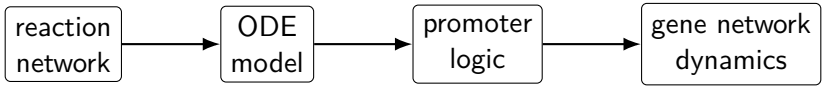
Mass Action Law and Gene Regulation Models

From reaction networks to transcriptional network dynamics

What these two lectures build toward

- The mass action law turns a reaction diagram into differential equations.
- Time-scale separation lets us replace fast promoter binding dynamics by algebraic input-output functions.
- Cooperativity turns smooth biochemical responses into switch-like responses.
- Once we have input-output functions for regulated promoters, we can build ODE models for transcriptional networks.



Lecture map

Lecture 1: Mass action and single genes

1. Assumptions behind mass action
2. Reaction fluxes
3. Converting reactions into ODEs
4. A constitutively expressed gene
5. Response time and degradation

Lecture 2: Regulation and networks

1. Repression by promoter binding
2. Time-scale separation
3. Cooperativity and Hill functions
4. Multiple inputs and AND gates
5. Two-node transcriptional networks

The modeling problem

We want a deterministic model for the mean concentration or abundance of each molecular species.

State variables

$$X_i(t) = \text{concentration or copy-number proxy for species } i, \qquad i = 1, \ldots, N.$$

- A stochastic model tracks probabilities of molecule counts.
- A deterministic mass action model tracks only the time evolution of the state variables $X_i(t)$.
- This is a good approximation when copy numbers are large enough that fluctuations are relatively small.

When mass action is a reasonable approximation

The mass action law assumes more than "molecules react randomly."

1. **Well-mixed system:** molecules encounter each other in proportion to their concentrations.
2. **Thermal equilibrium at the reaction scale:** microscopic collisions are not organized by spatial structure.
3. **Large-volume / large-copy-number limit:** fluctuations become small relative to the mean.

Important limitation

If a molecule exists in one or a few copies, its behavior is not well approximated by a deterministic concentration equation.

Reaction notation

A chemical reaction network is a list of reactions among molecular species.

$$X_1 + X_2 \xrightarrow{k_1} X_3.$$

- X_i names a molecular species and also denotes its modeled concentration.
- k_ℓ is the reaction rate constant for reaction ℓ.
- Each reaction contributes a signed term to every ODE: negative for consumed species, positive for produced species.

Flux: the speed of one reaction

For the elementary reaction

$$X_i + X_j \xrightarrow{k_\ell} \text{products},$$

mass action assigns the deterministic reaction flux

$$\nu_\ell = k_\ell X_i X_j.$$

- The flux is the rate at which this reaction occurs.
- It is deterministic: once X_i, X_j, and k_ℓ are specified, the reaction speed is fixed.
- It is related to the stochastic collision probability, but the deterministic constant k_ℓ need not equal the microscopic stochastic constant exactly.

General mass action flux

For a reaction with multiple reactants,

$$n_1 X_1 + n_2 X_2 + \cdots + n_m X_m \xrightarrow{k_\ell} \text{products},$$

the mass action flux is

$$\boxed{\nu_\ell = k_\ell X_1^{n_1} X_2^{n_2} \cdots X_m^{n_m}.}$$

- The exponent is the stoichiometric coefficient among the reactants.
- Example: if two copies of X_3 must collide, then $2X_3 \to$ products has flux kX_3^2.

How to build ODEs from reactions

Use the same procedure every time.

1. List every molecular species whose dynamics you want to track.
2. List every reaction and assign a rate constant.
3. Write the mass action flux for each reaction.
4. For each species, add production fluxes and subtract consumption fluxes.

Template

$$\frac{dX_i}{dt} = \sum_{\ell} (\text{net molecules of } X_i \text{ produced by reaction } \ell)\nu_\ell.$$

Worked example: reaction network

Consider three molecular species and three reactions:

$$X_1 + X_2 \xrightarrow{k_1} X_3, \qquad X_1 \xrightarrow{k_2} 0, \qquad 2X_3 \xrightarrow{k_3} X_2.$$

Worked example: compute the fluxes

The reactions are

$$X_1 + X_2 \xrightarrow{k_1} X_3, \qquad X_1 \xrightarrow{k_2} 0, \qquad 2X_3 \xrightarrow{k_3} X_2.$$

Therefore the fluxes are

$$\nu_1 = k_1 X_1 X_2, \qquad \nu_2 = k_2 X_1, \qquad \nu_3 = k_3 X_3^2.$$

- ν_1 consumes one X_1, consumes one X_2, produces one X_3.
- ν_2 consumes one X_1.
- ν_3 consumes two X_3, produces one X_2.

Worked example: write the ODEs

Add the signed contributions to each species.

$$\frac{dX_1}{dt} = -\nu_1 - \nu_2 = -k_1 X_1 X_2 - k_2 X_1,$$

$$\frac{dX_2}{dt} = -\nu_1 + \nu_3 = -k_1 X_1 X_2 + k_3 X_3^2,$$

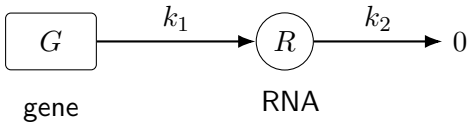
$$\frac{dX_3}{dt} = +\nu_1 - 2\nu_3 = k_1 X_1 X_2 - 2k_3 X_3^2.$$

Check the stoichiometry

If the reaction is $2X_3 \to X_2$, then each event consumes two copies of X_3. Therefore the X_3 equation has $-2\nu_3$, not $-\nu_3$.

A minimal model of constitutive expression

A gene produces RNA, and RNA is removed by degradation or dilution.



$$G \xrightarrow{k_1} G + R, \qquad R \xrightarrow{k_2} 0.$$

- The gene is a catalyst for RNA production: it appears on both sides of the production reaction.
- Therefore the gene copy number is not consumed by transcription.

Step 1: write the gene equation

For the transcription reaction

$$G \xrightarrow{k_1} G + R,$$

one copy of G is consumed and one copy of G is produced in the same reaction event.

$$\frac{dG}{dt} = -k_1 G + k_1 G = 0.$$

Thus $G(t) = G(0)$. If the gene copy number is constant, define

$$E = k_1 G.$$

Here E is the effective production rate of RNA.

Step 2: write the RNA equation

RNA is produced by transcription and removed by degradation or dilution:

$$\frac{dR}{dt} = k_1 G - k_2 R.$$

Using $E = k_1 G$, the equation becomes

$$\boxed{\frac{dR}{dt} = E - k_2 R.}$$

- E controls how much RNA is produced per unit time.
- $k_2 R$ says that removal is proportional to how much RNA is already present.

Step 3: solve the RNA equation

Assume $R(0) = 0$. The ODE is

$$\frac{dR}{dt} = E - k_2 R.$$

Its solution is

$$\boxed{R(t) = \frac{E}{k_2}\left(1 - e^{-k_2 t}\right).}$$

The steady-state RNA level is

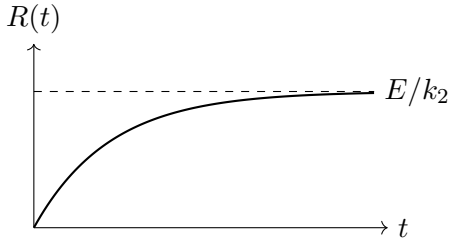
$$R_{\text{ss}} = \lim_{t\to\infty} R(t) = \frac{E}{k_2}.$$

What controls the final level?

The steady state is

$$R_{\mathrm{ss}} = \frac{E}{k_2}.$$

- Increasing production E increases the final RNA level.
- Increasing degradation or dilution k_2 decreases the final RNA level.
- The ratio E/k_2 is the nominal expression level of this simple gene.



What controls the speed of the response?

The halfway time T_{50} is defined by

$$R(T_{50}) = \frac{1}{2} R_{\mathrm{ss}}.$$

Substitute the solution:

$$\frac{E}{k_2}\left(1 - e^{-k_2 T_{50}}\right) = \frac{1}{2}\frac{E}{k_2}.$$

Cancel E/k_2:

$$1 - e^{-k_2 T_{50}} = \frac{1}{2}.$$

Therefore

$$\boxed{T_{50} = \frac{\log 2}{k_2}.}$$

Interpretation of the response time

$$T_{50} = \frac{\log 2}{k_2}.$$

- In this model, the response speed is controlled by degradation or dilution, not by production.
- To make the response faster, the cell must increase k_2. =k2A+k2Dil
- Increasing production E raises the final level, but it does not change the exponential time constant. =0+k2Dil

Biological point

For many stable molecules, effective loss is dominated by cell growth and division. In a per-cell measurement, dilution can reduce molecule number even when molecules are not actively degraded.

Dilution by cell division

Suppose there is no production and the cell divides after one cell-cycle time T_{cycle}. Then the amount per cell is approximately halved:

$$R(T_{\text{cycle}}) = \frac{1}{2} R(0).$$

For exponential dilution,

$$R(t) = R(0) e^{-k_2 t}.$$

Thus

$$e^{-k_2 T_{\text{cycle}}} = \frac{1}{2} \quad \Longrightarrow \quad k_2 = \frac{\log 2}{T_{\text{cycle}}}.$$

For molecules without active degradation, the response time can be limited by the cell-cycle time.

Adding regulation: a repressor binds the gene

Let P be a repressor protein. The gene can be free (G) or bound (G^*).

$$G + P \underset{k_4}{\overset{k_3}{\rightleftharpoons}} G^*, \qquad G \xrightarrow{k_1} G + R, \qquad R \xrightarrow{k_2} 0.$$

Only the free gene G produces RNA.

Step 1: write the full mass action equations

The reactions imply

$$\begin{aligned}
\frac{dG}{dt} &= -k_3 GP + k_4 G^*, \\
\frac{dR}{dt} &= k_1 G - k_2 R, \\
\frac{dP}{dt} &= -k_3 GP + k_4 G^*, \\
\frac{dG^*}{dt} &= k_3 GP - k_4 G^*.
\end{aligned}$$

- Binding consumes G and P, and produces G^*.
- Unbinding consumes G^*, and produces G and P.
- RNA production depends only on the free gene concentration G.

Step 2: use conservation laws

The total gene concentration is conserved:

$$G_T = G + G^*.$$

If no other processes produce or remove P, then total repressor is also conserved:

$$P_T = P + G^*.$$

For the input-output derivation, we treat P as an externally specified input. Then the useful conservation law is

$$G^* = G_T - G.$$

Substitute this into the G equation:

$$\frac{dG}{dt} = -k_3 G P + k_4 (G_T - G).$$

Step 3: identify fast and slow variables

Promoter binding and unbinding are usually much faster than RNA production and degradation.

Fast dynamics

$$\frac{dG}{dt} = -k_3 G P + k_4 (G_T - G)$$

Promoter occupancy changes on binding/unbinding time scales.

Slow dynamics

$$\frac{dR}{dt} = k_1 G - k_2 R$$

RNA changes on production/degradation time scales.

Time-scale separation idea

Let the fast promoter state equilibrate almost instantly relative to the slow RNA state.

Step 4: introduce the small parameter

Represent fast binding by scaling the rates:

$$k_3 = \frac{\tilde{k}_3}{\varepsilon}, \qquad k_4 = \frac{\tilde{k}_4}{\varepsilon}, \qquad 0 < \varepsilon \ll 1.$$

Then

$$\varepsilon \frac{dG}{dt} = -\tilde{k}_3 G P + \tilde{k}_4 (G_T - G).$$

In the limit $\varepsilon \to 0$, the left side vanishes:

$$0 = -\tilde{k}_3 G P + \tilde{k}_4 (G_T - G).$$

This is the quasi-steady-state equation for promoter occupancy.

Step 5: solve for the free gene concentration

Start with the quasi-steady-state equation:

$$0 = -k_3 G P + k_4 (G_T - G).$$

Move terms:

$$k_3 G P + k_4 G = k_4 G_T.$$

Factor out G:

$$G(k_3 P + k_4) = k_4 G_T.$$

Therefore

$$\boxed{G(P) = \frac{k_4 G_T}{k_3 P + k_4}.}$$

Step 6: write the regulated RNA model

Substitute the promoter input-output function into the RNA equation:

$$\frac{dR}{dt} = k_1 G(P) - k_2 R.$$

For a simple repressor,

$$\boxed{\frac{dR}{dt} = k_1 \frac{k_4 G_T}{k_3 P + k_4} - k_2 R.}$$

Equivalently, define the dissociation constant

$$K = \frac{k_4}{k_3}.$$

Then

$$\boxed{G(P) = \frac{G_T}{1 + P/K}.}$$

Interpreting the repression function

$$G(P) = \frac{G_T}{1 + P/K}.$$

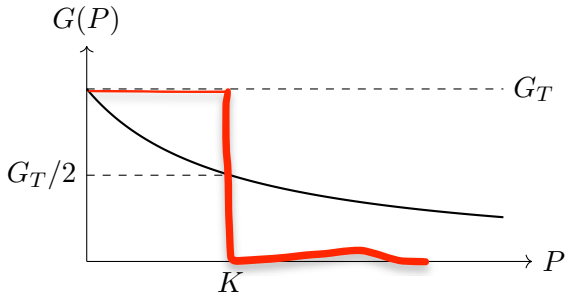
- When $P = 0$, the gene is fully free: $G(0) = G_T$.
- As $P \to \infty$, the free gene concentration approaches zero.
- The half-activation point is found from $G(P_{50}) = G_T/2$:

$$\frac{G_T}{2} = \frac{G_T}{1 + P_{50}/K} \quad \Longrightarrow \quad P_{50} = K.$$

Meaning of K

K is the input level at which half the gene copies are repressed.

Shape of simple repression



$$G(P) = \frac{G_T}{1 + P/K}.$$

The curve is monotone decreasing, but it is not very switch-like when binding is non-cooperative.

Why cooperativity matters

Earlier lectures may approximate regulation as a logic gate: below a threshold the gene is "on," above a threshold it is "off."

Question

When does a biochemical response actually become close to a logic gate?

The answer is cooperativity. If multiple molecules must bind together, the promoter response can become much steeper near its threshold.

Cooperative repression reaction

Suppose repression requires n copies of the repressor:

$$G + nP \underset{k_4}{\overset{k_3}{\rightleftharpoons}} G^*.$$

The mass action binding flux is

$$\nu_{\text{bind}} = k_3 G P^n.$$

The unbinding flux is

$$\nu_{\text{unbind}} = k_4 G^*.$$

At quasi-steady state,

$$k_3 G P^n = k_4 G^*.$$

Deriving the Hill repression function

Use gene conservation $G_T = G + G^*$, so $G^* = G_T - G$. Then

$$k_3 G P^n = k_4 (G_T - G).$$

Solve for G:

$$k_3 G P^n + k_4 G = k_4 G_T,$$

$$G(k_3 P^n + k_4) = k_4 G_T,$$

$$G(P) = \frac{k_4 G_T}{k_3 P^n + k_4}.$$

Define

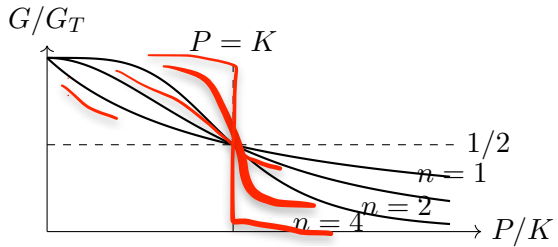
$$K = \left(\frac{k_4}{k_3}\right)^{1/n}.$$

Then

$$\boxed{G(P) = G_T \frac{1}{1 + (P/K)^n}.}$$

Cooperativity makes thresholds sharper

$$G(P) = G_T \frac{1}{1 + (P/K)^n}.$$



Higher n makes the transition from high expression to low expression more switch-like.

Activator version

G = expressing gene, Gstar = off gene, so that our RNA differential stays the same for act and rep

For an activator, expression is low without input and high when enough activator is present. A standard Hill activation function is

Acti

$$\boxed{G(P) = G_T \frac{(P/K)^n}{1 + (P/K)^n}.}$$

Compare the two common regulatory functions:

P=0, f+(P) = 0, as P goes to inf, f+(P) goes to GT

activation: $f_+(P) = \dfrac{(P/K)^n}{1 + (P/K)^n}$, repression: $f_-(P) = \dfrac{1}{1 + (P/K)^n}$.

Both have threshold K, and both become steeper as n increases.

Logic-gate approximation

When n is large, Hill functions approach threshold rules.

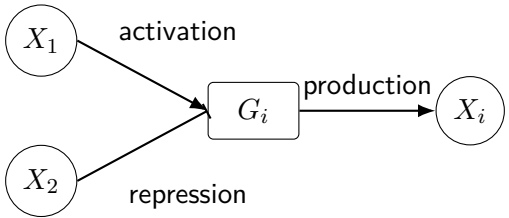
$$f_+(P) = \frac{(P/K)^n}{1+(P/K)^n} \approx \begin{cases} 0, & P < K, \\ 1, & P > K, \end{cases}$$

$$f_-(P) = \frac{1}{1+(P/K)^n} \approx \begin{cases} 1, & P < K, \\ 0, & P > K. \end{cases}$$

This is the mathematical reason that cooperative regulation can behave like a Boolean input in a coarse model.

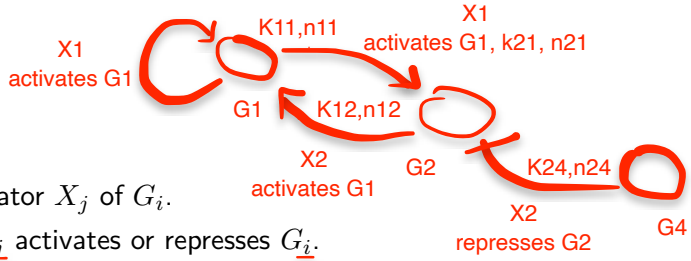
A single gene can have multiple regulatory inputs

Suppose gene G_i is regulated by more than one transcription factor.



The modeling task is to convert this regulatory diagram into a formula for the active gene fraction $G_i(X_1, X_2, \ldots)$.

Step-by-step: build a multi-input promoter model



For a target gene G_i:

1. Identify each regulator X_j of G_i.
2. Decide whether X_j activates or represses G_i.
3. Assign a threshold K_{ij} and cooperativity n_{ij} for each edge $j \to i$.
4. Write one Hill factor for each regulatory edge.
5. Combine the factors according to the promoter logic.
6. Use the resulting active gene level in the production-degradation ODE for X_i.

One regulatory edge gives one Hill factor

For edge $j \to i$, define

$$H_{ij}^{+}(X_j) = \frac{(X_j/K_{ij})^{n_{ij}}}{1 + (X_j/K_{ij})^{n_{ij}}}$$

for activation, and

$$H_{ij}^{-}(X_j) = \frac{1}{1 + (X_j/K_{ij})^{n_{ij}}}$$

for repression.

- K_{ij}: input level at which the effect is half-maximal.
- n_{ij}: steepness/cooperativity of the response.

AND logic using the minimum activity rule

There is no single universal formula for combining inputs; it depends on promoter biochemistry. A useful simple AND-like rule is the minimum activity formula. If all required regulatory conditions must be satisfied, define

$$\boxed{G_i(X) = G_{iT} \min_{j \in \mathcal{R}_i} H_{ij}(X_j).}$$

- $\mathcal{R}_i$ is the set of regulators of gene i.
- Each H_{ij} is either an activation factor or a repression factor.
- If any required input is low, the minimum is low, so the promoter is low.

Example: one activator and one repressor

Suppose gene G_1 is activated by X_1 and repressed by X_2.

$$X_1 \to G_1, \qquad X_2 \dashv G_1.$$

The activation factor is

$$H_{11}^{+}(X_1) = \frac{(X_1/K_{11})^{n_{11}}}{1 + (X_1/K_{11})^{n_{11}}}.$$

The repression factor is

$$H_{21}^{-}(X_2) = \frac{1}{1 + (X_2/K_{21})^{n_{21}}}.$$

The AND-like promoter activity is

$$\boxed{G_1 = G_{1T} \min\left(H_{11}^{+}(X_1),\, H_{21}^{-}(X_2)\right).}$$

The corresponding ODE

Once the active promoter level G_1 is known, write production and degradation for the output X_1:

$$\boxed{\frac{dX_1}{dt} = k_{p1}G_1 - k_{d1}X_1.}$$

This is the general expression for every gene. The differences between genes is in the epression of G_i

Substitute the promoter formula:

$$\frac{dX_1}{dt} = k_{p1}G_{1T} \min\left(\frac{(X_1/K_{11})^{n_{11}}}{1+(X_1/K_{11})^{n_{11}}}, \frac{1}{1+(X_2/K_{21})^{n_{21}}}\right) - k_{d1}X_1.$$

This is a complete dynamical equation for X_1.

General recipe for a transcriptional network

For each gene $i = 1, \ldots, N$:

1. Let $X_i(t)$ be the gene product concentration.
2. Let $G_i(X)$ be the active promoter level, computed from all inputs to gene i.
3. Write production proportional to active promoter level: $k_{pi} G_i(X)$.
4. Write degradation/dilution proportional to gene product concentration: $k_{di} X_i$.

$$\boxed{\frac{dX_i}{dt} = k_{pi} G_i(X) - k_{di} X_i.}$$

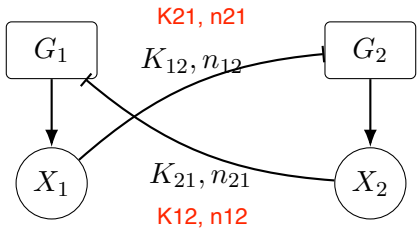
Parameter meanings

Parameter	Meaning
k_{pi}	Production rate scale for gene product X_i.
k_{di}	Degradation or dilution rate for X_i.
G_{iT}	Total promoter/gene copy level for gene i. Often set to 1 after nondimensionalization.
K_{ij}	Threshold for regulator X_j acting on target gene G_i.
n_{ij}	Cooperativity/steepness for the regulatory edge $j \to i$.

Common simplification: set $G_{iT} = 1$ and use k_{pi}, k_{di}, K_{ij}, and n_{ij} as the main model parameters.

How to read a regulatory network diagram

A diagram tells you the inputs to each gene and whether each input activates or represses.



Here X_1 represses G_2, and X_2 represses G_1.

Two-node mutual repression: build the model

Step 1: define variables.

$$X_1(t) = \text{product of gene } G_1, \qquad X_2(t) = \text{product of gene } G_2.$$

Step 2: identify regulatory inputs.

$$X_2 \dashv G_1, \qquad X_1 \dashv G_2.$$

Step 3: choose Hill repression functions.

$$G_1(X_2) = G_{1T}\frac{1}{1+(X_2/K_{21})^{n_{21}}},$$

$$G_2(X_1) = G_{2T}\frac{1}{1+(X_1/K_{12})^{n_{12}}}.$$

Two-node mutual repression: final ODEs

Use the production-degradation template:

$$\frac{dX_i}{dt} = k_{pi} G_i(X) - k_{di} X_i.$$

For the two-node network,

$$\boxed{\begin{aligned} \frac{dX_1}{dt} &= k_{p1} G_{1T} \frac{1}{1 + (X_2/K_{21})^{n_{21}}} - k_{d1} X_1, \\ \frac{dX_2}{dt} &= k_{p2} G_{2T} \frac{1}{1 + (X_1/K_{12})^{n_{12}}} - k_{d2} X_2. \end{aligned}}$$

If $G_{1T} = G_{2T} = 1$, these are the standard two-variable equations for mutual repression.

Parameter choices for the simple two-node example

A clean teaching example uses symmetric parameters:

$$G_{1T} = G_{2T} = 1, \qquad n_{12} = n_{21} = 4.$$

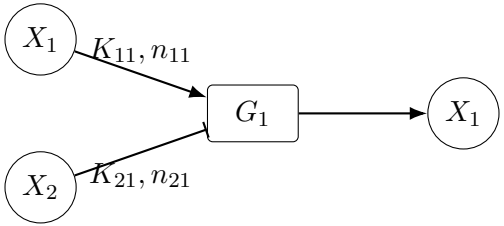
Then the variable parameters are

$$k_{p1}, k_{p2}, k_{d1}, k_{d2}, K_{12}, K_{21}.$$

If the production and degradation rates are also symmetric, the main qualitative behavior is controlled by the thresholds K_{12}, K_{21} and the steepness n.

Two-node network with activation and repression

Now suppose G_1 requires activation by X_1 and absence of repression by X_2.



The promoter activity is

$$G_1 = G_{1T} \min\left(\frac{(X_1/K_{11})^{n_{11}}}{1+(X_1/K_{11})^{n_{11}}}, \frac{1}{1+(X_2/K_{21})^{n_{21}}}\right).$$