

Transcription regulation

Model implementation. There are two aspects to modeling transcriptional regulation: (1) modeling the activation or inhibition of a transcription factor (e.g., by a ligand), and (2) given an active transcription factor, modeling its effect on RNA polymerase recruitment to a promoter site. The enhanced coverage of the regulatory network - 438 regulatory interactions described by 22 transcription factors that regulate 355 genes - is a significant difference from the *M. genitalium* model. To incorporate this network, regulation is represented by three different classes of transcription regulators: zero-component systems, one-component systems and two-component systems.

Modeling transcription factor activation. We consider three classes of transcription factors based on their mechanism of activation:

1. **Zero-component systems:** transcription factors that are considered to be active whenever they are expressed. Examples include the Fis and Hns proteins. These two proteins, for instance, are important in maintaining higher-order DNA structure and likely have complex feedback loops modulating their activity. Because this complexity is not yet fully understood, we make the simplifying assumption that these proteins are always active unless they are knocked out. Zero-component systems are modeled in the `TfBinding` process in the model, which handles transcription factor binding to promoters.
2. **One-component systems:** transcription factors that are directly activated or inhibited by a small molecule ligand. Examples of this class include the repressor TrpR which binds tryptophan, and the inducer AraC which binds arabinose. One-component systems are modeled in the `TfBinding` and `Equilibrium` processes in

the model, which handle transcription factor binding to promoters and transcription factor binding to ligands, respectively.

3. **Two-component systems:** transcription factors that are paired with a separate sensing protein that responds to an environmental stimulus (these are simple analogs to the vast, complicated signaling networks that exist in eukaryotic cells). The sensing protein phosphorylates the cognate transcription factor in a condition-dependent fashion. Examples include ArcA which is phosphorylated by its cognate ArcB in anaerobic conditions, and NarL which responds to the presence of nitrate when phosphorylated by its cognate sensor NarX. Two-component systems are modeled in the `TfBinding`, `Equilibrium`, and `TwoComponentSystems` processes in the model, which handle transcription factor binding to promoters, transcription factor binding to ligands, and phosphotransfer reactions of signaling pathways, respectively.

Zero-component systems. We assume all transcription factors of this class will bind to available promoter sites.

One-component systems. For a transcription factor with concentration T whose activity is directly modulated by a ligand with concentration L that binds with stoichiometry n , we assume that the two species achieve equilibrium on a short time scale and that the affinity of the two molecules can be described by a dissociation constant K_d :



where T^* represents the concentration of the ligand-bound transcription factor.

With the dissociation constant K_d defined as:

$$K_d = \frac{L^n \cdot T}{T^*} \quad (2)$$

we have:

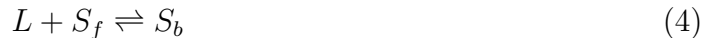
$$\frac{T^*}{T_T} = \frac{L^n}{L^n + K_d} \quad (3)$$

where T_T is the total concentration of the transcription factor, both ligand-bound and unbound. As we can see, the fraction of bound transcription factor is a function of ligand concentration and the dissociation constant. Importantly, if the ligand concentration is (approximately) constant over time, the fraction of bound transcription factor is (approximately) constant over time.

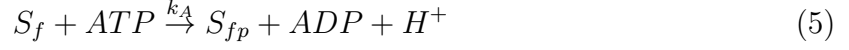
To computationally simulate this model we start with total counts of free transcription factor and ligand, completely dissociated from one another. We then form one molecule of the ligand-TF complex at a time and evaluate how close the ratio of $L^n \cdot T/T^*$ is to the actual K_d . We select the values of L , T and T^* that minimize the absolute difference between K_d and $L^n \cdot T/T^*$ (see Algorithm 1).

Two-component systems. For a transcription factor with concentration T ; a cognate sensing protein with concentration S ; a ligand with concentration L ; subscripts f denoting a free (unbound) form of a molecule, b denoting a ligand-bound form of a molecule, and p denoting a phosphorylated form of a molecule; and ATP , ADP , H^+ , and H_2O denoting concentrations of these molecules, we propose a system with the following:

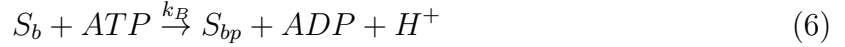
Free (unbound) cognate sensing protein at equilibrium with ligand-bound cognate sensing protein, described by dissociation constant K_d :



The autophosphorylation of a free (unbound) cognate sensing protein at a rate k_A :



The autophosphorylation of a ligand-bound cognate sensing protein at a rate k_B :



The phosphorylation of a transcription factor by its free, phosphorylated cognate sensing protein at a rate k_C :



The phosphorylation of a transcription factor by its bound, phosphorylated cognate sensing protein at a rate k_D :



The auto-phosphatase activity of a transcription factor at a rate k_E :



Ligand binding is simulated in a fashion identical to the one-component systems. By assuming mass-action kinetics, we can represent the rest of this system mathematically using ordinary differential equations:

$$\frac{dS_f}{dt} = -k_A \cdot S_f \cdot ATP + k_C \cdot S_{fp} \cdot T \quad (10)$$

$$\frac{dS_b}{dt} = -k_B \cdot S_b \cdot ATP + k_D \cdot S_{bp} \cdot T \quad (11)$$

$$\frac{dT}{dt} = -k_C \cdot S_{fp} \cdot T - k_D \cdot S_{bp} \cdot T + k_E \cdot T_p \cdot H_2O \quad (12)$$

$$\frac{dS_{fp}}{dt} = -\frac{dS_f}{dt} \quad (13)$$

$$\frac{dS_{bp}}{dt} = -\frac{dS_b}{dt} \quad (14)$$

$$\frac{dT_p}{dt} = -\frac{dT}{dt} \quad (15)$$

This system of equations is simulated using a numerical ODE integrator (see Algorithm 2).

Modeling the modulation of RNA polymerase recruitment. After modeling transcription factor activation, we need to model the probability that the transcription factor is bound to DNA, p_T , and, when the transcription factor is DNA-bound, its effect on RNA polymerase recruitment to the promoter site, Δr (see Algorithm 3). Recalling the notation used in the *Transcription* section (Algorithm ??), we want to modulate the j^{th} entry in the v_{synth} vector of RNA polymerase initiation probabilities such that:

$$v_{\text{synth},j} = \alpha_j + \sum_i p_{T,i} \Delta r_{ij} \quad (16)$$

where α_j represents basal recruitment of RNA polymerase and the second term is dependent on transcription factor activity: the probability that the i^{th} transcription factor is DNA-bound is $p_{T,i}$, and the recruitment effect of the i^{th} transcription factor on the j^{th} gene is Δr_{ij} . The α and Δr values are computed prior to simulation based on gene expression values from conditions that modulate transcription factor activity. Values for p_T are calculated as described in Table 1.

Transcription factor type	Promoter-bound probability
Zero-component system	$p_T = 1$ if TF is present, 0 otherwise
One-component system	$p_T = (T^*)/(T^* + T)$
Two-component system	$p_T = (T_p)/(T_p + T)$

Table 1: Formulas used to compute the probability that a transcription factor is promoter-bound. T^* is the active form of a one-component system transcription factor, while T_p is the phosphorylated form of a two-component system transcription factor, and T is the inactive or unphosphorylated form of a transcription factor.

Algorithm 1: Equilibrium binding

Input : c_m counts of molecules where $m = 1$ to $n_{molecules}$
Input : S matrix describing reaction stoichiometries where $S[i, j]$ describes the coefficient for the i^{th} molecule in the j^{th} reaction
Input : $reactants_j$ set of indices for c_m of reactant molecules that participate in the j^{th} reaction
Input : $product_j$ index for c_m of the product molecule formed by the j^{th} reaction
Input : f conversion factor to convert molecule counts to concentrations
Input : $K_{d,j}$ dissociation constant where $j = 1$ to $n_{reactions}$

1. Dissociate all complexes in c into constituent molecules to get total reactants (d) since some reactants participate in multiple reactions:

$$d = c$$
for each ligand-binding reaction, j do

for each molecule, i do

$d_i = d_i + c_{product_j} \cdot S[i, j]$

 $d_{product_j} = 0$
2. Find the number of reactions to perform (n_j) to minimize the distance from $K_{d,j}$, where r is a positive integer and not greater than the total products that can be formed by the reactants:
 for each ligand-binding reaction, j do

$$n_j = \underset{r}{\operatorname{argmin}} \left| \frac{\prod_{i \in reactants_j} (f \cdot (d_i - r))^{S[i, j]}}{f \cdot r} - K_{d,j} \right|$$
3. Update counts (c) based on number of reactions that will occur. Starting from the dissociated counts, reactants will decrease by the number of reactions and their stoichiometry and one product will be formed for each reaction.

$$c = d$$
for each ligand-binding reaction, j do

for each molecule in $reactants_j$, i do

$c_i = c_i - S[i, j] \cdot n_j$

 $c_{product_j} = n_j$

Result: Ligands are bound to or unbound from their binding partners in a fashion that maintains equilibrium.

Algorithm 2: Two-component systems

Input : Δt length of current time step

Input : c_m counts of molecules where $m = 1$ **to** $n_{molecules}$

Input : k_A rate of phosphorylation of free histidine kinase

Input : k_B rate of phosphorylation of ligand-bound histidine kinase

Input : k_C rate of phosphotransfer from phosphorylated free histidine kinase to response regulator

Input : k_D rate of phosphotransfer from phosphorylated ligand-bound histidine kinase to response regulator

Input : k_E rate of dephosphorylation of phosphorylated response regulator

Input : `solveToNextTimeStep()` function that solves two-component system ordinary differential equations to the next time step and returns the change in molecule counts (Δc_m)

1. Solve the ordinary differential equations describing phosphotransfer reactions to perform reactions to the next time step (Δt) using c_m , k_A , k_B , k_C , k_D and k_E .

$$\Delta c_m = \text{solveToNextTimeStep}(c_m, k_A, k_B, k_C, k_D, k_E, \Delta t)$$

2. Update molecule counts.

$$c_m = c_m + \Delta c_m$$

Result: Phosphate groups are transferred from histidine kinases to response regulators and back in response to counts of ligand stimulants.

Algorithm 3: Transcription factor binding

Input : c_a^i counts of active transcription factors where $i = 1$ **to** $n_{\text{transcription factors}}$
Input : c_i^i counts of inactive transcription factors where $i = 1$ **to** $n_{\text{transcription factors}}$
Input : P_i list of promoter sites for each transcription factor where $i = 1$ **to** $n_{\text{transcription factors}}$
Input : t_i type of transcription factor (either one of two-component, one-component, or zero-component) where $i = 1$ **to** $n_{\text{transcription factors}}$
Input : **randomChoice()** function that randomly samples elements from an array without replacement
for *each transcription factor, i* **do**
 if *active transcription factors are present* **then**
 1. Compute probability p of binding the target promoter.
 if t_i *is zero-component transcription factor* **then**
 transcription factor present $\rightarrow p_T = 1$
 transcription factor not present $\rightarrow p_T = 0$
 else

$$p_T = \frac{c_a^i}{c_a^i + c_i^i}$$

 2. Distribute transcription factors to gene targets.

$$P_i^{\text{bound}} = \text{randomChoice}(\text{from } P_i \text{ sample } p_T \cdot \text{len}(P_i) \text{ elements})$$

 3. Decrement counts of free transcription factors.
Result: Activated transcription factors are bound to their gene targets.

Associated data

Parameter		Symbol	Units	Value	Reference
Ligand::TF	dissociation constant	$k_d = k_r/k_f$	μM	[2e-15, 5e3]	See GitHub
Free HK	phosphorylation rate	k_A	$\mu\text{M/s}$	[1e-4, 5e2]	See GitHub
Ligand::HK	phosphorylation rate	k_B	$\mu\text{M/s}$	1.7e5	See GitHub
Phosphotransfer	rate from free HK-P to TF	k_C	$\mu\text{M/s}$	1e8	See GitHub
Phosphotransfer	rate from ligand::HK-P to TF	k_D	$\mu\text{M/s}$	1e8	See GitHub
Dephosphorylation	rate of TF-P	k_E	$\mu\text{M/s}$	1e-2	See GitHub
DNA::TF	dissociation constant	K_d	pM	[2e-4, 1.1e5]	See GitHub
Promoter sites		n	targets per chromosome	[1, 108]	See GitHub
Fold-change	gene expression	FC	$\log_2(a.u.)$	[-10.48, 9.73]	See GitHub

Table 2: Table of parameters for equilibrium binding, two-component systems, and transcription factor binding Processes. HK: histidine kinase, TF: transcription factor, HK-P: phosphorylated histidine kinase, TF-P: phosphorylated transcription factor. Note in this and future tables we reference the source code for our model, which will be freely available at GitHub as noted in the main text.

Associated files

wcEcoli Path	File	Type
wcEcoli/models/ecoli/processes	equilibrium.py	process
wcEcoli/models/ecoli/processes	tf_binding.py	process
wcEcoli/models/ecoli/processes	two_component_system.py	process
wcEcoli/reconstruction/ecoli/dataclasses/process	equilibrium.py	data
wcEcoli/reconstruction/ecoli/dataclasses/process	transcription_regulation.py	data
wcEcoli/reconstruction/ecoli/dataclasses/process	two_component_system.py	data
wcEcoli/reconstruction/ecoli/flat	equilibriumReactions.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	foldChanges.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	tfIds.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	tfOneComponentBound.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	twoComponentSystems.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	twoComponentSystemTemplates.tsv	raw data

Table 3: Table of files for transcription regulation.

References

- [1] Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S Tawfik, and Ron Milo. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–4410, 2011.
- [2] Bryson D Bennett, Elizabeth H Kimball, Melissa Gao, Robin Osterhout, Stephen J Van Dien, and Joshua D Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nature chemical biology*, 5(8):593–599, 2009.
- [3] Jonathan A Bernstein, Arkady B Khodursky, Pei-Hsun Lin, Sue Lin-Chao, and Stanley N Cohen. Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays. *Proceedings of the National Academy of Sciences*, 99(15):9697–9702, 2002.
- [4] Elsa W Birch, Madeleine Udell, and Markus W Covert. Incorporation of flexible objectives and time-linked simulation with flux balance analysis. *Journal of Theoretical Biology*, 345:12–21, March 2014.
- [5] F R Blattner, G Plunkett, C A Bloch, N T Perna, V Burland, M Riley, J Collado-Vides, J D Glasner, C K Rode, G F Mayhew, J Gregor, N W Davis, H A Kirkpatrick, M A Goeden, D J Rose, B Mau, and Y Shao. The complete genome sequence of Escherichia coli K-12. *Science (New York, N.Y.)*, 277(5331):1453–1462, September 1997.

- [6] H. Bremer and G Churchward. An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *Journal of Theoretical Biology*, 69(4):645–654, December 1977.
- [7] Hans Bremer and Patrick Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology*, 2:1553–1569, 1996.
- [8] Hans Bremer and Patrick P Dennis. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1), 2008.
- [9] Javier Carrera, Raissa Estrela, Jing Luo, Navneet Rai, Athanasios Tsoukalas, and Ilias Tagkopoulos. An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of escherichia coli. *Molecular systems biology*, 10(7):735, 2014.
- [10] W D Donachie. Relationship between cell size and time of initiation of DNA replication. *The American Journal of Gastroenterology*, 219(5158):1077–1079, September 1968.
- [11] Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krummenacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Molecular systems biology*, 3(1):121, 2007.
- [12] SYea Gerdes, MD Scholle, JW Campbell, G Balazsi, E Ravasz, MD Daugherty, AL Somera, NC Kyrpides, I Anderson, MS Gelfand, et al. Experimental determination and system level analysis of essential genes in escherichia coli mg1655. *Journal of bacteriology*, 185(19):5673–5684, 2003.

- [13] Jeremy Gunawardena. Silicon dreams of cells into symbols. *Nature biotechnology*, 30(9):838, 2012.
- [14] Anubhav Jain, Shyue Ping Ong, Wei Chen, Bharat Medasani, Xiaohui Qu, Michael Kocher, Miriam Brafman, Guido Petretto, Gian-Marco Rignanes, Geoffroy Hautier, Daniel Gunter, and Kristin A. Persson. Fireworks: a dynamic workflow system designed for high-throughput applications. *Concurrency and Computation: Practice and Experience*, 27(17):5037–5059, 2015. CPE-14-0307.R2.
- [15] Peter D Karp, Daniel Weaver, Suzanne Paley, Carol Fulcher, Aya Kubo, Anamika Kothari, Markus Krummenacker, Pallavi Subhraveti, et al. The EcoCyc database. *EcoSal Plus*, 2014.
- [16] Jonathan R. Karr, Jayodita C Sanghvi, Derek N Macklin, Miriam V Gutschow, Jared M Jacobs, Benjamin Bolival Jr., Nacyra Assad-Garcia, John I Glass, and Markus W Covert. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, 150(2):389–401, July 2012.
- [17] Ingrid M Keseler, Amanda Mackie, Martin Peralta-Gil, Alberto Santos-Zavaleta, Socorro Gama-Castro, César Bonavides-Martínez, Carol Fulcher, Araceli M Huerta, Anamika Kothari, Markus Krummenacker, Mario Latendresse, Luis Muñoz-Rascado, Quang Ong, Suzanne Paley, Imke Schröder, Alexander G Shearer, Pallavi Subhraveti, Mike Travers, Deepika Weerasinghe, Verena Weiss, Julio Collado-Vides, Robert P Gunsalus, Ian Paulsen, and Peter D Karp. EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research*, 41(Database issue):D605–12, January 2013.

- [18] Masanari Kitagawa, Takeshi Ara, Mohammad Arifuzzaman, Tomoko Ioka-Nakamichi, Eiji Inamoto, Hiromi Toyonaga, and Hirotada Mori. Complete set of orf clones of escherichia coli aska library (a complete s et of e. coli k -12 orf a rchive): Unique resources for biological research. *DNA Research*, 12(5):291, 2005.
- [19] Masanari Kitagawa, Takeshi Ara, Mohammad Arifuzzaman, Tomoko Ioka-Nakamichi, Eiji Inamoto, Hiromi Toyonaga, and Hirotada Mori. Complete set of orf clones of escherichia coli aska library (a complete s et of e. coli k-12 orf a rchive): Unique resources for biological research. *DNA research*, 12(5):291–299, 2005.
- [20] Bo Li and Colin N Dewey. Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC bioinformatics*, 12(1):323, 2011.
- [21] Bo Li and Colin N Dewey. Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC bioinformatics*, 12(1):323, 2011.
- [22] Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, 2014.
- [23] Michihisa Maeda, Tomohiro Shimada, and Akira Ishihama. Strength and regulation of seven rrna promoters in escherichia coli. *PloS one*, 10(12):e0144697, 2015.
- [24] Maurizi MR. Proteases and protein degradation in escherichia coli. *Experientia*, 48(2):178–201, 1992.
- [25] Kamalendu Nath and Arthur L. Koch. Protein degradation in escherichia coli : I. measurement of rapidly and slowly decaying components. *Journal of Biological Chemistry*, 245(11):2889–2900, 1970.

- [26] Frederick C Neidhardt, Philip L Bloch, and David F Smith. Culture medium for enterobacteria. *Journal of bacteriology*, 119(3):736–747, 1974.
- [27] Frederick Carl Neidhardt, John L Ingraham, and Moselio Schaechter. *Physiology of the bacterial cell: a molecular approach*. Sinauer Associates Sunderland, MA, 1990.
- [28] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of escherichia coli metabolism2011. *Molecular systems biology*, 7(1):535, 2011.
- [29] J B Russell and G M Cook. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiological reviews*, 59(1):48–62, March 1995.
- [30] John T Sauls, Dongyang Li, and Suckjoon Jun. Adder and a coarse-grained approach to cell size homeostasis in bacteria. *Current opinion in cell biology*, 38:38–44, 2016.
- [31] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nature biotechnology*, 34(1):104–110, 2016.
- [32] Ida Schomburg, Antje Chang, Sandra Placzek, Carola Söhngen, Michael Rother, Maren Lang, Cornelia Munaretto, Susanne Ulas, Michael Stelzer, Andreas Grote, Maurice Scheer, and Dietmar Schomburg. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic acids research*, 41(Database issue):D764–72, January 2013.

- [33] Yu Tanouchi, Anand Pai, Heungwon Park, Shuqiang Huang, Rumen Stamatov, Nicolas E Buchler, and Lingchong You. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature*, 523(7560):357, 2015.
- [34] J W Tobias, T E Shrader, G Rocap, and A Varshavsky. The N-end rule in bacteria. *Science (New York, N.Y.)*, 254(5036):1374–1377, November 1991.
- [35] Yoshihiro Toya, Nobuyoshi Ishii, Kenji Nakahigashi, Takashi Hirasawa, Tomoyoshi Soga, Masaru Tomita, and Kazuyuki Shimizu. ¹³c-metabolic flux analysis for batch culture of escherichia coli and its pyk and pgi gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnology progress*, 26(4):975–992, 2010.
- [36] Mats Wallden, David Fange, Ebba Gregorsson Lundius, Özden Baltekin, and Johan Elf. The synchronization of replication and division cycles in individual e. coli cells. *Cell*, 166(3):729–739, 2016.
- [37] Daniel S Weaver, Ingrid M Keseler, Amanda Mackie, Ian T Paulsen, and Peter D Karp. A genome-scale metabolic flux model of Escherichia coli K-12 derived from the EcoCyc database. *BMC Systems Biology*, 8:79, June 2014.