



**HAL**  
open science

# Stochastic modeling of gene activation and applications to cell regulation

G. Malherbe, David Holcman

► **To cite this version:**

G. Malherbe, David Holcman. Stochastic modeling of gene activation and applications to cell regulation. *Journal of Theoretical Biology*, 2011, 271 (1), pp.51. 10.1016/j.jtbi.2010.11.032 . hal-00657588

**HAL Id: hal-00657588**

**<https://hal.science/hal-00657588>**

Submitted on 7 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Author's Accepted Manuscript

Stochastic modeling of gene activation and applications to cell regulation

G. Malherbe, David Holcman

PII: S0022-5193(10)00627-2  
DOI: doi:10.1016/j.jtbi.2010.11.032  
Reference: YJTBI6254

To appear in: *Journal of Theoretical Biology*

Received date: 9 August 2010  
Revised date: 5 November 2010  
Accepted date: 24 November 2010

Cite this article as: G. Malherbe and David Holcman, Stochastic modeling of gene activation and applications to cell regulation, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2010.11.032](https://doi.org/10.1016/j.jtbi.2010.11.032)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

# Stochastic modeling of gene activation and applications to cell regulation

G. Malherbe David Holcman \*

November 5, 2010

## Abstract

Transcription factors (TFs) are key regulators of gene expression. Based on the classical scenario in which the TF search process switches between one-dimensional motion along the DNA molecule and free Brownian motion in the nucleus, we study the arrival time of several TFs to multiple binding sites. In the presence of a TF influx and competitive binding ligands, we derive the probability that a fixed number of target sites are simultaneously bound. We obtain analytic expressions for this probability as a function of the mean number of TFs. When there are multiple binding sites, because this probability is a sigmoidal curve, our analysis shows that a bistable regime is possible, which can be interpreted as a genetic switch, occurring without requiring cooperative binding (change in the binding probability depending on the previous bounds). Finally, we use our model to analyze fly embryo patterning and show that bicoid can induce a sharp hunchback concentration, resulting in the formation of a sharp boundary and stripes. To conclude, we have proposed here a general mechanism that allows cells to read a morphogenetic gradient. Thus activating the transcription of an auto-activated TF can lead to the conversion of a broad gradient of morphogens into a sharp boundary.

**Keywords:** stochastic binding, diffusion of transcription factor, gene activation, morphogenetic gradient, cell differentiation, genetic switch.

---

\*Department of Computational Biology, IBENS, Ecole Normale Supérieure, 46 rue d'Ulm 75005 Paris, France. This research is not supported by any grants.

# 1 Introduction

Transcription factors (TFs) are key regulators that can initiate or inhibit gene activation by binding to specific DNA sites. TFs enter the cell nucleus and search for their specific binding sites on the DNA molecule. In a context of competition for activation and inhibition, the search for target sites should not be too long otherwise, another gene might be activated. This is the case for olfactory gene activation [1], where a single G-coupled receptor out of hundreds is expressed in a single cell, while all other receptors are repressed. Thus cells might use various mechanisms to control gene activation including changing local properties of the DNA molecule (for example by methylation or acetylation) or changing the properties of the TF interaction as what happens when specific molecules bind to TFs and modify their affinity for the DNA molecule.

The mean time to reach a target site is thus a fundamental parameter of gene activation and several biophysical scenarios have been proposed to estimate it. Berg and Von Hippel [2, 3, 4] realized that the search time cannot be computed using a three dimensional random walk only, because the observed search time is indeed shorter. Using the property that the TF scans the DNA base pairs for potential binding sites, they proposed that the search consists of consecutive cycles of three and one dimensional random motion along the DNA molecule. DNA base pairs are not electrically charged and so no long distance interactions are involved, thus, the TF should physically come close to the DNA in order to generate a true interaction. During the one dimensional search, the TF is confined to a neighborhood of the DNA molecule and can detach due to thermal fluctuations and a new 1d-3d cycle resumes until the target is eventually reached. This basic scenario has been confirmed experimentally by single particle tracking experiments [21] and investigated theoretically, by accounting for the local base pair interactions, the mean number to scan the base pairs per cycle, the free diffusion time and the time the TF is bound to the DNA molecule [5, 6, 7, 8, 10].

However, it is still unclear how to go from the TF search time to the mechanism responsible for cell specialization. Cells in a living tissue are imbedded in a matrix of positional information, generated by morphogenetic gradients [11, 12, 13, 14, 15]. A first step consists of the ability of the cell to "read out" the local characteristics of the morphogenetic gradient so that the cell can be labeled and acquire its own identity. To address this question, we investigate the dynamics of binding sites interacting with TFs, generated by

an external steady state gradient. In particular, we are led to compute the mean time for some TFs to bind to many target sites. The number of bound sites can be seen as a digital-converter used by the cell to discriminate two different morphogenetic gradient concentrations depending whether or not all the binding sites are saturated. We further study the effect of competitive ligands that can be generated by the cell to modulate the TF action leading to gene repression.

First, we present our computations for the time for a single TF to bind multiple binding sites and estimate the distribution of the time for the TF to bind one of its binding sites. The mean arrival time is always a sum of two exponentials but, for a large number of free diffusion and DNA binding cycles, the arrival time is single exponentially distributed. We then expand our analysis to multiple TFs with multiple targets. We apply our results to estimate the number of active sites when the cell nucleus experiences a steady state TF influx. Finally, we estimate the steady state probability that a given number of binding sites are occupied. In our model, this probability describes the proportion of time a gene is actively transcribed. We apply our analysis to the initial patterning of the fly embryo by the bicoid (*bcd*) morphogenetic gradient. The *bcd* gradient regulates a number of downstream TFs involved in the gap gene network [16, 17], which determines the position of body sections along the anterior-posterior (A-P) axis in the drosophila embryo. Among these gap genes, hunchback (*hb*) is responsible for thoracic development [16, 17]. *Hb* activation leads to the formation of a sharp boundary and to the formation of stripes. We use our analysis of TF binding to determine the *hb* density induced by *bcd* activation, and we show that this *hb*-*bcd* interaction-modulation is sufficient to generate the transition from a smooth bicoid gradient into a sharp *hb* boundary in the middle of the drosophila embryo. Our approach provides a general scenario at a molecular level of TF interactions that lead to cell specialization.

## 2 Distribution and mean of the search time

We first summarize the properties associated with the TF's search process to its binding site. The TF switches between a free diffusion and random walk along the DNA molecule [2, 5, 9, 10] (Fig. 1):

1. Due to the interaction potential with the DNA backbone [18], the TF can bind unspecifically to the DNA molecule. The strength of the

interaction potential is around  $16kT$  [19, 5], larger than the thermal noise  $\sim kT$ . In this deep well approximation, the random time  $\tau_d$  a TF stays bound is exponentially distributed [20]. Experimental and theoretical estimates for the average time  $\bar{\tau}_d$  are in the range of a few milliseconds [21, 10].

2. A bound TF slides along the DNA molecule during a random time  $\tau_d$  scanning an average number  $\bar{n}(\tau_d)$  of base pairs (bp). The mean number  $\bar{n} = \mathbb{E}_{\tau_d}(\bar{n}(\tau_d))$  of base pairs scanned before detaching is on the order of 100 [21, 10].
3. A TF can detach from the DNA due to thermal noise and diffuse freely in the nucleus until it rebinds to the DNA. When the DNA molecule occupies a small fraction of the nucleus and can be approximated as long rods, the random time  $\tau_f$  a TF spends diffusing in the nucleus is exponentially distributed [10] with an average  $\bar{\tau}_f$ , which is on the order of a few milliseconds [21, 10]. However, for larger density and a complex DNA organization, the distribution time in general is a sum of exponentials and might even be more complicated.

We start with  $n_f$  copies of a TF, alternating independently between periods of free diffusion and random walks along the DNA until one of the  $n_s$  binding sites is found. We further consider an excess competitive ligands that can bind to the TF target sites, preventing the sites to be occupied by TFs. The ligand  $L$  binds to the target site  $S$  according to a first order reaction:



with an association and a dissociation rate  $k_a$  and  $k_d$  respectively. Although the competitive ligands can be TFs themselves and can have 1D/3D switch behavior, because such dynamics is of first order, we simply model them using a kinetics reaction and at equilibrium, for a concentration  $C$  of ligands, the probability that a binding site is not occupied is:

$$P = \frac{1}{1 + C \frac{k_a}{k_d}}. \quad (2)$$

## 2.1 Search time for a single TF

To compute the distribution of the time  $T(1, n_s)$  a TF binds to to one of  $n_s$  possible binding sites, we use the scenario where the TF is initially freely diffusing in the nucleus (Fig. 1).  $T(1, n_s)$  is the sum of the time the TF spends in one and three dimensions, and using the characteristic function of  $T(1, n_s)$ , the probability density function (pdf) is

$$p_T(t) = \frac{d}{dt} Pr\{T(1, n_s) < t\} \quad (3)$$

as follows (see appendix):

$$p_T(t) = \frac{r_2}{r_2 - r_1} \frac{e^{-r_1 t}}{r_1} + \frac{r_1}{r_1 - r_2} \frac{e^{-r_2 t}}{r_2}, \quad (4)$$

where  $r_1$  and  $r_2$  are the two positive roots of  $(1 - x\bar{\tau}_d)(1 - x\bar{\tau}_f) - 1 + p(n_s) = 0$  and  $p(n_s)$  is the probability to find a target during a single one dimensional walk along the DNA. The associated mean binding time is:

$$\bar{T}(1, n_s) = \int_0^{\infty} t p_T(t) dt = \frac{r_2}{r_1(r_2 - r_1)} + \frac{r_1}{r_2(r_1 - r_2)}. \quad (5)$$

In the limit  $p(n_s) \ll 1$ , using the expression for the two roots and approximating the pdf  $p_T$  (eq. 4) by a single exponential for a time  $t$  such that  $\left(\frac{1}{\bar{\tau}_d} + \frac{1}{\bar{\tau}_f}\right) t \gg 1$  (see appendix), we obtain that

$$p_T(t) = \frac{p(n_s)}{\bar{\tau}_d + \bar{\tau}_f} e^{-\frac{p(n_s)}{\bar{\tau}_d + \bar{\tau}_f} t}. \quad (6)$$

Since  $\bar{\tau}_d$  and  $\bar{\tau}_f$  are both on the order of a few ms [10, 21], the single exponential limit is valid for  $t$  larger than a few ms. The mean time  $\bar{T}(1, n_s)$  then reduces to

$$\bar{T}(1, n_s) \approx \frac{\bar{\tau}_d + \bar{\tau}_f}{p(n_s)}. \quad (7)$$

The mean number of free diffusions and DNA bindings before finding the target site is equal to  $\frac{1}{p(n_s)}$ . The limit  $p(n_s) \ll 1$  corresponds to TFs that find their target sites after a large number of cycles.

## 2.2 Search time for multiple TFs

When there are  $n_f$  TFs that can potentially bind to  $n_s$  identical binding sites (a site can bind to a single TF), in the single exponential limit, the time  $T(n_f, n_s)$  for the first TF to bind a site is the minimum of the  $n_f$  exponential laws of mean time  $\bar{T}(1, n_s)$ .  $T(n_f, n_s)$  is then exponentially distributed with mean:

$$\bar{T}(n_f, n_s) = \frac{\bar{T}(1, n_s)}{n_f}. \quad (8)$$

We shall consider  $n_s$  well separated sites (by at least a distance of  $\bar{n}$  base pairs). In this case, the probability to find each site during a DNA binding is  $\frac{\bar{n}}{N_{bp}}$  where  $N_{bp}$  is the total number of base pairs in the genome. Furthermore, in the presence of competitive ligands, there are  $Pn_s$  available binding sites. Thus, the probability of binding to one of the  $n_s$  sites is  $p(n_s) = Pn_s \frac{\bar{n}}{N_{bp}}$  and the mean binding time for  $n_s$  well separated sites with  $p(n_s) \ll 1$  is:

$$\bar{T}(n_f, n_s) \approx \frac{\bar{\tau}_d + \bar{\tau}_f}{n_f p(n_s)} = \frac{(\bar{\tau}_d + \bar{\tau}_f) N_{bp}}{n_f P n_s \bar{n}} \quad (9)$$

$$= \frac{\bar{T}_S}{n_f n_s}, \quad (10)$$

where

$$\bar{T}_S = \frac{(\bar{\tau}_d + \bar{\tau}_f) N_{bp}}{P \bar{n}} \quad (11)$$

is the search time for a single TF with a single target site.

**Remark 1.** Formula 10 describes the combined effect of multiple but well separated binding sites. When the sites are clustered, the mean time to find a target becomes a nonlinear function of the distribution [23, 24, 25, 26, 27] and has been approximated by the Berg-Purcell approximation formula [28]. When there are  $n_s$  binding sites of size  $a$ , located on an ensemble of DNA-molecules on a sphere of radius  $R$ , the mean time  $\tau_d$  in 3d to find a site is:

$$\tau_d \approx \frac{|\Omega|}{D_H} \left( \frac{1}{4\pi R} + \frac{1}{4n_s a} \right). \quad (12)$$

This formula can be improved [29, 30]. Here  $D_H$  is an effective diffusion constant that accounts for the switch between the 1D DNA motion and the



3D diffusion. When the one 1D excursion length is small compared to the 3D diffusion length,

$$D_H \approx \frac{D}{1 + \frac{\tau_d}{\tau_f}}. \quad (13)$$

In the other cases, one has to deal with random jumps.

**Remark 2.** For  $P = 1$  (no competitive ligand),  $\bar{n} \approx 100$  [21, 5, 10] and for a relatively small genome  $N_{bp} = 10^6$ ,  $p(n_s)$  is approximated by:

$$p(n_s) \approx n_s 10^{-4}. \quad (14)$$

Thus  $p(n_s) \ll 1$  is valid as long as the number of binding sites satisfies  $n_s \ll 10^4$ . We conclude that  $p(n_s) \ll 1$  is verified in most cases.

### 3 From a morphogenetic gradient to DNA site activation

When a nucleus is experiencing a steady influx of TFs, we shall now apply our previous results to estimate the number of occupied sites. This TFs steady influx can either be located outside the cell or be steadily produced in the cytoplasm of the cell. In the nucleus, these TFs can be degraded by an enzymatic activity or reach their target sites (Fig. 2). Competitive ligands can bind to the TF targets. Since the number of occupied binding sites controls gene expression, for a given TF influx, we shall now estimate, the mean proportion of time the binding sites are occupied.

#### Activation of a single binding site

We first compute the average occupation ratio  $\mathbb{P}_1$  of a single binding site before considering multiple sites in the following section. To compute  $\mathbb{P}_1$ , we use Bayes' law and sum over the number of TFs in the nucleus:

$$\mathbb{P}_1 = \sum_{n_f=0}^{+\infty} \mathbb{P}(1|n_f)\mathbb{P}(n_f), \quad (15)$$

where  $\mathbb{P}(n_f)$  is the probability of having  $n_f$  TFs in the nucleus and  $\mathbb{P}(1|n_f)$  is the conditional probability that a single binding site is occupied when there

are  $n_f$  TFs. To proceed with the computation of  $\mathbb{P}_1$ , we assume that TFs arrive in the nucleus at a Poissonian rate  $\lambda$  and are degraded (free or bound) by enzymes at a rate  $K$ . Thus, the number of TFs in the nucleus follows a birth and death process and is distributed according to a Poisson law with mean  $\alpha = \frac{\lambda}{K}$ :

$$\mathbb{P}(n_f) = \frac{\alpha^{n_f}}{n_f!} e^{-\alpha} \quad (16)$$

We now compute  $\mathbb{P}(1|n_f)$ . When a TF has found the target, it stays attached for a mean time  $\bar{T}_b$ . We consider that the rate of binding and unbinding to the sites is faster than the rate of TF turn over in the nucleus and that the steady state between binding and unbinding is reached, thus

$$\mathbb{P}(1|n_f) = \frac{\bar{T}_b}{\bar{T}_b + \bar{T}(n_f, 1)} = \frac{\bar{T}_b}{\bar{T}_b + \frac{\bar{T}_s}{n_f}} = \frac{n_f}{n_f + \beta}, \quad (17)$$

where  $\beta = \frac{\bar{T}_s}{\bar{T}_b}$ . Using equations (15), (16) and (17), we get:

$$\mathbb{P}_1 = e^{-\alpha} \sum_{n_f=1}^{\infty} \frac{n_f}{n_f + \beta} \frac{\alpha^{n_f}}{n_f!}. \quad (18)$$

Using  $\frac{\alpha^{n_f}}{n_f + \beta} = \alpha^{-\beta} \int_0^{\alpha} x^{\beta+n_f-1} dx$ , we obtain:

$$\begin{aligned} \mathbb{P}_1 &= e^{-\alpha} \alpha^{-\beta} \sum_{n_f=1}^{+\infty} \int_0^{\alpha} \frac{x^{\beta+n_f-1}}{(n_f-1)!} dx \\ &= e^{-\alpha} \alpha^{-\beta} \int_0^{\alpha} x^{\beta} \left( \sum_{n_f=1}^{+\infty} x^{n_f-1} \frac{1}{(n_f-1)!} \right) dx \\ &= e^{-\alpha} \alpha^{-\beta} \int_0^{\alpha} x^{\beta} e^x dx \\ &= \alpha \int_0^1 u^{\beta} e^{\alpha(u-1)} du, \end{aligned} \quad (19)$$

where  $x = \alpha u$ . We plot in figure 4a the occupation ratio  $\mathbb{P}_1$  as a function of the average number  $\alpha$  of TFs in the nucleus for different values of  $\beta$ . When the competitor ligand concentration  $C$  varies, the occupation ratio is modulated as described in figure 4b. Using Lac I data [21] and in the absence of DNA binding competitor ( $P = 1$ ), the total search time is  $\bar{T}_S = 6$  min [21, 10], while  $\bar{T}_b \approx 70$  min [32] and thus the ratio is  $\beta = \frac{\bar{T}_S}{\bar{T}_b} \approx 1/11$ . We conclude (red curve Fig. 4a) that for low  $\beta$ , the target site can be occupied for a significant proportion of time. In particular, small fluxes of TFs can induce significant modifications on gene expression in a target cell.

### Activation with multiple binding sites

When there are  $n_s$  binding sites, we shall now compute the proportion of time  $\mathbb{P}_k$  that  $k$  sites are occupied. Using Bayes' law, we have:

$$\mathbb{P}_k = \sum_{n_f=0}^{\infty} \mathbb{P}(k|n_f)\mathbb{P}(n_f), \quad (20)$$

where  $\mathbb{P}(n_f)$  is the probability to have  $n_f$  TFs given by expression (16). We now compute  $\mathbb{P}(k|n_f)$  by analyzing a Markov chain [33] which describes the probability  $\mathbb{P}_q(t)$  that  $q$  sites are occupied at time  $t$ .

When  $q$  sites are occupied, the total release rate is  $\frac{q}{\bar{T}_b}$  while the arrival rate to a site is given by  $\bar{T}^{-1}(n_f - q, n_s - q) = \frac{(n_f - q)(n_s - q)}{\bar{T}_S}$  with equation (11). The forward and backward rate of the Markov chain are given by:

$$F_q = \frac{(n_f - q)(n_s - q)}{\bar{T}_S} \quad (21)$$

$$B_q = \frac{q}{\bar{T}_b}, \quad (22)$$

and the Markov chain is given by [33]:

$$\frac{d}{dt}\mathbb{P}(q, t|n_f) = -(F_q + B_q)\mathbb{P}(q, t|n_f) + F_{q-1}\mathbb{P}(q-1, t|n_f) + B_{q+1}\mathbb{P}(q+1, t|n_f), \quad (23)$$

with the boundary conditions:

$$\frac{d}{dt}\mathbb{P}(n_f, t|n_f) = F_{n_f-1}\mathbb{P}(n_f-1, t|n_f) - B_{n_f}\mathbb{P}(n_f, t|n_f) \quad (24)$$

$$\frac{d}{dt}\mathbb{P}(0, t|n_f) = -F_0\mathbb{P}(0, t|n_f) + B_1\mathbb{P}(1, t|n_f). \quad (25)$$

We consider that the rate of binding and unbinding to the sites is faster than the rate of TF turn over in the nucleus and that the steady state is achieved quickly, thus:

$$0 = -(F_q + B_q)\mathbb{P}(q|n_f) + F_{q-1}\mathbb{P}(q-1|n_f) + B_{q+1}\mathbb{P}(q+1|n_f), \quad (26)$$

where  $\mathbb{P}(q|n_f) = \mathbb{P}(q, \infty|n_f)$ . By induction [33], for  $k \leq n' = \min(n_f, n_s)$  (the maximal number of sites occupied by TFs), we get:

$$\mathbb{P}(k|n_f) = \mathbb{P}(0|n_f) \frac{1}{\beta^k k!} \prod_{j=0}^{k-1} (n_f - j)(n_s - j), \quad (27)$$

where:

$$\beta = \frac{\bar{T}_S}{\bar{T}_b}. \quad (28)$$

For  $k > n' = \min(n_f, n_s)$ ,  $\mathbb{P}(k|n_f) = 0$  since there can be no more than  $n'$  TFs bound. Using the normalization condition,

$$\sum_{k=0}^{n'} \mathbb{P}(k|n_f) = 1, \quad (29)$$

we finally get for  $1 \leq k \leq n'$ :

$$\mathbb{P}(k|n_f) = \frac{\frac{1}{\beta^k k!} \prod_{j=0}^{k-1} (n_f - j)(n_s - j)}{1 + \sum_{l=1}^{n'} \frac{1}{\beta^l l!} \prod_{j=0}^{l-1} (n_f - j)(n_s - j)}. \quad (30)$$

and for  $k = 0$ :

$$\mathbb{P}(0|n_f) = \frac{1}{1 + \sum_{l=1}^{n'} \frac{1}{\beta^l l!} \prod_{j=0}^{l-1} (n_f - j)(n_s - j)}. \quad (31)$$

Using expressions (16), (20) and (27), we obtain for  $1 \leq k \leq n_s$ :

$$\mathbb{P}_k = \sum_{n_f=k}^{\infty} \frac{\alpha^{n_f}}{n_f!} e^{-\alpha} \frac{\frac{1}{\beta^k k!} \prod_{j=0}^{k-1} (n_f - j)(n_s - j)}{1 + \sum_{l=1}^{\min(n_f, n_s)} \frac{1}{\beta^l l!} \prod_{j=0}^{l-1} (n_f - j)(n_s - j)}, \quad (32)$$

and for  $k = 0$ :

$$\mathbb{P}_0 = e^{-\alpha} + \sum_{n_f=1}^{\infty} \frac{\alpha^{n_f}}{n_f!} e^{-\alpha} \frac{1}{1 + \sum_{l=1}^{\min(n_f, n_s)} \frac{1}{\beta^l l!} \prod_{j=0}^{l-1} (n_f - j)(n_s - j)}, \quad (33)$$

and  $\mathbb{P}_k = 0$  for  $k > n_s$  as there can not be more than  $n_s$  TFs bound to the target sites. We shall now derive asymptotic expressions for  $\mathbb{P}_k$  when  $\alpha \ll 1$  and  $\beta \ll 1$ , which correspond respectively to a small average number of TFs in the nucleus and to TFs that stay bound to the targets a long time compared to the search time.

### Asymptotics for $\alpha$ small

With the expression of  $\mathbb{P}(n_f)$  given in (16) and (20), only the terms  $n_f = 0, 1$  contribute to the first order asymptotic in  $\alpha \ll 1$ . With (27), we obtain that

$$\mathbb{P}(1|1) = \frac{n_s}{\beta} \mathbb{P}(0|n1) \quad (34)$$

and with  $\mathbb{P}(1|1) + \mathbb{P}(0|1) \approx 1$ ,

$$\begin{aligned} \mathbb{P}(1|1) &= \frac{n_s}{n_s + \beta} \\ \mathbb{P}(0|1) &= 1 - \mathbb{P}(1|1). \end{aligned}$$

With (20), for  $\alpha \ll 1$ ,

$$\mathbb{P}_1 \approx \alpha e^{-\alpha} \frac{n_s}{n_s + \beta} \approx \frac{\alpha n_s}{n_s + \beta}. \quad (35)$$

We conclude that the probability that only one site is occupied is given by the average number of TFs  $\alpha$  multiplied by the probability  $\frac{n_s}{n_s + \beta}$  to have one site occupied when there is a single TF in the nucleus.

### Asymptotic for $\beta$ small

We now compute the asymptotic for  $\beta \ll 1$ .

1. When the number of TFs is larger than the number of available sites ( $n_f \geq n_s$ ), using equation (30), for  $\beta \ll 1$ , to the first order in  $\beta \ll$

1, only the terms  $\mathbb{P}(n_s - 1|n_f)$  and  $\mathbb{P}(n_s|n_f)$  contribute. Using the normalization relation (29),

$$\mathbb{P}(n_s - 1|n_f) + \mathbb{P}(n_s|n_f) \approx 1.$$

Furthermore with (30),

$$\mathbb{P}(n_s - 1|n_f) = \frac{\beta n_s}{n_f - n_s + 1} \mathbb{P}(n_s|n_f),$$

we then obtain:

$$\mathbb{P}(n_s|n_f) \approx 1 - \frac{\beta n_s}{n_f - n_s + 1} \quad (36)$$

$$\mathbb{P}(n_s - 1|n_f) \approx \frac{\beta n_s}{n_f - n_s + 1}. \quad (37)$$

When  $n_f \geq n_s$  and  $\beta \ll 1$  almost all sites are occupied.

2. When there are less TFs than the number of available sites ( $0 < n_f < n_s$ ), then for  $\beta \ll 1$  only  $\mathbb{P}(k = n_f - 1|n_f)$  and  $\mathbb{P}(k = n_f|n_f)$  have a contribution in the leading order of equation (30). We obtain:

$$\mathbb{P}(n_f|n_f) = 1 - \frac{\beta n_f}{n_s - n_f + 1} \quad (38)$$

$$\mathbb{P}(n_f - 1|n_f) = \frac{\beta n_f}{n_s - n_f + 1}. \quad (39)$$

We neglect all other probabilities in the first order for  $\beta \ll 1$ .

Combining equations (16), (20) and the first order approximations in  $\beta$ , the probability  $\mathbb{P}_{n_s}$  that all sites are simultaneously occupied is:

$$\mathbb{P}_{n_s} = e^{-\alpha} \sum_{n_f=n_s}^{\infty} \left( 1 - \frac{\beta n_s}{n_f - n_s + 1} \right) \frac{\alpha^{n_f}}{n_f!}. \quad (40)$$

Using the partial sum:

$$S(x) = \sum_{k=0}^{n_s-1} \frac{x^k}{k!}, \quad (41)$$

and after some computations (see appendix) we can write:

$$\mathbb{P}_{n_s}(\alpha) = 1 - e^{-\alpha}S(\alpha) - \beta n_s e^{-\alpha} \int_0^1 \frac{e^{\alpha u} - S(\alpha u)}{u^{n_s}} du. \quad (42)$$

For  $\beta \ll 1$ ,  $\mathbb{P}_{n_s}$  is an increasing function of  $\alpha$  and a decreasing function of  $\beta$  (see appendix). Increasing the number of TFs  $\alpha$  leads to an increase in the probability that all sites are occupied while increasing  $\beta$  decreases the probability that all sites are occupied.

Similarly using equations (93), (92) and (42) the asymptotic expression of the occupation ratio  $\mathbb{P}_k$  for  $\beta \ll 1$  is given by (see appendix):

$$\mathbb{P}_k = \begin{cases} 1 - e^{-\alpha}S(\alpha) - \beta n_s e^{-\alpha} \int_0^1 \frac{e^{\alpha u} - S(\alpha u)}{u^{n_s}} du & \text{for } k = n_s \\ e^{-\alpha} \frac{\alpha^{n_s-1}}{(n_s-1)!} \left( 1 - \beta \frac{n_s-1}{2} \right) + \beta n_s e^{-\alpha} \int_0^1 \frac{e^{\alpha u} - S(\alpha u)}{u^{n_s}} du & \text{for } k = n_s - 1 \\ e^{-\alpha} \frac{\alpha^k}{k!} \left( 1 + \beta \left( \frac{\alpha}{n_s - k} - \frac{k}{n_s - k + 1} \right) \right) & \text{for } k \leq n_s - 2. \end{cases} \quad (43)$$

We plot in figure 5 (resp. 6)  $\mathbb{P}_k$  as a function of the steady state concentration  $\alpha$  of TF for  $n_s = 2$  (resp.  $n_s = 4$ ). For  $n_s > 1$   $\mathbb{P}_{n_s}(\alpha)$  is a sigmoid function. Indeed, at low concentration, the probability is locally convex starting like  $\alpha^{n_s}$ , while it saturated at high concentration.

## Consequence of the analysis for gene expression stability

We now use our results on the occupation ratios  $\mathbb{P}_k$  to show that at least two binding sites are required to produce a genetic switch. A genetic switch can be produced by an autoregulated gene. Indeed after a threshold concentration is reached, TFs regulate highly its own transcription but not before (Fig. 3). A genetic switch is characterized by two stable values of the transcription concentration [34]: an "on" position where the gene is transcribed and an "off" position where it is not. When the gene is "on", transcription is maintained at high levels through autoregulation. When it is "off", transcription remains at low level and does not turn on without an external signal. Genetic

switches play a central role in cellular differentiation, memory and plasticity [35, 36].

To show that a bistable genetic switch requires at least two binding sites, we first determine the steady state concentration of the TF due the balance of production and degradation. For a total number of  $N_{on}$  occupied binding sites, when the gene transcription occurs at a rate  $r$ , the TFs steady state production  $\lambda$  is the balance between production and degradation

$$\lambda = r\mathbb{P}_{N_{on}} = rf(\alpha), \quad (44)$$

where  $f(\alpha) = \mathbb{P}_{N_{on}}(\alpha)$  is given by formula (43) and it depends on  $N_{on}$  and  $n_s$ . The steady state value  $\alpha = \frac{\lambda}{K}$  satisfies the nonlinear equation:

$$Rf(\alpha) = \alpha, \quad (45)$$

where  $R = \frac{r}{K}$ . Bistability appears when equation (45) has two stable solutions, thus equation (45) must have three roots (two stable and one unstable in between). The number of solutions depends on the parameters  $N_{on}$  and  $R$ : For  $N_{on} = 0$ , as plotted in figures 5 and 6,  $f(\alpha) = \alpha/R$  has only one solution. For  $N_{on} = 1$ ,  $f(\alpha) = \alpha/R$  has one solution for  $R$  small and two solutions for  $R$  large. For  $N_{on} \geq 2$ , using formula (43) and as plotted in figures 5 and 6,  $f$  is a sigmoid type function. For  $R$  sufficiently large, equation (45) has three solutions (Fig. 5) and two of them are stable. A gene following such activation properties is a bistable switch. Conversely, for  $R$  sufficiently small,  $\alpha \approx 0$  is the only stable solution. The critical value of  $R$  can be characterized geometrically, as the point where  $\alpha/R$  is tangent to  $f(\alpha)$ . For this critical value there is a stable point at the origin and a saddle point at the tangent point. To conclude, a bistable switch can be obtained from two binding sites, regulating TF auto-expression when the parameter  $R$  is sufficiently large.

## 4 Formation of the Hunchback boundary by the Bicoid gradient

We shall now apply our analysis to determine the formation of the Hunchback TF (hb) boundary by the Bicoid (bcd) morphogen gradient in the drosophila embryo. The bcd gradient regulates a number of downstream TFs involved in the gap gene network [16, 17], which determine the position of body sections along the anterior-posterior (A-P) axis in the drosophila embryo. Among



these gap genes, hb is responsible for thoracic development [16, 17]. Given a bcd gradient, we propose to determine the spatial distribution of hb. Our analysis shows how a broad bcd gradient can trigger a sharp transition in the hb density in the middle of the embryo. We reproduce the bcd and hb density measured in vivo [16] in figure 9a. To distinguish the values of  $\alpha$  and  $\beta$  for the hb and bcd TFs required in our previous model, we shall use subscript  $_h$  for the hb TF and  $_b$  for the bcd TF. We approximate bcd gradient as exponential [16]:

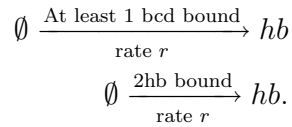
$$\alpha_b(x) = Be^{-kx}, \quad (46)$$

where  $x \in [0, 1]$  is the normalized A-P position ( $x = X/L$  where  $L$  is the length of the drosophila embryo). We use  $k = 5.5$ , corresponding to the best fit for the in vivo data [16]. The constant  $B$  cannot be obtained directly from in vivo data. However, since

$$\alpha_b = e^{-k(x - \frac{\ln(B)}{k})}, \quad (47)$$

changing the value of  $B$  is equivalent to an x-translation of the hb and bcd densities. We choose  $B$  such as the hb boundary is in the middle of the drosophila embryo (Fig. 7b).

hb transcription results from the binding of the hb TF and the bcd TFs to a promoter with 6 bcd binding sites and 2 hb sites [37, 38, 39]. hb transcription is activated by bcd and its auto-activation (Fig. 3). hb is transcribed at a rate  $r$  when there are two hb or at least one bcd bound to the sites, described as:



The hb density, triggered by a bcd concentration, is proportional to the steady state production of hb given by  $\lambda = r(1 - P)$ , where

$$P = \mathbb{P}_{0,b}(1 - \mathbb{P}_{2,h}) \quad (48)$$

is the probability that hb is not transcribed,  $\mathbb{P}_{0,b}$  is the probability that no bcd are bound to the promoter and  $1 - \mathbb{P}_{2,h}$  the probability that there are

not two hb bound. At equilibrium, using  $\alpha_h = \frac{\lambda}{K}$ , we obtain the steady state equation

$$\alpha_h = \frac{\lambda}{K} = R(1 - \mathbb{P}_{0,b}(1 - \mathbb{P}_{2,h})), \quad (49)$$

where  $R = \frac{r}{K}$  and  $K$  is the degradation constant for hb. Equation (49) is implicit for the mean number  $\alpha_h$  of hb, that we shall now compute. We will now evaluate separately expressions  $\mathbb{P}_{0,b}$  and  $\mathbb{P}_{2,h}$ . Along the A-P axis parameterized by the position  $x$ ,  $\mathbb{P}_{0,b}$  depends on the mean number  $\alpha_b(x)$  of bcd TFs and on the ratio  $\beta_b = \frac{\bar{T}_s}{\bar{T}_b}$  of the search time of bcd over the binding time. To evaluate  $\beta_b$ , we use the binding reaction of a bcd to its target site  $S$ :



where  $S.bcd$  is the bcd TF bound to its target site. The equilibrium constant  $K_d = \frac{[S.bcd]}{[S][bcd]}$  is the ratio of the forward to the backward rate of 50, equivalently:

$$K_d = \frac{\bar{T}_s}{\bar{T}_b N_a V} = \frac{\beta_b}{N_a V}. \quad (51)$$

For  $K_d = 0.24nM$  [40], a nucleus of volume  $V \approx 1\mu m^3$  and with  $N_a$  the Avogadro number we obtain  $\beta_b = K_d N_a V \approx 0.14$ .

To compute  $\mathbb{P}_{0,b}$ , we use formula (43) with  $k = 0$ ,  $n_s = 6$  and obtain:

$$\mathbb{P}_{0,b} = e^{-\alpha_b} \left( 1 + \frac{\beta_b \alpha_b}{6} \right). \quad (52)$$

We shall now evaluate the probability  $1 - \mathbb{P}_{2,h}$ . In the absence of any precise data on the dissociation constant of hb from its binding site, we consider that binding is fast enough so that  $\beta_h \approx 0$ . Using expression (43) for the probability  $\mathbb{P}_{2,h}$  with  $k = n_s = 2$ , we obtain

$$1 - \mathbb{P}_{2,h} = e^{-\alpha_h} + \alpha_h e^{-\alpha_h}. \quad (53)$$

Finally, at steady state, the equilibrium condition (49) reads:

$$R \left( 1 - e^{-\alpha_b} \left( 1 + \frac{\beta_b \alpha_b}{6} \right) (e^{-\alpha_h} + \alpha_h e^{-\alpha_h}) \right) = \alpha_h. \quad (54)$$

We solve equation (54) numerically (with Maple) to express  $\alpha_h$  as a function of  $\alpha_b$ . We plot in figure 7a-b several solutions associated with different values of  $R$  and  $B$ . As pointed out in equation (46) and plotted in figure 7a, changing the value of  $B$  is equivalent to a x-translation of the hb and bcd densities. To further study the different types of solutions, we will vary the parameter  $R$ . Following the discussion in section 3 on bistability, for  $N_{on} = n_s = 2$  the dynamics for hb can potentially be bistable. We show now that for  $n_s = 2$  and  $R < 3$ , hb is always monostable. To compute the critical value  $R_c$  after which bistability occurs, we shall use the functions:

$$P(x) = 1 - \mathbb{P}_{0,b} \quad (55)$$

$$f(\alpha_h) = \mathbb{P}_{2,h}, \quad (56)$$

where  $P(x)$  depends on  $x$  through  $\alpha_b$  (46). The function  $f(\alpha_h)$  is the fraction of time hb is autoactivated by the hb and  $P(x)$  is the fraction of time the gene is activated by the bcd gradient. Equation (54) can then be rewritten as:

$$R(1 - (1 - P(x))(1 - f(\alpha_h))) = \alpha_h. \quad (57)$$

We determine in the appendix the critical value for bistability given by  $R_c = 3$ . For  $R < R_c$  the gene is always monostable, while for  $R > R_c$  the gene is bistable for some values of  $P(x)$  and monostable for others:

- For  $R > 3$ , hb is monostable for  $x < x_c$  and bistable for  $x > x_c$  where  $x_c$  is a critical position. We represent the bifurcation diagram in figure 7d. Changing  $B$  is equivalent to an x-translation in the hb profile and thus  $B$  can be adjusted such as the bifurcation point is  $x_c = 0.5$  for example. If at time  $t = 0$  there is no hb, the hb density converges to the lower stable value, as represented in figure 7b. Nevertheless, for a bistable hb dynamic, cells located in  $x > x_c$  can switch from low to high (stable value) in response to a sufficient signal. In the absence of a repressor of hb on the posterior side of the embryo, the gene stays in the high stable state.
- For  $R < 3$ , hb is always monostable. When  $R$  becomes close to 3, there is already a boundary in the hb density (Fig. 7b). This boundary can be characterized by the point where  $f(\alpha_h)$  changes concavity and becomes tangent to a linear function (Fig. 8). At the point of concavity change, a small variation in  $P(x)$  induces a large variation in  $\alpha_h$  which produces a sharp transition in the hb density.

We conclude that for an auto-regulated hb, when the bifurcation parameter  $R$  is close but smaller than the critical value  $R_c$ , there is a sharp boundary of hb in the embryo and this boundary does not require a repressor in the posterior half of the embryo. As shown in figure 8, at the boundary hb synthesis is essentially due to autoactivation of hb (the activation  $P(x)$  due to bcd is  $\approx 10\%$  whereas the gene is autoactivated  $\approx 40\%$  of the time). To obtain a numerical estimation of  $R$ , we use the synthesis rate  $r$  generated by two hb bound to the target sites and the degradation rate  $K$  of hb. Using the values from the supplementary material of [37],  $r \approx 19$  and  $K \approx 7.08$  and we obtain

$$R \approx 2.7. \quad (58)$$

For  $R = 2.7$ , we observe a steep transition of the hb density at the middle of the embryo as in the in vivo data from [16] reproduced in figure 9a. The main difference between the theoretical density (Fig. 7b) and the in vivo data from [16] (Fig. 9a) is in the anterior edge where our model leads to an increase of the hb density instead of a decay as observed in vivo. This decay in the hb density at the anterior edge of the embryo is due to a repressive effect induced by the huckebein TF (hkb) [17] which we did not model in (54) and we shall examine now.

### Refining the gradient using hkb repressor

We now account for the repression induced by hkb and consider that the transcription of the hb gene is repressed when at least one hkb is bound to the promoter site (Fig. 3). hkb promoter sites are different from the ones of hb and bcd and thus hkb dynamics is independent of the two other TFs. Similarly to the analysis that lead us to equation (49), we obtain:

$$\alpha_h = R \mathbb{P}_{0,hkb} (1 - \mathbb{P}_{0,b} (1 - \mathbb{P}_{2,h})). \quad (59)$$

where  $\mathbb{P}_{0,hkb}$  is the probability that no hkb are bound. We assume hkb binds to its target fast enough and shall consider that  $\beta_{hkb} = 0$ . Finally,  $\mathbb{P}_{0,hkb}$  is then given by:

$$\mathbb{P}_{0,hkb} = e^{-\alpha_{hkb}}. \quad (60)$$

To evaluate the distribution  $\alpha_{hkb}$  we fit the measured hkb distribution [17] with an exponential function:

$$\alpha_{hkb} = C e^{k_{hkb} x}, \quad (61)$$

where  $k_{hkb} = 11.3$  (Fig. 9c). The value of  $C$  can not be obtained directly from experimental measurements. Changing the value of  $C$  is equivalent to an x-translation of the repression due to hkb. We calibrated  $C$  to have the same value for the hb density as in the vivo data (Fig. 9a and d)). We solve equation (59) numerically and obtain an hb density represented in figure 9b. This new theoretical density obtained is now close to the in vivo data (Fig. 9a), in particular we recover the sharp boundary of hb. The main differences between the theoretical and experimental densities are located at the posterior side of the hb boundary where we obtain a higher density compared to the vivo data and at the posterior edge where the density is lower. The difference after the hb boundary might be due to repression of hb by the knirps TF [17] which is not modelled here. As for the difference at the posterior edge, this can be due to activation of hb by the Caudal TF [17]. Taking into account these two regulation pathways should lead to a refined analysis of the hb density.

## 5 Conclusion

In this paper, we studied transcription activation by TFs starting from the stochastic nature of the search process for a DNA promoter site. Later on, we applied our computations to estimate the sharp boundary induced by a smooth gradient of TFs. In the first part, we focus on the kinetics of the binding of TFs to their target sites located on the DNA molecule: when the average number of cycles of free diffusions and DNA bindings before finding the target sites is large, the search time  $T(n_f, n_s)$  is exponentially distributed and we estimate the mean (relation 10). Next, we considered the case of a cell receiving a steady state TF concentration, and each TF can be enzymatically degraded. We model the dynamics of the TFs' binding and unbinding their target sites and we estimated the fraction of time  $\mathbb{P}_k$  that  $k$  out of  $n_s$  sites are occupied at steady state. For  $n_s = 1$  we obtain an explicit expression in equation (19). For  $n_s \geq 1$ , the general expression of  $\mathbb{P}_k$  is given by an implicit equation (32) and an asymptotic development for  $\beta \ll 1$  is provided in (43). We presented the different occupation ratios in figure 5 and 6 for two and four sites respectively. We consider that the transcription rate is proportional to the fraction of time a given number of sites is occupied. For a defined TF concentration entering the nucleus, our model provides a quantitative input-output relation in terms of the transcription rate.

we considered here two types of transcriptional repressors: 1) Repression by ligands that competitively bind to the TF binding sites (equation 1). Ligands were modeled as molecules in excess compared to the TF and we use an equilibrium approximation. 2) in section (4), we studied repression by *hkb*, where *hkb*, *hb* and *bcd* binding sites were different and hence bindings was independent. It would be interesting to expand our analysis to the case where two Tfs competitively bind to the same target sites, this situation remind us of the  $\lambda$ -Cro regulation [34].

When we apply our model to the regulation of hunchback by the bicoid morphogenic gradient, we focus on the sharp boundary in the *hb* density at the middle of the embryo. Several mechanisms accounting for the formation of sharp boundaries have been proposed: Some mechanisms [40, 43, 44, 45] result from cooperative binding while others include a bistable gene [37] or the antagonistic action of a repressor and activator gradient [46, 47, 48, 49]. Here, we use neither the repression of *hb* in the posterior half nor the cooperative binding of *bcd*, but we show that, in the absence of these two mechanisms, a smooth morphogenetic gradient can trigger a sharp boundary for an autoregulated gene. We also show that bistability of the autoregulated gene is not a requirement and that sharp boundaries can be generated by monostable autoregulated genes. We found the critical value for the transcription rate at which a bifurcation occurs and gave an estimate in equation (103). We further show that a bistable gene can produce a sharp boundary from a smooth gradient. Nevertheless, for a bistable *hb*, cells located on the posterior side of the embryo can switch from a low stable value to a high one in response to a sufficiently large perturbation. A repressor on the right hand side of the boundary would then be required to obtain a reliable boundary position. Our results show that an autoregulated gene close to bistability is sufficient to produce a sharp boundary.

Here, we focused on a minimal mechanism that allows a morphogenetic gradient to trigger a sharp boundary in an autoregulated gene. In order to focus on this minimal system that produces sharp boundaries, neither *hb*-repression in the posterior half nor cooperative binding of *bcd* are modeled. Both repression [46] and cooperative binding [43] are already known to play a key role in the formation of the sharp boundary of hunchback and it would thus be interesting to expand our model to take them into account. With autoregulation, it would then be interesting to see how these three mechanisms, which appear to be redundant, produce sharp and robust boundaries in the embryo.

## 6 Appendix

### 6.1 The pdf of $T(1, n_s)$

We compute here the pdf  $P_T(t)$  of the time  $T(1, n_s)$  a single TF takes to bind one of the  $n_s$  DNA specific targets. Decomposing the pdf by the event that the target is found after exactly  $k$  steps, we have:

$$P_T(t) = \sum_{k=0}^{\infty} Pr\{T(1, n_s) < t | k \text{ 1D walk}\} Pr\{k \text{ 1D walk}\}. \quad (62)$$

Using the probability  $p(n_s)$  to bind to one of the  $n_s$  sites during a one dimensional motion along the DNA molecule, the probability  $\tilde{P}_k = Pr\{k \text{ 1D walk}\}$  to find a site during the  $k^{th}$  one dimensional DNA motion is given by:

$$\tilde{P}_k = p(n_s)(1 - p(n_s))^{k-1}. \quad (63)$$

A cycle is the concatenation of one and three dimensional motions. Both periods are characterized by random exponential times. The conditional search time for  $k$  cycles of DNA binding and free diffusion is then:

$$T(1, n_s)\{\text{conditioned on } k \text{ cycle}\} = \sum_{j=1}^k (\tau_f(j) + \tau_d(j)), \quad (64)$$

where  $(\tau_d(1), \dots, \tau_d(k))$  and  $(\tau_f(1), \dots, \tau_f(2), \dots, \tau_f(k))$  are respective the times spent bound to the DNA and freely diffusing in the nucleus.

To compute  $P_T(t)$ , we will use the characteristic function  $F$  of  $T(1, n_s)$ ,

$$F(x) = \mathbb{E}_t(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} p_T(t) dt \quad (65)$$

$$= \sum_{k=1}^{\infty} G_k(x) \tilde{P}_k, \quad (66)$$

where  $G_k$  is the characteristic function of:

$$T\{(1, n_s) | k \text{ 1D walk}\} = \sum_{j=1}^k (\tau_f(j) + \tau_d(j)), \quad (67)$$

which is given in equation (64) and  $\tilde{P}_k = p(n_s)(1-p(n_s))^{k-1}$  is the probability to have  $k$  1D walks given in equation (63). Since the random times  $\tau_d(j)$  and  $\tau_f(j)$  are independent, the characteristic function of the sum (67) is the product of the characteristic functions:

$$G_k(x) = \prod_{j=1}^k F_{\tau_f(j)}(x) F_{\tau_d(j)}(x), \quad (68)$$

where  $F_{\tau_f(j)}(x)$  and  $F_{\tau_d(j)}(x)$  are respectively the characteristic functions of the free diffusion time  $\tau_f(j)$  and the time  $\tau_d(j)$  bound to the DNA. Since these times are exponentially distributed:

$$F_{\tau_f(j)}(x) = \frac{1}{1 - ix\bar{\tau}_f} \quad (69)$$

$$F_{\tau_d(j)}(x) = \frac{1}{1 - ix\bar{\tau}_d}. \quad (70)$$

Finally,

$$F(x) = \sum_{k=1}^{\infty} p(n_s) (1 - p(n_s))^{k-1} \frac{1}{(1 - ix\bar{\tau}_f)^k (1 - ix\bar{\tau}_d)^k} \quad (71)$$

$$= \frac{p(n_s)}{(1 - ix\bar{\tau}_d)(1 - ix\bar{\tau}_f) - 1 + p(n_s)}. \quad (72)$$

The poles are given by the two roots of  $(1 - y\bar{\tau}_d)(1 - y\bar{\tau}_f) - 1 + p(n_s) = 0$  with  $y = ix$ :

$$r_1 = \frac{(\bar{\tau}_d + \bar{\tau}_f) - \sqrt{(\bar{\tau}_d + \bar{\tau}_f)^2 - 4p(n_s)\bar{\tau}_f\bar{\tau}_d}}{2\bar{\tau}_f\bar{\tau}_d} > 0 \quad (73)$$

$$r_2 = \frac{(\bar{\tau}_d + \bar{\tau}_f) + \sqrt{(\bar{\tau}_d + \bar{\tau}_f)^2 - 4p(n_s)\bar{\tau}_f\bar{\tau}_d}}{2\bar{\tau}_f\bar{\tau}_d} > 0,$$

where, for  $p(n_s) \in [0; 1]$ , the two roots  $r_1$  and  $r_2$  are real positive. Decomposing the fraction (72) gives:

$$F(x) = \frac{p(n_s)}{\bar{\tau}_d\bar{\tau}_f(r_1 - r_2)(ix - r_1)} - \frac{p(n_s)}{\bar{\tau}_d\bar{\tau}_f(r_1 - r_2)(ix - r_2)} \quad (74)$$

$$= \frac{r_1 r_2}{(r_1 - r_2)(ix - r_1)} - \frac{r_1 r_2}{(r_1 - r_2)(ix - r_2)}, \quad (75)$$



where  $\frac{p(n_s)}{\bar{\tau}_d \bar{\tau}_f} = r_1 r_2$  comes from the equation satisfied by  $r_1$  and  $r_2$ . By inverting the characteristic function  $p_T(t) = \frac{1}{2\pi} \mathbb{E}_x(e^{-itx}) = \int_{-\infty}^{\infty} e^{itx} F(x) dx$  and since the inverse transform of  $-\frac{r_1}{ix-r_1}$  is an exponential distribution of mean  $\frac{1}{r_1}$ , we obtain:

$$p_T(t) = \frac{r_2}{r_2 - r_1} \frac{e^{-tr_1}}{r_1} + \frac{r_1}{r_1 - r_2} \frac{e^{-tr_2}}{r_2}. \quad (76)$$

We conclude that the distribution  $p_T$  is the sum of two decreasing exponentials.

## 6.2 Asymptotic pdf of $T(1, n_s)$ for $p(n_s) \ll 1$

We shall now study the approximation  $p(n_s) \ll 1$ , for which:

$$r_1 \approx \frac{p(n_s)}{\bar{\tau}_d + \bar{\tau}_f} \quad (77)$$

$$r_2 \approx \frac{1}{\bar{\tau}_d} + \frac{1}{\bar{\tau}_f}, \quad (78)$$

and

$$p_T(t) \approx (1 - \varepsilon) \frac{p(n_s)}{\bar{\tau}_d + \bar{\tau}_f} e^{-t \frac{p(n_s)}{\bar{\tau}_d + \bar{\tau}_f}} + \varepsilon \left( \frac{1}{\bar{\tau}_d} + \frac{1}{\bar{\tau}_f} \right) e^{-t \left( \frac{1}{\bar{\tau}_d} + \frac{1}{\bar{\tau}_f} \right)}, \quad (79)$$

with  $\varepsilon = p(n_s) \frac{1}{(\bar{\tau}_d + \bar{\tau}_f) \left( \frac{1}{\bar{\tau}_d} + \frac{1}{\bar{\tau}_f} \right)} \leq \frac{p(n_s)}{4}$ . Since  $p(n_s) \ll 1$  the second exponential converges faster to 0 than the first and is further multiplied by a small coefficient  $\varepsilon$ .

For a time  $\left( \frac{1}{\bar{\tau}_d} + \frac{1}{\bar{\tau}_f} \right) t \gg 1$ , we approximate the pdf  $p_T$  given in equation (76) by a single exponential:

$$p_T(t) = \frac{p(n_s)}{\bar{\tau}_d + \bar{\tau}_f} e^{-\frac{p(n_s)}{\bar{\tau}_d + \bar{\tau}_f} t}. \quad (80)$$

Since  $\bar{\tau}_d$  and  $\bar{\tau}_f$  are both on the order of a few ms [10, 21], the single exponential limit is valid for  $t$  larger than a few ms. The mean time  $\bar{T}(1, n_s)$  then reduces to:

$$\bar{T}(1, n_s) \approx \frac{\bar{\tau}_d + \bar{\tau}_f}{p(n_s)}. \quad (81)$$

### 6.3 Computation of $\mathbb{P}_k$ for $\beta \ll 1$

Combining equations (16), (20) and the first order approximations in  $\beta$ , the probability  $\mathbb{P}_{n_s}$  that all sites are simultaneously occupied is:

$$\mathbb{P}_{n_s} = e^{-\alpha} \sum_{n_f=n_s}^{\infty} \left(1 - \frac{\beta n_s}{n_f - n_s + 1}\right) \frac{\alpha^{n_f}}{n_f!}. \quad (82)$$

Using the partial sum:

$$S(x) = \sum_{k=0}^{n_s-1} \frac{x^k}{k!}, \quad (83)$$

and the relations:

$$\sum_{n_f=n_s}^{\infty} \frac{\alpha^{n_f}}{n_f!} = e^{\alpha} - S(\alpha) \quad (84)$$

$$\sum_{n_f=n_s}^{\infty} \frac{\alpha^{n_f}}{(n_f - n_s + 1)n_f!} = \alpha^{n_s-1} \int_0^{\alpha} \frac{e^x - S(x)}{x^{n_s}} dx, \quad (85)$$

we obtain:

$$\mathbb{P}_{n_s} = 1 - e^{-\alpha} S(\alpha) - \beta n_s e^{-\alpha} \alpha^{n_s-1} \int_0^{\alpha} \frac{e^x - S(x)}{x^{n_s}} dx. \quad (86)$$

Using the change of variable  $x = \alpha u$ , we can write:

$$\mathbb{P}_{n_s}(\alpha) = 1 - e^{-\alpha} S(\alpha) - \beta n_s e^{-\alpha} \int_0^1 \frac{e^{\alpha u} - S(\alpha u)}{u^{n_s}} du. \quad (87)$$

We shall now examine some properties of  $\mathbb{P}_{n_s}$ . For  $\alpha u \geq 0$ ,  $e^{\alpha u} - S(\alpha u) \geq 0$ , thus  $\mathbb{P}_{n_s}$  is a decreasing function of  $\beta$ . Indeed the partial derivative of  $\mathbb{P}_{n_s}$  in  $\beta$  is negative. Moreover,  $\mathbb{P}_{n_s}$  is an increasing function of  $\alpha$  for  $\beta \ll 1$ : starting from expressions 82 (which is equal to 87) and differentiating with respect to  $\alpha$ :

$$\frac{\partial \mathbb{P}_{n_s}}{\partial \alpha} = e^{-\alpha} \sum_{n_f=n_s}^{\infty} \left(1 - \frac{\beta n_s}{n_f - n_s + 1}\right) \left(\frac{\alpha^{n_f-1}}{(n_f-1)!} - \frac{\alpha^{n_f}}{n_f!}\right). \quad (88)$$

Using  $n_f \geq n_s$  and  $\varepsilon$  sufficiently small, then for  $\beta < \frac{1}{n_s}(1-\varepsilon)$ ,  $\left(1 - \frac{\beta n_s}{n_f - n_s + 1}\right) > \varepsilon$  and we obtain:

$$\frac{\partial \mathbb{P}_{n_s}}{\partial \alpha} > \varepsilon e^{-\alpha} \frac{\alpha^{n_s-1}}{(n_s-1)!} > 0. \quad (89)$$

and  $\partial \mathbb{P}_{n_s}$  is an increasing function of  $\alpha$ .

We now proceed with estimating  $\mathbb{P}_{n_s-1}$ . Using equations (16), (20), we obtain:

$$\mathbb{P}_{n_s-1} = e^{-\alpha} \sum_{n_f=n_s-1}^{\infty} \mathbb{P}(k = n_s - 1 | n_f) \frac{\alpha^{n_f}}{n_f!}. \quad (90)$$

For  $\beta \ll 1$ , using approximation (39) for the term in  $n_s - 1$  and (37) for the other terms:

$$\mathbb{P}_{n_s-1} = e^{-\alpha} \frac{\alpha^{n_s-1}}{(n_s-1)!} \left(1 - \beta \frac{n_s-1}{2}\right) + e^{-\alpha} \sum_{n_f=n_s}^{\infty} \frac{\beta n_s}{n_f - n_s + 1} \frac{\alpha^{n_f}}{n_f!}. \quad (91)$$

Using relation (87), for  $\beta \ll 1$ , we obtain:

$$\mathbb{P}_{n_s-1} = e^{-\alpha} \frac{\alpha^{n_s-1}}{(n_s-1)!} \left(1 - \beta \frac{n_s-1}{2}\right) + \beta n_s e^{-\alpha} \int_0^1 \frac{e^{\alpha u} - S(\alpha u)}{u^{n_s}} du. \quad (92)$$

Finally, when  $k \leq n_s - 2$  sites are occupied, using the first order approximations for  $\mathbb{P}(k|n_f)$  in formula (39), we shall only retain the probabilities associated with  $k$  or  $k+1$  TFs in the nucleus,

$$\begin{aligned} \mathbb{P}_k &\approx \mathbb{P}(k|k)\mathbb{P}(k) + \mathbb{P}(k|k+1)\mathbb{P}(k+1) \\ &= \left(1 - \frac{k\beta}{(n_s-k+1)}\right) e^{-\alpha} \frac{\alpha^k}{k!} + \frac{(k+1)\beta}{(n_s-k)} e^{-\alpha} \frac{\alpha^{k+1}}{(k+1)!} \\ &= e^{-\alpha} \frac{\alpha^k}{k!} \left(1 + \beta \left(\frac{\alpha}{n_s-k} - \frac{k}{n_s-k+1}\right)\right). \end{aligned} \quad (93)$$

## 6.4 Critical value for bistability

To compute the critical value  $R_c$  for which the profile  $\alpha_h(x)$  can be bistable, we use equation:

$$\alpha_h = R(1 - (1 - P(x))(1 - f(\alpha_h))). \quad (94)$$

For the critical value  $R_c$ , the function:

$$\alpha \rightarrow R_c(1 - (1 - P(x))(1 - f(\alpha))), \quad (95)$$

is tangent to  $\alpha \rightarrow \alpha$  in  $\alpha_c$  for some value of  $P(x)$ , where  $\alpha_c$  is the point where  $f$  changes concavity (Fig. 8). For  $k = n_s > 1$  and  $\beta = 0$ ,

$$f(\alpha) = \mathbb{P}_{n_s} = 1 - e^{-\alpha}S(\alpha). \quad (96)$$

$f''(\alpha_c) = 0$  is equivalent to  $S(\alpha_c) - 2S'(\alpha_c) + S''(\alpha_c) = 0$ , where:

$$S(\alpha) = \sum_{k=0}^{n_s-1} \frac{\alpha^k}{k!}. \quad (97)$$

After some computations, we find that:

$$\alpha_c = n_s - 1. \quad (98)$$

Now at the critical value  $R_c$ , the function (95) is tangent to  $\alpha$  in  $\alpha_c$  (Fig. 8) and we obtain the conditions:

$$R_c(1 - (1 - P(x))(1 - f(\alpha_c))) = \alpha_c \quad (99)$$

$$R_c \left( 1 - (1 - P(x)) \left( 1 - \frac{\partial f}{\partial \alpha}(\alpha_c) \right) \right) = 1. \quad (100)$$

After simplification,

$$1 - \frac{1 - f(\alpha_c)}{R_c \frac{\partial f}{\partial \alpha}(\alpha_c)} = \alpha_c / R_c. \quad (101)$$

We then obtain for  $R_c$ :

$$R_c = \alpha_c + \frac{1 - f(\alpha_c)}{\frac{\partial f}{\partial \alpha}(\alpha_c)} = \alpha_c + \frac{S(\alpha_c)}{S(\alpha_c) - S'(\alpha_c)}. \quad (102)$$

Finally, using (98) we obtain:

$$R_c = n_s - 1 + (n_s - 1)! \frac{S(n_s - 1)}{(n_s - 1)^{n_s - 1}}. \quad (103)$$

and for  $n_s = 2$ ,  $R_c = 3$ .

**Acknowledgment:** we thank K. Moya for critical reading of the manuscript.

## References

- [1] S. Lomvardas, G. Barnea, D. J. Pisapia, M. Mendelsohn, J. Kirkland, R. Axel, Interchromosomal interactions and olfactory receptor choice, *Cell*. 126(2): 403-13 (2006)
- [2] O. G. Berg, P. H. von Hippel, Selection of DNA binding sites by regulatory proteins, *Trends in Biochemical Sciences*. 13(6): 207-211 (1988)
- [3] O. G. Berg, R. B. Winter, P. H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids. Part 1: Models and theory, *Biochemistry*. 20(24): 6929-6948 (1981)
- [4] O. G. Berg, C. Blomberg, Association kinetics with coupled diffusion. An extension to coiled-chain macromolecules applied to the lac repressor-operator system, *Biophys Chem*. 7(1): 33-9 (1977)
- [5] M. Slutsky, L. A. Mirny, Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential, *Biophys. Journal*. 87(6): 4021-35 (2004)
- [6] Z. Wunderlich, L. A. Mirny, Spatial effects on the speed and reliability of protein-DNA search, *Nucleic Acids Res*. 36(11): 3570-8 (2008)
- [7] S. E. Halford, J. F. Mark, How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Research*. 32(10): 3040-3052 (2004)
- [8] T. Hu, A. Grosberg, B. Shklovskii, How proteins search for their specific sites on DNA: the role of DNA conformation, *Biophys J*. 90(8): 2731-44 (2006)
- [9] M. Coppey, O. Bénichou, R. Voituriez, M. Moreau, Kinetics of Target Site Localization of a Protein on DNA: A Stochastic Approach, *Biophys. Journal*. 87(3): 1640-9 (2004)
- [10] G. Malherbe, D. Holcman, Search for a DNA target site in the nucleus, *PLA*. 374(3): 466-471 (2010)

- [11] A. M. Turing, The Chemical Basis of Morphogenesis, *Phi. Tr. of the Royal Society of London. Series B, Biological Sciences*, Vol. 237, No. 641, pp.37-72.(1952)
- [12] L. Wolpert, One hundred years of positional information, *Trends Genet.* 12(9): 359-64 (1996)
- [13] H. Meinhardt, Models for the generation and interpretation of gradients. *Cold Spring Harb Perspect Biol.* 1(4):a001362. (2009)
- [14] N. A. Monk, Cell communities and robustness in development, *Bull Math Biol.* 59(6):1183-9 (1997)
- [15] D. J. Irons, A. Wojcinski, B. Glise, N. A. Monk NA, Robustness of positional specification by the Hedgehog morphogen gradient, *Dev Biol.* 342(2):180-93 (2010)
- [16] F. He, Y. Wen, J. Deng, X. Lin, L. J. Lu, R. Jiao, J. Ma, Probing intrinsic properties of a robust morphogen gradient in *Drosophila*, *Dev Cell.* 15(4): 558567 (2008)
- [17] M. Ashyraliyev, K. Siggins, H. Janssens, J. Blom, M. Akam, J. Jaeger, Gene circuit analysis of the terminal gap gene *huckebein*, *PLoS Comput Biol.* 5(10):e1000548 (2009)
- [18] K. Nadassy, S. J. Wodak, J. Janin, Structural features of protein-nucleic acid recognition sites, *Biochemistry.* 38(7): 1999-2017 (1999)
- [19] R. B. OGorman, M. Dunaway, K. S. Matthews, DNA Binding Characteristics of Lactose Repressor and the Trypsinresistant Core Repressor, *Journal Bio. Chem.* 255(21): 10100-6 (1980)
- [20] Z. Schuss, *Theory and applications of stochastic differential equations.* John Wiley Sons Inc, (1980)
- [21] J. Elf, G. W. Li, X. S. Xie, Probing transcription factor dynamics at the single-molecule level in a living cell, *Science.* 316(5828): 1191-4 (2007)
- [22] Z. Schuss, A. Singer, D. Holcman, The narrow escape problem for diffusion in cellular microdomains, *PNAS.* 104(41): 16098-103 (2007)

- [23] D. Holcman, Schuss Z, Diffusion through a cluster of small windows and flux regulation in microdomains, *Physics Letters A.* 372(21): 3768-3772 (2008)
- [24] D. Holcman, Schuss Z, Diffusion escape through a cluster of small absorbing windows, *Journal of Physics A.* 41: 155001 (2008)
- [25] S. Pillay, M. Ward, A. Peirce, T. Kolokolnikov, An Asymptotic Analysis of the Mean First Passage Time for Narrow Escape Problems: Part I: Two-Dimensional Domains, *SIAM Multiscale Modeling and Simulation.* 8(3): 803-835 (2010)
- [26] A. Cheviakov, M. Ward, R. Straube, An Asymptotic Analysis of the Mean First Passage Time for Narrow Escape Problems: Part II: The Sphere, *SIAM Multiscale Modeling and Simulation*, 8(3): 836-870 (2010)
- [27] A. Cheviakov, M. Ward, Optimizing the Principal Eigenvalue of the Laplacian in a Sphere with Interior Traps, *Mathematical and Computer Modeling*, In press, (2010)
- [28] H. C. Berg, E.M. Purcell, Physics of chemoreception, *Biophys J.* 20(2):193-219 (1977)
- [29] R. Zwanzig, Diffusion-controlled ligand binding to spheres partially covered by receptors: an effective medium treatment, *Proc. Nat. Acad. Sci.* 87(15):5856-5857 (1990).
- [30] A. M. Berezhkovskii, Y. A. Makhnovskii, M. I. Monine, V. Y. Zitserman, S. Y. Shvartsman, Boundary homogenization for trapping by patchy surfaces, *J Chem Phys.* 121(22):11390-4 (2004)
- [31] R. Baliga, E. E. Baird, D. M. Herman, C. Melander, P. B. Dervan, D. M. Crothers, Kinetic Consequences of Covalent Linkage of DNA Binding Polyamides, *Biochemistry*, 40(1):3-8 (2001)
- [32] S. Y. Lin, A. D. Riggs, Lac operator analogues: bromodeoxyuridine substitution in the lac operator affects the rate of dissociation of the lac repressor, *PNAS.* 69(9): 2574-76 (1972)
- [33] D. Holcman, Z. Schuss, Stochastic chemical reactions in microdomains, *Journal of Chemical Physics.* 122(11): 114710 (2005)

- [34] M. Ptashne, *A genetic switch: phage lambda revisited*. Cold Spring Harbor Laboratory Press (2004)
- [35] L. Wang, B. L. Walker, S. Iannaccone, D. Bhatt, P. J. Kennedy, W. T. Tse, Bistable switches control memory and plasticity in cellular differentiation, *PNAS*. 106(16): 6638-6643 (2009)
- [36] S. T. Crews, J. C. Pearson, Transcriptional autoregulation in development, *Curr Biol*. 19(6): 241-246 (2009)
- [37] J. P. Lopes, F. M. Vieira, D. M. Holloway, P. M. Bisch, A. V. Spirov, Spatial Bistability Generates hunchback Expression Sharpness in the *Drosophila* Embryo, *PLoS Comput Biol*. 4(9):e1000184.
- [38] G. Struhl, K. Struhl, P. M. Macdonald, The gradient morphogen bicoid is a concentration-dependent transcriptional activator, *Cell*. 57: 12591273 (1989)
- [39] W. Driever, C. Nusslein-Volhard, The bicoid protein is a positive regulator of hunchback transcription in the early *Drosophila* embryo, *Nature*. 337: 138143 (1989)
- [40] D. S. Burz, R. Rivera-Pomar, H. Jackle, S. D. Hanes, Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo, *EMBO J*. 17(20):5998-6009 (1998)
- [41] A. Pisarev, E. Poustelnikova, M. Samsonova, J. Reinitz, FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution, *Nucl. Acids Res*. 37: 560-566.(2009)
- [42] E. Poustelnikova, A. Pisarev, M. Blagov, M. Samsonova, J. Reinitz, A database for management of gene expression data in situ, *Bioinformatics*. 20: 2212-2221 (2004)
- [43] D. Lebrecht, M. Foehr, E. Smith, F. J. P.Lopes, C. E. Vanario-Alonso, J. Reinitz, D. S. Burz, S. D. Haneet, Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*. *PNAS* 102: 1317613181.
- [44] M. Ptashne, A. Gann *Genes and Signals* Cold Spring Harbor Laboratory Press, (2002).



- [45] T. Gregor, D. W. Tank, E. F. Wieschaus, W. Bialek, Probing the limits to positional information, *Cell*. 130(1):153-64 (2007)
- [46] U Lohr, H. R. Chung, M. Beller, H. Jackle, Antagonistic action of Bicoid and the repressor Capicua determines the spatial limits of *Drosophila* head gene expression domains, *PNAS*. 106(51):21695-700. (2009)
- [47] X. Wu, R. Vakani, S. Small, Two distinct mechanisms for differential positioning of gene expression borders involving the *Drosophila* gap protein giant, *Development*. 125(19):3765-74. (1998)
- [48] V. Kasatkin, A. Prochiantz, D. Holcman, *Bull Math Biol*, Morphogenetic gradients and the stability of boundaries between neighboring morphogenetic, 70(1):156-78 (2008)
- [49] D. Holcman, V. Kasatkin, A. Prochiantz, Modeling homeoprotein intercellular transfer unveils a parsimonious mechanism for gradient and boundary formation in early brain development, *J Theor Biol*. 249 (3): 503-17. (2007)

## 7 Figures

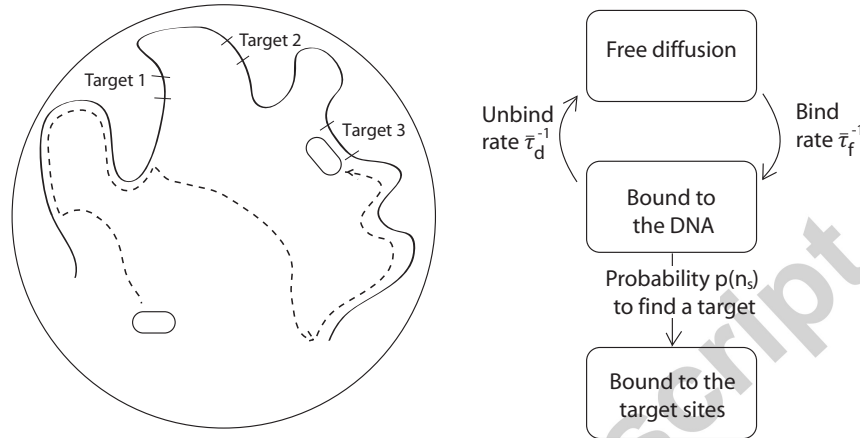


Figure 1: **Scheme of Transcription Factor binding.** TFs alternate between free diffusion in the nucleus and sliding along the DNA until one of the targets is eventually reached during a sliding period. TFs bind and detach to the DNA at a rate  $\bar{\tau}_f^{-1}$  and  $\bar{\tau}_d^{-1}$  respectively. During each sliding period, there are  $n_s$  target sites and the binding probability is  $p(n_s)$ .

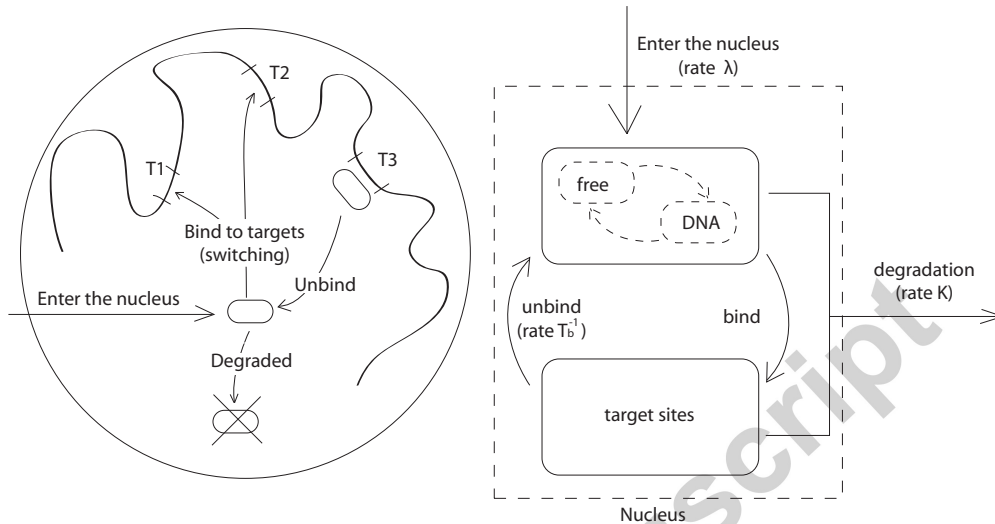


Figure 2: **Schematic representation for the target activation.** TFs enter the nucleus at a rate  $\lambda$  and are degraded at a rate  $K$  (free or bound). The binding rate to the target sites is given by equation (10) and depends on the number of unoccupied sites and on the number of unbound TFs. TFs unbind from the target sites at a rate  $T_b^{-1}$ .

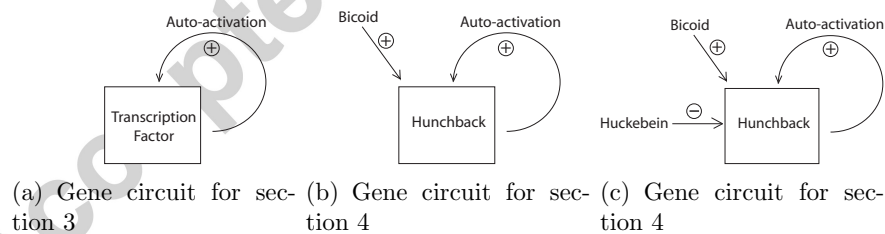


Figure 3: **Gene circuits used in this paper.** (a) Gene circuit for a genetic switch (section 3). (b) Regulation of hunchback (hb) transcription through auto-activation and activation by bicoid (bcd) (section 4). (c) Regulation of hb transcription through auto-activation and activation by bcd and repression by hucklebein (hkb) (section 4).

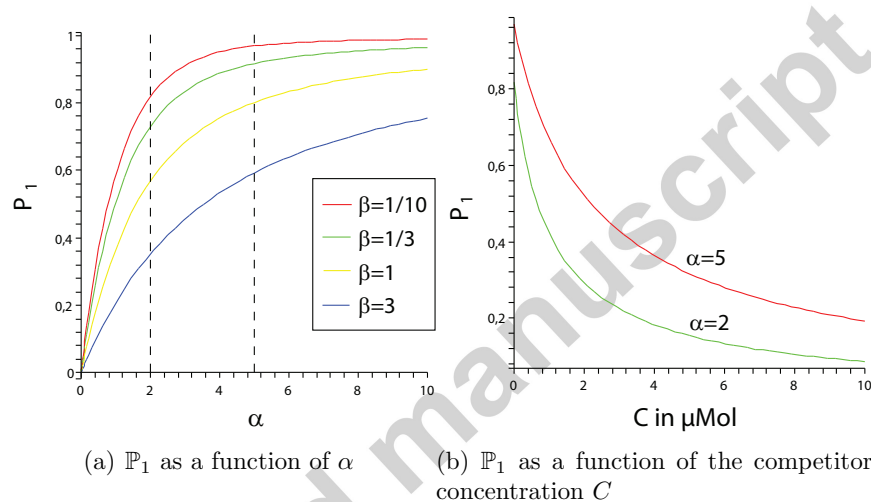
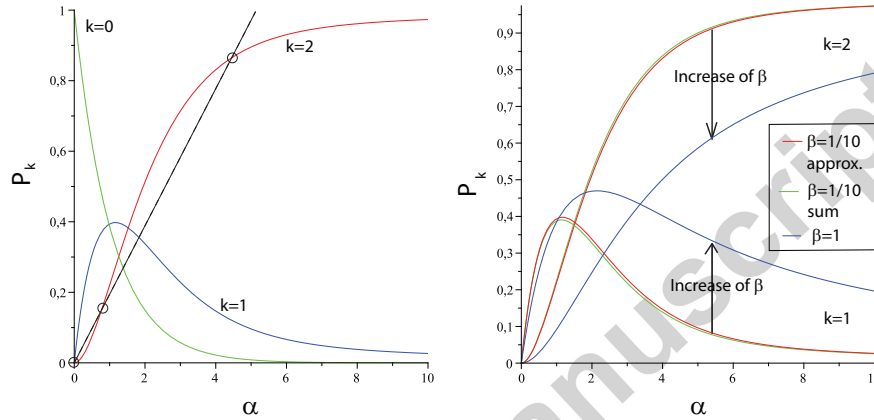
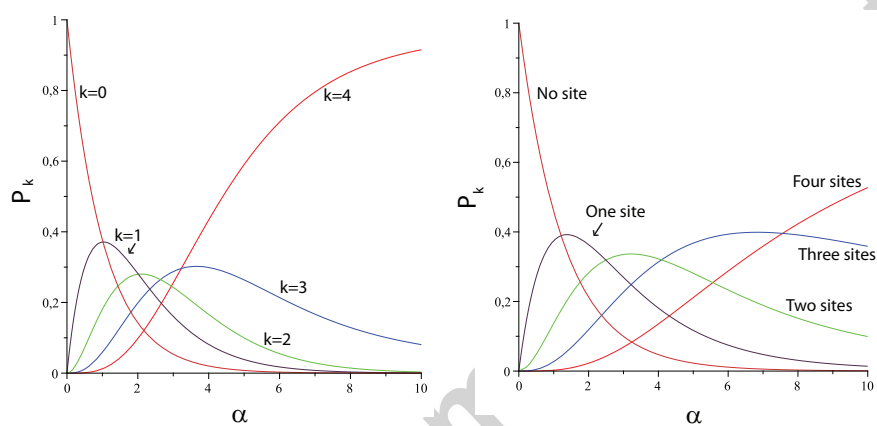


Figure 4: (a)  $\mathbb{P}_1$  as a function of  $\alpha$  for various values of  $\beta$ . From left to right,  $\beta$  increases  $1/10$  (red),  $1/3$  (green),  $1$  (yellow),  $3$  (blue). The upper curves correspond to fast search times and/or long binding times to the target site and no competitors. (b)  $\mathbb{P}_1$  as a function of the competitor concentration  $C$  in  $\mu\text{Mol}$ . The upper curve is obtained for  $\alpha = 5$ , the lower one is for  $\alpha = 2$ , where  $\beta = \beta_0 \left(1 + C \frac{k_a}{k_d}\right)$  with  $\beta_0 = \frac{1}{10}$  for  $C = 0$  and  $\frac{k_a}{k_d} = 20 \mu\text{Mol}^{-1}$  [31].



(a) Probability  $\mathbb{P}_k$  for  $n_s = 2$  and  $\beta = \frac{1}{10}$  (b) Probability  $\mathbb{P}_k$  for  $n_s = 2$  and  $\beta = 1$

Figure 5:  $\mathbb{P}_k$  for  $n_s = 2$ . (a)  $\mathbb{P}_k$  as a function of  $\alpha$  for  $\beta = \frac{1}{10}$ .  $\mathbb{P}_k$  is computed from 43. For a TF activating its own transcription when both sites are simultaneously occupied, the two stable values for  $\alpha$  (high and low values for  $\alpha$ ) and the unstable value (in the middle) are represented along the dotted line. (b) Effect of changing  $\beta$ . Red and green curves are given for  $\beta = 1/10$ , the blue curve for  $\beta = 1$ ,  $\frac{k_a}{k_d} = 20 \mu\text{Mol}^{-1}$  [31] and  $\beta = \frac{1}{10}$ .  $C = 0$  corresponds to a ligand concentration of  $0 \mu\text{Mol}$  (red and green) and  $0.5 \mu\text{Mol}$  (in blue). The red curve is computed from 43. Blue and green curve are computed by approximating equation 32 with 200 terms.



(a) Probability  $\mathbb{P}_k$  for  $n_s = 4$  and  $\beta = \frac{1}{10}$  (b) Probability  $\mathbb{P}_k$  for  $n_s = 4$  and  $\beta = 1$

Figure 6:  $\mathbb{P}_k$  for  $n_s = 4$  as a function of  $\alpha$  for (a)  $\beta = 1/10$  and (b)  $\beta = 1$ .  $\mathbb{P}_k$  is computed through finite sums of 32 (200 first terms). With  $\beta = \frac{1}{10}$  when  $C = 0$  and  $\frac{k_a}{k_d} = 20 \mu\text{Mol}^{-1}$  [31], this corresponds to a ligand concentration of  $0 \mu\text{Mol}$  (left) and  $0.5 \mu\text{Mol}$  (right).

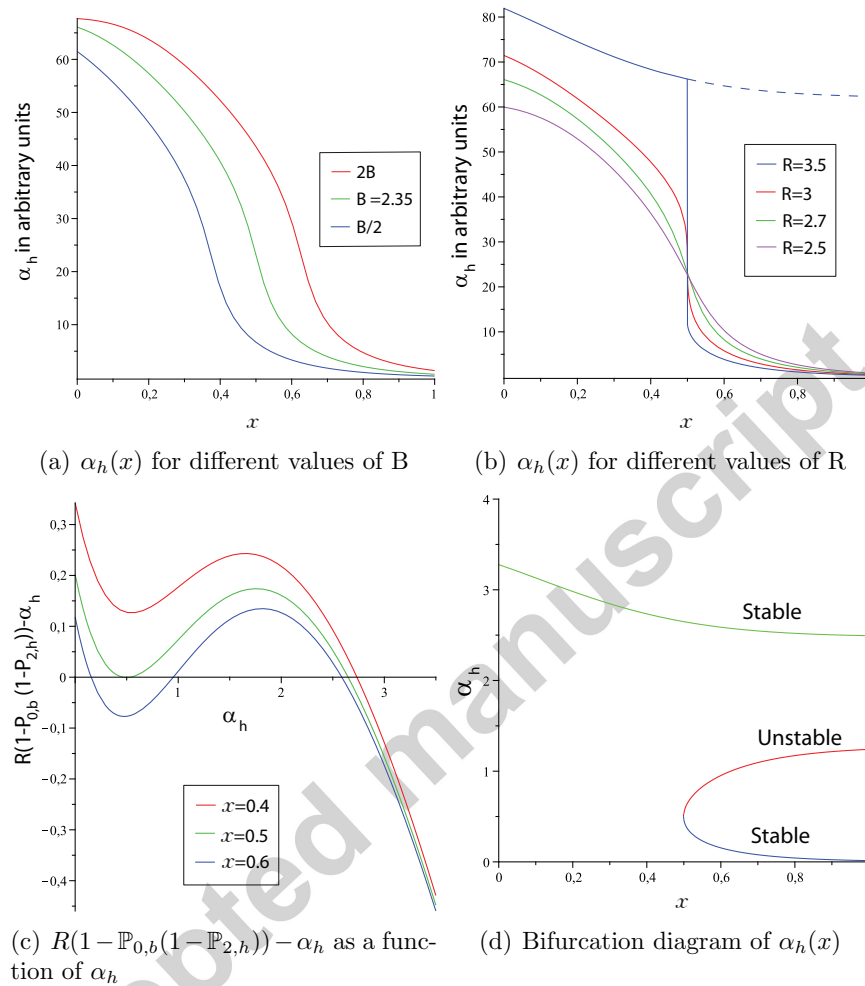
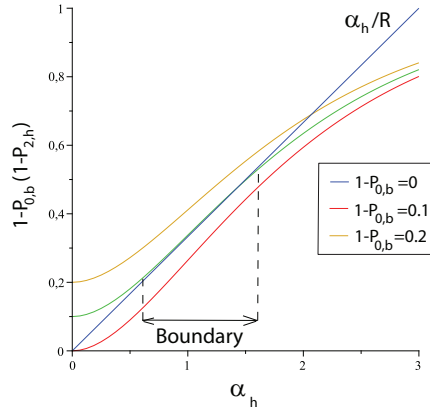


Figure 7: (a) Hb concentration  $\alpha_h(x)$  for different values of  $B$ :  $B = 2.35/2$  (blue),  $B = 2.35$  (green) and  $B = 2.35 * 2$  (red). Here, we use  $R = 2.7$ . All curves for  $\alpha_h$  were scaled by a factor 25 to obtain the same numerical values as the concentration in arbitrary units for in vivo data reproduced in figure 9a. (b)  $\alpha_h(x)$  for different values of  $R$ :  $R = 2.5$  (monostable),  $R = 2.7$  (monostable, value from [37]),  $R = 3$  (critical value for bistability) and  $R = 3.5$  (bistable). For  $R = 3.5$ , there are two stable points: the high (dotted lines) and the low (continue line) stable value.  $B$  in (46) was adjusted for each of the curves to cut 25 in  $x = 0.5$ :  $B = 3.3$  for  $R = 2.5$ ,  $B = 2.35$  for  $R = 2.7$ ,  $B = 1.58$  for  $R = 3$  and  $B = 0.95$  for  $R = 3.5$ . (c)  $R(1 - \mathbb{P}_{0,b}(\alpha_b(x))(1 - \mathbb{P}_{2,h}(\alpha_h))) - \alpha_h$  as a function of  $\alpha_h$  for  $R = 3.5$ . The curves are for  $x = 0.4, 0.5$  and  $0.6$ . We use  $B = 0.95$  as in figure 7b. (d) Bifurcation diagram of  $\alpha_h(x)$ . This bifurcation diagram is given by the solutions of  $R(1 - \mathbb{P}_{0,b}(\alpha_b(x))(1 - \mathbb{P}_{2,h}(\alpha_h))) - \alpha_h = 0$  as a function of  $x$ . We use  $B = 0.95$  and  $R = 3.5$  as in figure 7c.



(a) Boundary in the density of the autoregulated TF.

Figure 8: **Boundary in the density of the autoregulated TF.** Are representd  $\alpha_h/R$  (blue) and the proportion of time  $1 - \mathbb{P}_{0,b}(1 - \mathbb{P}_{2,h}(\alpha_h))$  the hb gene is active for  $P(x) = 1 - \mathbb{P}_{0,b} = 0$  (blue), 0.1 (red) and 0.2 (yellow). The curves are all for the critical value  $R = 3$  to amplify the boundary in hb. We use  $B = 1.58$  as in figure 7b. The boundary comes from  $1 - \mathbb{P}_{0,b}(1 - \mathbb{P}_{2,h}(\alpha_h))$  which is tangent to  $\alpha_h/R$  at the point where  $\mathbb{P}_{2,h}(\alpha_h)$  changes concavity. A small variation in  $P(x)$  then induces a large variation in  $\alpha_h$ .



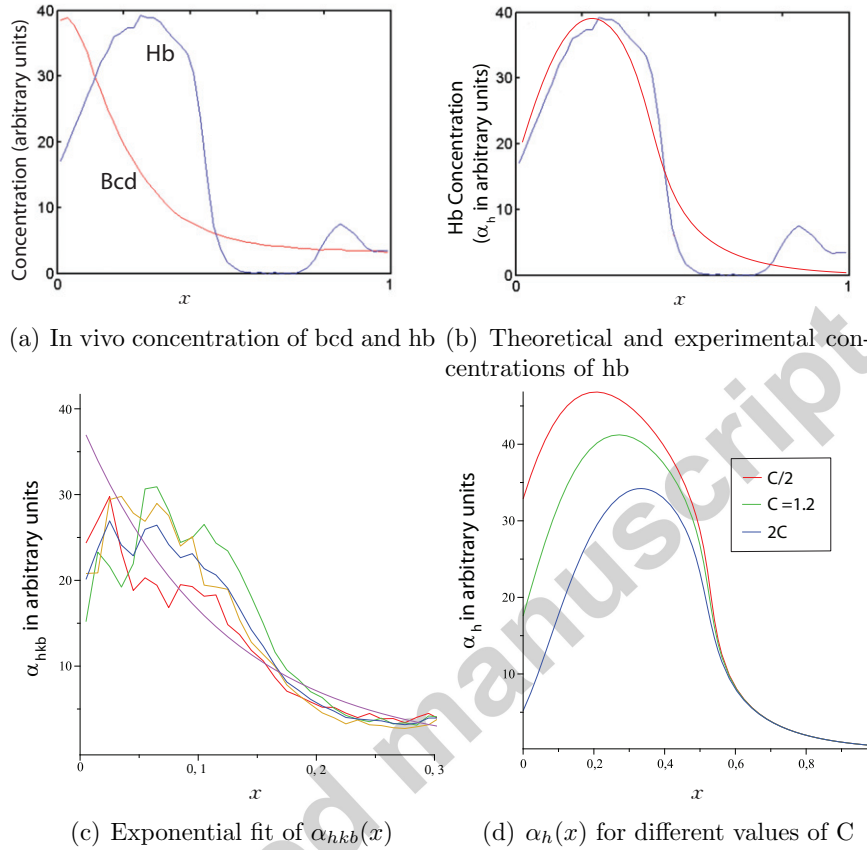


Figure 9: (a) **In vivo concentration of bcd and hb TFs as a function of  $x$ .** Figure reproduced from [16]. (b) Theoretical and experimental hb concentration as a function of  $x$ . Here to compute  $\alpha_h(x)$  we take into account bcd activation, hb autoregulation and hkb repression. The parameters  $B = 1.4$  and  $C = 1.1$  are used to fit the in vivo data which is reproduced from [16]. (c) In vivo hkb density as a function of  $x$  and exponential fit used in equation (61). The in vivo data is from the Flex database [17, 41, 42]. (Since the bcd and hb densities from [16] reproduced in figure (9)a are for the beginning of the 14 A cycle of the development of *Drosophila* [16], we use the in vivo hkb densities for the first half of the 14 A cycle (T1 to T4) from the Flex database to fit in equation 61.) (d) Hb density  $\alpha_h(x)$  for different values of  $C$ : for  $C=0.6$  (red), 1.2 (green), 2.4 (blue). The value of  $C$  is adjusted to have the same value for the hb density at  $x = 0$  as for the in vivo data.