

On Cooperative Quasi-Equilibrium Models of Transcriptional Regulation

Eric Mjolsness

Institute for Genomics and Bioinformatics
Departments of Computer Science and Mathematics
University of California, Irvine
and Biological Network Modeling Center, Caltech

TR# 06-13

28 October, 2006

Abstract

Mechanistic models for transcriptional regulation are derived using the methods of equilibrium statistical mechanics, to model equilibrating processes that occur at a fast time scale. These processes regulate slower changes in the synthesis and expression of transcription factors that feed back and cooperatively regulate transcription, forming a gene regulation network (GRN). We rederive and extend two previous quasi-equilibrium models of transcriptional regulation, and demonstrate circumstances under which they can be approximated by feed-forward artificial neural network (ANN) models. A single-level mechanistic model can be approximated for each transcription complex by a successfully used phenomenological model of GRNs which is based on single-layer analog-valued ANNs. A two-level hierarchical mechanistic model, with separate activation states for modules and for the whole transcription complex, can be approximated for each transcription complex by a two-layer feed-forward ANN in several related ways. The sufficient conditions demonstrated for the ANN approximations correspond biologically to large numbers of binding sites each of which have a small effect. A further extension to the single-level and two-level models allows one-dimensional chains of overlapping and/or energetically interacting binding sites within a module. Partition functions for these models can be constructed from stylized diagrams that indicate energetic and logical interactions between binary-valued state variables. All parameters in the mechanistic models, including the two approximations, can in principle be related to experimentally measurable free energy differences, among other observables.

Keywords: transcriptional regulation, cooperative activation, gene regulation network, transcription complex, quasi-equilibrium, statistical mechanics, feed-forward, neural network

1 Introduction

Transcriptional regulation is mediated by a multimolecular transcription complex, for which there is no unique accepted dynamical model. We explore a quasi-equilibrium approach to constructing models of the behavior of the complex in transcriptional regulation, starting from diagrams of hypothesized activation relationships within the transcription complex, and resulting in a series of propositions that formulate such models. In Section 2 introduces the necessary equilibrium statistical mechanics in the polynomial form in which we will use it. In Section 3 reexpresses a quasi-equilibrium family of models in which there is a single binary state variable corresponding to the global activation state of the complex, and derives a neural network-like approximation to that model with new, more realistic conditions of validity than in the original paper [2]. This derivation is generalized in the Appendix. Section 4 applies a similar treatment to a two-level hierarchical model (Hierarchical Cooperative Activation) in which the global activation state variable is supplemented by a set of sub-complex or module-level activation variables. An equilibrium model and a set of approximating feed-forward neural network-like models for the regulatory activity of such a complex is derived, along with conditions for their validity. Also a technique for adding one-dimensional chains of logical and energetic dependencies between binding site occupancy states to either the single-level or the two-level equilibrium models (Hierarchical Cooperative Activation with Chains) is derived in Section 4.3, and illustrated with a simple example. The method generalizes to more hierarchical levels, if necessary, and to more logical constraints on binding site occupancy. Section 5 discusses related work and summarizes.

2 Equilibrium statistical mechanics using polynomial partition functions

Suppose there is a molecular complex defined by a set of binary-valued occupancy variables and binary-valued molecular conformation variables, collectively designated $s_A \in \{0, 1\}$, all related through a higher-order Ising model. Then the partition function is

$$Z(\mu | J) = \sum_{\{s | s_A \in \{0,1\} \wedge P(s)\}} \exp(-\beta E), \quad \text{where}$$

$$E = \sum_A \mu_A s_B + \sum_{A < B} J_{AB} s_A s_B + \sum_{A < B < C} J_{ABC} s_A s_B s_C + \dots$$

Here $\beta = 1/(kT)$ where T is the temperature and k is Boltzmann's constant. $P(s)$ is a Boolean-valued function or predicate that restricts the allowed combinations of s_A values to those for which P is true. J_{AB} is an interaction energy between two variables, J_{ABC} between three variables and so on.

Each interaction J pertains to some subset of the variables s_A . This subset can be indicated by a new set of Boolean or $\{0, 1\}$ -valued parameters σ_A . For example, J_{234} corresponds to $\sigma = (0, 1, 1, 1, 0, 0, \dots)$. We can define the ‘‘compression’’ function ρ that maps σ to the ordered set of indices A for which $\sigma_A = 1$, in this case the triple $(2, 3, 4)$, and designate the corresponding interaction strength as

$$J_{\rho(\sigma)} = J_{(A(1) < A(2) < \dots < A(l))} \in \mathbb{R}.$$

If $J_{\rho(\sigma)}$ is zero it can be omitted from the sums in E . We define $Q(\sigma)$ to be a Boolean-valued function that is true just in case σ corresponds to a nonzero entry for $J_{\rho(\sigma)}$; Q serves to specify the bipartite graph of variables and their higher-order energetic interactions. With this notation the partition function becomes

$$Z(\mu) = \sum_{\{s | s_A \in \{0,1\} \wedge P(s)\}} \left[\prod_A \exp(-\beta \mu_A s_A) \right] \left[\prod_{\{\sigma | \sigma_A \in \{0,1\} \wedge Q(\sigma)\}} \exp\left(-\beta J_{\rho(\sigma)} \prod_A (s_A)^{\sigma_A}\right) \right]$$

The partition function can also be rewritten as a polynomial in fugacity-like parameters ζ_A and $\omega_{\rho(\sigma)}$ as follows. Defining

$$\begin{aligned} -\beta \mu_A &\equiv \log \zeta_A \\ -\beta J_{\rho(\sigma)} &= \log \omega_{\rho(\sigma)} \quad , \\ J_{\rho(\sigma)} &= \Delta G_{\rho(\sigma)} \end{aligned}$$

where $\Delta G_{\rho(\sigma)}$ is the change in Gibbs free energy (which may be zero) due specifically to setting the particular combination of state variables $s_A = 1$, for which $\sigma_A = 1$. In traditional Ising models, for example, only pairs of sites A and B which are spatial neighbors have nonzero $\Delta G(A, B)$ values; therefore nonneighboring pairs and all higher combinations have zero ΔG 's: $0 = \Delta G(A, B, C) = \Delta G(A, B, C, D) = \dots$. With this change to parameters ζ and constants ω , we find the general form for the polynomial partition function

$$Z(\zeta) = \sum_{\{s | s_A \in \{0,1\} \wedge P(s)\}} \left[\prod_A \zeta_A^{s_A} \right] \left[\prod_{\{\sigma | \sigma_A \in \{0,1\} \wedge Q(\sigma)\}} (\omega_{\rho(\sigma)})^{\prod_A (s_A)^{\sigma_A}} \right]. \quad (1)$$

For small systems and for those in contact with a chemical reservoir such as a solution, the reexpression of $Z(\zeta)$ as a polynomial can be very useful.

Suppose some of the logical variables s_A describe the binding or nonbinding of a particular substance S_i , that is otherwise in solution, at particular binding sites. Then the equilibrium grand partition function $Z(z, \zeta)$ can be obtained just by replacing the corresponding parameters ζ_A with $z_i \omega_A$ where z_i is a fugacity parameter z_i for substance S_i [1] [2]:

$$\zeta_A \mapsto z_i \omega_A \quad (2)$$

where s_A describes the binding of solute S_i to binding site b . The resulting $Z(z, \zeta)$ remains polynomial in the parameters z , but there may be many fewer z 's than ζ 's. In dilute solution the fugacity z_i is proportional to concentration $[S_i]$. In the special case that every binding site for S_i has the same ΔG_A , we can absorb their common ω_A into the proportionality constant of z_i and set $\omega_A = 1$.

Fortunately, new instrumentation holds the promise that binding energies ΔG_A of DNA-protein interactions can be measured more inexpensively and accurately than in the past (e.g. [3]).

The fugacity parameters z_i are assumed to be constant on the time scale of equilibration of $Z(z, \zeta)$, but may vary slowly on a much longer time scale if the system described by $Z(z, \zeta)$ itself gates the synthesis process for molecules of S_i . Thus, we seek a quasi-equilibrium model of a transcriptional regulatory network built from equilibrium models of transcriptional regulation.

3 Transcriptional regulation

We now apply Equation 1 and Equation 2 to modeling of transcriptional regulation and relate these models to previous work [4] [2] [5] [6] [7] [8]. We will illustrate the case of a two-level hierarchy of transcriptional activation, in which possible binding sites are grouped into nonoverlapping, separately activateable “modules” or stateful subcomplexes which can in turn activate the initiation of transcription, as shown in Figure 1.

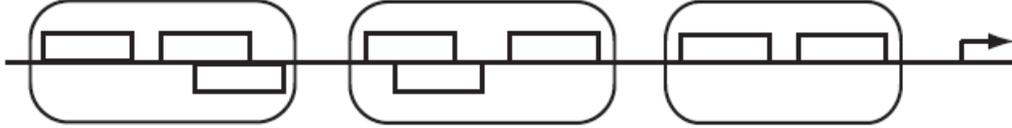


Figure 1. Schematic illustration of transcriptional regulation. Horizontal line: DNA regulatory region. Rectangles: binding sites on DNA, some of which overlap and cannot be simultaneously occupied. Ovals: activateable modules. Bent arrow: site of transcription initiation, influenced by activation state of modules within the ovals.

3.1 Quasi-Equilibrium

Suppose that concentrations of transcription factors (TFs) are $v_i \propto z_i$; that R_i is the maximum synthesis rate of TF i ; that λ_i is the autonomous decay rate for TF i ; that $s(i)$ is a binary random variable determining whether transcription can initiate or not; and that $\langle s(i) \rangle$ is the short-term average thereof at time t . A quasi-equilibrium model for a network of such regulatory interactions may then take a form such as the coupled differential equations

$$\frac{d v_i}{d t} = R_i \text{Activation}_i(v(t)) - \lambda_i v_i$$

where the level of activation of the i 'th transcription complex at time t is

$$\text{Activation}_i = \begin{cases} \langle s(i) \rangle & \text{for deterministic model} \\ \langle s(i) \rangle + \eta_i \sqrt{\text{Var}(s(i))} & \text{for Langevin (stochastic) model} \end{cases}$$

where η_i is a standard (normalized, zero mean) random process with a finite autocorrelation decay time. For the simple Bernoulli distributions on $s(i)$ that we will encounter, $\text{Var}(s(i)) = \langle s(i) \rangle (1 - \langle s(i) \rangle)$. It remains, then, to find $\langle s(i) \rangle$ as a function of the current TF concentrations v_i or fugacities z_i .

To begin, we consider the case of a single module.

3.2 Single-module structure

In the case of a single activateable module there is a set of binding sites with a global activation variable $s \in \{0, 1\}$ (Figure 2). This situation was described by a polynomial partition function and related to artificial neural network (ANN) quasi-equilibrium dynamics in [2]. Here we first rephrase that description in terms of Equation 1 and Equation 2, and then provide a new, much more general derivation of a neural network model as an approximation, as well as sufficient conditions for its validity.

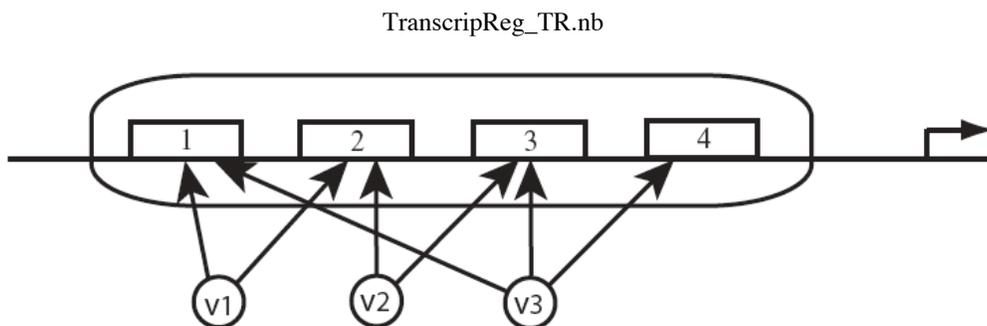


Figure 2 (a) Single-module model with transcription factors, binding sites, and global activation of a multimolecular complex regulating transcription initiation.

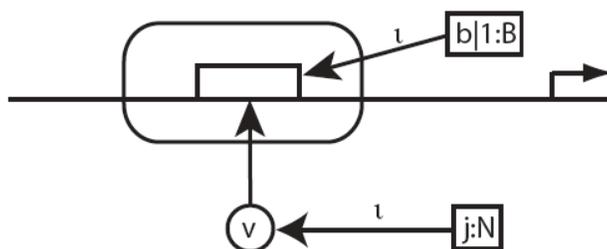


Figure 2(b) Equivalent diagram to 2(a) with replication (many sites and many transcription factors) indicated using index nodes (boxes) and indexing links (labelled by ι).

The binding sites are indexed by $b \in \{1 : B_{\max}\}$. The transcription factors (TF's) available in solution are indexed by $j \in \{1 : N\}$. Each transcription factor may bind to any binding site, but many (b, j) combinations will be energetically unfavorable.

3.3 Statistical mechanics model

To obtain a partition function for the foregoing single-module transcriptional model, we map the binary variable index A of Equation 1 to a pair (b, j) consisting of binding site index $b \in \{1 : B_{\max} \equiv "B"\}$ and a transcription factor j , together with a special value $A = \emptyset$ which corresponds to the single binary-valued global activation variable $s_{\emptyset} \equiv s$. Thus we have indices

$$\begin{aligned} i, j &\in \{1 : N\} \\ b &\in \{1 : B\} \\ A &\in \{\emptyset\} \cup \{(b, i)\} \cup \{(b, i, j)\} \end{aligned}$$

We explicitly enumerate the state variables as follows. There are two global states for transcription complex i , namely $s = 0$ or 1. Each binding site b can be occupied by at most one transcription factor (TF) at a time. However, a pair of mutually exclusively overlapping binding sites (as in Figure 1) can be modeled as a single binding site with multiple alternative occupiers (as in Figure 2), and a mutually exclusively synergistic or antisynergistic (energetically linked) pair of binding sites can be modeled as a single binding site with the ability to bind a dimer of transcription factors (so that the index A expands to include (b, j, k) triples as well, as shown

above). In the latter situation each of the two transcription factors, as well as their dimer, are regarded as mutually exclusive potential occupants of just one binding site. The partition function for site b that represents all these possible states, within a multimolecular transcription complex whose global activation state is s , is the sum over mutually exclusive alternatives with no internal interaction energies $\omega_{\rho(\sigma)}$:

$$\begin{aligned}\Xi_{(ib)}^s &= \sum_{\{s \mid \sum_j s_{(bj)} + \sum_{j,k} s_{(bjk)} \leq 1\}} \left[\prod_{bj} \zeta_{bj}^{s_{bj}} \right] \left[\prod_{bjk} \zeta_{bjk}^{s_{bjk}} \right] \\ &= 1 + \sum_{j=1}^J \zeta_{bj} + \sum_{j,k=1}^J \zeta_{bjk} .\end{aligned}$$

Substituting

$$\begin{aligned}\zeta_{(bj)} &\mapsto z_j \omega_{(ibj)}^{(s)} \\ \zeta_{(bjk)} &\mapsto z_j z_k \omega_{(ibjk)}^{(s)} ,\end{aligned}$$

this simplifies to

$$\Xi_{(ib)}^{(s)} = 1 + \sum_{j=1}^J \omega_{(ibj)}^{(s)} z_j + \sum_{j,k=1}^J \omega_{(ibjk)}^{(s)} z_j z_k . \quad (3)$$

The initial summand “1” represents the empty state of the binding site. If $\omega_{(ibjk)}^{(s)} \equiv 0$ then no pairwise energetic interaction or dimerization is allowed. In this model, exclusive pairwise synergy of bound TF’s is indistinguishable from TF dimerization in its effects on transcription. However, synergy of bound dimers would result in partition function terms of order 3 or 4 in z , not present in Equation 3.

Strictly for notational convenience, we will map the two values of binary variables $s \in \{0, 1\}$ that appear as superscript labels (s) , eg. in $\omega_{(ibj)}^{(s)}$, into the two symbols “+” and “-” for values 1 and 0 respectively, indicating “activation” and “inactivation”, eg. $\omega_{(ibj)}^+$ and $\omega_{(ibj)}^-$.

Conditioned on the value of s , the binding sites are all independent so their partition functions multiply, just as in the MWC and GMWC models of allosteric enzymes [9] [10]. Without loss of generality, the composite partition function is then:

$$\begin{aligned}Z_{\text{GRN } i}(\zeta_i) &= \zeta_i \omega_i \prod_{b=1}^B \Xi_{(ib)}^+ + \prod_{b=1}^B \Xi_{(ib)}^- \\ &= \zeta_i \omega_i \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(ibj)}^+ z_j + \sum_{j,k=1}^J \omega_{(ibjk)}^+ z_j z_k \right) + \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(ibj)}^- z_j + \sum_{j,k=1}^J \omega_{(ibjk)}^- z_j z_k \right)\end{aligned} \quad (4)$$

where

$$\begin{aligned}\omega_{(ibj)}^\pm &= \exp(-\Delta G_{(ibj)}^\pm / k T), \\ \omega_{(ibjk)}^\pm &= \exp(-\Delta G_{(ibjk)}^\pm / k T), \text{ and} \\ \omega_i &= \exp(-\Delta G_i / k T) .\end{aligned} \quad (5)$$

We now compute the activation function for the single-module equilibrium model:

$$\text{Activation}_i = \langle s(i) \rangle = \frac{\partial \log Z_i}{\partial \log \zeta_i} \Big|_{\zeta_i=1} = \frac{\omega_i \prod_{b=1}^B \Xi_{(i b)}^+}{\omega_i \prod_{b=1}^B \Xi_{(i b)}^+ + \prod_{b=1}^B \Xi_{(i b)}^-}$$

Thus we have

Proposition 1. The single-module statistical mechanical model for transcriptional regulation of Equation 4 above has activation function

$$\text{Activation}_i = \langle s \rangle = \frac{\left(\omega_i \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(i b j)}^+ z_j + \sum_{j,k=1}^J \omega_{(i b j k)}^+ z_j z_k \right) \right)}{\left(\omega_i \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(i b j)}^+ z_j + \sum_{j,k=1}^J \omega_{(i b j k)}^+ z_j z_k \right) + \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(i b j)}^- z_j + \sum_{j,k=1}^J \omega_{(i b j k)}^- z_j z_k \right) \right)} \quad (6)$$

which is a ratio of polynomials in the fugacity parameters, where the ω parameters are related to binding energies by Equation 5. In the absence of synergistic interactions and of dimerization of singly-active transcription factors, $\omega_{(i b j k)}^s = 0$ and this expression simplifies to

$$\text{Activation}_i = \frac{\omega_i \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(i b j)}^+ z_j \right)}{\omega_i \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(i b j)}^+ z_j \right) + \prod_{b=1}^B \left(1 + \sum_{j=1}^J \omega_{(i b j)}^- z_j \right)} \quad (7)$$

Proposition 1 reexpresses the quasi-equilibrium model in Section 4 of [2].

3.4 Neural network-like GRN model

Continuing the foregoing calculation from Equation 6,

$$\begin{aligned} \text{Activation}_i &= \frac{\omega_i \prod_{b=1}^B (\Xi_{(i b)}^+ / \Xi_{(i b)}^-)}{1 + \omega_i \prod_{b=1}^B (\Xi_{(i b)}^+ / \Xi_{(i b)}^-)} \\ &= \text{Hill} \left(\omega_i \prod_{b=1}^B (\Xi_{(i b)}^+ / \Xi_{(i b)}^-), n = 1 \right) = g \left[\log \left(\omega_i \prod_{b=1}^B (\Xi_{(i b)}^+ / \Xi_{(i b)}^-) \right) \right], \end{aligned} \quad (8)$$

where

$$\begin{aligned} \text{Hill}(x, n) &\equiv \frac{x^n}{1 + x^n}, \text{ and} \\ g(y) &\equiv 1 / (1 + \exp(-y)). \end{aligned}$$

Observation 1. If repression is disallowed, so all $\omega_{(i b j)}^- = 0$, and if all binding sites are identical, so that $\omega_{(i b j)}^+ = \omega_{(i j)}^+$ and $\omega_{(i b j k)}^+ = \omega_{(i j k)}^+$, then the activation function becomes

TranscripReg_TR.nb

$$\begin{aligned} \text{Activation}_i &= \frac{\omega_i \left(1 + \sum_{j=1}^J T_{ij} v_j + \sum_{j,k=1}^J T_{ijk} v_j v_k\right)^B}{1 + \omega_i \left(1 + \sum_{j=1}^J T_{ij} v_j + \sum_{j,k=1}^J T_{ijk} v_j v_k\right)^B} \\ &= \hat{g} \left(h_i + \sum_{j=1}^J T_{ij} v_j + \sum_{j,k=1}^J T_{ijk} v_j v_k \right) \end{aligned} \quad (9)$$

where v_j is the concentration of the j 'th transcription factor (which is proportional to fugacity z_j in dilute solution), and:

$$\begin{aligned} h_i &= \omega_i \\ T_{ij} v_j &= \omega_{(ij)}^+ (\omega_i)^{-1/B} z_j \\ T_{ijk} v_j v_k &= \omega_{(ijk)}^+ (\omega_i)^{-1/B} z_j z_k \\ \hat{g}(x) &= x^B / (1 + x^B) = \text{Hill}(x, B) \end{aligned}$$

This argument was made in [2]. Equation 9 has the form of a one-layer, feed-forward neural network which is a higher-order or sum-product neural network if some $T_{ijk} \neq 0$. However, it is derived under highly restrictive and biologically unrealistic conditions: there can be no repression (except for constitutive inhibition obtained by taking $\omega_i < 1$), and all $B > 1$ binding sites must be functionally identical so that \hat{g} is a sigmoidal function.

Continuing the calculation from Equation 8,

$$\text{Activation}_i = g \left[\log \omega_i + \sum_{b=1}^B \log \Xi_{(ib)}^+ - \sum_{b=1}^B \log \Xi_{(ib)}^- \right]. \quad (10)$$

Observation 2. Much more generally, if the average occupancies of all of the (heterogeneous, activating and/or repressing) binding sites are low, but there are many binding sites, then we may expand each logarithm in Equation 10 (whose radius of convergence is 1 and whose argument is independent of B) but not the sigmoidal function g (whose argument grows as $O(B)$):

$$\log \Xi_{(ib)}^{(s)} \cong \sum_{j=1}^J \omega_{(ibj)}^{(s)} z_j + \sum_{j,k=1}^J \omega_{(ibjk)}^{(s)} z_j z_k \ll 1 .$$

so

$$\text{Activation}_i \cong g \left[\sum_{j=1}^J (T_{(ij)}^+ - T_{(ij)}^-) z_j + \sum_{j,k=1}^J (T_{(ijk)}^+ - T_{(ijk)}^-) z_j z_k + h_i \right]$$

where

$$\begin{aligned} T_{(ij)}^\pm &= \sum_{b=1}^B \omega_{(ibj)}^\pm = \sum_{b=1}^B \exp(-\Delta G_{(ibj)}^\pm / kT), \\ T_{(ijk)}^\pm &= \sum_{b=1}^B \omega_{(ibjk)}^\pm = \sum_{b=1}^B \exp(-\Delta G_{(ibjk)}^\pm / kT), \text{ and} \\ h_i &= \log \omega_i = -\Delta G_i / kT \end{aligned}$$

This is again an Artificial Neural Network (ANN) model of gene regulation networks as proposed in [2], Section 2, where it was introduced as a ‘‘phenomenological modeling framework’’ for transcriptional regulation. It

was first applied to create a predictive experimental model of transcriptional regulation in a particular biological system in [11]. By the foregoing argument we now arrive at Proposition 2.

Proposition 2. The activation function of Equation 6 may be approximated by the feed-forward neural network equation

$$\text{Activation}_i = \langle s(i) \rangle \cong g \left(\sum_{j=1}^J T_{ij} v_j + \sum_{j,k=1}^J T_{ijk} v_j v_k + h_i \right), \quad (11)$$

where

$$\begin{aligned} g(x) &= 1 / (1 + \exp(-x)), \\ T_{(ij)} &= \sum_{b=1}^B \omega_{(ibj)}^+ - \sum_{b=1}^B \omega_{(ibj)}^- = \sum_{b=1}^B \exp(-\Delta G_{(ibj)}^+ / kT) - \sum_{b=1}^B \exp(-\Delta G_{(ibj)}^- / kT), \\ T_{(ijk)} &= \sum_{b=1}^B \omega_{(ibjk)}^+ - \sum_{b=1}^B \omega_{(ibjk)}^- = \sum_{b=1}^B \exp(-\Delta G_{(ibjk)}^+ / kT) - \sum_{b=1}^B \exp(-\Delta G_{(ibjk)}^- / kT), \text{ and} \\ h_i &= \log \omega_i = -\Delta G_i / kT \end{aligned} \quad (12)$$

by expanding the logarithm in Equation 10, which is valid under the following conditions:

- (a) the probability of occupancy for each binding site is ϵ , with $\epsilon \ll 1$, and
- (b) there are large number of binding sites, $B \gg 1$, and $B\epsilon = O(1)$.

Under Proposition 2, all connection matrix entries T and h in Equation 11 have micro-level interpretations in terms of binding energies in Equation 12. In the absence of pairwise exclusive synergies and dimerization, $T_{ijk} = 0$. The reduced model of Proposition 2 may be called ‘‘cooperative’’ because of conditions (a) and (b): changes in state require joint input from a number of binding sites. If the required number of sites is sufficiently large the model may even be called ‘‘collective’’ since in that case a global state variable (the argument to g) summarizes the effect of many local ones.

This reduced model assumes there is only very restricted (pairwise) binding site overlap. But even if general overlap is allowed, since occupancy probabilities are low, such overlaps will rarely actually occur and the first-order approximation is still valid. Only higher-order correction terms in ϵ are affected by the overlaps. Higher order terms in the expansion of the logarithm may of course be kept, at the cost of increasing the order of z in the argument to g .

Under restricted conditions it is also possible to derive a neural network form for the activation function if some sites have a dominant occupant which is present with probability $1 - \epsilon$ (as outlined in the Appendix). A further generalization of Proposition 2 would be to allow a few sites with nonextreme probabilities (so the full statistical mechanical model of Proposition 1 is needed for those sites) against a background of many sites with extreme (~ 0 or 1) probabilities.

The modeling framework of Equation 11 (with $T_{ijk} = 0$) has been fit to data and used to make testable predictions regarding *Drosophila* anterior-posterior axis in [11], [12], and [13]. In the absence of the grounding of its T and h parameters in measurable ΔG values provided by Equation 12, and in the absence of the achievable sufficient conditions for its validity provided by Proposition 2, this framework was a phenomenological rather than a mechanistic one, whose parameters could not be measured directly. As a phenomenological model it satisfies the criteria suggested in [2]: (1) it saturates both above (at a maximal initiation rate) and below (at zero initiation rate) due to the sigmoidal function g ; (2) each regulatory interaction is characterized by one real number

(the relevant T_{ij} or T_{ijk}); and (3) the effects of these numbers are approximately additive within the limits imposed by criterion (1). These criteria are not satisfied, for example, by linear or Boolean models, but they may be satisfied by piecewise linear approximations of g . They may be reasonable properties for a course-scale, phenomenological model that simplifies and approximates some finer-scale mechanistic model of transcriptional regulation.

With Proposition 2, however, the framework specified by Equation 11 can act not only as a phenomenological model but also as a mechanistic model, with primary mechanisms being transcription factor binding/unbinding in quasi-equilibrium, competitive inhibition by overlapped binding sites, and regulation of transcription initiation by average activation of a single molecular complex or module. Models expressed in this framework can be evaluated experimentally and quantitatively for particular systems alongside other, more elaborate mechanistic models such as those discussed below. If the approximation validity conditions of Proposition 2 are not satisfied, a quasi-equilibrium model of the same mechanisms is still described by Proposition 1.

4 HCA for transcriptional regulation

The single-module model can be extended straightforwardly to multiple, separately activateable modules or subcomplexes, each of which has a binary activation state variable $s_m \in \{0, 1\}$ which can interact with the global activation variable $s \in \{0, 1\}$ for transcription complex i . Each module contains a set of binding sites whose overlaps or synergistic interactions are limited to exclusive pairs, as for the previous single-module model. Now the numbering scheme for A is:

$$A \in \{\emptyset\} \cup \{m\} \cup \{(m, b, i)\} \cup \{(m, b, i, j)\}$$

The result will be a version of the Hierarchical Cooperative Activation model [1] for transcriptional regulation. Here we rephrase and extend the statistical mechanical expression of this model, and derive novel two-layer neural network approximations to it in a manner parallel to the foregoing single-module neural network approximation.

4.1 Hierarchical Cooperative Activation (HCA)

The HCA model is illustrated in Figure 3.

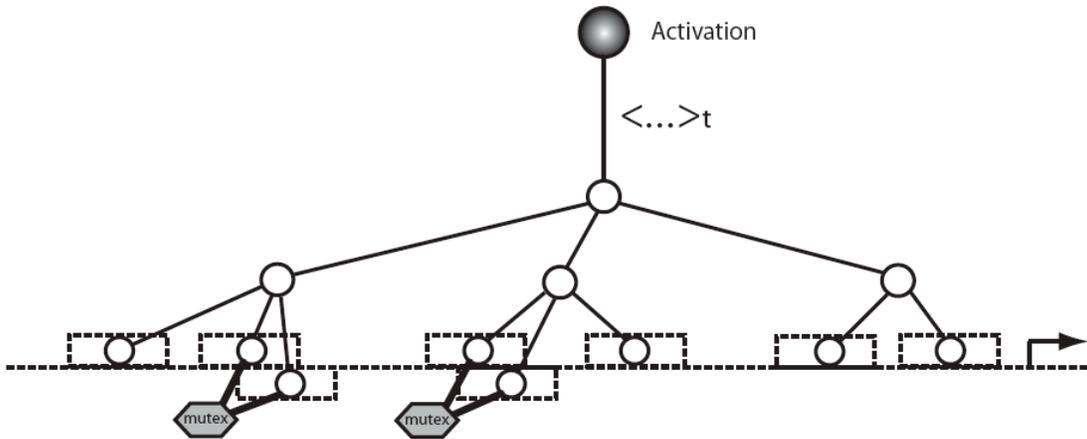


Figure 3. The transcriptional regulation model of Figure 1, redrawn to show the binary state variables (open circles), their “soft” energetic interactions (thin solid lines), and their “hard” logical interactions. Logical interactions are indicated by thick solid lines connecting to predicate-bearing hexagons. In this case the predicate P is “mutex” (mutual exclusion) or, what is equivalent for two Boolean arguments, “nand”. An additional labeled line connects the global activation state s , through ensemble- or time-averaging, to a real-valued average activation (shaded circle) which is proportional to the rate of transcription initiation. This diagram is similar to those used in machine learning applications. Biological objects (DNA and binding sites) are shown with dotted lines. Trees of energetic interactions for Ising models (spin systems, as in Equation 1) were introduced into statistical mechanics by Dyson [14] [15].

In this two-level hierarchical model, the partition function of Equation 2 can be built up by iterating the construction of the previous section, beginning with the single-site partition functions Ξ and allowing for both global activation state s and module-level activation states s_m where m indexes a set of separately activateable modules within the multimolecular transcription complex, as illustrated in Figures 1 and 3. Thus

$$Z_i(\zeta_i) = \zeta_i \omega_i \prod_{m=1}^M \left[\omega_{(m)}^+ \prod_{b=1}^{B(m)} \Xi_{(m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(m b)}^- \right] + \prod_{m=1}^M \left[\omega_{(m)}^- \prod_{b=1}^{B(m)} \Xi_{(m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(m b)}^- \right] \quad (13)$$

where

$$\Xi_{(m b)}^{(s_m)} = 1 + \sum_{j=1}^J \omega_{(m b j)}^{(s_m)} z_j + \sum_{j,k=1}^J \omega_{(m b j k)}^{(s_m)} z_j z_k .$$

Then we may calculate that the average activation is again a ratio of polynomials

$$\begin{aligned} \text{Activation}_i &= \frac{\partial \log Z_i}{\partial \log \zeta_i} \Big|_{\zeta_i=1} \\ &= \frac{\omega_i \prod_{m=1}^M \left[\omega_{(m)}^+ \prod_{b=1}^{B(m)} \Xi_{(m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(m b)}^- \right]}{\omega_i \prod_{m=1}^M \left[\omega_{(m)}^+ \prod_{b=1}^{B(m)} \Xi_{(m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(m b)}^- \right] + \prod_{m=1}^M \left[\omega_{(m)}^- \prod_{b=1}^{B(m)} \Xi_{(m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(m b)}^- \right]} \\ &= \text{Hill} \left[\omega_i \frac{\prod_{m=1}^M \left[\omega_{(m)}^+ \prod_{b=1}^{B(m)} \Xi_{(m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(m b)}^- \right]}{\prod_{m=1}^M \left[\omega_{(m)}^- \prod_{b=1}^{B(m)} \Xi_{(m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(m b)}^- \right]}, 1 \right] \end{aligned}$$

Thus

$$\begin{aligned} \text{Activation}_i &= \text{Hill} \left[\omega_i \prod_{m=1}^M \left\{ \frac{\left[\omega_{(m)}^+ \prod_{b=1}^{B(m)} \frac{\Xi_{(m b)}^+}{\Xi_{(m b)}} + 1 \right]}{\left[\omega_{(m)}^- \prod_{b=1}^{B(m)} \frac{\Xi_{(m b)}^+}{\Xi_{(m b)}} + 1 \right]} \right\}, 1 \right], \\ &= \text{Hill} \left[\omega_i \prod_{m=1}^M \left\{ 1 + \left(\frac{\omega_{(m)}^+}{\omega_{(m)}^-} - 1 \right) \text{Hill} \left[\omega_{(m)}^- \prod_{b=1}^{B(m)} \frac{\Xi_{(m b)}^+}{\Xi_{(m b)}}, 1 \right] \right\}, 1 \right]. \end{aligned} \quad (14)$$

Thus we have

Proposition 3. For a fully statistical HCA model,

TranscripReg_TR.nb

$$\text{Activation}_i = \langle s \rangle = \left(\omega_i \prod_{m=1}^M \left[\omega_{(i m)}^+ \prod_{b=1}^{B(m)} \Xi_{(i m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(i m b)}^- \right] \right) / \left(\omega_i \prod_{m=1}^M \left[\omega_{(i m)}^+ \prod_{b=1}^{B(m)} \Xi_{(i m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(i m b)}^- \right] + \prod_{m=1}^M \left[\omega_{(i m)}^- \prod_{b=1}^{B(m)} \Xi_{(i m b)}^+ + \prod_{b=1}^{B(m)} \Xi_{(i m b)}^- \right] \right).$$

Equivalently

$$\text{Activation}_i = \text{Hill} \left[\omega_i \prod_{m=1}^M \{1 + J_{i m} Y_{i m}\}, 1 \right],$$

where

$$Y_{i m} = \text{Hill} \left[\omega_{(i m)}^- \prod_{b=1}^{B(m)} \frac{\Xi_{(i m b)}^+}{\Xi_{(i m b)}^-}, 1 \right],$$

$$J_{i m} = \left(\frac{\omega_{(i m)}^+}{\omega_{(i m)}^-} - 1 \right), \text{ and}$$

$$\Xi_{(i m b)}^{(s_m)} = 1 + \sum_{j=1}^J \omega_{(i m b j)}^{(s_m)} z_j + \sum_{j,k=1}^J \omega_{(i m b j k)}^{(s_m)} z_j z_k .$$

Note that there are now two Hill functions composed with one another (one is nested as an argument inside the other), corresponding to the assumed two levels of activation in the transcription complex. Each Hill function also has as an argument a product, over modules or binding sites. Essentially the same model was proposed in [5] (with one redundant parameter compared to the present formulation) under a slightly weaker interpretation: $z_j z_k$ terms denoted bound dimers but not pairwise synergistic interactions. As in [5], a calculable observable is the average occupancy of each binding site. This can be expressed as a function of the elementary occupancies (omitting dimerization for simplicity)

$$f_{(i m b j)}^{s_m} = \frac{\partial \log \Xi_{(i m b)}^{(s_m)}}{\partial \log \zeta_{m b j}} \Bigg|_{\zeta_{m b j} \rightarrow \omega_{(i m b j)}^{s_m} z_j} = \frac{\omega_{(i m b j)}^{(s_m)} z_j}{1 + \sum_{l=1}^J \omega_{(i m b l)}^{(s_m)} z_l}$$

as follows:

$$f_{(i m b j)} = \frac{\partial \log Z_i(\zeta_i)}{\partial \log \zeta_{m b j}} \Bigg|_{\zeta_{m b j} \rightarrow \omega_{(i m b j)}^s z_j} \\ = \sum_s \frac{\partial \log Z_i(\zeta_i)}{\partial \log \Xi_{(i m)}^{(s)}} \sum_{s_m} \frac{\partial \log \Xi_{(i m)}^{(s)}}{\partial \log \Xi_{(i m b)}^{(s_m)}} \frac{\partial \log \Xi_{(i m b)}^{(s_m)}}{\partial \log \zeta_{m b j}} \Bigg|_{\zeta_{m b j} \rightarrow \omega_{(i m b j)}^s z_j}$$

TranscripReg_TR.nb

$$\begin{aligned}
&= \sum_s \frac{(\zeta_i \omega_i)^s \Xi_{(i m)}^{(s)}}{\zeta_i \omega_i \Xi_{(i m)}^+ + \Xi_{(i m)}^-} \sum_{s_m} \frac{(\omega_{(i m)}^{(s)})^{s_m} \Xi_{(i m b)}^{(s_m)}}{\omega_{(i m)}^{(s)} \Xi_{(i m b)}^+ + \Xi_{(i m b)}^-} \frac{\omega_{(i m b j)}^{(s_m)} z_j}{1 + \sum_{l=1}^J \omega_{(i m b l)}^{(s_m)} z_l} \\
&= \sum_s \Pr(s) \sum_{s_m=0,1} \Pr(s_m | s) \frac{\omega_{(i m b j)}^{(s_m)} z_j}{1 + \sum_{l=1}^J \omega_{(i m b l)}^{(s_m)} z_l} \\
&= \sum_{s_m=0,1} \Pr(s_m) \frac{\omega_{(i m b j)}^{(s_m)} z_j}{1 + \sum_{l=1}^J \omega_{(i m b l)}^{(s_m)} z_l}
\end{aligned}$$

whence

$$f_{(i m b j)} = \langle s_m \rangle f_{(i m b j)}^+ + (1 - \langle s_m \rangle) f_{(i m b j)}^- . \quad (15)$$

However, the ω 's themselves are also in principle observable as exponentials of measurable ΔG 's.

4.2 ANN-like approximation

Calculating from Equation 14 as in the single-module case,

$$\begin{aligned}
\text{Activation}_i &= g \left[\log \left(\omega_i \prod_{m=1}^M \left\{ 1 + \left(\frac{\omega_{(m)}^+}{\omega_{(m)}^-} - 1 \right) g \left[\log \left(\omega_{(m)}^- \prod_{b=1}^{B(m)} \frac{\Xi_{(m b)}^+}{\Xi_{(m b)}^-} \right) \right] \right\} \right) \right] \\
&= g \left[\log \omega_i + \sum_{m=1}^M \log \left\{ 1 + \left(\frac{\omega_{(m)}^+}{\omega_{(m)}^-} - 1 \right) g \left[\log \omega_{(m)}^- + \sum_{b=1}^{B(m)} \log \left(\frac{\Xi_{(m b)}^+}{\Xi_{(m b)}^-} \right) \right] \right\} \right] \\
&= g \left[h_i + \sum_{m=1}^M \log \left\{ 1 + (\exp(h_{(m)}^+) - h_{(m)}^-) - 1 \right\} g \left[h_{(m)}^- + \sum_{b=1}^{B(m)} (\log \Xi_{(m b)}^+ - \log \Xi_{(m b)}^-) \right] \right]
\end{aligned}$$

defining $h_* = \log \omega_*$. Therefore

$$\text{Activation}_i = g \left[h_i + \sum_{m=1}^M \tilde{g} \left[h_{(m)}^- + \sum_{b=1}^{B(m)} (\log \Xi_{(m b)}^+ - \log \Xi_{(m b)}^-); h_{(m)}^+ - h_{(m)}^- \right] \right]$$

where we define

$$\tilde{g}(x; h) \equiv \log[1 + (\exp(h) - 1) g(x)] \quad (16)$$

Using convexity and $0 \leq g(x) \leq 1$,

$$\begin{aligned}
\tilde{g}(x; h) &= \log[\exp(h) g(x) + (1 - g(x))] \\
&\geq h g(x) + (1 - g(x)) \log(1) = h g(x) .
\end{aligned}$$

If we minimize the approximation error

$$\mathcal{E}_2(h, c) = \frac{1}{2} \int_0^1 (h \lambda + c \lambda(\lambda - 1) - \log(e^h \lambda + (1 - \lambda)))^2 d\lambda \quad (17)$$

symbolically with respect to c for any given h , we find

$$c(h) = \frac{5}{6} (5 - 27 e^h + 27 e^{2h} - 5 e^{3h}) + \frac{3h(1 - 3 e^h - 3 e^{2h} + e^{3h})}{(1 - e^h)^3} . \quad (18)$$

Figure 4 quantifies the accuracy of this approximation as a function of h .

Other error integrands such as $(\text{fractional error} - 1)^2$ could also be used and numerically optimized, giving different values for $c(h)$. Using any such metric including Equation 17 we may approximate

$$\tilde{g}(x; h) \cong 1 + [h - 1 - c(h)] g(x) + c(h) g(x)^2$$

which is exact at the extreme values $g(x) = 0$ or $=1$. This formula may be seen as a small feed-forward neural network implementing $\tilde{g}(x; h)$ in terms of a previous layer that calculated $g(x)$.

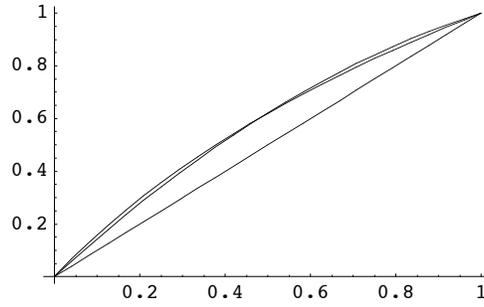


Figure 4(a). Approximation by linear and quadratic polynomials, for $h = 1$. Horizontal axis: $\lambda = g(x)$. Vertical axis: $\log(e^h \lambda + (1 - \lambda))$ and its approximants $h \lambda$ and $h \lambda + c \lambda(1 - \lambda)$, where $c(h)$ is given by Equation 18 .

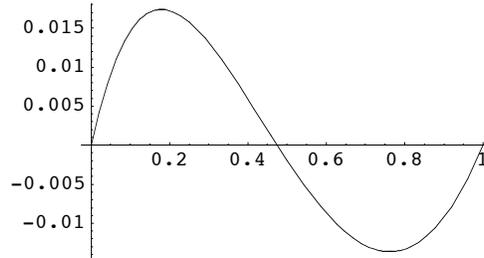


Figure 4(b). Approximation error for quadratic form, for $h = 1$. Horizontal axis: $\lambda = g(x)$. Vertical axis: $\log(e^h \lambda + (1 - \lambda)) - [h \lambda + c \lambda(1 - \lambda)]$, where $c(h)$ is given by Equation 18 . Further reductions in error ($\times 5$ for $h = 1$) are possible using higher order polynomials in λ , such as $h \lambda + c_2 \lambda(\lambda - 1) + c_3 \lambda(\lambda - 1/2)(1 - \lambda)$.

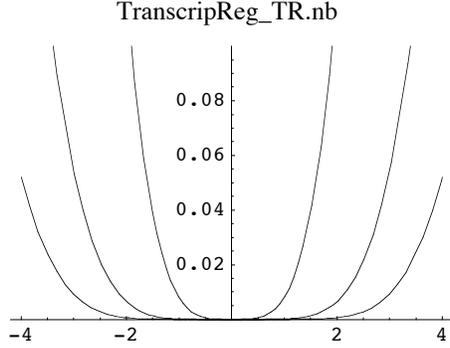


Figure 4(c). Integrated squared error $\mathcal{E}_1(h)$, $\mathcal{E}_2(h)$, and $\mathcal{E}_3(h)$ with optimal coefficients c , all computed symbolically and plotted numerically as a function of h , for polynomial approximations $h\lambda$, $h\lambda + c\lambda(1-\lambda)$, and $h\lambda + c_2\lambda(\lambda-1) + c_3\lambda(\lambda-1/2)(1-\lambda)$ of degree 1, 2, and 3 respectively. Errors decrease with increasing degree for each value of h . Empirically, it appears that $d\mathcal{E}_k(h, c(h))/dh \geq 0$ for $h \geq 0$ and $k \in \{1, 2, 3\}$, in which case decreasing $|h|$ always improves the accuracy of approximation and every plotted data point establishes an approximation validity condition for all smaller $|h|$. Numerical artifacts are encountered at $h \approx 0$ due to high-order degeneracy at zero, but analytically $\mathcal{E}_k(0) = 0$ and $\mathcal{E}_k(\epsilon) \sim d_k \epsilon^{2(k+1)}$ for $k \in \{1, 2, 3\}$, with coefficients $d_k > 0$. Larger values of h require more terms in such a polynomial expansion for a given level of accuracy.

If B is so large that we are well inside the radius of convergence for the logarithm around 1, which is 1, but outside the domain of validity for a low-order series approximation of $g(By + h_i)$, which is $O(1/B)$, then we can expand the logarithm but not the sigmoid function g and recover a feed-forward neural network equation:

$$\begin{aligned}
\log \Xi_{(mb)}^\pm &\cong \sum_{j=1}^J \omega_{(mbj)}^\pm z_j + \omega_{(mbjk)}^\pm z_j z_k \\
T_{(mj)}^\pm &= \sum_{b=1}^{B(m)} \omega_{(mbj)}^\pm \quad \text{and} \quad T_{(mjk)}^\pm = \sum_{b=1}^{B(m)} \omega_{(mbjk)}^\pm \\
\sum_{b=1}^{B(m)} \log \Xi_{(mb)}^\pm &\cong \sum_{j=1}^J T_{(mj)}^\pm z_j + \sum_{j,k=1}^J T_{(mjk)}^\pm z_j z_k \\
\sum_{b=1}^{B(m)} \log \Xi_{(mb)}^+ - \sum_{b=1}^{B(m)} \log \Xi_{(mb)}^- &\cong \sum_{j=1}^J (T_{(mj)}^+ - T_{(mj)}^-) z_j + \sum_{j,k=1}^J (T_{(mjk)}^+ - T_{(mjk)}^-) z_j z_k \\
&= \sum_{j=1}^J T_{(mj)} z_j + \sum_{j,k=1}^J T_{(mjk)} z_j z_k
\end{aligned}$$

This gives a series of approximations summarized in Proposition 4.

Proposition 4. For the hierarchical (HCA) model of Equation 13 the activation function is

$$\text{Activation}_i = \langle s \rangle = g \left(h_i + \sum_{m=1}^M \tilde{g} \left(h_{(m)}^- + \sum_{b=1}^{B(m)} (\log \Xi_{(mb)}^+ - \log \Xi_{(mb)}^-); h_{(m)}^+ - h_{(m)}^- \right) \right)$$

where $\tilde{g}(x; h) \equiv \log[1 + (\exp(h) - 1)g(x)]$ and $\Xi_{(m b)}^s$ is given by Equation 13. It has the following approximation:

$$\text{Activation}_i \cong g \left(h_i + \sum_{m=1}^M \tilde{g} \left(h_{(m)}^- + \sum_{j=1}^J T_{(i m j)} z_j + \sum_{j,k=1}^J T_{(i m j k)} z_j z_k; h_{(m)}^+ - h_{(m)}^- \right) \right) \quad (19)$$

as well as a second less accurate approximation (a two-layer sum-product neural network with particular weight formula)

$$\begin{aligned} \text{Activation}_i &\cong g \left(\tilde{h}_i + \sum_{m=1}^M T_{(i m)} v_{(m)} + \sum_{m=1}^M \tilde{T}_{(i m)} v_{(m)}^2 \right), \text{ where} \\ v_{(m)} &= g \left(h_{(m)}^- + \sum_{j=1}^J T_{(i m j)} z_j + \sum_{j,k=1}^J T_{(i m j k)} z_j z_k \right) \\ T_{(i m)} &= h_{(m)}^+ - h_{(m)}^- - 1 - \tilde{T}_{(i m)} \\ \tilde{T}_{(i m)} &= c(h_{(m)}^+ - h_{(m)}^-) \end{aligned} \quad (20)$$

where $c(h)$ is given by Equation 18. There is a third still less accurate approximation which also takes the form of a classic two-layer feed-forward sum-product neural network, obtained by setting $\tilde{T}_{(i m)} = 0$:

$$\begin{aligned} \text{Activation}_i &\cong g \left(\tilde{h}_i + \sum_{m=1}^M T_{(i m)} v_{(m)} \right), \quad v_{(m)} = g \left(h_{(m)}^- + \sum_{j=1}^J T_{(i m j)} z_j + \sum_{j,k=1}^J T_{(i m j k)} z_j z_k \right), \\ T_{(i m)} &= h_{(m)}^+ - h_{(m)}^- - 1. \end{aligned} \quad (21)$$

In these expressions the constants are defined by

$$\begin{aligned} T_{(i m j)} &= T_{(i m j)}^+ - T_{(i m j)}^- \quad \text{and} \quad T_{(i m j k)} = T_{(i m j k)}^+ - T_{(i m j k)}^- \\ T_{(i m j)}^\pm &= \sum_{b=1}^{B(m)} \omega_{(m b j)}^\pm = \sum_{b=1}^{B(m)} \exp(-\Delta G_{(m b j)}^\pm / k T), \\ T_{(i m j k)}^\pm &= \sum_{b=1}^{B(m)} \omega_{(m b j k)}^\pm = \sum_{b=1}^{B(m)} \exp(-\Delta G_{(m b j k)}^\pm / k T), \quad \text{and} \\ h_{(m)}^\pm &= \log \omega_{(m)}^\pm = -\Delta G_{(m)}^\pm / k T \\ h_i &= \log \omega_i = -\Delta G_i / k T \\ \tilde{h}_i &= \log \omega_i + M = -\Delta G_i / k T + M \end{aligned}$$

and $g(x) = 1 / (1 + \exp(-x))$. The conditions of validity for the first approximation are:

- (a) probability of occupancy for each binding site is ϵ , with $\epsilon \ll 1$.
- (b) large number of binding sites within each module, $B(m) \gg 1$ and $B(m)\epsilon = O(1)$

The conditions of validity of the second and third approximations depend also on Equation 17 and Equation 18.

It may also be possible to compute higher-order corrections to the large- B limit while retaining a major advantage of the neural net formulation, that many possible variations in structure are reduced to a fixed space of polynomially many parameters.

4.3 Interaction/overlap chains

The restrictions of pairwise mutually exclusive overlap and synergy allowed for simple analytic expressions of the polynomial grand partition function. At some cost in simplicity, more general one-dimensional chains of overlap and energetic interaction (synergistic or anti-synergistic) along the DNA such as those of Figure 5 can be expressed using such partition functions as well. Thus we will formulate an extension of the HCA model, that may be called Hierarchical Cooperative Activation with Chains (HCAC), to include such chains of 1D dependencies.

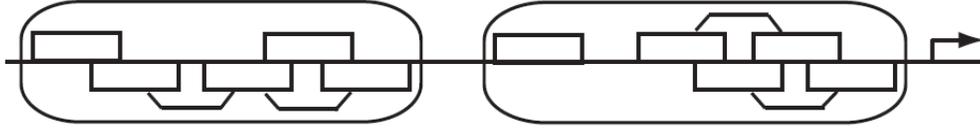


Figure 5. Chains of overlapping and energetically interacting sites, solvable with Equation 22. Symbols as in Figure 1, with energetic interactions added. Where adjacent sites overlap, only second nearest neighbors are allowed to interact energetically since nearest neighbors are never simultaneously present. An example of a model with a chain of energetic interactions occurs in [4].

For a single module, the partition function for a chain may be derived as a product of 3×3 matrices by considering pairs of overlapping sites ($b = 1, b = 2$), ($b = 3, b = 4$), and so on, each of which has three states: both empty, first site occupied, and second site occupied. In case there are some adjacent nonoverlapping sites, we “reduce” to the fully overlapping case: Simply add an extra intermediate site that overlaps with both adjacent sites and then set its $\zeta_b = 0$ so that it is never occupied.

The novel part of this model compared to HCA is the one-dimensional interactions through site overlap and synergy: second nearest neighbors (odd or even) interact energetically with factor ω . (See Figure 5). Therefore each successive pair of sites has three possible states. The model can be solved using 3×3 transfer matrices on site pairs:

$$Z(\zeta) = (1 \ 1 \ 1) \cdot \left\{ \prod_{p=\lceil B/2-1 \rceil}^1 \begin{pmatrix} 1 & 1 & 1 \\ \zeta_{2p+1} & \zeta_{2p+1} \omega_{2p-1, 2p+1} & 0 \\ \zeta_{2p+2} & \zeta_{2p+2} & \zeta_{2p+2} \omega_{2p, 2p+2} \end{pmatrix} \right\} \cdot \begin{pmatrix} 1 \\ \zeta_1 \\ \zeta_2 \end{pmatrix} \quad (22)$$

with Z , ζ , and ω all implicitly superscripted by the enclosing activation state (s) or (s_m) as the case may be, and also subscripted appropriately as in HCA. Any site can be omitted (removing its overlap constraints and interaction energies) by setting its ζ_b to zero and $\omega_{b*} = \omega_{*b} = 1$. Also as in HCA, the substitution Equation 2 of must be performed after $Z(\zeta)$ is calculated to get the grand partition function $Z(\mathbf{z}, \zeta)$.

As an example, the leftmost module in Figure 5 has overlapping pairs of sites 1 and 2 (numbering from the left); an interpolated nonbiological site 3 (not shown) that overlaps 2 and 4 and for which we take $\zeta_3 = 0$; overlapping sites 4 and 5; and overlapping sites 5 and 6. With this numbering, there are energetic interactions (synergistic or antisynergistic) between sites 2 and 4 and sites 4 and 6. The leftmost module’s partition function may therefore be calculated automatically as

TranscripReg_TR.nb

$$\begin{aligned}
Z(\zeta) &= (1 \ 1 \ 1) \cdot \begin{pmatrix} 1 & 1 & 1 \\ \zeta_5 & \zeta_5 & 0 \\ \zeta_6 & \zeta_6 & \zeta_6 \omega_{46} \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ \zeta_4 & \zeta_4 & \zeta_4 \omega_{24} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \zeta_1 \\ \zeta_2 \end{pmatrix} \\
&= 1 + \zeta_1 + \zeta_2 + \zeta_4 + \zeta_1 \zeta_4 + \zeta_2 \zeta_4 \omega_{24} + \zeta_5 + \zeta_1 \zeta_5 + \zeta_2 \zeta_5 \\
&\quad + \zeta_6 + \zeta_1 \zeta_6 + \zeta_2 \zeta_6 + \zeta_4 \zeta_6 \omega_{46} + \zeta_1 \zeta_4 \zeta_6 \omega_{46} + \zeta_2 \zeta_4 \zeta_6 \omega_{24} \omega_{46}
\end{aligned}$$

which can be hand-simplified taking advantage of the break at $b = 3$, to

$$Z(\zeta) = (1 + \zeta_1 + \zeta_2)(1 + \zeta_4 + \zeta_5 + \zeta_6 + \zeta_4 \zeta_6 \omega_{46}) + \zeta_2 \zeta_4 (\omega_{24} - 1)(1 + \zeta_6 \omega_{46}).$$

4.4 Further generalizations of HCAC

A solvable extension of HCAC is illustrated by the following diagram:

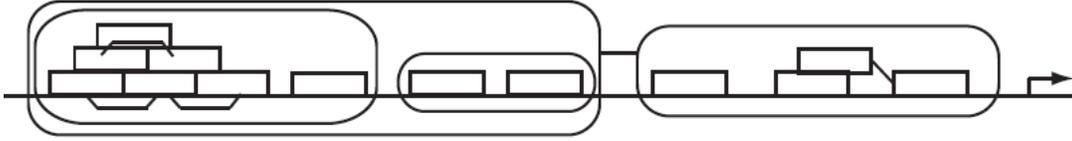


Figure 6. A solvable extension of HCAC. Symbols as in Figure 5.

The hierarchy can be made deeper, in a tree topology (representable with nested enclosures). A key point regarding the HCAC model is that the maximum height of the overlap stack, and therefore the dimension of the transfer matrix in the solution for the partition function, is bounded by the length of its constituent enhancer binding sites - from 6 to 18 or so. Typical overlaps will be much less. Coupled with a tree topology for the module structure, the result is still solvable by the methods used above. Multiple competing sites for transcriptional initiation can be accommodated in Equation 4 or Equation 13 by giving the top-level activation variable s_i more than just two discrete, mutually exclusive states, each with its own maximal transcription rate R_{i,s_i} , including $R_{i,s_i} = 0$ for the globally “inactive” state $s_i = 0$.

5 Discussion

5.1 Related models

Other models of transcriptional regulation have also relied on equilibrium statistical mechanics to compute partition functions and activation as a function of regulatory inputs, at least in particular cases. The models presented in this paper only use the predicate P to prevent simultaneous occupancy of overlapping sites. More extensive use of hard constraints P , and correspondingly less extensive use of the soft interaction parameters ω , appear in the abstract transcriptional logic examples of [6]. Likewise in [8], a particular GRN in *Drosophila* responsible for dorso-ventral patterning is studied using partition functions that encode extensive hard logical relationships between site occupancy and module activation and between module activation and activation of the transcription complex. These models can also be specified by diagrams in the style of Figure 3.

In [7] a transcriptional modeling framework is presented which has two levels of hierarchy as does HCA, and adds distance-dependent repressive “quenching” interactions at the module and whole complex levels. The framework is applied to the *Drosophila* anterior-posterior axis. Unlike the present models it is apparently not a quasi-equilibrium model because it is not derived from a global partition function; rather it is essentially feed-forward using a series of steps with individual formulae that may be derivable from local partition functions. Site occupancies are independent of activation state, unlike those of Equation 15. Whether neural net like feed-forward approximations to quasi-equilibrium models as developed in the present paper can or should approximate the models of [7] may be an interesting question.

5.2 Quasi-equilibrium models of transcriptional regulation

We have translated structural diagrams of hypothesised energetic interactions and simple Boolean constraints on the state variables of multimolecular transcription complexes into partition functions for equilibrium statistical mechanics in the fast component of a quasi-equilibrium approach to modeling transcriptional regulation. For single-level and two-level modular complexes with cooperative activation, we derived approximating feed-forward neural networks together with sufficient and plausible conditions for their validity. The sufficient conditions correspond biologically to having large numbers of binding sites for transcription factors, each of which have a small effect. These derivations included a single-layer neural network-like model class that has been successfully applied to model GRNs in several developmental models but has heretofore been a phenomenological rather than a mechanistic framework; now it can play both roles. They also included two-layer neural network approximations for two-level hierarchical cooperative activation (HCA) models of transcription complexes. We also showed how to add one-dimensional chains of energetic and mutual exclusion constraints among binding sites within the one- and two-level models, thus expanding the set of hypotheses that may be expressed and explored within a quasi-equilibrium framework.

There is no guarantee that quasi-equilibrium models are the correct way to model transcription complexes, although they are very convenient theoretically as shown here. It is even possible that suitable nonequilibrium statistical mechanical methods will be developed to the point that they are straightforward to apply in translating from model-structure diagrams to mathematical models, as are the equilibrium and quasi-equilibrium methods explored in the present work.

Acknowledgements. The author wishes to acknowledge Vitali Likhoshvai, Tarek Najdi, Bruce Shapiro, John Reinitz, Barbara Wold, and Chin-Ran Yang for valuable discussions. This work was supported in part by the Biomedical Information Science and Technology Initiative (BISTI) grant (number 4 R33 GM069013) from the National Institute of General Medical Sciences, USA and by the U.S. National Science Foundation, Frontiers in Integrative Biological Research Award EF-0330786 .

A Appendix: Highly occupied site partition function

Suppose one of the nonempty states of the site partition function $\Xi_{(i b)}^s$ dominates, say that corresponding to term $\omega_{(i b \hat{j})}^s z_{\hat{j}}$:

$$\Xi_{(i b)}^s = \omega_{(i b \hat{j})}^s z_{\hat{j}} + 1 + \sum_{j \neq \hat{j}(b)} \omega_{(i b j)}^s z_j + \sum_{j, k=1}^J \omega_{(i b j k)}^s z_j z_k$$

Then

$$\begin{aligned} \log \Xi_{(i b)}^s &= \log \left(1 + \omega_{(i b \hat{j}(b))}^s z_{\hat{j}(b)} \right) + \log \left(1 + \sum_{j \neq \hat{j}(b)} \frac{\omega_{(i b j)}^s z_j}{1 + \omega_{(i b \hat{j}(b))}^s z_{\hat{j}(b)}} + \sum_{j, k=1}^J \frac{\omega_{(i b j k)}^s z_j z_k}{1 + \omega_{(i b \hat{j}(b))}^s z_{\hat{j}(b)}} \right) \\ &\simeq \log \left(1 + \omega_{(i b \hat{j}(b))}^s z_{\hat{j}(b)} \right) + \sum_{j \neq \hat{j}(b)} \frac{\omega_{(i b j)}^s z_j}{1 + \omega_{(i b \hat{j}(b))}^s z_{\hat{j}(b)}} + \sum_{j, k=1}^J \frac{\omega_{(i b j k)}^s z_j z_k}{1 + \omega_{(i b \hat{j}(b))}^s z_{\hat{j}(b)}} \end{aligned}$$

To simplify the calculation assume that every site has such a dominant occupant. (To recover the mixed case in which some sites b are dominated by the empty states, set $\hat{j}(b) = 0$ and $z_{\hat{j}(b)} = 0$ for those sites). Then

$$\begin{aligned} \text{Activation}_i &= g \left[\log \omega_i + \sum_{b=1}^B \log \Xi_{(i b)}^+ - \sum_{b=1}^B \log \Xi_{(i b)}^- \right] \\ &= g \left[\log \omega_i + \sum_{b=1}^B \log \frac{1 + \omega_{(i b \hat{j}(b))}^+ z_{\hat{j}(b)}}{1 + \omega_{(i b \hat{j}(b))}^- z_{\hat{j}(b)}} \right. \\ &\quad \left. + \sum_{b, j \neq \hat{j}(b)} \frac{\omega_{(i b j)}^+ z_j}{1 + \omega_{(i b \hat{j}(b))}^+ z_{\hat{j}(b)}} - \sum_{b, j \neq \hat{j}(b)} \frac{\omega_{(i b j)}^- z_j}{1 + \omega_{(i b \hat{j}(b))}^- z_{\hat{j}(b)}} + \sum_{b, j, k=1}^J \frac{\omega_{(i b j k)}^+ z_j z_k}{1 + \omega_{(i b \hat{j}(b))}^+ z_{\hat{j}(b)}} - \sum_{b, j, k=1}^J \frac{\omega_{(i b j k)}^- z_j z_k}{1 + \omega_{(i b \hat{j}(b))}^- z_{\hat{j}(b)}} \right] \end{aligned}$$

In the special case that

$$\omega_{(i b \hat{j}(b))}^+ = \omega_{(i b \hat{j}(b))}^- = \omega_{(i b)}$$

the logarithm terms in the argument to g cancel, and we have a feed-forward neural network with “shunting” style inhibition from each nonzero $z_{\hat{j}(b)}$:

TranscripReg_TR.nb

$$\text{Activation}_i = g \left[\log \omega_i + \sum_{j \neq \hat{j}(b)} \frac{\omega_{(i b j)}^+ z_j}{1 + \omega_{(i b j)} z_{\hat{j}(b)}} - \sum_{j \neq \hat{j}(b)} \frac{\omega_{(i b j)}^- z_j}{1 + \omega_{(i b j)} z_{\hat{j}(b)}} \right. \\ \left. + \sum_{j, k=1}^J \frac{\omega_{(i b j k)}^+ z_j z_k}{1 + \omega_{(i b j)} z_{\hat{j}(b)}} - \sum_{j, k=1}^J \frac{\omega_{(i b j k)}^- z_j z_k}{1 + \omega_{(i b j)} z_{\hat{j}(b)}} \right]$$

If in addition $z_{\hat{j}(b)}$ is constant (for example because it is not involved in the GRN dynamics), then

$$\text{Activation}_i = g \left[\log \omega_i + \sum_{j \neq \hat{j}(b)} \tilde{\omega}_{(i b j)}^+ z_j - \sum_{j \neq \hat{j}(b)} \tilde{\omega}_{(i b j)}^- z_j + \sum_{j, k=1}^J \tilde{\omega}_{(i b j k)}^+ z_j z_k - \sum_{j, k=1}^J \tilde{\omega}_{(i b j k)}^- z_j z_k \right] \\ \tilde{\omega}_{(i b j)}^s = \omega_{(i b j)}^s / (1 + \omega_{(i b j)} z_{\hat{j}(b)})$$

and the recovered neural network is of the same form used in Proposition 2, without shunting inhibition but with an altered form for the connection matrices T .

References

- [1] Hill, T. L. (1985). *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*. Berlin: Springer-Verlag.
- [2] Mjolsness, E., Sharp, D. H., & Reinitz, J. (1991). *A Connectionist Model of Development*. *Journal of Theoretical Biology*, **152**(4), 429–454.
- [3] Maerkl, S. J., & Quake, S. R. (2006). *A Systems Approach to Measuring the Binding Energy Landscape of Transcription Factors*. Submitted manuscript, Bioengineering Department, Stanford University.
- [4] 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.
- [5] Mjolsness, E. (2001). *Gene Regulation Networks for Modeling Drosophila Development*. In J. M. Bower & H. Bolouri (Ed.), *Computational Methods in Molecular Biology*. MIT Press.
- [6] Buchler, N. E., Gerland, U., & Hwa, T. (2003). *On schemes of combinatorial transcription logic*. *Proc. Natl. Acad. Sci. USA*, **100**(9), 5136–5141.
- [7] Reinitz, J., Hou, S., & Sharp, D. H. (2003). *Transcriptional Control in Drosophila*. *ComplexUs*, **1**, 54–64.
- [8] Zinzen, R. P., Senger, K., Levine, M., & Papatsenko, D. (2006, July 11). *Current Biology*, **16**, 1–8.
- [9] Monod, J., Wyman, J., & Changeaux, J. P. (1965). *On the nature of allosteric transitions: A Plausible Model*. *Journal of Molecular Biology*, **12**, 88–118.
- [10] Najdi, T. S., et al. (2006). *Application of a Generalized MWC Model for the Mathematical Simulation of Metabolic Pathways Regulated by Allosteric Enzymes*. *Journal of Bioinformatics and Computational Biology*, **4**, 335–355.
- [11] Reinitz, J., Sharp, D. H., & Mjolsness, E. (1995). *Model for Cooperative Control of Positional Information in Drosophila by bcd and Maternal hb*. *Journal of Experimental Zoology*, **271**, 47–56.
- [12] Reinitz, J., & Sharp, D. H. (1995). *Mechanism of eve stripe formation*. *Mech. Dev.*, **49**, 133–158.
- [13] Jaeger, J., et al. (2004). *Dynamic control of positional information in the early Drosophila blastoderm*. *Nature*, **430**, 368–371.
- [14] Dyson, F. (1969). *Existence of a Phase-Transition in a One-Dimensional Ising Ferromagnet*. *Comm. Math. Phys.*, **12**, 91–107.
- [15] Dyson, F. (1971). *An Ising Ferromagnet with Discontinuous Long-Range Order*. *Comm. Math Phys*, **21**, 269–283.

Eric D. Mjolsness received his A.B. in Physics and Mathematics from Washington University in St. Louis and his Ph.D. in Physics and Computer Science from California Institute of Technology.

He was an Assistant Professor at Yale University, New Haven from 1985 to 1990, a faculty member of the Yale Program in Neuroscience from 1990 to 1994, a Research Scientist in the department of Computer Science and Engineering at University of California, San Diego from 1995 to 1998, a Computer Scientist — Principal in the Jet Propulsion Laboratory at California Institute of Technology from 1998 to 2002 and a Faculty Associate in Biology at California Institute of Technology from 2000 to 2002. He is currently an Associate Professor in the Departments of Computer Science and Mathematics at the University of California, Irvine, a Visiting Associate at the California Institute of Technology, and a group leader in Systems Biology in the Institute for Genomics and Bioinformatics at the University of California, Irvine.