

Bio-Logic: Gene Expression and the Laws of Combinatorial Logic

Maria J. Schilstra¹

Biological and Neural Computation Group
Science and Technology Research Institute
University of Hertfordshire
College Lane
Hatfield, Hertfordshire AL10 9AB
United Kingdom
Email: M.J.1.Schilstra@herts.ac.uk
Telephone: () 44 1707 281376
Fax: () 44 1707 284156

Chrystopher L. Nehaniv
Adaptive Systems & Algorithms Research Groups
School of Computer Science
University of Hertfordshire
College Lane
Hatfield, Hertfordshire AL10 9AB
United Kingdom
Email: C.L.Nehaniv@herts.ac.uk

Abstract

At the heart of the development of fertilized eggs into fully formed organisms and the adaptation of cells to changed conditions are genetic regulatory networks (GRNs). In higher multi-cellular organisms, signal selection and multiplexing is performed at the *cis*-regulatory domains of genes, where combinations of transcription factors (TFs) regulate the rates at which the genes are transcribed into mRNA. To be able to act as activators or repressors of gene transcription, TFs must first bind to target sequences on the regulatory domains. Two TFs that act in concert may bind entirely independently of each other, but more often binding of the first one will alter the affinity of the other for its binding site. This paper presents a

¹ Corresponding author

systematic investigation into the effect of TF binding dependencies on the predicted regulatory function of this “bio-logic”. Four extreme scenarios, commonly used to classify enzyme activation and inhibition patterns, for the binding of two TFs were explored: *independent* (the TFs bind without affecting each other’s affinities), *competitive* (the TFs compete for the same binding site), *ordered* (the TFs bind in a compulsory order), and *joint* binding (the TFs either bind as a preformed complex, or binding of one is virtually impossible in the absence of the other). The conclusions are: 1) the laws of combinatorial logic hold only for systems with independently binding TFs; 2) systems formed according to the other scenarios can *mimic* the functions of their Boolean logical counterparts, but cannot be combined or decomposed in the same way; and 3) the continuously scaled output of systems consisting of competitively binding activators and repressors can be more robustly controlled than that of single TF or (quasi-) logical multi-TF systems.

Keywords: Transcription regulation, Genetic regulatory networks, Enzyme kinetics, Combinatorial logic, Non-Boolean continuous logic, Modelling.

1 Introduction

Apart from blueprints for proteins - their so-called ‘code’ - many genes also carry information that ultimately directs where, when, and to what extent they will be expressed. A gene is said to be *expressed* when, at a particular time in a particular cell, transcripts of its code are detectably present. Because these transcripts are, in general, short lived, the fact that they can be detected indicates that the gene is currently being transcribed, and suggests that the transcripts are also being translated into protein.

Central to the control of gene expression is the interaction of so-called *trans*-regulatory factors with non-coding, regulatory domains of genes (see e.g. [7, 11] for recent reviews). The group of *trans*-regulatory factors considered here affect the transcription initiation rate, and are called transcription initiation factors, *transcription factors* (TFs) for short. A TF is a protein that recognizes and binds to a so-called *cis*-regulatory element, a particular nucleotide sequences in a gene’s regulatory domain. Bound TFs repress or stimulate the gene’s transcription. Some TFs act by modifying the assembly rate of RNA polymerase, the molecular motor that carries out the transcription itself. Others increase or decrease the rate at which fully assembled RNA

polymerase complexes escape from the assembly site and begin transcribing, or stimulate abortion of the transcription.

As TFs are gene products themselves, their own expression is also subject to TF control. Fine-regulation can take place on any post-transcriptional level, but primary control of gene expression is at the level of TF binding. Molecular signals of intra- or extra-cellular origin often bring about changes in the effectiveness, location, or lifetime of particular TFs, and as a result modify the expression levels of many genes. Different cells within the same organism are exposed to different signals (or signal strengths) and, because of their different history, may respond differently to the same signals, and this lies at the basis of differentiation and development. Genes, TFs, signal transduction chains, and the machinery that links these components form genetic regulatory networks (GRNs), which ultimately determine how cells divide, respond to stimuli, and differentiate; and how organisms develop and evolve.

In multi-cellular eukaryotes, *cis*-regulatory domains often have binding sites for many different TFs. Such arrangements function as “biological multiplexers” whose output (expression level) depends on the combined input (current local TF composition) that they receive through their multiple entry channels (TF binding sites).

Experimental techniques to record gene expression levels in different organisms, tissue types, or even in individual cells have undergone rapid development over the last decade. It is now possible to study the changes in gene expression after a perturbation of the status quo, or upon the contraction of a disease, simultaneously for hundreds of genes and many cell types. Statistical techniques are employed to correlate and classify expression patterns, but explaining the observed patterns, and making predictions that are more than mere interpolations requires models that relate gene expression to network structure and function. Ideally, such models capture the relevant aspects of the biochemistry of gene expression, but, faced with the absence of thermodynamic and mechanistic information, model builders often seek analogies outside the realm of biochemistry, in man-made systems such as electronic circuits. Similarly, engineers, computer scientists, and mathematicians look at biological systems for inspiration, to discover mechanisms that could enrich their gamut of design principles, control mechanisms, etc.

One of the greatest attractions of electronic systems is that their basic principles are, on the whole, well-understood, and that they come with “ready-made” rules and principles that beg to be applied to biological systems. Not surprisingly, biological and computational scientists alike have tried to apply rules that hold for digital and electronic networks to biological systems. The GRN modelling frameworks proposed by [4-6, 8-10, 12-14, 17, 19, 20] (see also the recent reviews by [18] and [15]) all apply at least some of the principles of combinatorial logic to describe the combined effects of a few *cis*-regulatory interactions per gene, whereas [20, 21] have sought to express a much more complex system, in which a dozen or so *cis*-regulatory elements determine the expression of the sea urchin *endo-16* gene, in this way.

However, “paradigm transplantation” – the wholesale transfer of theories and rules that were developed for one area to another field - may distort one’s vision, and unnecessarily restrict further development of the field that has adopted the foreign principles. In this paper, we analyse the “logic” behind the regulation of gene expression by TF combinations, point out parallels and differences with digital electronic logic, and make tacit assumptions explicit.

2 Transcriptional regulation and combinatorial logic

2.1 Overall dynamics of gene expression

A GRN is a structure that links the TF composition at particular times in particular cells to gene expression levels in those cells. In this paper, we will concentrate predominantly on the “static” characteristics of transcriptional regulation – the way in which TFs combine to generate a certain effect – but to set the scene we first give a brief overview of the most important dynamic aspects of GRNs.

[Figure 1 approximately here]

The diagram in Figure 1 shows the basic participants and processes in the expression of one gene, *gX*. Circles indicate molecular species; rectangles stand for processes. Solid arrows starting or ending at a vertical edge of the process rectangles indicate the production or consumption of material by the process; open arrows that end on a horizontal edge of a process rectangle denote a regulatory interaction. The molecular species that function as regulators modify the *rate* of a particular process, but are not themselves consumed in that

process. The transcription of gene gX is regulated by a number of transcription factors, whose interaction and effect we will concentrate on in the sections following this one. The transcription product is messenger RNA mX , which in turn functions as a regulator in its own translation. The translation product is the protein pX , which may function as a TF in the transcription of other genes. In this example, pX does not affect the rate of transcription of its “own” gene (gene X), but in reality many TFs do, directly or via longer feedback loops. Apart from genes, which, in this model, live forever, all other molecular species have finite lifetimes.

Technically, the transcription and translation processes consume material as well: the building blocks of RNA (nucleotides) and proteins (amino acids), and energy carriers (such as ATP). They also produce more than only mRNA and protein molecules (namely “waste”, such as pyrophosphate, water, and heat), but the building blocks are regenerated from the decay products, and the waste is reused or otherwise dealt with, and we shall assume that their overall effect is constant.

When gX is switched from “off” to “on”, for instance as a result of the reception of a specific signal at the cell surface, production of mX will start. Because mX molecules have a finite lifetime, t_{mX} , they will disappear after some time. Some time after the activation of gX , the system will reach a steady state, in which the number of mX molecules that is produced per unit of time equals the number of molecules that disappears. The steady state concentration of mX (number of molecules per unit of volume) will remain constant until the production or breakdown rates change, and the steady state concentration of mX , $[mX]_{ss}$, is equal to $v_x t_{mX}$, where v_x is the transcription rate of gene gX (concentration is the number of molecules or moles per unit of volume, and is conventionally indicated by square brackets). Similarly, $[pX]_{ss}$, the steady state concentration of pX , is $[pX]_{ss} = v_p t_{pX} [mX]$, where t_{pX} is the average lifetime of pX , and v_p is the rate at which pX is produced per molecule of mX . Thus, when both mX and pX have reached their steady state concentrations, $[pX]_{ss} = v_x v_p t_{mX} t_{pX}$.

It is useful to be aware of the following. 1) In simple irreversible production-decay sequences such as those in Figure 1, the transient time (the time required to reach a new steady state) depends only on the lifetimes of the species, not on their production rate. Typically, mRNA molecules live for at least a few hours, so that expression levels are not expected to stabilize for several hours after a change in the transcription rate has occurred. 2) Transcription and

translation, mRNA and protein breakdown, and, in eukaryotes, transport and modification of the transcripts, are complex multi-step processes, and their dynamics may be more accurately described by delay functions than by the type of equations used here and in most other studies. 3) The functionality of mRNA and proteins can be changed rapidly in specific post-transcriptional and post-translational processes. Rapid changes in the functionality of certain gene products may be coupled to much slower changes in gene expression levels. 4) Feedback loops are a prominent feature of all but the simplest GRNs (cf. [2, 6, 19]), and their presence, particularly in combination with delays, can give rise to much more complex behaviour - oscillatory, switch-like, or even chaotic – than the simple linear dynamics described above.

2.2 Transcription factor binding

From here on, we shall concentrate on the interactions of the TFs and the *cis*-regulatory domain of the gene and their regulatory effect. As stated above, TFs must be bound – directly, or via other TFs - to the gene’s regulatory domain to exert their effect on transcription initiation. Firstly we explain the basic thermodynamics of complex formation; more comprehensive treatments may be found in physical chemistry or enzyme kinetics textbooks.

2.2.1 Binding of one TF

Unless a covalent bond is created, the binding of one biomolecule to another is usually a dynamic process, and involves continuous association and dissociation of the molecules that are involved in the process. In the simplest possible situation, a single type of TF, A, forms a complex with a *cis*-regulatory domain D of a particular gene: $D + A \rightleftharpoons DA$. Complex formation occurs at a rate $k_a[A][D]$, and dissociation at a rate of $k_d[DA]$ (again, square brackets indicate concentration), k_a is the association rate constant, and k_d the dissociation rate constant of the complex. At equilibrium, the forward and reverse rates are equal: $k_a[A]_{\text{eq}}[D]_{\text{eq}} = k_d[DA]_{\text{eq}}$, where the subscript eq indicates equilibrium concentrations. The thermodynamic *equilibrium dissociation constant*² for the complex DA, $K_{DA} = k_d/k_a$, is a reciprocal measure for A’s affinity

² K_{DA} has units of concentration, and is equal to the concentration of A at which its binding site on D is half saturated if the total concentration of A is much larger than that of D. Its reciprocal, k_d/k_a , is known as the equilibrium association constant, also referred to as the binding constant.

for D. A smaller value indicates a higher affinity, and the higher a molecule's affinity for another, the more time they will on average spend as a complex. K relates to the Gibbs free energy (ΔG) of complex formation as $K_{DA} = \exp[-\Delta G_{DA}/RT]$, where R is the universal gas constant, and T the absolute temperature. Equilibrium constants are determined by the nature of the complexes (and by external factors such as temperature, pH, ionic strength, etc., which we shall assume to be constant), but not by the total concentrations of their constituents.

If A_T and D_T are the total concentration of A and D ($A_T = [DA] + [A]$; $D_T = [DA] + [D]$), and the equilibrium dissociation constant is known, then, at equilibrium, the fraction of D that has bound A, $[DA]_{eq}/D_T$ is obtained by solving the expression for K_{DA} with the appropriate substitutions for $[DA]_{eq}$. The general solution is a root of a quadratic equation, but if it is taken into account that the number of binding sites is very small (usually no more than a few per nucleus), and that the A is normally present in much larger quantities ($A_T \gg D_T$, so that $A_T - [A]_{eq} \cong A_T$), the expression reduces to a simple hyperbolic equation:

$$\alpha = \frac{a}{1 + a} \quad (1)$$

Here a is the normalized concentration of A (to the equilibrium dissociation constant for the complex AD, $a = A_T/K_{DA}$) and α is the "fractional saturation" of the binding site for A. If we further assume that averages over a sufficiently long period of time are equal to averages over many molecules (ergodicity), the fractional saturation is equal to the fraction of time A spends as a part of the complex AD, and the fraction of time that A is not bound is equal to $1 - \alpha$.

2.2.2 Binding of two TFs

If another TF, B, also binds to D, five more association reactions must be considered: $D + B \Leftrightarrow DB$; $DA + B \Leftrightarrow DAB$; $DB + A \Leftrightarrow DAB$, $A + B \Leftrightarrow AB$, and $D + AB \Leftrightarrow DAB$ (we disregard further isomerization reactions such as DNA looping). First, we consider the case in which A and B have an infinitely small tendency to form a complex "in solution", away from D (so that the last two reactions in the above list do not occur). We denote the equilibrium dissociation constant for DB as $K_{DB} = [D][B]/[DB]$, B's total concentration as B_T , and, assuming that $B_T \gg$

D_T , normalized [B] as $b = B_T/K_{DB}$. Furthermore, we need to take into account that the equilibrium dissociation constant for $DA + B \Leftrightarrow DAB$, $K_{DBA} = [DA][B]/[DAB]$, is not necessarily equal to K_{DB} , and we use the symbol r to indicate the ratio K_{DB}/K_{DBA} (which must be equal to K_{DA}/K_{DAB} where $K_{DBA} = [DB][A]/[DAB]$). A value of r that is different from 1 indicates a dependence in the binding of A and B to D: if $r > 1$, the presence of bound A stimulates the binding of B; if $r < 1$, bound A inhibits the binding of B to a certain extent. The fractional saturation α of the binding site for A is now equal to the fraction of time that the complex spends as DA or DAB, and the same holds for β , the fractional saturation of the binding site for B:

$$\mathbf{a} = \frac{a + ab}{1 + a + b + rab}, \quad \mathbf{b} = \frac{b + ab}{1 + a + b + rab} \quad (2)$$

In Table 1, we have listed the expressions for α and β for the extreme cases in which A and B bind independently ($r = 1$), competitively ($r = 0$), in compulsory order ($K_B \rightarrow \infty$, with b redefined as B_T/K_{DBA}), and in a highly dependent way ('coincident'), where A has virtually no affinity for the target in the absence of B, and vice versa ($K_A, K_B \rightarrow \infty$; a and b redefined as A_T/K_{DAB} and B_T/K_{DBA}). Also listed are the expressions for the fraction of time that neither A nor B is bound (φ_D), only A, but not B, or B but not A (φ_{DA} and φ_{DB}), and both A and B simultaneously (φ_{DAB}).

In scenario 1, in which the TFs bind independently, the fraction of time that, for example, only A and not B is bound, is calculated by multiplying the probability that A is found bound to the *cis*-regulatory domain (i.e. α) with the fraction of time that B is not bound ($1 - \beta$). The expressions for φ are readily generalized to n independently binding TFs. In scenario 4 (joint binding), the TFs act as if they are a single entity, and may function as such in the generalized expressions for scenario 1. However, these simple relationships do not hold in scenarios 2 and 3, and the fraction of time that, for instance, DA exists (φ_{DA}) is also dependent on the amount of B present.

If it is taken into account that A and B may form a complex away from, as well as on D, the general expressions for α and β become even more complex. However, in the extreme case

that the equilibrium dissociation constant for the ‘preformed’ complex AB, K_{AB} , is very small, the expressions can be significantly simplified. Under these conditions A and B form a very strong complex, and only ever bind to D as the complex AB, not separately, and therefore act essentially as a single species. In that case, here listed as scenario 5, the total amount of preformed AB is expressed as $\min(A_T, B_T)$, and the normalized concentrations a and b are both redefined as $\min(A_T, B_T)/K_{DAB}$ where K_{DAB} is the equilibrium dissociation constant associated with the reaction $AB + D \Leftrightarrow DAB$.

2.2.3 Binding of TFs of the same type

If A and B are equal, the equations of Table 1 of course still hold, although the expression for a in scenario 5 (in solution) becomes $a = A_T/2K_{DAB}$. Thus, if binding of the first molecule of A does not affect the binding characteristics of the second, $\varphi_{DA2} = \alpha^2 = a^2/(1+a)^2$; if the binding is highly dependent (“coincident” binding), $\varphi_{DA2} = \alpha = a^2/(1+a^2)$, and if A binds only as preformed A_2 , $\alpha = a/(1+a)$ again. These equations are easily generalized to n molecules of A, where n indicates the multiplicity. The equation $\varphi_{DA_n} = a^n/(1+a^n)$ is often used to generate sigmoid dependencies, particularly when some kind of threshold function is required to defend a particular point of view. The exponent n , which determines the steepness of the function, is often referred to as the Hill coefficient. It is useful to be aware of the fact that using the value of n implies that an $n + 1$ body collision (between n TFs and one target; hence the term coincident binding) is required to form the complex. If the TFs bind as a preformed homooligomer, the multiplicity, and therefore the Hill coefficient, is one.

2.2.4 Binding of multiple TFs

The expressions for α and β in equation 2 are readily generalized to more than two different TFs. Binding of each individual TF potentially affects the affinity of all others, and it will be necessary to define equilibrium constant ratios for all potential complexes. The expressions in the five scenarios may be expanded and combined, but in the following we shall focus on situations in which no more than two TFs are involved in the stimulation or repression of gene expression.

2.3 Regulation of gene expression

Above we have described and classified the various ways in which TFs interact with their binding sites and each other to form particular complexes, and how the fraction of time that each of the complexes exist depends on the concentrations of the individual TFs. In this section, we discuss how the observation that a TF, or a combination of TFs, act to stimulate or repress gene transcription could relate to these binding patterns.

Little is known about the precise way in which TF combinations influence the gene transcription rate, and here we will take for granted that some stimulate it, and others repress it. There is no rule that tells how the fractional saturation ϕ_{DX} for a complex DX relates to transcription rate. For practical purposes it is usually assumed that transcription rate is simply proportional to fractional saturation, and this approach was first used to quantitatively account for the observed control of gene expression by the bacteriophage λ repressor ([1, 16]; see also [2] and references therein). Although this assumption has – often tacitly - been made in many later studies, we would like to point out that there is, in fact, little evidence that it is entirely justified in more complex eukaryotic regulatory systems.

Quite independently of the precise relationship between TF binding and transcription rate, we argue that the number of output combinations that can be addressed is restricted by the binding scenario. In the following, we shall assume that there is a certain maximum rate, V_{max} , at which gene transcription can be initiated (which is, among other factors, determined by the size of the RNA polymerase that carries out the transcription, and by the speed at which it vacates its assembly site). We define as a “pure activator” a TF that, when present in sufficient quantities to achieve maximal saturation of its binding site (according to Table 1), is able to raise the transcription rate r_{ini} (and with it the expression level), from 0 to the maximum possible value V_{max} . A “pure repressor” is a TF that will decrease a gene’s expression level from the maximum to 0 when its binding site is fully saturated. Pure activators and repressors do not necessarily achieve their effect by themselves: they may do so only in combination with other TFs.

2.3.1 All-or-none stimulation or repression of gene expression

Firstly, we shall consider all-or-none situations, in which TFs are present in such quantities that their binding sites are fully saturated. In Figure 2, we have listed the output patterns that can be achieved by combinations of two TFs, each of which is a pure activator or pure repressor, for the independent, competitive, ordered, and joint (coincident, and as a complex) binding scenarios. Activators or repressors may achieve their effect alone (disjunctively) or in combination with other activators or repressors (conjunctively). A gene that needs activation is expressed only when the activator is present, whereas a gene that is regulated by repression is expressed when the repressor is absent.

[Figure 2 approximately here]

The symbols in Figure 2 represent the different configurations of activators and repressors and binding mode, and should be read as follows. Horizontal lines indicate genes, and the position where the bent arrow crosses the horizontal line symbolizes the position on the gene where the transcription begins. The portion of the gene to the left (“upstream”) of the promoter indicates its regulatory domain, and the region to the right is the “downstream” region that contains the actual blueprint for a protein. The gene’s transcription is controlled by two TFs, A and B, represented by the vertical lines that end in an arrowhead or a bar. An arrowhead indicates the TF is an activator, a bar represents repressing activity. TFs that end above two different positions on the regulatory domain bind independently of each other; TFs that end above the same position bind according to one of the other scenarios, as indicated in the header of the figure.

Each input pattern (00: both TFs absent; 10: A present, B absent; 01: A absent, B present; 11: both present) will result in a particular output pattern (0: gene not expressed, 1: gene expressed), and the output patterns for eight configurations, together with their Boolean logic formulation, are listed underneath the diagrams that represent the configurations in which the TFs bind independently. Configurations in the same box have the same output pattern. For TFs that bind independently, it is necessary to specify whether they act alone (disjunctively) or together with the other (conjunctively); in the other cases the action of combinations of activators and inhibitors is determined by their binding mode.

Thus, the third row of symbols in Figure 2, for example, symbolizes the configurations in which A is a repressor and B is an activator. If A and B bind independently, and act alone (first column), the gene will be transcribed (output 1) when either A is absent, or when B is present (even if A is there as well). The gene is only repressed (output 0) when A is present (1), and B is not (0). The output pattern is therefore 1011 for inputs 00, 10, 01, 11. (Formally, the absence of a repressor that is observed to work alone is sufficient to allow the gene to be transcribed). A similar behaviour is predicted to be generated by two TFs that bind in compulsory order, if the one that binds first, A, is a repressor, and B, the one that binds only when A is already bound, is observed to be an activator (and is indicated as such in the diagram, but is in fact an inhibitor of the repressive action of A). Therefore, A will repress gene expression by itself, but loses its repressive power when B is present. Therefore, the output pattern is for this configuration is 1011 as well.

If A and B bind independently, but act in conjunction (column 5), the gene will only be transcribed when B is present, but A is not (output pattern 0010). This illustrates the fact that the absence of a repressor that works in conjunction is formally insufficient to activate a gene. In column 7, where A and B bind in compulsory order, with B as the primary activator. A inhibits the activating capability of B, so that the gene is expressed only when B is present and A is absent, and the output pattern is, again 0010.

If two competitively binding TFs are both activators (as in row 1, column 2) the gene will be activated unless neither is present, and if both are repressors (row 4, column 6), the gene will be activated only when neither is present. However, in the case in which one is an activator and the other is a repressor (rows 2 and 3, columns 2 and 6), it is not possible to decide whether the TFs act alone or together, and resolve the output pattern without further information on the relative amounts of the TFs. The two configurations that give rise to this type of logic (rows 2 and 3) are depicted in grey, and situated in the appropriate columns (2 and 6).

2.3.2 The logic of gene expression

[Figure 3 approximately here]

Figure 3 shows that the fractional saturation for the 0100 (A and not B, Figure 2, row 2, column 5) output pattern is almost indistinguishable from that of and activator and repressor that bind in compulsory order (row 2, column 7). Apart from the scenarios in which an activator competes with a repressor (see below), this is true for all output patterns within the same group listed in Figure 2, and it would be tempting to suppose that combinations within a group are fully equivalent (cf. e.g. [9]).

However, the essence of combinatorial (Boolean) logic is formulated in the De Morgan's theorems: NOT A AND NOT B is equivalent with NOT (A OR B), and NOT A OR NOT B with NOT (A AND B). Inspection of Table 1 in combination with Figure 2 shows that the above theorems hold for independently binding TFs, but not for any of the other modes. For example: for independently binding TFs, NOT A AND NOT B, $(1 - \alpha)(1 - \beta)$, can be rewritten as $1 - (\alpha + \beta(1 - \alpha))$, which is equivalent to NOT (A OR B). However, NOT (A OR B) translates into $1 - \alpha - \beta$ when the TFs bind in competitive fashion, which is evidently not equal to NOT A AND NOT B, which for this scenario also translates as $(1 - \alpha)(1 - \beta)$. Furthermore, NOT A OR NOT B, $(1 - \alpha) + (1 - \beta) - (1 - \alpha)(1 - \beta)$ for independently binding TFs is rewritten as $1 - \alpha\beta$, i.e. NOT (A AND B), whereas for repressors that bind in compulsory order NOT (A AND B) translates into $1 - \beta$, and is, therefore, equivalent to NOT B. Thus, the rules that hold for Boolean logic may be applied to independently binding TFs, but cannot be applied in their entirety to TFs that bind according to the other scenarios. In "bio-logic", the logic of gene expression, the configurations listed in columns 2-4 and 6-8, must be considered as distinct non-Boolean operators.

2.3.3 Addressing intermediate expression levels

In general, the saturation level of the binding site for a particular TF (see Table 1) is a hyperbolic or sigmoid function of TF concentration. If its concentration in a particular cell is in the vicinity of its equilibrium constant, its binding site will not be occupied full time. If the level of activation or repression of transcription initiation is proportional the average fraction of time that some activating or repressing complex exists, the expression level may be regulated by varying the concentration of TFs. However, this process is likely to have a limited "dynamic range", for one or more of the following reasons. 1) If a TF has a high affinity for its binding site, a few molecules may be already sufficient to fully saturate the binding site. 2) If the

saturation curve is sigmoid, expression levels may change quite dramatically as a result of a relatively small change in TF concentration. 3) The relationship between fractional saturation modifier effect is, as pointed out above, likely to be non-linear, and full activity may be achieved well before the binding site is fully saturated. Therefore, variation of absolute TF concentrations is probably unsuitable as a robust general mechanism for fine-tuning of gene expression levels. Figure 3 shows intermediate output values are addressed over a relatively narrow normalized concentration range (a and b) for an activator-repressor pair that bind independently and act conjunctively, or bind in compulsory order (activator first). For both the repressor dominates when a and b are both high.

However, expression levels can be regulated in a potentially much more precise and robust way by varying the concentrations of two TFs with opposite effects that compete for a single binding site on the *cis*-regulatory domain. In that case (Figure 3, bottom row), intermediate values can be addressed by varying the *ratio* of a and b , a quantity that is more readily maintained, stabilized, and controlled (for instance by reversible modification) than absolute concentrations.

2.3.4 Truth tables

Activation and repression do not have to be extreme: each combination of bound TFs may modulate the rate of transcription initiation to a certain extent. To express this, we assign a *modulation coefficient*, e_0 , e_A , e_B , or e_{AB} , to each of the states D, DA, DB, and DAB. The value of a modulation coefficient is determined solely by the nature of the complex with which it is associated, not by the fraction of time that the complex is present. If we assume that the number of times gene transcription is initiated by a particular complex is proportional to the fraction of time that particular complex exists, and that the average initiation rate, r_{ini} is proportional to the sum of the rates of the individual contributions:

$$r_{ini} = V_{\max} (\epsilon_0 \varphi_D + \epsilon_A \varphi_{DA} + \epsilon_B \varphi_{DB} + \epsilon_{AB} \varphi_{DAB}) \quad (3)$$

Even though the use of Boolean logical operators to describe the combined effect of multiple TFs on gene transcription is not always appropriate, their action is probably captured most efficiently in truth tables. In these, the input is formed by the normalized concentration of all

TFs that interact with the regulatory site of the gene for which the table has been set up, and the output value is calculated by applying equation 3 as appropriate for the binding mode (generalized to the total number of TFs involved).

2.4 Discussion

There is an obvious similarity between the regulatory elements of genes and logic gates: two TFs may stimulate the expression of a gene only when both are present – in which case it is tempting to describe their interaction as logical conjunction ('AND'). Some TFs may appear to act irrespective of the presence of others – disjunctively; logical 'OR' -, and some clearly function to stop or repress gene expression: negation, or logical 'NOT'. Several earlier studies have categorized particular observed *cis*-regulatory functions as Boolean logic operations [4, 5, 9], but in this paper we have approached the subject from the opposite direction. Rather than classifying observed behaviour, we have explored all four extreme binding scenarios for two TFs, either one an activator or a repressor (with the binding scenario reflected in the denominators, and functionality in the numerators for α and β in Table 1), and found the following:

1. The rules of combinatorial logic only hold for combinations of independently binding TFs.
2. Two activators or repressors that bind jointly mimic logical AND or NAND operations, respectively, whereas competitively binding activators or repressors will mimic OR or NOR. The term *mimic* is used to indicate that De Morgan's laws do not strictly apply.
3. Two TFs that bind in compulsory order can mimic the behaviour of all combinations of independently binding activators and inhibitors. To predict the output pattern of two TFs that bind independently, it is necessary to know whether they act alone or in conjunction (Figure 2, left and right panels). Two TFs that bind in compulsory order always act together, but to predict their output pattern, it is necessary to know which one is the primary actor.
4. The situation in which an activator and a repressor compete for the same binding site produces a non-Boolean continuous logical aspect: the output cannot be predicted without knowing the relative normalized concentrations of activator and repressor. This scenario, which is probably the most interesting from an engineering point of view,

provides the system with the capability of robustly and precisely addressing intermediate expression levels, simply by varying the ratio of two TFs that have different effects (namely repression and activation). The competing TFs need not be the products of different genes: one can be a post-translationally modified version (e.g. phosphorylated), and such modification and its reversal may even occur whilst the TF is bound to its target site.

If the combined effect of two TFs is larger than the sum of their individual effects, the TFs are said to act in “synergy”. Synergy may be observed when binding of one TF strengthens the binding of the other, so that there is more of the fully formed complex than could be expected on the basis of the equilibrium dissociation constants of the individual TFs. It could also simply be observed when the fully formed complex initiates transcription more efficiently than any of the other complexes. In practice it may be almost impossible to distinguish between, or establish the relative contributions of these two mechanisms. If two TFs have similar effects, but act in different developmental stages or in different types of tissue, they probably bind and act truly independently. Their binding sites are likely to be part of different *cis*-regulatory modules [3], and the use of the true logical OR operator to describe their combined effect is almost certainly justified.

Acknowledgment

The authors would like to thank Hamid Bolouri (University of British Columbia, Vancouver) and Eric H. Davidson (Caltech, Pasadena, CA) for pioneering and establishing the concept of *cis*-regulatory logic, and for stimulating discussions. This work was supported by a Wellcome Trust Functional Genetics Project Grant (072930/Z/03/Z).

References

1. Ackers, G. K., Johnson, A. D., & Shea, M. A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the National Academy of Sciences USA* 79, 1129-1133.
2. Alon, U. (2006), *An introduction to systems biology. Design principles of biological circuits*. Mathematical and computational biology, London: Chapman & Hall/CRC.

3. Arnone, M. I. & Davidson, E. H. (1997). The hardwiring of development: Organization and function of genomic regulatory systems. *Development* 124, 1851-1864.
4. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., & Phillips, R. (2005). Transcriptional regulation by the numbers: Models. *Current Opinion in Genetics & Development* 15, 116.
5. Buchler, N. E., Gerland, U., & Hwa, T. (2003). On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences USA* 100, 5136-5141.
6. Chaouiya, C., Remy, E., Ruet, P., & Thieffry, D. (2004). Qualitative modelling of genetic networks: From logical regulatory graphs to standard petri nets. In J. Cortadella & W. Reisig (Eds.), *Lecture notes in computer science* (pp. 137-156). Springer-Verlag: Berlin.
7. Davidson, E. (2001), *Genomic regulatory systems.*, San Diego CA: Academic Press.
8. de Jong, H., Geiselman, J., & Thieffry, D. (2003). Qualitative modeling and simulation of developmental regulatory networks. In S. Kumar & P. J. Bentley (Eds.), *On growth, form, and computers* (pp. 109-143). Academic Press: London.
9. Istrail, S. & Davidson, E. H. (2005). Gene regulatory networks special feature: Logic functions of the genomic cis-regulatory code. *Proceedings of the National Academy of Sciences USA* 102, 4954-4959.
10. Kauffman, S. A. (1993), *The origins of order. Self-organization and selection in evolution*, New York, USA: Oxford University Press.
11. Lemon, B. & Tjian, R. (2000). Orchestrated response: A symphony of transcription factors for gene control. *Genes & Development* 14, 2551-2569.
12. McAdams, H. H. & Shapiro, L. (1995). Circuit simulation of genetic networks. *Science* 269, 650-656.
13. Mestl, T., Plahte, E., & Omholt, S. W. (1995). A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology* 176, 291-300.
14. Mjolsness, E. D., Sharp, D., & Reinitz, J. (1991). A connectionist model of development. *Journal of Theoretical Biology* 152, 429-453.
15. Schlitt, T. & Brazma, A. (2005). Modelling gene networks at different organisational levels. *FEBS Letters* 579, 1859.

16. Shea, M. A. & Ackers, G. K. (1985). The or control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of Molecular Biology* 181, 211.
17. Shmulevich, I., Dougherty, E. R., Kim, S., & Zhang, W. (2002). Probabilistic boolean networks: A rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18, 261-274.
18. Styczynski, M. P. & Stephanopoulos, G. (2005). Overview of computational methods for the inference of gene regulatory networks. *Computers & Chemical Engineering* 29, 519.
19. Thomas, R. & D'Ari, R. (1990), *Biological feedback*, Boca Raton, FLA: CRC Press.
20. Yuh, C.-H., Bolouri, H., & Davidson, E. H. (1998). Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science* 279, 1896-1902.
21. Yuh, C.-H., Bolouri, H., & Davidson, E. H. (2001). Cis-regulatory logic in the endo16 gene: Switching from a specification to a differentiation mode of control. *Development* 128, 617-629.

Figure 1. Schematic representation of the main processes (rectangles) and molecular participants (circles) in the expression of a single gene. Closed arrows indicate production or consumption of material; open arrows signify regulatory interactions. The product of the transcription of a gene (here gX) is a specific type of mRNA (mX), and the product of mX translation is the protein pX . The rate at which the gene is transcribed is regulated by so-called transcription factors (TFs). The transcription of a single gene may be controlled by many TFs, acting alone or in combination. TFs are proteins themselves, and are, therefore produced in a translation process. Apart from the gene itself, all participants have a limited lifespan, and disappear as a result of specific decomposition processes.

Figure 2. Gene expression patterns for combinations of two activators (row 1), an activator and a repressor (rows 2 and 3) and two repressors, in which the TFs bind and act as indicated in the header. Configurations in the same box are equivalent; grey symbols indicate configurations whose output is not fully specified. The scenarios in which the TFs bind jointly (either coincident or as a preformed complex, scenarios 4 and 5) are equivalent, and have been listed in a single column. See text for a full explanation.

Figure 3. Comparison of r_{ini}/V_{max} for the 3 configurations in the 0100 group of Figure 2. Row 1 (top): an independently binding activator and repressor that act in conjunction ($a/(1+a+b+ab)$, see Table 1); Row 2: an activator and repressor that bind in compulsory order ($a/(1+a+ab)$); Row 3: an activator and repressor that compete for a binding site ($a/(1+a+b)$). Column 1 (left) indicates the symbols for the three combinations, and columns 2 and 3 show the values of r_{ini}/V_{max} at low ($0 \leq a, b \leq 1$, column 2), and high ($0 \leq a, b \leq 1000$, column 3) normalized concentrations.

Table 1. Definition of TF binding scenarios and associated variables

	1. Independent	2. Competitive	3. Ordered	4. Joint (coincident)	5. Joint (as a complex)
	D+A \Leftrightarrow DA D+B \Leftrightarrow DB DA+B \Leftrightarrow DAB DB+A \Leftrightarrow DAB	D+A \Leftrightarrow DA D+B \Leftrightarrow DB	D+A \Leftrightarrow DA DA+B \Leftrightarrow DAB	D+A+B \Leftrightarrow DAB	D+AB \Leftrightarrow DAB
a	A_T/K_{DA}	A_T/K_{DA}	A_T/K_{DA}	A_T/K_{DAB}	$\min(A_T, B_T)/K_{DAB}$
b	B_T/K_{DB}	B_T/K_{DB}	B_T/K_{DBA}	B_T/K_{DBA}	$\min(A_T, B_T)/K_{DAB}$
α	$\frac{a}{1+a}$	$\frac{a}{1+a+b}$	$\frac{a+ab}{1+a+ab}$	$\frac{ab}{1+ab}$	$\frac{a}{1+a}$
β	$\frac{b}{1+b}$	$\frac{b}{1+a+b}$	$\frac{ab}{1+a+ab}$	$\frac{ab}{1+ab}$	$\frac{b}{1+b}$
φ_D	$(1-\alpha)(1-\beta)$	$1-\alpha-\beta$	$1-\alpha$	$1-\alpha$ or $1-\beta$	$1-\alpha$ or $1-\beta$
φ_{DA}	$\alpha(1-\beta)$	α	$\alpha-\beta$	-	-
φ_{DB}	$(1-\alpha)\beta$	β	-	-	-
φ_{DAB}	$\alpha\beta$	-	β	α or β	α or β

Table 2. Truth tables associated with TF binding scenarios 1-4. Columns A and B indicate the presence or absence of TFs A and B; expressions in brackets indicate the value of r_{in}/V_{max} for partially occupied binding sites; bold non-bracketed expressions those for fully occupied binding sites. ϵ_0 , ϵ_A , ϵ_B , and ϵ_{AB} are the modulation coefficients (see text) for the complexes D, DA, DB, and DAB; φ_D , φ_{DA} , φ_{DB} , φ_{DAB} represent the average fraction of time these complexes exist at given concentrations of A and B (see Table 1).

A	B	1. Independent	2. Competitive	3. Ordered	4. Joint
0	0	(ϵ_0)	ϵ_0	(ϵ_0)	ϵ_0
1	0	$(\epsilon_0 \varphi_D + \epsilon_A \varphi_{DA})$	ϵ_A	$(\epsilon_0 \varphi_D + \epsilon_A \varphi_{DA})$	ϵ_A
0	1	$(\epsilon_0 \varphi_D + \epsilon_B \varphi_{DB})$	ϵ_B	$(\epsilon_0 \varphi_D + \epsilon_B \varphi_{DB})$	ϵ_B
1	1	$(\epsilon_0 \varphi_D + \epsilon_A \varphi_{DA} + \epsilon_B \varphi_{DB} + \epsilon_{AB} \varphi_{DAB})$	ϵ_{AB}	$(\epsilon_0 \varphi_D + \epsilon_A \varphi_{DA} + \epsilon_B \varphi_{DB})$	ϵ_{AB}
				$\epsilon_A \alpha + \epsilon_B (1-\alpha)$	
				$(\epsilon_0 \varphi_D + \epsilon_A \varphi_{DA} + \epsilon_{AB} \varphi_{DAB})$	ϵ_{AB}
					$(\epsilon_0 \varphi_D + \epsilon_{AB} \varphi_{DAB})$
					ϵ_{AB}

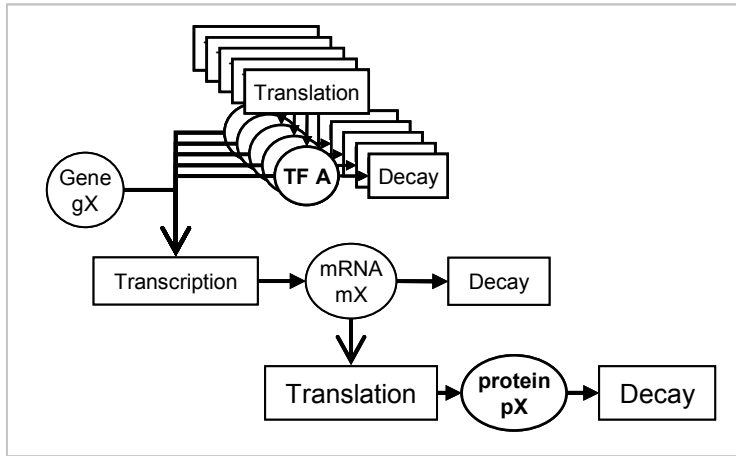


Figure 1

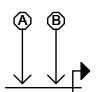
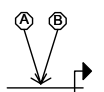
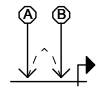
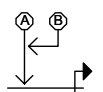
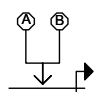

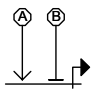
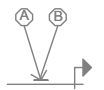
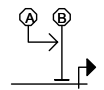
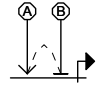
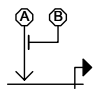


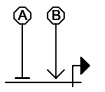
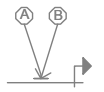
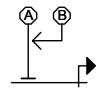
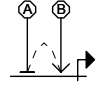
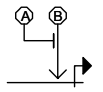


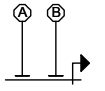
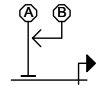
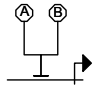
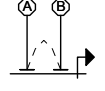

Independent (disjunctive)	Competitive	Ordered	Joint	Independent (conjunctive)	Competitive	Ordered	Joint
 0111 ($A \vee B$)				 0001 ($A \wedge B$)			
 1101 ($A \vee \neg B$)				 0100 ($A \wedge \neg B$)			
 1011 ($\neg A \vee B$)				 0010 ($\neg A \wedge B$)			
 1110 ($\neg A \vee \neg B$)				 1000 ($\neg A \wedge \neg B$)			

Figure 2

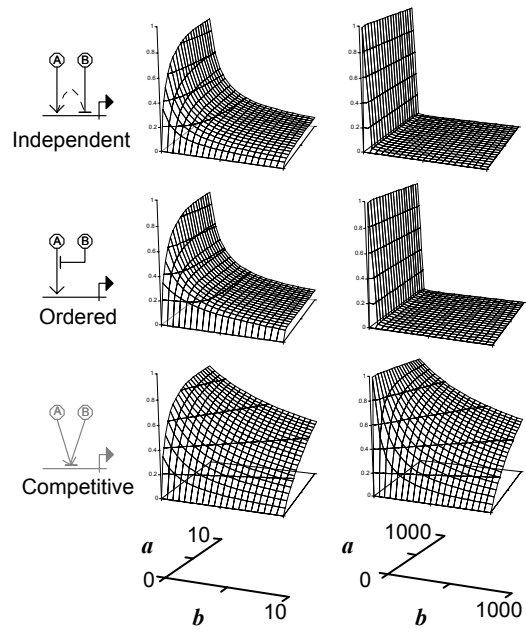


Figure 3