K_B(s) = \prod_{i=1}^{I} K_B(s_i)$$

or equivalently

$$E_s = \sum_{i=1}^{I} E_{s_i}.$$

The following theorem shows that within the Shea-Ackers function all the regulation of gene transcription occurs because of the interactions of binding energies and transcription initiation rates. Without this, the Shea-Ackers function reduces to a Hill function describing interaction of RNAP and the promoter. In particular, the transcription rate is constant with respect to concentrations of regulatory proteins.

Theorem 3.6 *Consider a gene with the set of regulatory states $\mathcal{S}$ which satisfies the following two conditions:*

1. *for each state $s \in \mathcal{S}$ its decomposition into its elementary states is independent,*

2. *the rate of transcription initiation $k(s)$ does not depend on the state, i.e. $k(s) = k$ for all the states $s \notin \mathcal{S}_0$.*

Then the transcription rate is given by

$$f([P], [r_1], \ldots, [r_m]) = k \frac{K_P[P]}{1 + K_P[P]}.$$

Proof. Consider a gene with a regulatory region that contains n distinct binding sites for m proteins, $\{r_1, \ldots, r_m\}$, and one binding site for RNAP. Let k be the state independent rate of transcription initiation. To simplify notation we will denote $K_B(s_i)$ by K_{s_i}. Since each state is decomposed independently into its elementary states

$$P(s) = k \frac{K_{s_1} K_{s_2} \ldots K_{s_n} K_P [r_1]^{\alpha_s^1} [r_2]^{\alpha_s^2} \ldots [r_m]^{\alpha_s^m} [P]}{Z},$$

and

$$Z([P], [r_1], \ldots, [r_m]) = 1 + \sum_{s \in \mathcal{S}} (K_{s_1} K_{s_2} \ldots K_{s_n} K_{s_{n+1}} [r_1]^{\alpha_s^1} [r_2]^{\alpha_s^2} \ldots [r_m]^{\alpha_s^m} [P]^{\alpha_s}),$$

where $s_i \in \{r_1, \ldots, r_m, \emptyset\}$ for $i = 1, \ldots n$ and either $s_{n+1}{=}s_P$ when RNAP is bound, or $s_{n+1}{=}s_\emptyset$ when the RNAP binding site is empty.

Therefore the transcription rate is of the form

$$f([P], [r_1], \ldots, [r_m]) = k \frac{A([P], [r_1], [r_2], \ldots, [r_m])}{A([P], [r_1], [r_2], \ldots, [r_m]) + B([r_1], [r_2], \ldots, [r_m])},$$

with $B = \sum_{s \in \mathcal{S}_0} K_{s_1} \dots K_{s_n} [r_1]^{\alpha^s_{r_1}} \dots [r_m]^{\alpha^s_{r_m}}$ and $A = K_P[P]B$.

Then

$$f([P], [r_1], \dots, [r_m]) = k \frac{K_P[P]B}{K_P[P]B + B} = k \frac{K_P[P]}{1 + K_P[P]},$$

which is a Hill function of the concentration of RNA polymerase, which is independent of the concentration $[r_i]$ of the transcription factors. ∎

Theorem 3.6 indicates that for control to occur there must be dependence of states and/or differences in the transcription initiation rate. To quantify these differences we introduce two new parameters.

Figure 3: The regulatory region of *gryB* as an example of $\phi < 1$. In the regulatory region of *gyrB* there are Fis binding sites upstream of the RNA polymerase binding site. When Fis is bound to the DNA it enhances RNA polymerase binding, but transcription fails to initiate.

Definition 3.7 Given a state $\{s_i \mid i = 1, \dots, I\}$ its *binding dependence constant* is defined by

$$\beta_s := \frac{K_B(s)}{\prod_{i=1}^{I} K_B(s_i)}$$

and if $s \in \mathcal{S} \setminus \mathcal{S}_0$ its *normalized transcription initiation constant* is

$$\phi_s := \frac{k(s)}{k_P}.$$

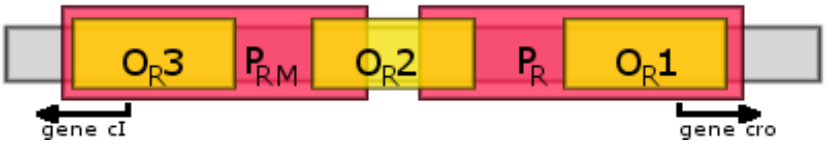


Figure 4: The right operator of phage λ as an example of $\phi > 1$. The cro gene is transcribed from the P_R promoter, while the cI gene is transcribed from P_{RM} promoter. The DNA regions O_R1, O_R2 and O_R3 are binding sites for either CI or Cro proteins.

Example 3.8 (Case of $\phi < 1$) Gyrase, an enzyme found in bacteria and plants, is composed of two subunits GyrA and GyrB, both of which are inhibited by Fis, a nucleoid protein [18, 19]. Fis inhibits GyrA by directly competing with RNAP for the *gyrA* promoter. The control of GyrB expression is more interesting. In the presence of Fis, RNAP stably binds the *gyrB* promoter, and even forms an open complex, but transcription still fails to initiate [20], see Figure 3. Because RNAP in the presence of Fis freely and stably binds to the *gyrB* promoter, but transcription fails, this is an example of $\phi < 1$.

Example 3.9 (Case of $\phi > 1$) The *right operator* O_R in phage λ has three regions designated O_R1, O_R2 and O_R3 (see Figure 4). The O_R region also contains two disjoint promoters P_R (*R*ight promoter) and P_{RM} (*R*epression *M*aintenance promoter). The promoter P_R completely overlaps O_R1 and partially overlaps O_R2; P_{RM} completely overlaps O_R3 and partially overlaps O_R2. The gene cI, that codes for the repressor protein, and a gene cro, that codes for Cro protein, flank the O_R region. The binding of RNA polymerase to P_R initiates transcription of cro gene, while RNAP binding to P_{RM} initiates transcription of the cI gene. When a CI_2 protein binds O_R2 it assists P_{RM} bound RNAP to isomerize from a closed complex to an open complex, increasing the transcription rate [15]. This is an example of $\phi > 1$.

4 Binding and Initiation Regulation

The key question we want to address is the correspondence between activation and repression on the biochemical level and on the macroscopic, or phenomenological level. We define a regulatory constant ρ which has a nonlinear dependence on both the binding dependence constant and the normalized transcription initiation constant. In Proposition 4.3 we characterize a phenomenological regulator using ρ and show that in the simplest of settings a phenomenological activator is equivalent to $\rho > 1$ and a phenomenological repressor is equivalent to $\rho < 1$.

In Section 4.2 we find that the constant ρ determines whether a regulator is an activator or a repressor in a situation where multiple regulators compete for the same binding site. In section 4.3 we explore in more depth the unequal effect of K_B and k, two key parameters of the Shea-Ackers function, on the rate of transcription. In section 4.4 and section 4.5 we discuss activation and repression for the operators with two binding sites and one, or two regulators, respectively. We illustrate our results on examples of phage λ and *lac* operons.

4.1 The Simple Regulatory Region

The simplest nontrivial regulatory region has one binding site for the RNAP and another for a single regulatory protein. We capture this in the following definition.

Definition 4.1 A *simple regulatory region* is defined by the set of states $\mathcal{S} = \{s_\emptyset, s_r, s_P, s_{rP}\}$ where $s_{rP} = \{s_r, s_P\}$.

The existence of the state s_{rP} implies that both RNAP and the protein r can be bound to the DNA simultaneously. To simplify the notation, let

$$\beta_r := \frac{K_B(s_{rP})}{K_B(r)K_P} \quad \text{and} \quad \phi_r := \frac{k(s_{rP})}{k_P}.$$

Observe that the Shea-Ackers function is

$$\begin{aligned} f([r],[P]) &= \frac{k_P K_P[P] + k(s_{rP})K_B(s_{rP})[r][P]}{1 + K_B(s_r)[r] + K_P[P] + K_B(s_{rP})[r][P]} \\ &= \frac{k_P K_P[P]}{Z}\left(1 + \phi_r \beta_r K_r[r]\right). \end{aligned} \tag{6}$$

Whether r is an activator or repressor is determined by the sign of the derivative of f. Differentiating (6) gives

$$\frac{\partial f}{\partial [r]}([r],[P]) = \frac{k_P K_r K_P [P]}{Z^2} \left(\phi_r \beta_r (1 + K_P[P]) - 1 - \beta_r K_P [P]\right). \qquad (7)$$

Failure of the regulatory protein to be an activator or a repressor at a particular concentration is equivalent to $\frac{\partial f}{\partial [r]}([r],[P]) = 0$, that is,

$$\frac{\phi_r \beta_r (1 + K_P[P])}{1 + \beta_r K_P [P]} = 1.$$

This leads to the following definition.

Definition 4.2 Consider a regulatory region with a regulatory protein r for which the state $s_{rP} = \{s_r, s_P\}$ exists. The *regulatory constant* of r is

$$\rho_r := \frac{\phi_r \beta_r (1 + K_P[P])}{1 + \beta_r K_P [P]}. \qquad (8)$$

As an example, consider an *E. coli* culture with growth rate $\mu \approx 0.02$ min^{-1}, which corresponds to a doubling time of 30 minutes, then there are approximately 1500 active RNA polymerase molecules per cell [21]. This corresponds to $[P] \approx 3.0\ \mu M$ and hence

$$\rho_r := \frac{\phi_r \beta_r (1 + 3.0 \cdot K_P)}{1 + 3.0 \cdot \beta_r K_P}.$$

Rewriting (7) in terms of the regulatory constant we obtain

$$\frac{\partial f}{\partial [r]}([r],[P]) = \frac{k_P K_r K_P [P]}{Z^2} \left(1 + \beta_r K_P [P]\right) (\rho_r - 1). \qquad (9)$$

Thus, as expected the sign of the derivative is determined by ρ_r. This gives the following result.

Proposition 4.3 *Consider a simple regulatory region with regulatory protein r. Then*

$$r \text{ is a phenomenological activator} \iff \rho_r > 1$$

and

$$r \text{ is a phenomenological repressor} \iff \rho_r < 1.$$

As immediate consequences of Proposition 4.3 we have the cases when either the binding constant β_r or the normalized transcription initiation constant ρ_r is one.

Corollary 4.4 *Consider a simple regulatory region and suppose that the rate of transcription initiation is independent of state, i.e.* $k(s_P) = k(s) = k$. *Then,*

$$r \text{ is a phenomenological activator} \iff \beta_r > 1$$

and

$$r \text{ is a phenomenological repressor} \iff \beta_r < 1.$$

Corollary 4.5 *Consider a simple regulatory region and assume that the decomposition of* s *into its elementary states is independent. Then,*

$$r \text{ is a phenomenological activator} \iff \phi_r > 1$$

and

$$r \text{ is a phenomenological repressor} \iff \phi_r < 1.$$

Observe that if $\rho_r = 1$, then $\frac{\partial f}{\partial [r]} \equiv 0$, and hence r has no regulatory impact. Figure 5 indicates the $\rho_r = 1$ isocline in the binding dependence constant and normalized transcription initiation constant plane.

One of the consequences of this result is that the regulatory protein r is either a phenomenological activator or phenomenological repressor, but it cannot take on both functions. As is indicated in later subsections, the introduction of multiple regulatory proteins or multiple binding sites changes this.

4.2 Multiple Regulators, One Binding Site

The examples of multiple proteins competitively binding to the same site are ubiquitous and range from λ phage Cro and CI proteins [15] to eukaryotes [22].

With this in mind we turn our attention to the setting of a gene with n regulatory proteins r_i, $i = 1, \ldots, n$, one regulator binding site, and one RNAP binding site. The associated collection of states is $\mathcal{S} = \{s_\emptyset, s_i, s_P, s_{iP} \mid i = 1, \ldots, n\}$, where $s_{iP} = \{s_i, s_P\}$. The initiation rate function is given by

$$f([r_1], \ldots, [r_n], [P]) = \frac{k_P K_P [P]}{Z} \left(1 + \sum_{i=1}^{n} \phi_i \beta_i K_i [r_i] \right) \tag{10}$$

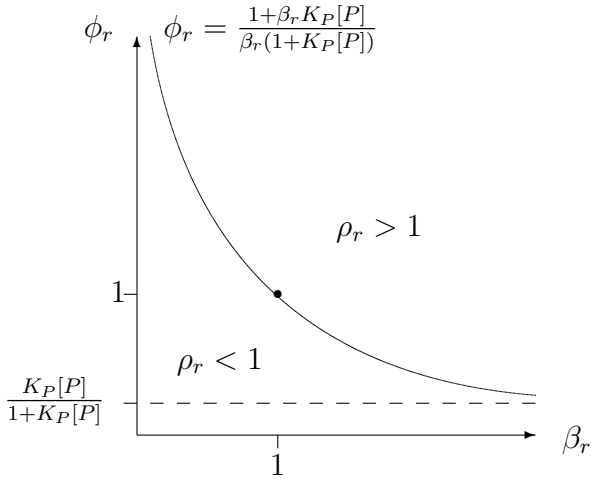


Figure 5: The relative importance of the dependence constant as opposed to the normalized transcription initiation constant in determining whether a single regulator is a phenomenological activator or repressor.

where $Z = 1 + \sum_{i=1}^{n} K_i[r_i] + K_P[P] + \sum_{i=1}^{n} \beta_i K_i K_P[r_i][P]$.

Straightforward differentiation and substitution of (8) gives

$$\frac{\partial f}{\partial [r_i]} = \frac{k_P K_i K_P[P]}{Z^2} \cdot \frac{1+\beta_i K_P[P]}{1+K_P[P]} \left((1+K_P[P])(\rho_i - 1) + \sum_{j=1}^{n} K_j (1+\beta_j K_P[P])[r_j](\rho_i - \rho_j) \right) \tag{11}$$

From an experimental point of view, perhaps the easiest test for the regulatory nature of the protein r_i is to measure whether production of mRNA increases or decreases with respect to $[r_i]$ in the absence of the other regulatory proteins. Observe that

$$\frac{\partial f}{\partial [r_i]} \mid_{[r_j]=0, j\neq i} = \frac{k_P K_i K_P[P]}{Z^2} \cdot (1+\beta_i K_P[P])(\rho_i - 1). \tag{12}$$

Therefore the sign of $\frac{\partial f}{\partial [r_i]}$ for low concentrations of other proteins $[r_j], j \neq i$, is determined by the regulatory constant ρ_i.

To determine whether or not a regulator can change between an activator and repressor requires identifying the regions where $\frac{\partial f}{\partial [r_i]} = 0$. Solving (12)

for zero gives rise to the following hyperplane

$$\sum_{j=1}^{n} K_j(1+\beta_j K_P[P])(\rho_i - \rho_j)[r_j] = (1+K_P[P])(1-\rho_i). \qquad (13)$$

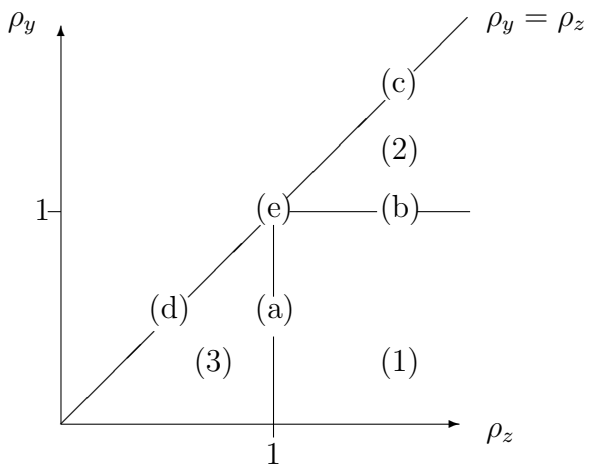


Figure 6: The regulatory constants ρ_y and ρ_z determine the roles of the regulatory proteins y and z as activators or repressors. In region (1) z is a phenomenological activator and y is a phenomenological repressor. For all values of ρ_z and ρ_y in region (2), there is a particular $[z]^*$ which designates whether y is an activator or repressor. Similarly, region (3) has a $[y]^*$ where z is an activator or repressor.

As a first application of this formalism we completely characterize the behavior of the regulatory proteins when $n = 2$. To make the notation more transparent let $y = r_1$ and $z = r_2$. Without loss of generality we restrict our attention to the case where $\rho_z \geq \rho_y$. (See Figure 6.)

Theorem 4.6 *Consider a gene with two regulatory proteins y and z, one regulator binding site, and one RNAP binding site.*

1. *If $\rho_z > 1 > \rho_y$, then z is a phenomenological activator and y is a phenomenological repressor.*

2. *If $\rho_z > \rho_y > 1$, then z is a phenomenological activator and there exists $[z]^* > 0$ such that if $[z] > [z]^*$ then y is an activator and if $[z]^* > [z]$ then y is a repressor.*

3. *If $1 > \rho_z > \rho_y$, then y is a phenomenological repressor and there exists $[y]^* > 0$ such that if $[y] > [y]^*$ then z is an activator and if $[y]^* > [y]$ then z is a repressor.*

We also have the following special cases.

Corollary 4.7 *Consider a gene with two regulatory proteins y and z, one regulator binding site, and one RNAP binding site.*

a. *If $1 = \rho_z > \rho_y$, then z is an activator for all $[y] > 0$ and y is a phenomenological repressor.*

b. *If $\rho_z > \rho_y = 1$, then z is a phenomenological activator and y is a repressor for all $[z] > 0$.*

c. *If $\rho_z = \rho_y > 1$, then z and y are phenomenological activators.*

d. *If $1 > \rho_z = \rho_y$, then z and y are phenomenological repressors.*

e. *If $\rho_z = \rho_y = 1$, then $\frac{\partial f}{\partial [z]} = \frac{\partial f}{\partial [y]} \equiv 0$. Thus neither z nor y are phenomenological activators or repressors.*

Proof of Theorem 4.6. In this simpler setting, (13) reduces to two equations which can be solved explicitly:

$$[y]^* = \frac{(1 + K_P[P])(\rho_z - 1)}{K_y(1 + \beta_y K_P[P])(\rho_y - \rho_z)} \tag{14}$$

and

$$[z]^* = \frac{(1 + K_P[P])(\rho_y - 1)}{K_z(1 + \beta_z K_P[P])(\rho_z - \rho_y)}. \tag{15}$$

Observe that there exists at most one positive value for $[y]^*$ and $[z]^*$.

1. Both $[y]^* < 0$ and $[z]^* < 0$, thus f is monotone in $[z]$ and $[y]$. The result follows from (12).

2. In this case, $[y]^* < 0$ and $[z]^* > 0$. Again, the signs of the derivatives are determined by (12).

Similar arguments prove *3* as well as Corollary 4.7 ∎

A presentation of the complete characterization of the behavior of more than two regulatory proteins is possible, but tedious. Instead we present a typical result in the case of three regulatory proteins.

Proposition 4.8 *Consider a gene with three regulatory proteins r_i, $i = 1, 2, 3$, one regulator binding site, and one RNAP binding site. Assume $\rho_3 > \rho_2 > 1 > \rho_1$. Then*

1. *r_3 is a phenomenological activator.*
2. *r_1 is a phenomenological repressor.*
3. *If*

$$[r_3] > \frac{K_1(1+\beta_1 K_P[P])(\rho_1-\rho_2)}{K_3(1+\beta_3 K_P[P])(\rho_2-\rho_3)}[r_1] + \frac{(1+K_P[P])(1-\rho_2)}{K_3(1+\beta_3 K_P[P])(\rho_2-\rho_3)}$$

then r_2 is a phenomenological repressor, and if

$$[r_3] < \frac{K_1(1+\beta_1 K_P[P])(\rho_1-\rho_2)}{K_3(1+\beta_3 K_P[P])(\rho_2-\rho_3)}[r_1] + \frac{(1+K_P[P])(1-\rho_2)}{K_3(1+\beta_3 K_P[P])(\rho_2-\rho_3)}$$

then r_2 is a phenomenological activator.

We note that the curve in the $[r_1], [r_3]$ plane that separates regions where r_2 is an activator and where r_2 is a repressor is a line where both the slope and the intercepts are functions of ρ_1, ρ_2 and ρ_3. This underscores the effectiveness of the regulatory constants in the characterization of activation and repression.

4.3 K_B versus k-cooperativity

Consider a regulatory region with states $\mathcal{S}$. Let r be a regulatory protein. Denote the elementary state in which r is bound to the DNA by s_r. Define $\mathcal{S}^r \subset (\mathcal{S} \setminus \mathcal{S}_0)$ to be the set of states s which contain the elementary state s_r in their decomposition. The protein r exhibits *K_B-cooperativity* (*k-cooperativity*) if $\beta(s) > 1$ ($\phi(s) > 1$) for all $s \in \mathcal{S}^r$. We wish to compare the relative effect of K_B-cooperativity against k-cooperativity.

The following theorem indicates that if a particular regulatory protein can produce either an "equal" amount of K_B-cooperativity or k-cooperativity, then the latter results in a greater rate of production of mRNA. Apart from

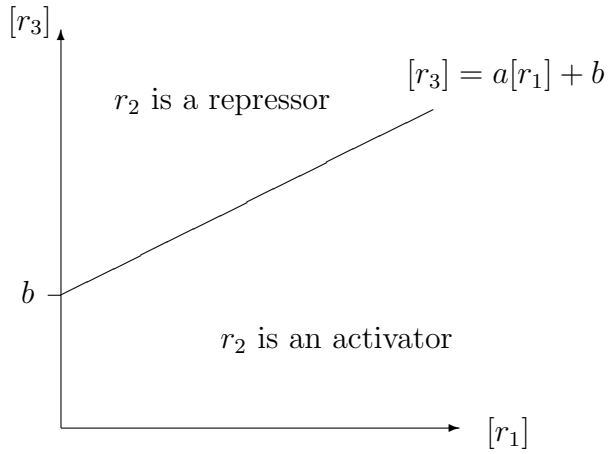


Figure 7: Under the assumption that $\rho_3 > \rho_2 > 1 > \rho_1$ we show that r_3 is a phenomenological activator and r_1 is a phenomenological repressor. A line in $[r_1], [r_3]$ plane separates regions where r_2 is an activator and where r_2 is a repressor, see Proposition 4.8. The slope $a = \frac{K_1(1+\beta_1 K_P[P])(\rho_1-\rho_2)}{K_3(1+\beta_3 K_P[P])(\rho_2-\rho_3)}$ and the intercept $b = \frac{(1+K_P[P])(1-\rho_2)}{K_3(1+\beta_3 K_P[P])(\rho_2-\rho_3)}$ are functions of ρ_1, ρ_2 and ρ_3.

the maximal production of mRNA there are very likely other evolutionary constraints imposed on the cell. A corollary of our result is that if a particular protein interacts with RNA polymerase by only K_B cooperativity, then there must be additional constraints worth the trade-off of decreased mRNA production.

Theorem 4.9 *Consider a regulatory region with states $\mathcal{S}$ and regulatory proteins $\{r, r_1, \ldots, r_n\}$. Let $s_{rP} = \{s_r, s_P\}$ where s_r is the elementary state where r is bound to the DNA. Assume that for all $s \in \mathcal{S}^r$, as defined above,*

$$K_B(s) = K_B(s_{rP}) \quad \textit{and} \quad k(s) = k(s_{rP}).$$

Let $f^{a,b}$ denote the initiation rate function under the assumption that $\beta_{s_r} = a$ and $\phi_{s_r} = b$. If $c > 1$, then

$$f^{1,c}([r], [r_1], \ldots [r_n], [P]) > f^{c,1}([r], [r_1], \ldots [r_n], [P]).$$

Proof. The initiation rate function $f^{a,b}$ has the form

$$\frac{ab\sum_{s\in\mathcal{S}^r} k_P K_P[r]^{\alpha_0}\cdots[P]^{\alpha} + \sum_{s\in(\mathcal{S}\setminus(\mathcal{S}^r\cup\mathcal{S}_0))} k(s)K_B(s)[r]^{\alpha_0}\cdots[P]^{\alpha}}{1 + a\sum_{s\in\mathcal{S}^r} K_B(s_{rP})[r]^{\alpha_0}\cdots[P]^{\alpha} + \sum_{s\in\mathcal{S}\setminus\mathcal{S}^r} K_B(s)[r]^{\alpha_0}\cdots[P]^{\alpha}}$$

Thus, $f^{1,c}$ is given by

$$\frac{c\sum_{s\in\mathcal{S}^r} k_P K_P[r]^{\alpha_0}\cdots[P]^{\alpha} + \sum_{s\in(s\in(\mathcal{S}\setminus(\mathcal{S}^r\cup\mathcal{S}_0)))} k(s)K_B(s)[r]^{\alpha_0}\cdots[P]^{\alpha}}{1 + \sum_{s\in\mathcal{S}^r} K_P[r]^{\alpha_0}\cdots[P]^{\alpha} + \sum_{s\in\mathcal{S}\setminus\mathcal{S}^r)} K_B(s)[r]^{\alpha_0}\cdots[P]^{\alpha}}$$

while $f^{c,1}$ is given by

$$\frac{c\sum_{s\in\mathcal{S}^r} k_P K_P[r]^{\alpha_0}\cdots[P]^{\alpha} + \sum_{s\in(s\in(\mathcal{S}\setminus(\mathcal{S}^r\cup\mathcal{S}_0)))} k(s)K_B(s)[r]^{\alpha_0}\cdots[P]^{\alpha}}{1 + c\sum_{s\in\mathcal{S}^r} K_P[r]^{\alpha_0}\cdots[P]^{\alpha} + \sum_{s\in\mathcal{S}\setminus\mathcal{S}^r} K_B(s)[r]^{\alpha_0}\cdots[P]^{\alpha}}$$

The numerators of the two terms are identical but the denominator of $f^{c,1}$ is larger since $c > 1$. ∎

Example 4.10 The difference between K_B- and k-cooperativity may have consequences for the function of the organism. On the P_{RM} promoter of the phage λ the CI_2 repressor interacts with RNAP using k-cooperativity [15]. By using a detailed model of the induction process, which was based on experimental data, Gedeon *et. al.* [10] predict that replacing k-cooperativity with the same amount of K_B-cooperativity yields a defective phage. This mutant phage induces at a much lower level of radiation and is inherently unstable to noise.

Finally, our result can be viewed in the context of design of tightly controllable promoters in synthetic biology. Lanzer and Bujard [23] studied which factors most affect repressibility of promoters. They found that both the association constant of the RNA polymerase K_P and the promoter clearance rate k_P play key roles. In a later paper Lutz and Bujard [24] put the emphasis on the rate K_P, since stronger binding of RNA polymerase puts the repressor at a competitive disadvantage and hence a gene with high K_P is difficult to repress. They construct tightly repressible and highly inducible synthetic operators from promoters with low to moderate K_P. Our results suggest that on highly inducible operators the cooperativity between the regulator and RNAP will be characterized by a moderate binding dependence constant β and a very high normalized transcription initiation constant ϕ.

4.4 One Regulator, Two Binding Sites

We now consider a gene with one regulator r, two regulator binding sites, and one RNAP binding site. The collection of non-empty states is

$$\mathcal{S} = \{s_1, s_2, s_{12}, s_P, s_{1P}, s_{2P}, s_{12P}\}$$

where s_i, $i = 1, 2$ denotes the elementary state of r bound in the i-th binding site and

$$s_{12} = \{s_1, s_2\}, s_{1P} = \{s_1, s_P\}, s_{2P} = \{s_2, s_P\}, s_{12P} = \{s_1, s_2, s_P\}.$$

To simplify the notation, let $K_* := K_B(s_*)$ and $\phi_* := \phi_{s_*}$.

Then the Shea-Ackers function takes the form

$$f([r], [P]) = \frac{k_P K_P [P]}{Z}(1 + (\phi_{1P}\beta_{1P}K_1 + \phi_{2P}\beta_{2P}K_2)[r] + \phi_{12P}\beta_{12P}K_1K_2[r]^2), \tag{16}$$

where

$$\begin{aligned} Z([r], [P]) &= 1 + K_P[P] + (K_1 + K_2)[r] + (\beta_{1P}K_1 + \beta_{2P}K_2)K_P[r][P] + \\ &\quad \beta_{12}K_1K_2[r]^2 + \beta_{12P}K_1K_2K_P[r]^2[P]. \end{aligned}$$

Recall the regulatory constants for each protein-binding site pair

$$\rho_1 := \frac{\phi_{1P}\beta_{1P}(1 + K_P[P])}{1 + \beta_{1P}K_P[P]} \quad \text{and} \quad \rho_2 := \frac{\phi_{2P}\beta_{2P}(1 + K_P[P])}{1 + \beta_{2P}K_P[P]}.$$

With two binding sites there is an additional state where both sites are occupied by the protein. We define a *regulatory constant for a pair* by

$$\rho_{12} := \frac{\phi_{12P}\beta_{12P}(1 + K_P[P])}{\beta_{12} + \beta_{12P}K_P[P]}. \tag{17}$$

Then after tedious computation the derivative of (16) can be written as

$$\begin{aligned} \frac{\partial f}{\partial [r]}([r], [P]) &= \frac{k_P K_P [P]}{Z^2}\Big(\alpha_1(\rho_1 - 1) + \alpha_2(\rho_2 - 1) + 2\alpha_{12}(\rho_{12} - 1)[r] \\ &+ [r]^2\frac{\alpha_{12}}{\alpha_P}\left(\alpha_1(\rho_{12} - \rho_1) + \alpha_2(\rho_{12} - \rho_2)\right)\Big), \end{aligned} \tag{18}$$

where

$$\begin{aligned}
\alpha_P &:= 1 + K_P[P] \\
\alpha_1 &:= K_1(1 + \beta_{1P} K_P[P]) \\
\alpha_2 &:= K_2(1 + \beta_{2P} K_P[P]) \\
\alpha_{12} &:= K_1 K_2(\beta_{12} + \beta_{12P} K_P[P])
\end{aligned}$$

are all positive constants.

Observe that $\frac{k_P K_P[P]}{Z^2}$ is nonzero for all $[r]$ and the remainder of the derivative is a quadratic function

$$\frac{df}{d[r]} = \frac{k_P K_P[P]}{Z^2}(A[r]^2 + B[r] + C) \tag{19}$$

where

$$\begin{aligned}
A &:= \frac{\alpha_{12}}{\alpha_P}(\alpha_1(\rho_{12} - \rho_1) + \alpha_2(\rho_{12} - \rho_2)) \\
B &:= 2\alpha_{12}(\rho_{12} - 1) \\
C &:= \alpha_1(\rho_1 - 1) + \alpha_2(\rho_2 - 1).
\end{aligned} \tag{20}$$

When the coefficients A, B, C have the same sign we have the following corollary.

Corollary 4.11 *Consider a gene with one regulator r, two regulator binding sites, and one RNAP binding site. If*

$$\rho_{12} \geq \rho_1 \geq 1 \text{ and } \rho_{12} \geq \rho_2 \geq 1,$$

and at least one inequality is strict, then r is a phenomenological activator. If

$$\rho_{12} \leq \rho_1 \leq 1 \text{ and } \rho_{12} \leq \rho_2 \leq 1$$

and at least one inequality is strict, then r is a phenomenological repressor.

However, if the signs of the coefficients are not the same then r may be an activator for small $[r]$ and a repressor for large $[r]$. This indeed is the case in phage λ $\mathrm{P_{RM}}$ promoter where CI acts as an activator at low concentrations and as a repressor at high concentrations [15].

Example 4.12 The center of the regulatory processes in the phage λ is the *right operator* O_R, see Figure 4.

The lysogenic pathway corresponds to the state of the O_R where CI dimers are bound to both O_R2 and O_R1, blocking the P_R promoter and thus transcription of the cro gene, while RNAP is free to bind P_{RM}, maintaining the transcription of the cI gene. In the lysogen O_R1 is almost always bound by a CI_2 protein and thus the production of Cro is very low. We simplify the situation by assuming that in fact O_R1 is always occupied by CI_2 and there is no production of Cro in lysogeny. Therefore we will only consider O_R2 and O_R3 binding sites and only the regulatory protein CI.

These assumptions imply that we are in the setting of a single regulatory protein with two binding sites. Let s_1 and s_2 correspond to the elementary states where CI is bound to O_R2 and O_R3, respectively. However, in the phage λ, O_R3 overlaps with P_{RM}. Mathematically this is incorporated by setting $\beta_{2P} = \beta_{12P} = 0$ which implies that $\rho_{12} = \rho_2 = 0$. This immediately implies that $A < 0$ and $B < 0$. The value of C can in principle be of both signs; if $C < 0$ then the CI would be a phenomenological repressor and if $C > 0$ then there is a unique positive value $[r]^*$ at which r switches from being an activator to being a repressor. At the value $[r]^*$ the transcription initiation rate is at its maximum and in this respect the promoter is at its peak performance. Based on the experimental values collected in Santillan and Mackey [11] $\beta_{1P} = 1$, $\phi_{1P} = 12.26$ and thus $\rho_1 = 12.26$. Further $K_1 > K_2$ and $(1 + K_P[P]) > 1$ and therefore $C > 0$ in phage λ.

The next question we address is whether it is possible to choose values of A, B, C in such a way that equation (19) has two positive roots. If it is possible and $A < 0$, then the regulatory protein r is an activator at low $[r]$, a repressor at intermediate $[r]$, and then an activator for large $[r]$. (For $A > 0$ the switch would be from repressor to activator and back to the repressor.)

While such A, B and C certainly exist for a general quadratic equation, as the next Proposition shows, for A, B and C as specified in (20) this is not possible.

Proposition 4.13 *Consider a gene with one regulator r, two regulator binding sites, and one RNAP binding site. Then there is at most one positive value $[r]^*$ at which the derivative of the transcription rate function $f([r])$ can change the sign.*

Proof. For (19) to have two positive zeros either $A < 0, C < 0$ and $B > 0$ or all signs are reversed. We will show that this cannot happen. Assume $A < 0, C < 0$ and $B > 0$, the other case being analogous. The condition $B > 0$ is equivalent to

$$\rho_{12} > 1. \tag{21}$$

The conditions $A < 0$ and $C < 0$ are equivalent to solving

$$\begin{aligned} \alpha_1(\rho_{12} - \rho_1) + \alpha_2(\rho_{12} - \rho_2) &< 0 \\ \alpha_1(\rho_1 - 1) + \alpha_2(\rho_2 - 1) &< 0. \end{aligned} \tag{22}$$

Since $\alpha_i > 0, i = 1, 2$ the terms $\rho_{12} - \rho_2$ and $\rho_{12} - \rho_1$ have opposite signs. The same is true of $\rho_1 - 1$ and $\rho_2 - 1$. Assume that $\rho_1 - 1 > 0$ and $\rho_2 - 1 < 0$, the opposite case being analogous. Since $\rho_{12} > 1$ this forces $\rho_{12} - \rho_2 > 0$ and thus $\rho_{12} - \rho_1 < 0$. Then the solution of the set of inequalities (22) is the region in the positive quadrant of the α_1, α_2 plane given by

$$\alpha_2 < \frac{-(\rho_{12} - \rho_1)}{\rho_{12} - \rho_2}\alpha_1, \quad \alpha_2 > \frac{\rho_1 - 1}{-(\rho_2 - 1)}\alpha_1$$

where both slopes are positive. This set has non-empty intersection in the the positive quadrant if and only if

$$\frac{-(\rho_{12} - \rho_1)}{\rho_{12} - \rho_2} > \frac{\rho_1 - 1}{-(\rho_2 - 1)}$$

which is equivalent to

$$(\rho_{12} - \rho_1)(\rho_2 - 1) > (\rho_{12} - \rho_2)(\rho_1 - 1).$$

After simplification this inequality is equivalent to

$$\rho_{12}(\rho_2 - \rho_1) > \rho_2 - \rho_1.$$

Since $\rho_2 < \rho_1$ this contradicts (21). This contradiction finishes the proof of the Proposition. ∎

Example 4.14 (Optimal transcription depends on K_B- vs. k-cooperativity.) We revisit our simplified λ phage lysogen maintenance model and discuss the dependence of the critical value $[r]^*$ on ϕ_{1P} and β_{1P} (see Section 4.3). Recall [25, 26] that the cooperativity between CI_2 and RNAP is accomplished

by O_R2 bound CI_2 increasing ϕ about 12-fold ($\phi_{1P} \approx 12$) without having any significant effect on binding probability of the polymerase ($\beta_{1P} \approx 1$).

When $\beta_{12P} = \beta_2 = 0$ the coefficients of the quadratic equation in 19 are

$$A = -K_1^2 K_2 \beta_{12} \phi_{1P} \beta_{1P}, \quad B = -2K_1 K_2 \beta_{12} \text{ and}$$

and

$$C = K_1(\phi_{1P}\beta_{1P} - 1) - K_2) + \beta_{1P} K_1(\phi_{1P} - 1) K_P[P].$$

Solving for the positive root of the quadratic equation in $[r]$ we get $[r]^* = \frac{1}{2A}(-B + \sqrt{B^2 - 4AC})$.

Now we discuss two cases. First, corresponding to the wild type phage, we let $\beta_{1P} = 1$ and $\phi_{1P} = \delta > 1$. Then

$$[r]^*_{wt} = \frac{1}{-K_1^2 K_2 \beta_{12} \delta}(-B + \sqrt{B^2 + 4K_1^2 K_2 \beta_{12} \delta (K_1(\delta - 1)(1 + \beta_{1P} K_P[P]) - K_2)}. \tag{23}$$

The second case we analyze is in certain sense the opposite of the first one. For this fictitious mutant we set $\beta_{1P} = \delta > 1$ and $\phi_{1P} = 1$, which means that there a CI-RNAP binding cooperation, but CI does not enhance the transcription initiation. Then the critical concentration value is

$$[r]^*_{mut} = \frac{1}{-K_1^2 K_2 \beta_{12} \delta}(-B + \sqrt{B^2 + 4K_1^2 K_2 \beta_{12} \delta (K_1(\delta - 1) - K_2)}. \tag{24}$$

Since $1 + \beta_{1P} K_P[P] > 0$ comparing (23) and (24) we see that

$$[r]^*_{wt} > [r]^*_{mut} \tag{25}$$

at the same δ.

Li *et. al.* [26] removed the positive control of the phage λ by an Arg to His change in the σ^{70} subunit of RNAP. This corresponds to $\beta_{1P} = 1$ and $\phi_{1P} = 1$ in our model. In the same paper Li *et. al* report that when the mutant RNAP was combined with the wild type CI, β_{1P} was increased, without significantly affecting ϕ_{1P}. By comparing the value of $[r^*]$ in such a mutant with the wild type value of $[r^*]_{wt}$ the prediction (25) can be verified experimentally.

Using k-cooperativity ($\phi_{1P} = \delta > 1$ and $\beta_{1P} = 1$) it takes a lower value of cooperation level δ to guarantee that CI_2 is an activator at low concentrations, compared to K_B-cooperativity ($\beta_{1P} = \delta > 1$ and $\phi_{1P} = 1$)

4.5 Two binding sites for two regulators

A canonical example for transcriptional control using two regulators is the E. coli *lac* operon. The transcription of the *lac* operon is controlled by lacI and CAP-cAMP complex (see Figure 1). This type of promoter has two regulators y and z, one binding site for regulator y, one binding site for regulator z and one RNAP binding site. The feasible states are

$$\mathcal{S} = \{s_\emptyset, s_z, s_y, s_P, s_{zP}, s_{zy}\},$$

where $s_{zP} = \{s_z, s_P\}$ and $s_{zy} = \{s_z, s_y\}$. In particular, we assume that the states $s = \{s_z, s_y, s_P\}$ and $s_{yP} = \{s_y, s_P\}$ are not possible because of the mutual overlap between the y and P binding sites. Then the transcription rate is (compare (5))

$$f([z],[y],[P]) = \frac{k_P K_P[P]}{Z}(1 + \phi_z \beta_{zP} K_z[z]), \tag{26}$$

where

$$Z([z],[y],[P]) = 1 + K_z[z] + K_y[y] + \beta_{zy} K_z K_y[z][y] + K_P[P](1 + \beta_{zP} K_z[z]).$$

Furthermore,

$$\begin{aligned} \frac{\partial f}{\partial [z]}([z],[y],[P]) &= \frac{k_P K_P K_z[P]}{Z^2}((\rho_z - 1)(1 + \beta_{zP} K_P[P]) \\ &\quad + (\phi_z \beta_{zP} - \beta_{zy}) K_y[y]) \end{aligned} \tag{27}$$

$$\frac{\partial f}{\partial [y]}([z],[y],[P]) = -\frac{k_P K_P[P]}{Z^2}(1 + \phi_z \beta_{zP} K_z[z])(K_y + \beta_{zy} K_z K_y[z]) \tag{28}$$

It is clear from expression (28) that the regulatory protein y is a phenomenological repressor. But in general we cannot label z as a phenomenological activator or repressor.

Example 4.15 Now we specialize further to the *lac* operon where y represents lacI and z represents the CAP-cAMP complex. By [27] it is known that $\phi_{zP} = 1$, and by [12] we know $\beta_{zy} = 1$ and $\beta_{zP} > 1$, therefore (27) simplifies to

$$\frac{\partial f}{\partial [z]}([z],[y],[P]) = \frac{k_P K_P K_z[P]}{Z^2}(\beta_{zP} - 1)(1 + K_y[y]) > 0.$$

It follows that z (CAP-cAMP) is a phenomenological activator. Therefore in the *lac* operon setup with the repressor blocking transcription by preventing RNA polymerase binding we recover the correspondence between biochemical and macroscopic markers of activation and repression.

An immediate consequence of having $\phi_{zP} = 1$ is that the sign of (27) has no dependence on RNA polymerase concentration. It is interesting to notice though that if $\phi_{zP} > 1$ instead of $\phi_{zP} = 1$, then z would still be a phenomenological activator and y a phenomenological repressor. Not only that, but z would be a more effective activator and y a more effective repressor since the derivatives would be greater in absolute value, but would keep the same signs. This situation could be achieved by moving the CAP binding region. This is the situation in the *gal* operon promoter P_1 [27]. It is not clear why this is not the regulation process adopted in the *lac* operon, since it seems that would be more effective regulation. It would be interesting to investigate whether there are other constraints that force *E. coli* to use this less than optimal regulator.

Another observation is that $\beta_{zy} = 1$ is also not a requirement for z to be a phenomenological activator and y a phenomenological repressor. Assuming $\beta_{zP} > 1$, it is sufficient to have $\beta_{zy} \leq \phi_{zP}\beta_{zP}$ to still have the same result. However, with these changes z would be less effective as an activator, and y less effective as a repressor. On the other hand, if $\beta_{zy} > \phi_{zP}\beta_{zP}$, then for high concentrations of y the regulator z would work as a repressor instead.

5 Discussion

Most of the conceptual models of transcriptional regulation assume monotonicity of the function relating product mRNA and the concentration of the regulator. This perhaps reflects the prevailing mode of data collection through knockout experiments where the absence of the putative regulator causes either an increase or a decrease of the mRNA production. On the modeling front this assumption leads to widespread use of Hill type response functions.

The Shea-Ackers model of transcriptional regulation was introduced more then 20 years ago. Using the chemical equilibrium assumption and using experimentally accessible parameters the resulting Shea-Ackers function relates concentrations of regulatory proteins, RNAP and the geometry of the promoter to the transcription rate. The model has been matched to experimental

data and the necessary parameters have been measured for at least a couple of canonical examples like *lac* operon in E. coli and phage λ switch.

The Shea-Ackers function reduces to a Hill function only in the case when there are no regulatory proteins. We show however that the Shea-Ackers function is still monotone for a promoter that contains a single binding site for a single regulatory protein in addition to a RNAP binding site. If there are more binding sites, or more regulatory proteins, then non-monotonicity is common. While this non-monotonicity is used by certain organisms (CI_2 control of its own expression in phage λ), it may be tightly controlled in other cases by keeping concentrations of regulatory proteins in monotone regions. This opens up many new questions about regulatory circuit design and perhaps points to a need to revisit results that were obtained using the Hill model response function.

For all but the simplest of operators the key parameters of the Shea-Ackers model have a complicated, nonlinear effect of monotonicity of the transcription rate. We define a new regulatory constant ρ which greatly simplifies characterization of activation and repression for several complicated promoter designs. Since the constant is experimentally accessible it provides a new tool for the understanding of existing operators as well as the design new ones.

Acknowledgments

T. G. was partially supported by NSF/NIH grant 0443863, NIH-NCRR grant PR16445 and NSF-CRCNS grant 0515290. K.M. was partially supported by NSF DMS 0443827 and grants from D.O.E. and DARPA. K.P. was supported by NSF/NIH grant 0443863. E.T. was partially supported by NSF DMS 0443827 and CAPES, Brazil.

References

[1] Lartigue C, Glass J, Alperovich N, Pieper R, Parmar P, et al. (2007) Genome transplantation in bacteria: changing one species to another. Science 317:632–638.

[2] Elowitz M, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338.

[3] Gardner T, Cantor C, Collins J (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342.

[4] Alon U (2007) An introduction to systems biology. Chapman & Hall/CRC.

[5] Ptashne M, Gann A (2001) Genes and Signals. Cold Spring Harbor Laboratory Press.

[6] Angeli D, Ferrell J, Sontag E (2004) Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. Proc Natl Acad Sci 101:1822–1827.

[7] Shea M, Ackers G (1985) The O_R control system of bacteriophage lambda : A physical-chemical model for gene regulation. J Mol Bio 181:211–230.

[8] Bintu L, Buchler N, Garcia H, Gerland U, Hwa T, et al. (2002) Transcriptional regulation by the numbers: models. Curr Opin Genet Dev 15:116–124.

[9] Ackers G, Johnson A, Shea M (1982) Qualitative model for gene regulation by λ phage repressor. Proc Natl Acad Sci 79:1129–1133.

[10] Gedeon T, Mischaikow K, Patterson K, Traldi E (2007) Binding cooperativity in phage λ is not sufficient to produce an effective switch. Preprint .

[11] Santillán M, Mackey M (2004) Why the lysogenic state of phage λ is so stable: A mathematical modeling approach. Biophysical Journal 86:75–84.

[12] Santillán M, Mackey M (2004) Influence of catabolite repression and inducer exclusion on the bistable behavior of the *lac* operon. Biophysical Journal 86:1282–1292.

[13] Mallet-Paret J, Smith H (1990) The Poincaré-Bendixson theorem for monotone feedback systems. J Dynam Diff Eq 2:367–421.

[14] Gedeon T (1998) Cyclic feedback systems, volume 637 of *Memoirs of AMS*. Americal Mathematical Society.

[15] Ptashne M (2004) A Genetic Switch. Cold Spring Harbor Laboratory Press.

[16] Hill T (1960) Introduction to Statistical Thermodynamics. Addison Wesley.

[17] Snyder L, Champness W (2003) Molecular Biology of Bacteria. ASM Press.

[18] M Nollman NC, Arimondo P (2007) Thirty years of Escherichia coli DNA gyrase: From in vivo function to single-molecule mechanism. Biochimie 89:490–499.

[19] Schroder O, Wagner R (2000) The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. J Mol Biol 298:737–747.

[20] Schneider R, Travers A, Kuteleladze T, Muskhelishvili G (1999) A DNA architectural protein couples cellular physiology and DNA topology in Escherichia coli. Molecular Microbiology 35:953–964.

[21] Bremer H, Dennis P (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In: et al FN, editor, In Escherichia coli and Salmonella thyphymurium: Cellular and Molecular Biology. volume 2.

[22] Watson J, Baker T, Bell S, Gann A, Levine M, et al. (2003) Molecular Biology of the Gene. Benjamin Cummings.

[23] Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor. Proc Natl Acad Sci 85:8973–8977.

[24] Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherishia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. Nucleic Acids Research 25:1203–1210.

[25] Hawley D, McClure W (1983) The effect of a lambda repressor mutation on the activation of transcription initiation from the lambda p_{RM} promoter. Cell 32:327–333.

[26] M Li WM, Susskind M (1997) Changing the mechanism of transcriptional activation by phage λ repressor. Proc Natl Acad Sci 94:3691–3696.

[27] Busby S, Ebright R (1999) Transcription activation by catabolite activator protein (CAP). J Mol Biol 293:199–213.