

Supporting Text

Model of Transcription Regulation. We model transcription regulation in bacteria through the thermodynamics of pairwise protein–DNA and protein–protein interactions (1). These interactions can be quantified by several parameters that are tuneable by the selection and placement of various protein-binding DNA sequences.

Protein–DNA interactions. The probability of transcription factor (TF)–DNA binding is of the Arrhenius form (2),

$$p_i = \frac{q_i}{1 + q_i}, \quad [\mathbf{M1}]$$

where $q_i = [\text{TF}_i]/K_i$ is the binding affinity of a site i , $[\text{TF}_i]$ is the cellular concentration of the activated TF targeted by this site, and K_i is the effective dissociation constant (relative to the genomic background) representing the concentration for half-maximal occupation of the site. It is known for a number of exemplary bacterial TFs (3) and expected on theoretical ground for a large class of bacterial TFs (4) that K_i can be readily tuned across a wide range of cellular concentrations from a low value of ≈ 1 molecule per cell (1 nM) to a high value of $\approx 10,000$ molecules per cell (10 μM) simply by adjusting the number of bases that match the strongest binding sequence for the TF. Thus, K_i is a parameter that is individually tuneable for each binding site. We describe RNA polymerase (RNAP)–promoter binding (in the absence of any TFs) by the same Arrhenius form of Eq. **M1**. Denoting the promoter as site “p,” we quantify the promoter affinity by q_p , which is also tuneable.

Protein–protein interaction. Interaction between a pair of proteins (TFs or RNAP) bound to two sites i and j is quantified by a cooperativity factor $w_{i,j}$, which is tuneable to a degree by the relative placement of these sites in the regulatory region. For example, the choice $w_{i,j} = 0$ can be implemented by the mutual exclusion between two proteins when their binding sites are made to overlap. Interaction can also be “turned off” (described by $w_{i,j} = 1$) by increasing the separation between two sites (but not placing them too far) so that physical contact cannot be made when both proteins are bound to their sites.

Cooperative interaction with $w_{i,j} > 1$ may be obtained if two proteins can contact each other while bound to their sites. Due to its structural complexity, RNAP can contact TFs over a range of TF-binding site positions extending, e.g., several tens of bases upstream of the promoter (5). This interaction is weak with typical binding free energy in the range of a few kilocalories per mole, corresponding to a cooperativity factor of $w = \approx 10\text{--}100$. Two TFs can interact cooperatively if they are bound to adjacent sites or if they can contact each other through DNA looping. In the former case, known binding free energies are again of the order ≈ 2 kcal/mol (1, 6) with $w \approx 20$. The interaction leading to DNA looping is necessarily much stronger. It can result from two distinct proteins that bind strongly to form a heterodimer (7) as indicated in Fig. 5a or from a single protein with two fused DNA-binding domains (8). Mathematically, we model this effect by a pair

of TFs, each with its own DNA-binding interaction as described by Eq. **M1**, along with a special cooperative interaction with cooperativity factor $\Omega \gg w$. Taken together, the protein–protein interaction described above can be summarized by

$$w_{i,j} = \{0, 1, w, \Omega\}, \quad [\text{M2}]$$

which is individually selectable for each pair of binding sites i,j . For simplicity, we report here only results with $w = 20$ and $\Omega = 100$. We have checked that the implementation of the logic functions in Fig. 1a are independent of the value of w in the range of ≈ 10 –100, and $\Omega \geq 100$.

Gene transcription. In bacteria, the rate of gene transcription is controlled in many instances by the amount of time the RNAP spends bound to the promoter. Following Shea and Ackers (1), we quantify the degree of gene transcription by the equilibrium probability P of RNAP-promoter binding due to interaction with bound TFs. For a single promoter, this quantity can be expressed as

$$P = \frac{Z_{ON}}{Z_{OFF} + Z_{ON}}, \quad [\text{M3}]$$

where Z_{ON} and Z_{OFF} are the partition sum of the Boltzmann weights W over all states of TF binding for the promoter bound and not bound, respectively, by the RNAP. In the simplest case involving a single TF-binding site (say site “1”), we have $Z_{OFF} = 1 + q_1$ and $Z_{ON} = q_p (1 + q_1 w_{p,1})$. With multiple TF-binding sites (labeled as sites $i = 1, \dots, L$), the Boltzmann weight for each configuration of site occupation is still a simple product of the q_i and $w_{i,j}$ values, under the assumption that the TF–TF interaction is glue-like (6, 9, 10). It is convenient to introduce a binary variable $\sigma_i \in \{0,1\}$ to denote the occupation of each site i . We have

$$W[\mathbf{s}_1, \dots, \mathbf{s}_L] = \prod_{i=1}^L q_i^{s_i} \prod_{i < j} w_{i,j}^{s_i s_j} \quad [\text{M4}]$$

as the weight for each configuration $\{\sigma_1, \dots, \sigma_L\}$, with Z_{OFF} obtained as the sum

$$Z_{OFF} = \sum_{\mathbf{s}_1=\{0,1\}} \dots \sum_{\mathbf{s}_L=\{0,1\}} W[\mathbf{s}_1, \dots, \mathbf{s}_L]. \quad [\text{M5}]$$

The expression for Z_{ON} can be generally written as

$$Z_{ON} = \sum_{\mathbf{s}_1=\{0,1\}} \dots \sum_{\mathbf{s}_L=\{0,1\}} Q[\mathbf{s}_1, \dots, \mathbf{s}_L] \cdot W[\mathbf{s}_1, \dots, \mathbf{s}_L], \quad [\text{M6}]$$

where $Q[\sigma_1, \dots, \sigma_L]$ is the Boltzmann weight due to the interaction of the RNAP with the bound TFs. This interaction is promoter-dependent and can be rather complicated for multiple TFs. For example, for the σ^{70} -promoters of *Escherichia coli*, the RNAP–TF interaction can be “synergistic” (5, 11, 12), because two subunits of the RNAP holoenzyme can simultaneously contact two different TFs bound to upstream locations, whereas for the σ^{54} -promoters, the interaction is “independent,” because activation of the RNAP involves binding with only one TF at a time (13). We have investigated both types of interactions and obtained similar conclusions. The response characteristics used in the text are produced by the (simpler but more restrictive) independent interaction model, given by the weight

$$Q = q_p \prod_{i=1}^L [1 - \mathbf{s}_i \mathbf{d}(\mathbf{w}_{0,i}, 0)] \cdot \left[1 + \mathbf{w} \sum_{j=1}^L \mathbf{s}_j \mathbf{d}(\mathbf{w}_{p,j}, \mathbf{w}) \right]. \quad [\text{M7}]$$

Here, the first bracket insures that the promoter cannot be occupied (i.e., $Q = 0$) if any one of the repressor sites (those with $\mathbf{w}_{p,i} = 0$) is occupied. The second bracket describes the additional weight gained by the interaction of the RNAP with all the bound TFs that it can interact cooperatively with (those with $\mathbf{w}_{p,j} = \mathbf{w}$) but only one at a time.

For those cases where a single gene is controlled by two promoters, we quantify the degree of gene transcription by the equilibrium probability P that the RNAP binds to at least one of the promoters. Assuming that there is no interaction between the two promoters (i.e., the TFs do not simultaneously interact with both polymerases in the unlikely case that both promoters are occupied), we can write the binding probability as

$$P = \frac{Z_{ON}^{(1)} \cdot Z_{OFF}^{(2)} + Z_{OFF}^{(1)} \cdot Z_{ON}^{(2)} + Z_{ON}^{(1)} \cdot Z_{ON}^{(2)}}{Z_{OFF}^{(1)} \cdot Z_{OFF}^{(2)} + Z_{ON}^{(1)} \cdot Z_{OFF}^{(2)} + Z_{OFF}^{(1)} \cdot Z_{ON}^{(2)} + Z_{ON}^{(1)} \cdot Z_{ON}^{(2)}}, \quad [\text{M8}]$$

where $Z_{ON}^{(i)}$ and $Z_{OFF}^{(i)}$ are the partition sum of the Boltzmann weights W over all states of TF binding when promoter p_i is bound and not bound by the RNAP, respectively.

Implementation of Logic Gates. Eqs. M3–M8 completely specify our model of transcription regulation. To use them to compute the response of a given gene, one needs to supply the cis-regulatory construct specifying all the pairwise protein interactions $\mathbf{w}_{i,j}$ as well as the affinities q_i of all the DNA sites in the regulatory region. The protein interactions can be represented graphically as shown in Figs. 2, 3, and 5b, with $\mathbf{w}_{i,j} = 0$ if two sites overlap, $\mathbf{w}_{i,j} = \mathbf{w} = 20$ if two sites are linked by a dashed line, $\mathbf{w}_{i,j} = \Omega = 100$ if linked by a solid line, and $\mathbf{w}_{i,j} = 1$ if otherwise. The analytical expression for each of the

response characteristics $P([A],[B])$ plotted in Figs. 2, 3, and 5b is then obtained by using Eq. M3 or M8, as appropriate, with the corresponding expressions for Z_{ON} and Z_{OFF} , and the values of the binding affinities as given below:

AND gate (Fig. 2a):

$$Z_{OFF} = 1 + q_A + q_B + \mathbf{w} q_A q_B$$

$$Z_{ON} = q_p (1 + \mathbf{w} q_A + \mathbf{w} q_B + 2\mathbf{w}^2 q_A q_B)$$

$$K_A = K_B = 3,500; q_p = 1/35$$

OR gate (Fig. 2b):

$$Z_{OFF} = 1 + q_A + q_B + q_A q_B$$

$$Z_{ON} = q_p (1 + \mathbf{w} q_A + \mathbf{w} q_B + 2\mathbf{w} q_A q_B)$$

$$K_A = K_B = 100; q_p = 1/20$$

NAND gate (Fig. 2c):

$$Z_{OFF} = 1 + q_A + q_B + \mathbf{w} q_A q_B$$

$$Z_{ON} = q_p$$

$$K_A = K_B = 100; q_p = 100$$

XOR gate, single promoter (Fig 3b):

$$Z_{OFF} = (1 + q_{A2} + q_{B2} + \mathbf{w} q_{A2} q_{B2}) \cdot (1 + q_{A1} + q_{B1} + q_{A1} q_{B1})$$

$$Z_{ON} = q_p (1 + \mathbf{w} q_{A1} + \mathbf{w} q_{B1} + 2\mathbf{w} q_{A1} q_{B1})$$

$$K_{A1} = K_{B1} = 200; K_{A2} = K_{B2} = 900; q_p = 1/10$$

XOR gate, double promoter (Fig 3c):

$$Z_{OFF}^{(1)} = (1 + q_{A1}) \cdot (1 + q_{B1})$$

$$Z_{OFF}^{(2)} = (1 + q_{A2}) \cdot (1 + q_{B2})$$

$$Z_{ON}^{(1)} = q_{p1} (1 + \mathbf{w} q_{A1})$$

$$Z_{ON}^{(2)} = q_{p2}(1 + \mathbf{w} q_{B2})$$

$$K_{A1} = K_{B2} = 500; K_{A2} = K_{B1} = 100; q_{p1} = q_{p2} = 1/20$$

EQ gate, long-distance repression (Fig. 5b):

$$Z_{ON} = q_p (Q_{R1}^+ + Q_{R1}^-) \cdot (Q_{R2}^+ + Q_{R2}^-)$$

$$Z_{OFF} = (Q_{R1}^+ + Q_{R1}^-) \cdot (Q_{R2}^+ + Q_{R2}^-) + q_S \left[Q_{R1}^- Q_{R2}^- + \Omega \cdot (Q_{R1}^+ Q_{R2}^- + Q_{R1}^- Q_{R2}^+) + 2\Omega Q_{R1}^+ Q_{R2}^+ \right]$$

$$Q_{R2}^- = (1 + q_{A2}) \cdot (1 + q_{B2}); Q_{R2}^+ = q_{R2}(1 + \mathbf{w} q_{B2})$$

$$Q_{R1}^- = (1 + q_{A1}) \cdot (1 + q_{B1}); Q_{R1}^+ = q_{R1}(1 + \mathbf{w} q_{A1})$$

$$K_{A1} = K_{B2} = 200; K_{A2} = K_{B1} = 50; q_{R1} = q_{R2} = 1/50; q_S = 10; q_p = 40$$

Note that the binding affinities q are directly selected for the promoters (site p) and the binding sites of the auxiliary TFs R and S, because the cellular concentrations of the RNAP and the auxiliary regulators are assumed to be only weakly variable. The remaining q values are defined through the variable controlling TF concentrations $[A]$ and $[B]$, i.e.,

$$q_A = \frac{[A]}{K_A}; q_B = \frac{[B]}{K_B}; q_{A1} = \frac{[A]}{K_{A1}}; q_{B1} = \frac{[B]}{K_{B1}}; q_{A2} = \frac{[A]}{K_{A2}}; q_{B2} = \frac{[B]}{K_{B2}}$$

where $K_A, K_B, K_{A1}, K_{B1}, K_{A2}, K_{B2}$ are the strengths of the various sites labeled in Figs. 2, 3, and 5b. For all the logic gates implemented above, we arbitrarily considered promoter occupancy of larger than 40% as sufficient for a gene to be “ON.”

Mapping to Neural Networks. The model of transcription regulation described by Eqs. **M3–M8** belongs to the class of “recurrent” neural networks (14). To highlight the connection, it is convenient to recast the partition function Z_{OFF} given in Eqs. **M4** and **M5** in terms of an “Ising Hamiltonian” \mathcal{H} , such that $Z_{OFF} = \sum_{\{\mathbf{s}\}} e^{-\mathcal{H}/RT}$. In this

framework where neural network models are often described (14), we have

$$\mathcal{H} = \sum_{j=1}^K h_j \mathbf{s}_j + \sum_{i \neq j} J_{i,j} \mathbf{s}_i \mathbf{s}_j$$

where each binding site j is identified as a “neuron,” \mathbf{s}_j

indicates the state of the j th neuron, $h_j = -RT \ln(q_j)$ is the “input” to that neuron, and

$J_{i,j} = -RT \ln(\mathbf{w}_{i,j})$ is the synaptic connection between the neurons i and j . Because h_j

biases the neuron to the “on” state ($\mathbf{s}_j = 1$) only if $q_j > 1$ or $[TF_j] > K_j$, we can identify

the binding strength K_j as the “firing threshold” of the j th neuron. Because this network

contains “hidden units,” which are “neurons” not linked directly to the controlling inputs

[A] and [B] (e.g., the binding sites of the auxiliary proteins R and S as shown in Fig. 7), the system is known as the “Boltzmann machine” (14).

The usual operation of neural networks (including the Boltzmann machine) amounts to finding the values of the connection matrix elements $J_{i,j}$ to implement the desired tasks, e.g., classification. The operation of the transcription control system we describe here is somewhat different: The $J_{i,j}$ values are constrained to take on one of the four discrete values corresponding to the form of protein–protein interaction described by Eq. M2. Instead, it is the firing thresholds that can be tuned continuously.

Promiscuity of Protein Interactions. At high protein concentrations, a nonspecific, glue-like interaction between TFs can lead to many spurious interactions that jeopardize the intended cis-regulatory control. Here, we provide a simple estimate of the range of TF concentrations over which this problem can be safely ignored. We will only consider spurious interactions that occur while the TFs are bound to DNA, because the TF molecules spend most of the time bound to the genome (either specifically or nonspecifically) due to electrostatic attraction (15, 16). For the same reason, we neglect possible spurious interactions between TFs and other, non-DNA-binding proteins. Let N denote the total number of activated TF molecules in a bacterial cell at a given instant in time. The average separation distance between two such TFs along the DNA is $\ell = \Gamma/N$ where a typical genome size Γ is 5×10^6 bp for bacteria. Two activated TF molecules will associate with each other if the interaction energy E_{int} overcomes the entropy cost of association. The latter is $\approx RT \ln(\ell/a)$, where the microscopic length a represents the range of interactions between activated TFs; we take 10 bp as a conservative upper bound for a . For a weak interaction energy of $E_{\text{int}} \approx 2$ kcal/mol, we can then safely ignore spurious interactions as long as N is less than or equal to 10^4 . Thus, at a typical cellular concentration of ≈ 100 molecules per cell, one species of activated TF can interact weakly with ≈ 100 other species before spurious interactions can affect cis-regulatory control at all.

1. Shea, M. A. & Ackers, G. K. (1985) *J. Mol. Biol.* **181**, 211–230.
2. Von Hippel, P. H. & Berg, O. G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1608–1612.
3. Stormo, G. D. & Fields, D. S. (1998) *Trends Biochem. Sci.* **23**, 109–113.
4. Gerland, U., Moroz, J. D. & Hwa, T. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 12015–12020.
5. Rhodius, V. A. & Busby, S. J. (1998) *Curr. Opin. Microbiol.* **1**, 152–159.

6. Ptashne, M. & Gann, A. (2002) *Genes and Signals* (Cold Spring Harbor Lab. Press, Plainview, NY).
7. Kornacker, M. G., Remsburg, B. & Menzel, R. (1998) *Mol. Microbiol.* **30**, 615–624.
8. Langdon, R. C., Burr, T., Pagan-Westphal, S. & Hochschild, A. (2001) *Mol. Microbiol.* **41**, 885–896.
9. Ptashne, M. & Gann, A. (1997) *Nature* **386**, 569–577.
10. Ptashne, M. & Gann, A. (1998) *Curr. Biol.* **8**, R812–R822.
11. Scott, S., Busby, S. & Beacham, I. (1995) *Mol. Microbiol.* **18**, 521–531.
12. Joung, J. K., Koepp, D. M. & Hochschild, A. (1994) *Science* **265**, 1863–1866.
13. Molina-Lopez, J. A. & Santero, E. (1999) *Mol. Gen. Genet.* **262**, 291–301.
14. Hertz, J., Krogh, A. & Palmer, R. G. (1991) *Introduction to the Theory of Neural Computation* (Addison–Wesley, Redwood City, CA).
15. Berg, O. G., Winter, R. B. & von Hippel, P. H. (1981) *Biochemistry* **20**, 6926–6948.
16. Winter, R. B., Berg, O. G. & von Hippel, P. H. *Biochemistry* **20**, 6948–6960.