On the Use of the Hill Functions in Mathematical Models of Gene Regulatory Networks

M. Santillán [1]

Centro de Investigación y de Estudios Avanzados del IPN
Unidad Monterrey Vía del Conocimiento 201
Parque de Investigación e Innovación Tecnológica
Autopista Monterrey-Aeropuerto km 9.5, 66600 Apodaca NL, México

Abstract. Hill functions follow from the equilibrium state of the reaction in which n ligands simultaneously bind a single receptor. This result if often employed to interpret the Hill coefficient as the number of ligand binding sites in all kinds of reaction schemes. Here, we study the equilibrium states of the reactions in which n ligand bind a receptor sequentially, both non-cooperatively and in a cooperative fashion. The main outcomes of such analysis are that: n is not a good estimate, but only an upper bound, for the Hill coefficient; while the Hill coefficient depends quite strongly on the cooperativity level among ligands. We finally use these results to discuss the feasibility and constrains of using Hill functions to model the regulatory functions in mathematical models of gene regulatory networks.

Key words: systems biology, molecular biology, cooperativity, Hill coefficient
AMS subject classification: 92C40

1. Introduction

The so called Hill functions were introduced by A. V. Hill [5] in 1910 to describe the binding of oxygen to hemoglobin. Subsequently, they have been widely used in biochemistry, physiology, and pharmacology to analyze the binding equilibria in ligand-receptor interactions [15]. Hill functions have also been employed in mathematical models of gene expression since the very early attempts in the late sixties [2, 3], and continue to be used for instance in mathematical models of

---

[1]E-mail: msantillan@cinvestav.mx

the somitogenesis segmentation clock [6, 7]

The Hill function parameters provide a measure of the ligand affinity for the receptor (parameter K in Equation 2.4), but Hill functions are commonly used as well to estimate the number of ligand molecules that are required to bind the receptor in order to produce a functional effect. The latter estimate is obtained from the Hill coefficient (n in Equation 2.4). Nevertheless, Hill himself recognized that the conditions under which the Hill coefficient provides an accurate estimate for the number of binding sites are very specific: only when extreme positive cooperativity is present between the binding of the first and subsequent ligand molecules. That is, the affinity of the binding has to be very asymmetric, with a much lower affinity of binding for the first ligand molecule than for the subsequent ones. This implies that the Hill coefficient is more appropriately described as an *interaction coefficient*, reflecting cooperativity, rather than as a reliable estimate for the number of binding sites.

This paper is devoted to tackling the above issue discussed in the previous paragraphs. We start by reproducing in Section 2. Hill's derivation of the functions that bear his name. To do this we analyse a reaction in which n ligand molecules simultaneously bind a single receptor. Then, we study in Section 3. the situation in which n ligands sequentially bind a receptor molecule, in a non-cooperative fashion. In that case, we demonstrate that increasing the number of ligand binding sites has very little effect on the Hill coefficient. Indeed, we prove there that the Hill coefficient tends to $2 \ln 2$ as the ligand count tends to infinity. We further investigate the cooperative sequential binding of n ligands to a single receptor in Section 4., and prove that Hill functions are recovered in the limit of very high cooperativity. Finally, we use the results of Sections 2.-4. to discuss the feasibility and constrains of using Hill functions in mathematical models of gene regulatory networks in Section 5..

2.   Derivation of the Hill function

Consider the reaction

$$A + nB \underset{}{\overset{K_D}{\rightleftharpoons}} A_{nB}, \tag{2.1}$$

in which n ligand molecules B bind simultaneously the receptor A. The following relation holds in the state of chemical equilibrium:

$$[A][B]^n = K_D[A_{nB}], \tag{2.2}$$

where K is the reaction dissociation constant, and $[X]$ denotes the concentration of chemical species X. Assume now a constant number of receptors:

$$[A] + [A_{nB}] = [A_{Tot}]. \tag{2.3}$$

It follows then from Equations (2.2) and (2.3) that the fractions of occupied and free molecules A are respectively given by

$$H^{(1)}([B]) = \frac{[A_{nB}]}{[A_{Tot}]} = \frac{[B]^n}{K^n + [B]^n} \quad \text{and} \quad H^{(2)}([B]) = \frac{[A]}{[A_{Tot}]} = \frac{K^n}{K^n + [B]^n}, \tag{2.4}$$

with $K^n = K_D$.

The functions $H^{(1)}([B])$ and $H^{(2)}([B])$, defined in equation (2.4), are both known as Hill functions. $H^{(1)}([B])$ is a monotonic growing function of $[B]$ that satisfies the following properties: $H^{(1)}(0) = 0$, $H^{(1)}(K) = 1/2$, and $\lim_{[B]\to\infty} H^{(1)}([B]) = 1$. Similarly, $H^{(2)}([B])$ is a monotonic decreasing function of $[B]$ that complies with: $H^{(2)}(0) = 1$, $H^{(2)}(K) = 1/2$, and $\lim_{[B]\to\infty} H^{(1)}([B]) = 0$. Finally $H^{(1)}([B]) + H^{(2)}([B]) = 1$.

By defining $x = [B]/K$, new renormalized Hill functions can be established as

$$h^{(1)}(x) = H^{(1)}(Kx) = \frac{x^n}{1 + x^n} \quad \text{and} \quad h^{(2)}(x) = H^{(2)}(Kx) = \frac{1}{1 + x^n} \tag{2.5}$$

It's straightforward to see that $h^{(1)}(1) = h^{(2)}(1) = 1/2$, and that $h^{(1)}(x) + h^{(2)}(x) = 1$. Functions $h^{(1)}(x)$ and $h^{(2)}(x) = 1$ are plotted in Figure 1 for various values of n.
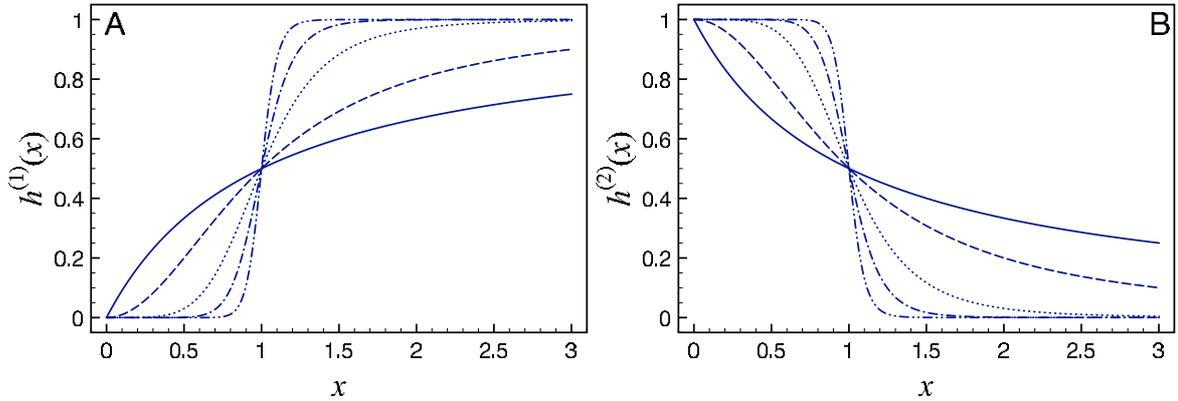

Figure 1: Plots of (A) $h^{(1)}(x)$ and (B) $h^{(2)}(x)$, as respectively defined by Equation (2.5), for different values of n. The solid lines correspond to $n = 1$, the dashed lines correspond to $n = 2$, the dotted lines correspond to $n = 5$, the dot-dashed lines correspond to $n = 10$, and the dot-dot-dashed lines correspond to $n = 20$.
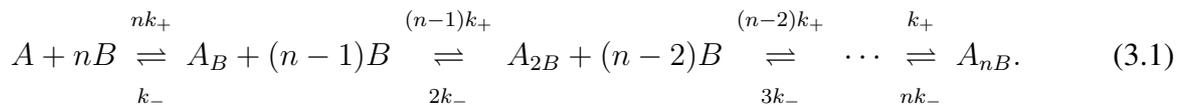
As can be seen in Figure 1, the *sigmoidicity* of both Hill functions increases with n. The absolute value of a Hill function slope at $x = 1$, $M = \text{abs}(dh^{(i)}/dx|_{x=1})$, can be regarded as an estimate for its sigmoidicity, and it is straightforward to show that $M = n/4$ for both $h^{(1)}(x)$ and $h^{(2)}(x)$. Furthermore, it can easily be proven that

$$\lim_{n\to\infty} h^{(1)}(x) = \begin{cases} 0 & x < 1, \\ 1/2 & x = 1, \\ 1 & x > 1, \end{cases} \quad \text{and} \quad \lim_{n\to\infty} h^{(2)}(x) = \begin{cases} 1 & x < 1, \\ 1/2 & x = 1, \\ 0 & x > 1. \end{cases}$$

3.   Sequential binding of n ligands B to a single receptor A

Consider the reaction where n ligands B sequentially bind a receptor A at n independent binding sites. Assume that the forward and backward reaction rates for the binding of a molecule B are k_+

and k_-, respectively. The sequence of reactions that lead to the complex formed by the receptor A bound by n ligands B (A_{nB}) are [10]:

$$A + nB \underset{k_-}{\overset{nk_+}{\rightleftharpoons}} A_B + (n-1)B \underset{2k_-}{\overset{(n-1)k_+}{\rightleftharpoons}} A_{2B} + (n-2)B \underset{3k_-}{\overset{(n-2)k_+}{\rightleftharpoons}} \cdots \underset{nk_-}{\overset{k_+}{\rightleftharpoons}} A_{nB}. \tag{3.1}$$

In the above reactions, the term A_{iB} ($i = 1 \ldots n$) denotes the complex formed by one receptor A bound by i ligands B. The following relations hold in the state of chemical equilibrium:

$$
\begin{aligned}
[A_B] &= nx[A], \\
[A_{2B}] &= \frac{n-1}{2}x[A_B], \\
[A_{3B}] &= \frac{n-2}{3}x[A_{2B}], \\
&\vdots \\
[A_{nB}] &= \frac{1}{n}x[A_{(n-1)B}],
\end{aligned}
\tag{3.2}
$$

where $x = [B]/K$ and $K = k_-/k_+$.

Take the equation set (3.2), solve for $[A_B]$ in the first equation and substitute in the second one; then, solve for $[A_{2B}]$ in the resulting equation and substitute in the third equation; repeat this process until $[A_{nB}]$ is solved for in the last equation. When this procedure is concluded, the concentrations of all partially and fully occupied molecules A come out to be:

$$[A_{iB}] = \binom{n}{i} x^i [A], \quad i = 0 \ldots n, \tag{3.3}$$

with

$$\binom{n}{i} = \frac{n!}{(n-i)!i!}.$$

The total concentration of A receptors can be calculated from Equation (3.3) as

$$[A_{Tot}] = \sum_{i=0}^{n} [A_{iB}] = (1+x)^n [A]. \tag{3.4}$$

Thus, from Equations (3.3) and (3.4), the fraction of receptors A bound by i molecules B is:

$$\frac{[A_{iB}]}{[A_{Tot}]} = \frac{\binom{n}{i} x^i}{(1+x)^n}. \tag{3.5}$$

In particular, the factions of totally occupied and free receptors A are respectively given by

$$S^{(1)}(x) = \frac{[A_{nB}]}{[A_{Tot}]} = \frac{x^n}{(1+x)^n} \quad \text{and} \quad S^{(2)}(x) = \frac{[A]}{[A_{Tot}]} = \frac{1}{(1+x)^n} \tag{3.6}$$

Let x^* be the value of x such that $S^{(1)}(x^*) = 1/2$. From Equation (3.6):

$$x^* = \frac{1}{2^{1/n} - 1}. \tag{3.7}$$

Notice that x^* increases with n. Indeed, $\lim_{n\to\infty} x^* = \infty$. In order to compare the behaviour of $S^{(1)}(x)$ for different values of n define

$$s^{(1)}(x) = S^{(1)}(x^* x) = \frac{x^n}{(x + 2^{1/n} - 1)^n}. \tag{3.8}$$

It is easy to see from the definition of Function $s^{(1)}(x)$ that its plots can be constructed by horizontally shrinking the corresponding $S^{(1)}(x)$ *vs.* x plots, so they reach half their maximum value at $x = 1$. Function $s^{(1)}(x)$ is plotted *vs.* x for various values of n in Figure 2A. Observe that all the $s^{(1)}$ *vs.* x curves are sigmoidal and that the their sigmoidicity increases with n. However, this increase is not as rapid as in the growing Hill function. To see this consider that

$$\lim_{n\to\infty} s^{(1)}(x) = g^{(1)}(x) = 2^{-1/x}. \tag{3.9}$$

The proof of this assertion is as follows:

$$
\begin{aligned}
\lim_{n\to\infty} s^{(1)}(x) &= \lim_{n\to\infty} \frac{x^n}{\left(x + 2^{1/n} - 1\right)^n}, \\
&= \exp\left(\lim_{n\to\infty} \frac{-\ln\left((x + 2^{1/n} - 1)/x\right)}{1/n}\right), \\
&= \exp\left(\lim_{n\to\infty} \frac{\frac{2^{1/n}\ln(2)}{n^2(x + 2^{1/n} - 1)}}{-\frac{1}{n^2}}\right), \\
&= 2^{-1/x}.
\end{aligned}
$$

Furthermore, a simple calculation leads to

$$\left.\frac{dg^{(1)}(x)}{dx}\right|_{x=1} = \frac{\ln 2}{2}.$$

Recall that the slope at $x = 1$ of the Hill function $h^{(1)}(x)$ is $n/4$. Therefore, Function $s^{(1)}(x)$ can be fitted by a Hill function with coefficient $n_H = 2\ln 2 \approx 1.39$ in the limit where the number of ligand binding sites tends to infinity. This last result proves that the sigmoidicity of Function $s^{(1)}(x)$ does not increase indefinitely with n but is upper bounded, and that the highest possible sigmoidicity corresponds to that of a Hill function with coefficient $n_H \approx 1.39$.

Let x^{**} be defined as

$$x^{**} = 2^{1/n} - 1. \tag{3.10}$$

Substitution of x^{**} into $S^{(2)}(x)$ —Equation (3.6)— renders

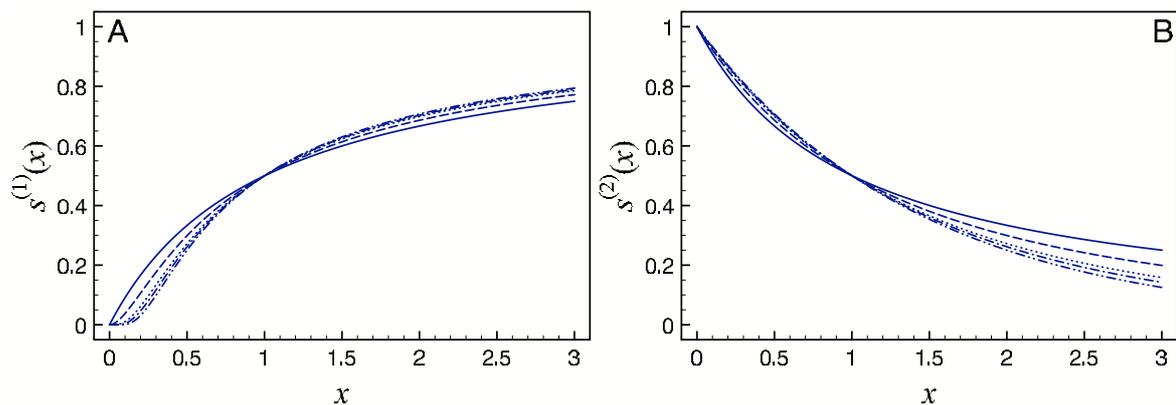$$S^{(2)}(x^{**}) = \frac{1}{(1 + x^{**})^n} = \frac{1}{2}.$$

Figure 2: Plots of $s^{(1)}(x)$ and $s^{(2)}(x)$, as defined by Equations (3.8) and (3.11), for different values of n. The solid lines correspond to $n = 1$, the dashed lines correspond to $n = 2$, the dotted lines correspond to $n = 5$, and the dot-dashed lines correspond to $n = 10$. Finally the dot-dot-dashed lines are the plots of (A) $g^{(1)}(x)$ and (B) $g^{(2)}(x)$, as respectively defined by Equations (3.9) and (3.12).

From its definition, x^{**} is a decreasing function of n. In fact,

$$\lim_{n \to \infty} x^{**} = 0.$$

Define

$$s^{(2)}(x) = S^{(2)}(x^{**}x) = \frac{1}{\left(1 + (2^{1/n} - 1)x\right)^n}. \tag{3.11}$$

That is, the $s^{(2)}(x)$ vs. x plots can be traced by horizontally stretching the corresponding $S^{(2)}(x)$ vs. x plots, so they reach half their maximum value at $x = 1$. With this, we can study the influence of the number n of B binding sites on function $S^{(2)}(x)$, while still working in a fixed scale.

Plots of $s^{(2)}(x)$ vs. x are shown in Figure 2B for different values of N. Note that, in all cases, $s^{(2)}(0) = 1$, $s^{(2)}(1) = 1/2$, and $\lim_{x \to \infty} s^{(2)}(x) = 0$. Furthermore, the sigmoidicity of the $s^{(2)}(x)$ vs. x curves increases with n, but only up to a certain point. To see this, take the limit of $s^{(2)}(x)$ as $n \to \infty$:

$$
\begin{aligned}
\lim_{n \to \infty} s^{(2)}(x) &= \lim_{n \to \infty} \frac{1}{\left(1 + (2^{1/n} - 1)x\right)^n}, \\[2mm]
&= \exp\left(\lim_{n \to \infty} \frac{-\ln\left(1 + (2^{1/n} - 1)x\right)}{1/n}\right), \\[2mm]
&= \exp\left(\lim_{n \to \infty} \frac{\frac{2^{1/n}\ln(2)x}{n^2\left(1 + (2^{1/n} - 1)x\right)}}{-\frac{1}{n^2}}\right), \\[2mm]
&= 2^{-x}.
\end{aligned}
$$

This proves that as $n \to \infty$, $s^{(2)}(x)$ tends to the function $g^{(2)}(x)$ defined as follows:

$$g^{(2)}(x) = 2^{-x}. \tag{3.12}$$

Moreover,

$$\left. \frac{dg^{(2)}(x)}{dx} \right|_{x=1} = -\frac{\ln 2}{2}.$$

It then follows from this and the fact that the slope at $x = 1$ of the Hill function $h^{(2)}(x)$ is $-n_H/4$— where n_H is the Hill coefficient— that Function $g^{(2)}(x)$ can be fitted by a decreasing Hill function with coefficient $n_H = 2 \ln 2 \approx 1.39$.

In conclusion, when n ligands B bind a receptor A sequentially, the fractions of free and fully receptors can be approximated by Hill functions. The number n of B binding sites has a very strong influence on the apparent dissociation constant: the larger n, the higher the concentration of B required to occupy all the binding sites, and the smaller the concentration of B required to have al least one binding site occupied. On the other hand, n has very little impact on the Hill coefficient. Even in the limit $n \to \infty$, the Hill coefficient only reaches the value 1.39.

4. Sequential cooperative binding of n ligands B to a single receptor A

Two ligands B bind cooperatively a receptor A if the binding of the first ligand to an available site increases the chances that a second ligand binds an empty nearby site. Assume that every molecule A has two identical binding sites for the B ligands, and that the reaction through which the first ligand binds the receptor if of the form:

$$A + B \underset{k_-}{\overset{2k_+}{\rightleftharpoons}} A_B.$$
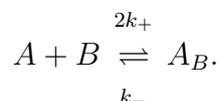
In the above reaction k_+ and k_- respectively represent the forward and backward reaction rates. The factor 2 multiplying k_+ accounts for the fact that there are two available binding sites. The equilibrium equation for this reaction is:

$$[A][B] = \frac{K}{2}[A_B], \tag{4.1}$$

where $K = k_-/k_+$ is the reaction dissociation constant.

Because of cooperativity either k_+ is increased or k_- is decreased in the reaction where the second ligand binds the receptor. In either case, the value of the corresponding dissociation constant diminishes. Hence, the corresponding reaction is
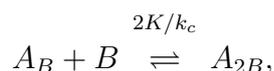
$$A_B + B \overset{2K/k_c}{\rightleftharpoons} A_{2B},$$

where the factor 2 accounts for the fact that either two of the bound ligands can detach from the complex A_{2B} at a given time, and $k_c > 1$ is a constant accounting for cooperativity. The equilibrium equation for this reaction is

$$[A_B][B] = \frac{2K}{k_c}[A_{2B}]. \tag{4.2}$$
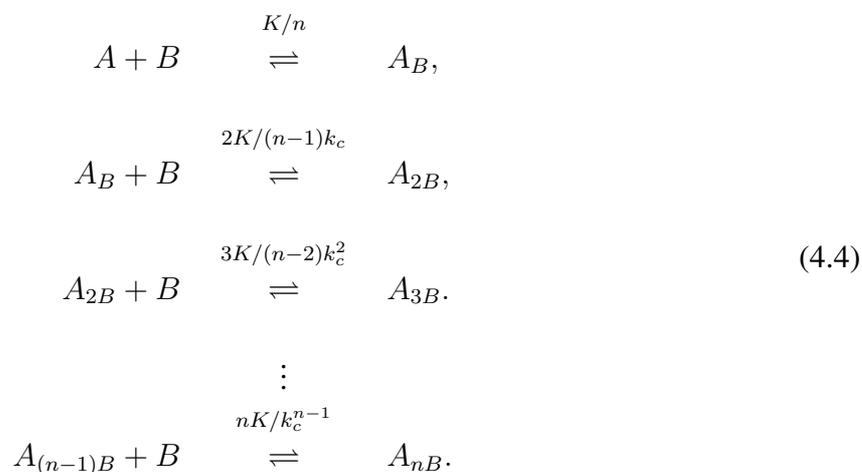
Assume that the count of A molecules is constant:

$$[A] + [A_B] + [A_{2B}] = [A_{Tot}]. \tag{4.3}$$

Then, it is possible to solve for $[A]$, $[A_B]$, and $[A_{2B}]$ from Equations (4.1), (4.2), and (4.3). In particular:

$$\frac{[A]}{[A_{Tot}]} = \frac{1}{1 + 2\dfrac{[B]}{K} + k_c \left(\dfrac{[B]}{K}\right)^2},$$

$$\frac{[A_{2B}]}{[A_{Tot}]} = \frac{k_c \left(\dfrac{[B]}{K}\right)^2}{1 + 2\dfrac{[B]}{K} + k_c \left(\dfrac{[B]}{K}\right)^2}.$$

Consider the situation in which a receptor A that can be bound by up to n ligands B. Assume that, when the i-th ($i = 2, 3 \ldots n - 1$) ligand binds the receptor, it interacts cooperatively with the $i - 1$ ligands already bound. The corresponding reaction sequence is:

$$
\begin{aligned}
A + B \quad &\overset{K/n}{\rightleftharpoons} \quad A_B, \\[2mm]
A_B + B \quad &\overset{2K/(n-1)k_c}{\rightleftharpoons} \quad A_{2B}, \\[2mm]
A_{2B} + B \quad &\overset{3K/(n-2)k_c^2}{\rightleftharpoons} \quad A_{3B}. \\[2mm]
&\qquad\vdots \\[2mm]
A_{(n-1)B} + B \quad &\overset{nK/k_c^{n-1}}{\rightleftharpoons} \quad A_{nB}.
\end{aligned}
\tag{4.4}
$$

In the above reaction, K denotes the dissociation constant for the reaction in which a molecule B binds its corresponding site on molecule A, while the constant $k_c > 1$ accounts for cooperativity.

The equilibrium equations for the reactions in Equation (4.4) are:

$$
\begin{aligned}
[A_B] &= n\frac{[B]}{K}[A], \\
[A_{2B}] &= \frac{n-1}{2}k_c\frac{[B]}{K}[A_B], \\
[A_{3B}] &= \frac{n-2}{3}k_c^2\frac{[B]}{K}[A_{2B}], \\
&\vdots \\
[A_{nB}] &= \frac{1}{n}k_c^{n-1}\frac{[B]}{K}[A_{(n-1)B}].
\end{aligned}
\tag{4.5}
$$

It is straightforward to demonstrate after a little algebra that Equation (4.5) reduces to:

$$
[A_{iB}] = \binom{n}{i}\left(\frac{x}{k_c^{(n-i)/2}}\right)^i [A],
\tag{4.6}
$$

where

$$
x = k_c^{(n-1)/2}\frac{[B]}{K}.
$$

Hence, under the assumption that the receptor count remains constant, the corresponding conservation equation is:

$$
\sum_{i=0}^{n}[A_{iB}] = \sum_{i=0}^{n}\binom{n}{i}\left(\frac{x}{k_c^{(n-i)/2}}\right)^i [A] = [A_{Tot}].
\tag{4.7}
$$

Finally, the equilibrium concentrations $[A_{iB}]$ $(i = 0 \dots n)$ can be calculated from Equations (4.6) and (4.7) as

$$
[A_{iB}] = \frac{\binom{n}{i}\left(\frac{x}{k_c^{(n-i)/2}}\right)^i}{\sum_{j=0}^{n}\binom{n}{j}\left(\frac{x}{k_c^{(n-j)/2}}\right)^j}[A_{Tot}].
\tag{4.8}
$$

Define

$$
\Xi_i(x) = \binom{n}{i}\left(\frac{x}{k_c^{(n-i)/2}}\right)^i.
$$

Notice that $\Xi_0(x) = 1$ and $\Xi_n(x) = x^n$, for all $k_c \geq 1$. Furthermore, it is not hard to prove that

$$
\lim_{k_c \to \infty} \Xi_i(x) = 0, \quad \forall i = 1, 2 \dots n-1,
$$

as long as x does not diverge. Thus, we have from equation (4.8) that, in the limit of very high

cooperativity:

$$[A] \approx \frac{1}{1+x^n}, \tag{4.9}$$

$$[A_{nB}] \approx \frac{x^n}{1+x^n}, \tag{4.10}$$

$$[A_{iB}] \approx 0, \quad \forall i \neq 0, n. \tag{4.11}$$

Thus, according to these last results, the situation in which n ligands simultaneously bind a receptor is recovered, in the limit of very high cooperativity, from the model where n ligands sequentially bind a receptor in a cooperative fashion.

5. Discussion and conclusions

In the model in which n ligands B bind a receptor A, the fractions of empty and occupied molecules A are respectively given by the decreasing and increasing Hill functions $H^{(2)}([B])$ and $H^{(1)}([B])$ —defined in Equation (2.4). These equations involve two parameters: K and n_H. Parameter K is the value of $[B]$ at which functions $H^{(2)}([B])$ and $H^{(1)}([B])$ reach half their maximum value, while parameter n_H determines the *sigmoidicity* of the $H^{(2)}$ *vs.* $[B]$ and $H^{(1)}$ *vs.* $[B]$ curves: the larger n_H, the steeper the curves at $[B] = K$. Furthermore, in this particular model, $K^n = K_D$, where K_D is the dissociation constant of the reaction $A + nB \rightleftharpoons A_{nB}$, while n_H is the number of molecules B binding A.

Consider a gene that is regulated by a transcription factor P, and let M denote the RNA molecular species resulting from this gene transcription. Under the assumption that chemical kinetics can be used to model gene expression and gene regulation, the equation governing the dynamics of the RNA concentration $[M]$ is:

$$\frac{d[M]}{dt} = k_M [D] \mathcal{R}([P]) - \gamma_M [M]. \tag{5.1}$$

In the above equation $[D]$ represents the concentration of gene copies, k_M is the maximum transcription rate of a gene copy, γ_M is the RNA degradation rate, and $\mathcal{R}([P])$ accounts for transcription regulation by the transcription factor P.

Although in some cases a transcription factor can act as both an activator and a repressor for different concentration ranges, most transcription factors can be classified as either one or the other. If P is an activator (repressor), the regulatory function $\mathcal{R}([P])$ comes out to be monotonically increasing (decreasing).

The transcription rate of every gene is upper bounded and this bound is determined by the promoter sequence and the type of polymerase carrying out transcription. From this and the fact that k_M is the gene maximum transcription rate, $\mathcal{R}([P])$ can be interpreted as the probability that a gene copy is transcribed at a given time, as a function of $[P]$.

Most people believe that transcription factors regulate gene expression in a switch-like fashion. Indeed, this supposition is behind Boolean-like models of gene regulatory networks, which are quite popular for networks involving a large number of genes [1]. If true, this means that the regulatory function $\mathcal{R}([P])$ is highly sigmoidal.

Thus, from the considerations in the previous paragraphs, the gene regulation function $\mathcal{R}([P])$ can in general be modelled via Hill-type equations.

Transcription factors (TFs) usually act as polymers (dimers or tetramers). Moreover, there are frequently multiple TF binding sites on the DNA regulatory region, or a single TF can simultaneously bind two different sites. Not only that but, TFs often need to be activated or they can be inactivated by the binding of two or more metabolites. From these considerations and the fact that the Hill coefficient corresponds to the number of ligand molecules simultaneously binding a receptor, as shown in Section 2., the coefficient n_H of the Hill function employed to model $\mathcal{R}([P])$ is often estimated by simply counting the number of binding sites for a given molecule.

In Section 3. we demonstrated that, when n ligands B sequentially bind a receptor A in a noncooperative manner, the fractions of free and totally occupied molecules A as functions of $[B]$ can be fit by decreasing and increasing Hill functions, respectively. The values of the resulting half saturation constants K are quite sensible to n, in opposition to the behaviour of the Hill coefficient. Indeed, in both cases the the Hill coefficient tends to 1.39 when $n \to \infty$. This last result proves that, although Hill functions are quite useful to model gene regulatory functions, the Hill coefficient cannot be estimated by simply counting the number of binding sites for a given ligand. At best, the binding site count is an upper bound for n_H, but real situations are usually far from this bound.

The reactions through which n ligands B sequentially bind a receptor A in a cooperative fashion are analysed in Section 4.. In the absence of cooperativity, the results of Section 3. are recovered. On the other hand, the situation where n ligands bind a receptor simultaneously is recovered in the limit of very high cooperativity. Therefore, contrarily to the number n of binding sites, the cooperativity level has a quite strong effect on the sigmoidicity of the functions that determine the fraction of free ($[A]/[A_{Tot}]$) and fully occupied ($[A_{nB}]/[A_{Tot}]$) receptors. In conclusion, when n ligands bind a receptor sequentially, n represents an upper bound for the coefficient of the Hill functions that best fit functions $[A]/[A_{Tot}]$ and $[A_{Tot}]/[A_{Tot}]$ vs. $[B]$, and the level of cooperativity determines how much the real system approaches this ideal situation.

Interestingly, the most extensively studied gene regulatory systems (the tryptophan and lactose operons of *Escherichia coli*) make use of cooperativity to increase the sigmoidicity of the regulatory functions.

The tryptophan (*trp*) operon consists of the genes whose encoded proteins constitute the enzymes necessary to synthesize the amino acid tryptophan, and of an upstream regulatory region. This regulatory region contains a binding site for mRNA polymerases (which carry out transcription) as well three binding sites (operators) for repressor molecules. A repressor needs to be activated by the binding of two tryptophan molecules so it can bind an operator. Therefore, the higher the intracellular tryptophan concentration, the higher the amount of active repressors. In turn, these active repressors bind the operators and inhibit transcription initiation by blocking the polymerase binding site. The above described mechanism receives the name of repression, and is one of the three regulatory processes found in the *trp* operon. The binding of active repressors to the first

95

two operators occurs cooperatively and this increases the sigmoidicity of the regulatory function accounting for repression [4, 14]. Furthermore, it has been shown that a high level on sigmoidicity is necessary for the the bacteria to maintain a constant intracellular tryptophan concentration, regardless of variations in the amino acid concentration in the extracellular medium [16].

The genes in the lactose (*lac*) operon code for the proteins necessary to transport and metabolize the sugar lactose. Repression is also one of the regulatory mechanisms in this operon. In this case, the repressor is active in its free form and is inactivated when bound by allolactose —a by-product of lactose metabolism [9]. The *lac* operon comprises three different repressor binding sites, and a single active repressor can simultaneously bind two of them by way of DNA folding [8]. Given that the binding energy of the complex formed by a repressor simultaneously binding two different operators is larger than the sum of binding energies associated to the complexes formed by one repressor bound to each operator, this simultaneous binding is a kind of cooperativity. Furthermore, it has been shown that this cooperativity increases the sigmoidicity of the corresponding regulatory function, which in turn is essential to explain the operon bistable behaviour [12]. Finally, bistability has been argued to help *E. coli* consume glucose and lactose in the most efficient possible way [13, 11].

In conclusion, Hill functions of the form

$$H^{(1)}([X]) = \frac{[X]^{n_H}}{K^{n_H} + [X]^{n_H}} \quad \text{and} \quad H^{(2)}([X]) = \frac{K^{n_H}}{K^{n_H} + [X]^{n_H}}$$

are extremely useful to model gene regulatory functions, like $\mathcal{R}([P])$ in Equation (5.1), because they have many of the experimentally-observed required characteristics. However, modellers should be extremely careful when interpreting the biological significance of parameters K and n_H. K represents the value of $[X]$ at which both Hill functions reach half their maximum value, while n_H determines the sigmoidicity or steepness at $[X] = K$ of the $H^{(i)}$ ($i = 1, 2$) *vs.* $[X]$ curves. However, it is not possible to deduce the details of the DNA-transcription factor chemical interactions from the values of K or n_H. In particular, a high n_H value implies a high level of co-operativity, but no further assertions can be made from this fact. Conversely, if the gene regulatory mechanisms are known in detail, it is not possible to deduce by simple inspection the values of the best fitting Hill function parameters.

Acknowledgements

References

[1] S. Bornholdt. *Boolean network models of cellular regulation: prospects and limitations*. J. Roy. Soc. Interface, 5 (2008), Suppl. 1:S85-S94.

[2] J. S. Griffith. *Mathematics of cellular control processes. I. Negative feedback to one gene.* J. Theor. Biol., 20 (1968), 202-208.

[3] J. S. Griffith. *Mathematics of cellular control processes. II. Positive feedback to one gene.* J. Theor. Biol., 20 (1968), 209-216.

[4] A. O. Grillo, M. P. Brown, C. A. Royer. *Probing the physical basis for trp repressor- operator recognition.* J. Mol. Biol., 287 (1999), 539-554.

[5] A. V. Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J. Physiol, 40 (1910) (Suppl.), iv-vii.

[6] J. Lewis. *Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator.* Curr. Biol., 13 (2003), 1398-1408.

[7] J. Lewis, E. M. Özbudak. *Notch signalling synchronizes the zebrafish segmentation clock but is not needed to create somite boundaries.* PLoS Genet., 4 (2008), e15.

[8] S. Oehler, E. R. Eismann, H. Krämer, B. Müller-Hill. *The three operators of lac operon cooperate in repression.* EMBO J., 9 (1990), 973-979.

[9] W. S. Reznikoff. *The lactose operon-controlling elements: A complex paradigm.* Mol. Microbiol., 6 (1992), 2419-2422.

[10] S. I. Rubinow. Introduction to mathematical biology. Dover, New York, 2003.

[11] M. Santillán. *Bistable behaviour in a model of the lac operon in Escherichia coli with variable growth rate.* Biophys. J., in press, 2008.

[12] M. Santillán, M. C. Mackey. *Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon.* Biophys. J., 86 (2004), 1282-1292.

[13] M. Santillán, M. C. Mackey, E. S. Zeron. *Origin of bistability in the lac operon.* Biophys. J., 92 (2007), 3830-3842.

[14] M. Santillán, E. S. Zeron, M. C. Mackey. *Systems biology of the tryptophan operon.* In L. B. Wilson, editor, Mathematical biology research trends, page In Press. Nova Science Pulishers, Hauppauge, NY, 2008.

[15] J. N. Weiss. *The Hill equation revisited: uses and misuses.* FASEB J., 11 (1997), 835841.

[16] E. S. Zeron. *Positive and negative feedback in engineering and biology.* Math. Mod. Nat. Phenom., in press, 2008.