

Catching transcriptional regulation by thermostatistical modeling

Till D. Frank^{1,2*}, Alex Cheong¹, Mariko Okada-Hatakeyama³ and Boris N. Kholodenko^{1*}

¹Systems Biology Ireland, University College Dublin, Belfield, Dublin 4, Ireland, ²UCD School of Physics, University College Dublin, Belfield, Dublin 4, Ireland, ³Laboratory for Cellular Systems Modeling, RIKEN Research Center for Allergy and Immunology W518, 1-7-22 Tsurumi-ku, Yokohama, 230-0045, Japan.

*Correspondence to: Till.Frank@ucd.ie, Boris.Kholodenko@ucd.ie.

Abstract

Gene expression is frequently regulated by multiple transcription factors (TFs). Thermostatistical methods allow for a quantitative description of interactions between TFs, RNA polymerase, and DNA, and their impact on the transcription rates. We illustrate three different scales of the thermostatistical approach: the microscale of TF molecules, the mesoscale of promoter energy levels, and the macroscale of transcriptionally active and inactive cells in a cell population. We demonstrate versatility of combinatorial transcriptional activation by exemplifying logic functions, such as AND and OR gates. We discuss a metric for cell-to-cell transcriptional activation variability known as Fermi entropy. Suitability of thermostatistical modeling is illustrated by describing the experimental data on transcriptional induction of NF κ B and the c-Fos protein.

1. Introduction

Biology has witnessed a dramatic revolution in both technology and scientific thinking. While a few decades ago, students chose biology because they did not like mathematics, current biology is increasingly becoming a quantitative science. Now we know that cells perceive and process information using networks of signaling proteins and genes. The intricate spatiotemporal dynamics of these networks specify cell-fate decisions, and mathematical modeling has a key role in understanding cellular responses to signals [1]. Previously, signaling pathways were viewed as linear information pipelines that relate extracellular cues to specific genes. For instance, growth factor stimulation pivotal for cell division and survival was thought to activate the linear mitogen-activated protein kinase (MAPK) pathway where the terminal extracellular signal-related kinase (ERK) led to the induction of the immediate early gene *c-Fos*. However, now we know that pathways emanating from multiple cell-surface receptors crosstalk with each other. Consequently, signals propagate via highly interconnected protein networks; each of effector kinases phosphorylates numerous transcription factors (TFs), and multiple factors are integrated on a promoter of each of a multitude of stimulated or repressed genes.

In an organism, cells are exposed to myriad of external cues that activate a variety of cell-surface receptors. How do cells process multiple signals and exhibit signal-dependent phenotypic changes? One possibility is that graded, analogue signalling by receptors and kinases are converted into discrete, digital outputs of target genes such as all-or-none responses that can be described using 1 or 0 values [2-4]. Here we show that under the proper conditions, logic functions, such as OR and AND gates, can arise from combinatorial transcriptional regulation. Thus, the transcriptional machinery can deconvolute complex analog signals by kinases into digital gene responses. To describe simultaneous regulation of gene expression by several TFs we will use a thermostatistical approach [5-16]. We also address other cases of transcriptional responses, such as multiplicative, independent regulation by TFs.

We will discuss a mathematical model for combinatorial transcriptional activation which is based on a fundamental thermostatistical concept: the microstate Boltzmann probability that is defined on the level of TF molecules (Sections 2 and 3). We will show how this microstate description can be coarse-grained to yield a mesoscale perspective for promoter states and a macroscale perspective of active and inactive cells. In Section 4, we will show that a thermostatistical viewpoint allows us to approximate many integrative promoter layouts using elementary logic operations such as AND and

OR. A thermostatics-based metric for cell-to-cell variability will be introduced in Section 5. Subsequently, we will demonstrate how to relate gene expression data obtained from cellular dose response experiments to mathematical, thermostatical models (Section 6).

2. Micro-, meso- and macrolevels of the thermostatical modeling of the combinatorial regulation of transcriptional activity

In this section, we outline the ingredients of a thermostatical model of combinatorial transcriptional activation. Thermostatical models involve three levels: the single cell microscopic level, the single cell mesoscopic level and the macroscopic level on the scale of cell population, see Fig 1.

<Insert Figure 1 about here>

On the single-cell microscopic level, we consider TFs and RNA polymerase II (RNAP) that may or may not be bound at the DNA promoter specific for a particular gene. There is a large number of possible microstate configurations of the promoter. A microstate configuration indicates if a particular set of TF molecules is bound at certain promoter sites and if RNAP is or is not bound at the promoter. For example, let us consider the *c-fos* promoter with binding sites for the TFs pCREB, which is an activated form of CREB, and pELK1, which is an activated form of ELK1 [17,18]. Let us assume that there are 10000 molecules of pCREB, 5000 molecules of pELK1 and 1000 RNAP molecules in the nucleus, which are available for binding at the promoter. Let us label these molecules with numbers 1 to 10000, 1 to 5000, and 1 to 1000, respectively. An example of a microstate configuration can be that the pCREB molecule with label 3400 and the pELK1 molecule with label 200 are bound, while none of the RNAP molecules is bound at the promoter. Another example of a microstate configuration is the pCREB molecule #3333 is bound, the RNAP molecule #400 is bound, and none of the pELK1 molecules is bound. Each microstate configuration has a specific free energy relative to the free energy of the plain promoter, defined as the microstate in which TFs and RNAP are not bound. Since thermostatical approaches are based on free energy changes rather than absolute levels, it is convenient to put the free energy level of the plain promoter equal to zero. The relative free energy of a microstate configuration is then given by the sum of the following terms: (i) the binding energy E_{RNAP} of RNAP if RNAP is bound, (ii) the binding energy E_k

of bound TF k , (iii) the energy shift E_{Rk} of RNAP induced by binding both RNAP and TF k , (iv) the interaction energy E_{km} between two TFs k and m , and (v) the interaction energy E_{Rkm} between RNAP and two TFs k and m bound at the promoter. Mathematically speaking, the microstate configuration energy is given by

$$E(\text{micro}) = E_{RNAP} + \sum E_k + \sum E_{Rk} + \sum E_{km} + \sum E_{Rkm} \quad (1).$$

For example, for the aforementioned configuration with pCREB molecule #3333 bound, RNAP molecule #400 bound, and all pELK1 molecules unbound we have

$$E(\text{pCREB}\#3333, \text{RNAP}\#400) = E_{RNAP} + E_{CREB} + E_{R,CREB} \quad (2).$$

The probability to find in a single cell a particular microstate configuration is determined by the Boltzmann distribution of the free energies of microstates, as follows

$$p(\text{micro}) = \frac{1}{Z_1} \exp\left(-\frac{E(\text{micro})}{RT}\right) \quad (3)$$

where Z_1 is a normalization constant such that the probabilities of all microstate configurations $p(\text{micro})$ add up to unity. R is the Boltzmann gas constant, T is temperature. The exponential term $\exp(-E/(RT))$ occurring in Eq. (3) is called the Boltzmann factor [19-21].

Let us turn next to the mesoscale of thermostatistical modeling of transcriptional activity. Clearly, there are many microstate configurations with exactly the same energy, $E(\text{micro})$. Consequently, each of these configurations has the same probability to occur in a cell. For example, the configuration in which pCREB # 3333 and RNAP # 400 is bound has the same energy and probability as the configuration in which pCREB # 1000 and RNAP # 100 is bound. In general, as far as the biological functioning of a promoter is concerned it is irrelevant which particular TF molecule is bound at the promoter. Rather it is important, whether TFs sites are occupied or not, and whether a RNAP molecule is bound at the promoter or not. In line with this remark, we group all microstates with equal free energy into separate classes. In thermostatics, these classes are referred to as energy levels. Consequently, from a coarse grained point of view, we ask the question what is the probability to find a DNA promoter in a particular energy level that can be realized by a multitude of microstate configurations. Let $g(E)$ denote the number of microstate configurations with the same energy level E [19-21]. The factor g is called the degeneracy factor. For example, let us consider the *c-fos* promoter with TF binding sites for pCREB and pELK1 in the hypothetical case when only 3 pCREB molecules and 2 pELK1 molecules are available in the nucleus for binding at

the promoter. The energy level defined by microstate configurations where pCREB1 and pELK1 molecules are bound but no RNAP is bound is given by $E(\text{pCREB}, \text{pELK1}) = E_{\text{CREB}} + E_{\text{ELK1}}$. There are 6 microstate configurations with this energy level: 11, 21, 31, 21, 22, 32 (where the first number is the label of the pCREB molecule, the second the label of the pELK1 molecule). Using Eq. (3), the probability $p(\text{meso})$ of having the energy level $E(\text{meso})$ on the mesoscale can be calculated as,

$$p(\text{meso}) = \frac{g(E)}{Z_1} \exp\left(-\frac{E(\text{meso})}{RT}\right) \quad (4).$$

The degeneracy factor g acts as weight of the exponential factor [19-21]. The more microstate configurations exist for a given energy level, the higher is the probability that the DNA promoter is at that energy level. Let us make Eq. (4) more explicit. Let m_k denote a binary variable with index k that only assumes the value 0 and 1. If the TF k is bound at a promoter binding site, $m_k=1$, otherwise $m_k=0$. Likewise, we introduce the variable m_R such that $m_R=1$ if RNAP is bound and $m_R=0$ otherwise. The energy level of a promoter can then be described by means of the variables m_R and m_k with $k=1, \dots, M$ (where M is the number of different TFs). Let n_R and n_k denote the numbers of RNAP molecules and TFs of type k in the nucleus available for binding at the promoter. Then, the multiplicity or degeneracy of a promoter energy level is given by

$$g(m_R, m_1, \dots, m_M) = (n_R)^{m_R} \cdot (n_1)^{m_1} \dots (n_M)^{m_M} \quad (5).$$

It is more convenient to use RNAP and TF concentrations, $[\text{RNAP}]$ and $[\text{TF}_k]$, respectively, rather than molecule numbers. Let denote the respective concentrations. Then Eq. (5) becomes

$$g(m_R, m_1, \dots, m_M) \propto [\text{RNAP}]^{m_R} \cdot [\text{TF}_1]^{m_1} \dots [\text{TF}_M]^{m_M} \quad (6).$$

Note that the proportionality factor in Eq. (6) depends on the units in which the concentrations are measured. Interestingly, Eq. (6) also holds in the more general case for promoters that may exhibit more than one binding site for a particular type of a TF. In this case, the parameters m_k may assume integer values larger than 1. In this context, however, Eq. (6) holds only when the molecule numbers n_k are large. If only relatively few RNAP and TF molecules are available for binding at the promoter, the degeneracy must be computed in a slightly different way [5]. Taking Eqs. (4) and (6) together, the probability that a promoter exhibits a particular mesoscale state defined by the parameters m_R, m_1, \dots, m_M is given by

$$p(m_R, m_1, \dots, m_M) = \frac{[\text{RNAP}]^{m_R} \cdot [\text{TF}_1]^{m_1} \dots [\text{TF}_M]^{m_M}}{Z} \exp\left(-\frac{E(m_R, m_1, \dots, m_M)}{RT}\right) \quad (7)$$

where Z is a normalization constant that also accounts for the measurement units in which concentrations are measured.

On the level of a cell population there are two situations of interests. Given a particular concentration pattern of TFs specifying a particular gene, there will be some cells expressing that gene and other cells not expressing that gene. Thus, we distinguish two cell states, active or inactive, and refer to these two states as macrostates. Thermostatistical approaches to determine transcriptional activity are based on the assumption that a cell with a promoter at which RNAP is bound is engaged in transcriptional initiation [5-11]. The higher the probability is that cells have an active promoter (i.e., exhibiting a RNAP molecule bound at the promoter), the higher is the transcriptional activity of the cell population. The probability (P) of cells with active promoters is given by the sum of all probabilities $p(m_R, m_1, \dots, m_M)$ of mesostates in which RNAP is bound at the promoter ($m_R=1$)

$$P = \sum_{m_1, \dots, m_M} p(m_R = 1, m_1, \dots, m_M) \quad (8).$$

The transcriptional activity r is assumed to be proportional to the probability P [5-11]

$$r = \kappa \cdot P \quad (9)$$

Depending on the context in which the thermostatistical model is used, the parameter $\kappa > 0$ refers to the rate with which messenger RNA is produced or the rate of protein production.

3. Transcriptional activity regulated by two activators

We next illustrate the thermostatistical modeling approach for a promoter with two activators. The micro-, meso-, and macrostates of the thermostatistical model for a population of cells whose gene expression levels are regulated by two activators are summarized in Table 1.

<Insert Table 1 about here>

For a system with two TFs, the corresponding energy terms are E_1 , E_2 , E_{RNAP} , E_{R1} , E_{R2} , E_{12} , and E_{R12} (see Section 2). Using these energy terms, the probability P of transcriptional active cells can be computed from Eqs. (1) and (7). In order to obtain a concise description, it is useful to introduce normalized variables and parameters determined by the energy contributions. First, the exponential

terms $\exp(-E_k/(RT))$ for $k=1,2$ always occur together with the concentrations $[TF_k]$, as $[TF_k] \exp(-E_k/(RT))$. Therefore, we introduce the normalized variables,

$$q_k = [TF_k] \exp\left(-\frac{E_k}{RT}\right), k=1,2 \quad (10).$$

Likewise, we define the variable,

$$q_R = [RNAP] \exp\left(-\frac{E_{RNAP}}{RT}\right) \quad (11).$$

The variables q_1 , q_2 , and q_R describe TF and RNAP concentrations, normalized by the corresponding dissociation constants presented in terms of the free energy changes. Second, we introduce the parameters $\Omega_{Rk}=\exp(-E_{Rk}/(RT))$, $\Omega_{12}=\exp(-E_{12}/(RT))$, $\Omega_{R12}=\exp(-E_{R12}/(RT))$. It is convenient to define the probabilities of mesostates, in which RNAP is bound to (P_{on}) or unbound of (P_{off}) the promoter. These probabilities are the sum of mesoscale probabilities given by Eq. (7). By eliminating the unknown normalization constant Z , we obtain an expression for the probability P of having active transcription of the gene of interest in a cell within a population,

$$\begin{aligned} P &= \frac{P_{on}}{P_{off} + P_{on}} \\ P_{on} &= q_R (1 + \Omega_1 \cdot q_1 + \Omega_2 \cdot q_2 + \Omega_1 \Omega_2 \Omega_{12} \Omega_{R12} \cdot q_1 q_2) / Z \\ P_{off} &= (1 + q_1 + q_2 + \Omega_{12} \cdot q_1 q_2) / Z \\ r &= \kappa \cdot P \end{aligned} \quad (12).$$

From Eq. (12) it follows that in the absence of TFs the basal binding probability P_0 of RNAP and the basal transcriptional activity are given by:

$$P_0 = \frac{q_R}{1 + q_R}, \quad r_0 = \kappa \frac{q_R}{1 + q_R} \quad (13).$$

The function binding probability P_0 as a function of RNAP concentration q_R can be determined experimentally and fits well with the theoretical function $P_0(q_R)$ shown above [22]. Having obtained the basal activity r_0 , we defined the fold change A in transcriptional activity by

$$A = \frac{r}{r_0} \quad (14).$$

4. Fundamental special cases of TF regulation and logic gates

Regulation of transcriptional activity by multiple TFs may operate like logic functions [2,23-30]. We demonstrate this for the OR and AND functions and in addition will address another fundamental case of transcriptional regulation: multiplicative, independent regulation.

4.1. The OR Gate

The Boolean table (truth table) of the OR gate is shown below (Table 2).

<Insert Table 2 about here>

The entry “0” means that TF concentrations are negligibly low. An entry “SAT” means that the normalized concentration q_k of TF k is at the saturation level that is achieved when q_k is much larger than unity. The characteristic properties of an OR gate are described in columns 3 and 4. The entry “SAT” refers to the maximal possible transcriptional activity of the promoter under consideration. Note that this saturation value is the same for the three cases (i) $TF_1=0$ and $TF_2=SAT$, (ii) $TF_1=SAT$ and $TF_2=0$, and (iii) $TF_1=SAT$ and $TF_2=SAT$ listed in Table 2. A mechanistic explanation of an OR TF gate is sketched in Table 3. An OR gate implies that the two binding sites for TFs 1 and 2 overlap or are in close proximity such that when one binding site is occupied, the other binding site cannot be occupied [24]. That is, both TFs compete for the promoter

<Insert Table 3 about here>

Using Eq. (12), the probability of cells in the active state can be computed as a function of the activator concentrations and the concentration of RNAP. For an OR gate, we assume that the maximal activation of transcription by TF_1 or TF_2 is similar, which requires the parameters $\Omega_1=\Omega_2$. It is useful to consider the summed concentration of both TFs as a new variable, $S = q_1 + q_2$. The transcriptional activity r is given by,

$$r = \kappa \frac{q_R(1 + \Omega \cdot S)}{1 + S + q_R(1 + \Omega \cdot S)}, \quad \Omega = \Omega_1 = \Omega_2 \quad (15).$$

From Eq. (15) it follows that the saturation value of the transcriptional rate r and the variable A that specifies a fold change are:

$$r_{sat} = \kappa \frac{\Omega \cdot q_R}{1 + \Omega \cdot q_R} \quad A_{sat} = \frac{\Omega \cdot (1 + q_R)}{1 + \Omega \cdot q_R} \quad (16).$$

It is crucial for our understanding of the OR gate that these saturation levels hold for $S \gg 1$ irrespective of how a large value of S is achieved. Consequently, the saturation values (16) hold if (i) $q_2 \gg 1$ and $q_1 = 0$, (ii) $q_1 = 0$, $q_2 \gg 1$, and (iii) $q_1 \gg 1$ and $q_2 \gg 1$, which corresponds to the cases described in the three bottom rows of the Boolean OR table (Table 2). The OR gate exhibits a characteristic regulation function given by

$$A(s) = \frac{1 + A_{sat} \cdot s}{1 + s} \quad (17).$$

where s is a scaled sum of the normalized TF doses.

$$s = \frac{\Omega}{A_{sat}} S = \left(\frac{1 + \Omega \cdot q_R}{1 + q_R} \right) \cdot (q_1 + q_2) \quad (18).$$

The fold change A as a function of s is shown in Figs. 2A, 2B. $A(s)$ is a monotonic function of s and in particular has a sigmoid shape when plotted in a lin-log plot. As such the OR gate is defined on the two-dimensional plane spanned by the normalized doses q_1 and q_2 of the two transcriptional factors. That is, in terms of the independent variables q_1 and q_2 , Eq. (17) reads

$$A(q_1, q_2) = \frac{1 + A_{sat} \cdot \left(\frac{1 + \Omega \cdot q_R}{1 + q_R} \right) \cdot (q_1 + q_2)}{1 + \left(\frac{1 + \Omega \cdot q_R}{1 + q_R} \right) \cdot (q_1 + q_2)} \quad (19).$$

The function is shown in Figs. 2C and 2D. Panel 2D shows this function in a lin-log plot. Consistent with our analytical consideration, we see that the levels of transcriptional activity are the same for (i) $q_1 = 0$ and $q_2 = \text{SAT}$, (ii) $q_1 = \text{SAT}$ and $q_2 = 0$, and (iii) $q_1 = \text{SAT}$ and $q_2 = \text{SAT}$

<Insert Figure 2 about here>

4.2. The AND Gate

The Boolean table for the AND function is shown in Table 4. Only if both TFs are at their saturation levels, then the promoter exhibits maximal transcriptional activity. In contrast, if the concentration of one of the TFs is negligibly low, then the gene expression is only at the basal level.

<Insert Table 4 about here>

The simplest AND gate mechanism occurs if the promoter binds the two TFs only as a pair. For instance, there might be a glue-like (attractive) interaction between two TFs (as described above in terms of the energy shift E_{km}) such that the binding energy for the pair of transcription factors is much lower than the binding energies of the individual molecules [5,23,31]. In addition or independent of the aforementioned TF-TF interactions, in the other AND gate mechanism both TFs may bind individually to the promoter, but a greater facilitation of RNAP binding occurs only when both TFs are bound at the promoter. This might occur due to cooperative TF interactions or synergistic interactions with the DNA [27,29]. These two variations of AND gates have for example been addressed, for example, in the context of the interaction between nuclear factor Y and E2 (transcriptional) factors (see Figure 7 in [32]). In any case, effectively we need to distinguish only between (i) cells with unoccupied TF binding sites and (ii) cells where both TFs are bound to the promoter. These two possibilities can be combined with the two possibilities to have RNAP bound or not bound at the promoter. As a result, we obtain a total of four mesostates with two mesostates describing cells in the active macrostate and two mesostates reflecting cells in the non-active macrostate. The layout of an AND gate is given in Table 5.

<Insert Table 5 about here>

The thermostatistical model can be cast into mathematical expression by exploiting Table 5 and Eq. (12). In order to uncover the underlying nature of the AND gate, it is useful to introduce the auxiliary variable, which is the product of the normalized TF doses.

$$M = q_1 \cdot q_2 \tag{20}.$$

In terms of the product variable M , the transcriptional activity r is given by

$$r = \kappa \frac{q_R(1 + w \cdot M)}{1 + M + q_R(1 + w \cdot M)} \quad w = \Omega_1 \cdot \Omega_2 \cdot \Omega_{12} \cdot \Omega_{R12}. \quad (21).$$

The saturation values of the transcriptional rate r and the variable A that specifies a fold change are given by

$$r_{sat} = \kappa \frac{w \cdot q_R}{1 + w \cdot q_R} \quad A_{sat} = \frac{w \cdot (1 + q_R)}{1 + w \cdot q_R} \quad (22)$$

Clearly, this saturation value can only be achieved when $M \neq 0$. More precisely, when one of the concentrations equals 0, then the product M vanishes and the promoter shows gene expression at a basal level. In contrast, when both activators exhibit their saturation values, then M and consequently, the transcriptional activity (as measured by r and A) is maximal. In short, the conditions of the Boolean AND table (see Table 4) are satisfied. Transcriptional activity regulated by an AND gate exhibits a characteristic regulation function defined by,

$$A(m) = \frac{1 + A_{sat} \cdot m}{1 + m} \quad (23)$$

with the scaled product of TF doses,

$$m = \frac{w}{A_{sat}} M = \left(\frac{1 + w \cdot q_R}{1 + q_R} \right) \cdot (q_1 \cdot q_2) \quad (24)$$

The function $A(m)$ is shown in Figs. 3A and 3B. It is equivalent to the function $A(s)$ for an OR gate Eq. (17) when the scaled sum of the normalized TF doses s is replaced by the scaled product of TF doses m . The AND function in the q_1 - q_2 space follows from Eqs. (23) and (24) and reads

$$A(q_1, q_2) = \frac{1 + A_{sat} \cdot m(q_1 \cdot q_2)}{1 + m(q_1 \cdot q_2)} \quad (25)$$

Fig. 3 illustrates an AND gate. In particular, the log-lin plot reveals that the basal level of activity is observed when at least one of the two TFs is at very low dose.

<Insert Figure 3 about here>

4.3. Multiplicative, independent regulation of transcriptional activity

Regulation of transcriptional activity by more than one activator frequently results in synergy effects (see e.g., [33-35]). Two distinct mechanisms have been proposed [36]. First, multiple activators may induce transcription by means that are impossible for a single activator (e.g., an assembly of activation complex and DNA looping). Second, the sigmoidal dependency of transcriptional activity on TF concentrations (see Figs. 2B-D, 3B-D) may lead to synergy effects even when the activators independently promote RNAP binding to the promoter region. In this context, the model of multiplicative, independent regulation by two activators provides a quantitative approach to study synergy effect without interactions within TFs. The model is described in Table 6. Each TF induces its own saturation level of transcriptional activity (α and β). When both TFs are at saturation levels, transcriptional activity is at a third saturation level γ that does not necessarily correspond to any of the two other saturation levels α and β .

However, at low RNAP concentrations gene expression is low as well. This is due to the fact that according Eq. (12) the transcriptional activity is proportional to the RNAP concentration. For this scenario, the model defined by Eq. (12) under the assumption of independent activation predicts that the transcriptional activity of combined stimulation (when both activators are at saturation) is the product of the activities induced by the individual activations (last column in Table 6).

<Insert Table 6 about here>

The thermostatistical model for independent regulation by two TFs follows immediately from the general two-activator-model defined by Eq. (12) when we assume the interaction energies between TF (E_{12}) and between two TFs and RNAP (E_{R12}) to be zero. This implies that $\Omega_{12}=1$ and $\Omega_{R12}=1$. In this case, Eq. (12) can be equivalently expressed as,

$$\begin{aligned}
 r &= \kappa \frac{q_R \cdot f_1 \cdot f_2}{1 + q_R \cdot f_1 \cdot f_2} \\
 f_1 &= \frac{1 + \Omega_1 \cdot q_1}{1 + q_1} \\
 f_2 &= \frac{1 + \Omega_2 \cdot q_2}{1 + q_2}
 \end{aligned} \tag{26}.$$

From Eq. (26) it follows that the saturation levels listed in Table 6 are given by

$$\begin{aligned}
\alpha &= \kappa \frac{q_R \cdot \Omega_2}{1 + q_R \cdot \Omega_2} \\
\beta &= \kappa \frac{q_R \cdot \Omega_1}{1 + q_R \cdot \Omega_1} \\
\gamma &= \kappa \frac{q_R \cdot \Omega_1 \cdot \Omega_2}{1 + q_R \cdot \Omega_1 \cdot \Omega_2}
\end{aligned}
\tag{27}$$

From these relationships the fold change saturation levels can be computed. In particular, for low RNAP concentrations we obtain approximately

$$\begin{aligned}
A_{sat}(TF_1 = 0, TF_2 = SAT) &= \lambda = \Omega_2 \\
A_{sat}(TF_1 = SAT, TF_2 = 0) &= \sigma = \Omega_1 \\
A_{sat}(TF_1 = SAT, TF_2 = SAT) &= \Omega_1 \cdot \Omega_2
\end{aligned}
\tag{28}$$

That is, the fold change induced by a combined stimulation with both TFs is the product of the fold changes induced by the corresponding individual stimulations [5,6].

5. Thermostatistics-based metric of cell-to-cell variability

While the primary objective of thermostatistical modeling is to provide insights into the functional relationships between the impacts of various TFs, the approach allows us to address a variety of related regulatory properties. First of all, since the model is nonlinear, sensitivity analysis can provide information about which of the model parameters are of particular importance for a given promoter [12,13]. Second, since the modeling approach links systems biology with statistical mechanics it invites us to take advantage of the statistical mechanics approach in a broader sense. For example, cell-to-cell variability can be addressed by taking a quantitative, model-based perspective. According to Eq. (12), the function P describes the probability of cells in the cell populations to have an active transcription at a given promoter. Consequently, the probability of having inactive cells is $1-P$. In general, there will be both active and inactive cells in a population. When studying the expression of a certain protein, the expression level of the protein will be relatively low in the inactive cells and relatively high in the active cells. The population will exhibit a large degree of heterogeneity or cell-to-cell variability. A crude measure for cell-to-cell variability is the entropy S defined on the macrostate probabilities P and $1-P$ of the active and inactive states, respectively. We have

$$S = -P \ln P - (1 - P) \ln(1 - P)
\tag{29}$$

which is also known in physics as Fermi entropy [37,38]. The cell-to-cell variability as measured by S exhibits a maximum at $P=1/2$ and is minimal at $P=0$ and $P=1$, see Fig. 4A. For the logic functions OR and AND, this implies that if the ‘OFF’ and ‘ON’ states (i.e., the four corners in the two-dimensional spaces spanned by the normalized concentrations q_1 and q_2) correspond to the RNAP binding probabilities that are relatively low and high, respectively, then the variability will be lowest in these states. This scenario is illustrated for the OR gate in Fig. 4B-C. We conclude that a thermodynamic metric (macrostate entropy S) can be linked to a biologically relevant quantity (transcriptional activity). Other thermodynamics metrics defined on other scales (not necessarily the macroscale) may provide similar useful information.

<Insert Figure 4 about here>

6. Modeling the dose responses.

NFκB induction. In this section we illustrate the suitability and accuracy of the thermostatistical model to describe experimental dose response data on NFκB induction. For the purpose of clarity, we focus on the impact of a single TF only. From Eq. (12) it follows that the transcriptional activity r of a TF at a normalized dose q_T is given by

$$r = \kappa \cdot \frac{q_R(1 + \Omega \cdot q_T)}{1 + q_T + q_R(1 + \Omega \cdot q_T)} = \kappa \cdot \frac{q_R(1 + \Omega \cdot q_T)}{1 + q_R + (1 + q_R \cdot \Omega)q_T} \quad (30).$$

Frequently, the TF dose is measured in relative units such that we put $q_T = \eta_1 \cdot x$, where x is the dose and $\eta_1 > 0$ a conversion factor. In addition, the TF dose x might be the dependent variable in the experiment design, while the independent variable corresponds to a stimulus Y with dose y . Assuming that the signal transduction network can be linearized at zero doses [39,40], we obtain a linear steady-state relationship between x and y like $x = \eta_2 \cdot y$ with $\eta_2 > 0$ such that Eq. (30) becomes

$$r = \kappa \cdot \frac{q_R(1 + \Omega \cdot \eta \cdot y)}{1 + q_R + (1 + q_R \cdot \Omega) \cdot \eta \cdot y} \quad (31)$$

with $\eta = \eta_1 \cdot \eta_2$. Finally, transcriptional activity may be measured in relative units (e.g., see Eq. 14). Below we will use the data on the transcriptional activity r normalized with respect to the activity measured at a reference dose $y = y(\text{ref})$, as follows

$$a(y) = \frac{r(y)}{r(y(\text{ref}))} \quad (32)$$

Note that the function $a(y)$ reflecting the relative transcriptional activity is independent of the model parameter κ and depends only on the three parameters q_R , Ω , and η .

We fitted dose response data from a recent study [35]. HEK293 cells were transfected with a vector encoding the Gaussia luciferase under the control of NF κ B (pGluc-NRE) and stimulated with NF κ B-inducing cytokines TNF α or IL-1 β (0.1 to 5 ng/ml). Gaussia luciferase differs from firefly or renilla luciferases in that it is secreted into the culture media, thus allowing temporal measurement of transcriptional activity from the same population of cells. The experimental data (sample means and standard deviations) are shown in Fig. 5 (dots and error bars). Using nonlinear regression analysis, the data were fitted to Eqs. (31) and (32). The dependences $a(y)$ of transcriptional activities on the stimuli (y) for the best-fit parameters are depicted in Fig. 5, as lin-lin and semi-logarithmic plots for TNF α (A, B) and IL-1 β (D,E) stimuli. We conclude that a thermostatistical model can describe both data sets reasonably well with somewhat better fidelity for IL-1 β than for TNF α .

<Insert Figure 5 about here>

An AND gate description of c-Fos protein expression. Here we illustrate the thermostatistical modeling of an AND gate for the *c-fos* promoter and its two activator TFs, pCREB and pELK1. We use the time course data for pCREB, pELK1 and the c-Fos protein obtained for MCF-7 cells stimulated with 10 nM heregulin (HRG) for 2 hours [2]. We normalized the product of the concentrations pCREB and pELK1 by its maximum value ($z = m/m(\max)$, see Eqs. 23 and 24), as shown in Fig. 6A. Then Eq. (23) becomes,

$$A(z) = \frac{1 + A_{sat} \cdot m(\max) \cdot z}{1 + m(\max) \cdot z} \quad (33).$$

c-FOS protein expression was modeled as a two-step process involving the activity of messenger RNA. The two-step model reads,



$$\begin{aligned}
\frac{d [mRNA]}{dt} &= A(z) - \frac{k_1 \cdot [mRNA]}{r_0} \\
\frac{d [c-Fos]}{dt} &= \frac{\gamma \cdot [mRNA]}{[c-Fos]_{\max}} - \frac{k_2 \cdot [c-Fos]}{[c-Fos]_{\max}}
\end{aligned}$$

For the sake of simplicity, mRNA concentrations are measured in units of the basal promoter activity (r_0), implying that the transcriptional activity A can be used as the production rate of the first reaction step, Eq. (34). The parameters k_1 and k_2 are taken from the literature. Half-lives for *c-fos* mRNA and the c-Fos protein were reported to be in the order of a few minutes [41,42] and in the range of 30 to 45 min, respectively [3]. We used $k_1=0.12/\text{min}$ (half-life of 6min) and $k_2=0.015/\text{min}$ (half-life of 46min). There are three unknown parameters in Eqs. (32) and (33), A_{sat} , the rate constant γ and the normalization parameter $m(\text{max})$. The characteristic promoter function A that is calculated for the best-fit parameters is shown in Fig. 6B. The dots indicate the position of the experimental data points. The solid line in panel 6C shows the normalized c-FOS expression time course predicted by the model, using the experimental data shown in 6A as input.

<Insert Figure 6 about here>

7. Discussion

Gene expression in general depends on a variety of extracellular signals. These signals are transmitted by means of intracellular signaling molecules to targets, such as TFs. Intracellular signaling operates via multiple pathways. In order to respond appropriately to several extracellular cues, the cell must integrate the extracellular information. In addition to the integration on the level of the intracellular signaling molecules (interaction between pathways), the information can be integrated on the promoter level. The thermostatistical model for combinatorial transcriptional

activation described in the previous sections is a useful tool to study how the cell can integrate information from various pathways and extracellular sources on the level of the transcriptional machinery. In particular, the thermostatistical approach provides a framework to address quantitatively information integration on the transcriptional level in form of logic functions such as AND and OR functions.

While reviewing the thermostatistical approach to transcriptional regulation in Section 2, we distinguished between three different scales: the micro-, meso-, and macro-scales. The promoter states on the microscale reflect physical states. In contrast, on the meso- and macro-scales we are dealing with mathematical representations of certain sets of microscale states. Consequently, microscale states may be regarded as the physical ‘backbone’ of the thermostatistical approach. The mesoscale is of particular interest, since the regulation of transcriptional activity by the transcription factor abundance happens on this scale. More precisely, while the probability of microscale states is *independent* of transcription factor concentrations (see Eq. 3), the probability of mesoscale states in fact *depends* on transcription factor concentrations (see Eqs. 4-6). Finally, on the level of cell populations there are two macroscale probabilities: the probabilities of observing transcriptionally active or inactive cells in a population. These macroscale probabilities connect theory with experiment because experimental results in cell biology are frequently obtained by studying cell population rather than single cells.

In Section 5 we introduced the entropy measures $S(P)$ to quantify cell-to-cell variability for the cell populations. In order to derive $S(P)$ from the transcriptional activity r measured at different time points or concentrations, we need a fit to the predicted graph r that provides us with values of the basal activity r_0 and the saturation activity $r(\text{sat})$ and we need to know the proportionality factor κ occurring in Eq. (9). Then, we will be able to identify the relationship between $S(P)/S_{\text{max}}$ and $r/r(\text{sat})$ because P can be calculated from r like

$$P = r \cdot \frac{P(\text{max})}{r(\text{sat})} \quad (35)$$

From Figure 4A it follows that there are three possible cases. If $P_0 = \kappa \cdot r_0$, $P(\text{max}) = \kappa \cdot r(\text{sat}) < 1/2$ holds, then the thermostatistical approach predicts that the transcriptional machinery exhibits a monotonically increasing relationship between stochasticity and determinism; if $P_0 < 1/2 < P(\text{max})$ holds, then a non-monotonic relationship is predicted; finally for $P_0, P(\text{max}) > 1/2$ a monotonically decreasing relationship is predicted. While an in-depth discussion of the implementation of this

theoretical framework is beyond the scope of the present manuscript, future work may be dedicated to demonstrate in more detail the application of Eqs. (29) and (35) to experimental data.

Acknowledgements

This work was supported by Science Foundation Ireland under Grant No. 06/CE/B1129. The study was in part supported by Cell Innovation Program of MEXT, Japan, to MO.

References

- [1] Kholodenko, B. N. (2006). Cell-signalling dynamics in time and space, *Nat Rev Mol Cell Biol.* 7, 165-176.
- [2] Nakakuki, T., Birtwistle, M. R., Saeki, Y., Yumoto, N., Ide, K., Nagashima, T., Brusch, L., Ogunnaike, B. A., Okada-Hatakeyama, M., Kholodenko, B. N. (2010). Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics, *Cell.* 141, 884-96.
- [3] Murphy, L. O., Smith, S., Chen, R. H., Fingar, D. C. & Blenis, J. (2002). Molecular interpretation of ERK signal duration by immediate early gene products, *Nat Cell Biol.* 4, 556-64.
- [4] Kholodenko, B. N., Hancock, J. F. & Kolch, W. (2010). Signalling ballet in space and time, *Nat Rev Mol Cell Biol.* 11, 414-26.
- [5] Bintu, L. Buchler, N.E., Garcia, H. G., Gerland, U., Hwa, T. et al. (2005). Transcriptional regulation by the numbers: models. *Current opinion in Gen & Dev* 15: 116-124
- [6] Bintu, L. Buchler, N.E., Garcia, H. G., Gerland, U., Hwa, T. et al. (2005). Transcriptional regulation by the numbers: applications. *Current opinion in Gen & Dev* 15: 125-135
- [7] Ackers, G.K., Johnson, A.D., Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci USA* 79: 1129-1133
- [8] Shea, M.A., Ackers, G.A. (1985). The sigma-r control system of bacteriophage lambda: a physical-chemical model for gene-regulation. *J. Mol. Biol.* 181: 211-230
- [9] Wang, J., Ellwood, K. Lehman, A., Carey, M.F. (1999). A mathematical model for synergistic Eukaryotic gene activation. *J. Mol. Biol.* 286: 315-325
- [10] Segal, E., Raveh-Sadka, Schroeder M., Unnerstall U., Gaul, U. (2008). Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451: 535-541
- [11] Veitia, R.A. (2003). A sigmoidal transcriptional response: cooperativity, synergy, and dose effects. *Biol. Rev.* 78: 149-170
- [12] Ahment A, Arnost, D.N. (2011). Mathematical modeling of gene expression: a guide for the perplexed biologist, *Critical Reviews Biochem. Mol. Biology*, 46: 137-151
- [13] Dresch, J.M., Liu X, Arnosti, D.N., Ay, A. (2010). Thermodynamic modeling of transcription: sensitivity analysis differentiates biological mechanisms from mathematical model-induced effects, *BMC Systems Biology*, 4: 142 (11 pp)
- [14] Zinsen, R.P., Senger, K., Levine, M., Papatsenko, D. (2006). Computational models for neurogenic gene expression in the *Drosophila* embryo. *Current Biology*, 16: 1358-1365

- [15] Gertz, J., Siggia, E.D., Cohen, B.A. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, 457: 215-219.
- [16] He, X., Samee, A.H., Blatti, C., Sinha, S. (2012). Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Computational Biology*, 6: e1000935 (15 pp)
- [17] Buchwalter, G., Gross, C. & Wasylyk, B. (2004). Ets ternary complex transcription factors, *Gene*. 324, 1-14.
- [18] Wang, Y. & Prywes, R. (2000). Activation of the c-fos enhancer by the erk MAP kinase pathway through two sequence elements: the c-fos AP-1 and p62TCF sites, *Oncogene*. 19, 1379-85.
- [19] Chandler, D. (1987). *Introduction to modern statistical mechanics*. Oxford University Press, New York.
- [20] Huang, K. (1990). *Statistical mechanics*. John Wiley and Sons, New York.
- [21] Reif, F. (1965). *Fundamentals of statistical and thermal physics*. McGraw-Hill Book Company, New York.
- [22] Vogel, S.K., Schulz, A., Rippe, K. (2002). Binding affinity of Escherichia coli RNA polymerase sigma54 holoenzyme for the glnAp2, nifH, and nifL promoters. *Nucleic Acids Research*, 30: 4094-4101
- [23] Buchler, N.E., Gerland, U., Hwa, T. (2003). On schemes of combinatorial transcriptional logic. *Proc Natl. Acad Sci USA* 100: 5136-5141
- [24] Hermsen, R., Tans, S., ten Wolde, P. R. (2006). Transcriptional regulation by competing transcriptional factor modules. *PLoS Computational Biology* 2: e164 (9pp)
- [25] Setty, Y., Mayo, A.E., Surette, M.G., Alon, U. (2003). Detailed map of the cis-regulatory input function. *Proc Natl. Acad Sci USA* 100: 7702-7707
- [26] Mayo, A.E., Setty, Y., Shavit, S., Zaslaver, A., Alon, U. (2006). Plasticity of the cis-regulatory input function of a gene. *PLoS Biol.* 4: e45 (7 pp)
- [27] Istrail, S., Davidson, E.H. (2005). Logic functions of the genomic cis-regulatory code. *Proc Natl. Acad Sci USA* 102: 4954-4959
- [28] Istrail, S., De-Leon, S.B., Davidson, E.H. (2007). The regulatory genome and the computer. *Developmental Biology*, 310: 187-195.

- [29] Yu, C.H., Dorman, E.R., Howard, M.L., Davidson, E.H. (2004). An *otx* cis-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. *Genomes & developmental control* 269: 536-551.
- [30] Mangan, S., Alon U. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl. Acad Sci USA* 100: 11980-11985
- [31] Hermesen, R., Tans, S., ten Wolde, P. R. (2010). Combinatorial gene regulation using auto-regulation. *PLoS Computational Biology* 6: e1000813 (13pp)
- [32] van Ginkel, P.R., Hsiao, K.M., Schjerven, H., Farnham, P.J. (1997). E2F-mediated growth regulation requires transcription factor cooperation. *Journal of Biological Chemistry* 272: 18367-18374
- [33] Joung, J.K., Koepp, D.M., Hochschild, A. (1994). Synergistic activation of transcription in *Escherichia coli* cAMP receptor protein. *Proc Natl Acad Sci USA* 90: 3083-3087
- [34] Lee, H.S., Bae, E.J., Yi, S.H., Shim, J.W., Jo A.Y. et al. (2010). *Foxa2* and *nurr1* synergistically yield A9 nigral dopamine neurons exhibiting improved differentiation, function, and cell survival. *Stem Cells* 28: 501-512
- [35] Bruning, U., Fitzpatrick, S.F., Birtwistle, M., Frank, T., Taylor, C., Cheong, A. (2012) *NFkappaB* and *HIF* display synergistic behavior during hypoxic stimulation. *Cellular and Molecular Life Sciences* 69: 1319-1329.
- [36] Carey, M. (1998). The enhanceosome and transcriptional synergy. *Cell* 92: 5-8
- [37] Frank, T.D. (2005). *Nonlinear Fokker-Planck equations*, Springer, Berlin.
- [38] Landsberg, P.T. (1961). *Thermodynamics*, Interscience Publishers, London
- [39] Lorenz, D.R., Cantor, C. R., Collins, J.J. (2009). A network biology approach to aging in yeast. *Proc Natl Acad Sci USA* 106: 1145-1150
- [40] Kholodenko B.N, Kiyatkin A, Bruggeman, F., Sontag, E, Westerhoff, H.V. & Hoek, JB. (2002). Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. USA* 20: 12841-12846.
- [41] Rahmsdorf, H.J., Schonthal, A., Angel, P. et al. (1987). Posttranscriptional regulation of *c-fos* mRNA expression. *Nucleic Acids Research*, 15: 1643-1659
- [42] Sariban, E. Luebbers, R. Kufe D. (1988). Transcriptional and posttranscriptional control of *c-fos* gene expression in human monocytes. *Mol. Cell. Biol.* 8: 340-346

Table 1. Schematic description of the thermostatistical model for transcriptional activity induced by two activators

RNAP m_R	TF1 m_1	TF2 m_2	Number of microstates	Mesostate label	Macrostates
0	0	0	1	1	Inactive
0	0	1	$\sim [TF_2]$	2	Inactive
0	1	0	$\sim [TF_1]$	3	Inactive
0	1	1	$\sim [TF_1] \cdot [TF_2]$	4	Inactive
1	0	0	$\sim [RNAP]$	5	Active
1	0	1	$\sim [RNAP] \cdot [TF_2]$	6	Active
1	1	0	$\sim [RNAP] \cdot [TF_1]$	7	Active
1	1	1	$\sim [RNAP] \cdot [TF_1] \cdot [TF_2]$	8	Active

Table 2. Truth table of the OR gate realized by a promoter with two activators TF_1 and TF_2

TF_1	TF_2	Transcriptional activity r	Fold change A
0	0	Basal level	1
0	SAT	SAT	A_{SAT}
SAT	0	SAT	A_{SAT}
SAT	SAT	SAT	A_{SAT}

Table 3. Schematic description of the thermostistical model of the OR gate

RNAP	TF ₁	TF ₂	Number of microstates	Mesostate label	Macrostates
0	0	0	1	1	Inactive
0	0	1	$\sim [TF_2]$	2	Inactive
0	1	0	$\sim [TF_1]$	3	Inactive
1	0	0	$\sim [RNAP]$	4	Active
1	0	1	$\sim [RNAP] \cdot [TF_2]$	5	Active
1	1	0	$\sim [RNAP] \cdot [TF_1]$	6	Active

Table 4. Truth table of the AND gate of a promoter with two activators TF₁ and TF₂

TF ₁	TF ₂	Transcriptional activity r	Fold change A
0	0	Basal level	1
0	SAT	Basal level	1
SAT	0	Basal level	1
SAT	SAT	SAT	A _{SAT}

Table 5. Schematic description of the thermostistical model of the AND gate

RNAP	TF ₁	TF ₂	Number of microstates	Mesostate label	Macrostates
0	0	0	1	1	Inactive
0	1	1	$\sim [TF_1] \cdot [TF_2]$	2	Inactive
1	0	0	$\sim [RNAP]$	3	Active
1	1	1	$\sim [RNAP] \cdot [TF_1] \cdot [TF_2]$	4	Active

Table 6. Input-output relationship of a promoter exhibiting multiplicative, independent regulation by means of two activators TF_1 and TF_2

TF_1	TF_2	Transcriptional activity r	A	A for low RNAP concentrations
0	0	Basal level	1	1
0	SAT	SAT level α	A_{SAT} level λ	A_{SAT} level λ
SAT	0	SAT level β	A_{SAT} level σ	A_{SAT} level σ
SAT	SAT	SAT level γ	A_{SAT} level ε	A_{SAT} level $\sigma \cdot \lambda$

Figure captions

Fig. 1: Different scales involved in the themostatistical modeling of transcriptional activity

Fig. 2. The regulation of transcriptional activity for an OR gate. A,B: The transcriptional activity (fold changes) as function of the scaled sum of the normalized TF doses. The function $A(s)$ was computed from Eq. (17) for $A_{\text{sat}}=10$. A: lin-lin plot. B: semi-logarithmic plot. C,D: Transcriptional activity A as a function of the normalized activator doses q_1 and q_2 . The function $A(q_1, q_2)$ was computed from Eqs. (18) and (19) for $A_{\text{sat}}=10$ and $\Omega=1.5$. C: lin-lin plot. D semi-logarithmic plot.

Fig. 3. The cooperative regulation of transcriptional activity for an AND gate. A,B: The function $A(m)$ is shown for $A_{\text{sat}}=10$ as defined by Eq. (23). A: lin-lin plot. B: semi-logarithmic plot. C,D: The function $A(q_1, q_2)$ is shown for $A_{\text{sat}}=10$ and $w=4$ as was defined by Eq. (24) and (25). C: lin-lin plot. D: semi-logarithmic plot.

Fig. 4. Cell-to-cell variability as quantified by the entropy S defined on the macrostate probabilities of active and inactive cells. A: functional relationship between S and P as computed from Eq. (34). B: the OR gate binding probability P computed from Eqs. (9) and (15). C: Cell-to-cell variability S for OR gate shown in panel B computed by substituting the function values P of panel B into the function S shown in panel A. B-C: Parameters: $q_R=0.01$, $\Omega=2000$.

Fig. 5: Transcriptional dose responses of HEK293 cells to stimulation with $\text{TNF}\alpha$ (panels A&B) and $\text{IL-1}\beta$ (panels C&D) [35]. The normalized transcription activity $a(y)$ calculated by the model defined by Eqs. (31) and (32) is shown for the best-fit estimates: $q_R \approx 0$, $q_R \cdot \Omega = 1.1$, and $\eta = 1.7$ ($\text{TNF}\alpha$) and $q_R = 0.0072$, $q_R \cdot \Omega = 1.7$, and $\eta = 0.63$. A&C: lin-lin plots. B&D: semi-logarithmic plots.

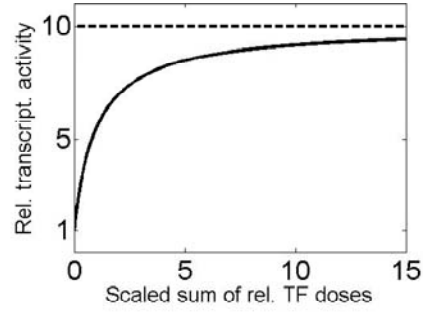
Fig. 6. An AND gate description of c-Fos protein expression induced by two activator TFs, pCREB and pELK-1. A: The model input: the mathematical product of the concentrations pCREB and pELK1 normalized by its maximal value. B: The themostatistical characteristic function A calculated for the best-fit estimates, $A_{\text{sat}}=25$, $m(\text{max})=2.0$ and $\gamma=0.08/\text{min}$. C: The time course of c-FOS expression predicted by the model.

Figure 1

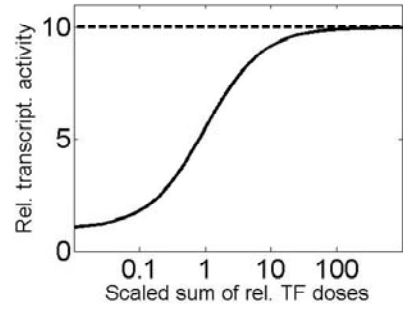
Single cell / Microscale		
Probability of a promoter microstate	~	Boltzmann factor of promoter energy
Single cell / Mesoscale		
Probability of a promoter mesostate	~	microstate degeneracy × Boltzmann factor
Cell population / Macroscale		
Transcriptional activity	~	sum of mesostate probability of active promoters

Figure 2

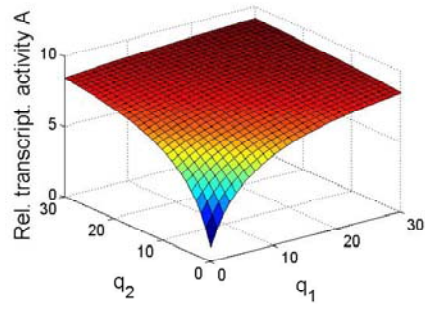
A



B



C



D

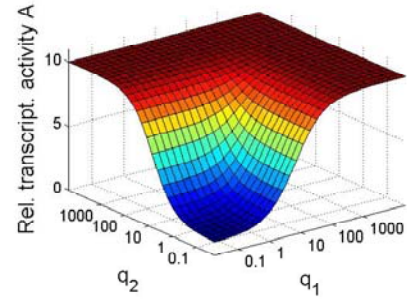
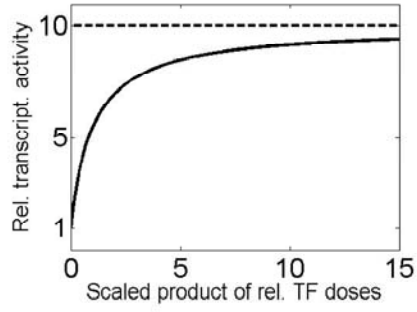
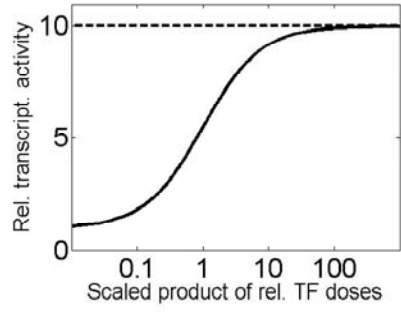


Figure 3

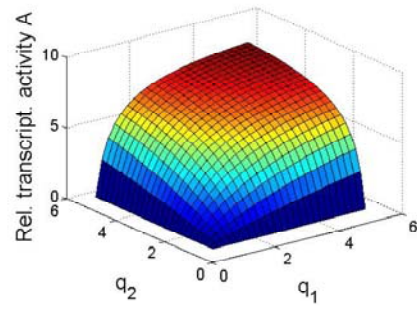
A



B



C



D

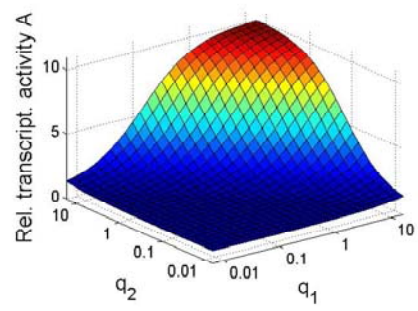


Figure 4

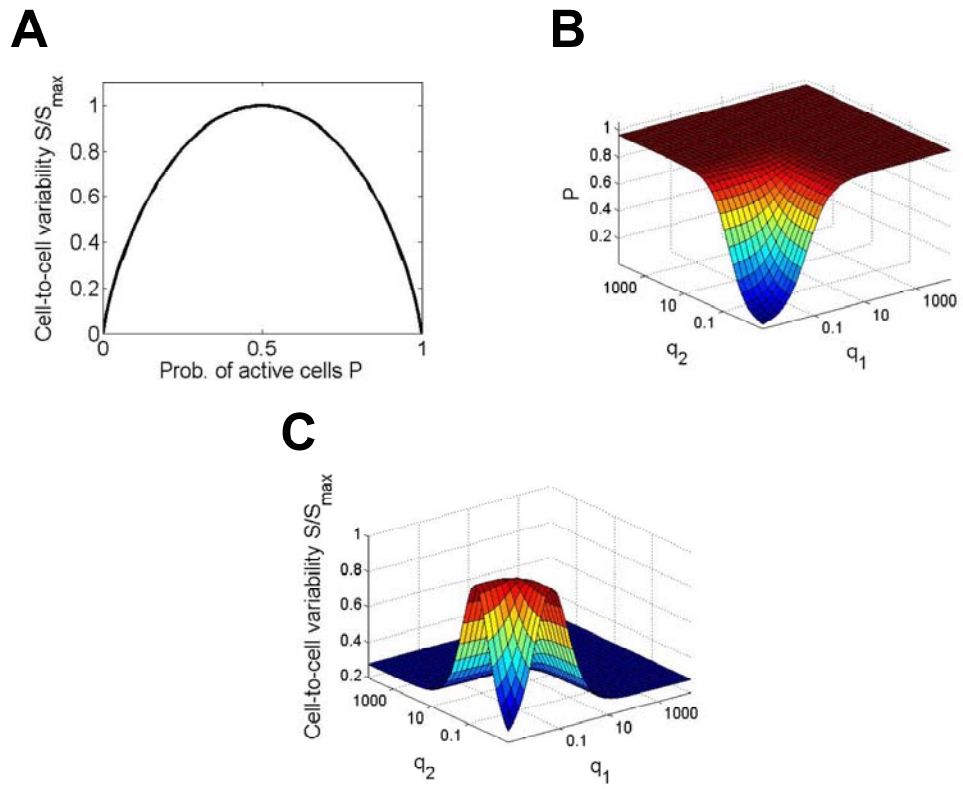


Figure 5

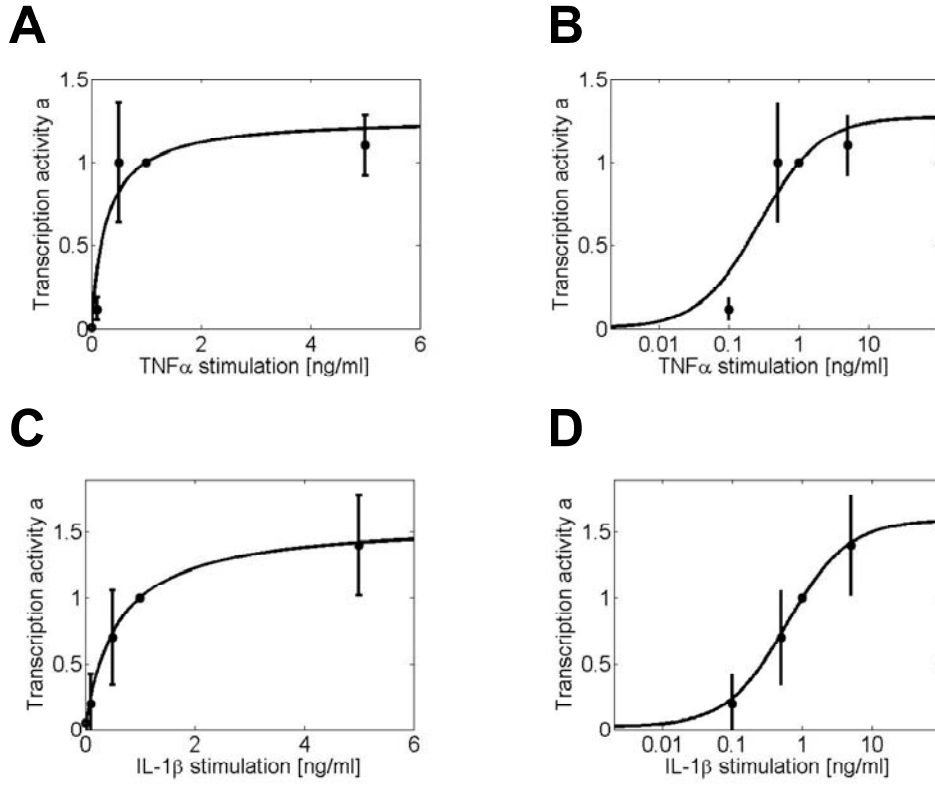


Figure 6

