Source: Godel, Escher, Bach (Hofstadter).

Systems Biology
Part 1: Motifs in transcription factor networks
Types of Biological Networks

Yeast transcription network
Yeast Protein-Protein interaction network
Yeast Phosphorylation network
E. coli metabolic network
Yeast SSL network


Are there any recurring “subgraphs”, or network motifs, that occur surprisingly often?
Recall: Transcription Factors

A transcription factor can either cause the cell to increase (activate) or decrease (repress) the production of RNA/protein corresponding to a given gene.
Transcription Factor Networks

**Transcription factor network:**

- **Nodes:** genes
- **Edges:** $x$ is connected to $y$ if $x$ is a transcription factor that regulates the expression of $y$.

_E. coli_ transcription factor network
Transcription Factor Networks

Transcription factor network:

- Nodes: genes
- Edges: $x$ is connected to $y$ if $x$ is a transcription factor that regulates the expression of $y$.

Notes:

1. Some of the nodes are transcription factors; others aren’t.
2. Edge $x \rightarrow y$ is also labeled +/- according to whether $x$ activates/represses $y$. 
That sure is a lot of (feedback) loops!

**Autoregulation:** a transcription factor $Y$ regulates *its own* transcription.

Source: https://journals.aps.org/pre/abstract/10.1103/PhysRevE.97.062407
Two Questions

**Question 1:** How can we argue that the number of loops in a TF network is "surprisingly large"?

**Question 2:** If autoregulation is so common, then why did such a strange mechanism evolve?

“Nothing in biology makes sense except in the light of evolution.”

Theodosius Dobzhansky
THE NEGATIVE AUTOREGULATION MOTIF
STOP: What does it mean for a biological network motif to occur “surprisingly often”?

E. coli transcription factor network
STOP: What does it mean for a biological network motif to occur "surprisingly often"?

Answer: It occurs more often than it would if the network were random! (Nothing new here 😊.)
Recall the Definition of TF Network

**Transcription factor network:**

- **Nodes:** genes
- **Edges:** $x$ is connected to $y$ if $x$ is a transcription factor that regulates the expression of $y$. 

Recall the Definition of TF Network

**Transcription factor network:**

- **Nodes:** genes
- **Edges:** $x$ is connected to $y$ if $x$ is a transcription factor that regulates the expression of $y$.

The *E. coli* transcription factor network contains thousands of genes, most of which are not transcription factors. Our random “decoy network” should not have nodes $x \rightarrow y$ if $y$ is not a transcription factor.
Gilbert model for random graphs: given an integer \( n \) and a number \( p \) between 0 and 1, define \( G(n, p) \):

- Form \( n \) nodes.
- For all \( n^2 \) choices of starting node \( x \) and ending node \( y \), connect \( x \) to \( y \) with probability \( p \).
Gilbert model for random graphs: given an integer \( n \) and a number \( p \) between 0 and 1, define \( G(n, p) \):

- Form \( n \) nodes.
- For all \( n^2 \) choices of starting node \( x \) and ending node \( y \), connect \( x \) to \( y \) with probability \( p \).

We limit ourselves to the TF network comprising only transcription factors that regulate each other. This network has 197 nodes and 477 edges.
Gilbert model for random graphs: given an integer $n$ and a number $p$ between 0 and 1, define $G(n, p)$:

- Form $n$ nodes.
- For all $n^2$ choices of starting node $x$ and ending node $y$, connect $x$ to $y$ with probability $p$.

We limit ourselves to the TF network comprising only transcription factors that regulate each other. This network has 197 nodes and 477 edges.

STOP: What should $n$ be in our decoy network?
Constructing Random Networks

**Gilbert model for random graphs:** given an integer \( n \) and a number \( p \) between 0 and 1, define \( G(n, p) \):

- Form \( n \) nodes.
- For all \( n^2 \) choices of starting node \( x \) and ending node \( y \), connect \( x \) to \( y \) with probability \( p \).

We limit ourselves to the TF network comprising only transcription factors that regulate each other. This network has 197 nodes and 477 edges.

**Answer:** \( n = \) the number of transcription factors.
Gilbert model for random graphs: given an integer $n$ and a number $p$ between 0 and 1, define $G(n, p)$:

- Form $n$ nodes.
- For all $n^2$ choices of starting node $x$ and ending node $y$, connect $x$ to $y$ with probability $p$.

We limit ourselves to the TF network comprising only transcription factors that regulate each other. This network has 197 nodes and 477 edges.

STOP: OK, but what should $p$ be?
Gilbert model for random graphs: given an integer \( n \) and a number \( p \) between 0 and 1, define \( G(n, p) \):

- Form \( n \) nodes.
- For all \( n^2 \) choices of starting node \( x \) and ending node \( y \), connect \( x \) to \( y \) with probability \( p \).

Answer: If we were to set \( p \) equal to \( 1/n^2 \), then we would on average only see a single edge in the random network. To get 477 edges on average, we set \( p \) equal to \( 477/n^2 \).
Real vs. Random *E. coli* TF Network

Real

Random

130 loops!

5 loops
Real vs. Random *E. coli* TF Network

And of the 130 loops, 95 are downregulation.
Negative autoregulation: a transcription factor $Y$ represses its own expression.

www.nature.com › letters › article

Engineering stability in gene networks by autoregulation ...

by A Becskei - 2000 - Cited by 1559 - Related articles

Jun 1, 2000 - The genetic and biochemical networks which underlie such things as homeostasis in metabolism and the developmental programs of living ...
Negative autoregulation: a transcription factor $Y$ represses its own expression.

Question 2: If autoregulation is so common, then why did such a strange mechanism evolve?
AN EVOLUTIONARY BASIS FOR NEGATIVE AUTOREGULATION
Simulating a Race to Steady State

Say that a TF $X$ regulates another transcription factor $Y$, and consider two cells. In both cells, $X$ upregulates the transcription of $Y$, but in the second cell, $Y$ also negatively autoregulates.

Cell 1: $X \rightarrow Y$

Cell 2: $X \rightarrow Y \rightarrow -$
We will simulate a “race” to the steady-state concentration of $Y$ in the two cells. The cell that reaches this steady state faster can respond more quickly to its environment and is therefore more fit for survival.
Simulating Particle Level Simulations

**MCell**: A software program that simulates reaction-diffusion models, in which particles interact with each other as they diffuse randomly.

**Monte Carlo methods for simulating realistic synaptic microphysiology using MCell**

JR Stiles, TM Bartol - Computational neuroscience: realistic ..., 2001 - books.google.com

... The MDL and MCell's program flow are summarized in Figure 4.1 and Boxes 4.1 and 4.2. Specific examples follow in Sections 4.5 and 4.6 ... Page 115. Monte Carlo Synaptic Models 93 FIGURE 4.1 General Overview of MCell Simulations ...

☆ 99 Cited by 307  Related articles  All 4 versions

© 2021 Phillip Compeau
Initialization: Start with a constant number of $X$ particles and no $Y$. 
Simulating Cell 1

\[ \text{X} \quad + \quad \text{Y} \]

**Initialization:** Start with a constant number of X particles and no Y.

**Diffusion:** Both X and Y diffuse at the same rate.
Initialization: Start with a constant number of $X$ particles and no $Y$.

Diffusion: Both $X$ and $Y$ diffuse at the same rate.

$X$ activating $Y$: we add the reaction $X \rightarrow X + Y$. In any interval of time, there is some probability that an $X$ particle will produce a $Y$. 

Simulating Cell 1

$X + Y$
Over time, proteins are degraded by proteases so that proteins at high concentration can be removed.
Simulating Cell 1

\[ X + Y \]

Over time, proteins are \textit{degraded} by proteases so that proteins at high concentration can be removed.

**Kill reaction:** $Y$ are removed at some rate.
Simulating Cell 1

\[
\begin{align*}
X & \rightarrow Y
\end{align*}
\]

Over time, proteins are *degraded* by proteases so that proteins at high concentration can be removed.

**Kill reaction:** \(Y\) are removed at some rate.

**Note:** we will assume that \(X\) is at steady-state, so the rate of production of \(X\) balances its removal and we do not need such reactions for \(X\).
Simulating Cell 2

\[ X + Y \xrightarrow{\text{-}} \]

STOP: What reaction could be used to add a simulation of the negative autoregulation of \( Y \)?
**STOP:** What reaction could be used to add a simulation of the negative autoregulation of $Y$?

**Answer:** We will use $Y + Y \rightarrow Y$. When two $Y$ particles encounter each other, there is some probability that one of the particles is removed, which mimics the process of a transcription factor turning off another copy of itself during negative autoregulation.
Comparing concentration of Y in the two simulated cells

STOP: It seems like Cell 1 is winning. What are we missing?
Comparing the Two Approaches

Mathematically controlled comparison (Savageau 1976): we can only compare models on a mathematically level playing field.

Key Point: if we are comparing the two cells, then the steady-state concentration of $Y$ in the two cells should be approximately the same.
Recall Cell 1’s Model

\[ X + Y \]

\textbf{Initialization:} Start with a constant number of \( X \) particles and no \( Y \).

\textbf{Diffusion:} Both \( X \) and \( Y \) diffuse at the same rate.

\textbf{\( X \) activating \( Y \):} we add the reaction \( X \rightarrow X + Y \). In any interval of time, there is some probability that an \( X \) particle will produce a \( Y \).
Recall Cell 1’s Model

\[ X \xrightarrow{+} Y \]

**Initialization:** Start with a constant number of \( X \) particles and no \( Y \).

**Diffusion:** Both \( X \) and \( Y \) diffuse at the same rate.

**\( X \) activating \( Y \):** we add the reaction \( X \rightarrow X + Y \). In any interval of time, there is some probability that an \( X \) particle will produce a \( Y \).

**STOP:** How can we ensure that Cell 2 has a higher steady-state concentration of \( Y \)?
Recall Cell 1’s Model

\[ X + Y \]

**Initialization:** Start with a constant number of \( X \) particles and no \( Y \).

**Diffusion:** Both \( X \) and \( Y \) diffuse at the same rate.

**\( X \) activating \( Y \):** we add the reaction \( X \rightarrow X + Y \). In any interval of time, there is some probability that an \( X \) particle will produce a \( Y \).

**Answer:** The only thing that we can change is increasing the *rate* of \( X \rightarrow X + Y \) in Cell 2.
STOP: Now, why do you think nature has evolved negative autoregulation?
Ensuring that $Y_{st} = X_{st}$

**Answer:** To help a TF respond to a stimulus and reach steady-state faster.
Ensuring that $Y_{st} = X_{st}$

Analogy (Alon): negative autoregulation is like a sportscar with a powerful engine and sensitive brakes.
Ensuring that $Y_{st} = X_{st}$

Cell 2’s “response time is much faster than cell 1’s.

But if the protein is not a TF, can it be turned on faster than simple regulation?
The "feed-forward loop" motif allows a non-TF Z to be turned on quickly

“Type 1” incoherent feed-forward loop

STOP: Why might this motif allow Z to be turned on faster than under simple regulation?
The "feed-forward loop" motif allows a non-TF Z to be turned on quickly.

Answer: $X$ serves to "ramp up" $Z$ quickly, and once $X$ builds up $Y$, $Y$ serves as a delayed-action "brakes" for $Z$. 
Plotting a FFL-regulated protein $Z$ against one regulated by $X \rightarrow Z$
Damped Oscillations

**Note:** The shape of the FFL-regulated protein concentration is similar to a “damped” oscillation.

Source: https://www.toppr.com/guides/physics/oscillations/damped-simple-harmonic-motion/
Real oscillations are everywhere in biology

https://www.thoughtco.com/stages-of-mitosis-373534

https://www.narayanahealth.org/blog/what-is-abnormal-heartbeat/

https://carex.com/blogs/resources/circadian-rhythm
We will build an oscillator in recitation!

**Repressilator**: a three-element synthetic “cycle” motif of repression that produces oscillations.

**A synthetic oscillatory network of transcriptional regulators**

MB Elowitz, S Leibler - Nature, 2000 - nature.com

Networks of interacting biomolecules carry out many essential functions in living cells, but the 'design principles' underlying the functioning of such intracellular networks remain poorly understood, despite intensive efforts including quantitative analysis of relatively simple ...  

☆  Cited by 4909  Related articles  All 82 versions

Oscillator motifs in nature are much more complicated than this!
Part 2: Modeling Bacterial Chemotaxis
MODELING BACTERIAL CHEMOTAXIS
If two immortal people were placed on opposite sides of an uninhabited Earth-like planet, how long would it take them to find each other?
100,000 years?
1,000,000 years?

STOP: Any thoughts?
Munroe’s Answer: “Be an Ant”

• If you have no information, walk at random, leaving a trail of stone markers, each one pointing to the next. For every day that you walk, rest for three. Periodically mark the date alongside the cairn. It doesn’t matter how you do this, as long as it’s consistent. You could chisel the number of days into a rock, or lay out rocks to plot the number.

• If you come across a trail that’s newer than any you’ve seen before, start following it as fast as you can. If you lose the trail and can’t recover it, resume leaving your own trail.

• You don’t have to come across the other player’s current location; you simply have to come across a location where they’ve been. You can still chase one another in circles, but as long as you move more quickly when you’re following a trail than when you’re leaving one, you’ll find each other in a matter of years or decades.

• And if your partner isn’t cooperating—perhaps they’re just sitting where they started and waiting for you—then you’ll get to see some neat stuff.
Bacteria employ a similar randomized algorithm to find food

https://www.youtube.com/watch?v=F6QMU3KD7zw
Bacterial runs and tumbles

An *E. coli* cell has 5-12 flagella on its surface, which can rotate both clockwise and counter-clockwise.

**Chemotaxis:** The movement of an organism in response to a chemical stimulus.

https://www.sciencephoto.com/media/659604/view/e-coli-bacterium-illustration
Bacterial runs and tumbles

An *E. coli* cell has 5-12 flagella on its surface, which can rotate both clockwise and counter-clockwise.

When the flagella are all rotating CCW, they form a bundle and propel the cell forward at 20 µm/s.
Bacterial runs and tumbles

Note: This is about 10x the length of the cell per second, like a car traveling at 160 kph (100 mph).

When the flagella are all rotating CCW, they form a bundle and propel the cell forward at 20 µm/s.
Bacterial runs and tumbles

When any flagellum rotates CW, the flagella are uncoordinated, and the bacterium stops and rotates.

Source: https://courses.lumenlearning.com/microbiology

© 2021 Phillip Compeau
Bacterial runs and tumbles

**Run and tumble model:** when we zoom out, *E. coli* alternates between running and tumbling in place.

Source: https://courses.lumenlearning.com/microbiology

© 2021 Phillip Compeau
*E. coli*’s movement looks like a random walk!

**Question:** what is the *molecular* basis for the random walk movement?

https://commons.wikimedia.org/wiki/File:Random_walk_25000.gif
How *E. coli* detects attractants

*E. coli* has **receptor proteins** that detect attractants such as glucose by binding to and forming a complex with these attractant **ligands**.
How *E. coli* detects attractants

The “signal” of the binding is then “transduced” via a series of internal chemical processes that leads to a change in the flagellar rotation.
The Need for a "Particle-Free" Model

The *E. coli* cell is so small that we will assume that the concentration of any particle in its immediate surroundings is **well-mixed** (i.e., uniform).
The Need for a “Particle-Free” Model

The *E. coli* cell is so small that we will assume that the concentration of any particle in its immediate surroundings is well-mixed (i.e., uniform).

Our model of chemotaxis will have many particles and reactions that depend on each other, and so a “particle-free” model that does not track the diffusion of individual particles will greatly increase efficiency.
THE GILLESPIE ALGORITHM
Say that you own a store and have noticed that on average, there are $\lambda$ customers entering your store in a single hour. Let $X$ be a random variable denoting the number of customers that enter the store in the next hour.
Say that you own a store and have noticed that on average, there are $\lambda$ customers entering your store in a single hour. Let $X$ be a random variable denoting the number of customers that enter the store in the next hour.

$X$ follows a **Poisson distribution**; it can be shown that the probability that exactly $n$ customers arrive in the next hour is

$$\Pr(X = n) = \frac{\lambda^n e^{-\lambda}}{n!}$$
Furthermore, the probability of observing exactly $n$ customers in $t$ hours is

$$\frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

$X$ follows a **Poisson distribution**; it can be shown that the probability that exactly $n$ customers arrive in the next hour is

$$\Pr(X = n) = \frac{\lambda^n e^{-\lambda}}{n!}$$
If we let $T$ be the random variable corresponding to the wait time on the next customer, then the probability of waiting at least $t$ hours is the probability of seeing zero customers in $t$ hours:

$$
\Pr(T > t) = \Pr(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}
$$
If we let $T$ be the random variable corresponding to the wait time on the next customer, then the probability of waiting at least $t$ hours is the probability of seeing zero customers in $t$ hours:

$$\Pr(T > t) = \Pr(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

That is, $\Pr(T > t)$ decays exponentially as $t$ increases; thus, random variable $T$ follows an **exponential distribution**. (Mean wait time: $1/\lambda$).
If we let $T$ be the random variable corresponding to the wait time on the next customer, then the probability of waiting at least $t$ hours is the probability of seeing zero customers in $t$ hours:

$$\Pr(T > t) = \Pr(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

**STOP:** What is the probability $\Pr(T < t)$?
From Poisson to Exponential

If we let $T$ be the random variable corresponding to the wait time on the next customer, then the probability of waiting at least $t$ hours is the probability of seeing zero customers in $t$ hours:

$$\Pr(T > t) = \Pr(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

**STOP:** What is the probability $\Pr(T < t)$?

**Answer:** $1 - e^{-\lambda t}$. 
An overview of the Gillespie algorithm

Given a well-mixed environment and a reaction taking place at some known average rate, how long should we expect to wait before this reaction occurs somewhere in the environment?
An overview of the Gillespie algorithm

Given a well-mixed environment and a reaction taking place at some known average rate, how long should we expect to wait before this reaction occurs somewhere in the environment?

STOP: Remind you of anything?
An overview of the Gillespie algorithm

Given a well-mixed environment and a reaction taking place at some known average rate, how long should we expect to wait before this reaction occurs somewhere in the environment?

**STOP:** Remind you of anything?

**Answer:** We will model each of our reactions using an exponential distribution!
Given a well-mixed environment and a reaction taking place at some known average rate, how long should we expect to wait before this reaction occurs somewhere in the environment?

STOP: Remind you of anything?

Answer: We will model each of our reactions using an exponential distribution!

This idea is the engine of Gillespie’s stochastic simulation algorithm (SSA).
To model ligand-receptor dynamics, we will use a **reversible reaction** in which a ligand $L$ and receptor $T$ bond and dissociate at different rates.

$$T + L \rightleftharpoons LT$$

**STOP:** Why would a ligand and receptor need to dissociate?
Modeling a single ligand-receptor reaction

To model ligand-receptor dynamics, we will use a **reversible reaction** in which a ligand \( L \) and receptor \( T \) bond and dissociate at different rates.

\[
T + L \rightleftharpoons LT
\]

**Answer:** We don’t want to detect temporary changes permanently.
Say that the rate of $T + L \rightarrow LT$ is $k_{\text{bind}}$ and the rate of $LT \rightarrow T + L$ is $k_{\text{dissociate}}$. 
Return to Gillespie

Say that the rate of $T + L \rightarrow LT$ is $k_{\text{bind}}$ and the rate of $LT \rightarrow T + L$ is $k_{\text{dissociate}}$.

Repeat the following steps for the entire simulation.
1. Pick a wait time according to an exponential distribution with $\lambda = k_{\text{bind}} + k_{\text{dissociate}}$. 
Say that the rate of $T + L \rightarrow LT$ is $k_{\text{bind}}$ and the rate of $LT \rightarrow T + L$ is $k_{\text{dissociate}}$.

Repeat the following steps for the entire simulation.
1. Pick a wait time according to an exponential distribution with $\lambda = k_{\text{bind}} + k_{\text{dissociate}}$.
2. The probability that the reaction is the forward reaction is $\Pr(L + T \rightarrow LT) = \frac{k_{\text{bind}}}{k_{\text{bind}} + k_{\text{dissociate}}}$.
Say that the rate of $T + L \rightarrow LT$ is $k_{\text{bind}}$ and the rate of $LT \rightarrow T + L$ is $k_{\text{dissociate}}$.

Repeat the following steps for the entire simulation.

1. Pick a wait time according to an exponential distribution with $\lambda = k_{\text{bind}} + k_{\text{dissociate}}$.

2. The probability that the reaction is the forward reaction is $\Pr(L + T \rightarrow LT) = \frac{k_{\text{bind}}}{k_{\text{bind}} + k_{\text{dissociate}}}$.

3. The probability that it is the reverse reaction is $\Pr(LT \rightarrow L + T) = \frac{k_{\text{dissociate}}}{k_{\text{bind}} + k_{\text{dissociate}}}$.
Running Gillespie on a real example

\[ L_0 = 10,000; \ T_0 = 7,000; \ LT_0 = 0 \]
\[ k_{\text{bind}} = 0.0146(\text{molecules/}\mu\text{m}^3)^{-1}s^{-1}; \ k_{\text{dissociate}} = 35s^{-1} \]
BUILDING A COMPLETE MODEL OF CHEMOTAXIS
The engine of signal transduction is a “cascade” of phosphorylation events, attaching a phosphoryl group (PO$_3^-$) to an organic molecule.
Phosphoryl can be broken off an adenosine triphosphate (ATP) molecule, or exchanged as part of *dephosphorylation* of a phosphorylated molecule.
Explaining the molecular basis for signaling

Receptors form a complex on the inside of the cell with CheA and CheW proteins, which is more stable without ligand binding. Remember this fact!
When bound, CheA **autophosphorylates**, adding a phosphoryl group to itself – not a strange concept after autoregulation 😊
Explaining the molecular basis for signaling

When phosphorylated, CheA can pass on the phosphoryl group to a molecule called CheY.
Explaining the molecular basis for signaling

When phosphorylated CheY interacts with the **flagellar motor switch** protein complex on the flagellum, it changes rotation from CCW to CW.
Bacterial runs and tumbles

Recall: when the flagella are all rotating CCW, they form a bundle and propel the cell forward at 20 µm/s.

STOP: What happens when the flagellum rotates CW instead?

https://www.sciencephoto.com/media/659604/view/e-coli-bacterium-illustration
Bacterial runs and tumbles

Recall: when the flagella are all rotating CCW, they form a bundle and propel the cell forward at 20 µm/s.

STOP: What happens when the flagellum rotates CW instead?

Answer: Tumble!

https://www.sciencephoto.com/media/659604/view/e-coli-bacterium-illustration
Explaining the molecular basis for signaling

Remember: this whole process is more likely when ligand is not present. So, less ligand means more tumbling, and more ligand means more running.
Remember: this whole process is more likely when ligand is *not* present. So, less ligand means more tumbling, and more ligand means more running.

**Note:** The default tumbling frequency of *E. coli* is once every 1-1.5 seconds, which is similar for many other bacteria as well.
Explaining the molecular basis for signaling

Remember: this whole process is *more likely* when ligand is *not* present. So, less ligand means more tumbling, and more ligand means more running.

**Note:** The default tumbling frequency of *E. coli* is once every 1-1.5 seconds, which is similar for many other bacteria as well.

In recitation, we will apply the Gillespie algorithm to implement all of the reactions; it turns out that chemotaxis is even more complex than it seems!
A note on Gillespie with multiple different reactions

Recall that the Gillespie algorithm works for a reversible reaction in the following way.

Repeat the following steps for the entire simulation.
1. Pick a wait time according to an exponential distribution with \( \lambda = k_{\text{bind}} + k_{\text{dissociate}} \).
2. The probability that the reaction is the forward reaction is \( \Pr(L + T \rightarrow LT) = \frac{k_{\text{bind}}}{k_{\text{bind}} + k_{\text{dissociate}}} \).
3. The probability that it is the reverse reaction is \( \Pr(LT \rightarrow L + T) = \frac{k_{\text{dissociate}}}{k_{\text{bind}} + k_{\text{dissociate}}} \).
A note on Gillespie with multiple different reactions

STOP: The question is how to generalize this idea to $n$ reactions, having rates $k_1, k_2, \ldots, k_n$. Ideas?

Repeat the following steps for the entire simulation.

1. Pick a wait time according to an exponential distribution with $\lambda = k_{\text{bind}} + k_{\text{dissociate}}$.
2. The probability that the reaction is the forward reaction is $\Pr(L + T \rightarrow LT) = \frac{k_{\text{bind}}}{k_{\text{bind}} + k_{\text{dissociate}}}$.
3. The probability that it is the reverse reaction is $\Pr(LT \rightarrow L + T) = \frac{k_{\text{dissociate}}}{k_{\text{bind}} + k_{\text{dissociate}}}$.
A note on Gillespie with multiple different reactions

Answer: It’s easier than you might imagine! We just extend the sums over $n$ terms.

Repeat the following steps for the entire simulation.
1. Pick a wait time according to an exponential distribution with \( \lambda = k_1 + k_2 + \ldots + k_n \).
2. The probability that the $i$-th reaction is the current reaction is \( k_i / (k_1 + k_2 + \ldots + k_n) \).

Exact stochastic simulation of coupled chemical reactions

DT Gillespie - The journal of physical chemistry, 1977 - ACS Publications

There are two formalisms for mathematically describing the time behavior of a spatially homogeneous chemical system: The deterministic approach regards the time evolution as a continuous, wholly predictable process which is governed by a set of coupled, ordinary …

☆ 599  Cited by 10346  Related articles  All 47 versions

© 2021 Phillip Compeau
TOWARD A COMPLETE MODEL OF THE BACTERIAL CELL
Once we have a model of chemotaxis, what to do with it?

We could model every process for a very simple bacterium (*M. genitalium*, only 525 genes).
Once we have a model of chemotaxis, what to do with it?

We could model every process for a very simple bacterium (*M. genitalium*, only 525 genes).

Then build a “super-model” that links up these smaller models into a model of the cell.

---

A whole-cell computational model predicts phenotype from genotype

JR Karr, JC Sanghvi, DN Macklin, MV Gutschow... - Cell, 2012 - Elsevier

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental …

🌟 Cited by 1249  Related articles  All 39 versions
Remember replication?

The whole cell model showed that the lengths of *initiation* and *replication* are inversely correlated. In other words, the length of replication is *robust* to small stochastic changes in the cell.
Remember replication?

Why? If initiation of replication is slow, the cell builds up a larger surplus of dNTP molecule used by DNA polymerase during replication.
Doing biological research with a computational model

Key point: This was a new biological observation made by a purely computational model that was outside known research at the time.
And yet biology remains difficult ...

**Key point:** This was a *new biological* observation made by a purely *computational* model that was outside known research at the time.

Unfortunately, no model of this sophistication for *E. coli* has been published.

*A whole-cell computational model predicts phenotype from genotype*

**JR Karr, JC Sanghvi, DN Macklin, MV Gutschow...**  *Cell, 2012*  Elsevier

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental ...