



**HAL**  
open science

## How transcription factors can adjust the gene expression floodgates.

Denis Michel

► **To cite this version:**

Denis Michel. How transcription factors can adjust the gene expression floodgates.. Progress in Biophysics and Molecular Biology, 2010, 102 (1), pp.16-37. 10.1016/j.pbiomolbio.2009.12.007 . hal-00464812

**HAL Id: hal-00464812**

**<https://hal.science/hal-00464812>**

Submitted on 18 Mar 2010

**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.

## How transcription factors can adjust the gene expression floodgates

**Denis Michel**

Université de Rennes1, Molecular and Cellular Interactions UMR6026, Hip, IFR140 GFAS, Bâtiment 13, Campus de Beaulieu, 35042 Rennes Cedex, France

### **Abstract**

The rate of transcription initiation is the main level of quantitative control of gene expression, primarily responsible for the accumulation of mRNAs in the cell. Many, if not all, molecular actors involved in transcription initiation are known but the mechanisms underlying the frequency of initiations, remain elusive. To make the connection between transcription factors and the frequency of transcription initiation, intricate aspects of this complex activity are classified i) depending on whether or not the DNA-bound transcription factors directly activate the commitment to transcription and ii) on the destructive or non-destructive effect of transcription initiation on the stability of promoter complexes. Two possible sources of synergy allowing the combinatorial specificity of transcription factors action are compared, for binding to DNA and for recruiting transcription machineries. Tentative formulations are proposed to discriminate the different micro-reversible modes of DNA binding cooperativity modulating the specificity and dosage of transcription initiation.

*Keywords:* Transcription initiation frequency; Combinatorial regulation; Cooperativity; Steady-state rates.

**Abbreviations** BCD: Bicoid protein; DBD: DNA-binding domain; FRAP: fluorescence recovery after photobleaching; GR: Glucocorticoid receptor; GTF: general transcription factor, component of the PIC; HB: Hunchback protein; HSF : Heat shock factor; PIC: transcription preinitiation complex; Pol II: Eukaryotic RNA-polymerase which transcribes protein-coding genes; RNAP: RNA polymerase of E. Coli; TAD: transcription activation domain; TBP: TATA-binding protein; TCC: transcription commitment complex; TF: transcription factor; TRC: transcription reinitiation complex; Y: fractional time or space saturation.

Unless otherwise specified, the equilibrium constants used in this study are binding constants. To differentiate the traditional notion of reversibility from the physical principle of microscopic reversibility, the latter will be called micro-reversibility.

E-mail addresses: [denis.michel@univ-rennes1.fr](mailto:denis.michel@univ-rennes1.fr); [denis.michel@live.fr](mailto:denis.michel@live.fr)

## 1. Introduction

A wide variety of chromatin components are described, with different variants and modifications. The same inventory exists for the constituents of the transcription initiation complexes and the correlations between these molecules and the status of the cell (developmental stage or differentiation state). However, the innermost principles underlying the transcription initiation rate, which is the preponderant level of regulation of gene expression, are rarely addressed. The nature of the arrow traditionally drawn between TFs and the transcription start site to symbolize a more or less strong activation, is not really defined. For example, relatively few articles deal with the interesting concept of transcription reinitiation, of fundamental importance for setting the transcription rates. Studies are now more focused on the time distribution than the regulation of the average frequency of transcriptional events, probably because of the general interest for biological randomness. Besides, though the transcriptional synergy between transcription factors (TFs) was first assumed to concern the recruitment of transcriptional machineries, this level of cooperativity is now neglected in the literature. A series of elementary principles of transcription initiation are revisited in the present study, more focused on general mechanisms than on particular actors, and starting from the reinterpretation of key data from the literature. Only activated transcription activators will be considered, bearing the appropriate modifications allowing them to reach the nucleus, to bind DNA and to activate transcription. Given the variety of reports on the regulation of transcription, this review cannot be fully comprehensive. Rather, particular examples more thoroughly documented, are selected to illustrate generalizations, such as the Bicoid (BCD)/Hunchback (HB) system. Attention is mainly focused on the genes encoding proteins and transcribed by RNA polymerase II (Pol II) in eukaryotes and particularly metazoans. Only a few simple examples from eubacteria will be given for comparison.

## 2. Archiving vs signal genome functions

DNA is commonly envisioned as an inert support encoding collections of proteins, easily duplicatable and transmittable to the different cells and generations of organisms. This view is correct but obviously incomplete. An unexpected lesson from genomes sequencing is that the mouse and human genomes encode nearly the same proteins, suggesting that the same proteins can make, in a programmed manner, either a mouse or a man. The solution of this apparent enigma has been early and clearly explained by Carroll (Carroll, 2000), who proposed that the fundamental parameter to add to the qualitative series of proteins, is their quantitative spatio-temporal profile of expression, dictating the relative amounts of the different proteins in every cell and at every point of time. This parameter is inscribed in the genome in the form of signal DNA, present around coding DNA. Signal DNA can evolve much more rapidly than protein-coding DNA, which can explain strong phylogenetic evolutions with moderate protein changes. Signal DNA is mainly represented by the *cis*-regulatory

elements present in gene promoters (Carroll, 2000), which are *trans*-activated by proteins which themselves derive from genes, leading to the concept of gene network. This concept is a profound revision of the classical view of the causal relationships supposed to exist between individual gene regulations and signalling pathways, which now appear much more diffuse and indirect than initially assumed. The work of Huang (Huang et al., 2005) is illuminating in this respect, since it shows that two different signals leading to different and often opposite short-term genetic responses, can ultimately result in the same long-term gene expression profile. This very instructive study i) proves that individual gene regulations have no predictive meaning if they are disconnected from the entire network and ii) strongly reduces the interest of gene array technologies in non-steady-state conditions. This network behaviour could have been anticipated from the precursor approaches of somatic cell fusion, in which different cell types are fused together, allowing intergenomic cross-regulation by diffusing factors. These experiments revealed the conflicting and concurrent repression of certain genes and the activation of others, so that the juxtaposition of two pre-existing stationary states leads to a novel stationary state (Mével-Ninio and Weiss, 1981; Chiu and Blau, 1984; Chin and Fournier, 1987). These results also definitively support the fundamental reversibility of gene regulation mechanisms. The fruitful analogy between gene network interactions and differential equations, suggests that the genome is not only a collection of genes, but a series of preset programs, evolutionary selected and governed by appropriate combinations of TF-target sites in gene promoters. TFs and DNA modules were previously envisioned as exclusively specialised in particular cellular activities (for example NF $\kappa$ B in inflammation, or AP1 in cell growth and migration etc). This simplistic causal view is not compatible with the limited number of TFs involved in the regulation of a lot of genes. In turn, if the regulation of any individual gene is ensured by the combinatorial and cooperative action of a set of TFs, a limited number of TFs is sufficient for the differential quantitative regulation of a huge number of genes. According to this network view, a given TF can ensure different roles depending on the cocktail of other TFs present in the cell. It is thus essential, for elucidating cellular programs, to understand how combinations of TFs bound to a given gene promoter, can quantitatively regulate the transcription of this gene. To this end, realistic functions should be defined to describe the action of TFs on gene expression. Given that gene expression cannot be unlimitedly increased, the simplest relevant function enclosing saturability to describe the action of TFs, is the hyperbolic saturation function, or Michaelis-Menten equilibrium hyperbola.

### **3. Hyperbolic saturation function for describing simple binding**

This hyperbola is a fundamental basis of molecular interactions (Biot, 1838). It is obviously applicable to gene expression since the shape of a lot of gene response curves, such as the inducer concentration-dependent responses of bacterial operons, is hyperbolic.

### 3.1. Gradation, saturability and ergodicity, of the hyperbolic saturation function

This function corresponds to the steady-state average of first order cycles, whatever these cycles are micro-irreversible, as catalytic kinetics (Briggs and Haldane, 1925), or micro-reversible. The equilibrium approximation can be retained if the variations of TF concentrations are slow relative to their time of equilibration with DNA, so that the parameter  $k_a[\text{TF}]$ , where  $k_a$  is the association kinetic constant and  $[\text{TF}]$  is the concentration of diffusing TF, can be considered as a pseudo-first order constant ( $\text{s}^{-1}$ ). Under these conditions, hyperbolic saturation can be uncovered as well from the ensemble and single molecule time perspective, reflecting its fundamental ergodicity. The average turnover of a single TF-binding site is

$$T = \frac{1}{k_a[\text{TF}]} + \frac{1}{k_d} \quad \text{Eq. 1}$$

so that the frequency of DNA-TF interaction (inverse of the period length), is

$$\frac{1}{T} = \frac{k_d k_a[\text{TF}]}{k_d + k_a[\text{TF}]} \quad \text{Eq. 2}$$

Hence, given that the mean occupancy per cycle is  $1/k_d$ , the average occupancy of this DNA module is:

$$Y = \frac{1}{T} \cdot \frac{1}{k_d} = \frac{K_d[\text{TF}]}{1 + K_d[\text{TF}]} \quad \text{Eq. 3}$$

which is the typical hyperbolic concentration-dependence formula. In Eq. 3, the concentration of free TFs is assimilated to that of total TFs. This approximation is not always valid in biochemistry. If two molecules A and B interacting with an equilibrium dissociation constant  $K_d$ , are present in significant concentration, only the free concentrations  $[A_F]$  and  $[B_F]$  should be incorporated in the hyperbolic saturation functions.

$$Y_A = \frac{[A_F]}{K_d + [A_F]} = \frac{[AB]}{[B]} \quad \text{Eq. 4a}$$

$$Y_B = \frac{[B_F]}{K_d + [B_F]} = \frac{[AB]}{[A]} \quad \text{Eq. 4b}$$

with

$$[AB] = \left( [A] + [B] + K_d - \sqrt{([A] + [B] + K_d)^2 - 4[A][B]} \right) / 2 \quad \text{Eq. 4c}$$

Fortunately, this complicated situation does not apply to genes, considering that a given DNA module in a given gene, can fix only one TF, or two if considering the two alleles.

In turn, the fraction of TFs sequestered in other DNA sites should not be neglected (von Hippel and Berg, 1986). However, the importance of TF sequestration by DNA is difficult to evaluate and will be neglected in this study, by assuming that upon activation, the amount of TFs diffusing in the nucleoplasm is high enough. This assumption is supported by fluorescence recovery after photobleaching (FRAP) experiments. Even when engineered TFs fused to fluorescent proteins are expressed at roughly the same level than their endogenous counterparts, as monitored by immunoblotting comparison, FRAP experiments generally reveal very high on rates, suggesting that the amount of TFs available for association is significant. This approximation is however likely to not hold for all TFs and to be dramatically dependent on: i) their concentration, given that certain TFs are known to be present in clearly limiting concentrations and ii) on their specific modes of spatial inspection in the nucleus, through scanning, hopping or jumping. For example, a recent comparative FRAP analysis shows that the low mobility of the Fox TFs results from their unidimensional sliding along chromatin, while others such as c-Myc, rapidly re-equilibrate after photobleaching (Sekiya et al., 2009). The TF concentrations considered in the following developments will be those of diffusing TFs, but conversely, sequestration of TFs by overexpressed diffusing partner proteins will be taken into account.

### **3.2. Illustration of the capacity of the hyperbolic saturation function to describe transcriptional responses**

The concentration of active TFs can vary with time, for example following hormone addition, but also in space at one time, as in the case of morphogenetic gradients. A typical example, which will be extensively used in this study, is the transcriptional response to Bicoid (BCD), whose gradient of concentration, decreasing exponentially from the anterior pole, is responsible for the antero-posterior axis of the *Drosophila* embryo (Wu et al., 2007). Hunchback (HB) is a BCD-responsive gene but its expression level does not parallel the concentration of BCD. It is high and constant in the anterior part of the embryo and sharply decreases around the middle part of the embryo. This pattern can be readily modeled in equilibrium by using the hyperbolic saturation function and replacing the concentration of BCD by the distance from the anterior end ( $L_0$ ) (Eq. 5).

$$[\text{BCD}]_L = [\text{BCD}]_{L_0} e^{-hL} \quad \text{Eq. 5}$$

where  $h$  is the constant of decrease of BCD ( $\text{distance}^{-1}$ ). Hence, the resulting fractional saturation of a BCD-binding site for which the affinity of BCD is  $K$  ( $M^{-1}$ ), is the logistic-like function:

$$Y_{(L)} = \frac{K [\text{BCD}]_{L_0} e^{-hL}}{1 + K [\text{BCD}]_{L_0} e^{-hL}} \quad \text{Eq. 6}$$

The saturability inherent to the hyperbolic saturation function generates a plateau of BCD-responsiveness, whose length depends on the DNA-binding affinity of BCD (Fig. 1). The boundary of HB expression, which is a narrow gradient, is expected at the distance

$$L_{1/2} = \frac{1}{h} \text{Ln}(K[\text{BCD}]_{L_0}) \quad \text{Eq. 7}$$

According to Eq. 7, increasing either the global production of BCD or its affinity for DNA, would shift the HB expression boundary to the posterior side, as experimentally observed (Struhl et al., 1989). The cooperativity of BCD fixation to the Hb gene promoter has been proposed to further increase the sharpness of the boundary.

#### 4. Adair equations are suitable for describing multiple binding of transcription activators

Genes encoding TFs are the minority. Accordingly, to increase the number of differentially regulated TF targets, full gene promoter activation is generally not triggered by a single TF but by a combination of TFs, different or identical. This organization allows cell-type specific gene expression through the combinatorial distribution of TFs. Several approximations are generally retained to simplify the formal description of this situation.

- Gene promoters are made such that their maximal activity is obtained when all active TFs capable of binding to the promoter are present (opposite hypotheses are however encountered in the literature and mentioned in section 7.2.2.3).
- The overall promoter activity is the sum of all individual TFs bound to the promoter (discussed and refuted in section 7.2).

For intricate binding phenomena, the saturation hyperbola can be usefully replaced by the Adair function, established in the context of hemoglobin oxygenation (Adair, 1925) and based on the mass action law. The fractional activity of a macromolecule  $Y$  is

$$Y = \frac{\text{number of occupied binding sites participating to the considered activity}}{\text{number of possible binding sites}} \quad \text{Eq. 8}$$

Gilbert Adair expressed the fractional saturation  $Y$  of the macromolecule by using the concentrations of the different liganded microstates. For example, if the macromolecule  $M$  contains two binding sites for the ligands  $A$  and  $B$ :

$$Y = \frac{[MA] + [MB] + 2[MAB]}{2([M] + [MA] + [MB] + [MAB])} \quad \text{Eq. 9}$$

Given that at equilibrium,

$$[MA] = [M] K_A [A] \text{ and } [MB] = [M] K_B [B] \quad \text{Eq. 10}$$

One obtains

$$Y = \frac{K_A [A] + K_B [B] + 2 K_A [A] K_B [B]}{2(1 + K_A [A] + K_B [B] + K_A [A] K_B [B])} \quad \text{Eq. 11}$$

Since the transcriptional influence of the different TFs bound to the promoter can be different, it is sometimes necessary to split the global promoter fractional occupancy  $Y$  by  $A$  and  $B$  into individual fractional saturations (Eq. 12).

$$Y_A = \frac{K_A [A] + K_A [A] K_B [B]}{1 + K_A [A] + K_B [B] + K_A [A] K_B [B]} \quad \text{Eq. 12a}$$

$$Y_B = \frac{K_B [B] + K_A [A] K_B [B]}{1 + K_A [A] + K_B [B] + K_A [A] K_B [B]} \quad \text{Eq. 12b}$$

Interestingly all these ergodic fractional saturations, in the hyperbola or Adair form, can be envisioned from several perspectives which are all equivalent: i) for a population of macromolecules, they represent the proportion of molecules bound at a time; ii) for a single macromolecule, they correspond to probability of binding at one time and hence, iii) they can also be considered as the fraction of time spent in the bound form.

## 5. The micro-irreversible transition in transcription initiation

The fractional saturations  $Y$  described above, fully comply with the principle of microscopic reversibility and accordingly, are describable using equilibrium binding constants. It is however clear that the chain of events taking place between TF binding and transcription elongation, necessarily includes at least one “one-way arrow” reflecting a micro-irreversible process. If the simple binding of

TFs to DNA or stable protein complexes can be micro-reversible, the no-return escape of Pol II from the promoter is by nature micro-irreversible.

A critical point for defining the transcription initiation rate is the place of the first micro-irreversible step leading to transcription initiation. An unsolved question is to determine if DNA-bound TFs can directly activate an initiation catalyst, or should stabilize an intermediate transcription reinitiation complex. It is difficult, from the current literature, to decide which case is the right one and in fact, as discussed later, both could coexist. Let us define two possible types of transcription complexes recruitable by DNA-bound TFs, called TRC and TCC. Once bound to DNA, the TFs could directly trigger the action of a transcription commitment complex (TCC). Alternatively, the TFs could micro-reversibly recruit a transcription reinitiation complex (TRC) (Fig. 2A), supposed to be both: i) autonomous for reinitiating transcription as long as it is linked to the promoter and ii) requiring TFs for its recruitment. The rate ( $s^{-1}$ ) corresponding to the first one-way arrow in the two schemes of Fig. 2, could be assimilated to a pseudo-first order constant if supposing the recruitment of diffusing TCC ( $k[\text{TCC}]$ ). But it could also correspond to a true first order constant if the contact between TFs and TCC occurs upon conformational fluctuations of the pre-loaded promoter. Finally, it could be the rate of interaction with an active dissociating machinery, disrupting the association between TFs and transcription elongation complexes. This case, in which transcription is triggered by a dissociation phenomenon, is suggested by one-shot TFs described later (section 8.2) and by the disruptive action of HSF1 on the transcription preinitiation complex (PIC) (section 9.2.4). To incorporate all these possible mechanisms, the rate constant of the first irreversible transition will be simply written  $k_S$  (start rate) in the rest of the manuscript.

## **6. Time connection between enhancers and transcription complexes**

### **6.1. The bridging nature of TFs**

TFs have an obviously modular structure reflecting their dual capacity to bind both to DNA, at the level of TF-binding sites (enhancers) through their DNA-binding domains (DBD), and to transcription complexes through their transactivation domains (TAD), making the link between DNA and transcription machineries. This functional organization is also confirmed by the remarkable efficacy of engineered chimeric TFs made of heterologous or synthetic DBDs and TADs. Other bridging functions have been proposed for TFs, including the localization of the gene inside nuclear domains called transcription factories. This role is however less consistent with certain properties of TFs such as the very dynamic nature of their interactions with DNA, which would become puzzling once the gene is present in a transcription factory. Transcription factories can enhance transcription by increasing the local concentration of diffusing molecular actors. They are discussed in a recent review (Sutherland and Bickmore, 2009) and will not be considered here. Since several TFs can bind to the

same promoter and to the same transcription complex, a first important point is to determine if there is space enough to allow the simultaneous contribution of several TFs in the promoter-transcription connection. A clear difference appears between the two models previously defined (Fig. 2).

### 6.2. Unlimited connection to a transcription commitment complex

Molecular association reactions are instantaneous jumps, preceded by given waiting times but devoid of intrinsic durations, so that the same TCC can be bound without restriction by many different DNA-bound TFs. The resulting rate  $k_S$  is simply the sum of that of all participating TFs.

### 6.3. Competition for the same surfaces

Competition effects appear if postulating the micro-reversible recruitment/stabilization of a TRC. If  $n$  ligands  $A_i$  can bind independently and simultaneously to the same macromolecule, then, the fractional saturation of this macromolecule is

$$Y = \frac{1}{n} \sum_{i=1}^n \frac{A_i}{1+A_i} \quad \text{Eq. 13}$$

where  $A_i$  is a dimensionless normalized concentration:  $A_i = K_{A_i} [A_i]$  and  $K_{A_i}$  is an equilibrium binding constant. Conversely, if the  $n$  ligand molecules have overlapping interaction surfaces on the macromolecule, they should share time to associate with it (Eq. 14).

$$Y = \frac{\sum_{i=1}^n A_i}{1 + \sum_{i=1}^n A_i} \quad \text{Eq. 14}$$

Both the full-time and time-share modes of residence can apply for TF binding to DNA as well as to transcription complex, but the full-time mechanism is more likely for gene promoter occupancy since many DNA modules can be juxtaposed in a promoter. The situation is more problematic for the TRC. Promoter-bound TFs can contact different components of the transcription complex or different surfaces of the same component, according to the multiacceptor hypothesis (Herbomel, 1990), but the number of possible interaction surfaces appears at first analysis more limited in a TRC than in DNA. Indeed, a comparative survey of the different TF families shows that the structural diversity of the DBDs largely exceeds that of the TADs, which fall into a limited number of categories (acidic,

glutamine- or proline-rich). Moreover, many examples show that different TFs binding to the same promoter have the same TAD.

## 7. The origin of the synergistic action of TFs

The specificity of gene expression ensured by unique combinations of TF-binding sites in gene promoters, can work if the TFs act synergistically. In the bridging model of TF action (DNA binding and transactivation), two main mechanisms can be involved in the non-additive action of several TFs: i) non-independent fixation to DNA or ii) non-independent activation of transcription by DNA bound-TFs.

### 7.1. Non-independent binding of the TFs on the promoter

The synergy of promoter occupancy by TFs is abundantly documented and widely used in modeling studies. It offers many possible cooperative mechanisms, detailed later and which can be classified in two main categories: reciprocal and hierarchical cooperativity. DNA-binding cooperativity is the only source of cooperativity possible if assuming the direct irreversible recruitment of the TCC by DNA-bound TFs (Fig. 2B). In turn, the micro-reversible stabilization of a TRC allows an additional layer of cooperativity.

### 7.2. Additive vs non-additive actions of DNA-bound TFs.

#### 7.2.1 Additive activation of transcription machineries

According to this possibility, the transcription-triggering activity of TFs is mainly kinetic and based on their binding to the hypothetical TCC (initiation scheme described in section 9.2.1). The resultant global recruitment rate of the promoter is then the sum of the individual TF recruitment strengths.

$$k_{\text{init}} = n Y_{\text{TF1}} k_{\text{TF1}} + m Y_{\text{TF2}} k_{\text{TF2}} + \dots \quad \text{Eq. 15}$$

where  $Y_{\text{TF1}}$  and  $Y_{\text{TF2}}$  are the fractional saturations of the  $n$  and  $m$  binding sites in the promoter, for TF1 and TF2 molecules respectively.  $k_{\text{TF1}}$  and  $k_{\text{TF2}}$  are the rates of association with  $\text{TCC}_1$  and  $\text{TCC}_2$  (which are, most probably, the same), by the TADs of DNA-bound TFs. The involvements of TADs in triggering a micro-irreversible step can appear in fact logical, considering that stable interactions between TFs and the transcription machineries would be counterproductive for the escape of Pol II.

### 7.2.2. Non-additive recruitment of transcription machineries by DNA-bound TFs

The alternative possibility that DNA-bound TFs could have non-additive effects on the recruitment of the transcription machinery, is suggested by certain experimental data and is (implicitly or unintentionally) assumed in theoretical postulates encountered in modeling articles, which propose non-additive actions of multiple TFs.

**7.2.2.1. Experimental data.** Authors showed that increasing the number of interaction surfaces between TFs and components of the transcription preinitiation complex (PIC), increases the resulting transcription in a non-additive fashion (Sauer et al., 1995). The transcriptional activity of HB and BCD are mediated by TAFII60 or TAFII110 respectively (TBP-associated factors). While HB and BCD alone cause a 7-fold and 6-fold increase respectively of the activity of a promoter containing HB and BCD-binding sites, the conjunction HB + BCD leads to a 65-fold increase of this activity (more than additive and more than multiplicative). This synergistic action is destroyed if either TAFII110 or TAFII60 is missing (Sauer et al., 1995). This result suggests that the rebinding time of TFs to the PIC is strongly shortened when there is more than one TF under the full-time/full-time condition.

**7.2.2.2. The BOTH mechanism with several DNA modules.** The non-additive action of different TFs is assumed when supposing that the simultaneous presence of several TFs is required to trigger transcription (Fig. 4B line 2). This postulate is somehow related to the work of (Sauer et al., 1995), if further decreasing the individual actions of HB and BCD when alone.

**7.2.2.3 The OR mechanism with several DNA modules.** According to this model, full promoter activity can be obtained even when some TF-binding sites are still empty. This postulate sharply contradicts the previous assumption that maximum activity requires the whole TF combination on the promoter. The OR gate is confined in theoretical articles but is not well supported by experimental observations. Indeed, most data show that adding more TFs in a synthetic promoter, increases the maximal strength of a promoter. A typical OR equation is:

$$Y = \frac{K_A[A] + K_B[B] + K_A[A]K_B[B]}{1 + K_A[A] + K_B[B] + K_A[A]K_B[B]} \quad \text{Eq. 16}$$

This global transcriptional activity can tend to 1 even when [B]=0. More puzzling schemes found in the literature will not detailed here, such as the exclusive OR gate (XOR) in which, when a gene is governed by two DNA modules, full expression is obtained when only one of them is occupied but, intriguingly, no longer when both are occupied. The last term of the numerator of Eq. 16 is eliminated in this case. If the OR gate is realistic for a single DNA module, the inventive XOR gate is biologically doubtful in all cases, for a single as well as for several DNA modules.

**7.2.2.4. The "two-states" promoter modeling.** The idea of "on" and "off" states postulated in binary models, denies the importance of the combinatorial status of promoter occupancy for the quantitative adjustment of gene expression. If the "off" state (0) can result from transcriptional

repression, the "on" state has generally not the unique value of 1. Once again, these two states are not well defined in term of structural chemistry.

### **7.3. Comparative plausibility of the irreversible vs reversible recruitment of transcription complexes by TFs.**

Taking simultaneously into consideration the two layers of cooperativity described above, i) for binding to DNA and ii) for recruiting the transcription machinery, complicates the quantitative modeling of transcription and accordingly, a single aspect is generally addressed at the same time. For instance, in the study of Sauer (Sauer et al., 1995), near-saturating concentrations of BCD and HB were used to minimize their DNA-binding synergy. At the opposite, in more recent studies such as (Lopes et al., 2008), the influence of the HB/BCD combination on transcription is exclusively examined at the DNA-binding level. The choice to favour DNA-binding synergy is now predominant (Bolouri and Davidson, 2002; Bintu et al., 2005; Zinzen et al., 2006), but is not justified. Some speculative arguments can be tentatively proposed to support it. Transcription initiation includes by essence irreversible steps such as Pol II escape from the preinitiation complex, without possibility to go back. One should postulate that TFs can induce re-iterative transcription initiations through mobilizing a factor, called TCC in this study, irreversibly triggering transcription initiation once recruited. The important point, from a physical chemistry perspective, is that an irreversible process can be treated only using kinetic approaches. This is not the case for the interactions between DNA and TFs, which are reversible by nature and can be (at least in part) thermodynamically modeled using equilibrium constants.

On reflexion, several observations suggest that the downstream action of DNA-bound TFs is mainly kinetic. If assuming that the number of contacts between the different TFs and the transcription initiation machinery is responsible for the synergistic action of TF combinations, a series of DNA modules for different TFs would have a greater efficiency than a series of DNA modules for the same TF, which would compete for binding to the same domain of the transcription complex. Real examples suggest that this is not the case. Every element from an array of DNA modules for the same TF, contributes to the global promoter activity, as long observed for the six Sp1 binding sites present in the early promoter of the simian virus 40 (Barrera-Saldana et al., 1985). This non-competitive behaviour has then been generalized to arrays of DNA elements for the same TFs, either natural or synthetic (Strähle et al., 1988; Koutroubas et al., 2008 and many personal data not shown). The fact that synergy can be obtained as well with different and identical TFs (unique surface) and the low diversity of the TAD structures, concur to support the model of direct micro-irreversible activation of transcription initiating factors by TFs (Fig. 2B).

#### **7.4. Transcription initiation and reinitiation: multiplying affinities and summing frequencies**

Synergistic recruitment of transcription complexes through different surfaces by several DNA-bound TFs (Sauer et al., 1995) cannot be ruled out and has been recently supported by structural studies showing that the composition of TFIID (the PIC subcomplex including TBP and TAFs), depends on the upstream activators (Liu et al., 2009). To reconcile these observations with the numerous studies neglecting this layer of cooperativity, one can propose that the two possibilities shown in Fig. 2 could coexist. Transcription initiation has long been established to be preceded by the formation of stable PIC but the relative frequencies of PIC formation and transcription initiation remain elusive. The synergistic activity of BCD and HB shown in (Sauer et al., 1995) concerns the assembly and stability of the PIC, which can include the TRC or participate to the recruitment of the TRC. The PIC is a pile of general TFs (GTFs) clustered on the proximal promoter (Krishnamurphy and Hampsey, 2009). Their recruitment is sequential and cooperative and can be nucleated under the dual influence of pre-bound TFs and of proximal DNA elements. The  $n^{\text{th}}$  component is cooperatively recruited by the  $(n-1)$  components previously clustered and reciprocally, once recruited, can further stabilize or lock them. Given that mRNAs can be generated every 3-10 s, it seems logical that the same PIC platform could remain assembled to allow repetitive recruitments or activations of a few actors including Pol II and TCC. One can easily conceive that preinitiation is mainly regulated by the strengths of interactions with DNA-bound TFs. The overadditive effect of the BCD/HB conjunction shown in (Sauer et al., 1995), is acceptable from this perspective. The TADs of active TFs could be involved both in the recruitment of the PIC and in the subsequent re-iterative activations of the TCC. Besides, for the genes whose expression is established continuously, transcription reinitiation could be further facilitated by Pol II recycling through the coupling of transcription termination and reinitiation (Woodard et al., 2001; Orphanides and Reinberg, 2002). The random recruitment of Pol II can be avoided by bridging the 5' and 3' ends of the gene (El Kaderi et al., 2009) (section 9.5).

#### **7.5. Nature of the transcription reinitiation and commitment complexes**

The TRC and TCC complexes postulated above for illustrating theoretical principles, could in practice include previously known components of the PIC, mediator or elongation complexes. For simplicity, they are approximated as preformed, using a time scale separation hypothesis. Their precise composition can depend on the type of recruiting TF and on the promoter architecture.

The TRC could be represented by GTFs and some components of the PIC, such as the scaffold reinitiation complex postulated in (Yudkovsky et al., 2000), which are: i) not self-sufficient for binding to the proximal promoter and ii) remaining autonomous for reinitiating transcription after initial recruitment and stabilization by TFs. The idea that GTFs are recruited only during the time

windows of TF residence, even if they are short, has also been proposed by (Bosisio et al., 2006). A dilemma in this functional dissection attempt, is that the classical PIC can be recruited and stabilized through both TF binding and direct interactions with DNA, when the modules of the proximal promoter and 5' untranslated region are consensual enough (section 7.6).

The TCC can be even more diverse in term of composition and of mode of action, and only a few examples will be provided here. The TCC could correspond to the few components including the Pol II core, leaving the promoter at every round of transcription initiation in the mechanism proposed by (Yudkovsky et al., 2000). The physical separation between Pol II and the promoter can be mediated by enzymatic components of the TFIID complex which include DNA helicases, possibly involved in promoter melting and a kinase phosphorylating the carboxy-terminus of Pol II. An elongation factor has also been involved in transcription reinitiation (Szentirmay and Sawadogo, 1993) and the CDK9/P-TEFb kinase (Positive Transcription Elongation Factor b), has been shown to convert Pol II into an elongation-competent form (Ni et al., 2008). The start rate  $k_S$  used here, can thus be defined as the reciprocal of the waiting time necessary to restore the conditions allowing the next initiation, including the recruitment of the fresh components listed above. The action of TFs can be more based on induced dissociation than association. For example, an interaction between TBP and Pol II has been proposed to be broken by some TFs such as VP16 and the heat shock factor HSF1, competing for the same interaction surface (Mason and Lis, 1997). The release of Pol II can also go with the partial or complete disassembly of the PIC (section 8.1). Furthermore, the displacement of the triggering TF itself could be required for Pol II escape, as described in section 8.2. The status of Pol II prior to its escape is variable. Pol II can be: i) entrapped in the PIC, ii) continuously recycling in the proximal transcribed region of the genes submitted to the so-called abortive transcription mechanism (Plet et al., 1995), or iii) transcriptionally engaged but pausing in the upstream transcribed region (Gilmour, 2009). This diversity of situations is not of fundamental mechanistic importance since it only affects the rapidity of the first initiation round, but not the reinitiation rate for established gene expression. It confirms that the micro-irreversible commitment step can involve various actors.

### **7.6. Recruitment of transcription machineries by both proximal promoter DNA and TFs**

Transcription initiation machineries are often represented in a triangular interaction diagram including: i) the transcription machineries containing Pol II, ii) the PIC, bound to proximal DNA and iii) TFs bound to upstream promoter elements, which participate to the recruitment and stabilization of the PIC and then to transcription initiation. The triggering role of TFs on transcription start is evidenced by the very low expression level obtained with a basal promoter, as ideal as it can be. Conversely, enhancers and TFs need proximal elements to work efficiently. This functional organization further suggests that the two theoretical mechanisms proposed here for the TF action: micro-reversible recruitment of a TRC and micro-irreversible mobilisation of a TCC can be conjointly involved in regulating

transcription. As assumed in the context of the control of accuracy, the specificity of interactions and substrate discrimination, depends more on dissociation than association rates. The TRC recruitment results from the balance between association and dissociation, while TCC recruitment depends only on the efficacy of associations. Hence, the micro-reversible stabilization of an indispensable complex by TFs could be a double-check security mechanism, in addition to TF affinity for DNA, filtering non-specific transcription initiations. Besides, the unicity of the TRC recruited by different TFs can be ensured if the TRC is a conditional component of the PIC. Two schematic transcription initiation schemes will be retained here to illustrate the cases in which the basal complexes can directly bind to the proximal promoter elements without the help of an upstream TF. Once built, this proximal complex has to interact with a DNA-bound TF to initiate transcription, in a conservative (section 9.2.3, scheme RC2) or destructive manner (section 9.2.4, scheme IC3).

A tentative unifying explanation could be that the recruitment of transcription machineries by dual interactions with the basal PIC and TFs, and the mutual stabilization between the PIC and the TFs, is beneficial to mix micro-reversible interactions ensuring specificity, and micro-irreversible interactions allowing the additive contribution of many TFs. The stabilization of the PIC can be assured by only a subset of promoter-bound TFs while the recruitment of the TCC can involve all the TFs endowed with TADs.

## **8. Partial or complete disassembly of TFs and/or transcription complexes upon transcription initiation**

Observations suggest that the firing of Pol II can lead to the dismantlement of all or part of the PIC, and to the displacement of the triggering TF. The forced dissociation of TFs from DNA upon transcription initiation (one-shot TFs), can be coupled to their destruction (single-use TFs) or not (actively recycling TFs).

### **8.1. Reassembly of the PIC before every initiation cycle**

This possibility has been suggested for hormone-regulated genes which appear periodically deactivated, giving rise to transcriptional cycles visible at the cell population level thanks to the synchronizing effect of hormone addition. In the case of the pS2 gene whose expression is triggered by hormone-activated estrogen receptor (ER), the recruitment of Pol II in the promoter is included in the cyclical recruitment of a series of 200 components (Lemaire et al., 2006; Métivier et al., 2003). The oscillations with a 40 min-long period observed at the level of a cell population after hormone addition and fading with time, are compatible with the transient solution of a large loop of linear differential equations (Lemaire et al., 2006). The interpretation of nuclear receptor-induced cycles has recently been completed by introducing a few energy-driven micro-irreversible transitions between

small complexes (Degenhardt et al., 2009). These results however reflect cycles of transcriptional competence rather than authentic transcription initiation cycles. Indeed, given the limited lifetime of mRNAs, the recruitment of Pol II on the promoter every 40 min would logically result in undetectable mRNA accumulation. The study of (Degenhardt et al., 2009) opens the possibility of multiple Pol II recruitments per cycle.

True disassembly and reassembly of all or part of the PIC in transcriptionally-competent chromatin is suggested by the recycling of GTFs (Zawel et al., 1995) and sometimes TBP (Dasgupta et al., 2005), which is considered as a pivotal component of the PIC, even for promoters that lack a TATA box. While the TATA box/TBP interaction is very stable and exceeds two hours *in vitro*, TBP can be actively displaced *in vivo*, in yeast by the ATPase Mot1. Interestingly, the inactivation of Mot1 has different outcomes depending on the genes. For a subset of genes, Mot1 defect intriguingly leads to increased TBP binding but inhibition of expression (Dasgupta et al., 2005). A speculative explanation of this apparent paradox could be that this subset of genes is dependent on TBP turnovers for transcription reinitiation. Accordingly, a mutation of TBP capable of rescuing Mot1 defect, renders TBP less stably bound to DNA (Sprouse et al., 2009). The need for a displacement of TBP is also supported by the observation that high concentrations of TBP lead to a single round of initiation but forbid reinitiations (Szentirmay and Sawadogo, 1993).

Finally, PIC turnovers are suggested for heat shock genes, whose activation is coupled to a marked instability of proximal promoter complexes (Lebedeva et al., 2005). In this case, the destruction, rather than the recruitment of the PIC, could be dependent on the transcription-triggering TF. Consistent with this possibility, HSF1-dependent genes generally have an ideal TATA box, likely to be sufficient for recruiting TBP and nucleating the PIC platform without the help of a TF. This example shows that the TRC postulated here cannot be completely assimilated to the PIC. This situation could be convenient for emergency genes, whose promoters are pre-loaded with general machineries for rapid induction in case of necessity. In such a system, the role of the upstream TF could be to regulate exclusively the reinitiation rate, as suggested for HSF1 (Sandaltzopoulos and Becker, 1998).

## **8.2. Necessarily recycling TFs**

There are clearly several classes of TF behaviours with respect to the number of reinitiation rounds allowed at each TF-binding cycle. HSF1 is stable when active (Yao et al., 2006), suggesting that multiple rounds of reinitiation occur while it remains bound. Conversely, preventing the degradation of certain TFs abrogates their activity. This is the case of ATF6 whose domains involved in transactivation and degradation are the same, suggesting that both activities are coupled (Thuerlauf et al., 2002). Accordingly, continuous exchanges of this TF are required for sustained target gene

expression (unpublished results). The dissociation of an intensely recycling TF from DNA is of primary importance for setting  $k_{\text{init}}$ .

Active dissociation of TFs from DNA can be ensured by protein chaperones (Elbi et al., 2004) or degradation by the ubiquitine-proteasome (Molinari et al., 1999; Thomas and Tyers, 2000; Salghetti et al., 2001). The TADs of one-shot TF are involved in the functional coupling between TF degradation and transactivation. This is further supported by the belonging of certain transcription coactivators to the ring-finger family of proteins, involved in ubiquitin conjugation. A singular and somewhat paradoxical property of one-shot TFs is that they behave as repressors when prevented to dissociate. Remarkably, this property seems to have been exploited physiologically: ATF6 $\beta$ , a less degradable variant of ATF6 involved in the endoplasmic reticulum stress (ER stress), has been shown to attenuate ER stress (Thuerauf et al., 2007).

The case of one-shot TFs suggests an alternative view of the micro-irreversible step of transcription initiation, which can be envisioned more as a dissociation, releasing Pol II from a complex, than an association mechanism. The definition of  $k_s$  in this case would be the rate of active dissociation between TF and the transcription elongation complex. Transcription initiation concomitant to disassembly of promoter-bound proteins or complexes, will be called destructive initiation. The destructive *vs* non-destructive nature of transcription initiation, can be combined with the classification proposed earlier of micro-reversible or irreversible actions of DNA-bound TFs (Fig. 2).

## **9. Formulations for one TF, of transcription reinitiation frequencies**

The classical transcription initiation schemes similar to that shown in (Krishnamurphy and Hampsey, 2009), have remained unchanged for several decades. If they satisfactorily represent the different actors, they do not provide clear pictures of the underlying physical mechanisms. In an attempt to codify the different observations previously introduced, a non-exhaustive list of simplified formulations of the transcription initiation frequency ( $k_{\text{init}}$ ) will be explicated.

### **9.1. Combination of micro-reversible *vs* irreversible and destructive *vs* non-destructive transcription initiation schemes**

Four different types of transcription initiation schemes are expected from the combination of the mechanisms reported above. They are symbolized: IC, ID, RC and RD, where the first term, I or R, refers to the micro-Irrversible and micro-Reversible action of the DNA-bound TF respectively (sections 5 and 6), and the second term, C or D, refers to the Conservative or Destructive mode of transcription initiation (section 8). To make concrete these notions, steady-state equations are proposed in the case of a single TF (Fig. 3), since extensions to TF combinations can lead to very

complex equations for certain schemes. Synergy and cooperativity are of course excluded for a single TF, except in the case of driven cooperativity which will not be described here. Transient evolutions, experimentally inscrutable and subject to ill-conditioning, are not examined. The definitions of initiation rate  $k_{init}$  proposed below include pseudo-first order rates which depend on the local concentration of the components to be recruited. Hence,  $k_{init}$  can be increased by increasing these concentrations. This can be achieved by targeting the gene to transcription factories or nuclear pores.

## 9.2. Conservative initiation schemes

A common feature of the IC and RC schemes, is that the average transcription rate is independent of the TF-binding cycles.

### 9.2.1. Irreversible conservative scheme (IC)

In this scheme, Pol II is irreversibly committed to transcribe once the TCC is activated by DNA-bound TF, and transcription can be reinitiated as long as the TF remains bound to DNA. The IC equation shown in Fig. 3B is common to several types of mechanisms (IC1, 2 and 3). This particular definition of  $k_{init}$  is the most widely, if not the sole, used in transcription modeling studies. It is formally equivalent to the asymptotic rate of accumulation of mRNAs expected in the Markovian modeling of Peccoud and Ycart (Peccoud and Ycart, 1995), if assimilating the concept of active state used in this study, to the presence of a unique TF sufficient for triggering initiation (i.e.  $\lambda=k_a[TF]$ ;  $\mu=k_d$ ;  $v=k_s$ ). In this scheme, the relationship  $Y=v/V_{max}$  is obtained without need for the rapid pre-equilibrium hypothesis, contrary to the classical rule of enzymology. This scheme can be generalized to more than one TF, by using the general Eq. 17. It allows to incorporate the synergy of DNA binding but not of transcription complex recruitment, which is assumed to be kinetically additive.

$$k_{init}=Y_{DNA}.k_s \quad \text{Eq. 17}$$

As previously mentioned, if  $k_s$  is different for the different TFs, the fractional saturation  $Y$  should be individualized as in Eqs. 12 and 15.

### 9.2.2. Reversible conservative scheme with TRC recruitment

The DNA-bound TF should bind to a TRC to initiate transcription. In this case,

$$k_{init}=Y_{DNA}.Y_{TRC}.k_s \quad \text{Eq. 18}$$

Various developments of Eq. 18 are possible depending on the full-full or full-share mode of fractional saturation or other cooperative systems detailed later. The fundamental difference with the previous scheme (Eq. 17), is that Eq. 18 includes the product of two TF dissociation rates. This situation is likely to strongly increase TF-type selectivity, as in the mechanism of kinetic amplification of substrate discrimination (Hopfield, 1974; Ninio, 1986), but in the present case in equilibrium conditions. In the two latter schemes (Eqs. 17 and 18), the frequency of the TF-binding cycles does not interfere with the average transcription rate. When bound to DNA (average duration  $1/k_d$ ), the TF can contact the TCC after a waiting time of  $1/k_s$ , but somewhat counter-intuitively,  $k_{init}$  is independent of the  $k_s/k_d$  ratio, since even for very low  $k_s$ , the waiting time for TCC recruitment can be distributed over several TF-binding cycles. If the ratio between the frequency of the TF-binding cycles and the recruitment rate is not important for the average transcription rate, one should however keep in mind that it profoundly influences the steadiness of mRNA synthesis. When the TF-binding cycles are very slow relatively to  $k_s$ , transcription initiation would become burst-like (section 9.6).

### **9.2.3. Interaction between a DNA-bound TF and a pre-recruited proximal complex (RC2)**

In this scheme and the following one (section 9.2.4), a proximal complex (PC) is supposed to bind autonomously to the proximal *cis*-elements (TATA box, Inr). In the RC2 scheme, transcription can start only when the DNA-bound TF and PC interact each other through a first order micro-reversible reaction ( $k_3/k_{-3}$  in Fig. 3B) involving DNA folding. Similar equations are expected for slight variations of this scheme, such as when the PC, including GTFs, first binds to TFs and enhancers before interacting with the proximal promoter (Orphanides and Reinberg, 2002).

### **9.2.4. Irreversible conservative scheme, with conservation of the TF but destruction of the PIC (IC3)**

This particular case has been classified in the conservative modes of initiation since: i) transcription initiation is directly triggered by the DNA-bound TF and ii) the TF is not displaced upon initiation. Nevertheless, a destruction does occur but it concerns the basal transcription preinitiation complex. This scheme is likely to apply to heat shock stress and to describe the mode of action of HSF1, as explained in section 8.1. The transcription initiation rate expected from this hypothesis is similar to the general IC (section 9.2.1).

A burst-like regime is obtained for slow HSF1 binding cycles (low  $k_1k_{-1}/(k_1+k_{-1})$  ratio). Though it is not definitely proven, this mechanism would allow to reconcile two seemingly contradictory observations:

- **i)** Hsp gene promoters (targets of HSF1), have very consensual TATA boxes and Inr.
- **ii)** The TATA box appears unoccupied through conventional techniques during intense transcriptional activity.

According to the IC3 scheme, the reconstruction of the PIC is favoured by consensual promoter elements and then destroyed by functional interference with HSF1, in sharp contradiction with the classically admitted role of TFs in building basal transcription complexes.

### 9.3. Destructive initiation schemes

#### 9.3.1. One-shot TFs (ID1 and ID2)

The main feature of these modes of initiation, contrary to the previous ones, is that the average transcription rate is strikingly dependent on the TF-binding cycle duration. A single initiation of transcription is allowed per TF-binding cycle since the TF-DNA interaction is destroyed upon initiation. Two possible cases of one-shot TFs can be distinguished: those which are only displaced upon initiation and those which are destroyed upon initiation (single-use TFs). As explained in section 8.2, a single-use TF candidate is ATF6, whose activity is prevented if it cannot be degraded. VP16, the most potent TF so far identified, is also likely to belong to this category. Two formulations of  $k_{\text{init}}$  can be proposed for one-shot TFs.

**9.3.1.1. Competition between TF-induced initiation and TF dissociation from DNA (ID1)** The steady-state transcription rate ID1 in Fig. 3B, can be obtained using either the probabilistic waiting time approach or a steady-state method. The fundamental frame of this equation is the same that the classical Briggs and Haldane enzymatic reaction (Briggs and Haldane, 1925), in which the micro-irreversible catalytic step generates a reaction product letting the enzyme empty when leaving it.

**9.3.1.2. Initiation irrevocably following DNA-binding (ID2)** If TF binding to DNA systematically leads to a series of events directing the departure of Pol II, the resulting transcription rate is a simplification of the ID1 equation when  $k_s \gg k_1, k_{-1}$ . This situation can be obtained if transcription initiation is coupled to the displacement of the TF, by machineries already present on the promoter. For the TFs whose transactivation activity and degradation are coupled, this system can allow the simple programming of total transcription by single-use TFs (section 13.3). In practice, this mode of transcription initiation could be evaluated experimentally through its expected Fano factor approaching 1. Such an action of a TF, only based on an association step and without competition between transcription initiation and TF dissociation, is subject to a lack of specificity, which is compensated in the present scheme, by the need for a recognition by degradation and dissociation machineries.

#### 9.3.2. Destruction of a basal transcription complex upon initiation (RD1 and RD2).

This possibility is somehow related to IC3 in that a basal transcription complex is repeatedly broken and reconstructed at every reinitiation cycle. But while this reconstruction is autonomous and self-grounded on proximal promoter elements in the case of IC3, it is driven by DNA-bound TFs in the RD

case. The initiation-associated destruction can concern only the TRC (RD1) or both TF+TRC (RD2). In the RD2 scheme, the TF has an apparent behaviour of one-shot TF, but its dissociation can be secondary to that of the complex to which it is bound.

#### 9.4. Elongation following RNAP binding

The simple hit-and-run situation of section 9.3.1.2, applies to vegetative genes from *E. coli*, not regulated by specific TFs and for which the RNAP, associated to a  $\sigma$  initiation factor, is autonomous for DNA binding, so that transcription directly follows RNAP binding to the promoter. The molecular events following RNAP fixation such as double-helix opening, do not influence the transcription initiation rate since they are obligatory. The resulting transcription rate is identical to ID2.

$$k_{\text{init}}=k_a[\sigma/\text{RNAP}] \qquad \text{Eq. 19}$$

If the  $\sigma/\text{RNAP}$  complex is recruited from the cytoplasmic solution, the average rate is linearly dependent of the  $\sigma/\text{RNAP}$  concentration in the cell. This Poissonian rate does not explain the transcriptional bursts sometimes reported, which can result from the stages following initiation, such as spasmodic transcription elongation.

#### 9.5. Capping and clicking mechanisms

“Capping” components can prolong the half-time of complexes by inhibiting their depolymerisation, thereby generating relatively high steady-state activities with lower concentrations of constituent proteins. This mechanism can concern for example the connection between DNA-bound TFs and DNA-bound proximal complexes illustrated in the RC2 scheme (section 9.2.3), or the stabilization of the PIC. Once built, the PIC platform has been proposed to remain stabilized by Ku in the reinitiation model of (Woodard et al., 2001). A locking effect is of course also expected for covalent marking of DNA or nucleosomes (section 13.2), and chromosomal organization into functional domains (Gaszner and Felsenfeld, 2006).

Linking both sides of genes has also been proposed to favour sustained transcription reinitiation through Pol II recycling. Many genes encoding housekeeping proteins are TATA-less and expressed at low rate and with lower stochasticity. These genes are not dependent on potent TFs and could necessitate only Pol II recycling by linking transcription initiation and termination. If transcription elongation is straight, this mechanism is expected to render the transcription rate less stochastic and relatively independent of transactivator concentrations, by clicking gene expression in a constant regime. This mechanism appears convenient for housekeeping genes but it is also applicable

to quantitatively regulated genes whose expression can be conditionally induced and then established for given durations. In this case, the initiation and termination sites of the gene could be bridged only upon TF activation (El Kaderi et al., 2009) and the link between the 5' and 3' parts of the gene is dependent on the presence of the TF. The reiterative use of the same group of Pol IIs through recycling from termination to reinitiation, is consistent with the fact that transcriptionally active Pol IIs are sequestered at the gene locus, as suggested by the slow fluorescence recovery after photobleaching of fluorescent protein-tagged Pol II (Kimura et al., 2002).

### **9.6. Time distributions of transcription initiations**

Though the  $k_{\text{init}}$  functions of Fig. 3B are only based on Poisson processes, the resulting time distributions (vertical bars in Fig. 3B), are not exponentially distributed. They can generate transcriptional bursts and Fano factors higher than unity. This phenomenon can be responsible for the well established cell-to-cell heterogeneity of gene expression (Peccoud and Ycart, 1995; Chubb et al., 2006; Raj et al., 2006).

In the TF-driven modes of transcription initiation examined here, shortening the TF-binding cycles is a mean to overcome this problem (Ko, 1991). Short TF-binding cycles are however unable to counteract the bursting effects of non-exponentially distributed chromatin alterations and repression mechanisms described later.

## **10. Example of reversible combinations of activating TFs**

When the average transcription rate is proportional to fractional promoter occupancy (Eqs. 15 and 17), plausible equations of this occupancy can be defined for modeling combinatorial TFs actions. Only simple micro-reversible modes of promoter occupancy are considered in the examples listed in Fig. 4. Given that energy-driven and covalent modifications are in fact involved in certain activation (chromatin remodeling) and active repression systems (section 13), the micro-reversible modeling proposed here is based on the very approximate assumption that these micro-irreversible processes are rapid and obligatory. The consequences of time-irreversible processes will be examined in a forthcoming study.

### **10.1 A single enhancer DNA module recognized by a single TF (Fig. 4A, line 1).**

This is the most elementary hyperbolic saturation function, describing the hyperbolic responses to inducer concentrations or the logistic response to an exponential gradient (section 3.2).

### 10.2. Two different TFs can bind to the same DNA module (Fig. 4A, line 2).

Full promoter activation is obtained at high concentration of either A1 or A2. This time-share situation, frequent in the nuclear receptor family, applies when different TFs have the same DNA target site. For example, the mouse mammary tumor virus (MMTV) promoter can be activated by either the progesterone receptor (PR) or the glucocorticoid receptor (GR). Though this situation is verified experimentally, the simultaneous presence of high amounts of both receptors in the same cells is however doubtful.

### 10.3. Dimerisation is necessary for binding to the promoter (Fig. 4B, line 1).

This case, in which DNA binding activity only appears upon monomer to dimer transition, is related to the previous one since the DNA binding sites of dimers are often bipartite, made of two related half-sites arranged either in the same (direct repeats) or opposite orientation (inverted repeats). This very frequent organization concerns the bZIP, nuclear receptor and bHLH families. For TF homodimers with identical monomers contacting each other through the same interaction surface, the DNA module is often palindromic as expected. Conversely, direct repeats are possible for dimers linked together through a different surface, unless a flexible linker exists between the dimerisation domain and the DBD. The TF concentration-dependent transcriptional activation by a dimer is

$$Y = [PAA] / ([P] + [PAA]) \quad \text{Eq. 20}$$

where P is the promoter,

$$\text{with } [PAA] = [P] K_{DNA} [AA] \quad \text{Eq. 21}$$

$$\text{and with } [AA] = K [A_F]^2 \quad \text{Eq. 22}$$

where K is the dimerisation constant.

Hence,

$$Y = \frac{K_{DNA} K [A_F]^2}{1 + K_{DNA} K [A_F]^2} \quad \text{with } [A_F] = (\sqrt{1 + 8K[A]} - 1) / 4K \quad \text{Eq. 23}$$

where [A] is the total concentration of A diffusing in the nucleosol. TF multimerisation prior to DNA binding is a particular case of cooperativity, in which the mutual aid between TFs, occurs in solution rather than upon fixation to DNA. The square concentration in Eq. 23 indicates that TF dimerisation is a simple way to generate useful sigmoidal saturation curves (Fig. 5), with an inflection point at  $[A_F] = 1/\sqrt{3} \sqrt{K_{DNA} K}$  and capable of generating sensitive responses (section 16).

#### 10.4. Independent binding but simultaneous presence necessary to induce transcription (Fig. 4B, line 2).

This "both" system, mentioned here because it is often proposed in the literature, relies on the hypothesis of non-additivity of the transactivation of the different TFs (section 7.2.2.2). The sigmoidicity generated in this situation, with an inflexion point at  $[TF] = K_d/2$  and reminiscent of condensation enzymes whose reaction can occur only upon simultaneous filling of the two substrate binding sites, can be distinguished from the other types of cooperative sigmoidicity through the atypical shape of the Hill plots (Fig. 5). The fractional occupancy of this system shown in Fig. 4B, line 2, can be alternatively obtained through a sequential approach. In the case of  $n$  identical TFs whose concomitant fixation is necessary to trigger transcription,

- $[P*TF] = n K[P][TF]$  Eq. 24a

- $[P*TF_i] = ((n-i+1)/i) K[P*TF_{i-1}][TF] = \binom{n}{i} [P](K[TF])^i$  Eq. 24b

- $[P*TF_n] = (1/n) K[P*TF_{n-1}][TF] = [P](K[TF])^n$  Eq. 24c

Hence,

$$Y = \frac{[P*TF_n]}{\sum_{i=0}^n [P*TF_i]} = \left( \frac{K[TF]}{1 + K[TF]} \right)^n \quad \text{Eq. 25}$$

#### 10.5. Reciprocal binding cooperativity (Fig. 4B, line 4).

This system, in which the first binding step favours the second one by an interaction constant  $K_C$ , is the most classical mode of cooperativity encountered in the literature.

#### 10.6. Hierarchical binding cooperativity (Fig.4B, line 6).

The binding of a TF can be dependent on the previous binding of another one. This dependence can involve direct protein-protein interactions, such as in the case of the herpes simplex virus protein VP16 which can for example bind to DNA-bound Oct-1. But hierarchical cooperativity is most often indirect. The indirect dependence between different TFs, mediated by DNA distortion or chromatin movements, has several interesting features:

- It can generate strongly inducible expression with non-inducible, constitutively active TFs.
- It does not require direct protein-protein contacts between the cooperating TFs.

Hierarchical promoter occupancy can concern either the same TF (Lacal et al., 2008) or different types of TFs (Archer et al., 1992; Chávez and Beato, 1997). Hierarchical equilibration has been proposed in the very different context of hemoglobin oxygenation (Michel, 2008) but it is likely to be relatively

rare for carriers and enzymes. Conversely, given the flexibility of DNA and the existence of chromatin remodeling factors, this phenomenon could be very frequent for DNA occupancy. The indirect influence between the sequential binding steps postulated in the hierarchical equilibration system, can be mediated by different mechanisms listed below.

### 10.6.1 Hierarchical equilibration mediated by DNA allostery

The TodT TFs bind to the three DNA modules present in the Tod gene of *Pseudomonas putida* in a hierarchical manner (Lacal et al., 2008). Since the TodT monomers do not interact one another, this ordered occupancy of the promoter should involve conformational changes of DNA. In this case, the Adair equation, which can correspond to a sigmoid (Fig. 5C), is:

$$Y = \frac{K_1[A] + 2K_1K_2[A]^2 + 3K_1K_2K_3[A]^3}{3(1 + K_1[A] + K_1K_2[A]^2 + K_1K_2K_3[A]^3)} \quad \text{Eq. 26}$$

### 10.6.2 Hierarchical equilibration mediated by chromatin remodeling

The MMTV promoter has been widely used as a model for studying the role of chromatin remodeling in transcription. The initial binding of the GR induces a re-positioning of the nucleosome, exposing the DNA modules of NF1, thereby allowing NF1 to access DNA (Archer et al., 1992; Chávez and Beato, 1997). This system of induction through a non-inducible TF, amplifies the glucocorticoid hormone action on MMTV through the participation of NF1 which is constitutively active. The role of nucleosome positioning is general in eukaryotic transcription (Polach and Widom, 1996; Vashee et al., 1998; Raveh-Sadka et al., 2009). Nucleosomes are positively charged beads, because of the richness in basic amino-acids of the histone proteins. The wrapping of negatively charged DNA around nucleosomes induces strong curvatures. Hence, since all local DNA sequences are not equally prone to bend, the positioning of the nucleosomes is mainly determined by DNA sequence. When one particular positioning is strongly favoured thermodynamically, a phasing is observed between the cells, yielding a clear ladder pattern in footprinting experiments starting from cell populations, as in the case for the MMTV promoter. This mode of transcriptional regulation of TFs could be very important in eukaryotes considering the abundance of chromatin remodeling factors (Rippe et al., 2007), and can be extended to the more general principle of chromatin pre-setting. In addition to chromatin remodeling, pioneer TFs can indirectly regulate the responsiveness of genes by recruiting enzymes that modify the chromatin architecture, but these aspects will be described later since they also concern repression mechanisms.

## 11. Independent DNA module occupancy

Independent fixation of two activators A1 and A2 (Fig. 4B, line 3) generates a concentration-dependence curve resembling the classical hyperbolic saturation if the two TFs have roughly the same activity. Let us now examine the case of a series of independent DNA modules.

### 11.1. Independent occupancy of a series of DNA modules, with possible full-time residence

For easier treatment, a series of  $n$  DNA modules for the same TF is examined. If every individual saturation is hyperbolic, the global fractional saturation of the cluster is

$$Y = \frac{1}{n} \sum_{i=1}^n \frac{K_i[\text{TF}]}{1 + K_i[\text{TF}]} \quad \text{Eq. 27}$$

This sum of hyperbolas can be rearranged into a single ratio

$$Y = \frac{\sum_{i=1}^n i \left( \sum \prod_{i/n} K \right) [\text{TF}]^i}{n \left( 1 + \sum_{i=1}^n \left( \sum \prod_{i/n} K \right) [\text{TF}]^i \right)} \quad \text{Eq. 28}$$

where  $\prod_{i/n} K$  is the product of the circular permutations of  $i$   $K$  out of  $n$ , whose number is  $\binom{n}{i}$ .

### 11.2. Binding to the $n$ modules with the same affinity

When all the binding sites are identical (with the same  $K$ ) and equally accessible, Eq. 28 becomes

$$Y = \frac{\sum_{i=1}^n i \binom{n}{i} (K[\text{TF}])^i}{n \left( 1 + \sum_{i=1}^n \binom{n}{i} (K[\text{TF}])^i \right)} \quad \text{Eq. 29}$$

which reduces to the elementary hyperbola (Eq. 3). This reduction can appear strange if using the classical sequential Adair approach with statistical balancing of microscopic equilibrium constants, but is obviously expected here when starting from a sum of hyperbolas (Eq. 27). The shape of the fractional curve does not depend on the number of binding sites if they are independent and equivalent.

### 11.3. Binding of TFs to the $n$ modules with very different affinities

If certain sites have a higher affinity for the TF than others, reduction of Eq. 28 is no longer possible and a seemingly anti-cooperative response appears (Fig. 5D). The envelope of this curve is an hyperbola shifted to the left. This curve could reflect the regulation of a gene expressed at high level at

low TF concentration, but which can be further stimulated. This anti-cooperative-like function obtained with completely independent binding sites, shows that non-hyperbolic behaviours do not necessarily imply the existence of mutual influences between the different binding steps, as sometimes claimed to emphasize the superiority of cooperativity models enclosing negative cooperativity.

## **12. Master TF vs collective TF gene regulation**

Two parameters are differently affected by the number of participating TFs: quantity and stochasticity.

### **12.1. Single TF decision but collective strength**

Depending on the cooperative system, gene expression can be triggered by either a single (master TF) or a combination of TFs. The decision of gene expression is particularly stringent when ensured by a unique, regulatable master TF. But unique or multiple triggering TFs have the same overall resulting strength if the master TF induces a cascade of secondary TFs amplifying its effect through hierarchical cooperativity.

### **12.2. Concentration-dependent vs time-dependent determinism**

A dramatic difference between master TF and collective TF activation, is the degree of stochasticity of gene expression, which is not apparent in steady-state representations, but is of primary importance for the cell. Stochasticity of promoter occupancy can be lowered by two means: i) either by increasing the number of linked DNA modules or ii) by accelerating the TF exchanges with a single DNA module. These two mechanisms allow to minimize the binary status of the occupancy of every individual DNA module, by averaging many DNA modules, or events respectively. Only the latter mechanism is possible for master TF regulation such as in the case of GR-mediated regulation of the MMTV promoter (Fig. 4B, line 5). This could explain the high frequency of interaction observed between GR and the MMTV promoter (Sprague et al., 2004).

## **13. Transcriptional repression**

Two main types of transcriptional repression can be distinguished. Repression can be i) graduated and participate to the dosage of gene expression, or ii) binary and dictate the transcriptional competence of genes. More frequently, repressors act by preventing the action of activators rather than by lowering their activity. This inhibitory action can be direct, through competition for binding to the same DNA modules, or indirect through chromatin modifications.

### 13.1. Graded repression

This mode of transcriptional repression, contrary to activation, is not a self-sufficient activity but is a tributary of a stimulated state. Hence, the probabilistic formulation  $x_{1/2}/(x_{1/2} + x)$  alone is meaningless since it would mean that the gene is fully activated in absence of the repressor, which is rarely the case.

#### 13.1.1. Competition for binding to DNA (Fig. 6, line 1).

Though this study is not focused in prokaryotes, bacteria should be reminded in this context since in these organisms, repressors are classically reported to work through preventing TFs or RNAP to bind to DNA. The extensively studied archetypal example is the lactose operon and its repressor LacI. In the regulatory region of this operon, the DNA-binding sites for the RNAP, called promoter (P), and that of LacI, called operator (LacO), are overlapping, so that the binding of RNAP and of LacI are mutually exclusive. This type of bacterial repressor has generally a low dissociation rate, leading to long interaction cycles. Slow stochastic cycles generate a high degree of intercellular heterogeneity, particularly appropriate for increasing the adaptability of unicellular populations to changing conditions. In addition, the binding to DNA of the repressor molecule is often conditional and involved in bistability. For the lactose operon, the repressor LacI is converted into a non-DNA binding form upon allosteric fixation of a beta-galactoside inducer. The opposite can be encountered for other operons. For example the DNA-binding activity of the tryptophane operon repressor is stimulated by fixation of a co-repressor. Fine regulations of transcription initiation in bacteria are not described here but are available in (Minchin and Busby, 2009). Competition for DNA-binding sites also exists in eukaryotes, but the role of repressor is more generally ensured by variant isoforms of TFs, instead of pure repressors structurally unrelated to TFs.

#### 13.1.2. Passive repression: dominant-negative TFs (Fig. 6, lines 2-3)

In the lines 2 and 3 of Fig. 6,  $[A_F]$  and  $[R_F]$  are the concentrations of free activator and repressor capable of dimerizing together.  $K_{DNA}$  is the DNA-binding affinity of the dimer and  $K$  is the dimerisation constant. If  $[A]$  and  $[R]$  are the total concentrations of activator and repressor circulating in the nucleosol,  $[A_F]$  and  $[R_F]$  are such that

$$[A] = [A_F] + 2[AA] + [AR] \quad \text{Eq. 30}$$

$$\text{with } [AA] = K[A_F]^2 \quad \text{Eq. 31}$$

$$\text{and } [AR] = \frac{K[A_F]}{1+K[A_F]} [R] \quad \text{Eq. 32}$$

The real number solution of this system is

$$[A_F] = Q^{1/3} + \frac{1+2K([A]-[R])}{12K^2} Q^{-1/3} - \frac{1}{2K} \quad \text{Eq. 33a}$$

$$[R_F] = \frac{[R]}{1+K[A_F]} \quad \text{Eq. 33b}$$

$$\text{with } Q = \frac{\sqrt{\Delta}}{24\sqrt{3}K^3} + \frac{[A]+[R]}{8K^2} \quad \text{Eq. 33c}$$

where

$$\Delta = 8K^3([R]-[A])^3 + 3K^2(5[R]^2+5[A]^2+26[A][R]) + 6K([R]-[A]) - 1 \quad \text{Eq. 33d}$$

This result is uselessly complicated since it is still a strong simplification based on the assumption of well-mixed cellular contents. Dominant-negative transcription-incompetent TFs, also called poison partners, can repress transcription through heterodimerizing with active TFs (Fig. 6, line 2). For example, ICER, a small truncated isoform of bZIP TF, can inactivate its long counterpart CREM by this way (Stehle et al., 1993). This sequestration mechanism is proposed to generate ultrasensitive responses (Buchler and Cross, 2009). Dominant-negative TFs, directly encoded as truncated proteins or generated by alternative splicing, can abrogate the fixation to DNA of their full length partners (Fig. 6, line 3). This is the case for Id, that contains the dimerisation domain to interact with MyoD but lacks a functional DNA-binding domain (Benezra et al., 1990). Examples of this mode of regulation are also found in the nuclear receptor family (Yen and Chin, 1994). Sequestration can apply directly to a TF, but also to a coactivator. In this respect, the participation of common coactivators to multiple transcription signalling pathways seems to have been selected as a mean to coordinate and limit the overall transcriptional activity in the cell. This is the case of CBP, also called integrator, present in limiting concentration in the cell and required for the transcriptional activity of several types of TFs, from the bZIP family (AP1), NFκB and nuclear receptors (Kamei et al., 1996). TBP is also considered as a global integrator since it is involved in almost all transcription initiation complexes and its concentration is also limiting in the cell. The graded modes of repression through DNA occupancy described above are fully micro-reversible, contrary to the following ones.

### 13.1.3. Active repression (Fig. 6, line 4)

Certain eukaryotic TFs have an intrinsic repressive activity by preventing the participation of other DNA modules in the promoter. An example is provided by unliganded nuclear receptors such as the thyroid receptor (TR). TR can shift from a role of activator, when stimulated by thyroid hormone, to a role of repressor in absence of the hormone. The repressive action is mediated by co-repressors such as SMRT or N-coR, which can deacetylate histones (enzymatic activity called histone deacetylase HDAC). Chromatin deacetylation then prevents the binding of other TFs by tightening the electrostatic

contact between DNA (negatively charged) and nucleosomes (positively charged). The list of post-translational modifications of histones influencing the transcriptional competence of chromatin, is not restricted to acetylation and can also involve coordinated methylation, phosphorylation, and mono-ubiquitination (Kouzarides, 2007). Chromatin-mediated repression can be reversed. Deacetylated nucleosomes can be re-acetylated, either after fixation of the hormone to the nuclear receptor, or thanks to the appearance of an additional TF capable of binding to deacetylated chromatin and recruiting histone acetylases called histone acetyl transferase: (HAT), with an activity opposite to that of HDAC. In principle, the simultaneous binding of factors capable of recruiting HAT and HDAC would lead to conflicting influences. Chromatin “openers” are not necessarily classical TFs with TADs. The scheme of Fig. 6, line 6, represents a chromatin opener A1 with no intrinsic transactivation activity, counteracting the repressing action of a chromatin compactor R. Another way to actively repress transcription, not based on DNA occupancy, is to disrupt the interaction between DNA-bound activators and the transcription machinery. For example, the transactivating activity of Sp1 can be abrogated by the E2F-HR complex (Zwicker and Müller, 1997) (Fig. 6, line 4, blunt arrow).

### **13.2. Gene occlusion through chromatin condensation**

Certain repression systems appear more specific of eubacteria and others of eukarya. Noticeably, the importance of chromatin in eukaryotes has changed the principles of transcription regulation, which is more based on the transcriptional competence of the genes than on classical transcription activation. The equations of Fig. 6 do not apply to genes shut off through long-term covalent locking.

#### **13.2.1. Long-term heterochromatinization**

The reversible formation of facultative heterochromatin is a frequent mode of repression in higher eukaryotes, particularly for reducing certain genes to silence. Heterochromatin is a transcriptionally-incompetent form of DNA compaction, regulated by a complex interplay between DNA and histone modifications (Rottach et al., 2009). The picture is further complicated by the existence of several variants of linker and core histones, certain of which interfering with chromatin remodeling (Altaf et al., 2009). Heterochromatinization can be triggered by enzymatic methylation of cytosine residues in the context of the 5'-CG-3' dinucleotides. Facultative heterochromatinization is assumed to close the genes whose expression is no longer desirable in the course of cell differentiation. By this way, the number of accessible binding sites for a given TF in a given cell state, would be much lower than predicted from the genome sequence, thus solving in part the question of the available concentration of diffusing TFs (section 3.1). The three fulfilled following conditions: 1) the palindromic structure of the 5'-CG-3' dinucleotides, 2) the enzymatic activity of the DNA methylase DNMT1 which can methylate hemimethylated sites but not non-methylated sites, and 3) the semi-conservative mechanism of DNA replication, concur together to render DNA methylation potentially appropriate and sufficient for

transmitting the embryonic methylation profiles over cellular generations. By this way, DNA methylation would appear as an elegant and economical way of management of large genomes, allowing the stability of cell fates over time and avoiding the synthesis of myriads of classical repressors all along the organism's life. This elegant view has however been recently questioned (Huang, 2009), as is the causal role of chromatin marks on gene occlusion (Lee et al., 2009). In fact, DNA methylation appears redundant with non-covalent intergenic constraints for stabilizing cell fates (Huang, 2009). In addition, covalent epigenetic marking is very transient and reversible, for DNA methylation (Kangaspeska et al., 2008) as well as for histone modifications (Trojer and Reinberg, 2006). Precarious epigenetic marking is also suggested by the old experiments of somatic cell fusion, in which numerous gene activations and occlusions were observed even in absence of DNA replication (Mével-Ninio and Weiss, 1981; Chiu and Blau, 1984; Chin and Fournier, 1987). The fact that DNA methylation can be erased after cell fusion in absence of cell division (Zhang et al., 2007), shows that demethylation is a directed activity, not merely due to a defective methylation of newly synthesized DNA strands. One can suppose that that the set of genes reactivated after cell fusion are those for which *trans*-acting reprogramming factors are provided by the foreign cytoplasm. According to this view, the egg cytoplasm should be particularly enriched in such factors to activate early embryonic genes and explain the success of nuclear transfer experiments. Its instability renders DNA methylation poorly appropriate for ensuring long-term memory and accordingly, the methylation profiles become heterogeneous between cells deriving from a single precursor (Silva et al., 1993). In addition, DNA methylation can occur in non-CG contexts which are less appropriate for the perpetuation of methylation profiles, and many other roles have been attributed to DNA methylation, including several aspects of chromatin architecture not necessarily related to transcriptional repression (Kaneko et al., 2004), the position of nucleosomes, transcription elongation rate and splicing (Choi et al., 2009), DNA repair (Wang and James Shen, 2004), replication, chromosome stability, or the mutational divergence of DNA repeats (Krickler et al., 1992). These pleiotropic roles blur the simple causal implication of DNA methylation in gene occlusion. Finally, the control of DNA methylation involved in transcription, is ultimately ensured by TFs capable of targeting to specific DNA sites, *de novo* DNA methyltransferases or, conversely, demethylation machineries. Hence, if the results about instable methylation are not biased by the fact that they are obtained using cancer cell lines with erratic epigenomes, DNA methylation should perhaps be no longer considered as a primary upstream regulator of the TF action, as initially assumed, but rather as an additional device in the TF toolbox, allowing the pioneer TFs to continuously reconfigure covalent epigenetic patterns.

### **13.2.2. Mitotic chromatin compaction**

Dynamic gene networks are generally envisioned in the simplified context of a well mixed cell content, with unbroken interactions avoiding to repeatedly reset the initial conditions. This is clearly not the case for dividing cells because of the highly condensed state of chromatin at mitosis,

unsuitable for gene expression. Mechanisms can be proposed to ensure short-term memory during mitosis and perpetuate preset programs across cell generations.

**13.2.2.1. Memory of the set of genes to be repressed.** Nucleosomes are not covalently linked to DNA, but mechanisms exist to retain nucleosomal marks during mitosis, since nucleosome composition and modifications can survive their partial disassembly during DNA replication (Probst et al., 2009). Though its capacity to imprint long-term memory has been questioned (section 13.2.1), DNA methylation appears ideally stable during cell division. DNA methylation and insulators, supposed to determine chromosomal domains, are clearly involved in the phenomenon of parental imprinting (Gaszner and Felsenfeld, 2006). Hence, it could also preserve the non-expression information across mitosis and meiosis.

**13.2.2.2. Memory of the set of genes to be expressed.** In addition to escape repressive marking, active genes can be prepared for immediate re-expression following cell division. While most TFs are no longer capable of reaching DNA when buried in the highly condensed mitotic form of chromatin, some ones have the remarkable capacity to remain bound to mitotic chromosomes, thereby allowing the cell to rapidly identify the genes to be reexpressed after mitosis and chromatin decompaction (Michelotti et al., 1997). This is the case of HSF2 (Xing et al., 2005) for stress genes whose rapid expression is crucial for cell viability (John and Workman, 1998). More generally, TBP can also behave as a genome bookmarker (Chen et al., 2002), through inhibiting condensin near active promoters (Xing et al., 2008). One can also cite the case of Pol I in rDNA (Scheer and Rose, 1984).

**13.2.2.3. Mitotic reprogramming.** Chromatin disassembly during replication can provide a window of opportunity for certain proteins to reach DNA. It has recently been suggested that important cellular decisions are taken during cell division. The memory of the gene expression competence status is preserved during mitosis, but at the opposite, it can also be reprogrammed during this period (Zhang et al., 2002; Hiratani and Gilbert, 2009). Correlations have been observed between the timing of replication and the types of chromatin present in the different chromosome regions, but the causal relationships behind these observations are far from clear. Cell fusion experiments suggest that mitosis is not an obligatory checkpoint for gene expression changes (Chiu and Blau, 1984; Zhang et al., 2007).

### **13.3. Programmed cessation of gene expression**

The transcriptional induction of certain genes should be transient. Stress genes are finely regulated and strongly inducible upon cellular disturbances, in a reparative attempt. An important feature of stress genes is their capacity to be expressed transiently at high level, in proportion to the intensity of the deleterious conditions, and then to rapidly return to basal levels once the insults are repaired. This recovery is important since in absence of disordered substrates, sustained overexpression of certain anti-apoptotic stress gene products is a risk of cancer promotion. Recovery can be achieved through

quite diverse mechanisms, including classical negative feed-back loops. For example, after an heat shock stress, HSF1 is sequestered by proteins derived from HSF1-dependent genes (Shi et al., 1998). In this system, HSF1 is reversibly inactivated but remains usable in future heat shock responses. An alternative and more original mechanism is provided by single-use TFs, appropriately called kamikaze TFs (Thomas and Tyers, 2000), whose transcriptional activity and degradation are coupled. This is the case for ATF6, a bZIP TF essential for inducing endoplasmic reticulum (ER) stress gene, presynthesized in normal conditions but inactive because it is anchored in the ER membrane and prevented to enter the nucleus. Upon ER stress, a given amount of ATF6, quantitatively depending on the intensity of the stress, is cleaved and released from the ER membrane. This truncated form of ATF6 (tATF6) immediately stimulates the expression of target ER stress genes (Shen and Prywes, 2005). ATF6 cleavage results from a desequestration mechanism, in which the ER chaperone BiP is diverted, thereby allowing the translocation of ATF6 in the Golgi apparatus and its processing (Shen and Prywes, 2005). The initial burst of tATF6 determines the level of tATF6-dependent ER stress gene expression. The transcription initiation rate proposed in section 9.3.1.2 (scheme ID2), is

$$k_{\text{init}} = k_a[A] \quad \text{Eq. 34}$$

where  $[A]$  is the concentration of tATF6. This rate progressively diminishes with time since every tATF6 molecule is degraded when initiating transcription. For a single responsive gene in a compartment of volume  $V$ ,

$$k_{\text{init}} = k_a[A]_0 e^{-\frac{k_a t}{V}} \quad \text{Eq. 35}$$

In addition, tATF6-dependent transcription is cut short through a negative feed-back mechanism involving, once again, the ER chaperone BiP. BiP first acts as a sensor of the ER stress since it is responsible for the release from the ER membrane of a certain amount of ATF6 upon stress. Given that BiP is itself an ER stress gene, its own induction allows: i) to repair the malformed ER proteins at the origin of the stress and ii) to massively sequester the ATF6 precursor at the ER membrane, thereby cancelling the release of tATF6 (Shen and Prywes, 2005).

## 14. Different subsets of TF behaviours

### 14.1. TFs differentially affected by *cis*-repressive conditions

The specificity of combinatorial regulation of promoter occupancy is favoured by the diversity of TF behaviours, with respect to their differential ability to recruit transcription machineries and their different susceptibility to chromatin configurations.

#### **14.1.1. TFs capable or not to bind to compacted chromatin**

The reversibility of heterochromatinization suggests the existence of pioneer TFs capable of accessing it and initiating the reversion to transcription-competent chromatin. This possibility is illustrated by the capacity of a master TF to convert a fibroblast into a myoblast (Davis et al., 1987) and by TFs capable to overcome epigenetic *cis*-repression (Koutroubas et al., 2008). This activity must not be confused with the phenomenon of gene bookmarking mentioned above (section 13.2.2.2), which is a local inhibition of chromatin compaction by pre-bound TFs.

#### **14.1.2. TFs capable or not to bind to nucleosomal DNA**

The position of nucleosomes, regulated by both DNA sequence and the activity of SWI-SNF ATPases, is of primary importance for regulating promoter occupancy (Raveh-Sadka et al., 2009). If considering the example of the MMTV promoter, the initiating role of GR and the amplifying role of NF1 (section 10.6.2), are explained by their differential capacity to bind to chromatin. GR contacts one side of the double-helix so that it can bind to DNA wrapped around nucleosomes, whereas NF1 has to surround the double-helix, preventing it to bind to nucleosomal DNA (Piña et al., 1990; Eisfeld et al., 1997). Finally, the balance between chromatin-mediated repression or competence, can be triggered by different TF heterodimeric combinations and the mechanisms listed in Figs. 4 and 6 can be mixed in the same promoter regulation.

### **14.2. Different possible roles of short TF-DNA binding cycles**

#### **14.2.1. Maintenance of free TF concentrations**

Certain TFs can access to many binding sites in the nucleus, thus lowering the nucleosolic concentration of free TFs if they are not expressed at high level. Their rapid and active dissociation could restore a sufficient amount in the soluble fraction, as proposed for TBP (Muldrow et al., 1999). This conclusion is supported by the observation that the lethality caused by a defect of Mot1, the enzyme responsible for the displacement of TBP, is rescued by overexpression of TBP (Auble et al., 1994).

#### **14.2.2. Reactivity to signal fluctuations**

This possibility has long been proposed for hormone-responsive TFs whose rapid recycling allows to adjust their activity to rapid variations of their ligands, such as in the case of GR-mediated gene

expression in response to the cycling of cortisol levels (Freeman and Yamamoto, 2001). This view is supported in (Stavreva et al., 2009).

#### **14.2.3. Noise reduction**

This possibility, first pointed by (Ko, 1991), can mainly concern the genes whose dosage is critical for the organism's health and development. Candidates genes are those involved in dosage pathologies such as trisomies, and the genes regulated by single master TFs.

#### **14.2.4. Coupling TF dissociation and transactivation**

TF dissociation could be a prerequisite for Pol II escape from the promoter. This interesting idea whose underlying molecular mechanisms remain to be elucidated, has been proposed in the context of kamikaze TFs (Thomas and Tyers, 2000). Destructive transcription initiation could concern intensely recycling TFs, but also the GTFs in the case of stable upstream TFs. For instance, no stable PIC can be detected on the promoter of an active heat shock gene responsive to HSF1 (Lebedeva et al., 2005). In this respect, the high dynamics of TBP, the pivotal component of the PIC, also seems necessary for the expression of a subset of genes in yeast. TBP-promoter interactions can be regulated and remodeled by several factors such as Mot1 and NC2. This would explain the dual role of the TBP-displacing enzyme Mot1, which is, depending on the cases, classified as a repressor or conversely, as an activator (Dasgupta et al., 2005).

#### **14.2.5. Concomitant binding to mutually exclusive sites**

Interestingly, DNA elements whose occupation are mutually exclusive, can be all necessary for gene activation (Schaufele et al., 1990). This puzzling behaviour, interpreted by invoking sequential steps in the process of transcription initiation, could also be envisioned as a single time-share process. This seemingly paradoxical goal can be functionally achieved if the mutually exclusive binding events, rapidly equilibrate. Conversely, to make the occupations really mutually exclusive, as observed for the bacterial repressor systems (section 13.1.1), the sticking time of the repressor should be long enough.

This non-limitative list of possible functions shows that it is virtually impossible to assign precise roles to the increasing number of reports about TFs with short DNA interaction cycles. This example illustrates the difficulty to identify relevant observations among the lot of interactions existing in the cell. Exhaustive quantitative modeling of gene expression is obviated, not because of the huge number of molecular actors involved, whose incorporation in the system is only limited by the available calculation power, but because of our ignorance. On the one hand, the difficulty to discern relevant causal relationships in the abundant postgenomic data, could be bypassed by objectively treating the system as a whole and drawing interpretations only from the resulting solutions; but on the other hand, many parameters necessary to this end remain elusive.

## 15. The essential condition of time scale separation

This condition has been used several times in this study. An interest of time scale separation is to solve the problem of stochasticity inherent to low copy number molecules when ensemble averaging no longer holds. This point is essential in the field of gene expression since genes are present in only one or two copies per cell. The probabilistic treatment of biochemistry is now widely admitted and used in the literature, but except in the quantum world, it remains that probability is definitely not reality. Hence, even if it is hardly scalable to the size of eukaryotic cells, the discrete treatment of gene networks is often proposed as a convenient alternative to smooth differential equations, since it appears appropriate to account for random cellular phenotypes. In fact, when applicable, the condition of time scale separation allows to equate probability and reality (Michel, 2009), thus restoring their full relevance to continuous differential equations which accurately describe probability evolutions. The participation of this mechanism is suggested by several observations. A direct way to appreciate intercellular heterogeneity is provided by the spatial response to a signal gradient, such as that of BCD. Though the bitmap mode of BCD-response would have been fully convenient to generate a sharp boundary of HB expression, the images obtained for the BCD-responsive gene product HB detected by *in situ* mRNA hybridization, immunochemistry, or expression of a genome-integrated reporter gene driven by a synthetic BCD-responsive promoter, clearly show halftone patterns at the boundary of BCD-response, with nuclei presenting all the intermediate intensities of staining. This example shows that the binary fixation of BCD to DNA can be converted into a graded response (bitmap vs greyscale patterns in Fig. 1C). Though widely used in biophysical studies, the locution time scale separation can appear somewhat terrific and could be replaced by differential kinetics. It is also known as the principle of rapid pre-equilibrium in enzymology, when the rate of the catalytic step is much lower than those of substrate exchanges. In the context of the present study, this principle is important to generate smooth Adair curves and for destructive initiation schemes.

Finally, an overlooked aspect of time scale separation is its role in biological accuracy. To various extents, the time-share situation described in this report is likely to be very general. Any accessible molecular site can be frequently but shortly engaged by a wide variety of diffusing ligands. A principle of kinetic accuracy is that the correct and false ligands should equilibrate with the macromolecule before firing the micro-irreversible reaction, to allow their discrimination based on the off rates.

## 16. Sigmoidicity, specificity and dosage of gene expression

Certain combinations of transcription regulators described above can generate sigmoidal transcriptional responses (Fig. 5 for curves in 2D; right column of Figs 4 and 6 for surfaces in 3D).

The sigmoidal shape of response curves is of utmost physiological interest, when compared to hyperbolic responses, to ensure the specificity and dosage of gene expression. i) the low slope near the origin allows to buffer stochastic fluctuations of activators, thus avoiding non-specific expression; ii) steep sigmoids create threshold effects, such that gene expression governed by different types of TFs is triggered only when several conditions are reunited; and iii) soft sigmoids allow to finely graduate the response. The position of the roughly linear section of the curve around the inflection point, relatively to the concentration range of activators, is critical for optimizing the response. Combinatorial interactions with gene promoters have been proposed to participate to transcriptional heterogeneity (Voss et al., 2009) but in turn, they can also ensure, when coupled to time scale separation, precisely graded expression. The examples of sigmoidicity shown here are generated in a micro-reversible fashion (cooperativity “through space”), thanks to the juxtaposition of multiple *cis*-elements in a promoter; but sigmoidicity can also result from driven cooperativity, not described here.

## 17. Conclusions

Despite the number of reports on TFs, the very physical principles of transcription initiation remain largely obscure, and the wealth of descriptive observations has to be integrated in a coherent conceptual framework to facilitate their analysis and to plan rational future developments. The available data already indicate that the mechanism of transcription initiation is not unique. Hence, a comprehensive physical picture of transcription initiation will emerge from the comparative examination of different genes rather than from the functional dissection of a single model gene, as thorough as it can be. Differences between the transcription initiation schemes, allow to perceive how they have been selected in harmony with the function of the derived gene products, such as for example, strong but restrained in time for rapidly accumulating stress proteins, finely tuned for concentration-dependent regulatory proteins, or erratic to generate intercellular heterogeneity, propitious for cell fate determination during embryogenesis (Chang et al., 2008), etc. The present essay is aimed at proposing a tentative classification of the different categories of transcription initiation schemes into which the different genes are likely to fall into. The few formulas proposed here could be incorporated in the production functions of the differential equations (unfortunately analytically unsolvable), used for the quantitative modeling of gene networks, instead of the traditional but irrelevant Hill function.

This study also points to disparate reports to be reconciled and to fundamental questions remaining open. A major uncertainty concerns the precise role of TADs in transcription initiation. The only source of transcriptional cooperativity between several TF currently retained is based on DNA occupancy, in which the TADs are not involved. Several arguments are proposed here to support this view, such as the poor variety of the TADs suggesting a time-share mechanism unfavorable to the synergistic recruitment of transcription complexes. But reports clearly show that components of the

PIC are conjointly recruited by different TFs (Liu et al., 2009). This recruitment could concern only the stabilization of a basal transcriptional platform but not a transcription-triggering complex. These points are intriguingly eluded in all mechanistic studies on transcription initiation synergy.

The time distribution of transcriptional events attracted much more interest than the average frequency of transcription initiation (Métivier et al., 2003; Chubb et al., 2006; Raj et al., 2006; Degenhardt et al., 2009; Stavreva et al., 2009; Voss et al., 2009). But once again in this aspect of gene expression, the underlying events remain uncertain since resembling behaviours can arise from different mechanisms. Several transcription initiation schemes are compatible with the clustering of initiation events into finite windows. The loose notions of transcriptional pulses, cycles or bursts, widespread in the current literature, can reflect either: i) TF exchanges with accessible DNA, ii) periodic windows of chromatin accessibility, iii) threshold effects generated by steep sigmoidal responses whose inflection point corresponds to fluctuating ranges of TF concentrations, or iv) slow cycles of capped complex formation at low concentration of constituents (section 9.5). Hence, beside imaging and descriptive technologies, hypotheses-driven investigations could be helpful for the progress of knowledge about gene expression.

The gene expression profile of a given cell is sometimes considered as a series of 0 and 1 because this approximation is suitable for Boolean approaches and for computerization, but in eukaryotes, gene expression is not a binary but a graded activity. In addition to the open/closed gene status, the accurate dosage of active genes is also crucial in multicellular eukaryotes. A near-binary gene expression can appear acceptable for bacteria, since numerous reports convincingly describe stochastic bacterial phenotypic switches between repressed and non-repressed states. This situation is favourable to the adaptability of populations of unicellular organisms to changing environments, but it is possible only when a certain level of cellular waste is allowed. This is the case for bacteria, yeast or proliferating metazoan cells. But the degree of heterogeneity could be overestimated in mammals when evaluated using cultured cell lines instead of normal tissue cells. Indeed, cell-to-cell heterogeneity is high in multipotent progenitor cells but decreases upon differentiation (Chang et al., 2008). Moreover, when measured at the level of a healthy post-mitotic tissue, the degree of cell-to-cell variation of gene expression appears relatively low and then increases with age (Bahar et al., 2006). Elucidating the parameters governing the average frequency and time distribution of transcription initiations, will be critical for understanding the fundamental bases of pluricellularity embodied in the complexity of gene promoters.

## 18. References

- Adair, G.S. 1925. The hemoglobin system. VI. The oxygen dissociation curve of hemoglobin. *J. Biol. Chem.* 63, 529-545.
- Altaf, M., Auger, A., Covic, M., Côté, J. 2009. Connection between histone H2A variants and chromatin remodeling complexes. *Biochem. Cell. Biol.* 87, 35-50.
- Archer, T.K., Lefebvre, P., Wolford, R.G., Hager, G.L. 1992. Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. *Science* 255, 1573-1576.
- Auble, D.T., Hansen, K.E., Mueller, C.G., Lane, W.S., Thorner, J., Hahn, S. 1994. Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.* 15, 1920-1934.
- Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttill, R.A., Dollé, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., Vijg, J. 2006. Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* 441, 1011-1014.
- Barrera-Saldana, H., Takahashi, K., Vigneron, M., Wildeman, A., Davidson, I., Chambon, P. 1985. All six GC-motifs of the SV40 early upstream element contribute to promoter activity in vivo and in vitro. *EMBO J.* 4, 3839-3849.
- Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L., Weintraub, H. 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* 61, 49-59.
- Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Phillips, R. 2005. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* 15, 116-124.
- Biot, J.B. 1838. Méthodes mathématiques et expérimentales pour discerner les mélanges et les combinaisons chimiques. *Mem. Acad. Sci.* 15, 93-279.
- Bolouri, H., Davidson, E.H. 2002. Modeling transcriptional regulatory networks. *Bioessays* 24, 1118-1129.
- Bosisio, D., Marazzi, I., Agresti, A., Shimizu, N., Bianchi, M.E., Natoli, G. 2006. A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *EMBO J.* 25, 798-810.
- Briggs, G.E., Haldane, J.B. 1925. A Note on the Kinetics of Enzyme Action. *Biochem. J.* 19, 338-339.
- Buchler, N.E., Cross, F.R. 2009. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* 5, 272.
- Carroll, S.B. 2000. Endless forms: The evolution of gene regulation and morphological diversity. *Cell* 101, 577-580.
- Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., Huang, S. 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453, 544-547.
- Chávez, S., Beato, M. 1997. Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus promoter. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2885-2890.
- Chen, D., Hinkley, C.S., Henry, R.W., Huang, S. 2002. TBP dynamics in living human cells: constitutive association of TBP with mitotic chromosomes. *Mol. Biol. Cell.* 13, 276-284.
- Chin, A.C., Fournier, R.E. 1987. A genetic analysis of extinction: trans-regulation of 16 liver-specific genes in hepatoma-fibroblast hybrid cells. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1614-1618.

- Chiu, C.P., Blau, H.M. 1984. Reprogramming cell differentiation in the absence of DNA synthesis. *Cell* 37, 879-887.
- Choi, J.K., Bae, J.B., Lyu, J., Kim, T.Y., Kim, Y.J. 2009. Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biol.* 10, R89.
- Chubb, J.R., Treck, T., Shenoy, S.M., Singer, R.H. 2006. Transcriptional pulsing of a developmental gene. *Curr Biol* 16, 1018-1025.
- Dasgupta, A., Juedes, S.A., Sprouse, R.O., Auble, D.T. 2005. Mot1-mediated control of transcription complex assembly and activity. *EMBO J.* 24, 1717-1729.
- Davis, R.L., Weintraub, H., Lassar, A.B. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987-1000.
- Degenhardt, T., Rybakova, K.N., Tomaszewska, A., Moné, M.J., Westerhoff, H.V., Bruggeman, F.J., Carlberg, C. 2009. Population-level transcription cycles derive from stochastic timing of single-cell transcription. *Cell* 138, 489-501.
- Eisfeld, K., Candau, R., Truss, M., Beato, M. 1997. Binding of NF1 to the MMTV promoter in nucleosomes: influence of rotational phasing, translational positioning and histone H1. *Nucleic Acids Res.* 25, 3733-3742.
- El Kaderi, B., Medler, S., Raghunayakula, S., Ansari, A. 2009. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. *J. Biol. Chem.* 284, 25015-25025.
- Elbi, C., Walker, D.A., Romero, G., Sullivan, W.P., Toft, D.O., Hager, G.L., DeFranco, D.B. 2004. Molecular chaperones function as steroid receptor nuclear mobility factors. *Proc. Natl. Acad. Sci. USA* 101, 2876-2881.
- Freeman, B.C., Yamamoto, K.R. 2001. Continuous recycling: a mechanism for modulatory signal transduction. *Trends Biochem. Sci.* 26, 285-290.
- Gaszner, M., Felsenfeld, G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* 7, 703-713.
- Gilmour, D.S. 2009. Promoter proximal pausing on genes in metazoans. *Chromosoma* 118, 1-10.
- Herbomel, P. 1990. Synergistic activation of eukaryotic transcription: the multiacceptor target hypothesis. *New Biol.* 2, 1063-1070.
- Hiratani, I., Gilbert, D.M. 2009. Replication timing as an epigenetic mark. *Epigenetics* 4, 93-97.
- Hopfield, J.J. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U.S.A.* 71, 4135-4139.
- Huang, S. 2009. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* 31, 546-560.
- Huang, S., Eichler, G., Bar-Yam, Y., Ingber, D.E. 2005. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94, 128701.
- John, S., Workman, J.L. 1998. Bookmarking genes for activation in condensed mitotic chromosomes. *Bioessays* 20, 275-279.
- Kamei, Y., Xu, L., Heinzl, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S.C., Heyman, R.A., Rose, D.W., Glass, C.K., Rosenfeld, M.G. 1996. A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* 85, 403-414.
- Kaneko, K.J., Rein, T., Guo, Z.S., Latham, K., DePamphilis, M.L. 2004. DNA methylation may restrict but does not determine differential gene expression at the Sgy/Tead2 locus during mouse development. *Mol. Cell. Biol.* 24, 1968-1982.

- Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., Reid, G. 2008. Transient cyclical methylation of promoter DNA. *Nature* 452, 112-115.
- Kimura, H., Sugaya, K., Cook, P.R. 2002. The transcription cycle of RNA polymerase II in living cells. *J. Cell. Biol.* 159, 777-782.
- Ko, M.S.H. 1991. A stochastic model for gene induction. *J. Theor. Biol.* 153, 181-194.
- Koutroubas, G., Merika, M., Thanos, D. 2008. Bypassing the requirements for epigenetic modifications in gene transcription by increasing enhancer strength. *Mol. Cell. Biol.* 28, 926-938.
- Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* 128, 693-705.
- Kricker, M.C., Drake, J.W., Radman, M. 1992. Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1075-1079.
- Krishnamurphy, S., Hampsey, M. 2009. Eukaryotic transcription initiation. *Curr. Biol.* 19, R153-R156.
- Lacal, J., Guazzaroni, M.E., Busch, A., Krell, T., J.L., R. 2008. Hierarchical binding of the TodT response regulator to its multiple recognition sites at the *tod* pathway operon promoter. *J. Mol. Biol.* 376, 325-337.
- Lebedeva, L.A., Nabirochkina, E.N., Kurshakova, M.M., Robert, F., Krasnov, A.N., Evgen'ev, M.B., Kadonaga, J.T., Georgieva, S.G., Tora, L. 2005. Occupancy of the *Drosophila hsp70* promoter by a subset of basal transcription factors diminishes upon transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18087-18092.
- Lee, J.H., Gaetz, J., Bugarija, B., Fernandes, C.J., Snyder, G.E., Bush, E.C., Lahn, B.T. 2009. Chromatin analysis of occluded genes. *Hum. Mol. Genet.* 18, 2567-2574.
- Lemaire, V., Lee, C.F., Lei, J., Métivier, R., Glass, L. 2006. Sequential recruitment and combinatorial assembling of multiprotein complexes in transcriptional activation. *Phys. Rev. Lett.* 96, 198102.
- Liu, W.L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., Tjian, R. 2009. Structures of three distinct activator-TFIIID complexes. *Genes Dev.* 23, 1510-1521.
- Lopes, F.J., Vieira, F.M., Holloway, D.M., Bisch, P.M., Spirov, A.V. 2008. Spatial bistability generates hunchback expression sharpness in the *Drosophila* embryo. *PLoS Comput. Biol.* 4, e1000184.
- Mason, P.B.J., Lis, J.T. 1997. Cooperative and competitive protein interactions at the *hsp70* promoter. *J. Biol. Chem.* 272, 33227-33233.
- Métivier, R., Penot, G., Hübner, M.R., Reid, G., Brand, H., Kos, M., Gannon, F. 2003. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* 115, 751-763.
- Mével-Ninio, M., Weiss, M.C. 1981. Immunofluorescence analysis of the time-course of extinction, reexpression, and activation of albumin production in rat hepatoma-mouse fibroblast heterokaryons and hybrids. *J. Cell. Biol.* 90, 339-350.
- Michel, D. 2008. An alternative theoretical formula for hemoglobin oxygenation. *Eur. Biophys. J.* 37, 823-827.
- Michel, D. 2009. Fine tuning gene expression through short DNA-protein interaction cycles. *Biochimie* 91, 933-941.
- Michelotti, E.F., Sanford, S., Levens, D. 1997. Marking of active genes on mitotic chromosomes. *Nature* 388, 895-899.
- Minchin, S.D., Busby, S.J. 2009. Analysis of mechanisms of activation and repression at bacterial promoters. *Methods* 47, 6-12.

- Molinari, E., Gilman, M., Natesan, S. 1999. Proteasome-mediated degradation of transcriptional activators correlates with activation domain potency in vivo. *EMBO J.* 18, 6439-6447.
- Muldrow, T.A., Campbell, A.M., Weil, P.A., Auble, D.T. 1999. MOT1 can activate basal transcription in vitro by regulating the distribution of TATA binding protein between promoter and nonpromoter sites. *Mol. Cell. Biol.* 19, 2835-2845.
- Ni, Z., Saunders, A., Fuda, N.J., Yao, J., Suarez, J.R., Webb, W.W., Lis, J.T. 2008. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Mol. Cell Biol.* 28, 1161-1170.
- Ninio, J. 1986. Kinetic and probabilistic thinking in accuracy. In *Accuracy in Molecular Processes*, Kirkwood. T.B.L., R. Rosenberger, D.J. Galas, eds. (London, Chapman & Hall), pp. 291-328.
- Orphanides, G., Reinberg, D. 2002. A unified theory of gene expression. *Cell* 108, 439-451.
- Peccoud, J., Ycart, B. 1995. Markovian modelling of gene products synthesis. *Theor. Pop. Biol.* 48, 222-234.
- Piña, B., Brüggemeier, U., Beato, M. 1990. Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter. *Cell* 60, 719-731.
- Plet, A., Eick, D., Blanchard, J.M. 1995. Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene* 10, 319-328.
- Polach, K.J., Widom, J. 1996. A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* 258, 800-812.
- Probst, A.V., Dunleavy, E., Almouzni, G. 2009. Epigenetic inheritance during the cell cycle. *Nat Rev Mol. Cell Biol.* 10, 192-206.
- Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., Tyagi, S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 10, e309.
- Raveh-Sadka, T., Levo, M., Segal, E. 2009. Incorporating Nucleosomes into Thermodynamic Models of Transcription Regulation. *Genome Res.* 19, 1480-1496.
- Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., Längst, G. 2007. DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15635-15640.
- Rottach, A., Leonhardt, H., Spada, F. 2009. DNA methylation-mediated epigenetic control. *J. Cell. Biochem.* 108, 43-51.
- Salghetti, S.E., Caudy, A.A., Chenoweth, J.G., Tansey, W.P. 2001. Regulation of transcriptional activation domain function by ubiquitin. *Science* 293, 1651-1653.
- Sandaltzopoulos, R., Becker, P.B. 1998. Heat shock factor increases the reinitiation rate from potentiated chromatin templates. *Mol. Cell. Biol.* 18, 361-367.
- Sauer, F., Hansen, S.K., Tjian, R. 1995. Multiple TAFII directing synergistic activation of transcription. *Science* 270, 1783-1788.
- Schaufele, F., West, B.L., Reudelhuber, T.L. 1990. Overlapping Pit-1 and Sp1 binding sites are both essential to full rat growth hormone gene promoter activity despite mutually exclusive Pit-1 and Sp1 binding. *J. Biol. Chem.* 265, 17189-17196.
- Scheer, U., Rose, K.M. 1984. Localization of RNA polymerase I in interphase cells and mitotic chromosomes by light and electron microscopic immunocytochemistry. *Proc. Natl. Acad. Sci. U.S.A.* 81, 1431-1435.

- Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A.V., Zaret, K.S. 2009. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* 23, 804-809.
- Shen, J., Prywes, R. 2005. ER stress signaling by regulated proteolysis of ATF6. *Methods* 35, 382-389.
- Shi, Y., Mosser, D.D., Morimoto, R.I. 1998. Molecular chaperones as HSF1-specific transcriptional repressors. *Genes Dev.* 12, 654-666.
- Silva, A.J., Ward, K., White, R. 1993. Mosaic methylation in clonal tissue. *Dev. Biol.* 156, 391-398.
- Sprague, B.L., Pego, R.L., Stavreva, D.A., McNally, J.G. 2004. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.* 86, 3473-3495.
- Sprouse, R.O., Wells, M.N., Auble, D.T. 2009. TATA-binding protein variants that bypass the requirement for Mot1 In Vivo. *J. Biol. Chem.* 284, 4525-4535.
- Stavreva, D.A., Wiench, M., John, S., Conway-Campbell, B.L., McKenna, M.A., Pooley, J.R., Johnson, T.A., Voss, T.C., Lightman, S.L., Hager, G.L. 2009. Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nat. Cell. Biol.* 11, 1093-1102.
- Stehle, J.H., Foulkes, N.S., Molina, C.A., Simonneaux, V., Pévet, P., Sassone-Corsi, P. 1993. Adrenergic signals direct rhythmic expression of transcriptional repressor CREM in the pineal gland. *Nature* 365, 314-320.
- Strähle, U., Schmid, W., Schütz, G., 1988. Synergistic action of the glucocorticoid receptor with transcription factors. *EMBO J.* 7, 3389-3395.
- Struhl, G., Struhl, K., Macdonald, P.M. 1989. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* 57, 1259-1273.
- Sutherland, H., Bickmore, W.A. 2009. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* 10, 457-466.
- Szentirmay, M.N., Sawadogo, M. 1993. Synthesis of reinitiated transcripts by mammalian RNA polymerase II is controlled by elongation factor SII. *EMBO J.* 12, 4677-4684.
- Thomas, D., Tyers, M. 2000. Transcriptional regulation: Kamikaze activators. *Curr. Biol.* 10, R341-R343.
- Thuerauf, D.J., Marcinko, M., Belmont, P.J., Glembotski, C.C. 2007. Effects of the isoform-specific characteristics of ATF6 alpha and ATF6 beta on endoplasmic reticulum stress response gene expression and cell viability. *J. Biol. Chem.* 282, 22865-22878.
- Thuerauf, D.J., Morrison, L.E., Hoover, H., Glembotski, C.C. 2002. Coordination of ATF6-mediated transcription and ATF6 degradation by a domain that is shared with the viral transcription factor, VP16. *J. Biol. Chem.* 277, 20734-20739.
- Trojer, P., Reinberg, D. 2006. Histone lysine demethylases and their impact on epigenetics. *Cell* 125, 213-217.
- Vashee, S., Melcher, K., Ding, W.V., Johnston, S.A., Kodadek, T. 1998. Evidence for two modes of cooperative DNA binding in vivo that do not involve direct protein-protein interactions. *Curr. Biol.* 8, 452-458.
- von Hippel, P.H., Berg, O.G. 1986. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 83, 1608-1612.
- Voss, T.C., Schiltz, R.L., Sung, M.H., Johnson, T.A., John, S., Hager, G.L. 2009. Combinatorial probabilistic chromatin interactions produce transcriptional heterogeneity. *J. Cell. Sci.* 122, 345-356.
- Wang, K.Y., James Shen, C.K. 2004. DNA methyltransferase Dnmt1 and mismatch repair. *Oncogene* 23, 7898-7902.

- Woodard, R.L., Lee, K.J., Huang, J., Dynan, W.S. 2001. Distinct roles for Ku protein in transcriptional reinitiation and DNA repair. *J. Biol. Chem.* 276, 15423-15433.
- Wu, Y.F., Myasnikova, E., Reinitz, J. 2007. Master equation simulation analysis of immunostained Bicoid morphogen gradient. *BMC Syst. Biol.* 1, 52.
- Xing, H., Vanderford, N.L., Sarge, K.D. 2008. The TBP-PP2A mitotic complex bookmarks genes by preventing condensin action. *Nat. Cell. Biol.* 10, 1318-1823.
- Xing, H., Wilkerson, D.C., Mayhew, C.N., Lubert, E.J., Skaggs, H.S., Goodson, M.L., Hong, Y., Park-Sarge, O.K., Sarge, K.D. 2005. Mechanism of hsp70i gene bookmarking. *Science* 307, 421-423.
- Yao, J., Munson, K.M., Webb, W.W., Lis, J.T. 2006. Dynamics of heat shock factor association with native gene loci in living cells. *Nature* 442, 1050-1053.
- Yen, P.M., Chin, W.W. 1994. Molecular mechanisms of dominant negative activity by nuclear hormone receptors. *Mol. Endocrinol.* 8, 1450-1454.
- Yudkovsky, N., Ranish, J.A., Hahn, S. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408, 225-229.
- Zawel, L., Kumar, K.P., Reinberg, D. 1995. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* 9, 1479-1490.
- Zhang, F., Pomerantz, J.H., Sen, G., Palermo, A.T., Blau, H.M. 2007. Active tissue-specific DNA demethylation conferred by somatic cell nuclei in stable heterokaryons. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4395-4400.
- Zhang, J., Xu, F., Hashimshony, T., Keshet, I., Cedar, H. 2002. Establishment of transcriptional competence in early and late S phase. *Nature* 420, 198-202.
- Zinzen, R.P., Senger, K., Levine, M., Papatsenko, D. 2006. Computational models for neurogenic gene expression in the *Drosophila* embryo. *Curr. Biol.* 16, 1358-1365.
- Zwicker, J., Müller, R. 1997. Cell-cycle regulation of gene expression by transcriptional repression. *Trends Genet.* 13, 3-6.

## Figures legends

Fig. 1. **(A)** and **(B)**. Spatial variations of the concentrations of BCD and of a BCD-responsive gene product. Plain line: exponential decrease of BCD with the distance ( $L$ ) from the anterior pole. Dotted line: BCD concentration-dependence curve for only one binding site (Eq. 6). Dashed line: cooperative fixation of BCD to two binding sites (the fixation of the first BCD molecule leads to a 20-fold increase of the constant for the second BCD binding step). The slope can be further steepened through additional mechanisms. **(C)** Enlarged scheme of the nuclei at the boundary, showing two possible BCD-responsive patterns (explained in section 15). The bitmap pattern is expected from the binary hypothesis, in which the nuclei located around the mid-saturating concentration of BCD fall into two equal populations (50% negative and 50% positive). In the greyscale pattern, the boundary is progressive and parallels the slope of the curve in A.

Fig. 2. Two possible situations with respect to the role of DNA-bound TFs in the commitment to transcription. The transcription reinitiation complex (TRC) is autonomous for reinitiating transcription as long as present in the promoter, while the transcription commitment complex (TCC) irreversibly triggers one round of transcription initiation once connected. Double and single arrows represent microreversible and micro-irreversible processes respectively. The rate constant  $k_S$  has been attributed to the first micro-irreversible step.

Fig. 3. **(A)** Schematic representation of the different modes of transcription initiation triggered by a single TF and corresponding to the classification of section 9.1, with respect to the micro-reversible or micro-irreversible action of the TADs (first term R or I) and to the conservative vs destructive impact (second term C or D) of transcription start on the stability of TF and/or TRC. In RC2, PC is a basal PIC autonomously built on the proximal promoter (P) while the TF binds to a distal (D) element. **(B)** Average transcription initiation rates and schematic illustrations of the corresponding time distribution. Red vertical bars: transcription starts; black crenels: TF cycles; blue crenels: TRC cycles. The pseudo-first order constants  $k_2[\text{TRC}]$  can be replaced by true first order constants if the TRC is already present in the promoter but non conveniently arranged for initiating transcription, and if the TF-TRC interaction results from conformational fluctuations of the loaded promoter. The causal dependencies between the cycles in the time distribution schemes are bottom-up.

Fig. 4. Fractional occupancy of a promoter containing one **(A)** or two DNA modules **(B)**. The dashed top arrows represent the recruitment activity of DNA-bound activators. In the equations,  $A$  (where the concentration symbol is omitted), is the normalized concentration balanced by the affinity for the binding site: ( $A = K_A[A]$ , where  $K_A$  is the equilibrium binding constants for the first binding step).

When concentrations are used, they are associated to dimerisation constant ( $K$ ) and affinity for DNA ( $K_{\text{DNA}}$ ).  $[A_{\text{F}}]$  is the concentration of free diffusing activator, defined in the text.

Fig. 5. Examples of concentration-dependence curves (top) and corresponding Hill plots (bottom). **(A)** Dimeric TF, yielding a straight Hill plot of slope 2. **(B)** Simultaneous binding of two TFs (both system, section 10.4). **(C)** Hierarchical binding of three identical TFs, similar to (Lacal et al., 2008) (section 10.6.1), with 2 and 3-fold increases of the binding affinities at the second and third binding step respectively. **(D)** Four independent binding sites for the same TF, one of which having a 2000-fold higher affinity (section 11.3). Since linear coordinates are not appropriate for low concentration values and for distinguishing by eye between the different types of sigmoids, the corresponding Hill plots with clearly identifiable shapes, are shown.

Fig. 6. As for Fig. 4,  $A$  and  $R$  are normalized concentrations, and the fractional saturations  $Y$  can be incorporated in Eqs. 15 and 17 to calculate transcription rates.  $K_{\text{DNA}}$  and  $K$  have the same meaning than in Fig. 4.  $[A_{\text{F}}]$  and  $[R_{\text{F}}]$  are real concentrations of free  $A$  and  $R$ , whose values are defined in section 13.1.2.

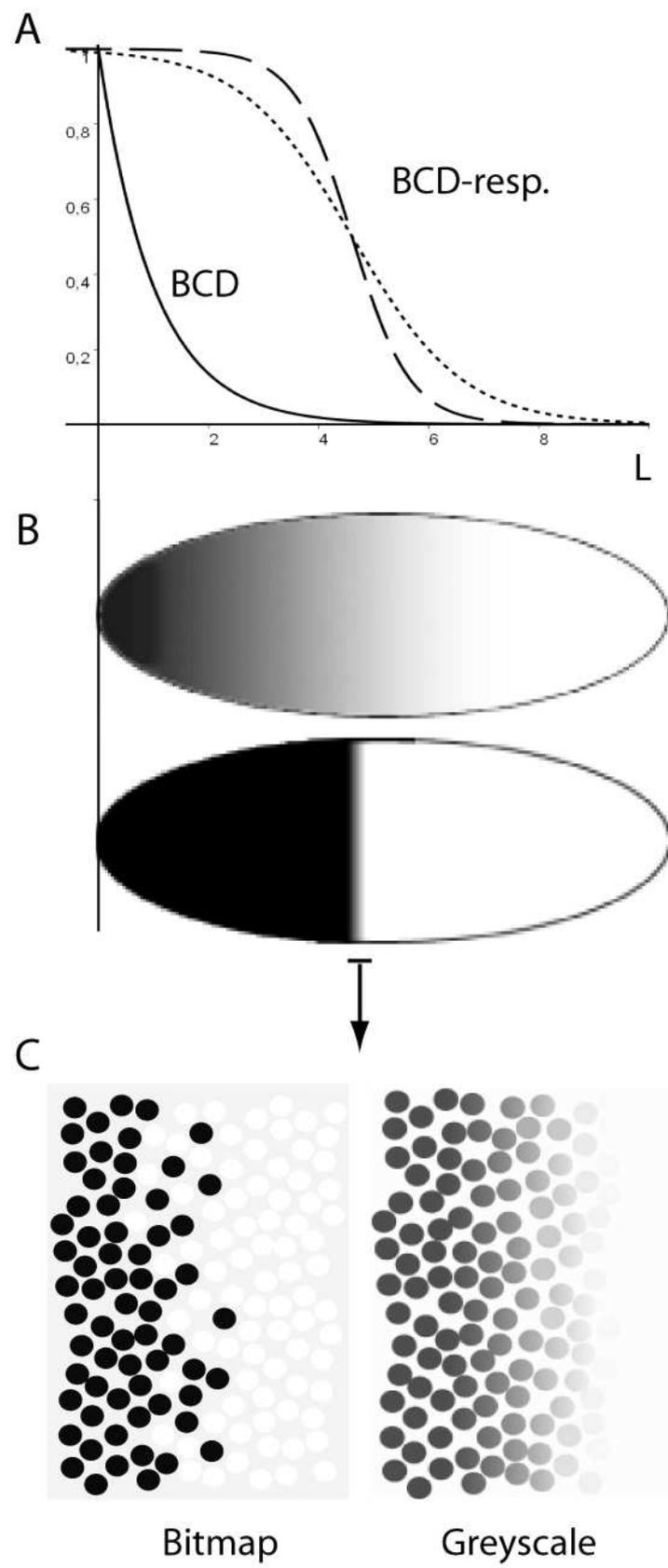


Fig. 1

A



B

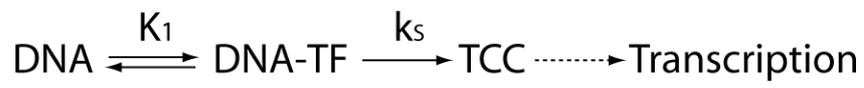


Fig. 2

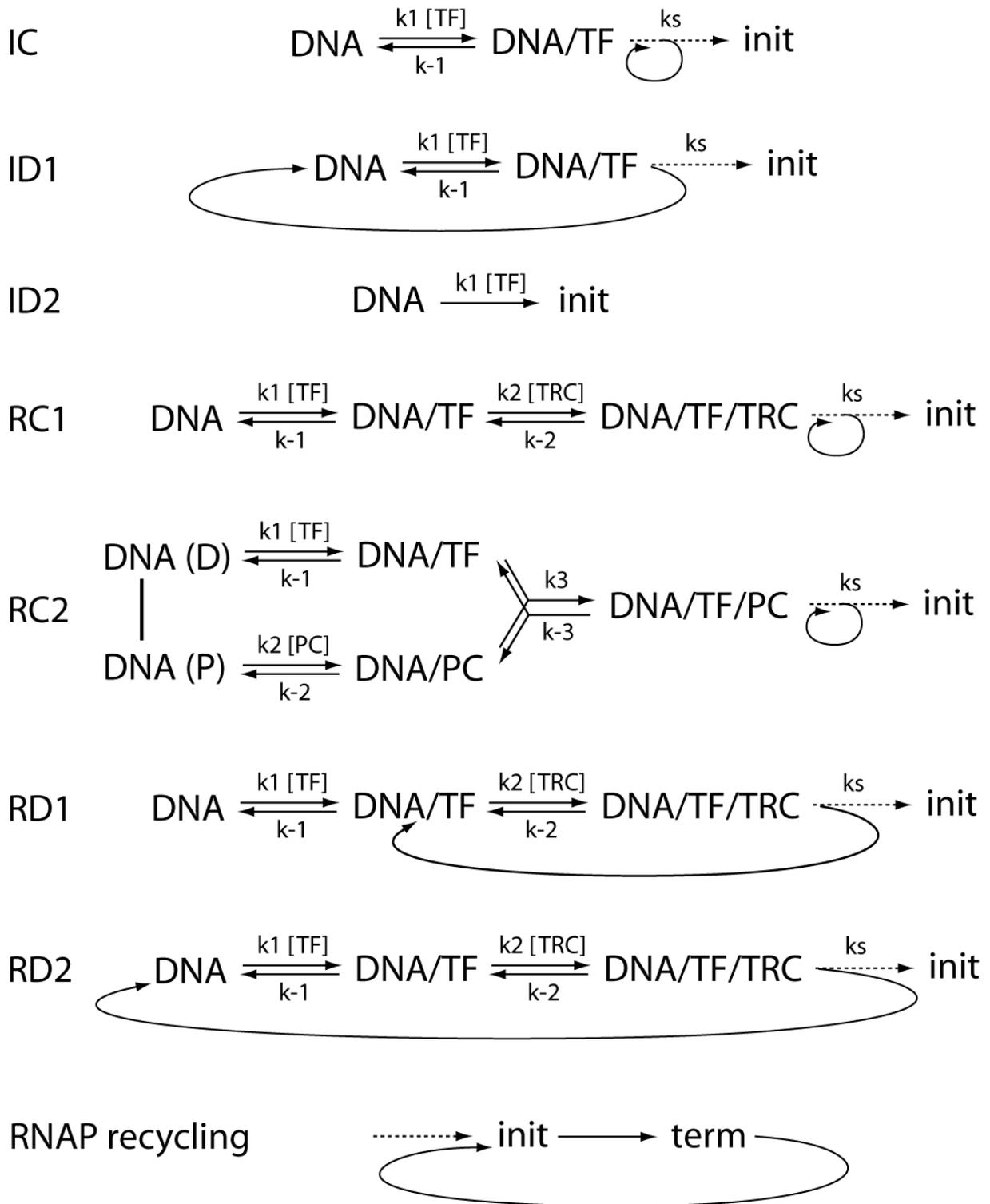


Fig. 3A

Type		Time distribution	Average transcription initiation rate
IC	IC1		$k_{init} = \frac{k_1[TF] k_s}{k_{-1} + k_1[TF]}$
	IC2		
	IC3		
ID	ID1		$k_{init} = \frac{k_1[TF] k_s}{k_{-1} + k_1[TF] + k_s}$
	ID2		$k_{init} = k_1[TF]$
RC	RC1		$k_{init} = \left( \frac{k_1[TF]}{k_{-1} + k_1[TF]} \right) \left( \frac{k_2[TRC]}{k_{-2} + k_2[TRC]} \right) k_s$
	RC2		$k_{init} = \left( \frac{k_1[TF]}{k_{-1} + k_1[TF]} \right) \left( \frac{k_2[PC]}{k_{-2} + k_2[PC]} \right) \frac{k_3}{k_{-3} + k_3} k_s$
RD	RD1		$k_{init} = \left( \frac{k_1[TF]}{k_{-1} + k_1[TF]} \right) \frac{k_2[TRC] k_s}{k_{-2} + k_2[TF] + k_s}$
	RD2		$k_{init} = \frac{k_1[TF] k_2[TRC] k_s}{k_{-1}(k_{-2} + k_s) + k_s k_2[TRC] + (k_s + k_{-2} + k_2[TRC]) k_1[TF]}$

Fig. 3B

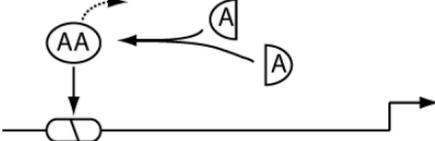
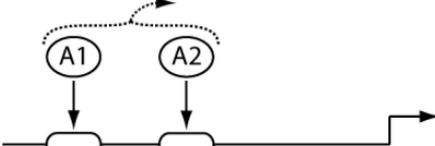
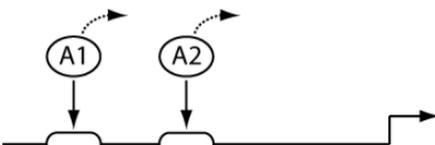
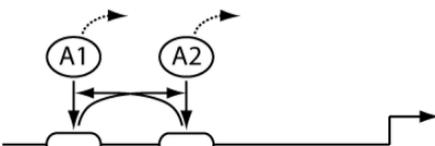
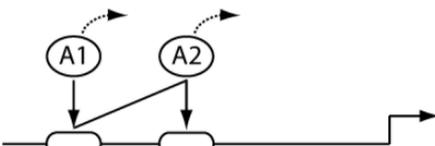
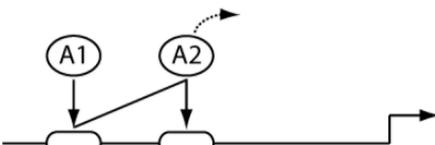
<b>A</b> Single DNA module	Fractional promoter occupancy	Possible sigmoidicity
<b>1. Hyperbolic activation</b> 	$Y = \frac{A1}{1+A1}$	-
<b>2. Competition for the same site</b> 	$Y = \frac{A1+A2}{1+A1+A2}$	-
<b>B</b> Two DNA modules	Fractional promoter occupancy	Possible sigmoidicity
<b>1. Dimeric TF</b> 	$Y = \frac{K_{DNA} \cdot K[A_F]^2}{1 + K_{DNA} \cdot K[A_F]^2}$	+
<b>2. Active only when both are bound</b> 	$Y = \left(\frac{A1}{1+A1}\right)\left(\frac{A2}{1+A2}\right)$	+
<b>3. Independent TFs</b> 	$Y = \frac{1}{2} \left[ \left(\frac{A1}{1+A1}\right) + \left(\frac{A2}{1+A2}\right) \right]$	-
<b>4. Reciprocal cooperativity</b> 	$Y = \frac{A1 + A2 + 2K_c A1A2}{2(1 + A1 + A2 + K_c A1A2)}$	+
<b>5. Hierarchical cooperativity</b> 	$Y = \frac{A1 + 2A1A2}{2(1 + A1 + A1A2)}$	+
<b>6. Promoter presetting</b> 	$Y = \frac{A1A2}{1 + A1 + A1A2}$	+

Fig. 4

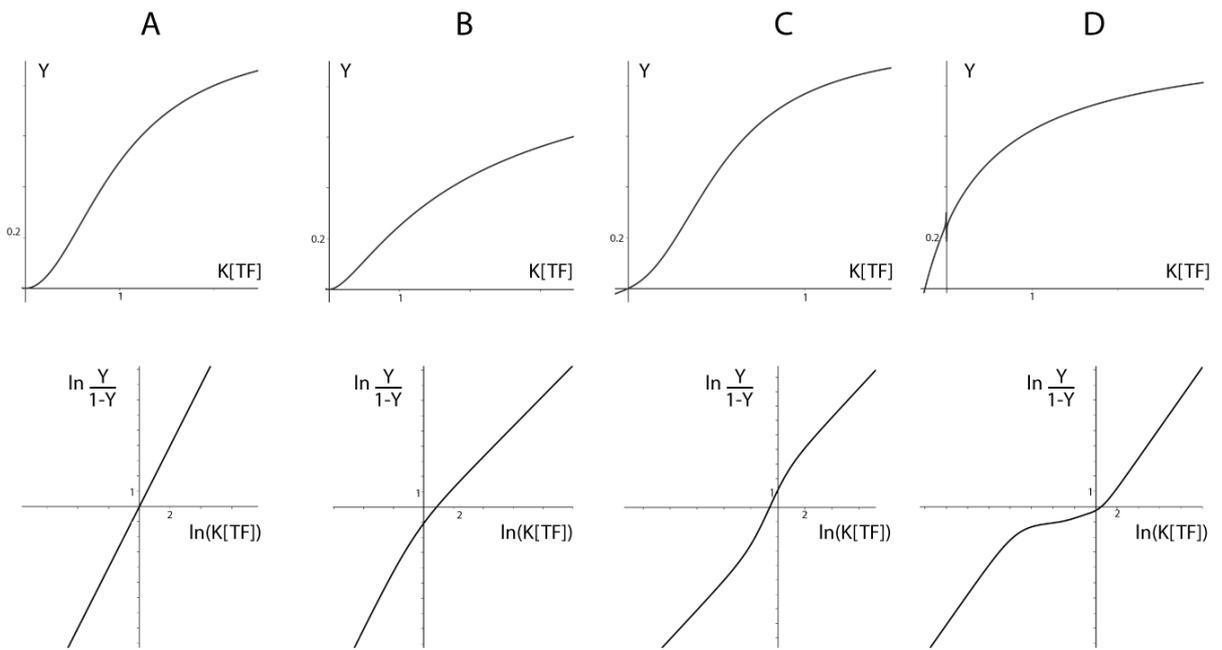


Fig.5

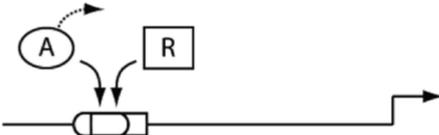
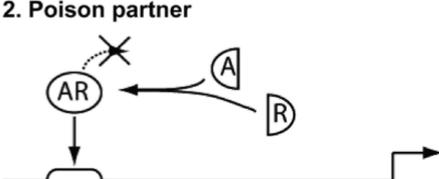
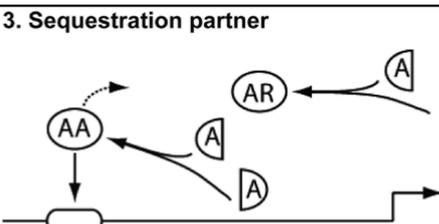
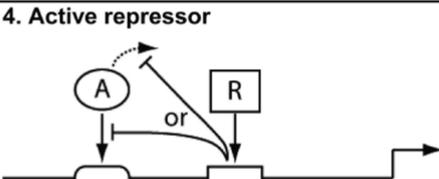
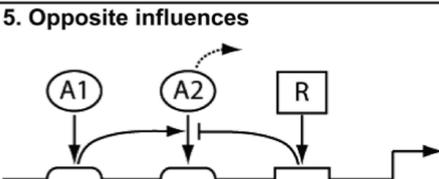
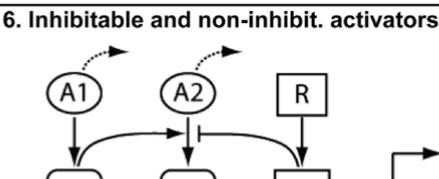
	Fractional promoter occupancy	Sigmoidicity
<b>1. Competition for DNA</b> 	$Y = \frac{A}{1+A+R}$	-
<b>2. Poison partner</b> 	$Y = \frac{K_{DNA}K[A_F]^2}{1+K_{DNA}K[A_F]^2+K_{DNA}K[A_F][R_F]}$	+
<b>3. Sequestration partner</b> 	$Y = \frac{K_{DNA}K[A_F]^2}{1+K_{DNA}K[A_F]^2}$	+
<b>4. Active repressor</b> 	$Y = \left(\frac{1}{1+R}\right)\left(\frac{A}{1+A}\right)$	-
<b>5. Opposite influences</b> 	$Y = \left(\frac{1}{1+R}\right)\left(\frac{A1}{1+A1}\right)\left(\frac{A2}{1+A2}\right)$	+
<b>6. Inhibitible and non-inhibit. activators</b> 	$Y = \frac{1}{2}\left(\frac{A1}{1+A1}\right)\left[1 + \left(\frac{1}{1+R}\right)\left(\frac{A2}{1+A2}\right)\right]$	+

Fig. 6