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. We address these topics sequentially below.

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

1. **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.

2. **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.

3. **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.

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:

$$nL + T \rightleftharpoons T^* \tag{1}$$

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

With the dissociation constant K_d defined as:

$$K_d = \frac{L^n \cdot T}{T^*} \tag{2}$$

we have:

$$\frac{T^*}{T_T} = \frac{L^n}{L^n + K_d} \tag{3}$$

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:

$$L + S_f \rightleftharpoons S_b \tag{4}$$

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

$$S_f + ATP \overset{k_A}{\rightarrow} S_{fp} + ADP + H^+ \tag{5}$$

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

$$S_b + ATP \overset{k_B}{\rightarrow} S_{bp} + ADP + H^+ \tag{6}$$

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

$$S_{fp} + T \overset{k_C}{\rightarrow} S_f + T_p \tag{7}$$

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

$$S_{bp} + T \overset{k_D}{\rightarrow} S_b + T_p \tag{8}$$

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

$$T_p + H_2O \xrightarrow{k_E} T + P_i \tag{9}$$

By assuming mass-action kinetics, we can represent this system mathematically using ordinary differential equations. Ligand binding is simulated in a fashion identical to the one-component systems and the rest of the sub-model is simulated using a numerical ODE integrator (see Algorithm 2).

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

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:

$$v_{\text{synth},j} = \alpha_j + \sum_i P_{T,i} \Delta r_{ij} \tag{10}$$

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: Algorithm for 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 : K_d^r dissociation constant where $r = 1$ **to** $n_{reactions}$

for *each ligand-binding reaction* j **do**

1. Dissociate all complexes in c formed by reaction j into constituent molecules

while *True* **do**

1. Form complex described by $S[:, j]$

if $\left| \frac{c_{reactant1}^{S[reactant1,j]} \cdot c_{reactant2}^{S[reactant2,j]} \cdot \ldots \cdot c_{reactantm}^{S[reactantm,j]}}{c_{complex}} - K_d \right|$ *has reached a minimum (i.e., the ratio of reactants to products is as close as possible to the dissociation constant)* **then**

1. Set reactant and product values in c to these levels

2. Break out of while loop

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

Algorithm 2: Algorithm for 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.

$\Delta c_m =$ `solveToNextTimeStep(`c_m, k_A, k_B, k_C, k_D, k_E, Δt`)`

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: Algorithm for 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* **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 $\rightarrow p = 1$
      transcription factor not present $\rightarrow p = 0$
    else
      $p = \frac{c_a^i}{c_a^i + c_i^i}$
    2. Distribute transcription factors to gene targets.
      $P_i^{bound} =$ `randomChoice(`*from* P_i *sample* $p \cdot len(P_i)$ *elements*`)`
    3. Decrement counts of free transcription factors.
  else
    move on to next transcription factor

Result: Activated transcription factors are bound to their gene targets.

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

Table 2: Table of files for transcription regulation.

Difference from *M. genitalium* model. The most significant difference from the *M. genitalium* model is the enhanced coverage of the regulatory network; 438 regulatory interactions are described by 22 transcription factors that regulate 355 genes.

Accordingly, regulation is represented by three different classes of transcription regulators: two-component system, one-component system and zero-component systems. While the phosphotransfer reactions of two-component signaling pathways are modeled in `TwoComponentSystems`, one-component systems (which bind directly to the transcription factor) and zero-component systems (whose presence or absence determines activity) are modeled by the `EquilibriumBinding` and `TranscriptionFactorBinding` Processes.

Associated data

Parameter	Symbol	Units	Value	Reference
Ligand::TF dissociation constant	$k_d = k_r/k_f$	μM	[2e-15, 5e3]	See Source Code
Free HK phosphorylation rate	k_A	μM/s	[1e-4, 5e2]	See Source Code
Ligand::HK phosphorylation rate	k_B	μM/s	1.7e5	See Source Code
Phosphotransfer rate from free HK-P to TF	k_C	μM/s	1e8	See Source Code
Phosphotransfer rate from ligand::HK-P to TF	k_D	μM/s	1e8	See Source Code
Dephosphorylation rate of TF-P	k_E	μM/s	1e-2	See Source Code
DNA::TF dissociation constant	K_d	pM	[2e-4, 1.1e5]	See Source Code
Promoter sites	n	targets per chromosome	[1, 108]	See Source Code
Fold-change gene expression	FC	$log_2(a.u.)$	[-10.48, 9.73]	See Source Code
Gene expression profile shifts	-	shifts	294*	See Source Code

Table 3: 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. *We found 144 pairs of comparable shifts (see Fig. S2a). Note in this and future tables we reference the source code for our model, which will be freely available at GitHub and SimTK as noted in the main text.

References