Problems with Reaction Rate Equations

- Chemical reaction network model can be transformed using law of mass action into ODEs known as reaction rate equations.
- Assume concentrations vary continuously and deterministically.
- Chemical systems satisfy neither of these assumptions.
- Number of molecules of a species is a discrete quantity.
- Chemical reactions occur after two molecules collide.
- Unless track exact position and velocity of every molecule, not possible to know when a reaction may occur.
- Should consider occurrence of reactions to be stochastic.
- For systems which involve large molecular counts, ODE models give accurate picture of their behavior.

Stochastic Process Description

- If molecular counts are small, discrete and stochastic nature may have significant influence on observed behavior.
- Genetic circuits typically involve small molecule counts.
- Often only one strand of DNA and a few 10s or 100s of molecules of each transcription factor.
- Accurate analysis requires a stochastic process description.
- This lecture presents one such description, the chemical master equation, and algorithms to analyze it.

A Stochastic Chemical Kinetic Model

- Composed of $n$ chemical species $\{S_1, \ldots, S_n\}$ and $m$ chemical reaction channels $\{R_1, \ldots, R_m\}$.
- Assume species contained within constant volume $\Omega$.
- Assume system is well-stirred to neglect spatial effects.
- Assume system is in thermal equilibrium, but not necessarily chemical equilibrium.
State Updates

- $X_i(t)$ is the number of molecules of $S_i$ at time $t$.
- $X(t) = (X_1(t), \ldots, X_n(t))$ is state of system at time $t$.
- $X(t_0) = x_0$ is initial number of molecules at initial time $t_0$.
- After $R_{\mu}$, the new state found by $x' = x + v_{\mu}$.
- $v_{\mu} = (v_{\mu1}, \ldots, v_{\mu n})$ is the state-change vector where $v_{i\mu}$ is change in $S_i$ due to $R_{\mu}$.
- $\{v_{\mu}\}$ is the stoichiometric matrix.
- $R_{\mu}$ is elemental if it can be considered a distinct physical event that happens nearly instantaneously.
- For elemental $R_{\mu}$, values of $v_{i\mu}$ are constrained to $0, \pm 1, \pm 2$.

A Bimolecular Reaction Channel

- A typical bimolecular reaction channel $R_{\mu}$ has form:
  
  \[ S_1 + S_2 \xrightarrow{c_{\mu}} S_3 + \ldots \]

- $c_{\mu}$ is probability that a $S_1$ molecule and $S_2$ molecule collide and react within next $dt$ time units.
- Assume molecules hard spheres with masses $m_i$ and radii $r_i$.
- Thermal equilibrium means that a selected $S_i$ can be found uniformly distributed within $\Omega$.
- Also means avg. relative speed in which $S_2$ sees $S_1$ moving is:
  
  \[ \nu_{12} = \sqrt{8k_BT/\pi m_{12}} \]

  where $k_B$ is Boltzmann’s constant & $m_{12} = m_1m_2/(m_1 + m_2)$.

Specific Probability Rate Constant

- Every $R_{\mu}$ has a specific probability rate constant, $c_{\mu}$, which is related to the reaction rate constant, $k_{\mu}$.
- $c_{\mu}dt$ is the probability that a randomly chosen combination of reactant molecules react as defined by $R_{\mu}$ inside $\Omega$ in $[t, t + dt]$.
- Multiplying $c_{\mu}$ by the number of possible combinations of reactant molecules for $R_{\mu}$ in a state $x$ yields the propensity function, $a_{\mu}$.
- $a_{\mu}(x)dt$ is the probability that $R_{\mu}$ occurs in the state $x$ within $\Omega$ in the next infinitesimal time interval $[t, t + dt]$.

A Bimolecular Reaction Channel (cont)

- In next $dt$, $S_2$ molecule sweeps a collision cylinder relative to $S_1$ which has a height $\nu_{12}dt$ and base area $\pi(r_1 + r_2)^2$.
- Probability that $S_1$ is within the collision cylinder is ratio of its volume to $\Omega$, so $c_{\mu}$ is:
  
  \[ c_{\mu} = \Omega^{-1}\pi(r_1 + r_2)^2\nu_{12}p_{\mu} \]

  where $p_{\mu}$ is probability that $S_1$ and $S_2$ react when they collide.
- If assume that $S_1$ and $S_2$ react only when their kinetic energy exceeds the activation energy, $\epsilon_{\mu}$, then $c_{\mu}$ is:
  
  \[ c_{\mu} = \Omega^{-1}\pi(r_1 + r_2)^2\left(\frac{8k_BT}{\pi m^*}\right)^{1/2} \exp(-\epsilon_{\mu}/k_BT). \]
A Bimolecular Reaction Channel (cont)

- Number of combinations of $S_1$ and $S_2$ molecules is $x_1x_2$, so propensity function for $R_\mu$ is $a_\mu(x) = c_\mu x_1x_2$.
- If $S_1 = S_2$ then number of combinations is $x_1(x_1 - 1)/2$, and $a_\mu(x) = c_\mu x_1(x_1 - 1)/2$.

Monomolecular Reactions

- *Monomolecular* reactions are of this form:
  \[ S_1 \xrightarrow{c_\mu} S_2 + \ldots \]
- $S_1$ makes a spontaneous change in its internal structure.
- $c_\mu$ must be found from quantum mechanical considerations.
- Propensity function is simply $a_\mu(x) = c_\mu x_1$.
- If it is actually an enzymatic reaction of the form:
  \[ E + S_1 \xrightarrow{c_\mu} E + S_2 + \ldots \]
  where $E$ is an enzyme, should be considered as a bimolecular reaction.

Trimolecular Reactions

- *Trimolecular* reactions are of this form:
  \[ S_1 + S_2 + S_3 \xrightarrow{c_\mu} S_4 + \ldots \]
- Probability is very small, typically used as approximation for:
  \[ S_1 + S_2 \xrightarrow{c_3}{S^*} \text{ and } S^* + S_3 \xrightarrow{c_\mu} S_4 + \ldots \]
- Approximation reasonable when the lifetime of $S^*$ is very short (i.e., $1/c_2$ is very small).
- Probability that a molecule of $S^*$ reacts with a randomly chosen molecule of $S_3$ is approximately $c_3(1/c_2)$.

Trimolecular Reactions (cont)

- Consider a small but finite time interval $\Delta t$ which is still much larger than the lifetime of $S^*$ (i.e., $\Delta t >> 1/c_2$).
- If $\Delta t$ is sufficiently small, then the probability that $S_1$ and $S_2$ react in that time interval to form $S^*$ is $c_1\Delta t$.
- Probability of both reactions occurring in $\Delta t$ is $(c_1c_3/c_2)\Delta t$, so $c_\mu$ for the trimolecular reaction approximation is:
  \[ c_\mu = c_1c_3/c_2 \]
- Approximation because $\Delta t$ is not a true infinitesimal.
- Propensity function for this reaction is $a_\mu(x) = c_\mu x_1x_2x_3$. 
Relationship Between $c_\mu$ and $k_\mu$

- For bimolecular reactions, $c_\mu$ is proportional to $\Omega^{-1}$.
- For monomolecular reactions, it is independent of volume.
- For trimolecular reactions, it is proportional to $\Omega^{-2}$.
- In general, if $m$ is the number of reactant molecules in $R_\mu$:
  
  $c_\mu \propto \Omega^{-(m-1)}$

- Key to understanding relationship between $c_\mu$ and $k_\mu$.
- For monomolecular reactions, $c_\mu$ is equal to $k_\mu$.
- For bimolecular reactions, $c_\mu$ is equal to $k_\mu/\Omega$ if the reactants are different species and $2k_\mu/\Omega$ if the same species.

Time Evolution of Probability

- Not possible to know the exact state $X(t)$.
- Only can know probability of being in a given state at time $t$ starting from a state $X(t_0) = x_0$ (i.e., $P(x, t|x_0, t_0)$).
- Probability using a time-evolution of step $dt$ is shown below:
  
  $P(x, t + dt|x_0, t_0) = P(x, t|x_0, t_0) \times \left[ 1 - \sum_{j=1}^{m} (a_j(x) dt) \right]$
  
  $+ \sum_{j=1}^{m} P(x - v_j, t|x_0, t_0) \times (a_j(x - v_j) dt).$

- $dt$ chosen small enough that at most one reaction occurs during this time period.

Jump Markov Processes

- Stochastic model is a jump Markov process.

- A Markov process is one where the next state is only dependent on the present state and not the past history.

- A jump Markov process is one in which the state updates occur in discrete amounts.

Chemical Master Equation

- Chemical master equation (CME) defines time evolution of state probabilities, $P(x, t|x_0, t_0)$:
  
  $\frac{\partial P(x, t|x_0, t_0)}{dt} = \lim_{dt \to 0} \frac{P(x, t + dt|x_0, t_0) - P(x, t|x_0, t_0)}{dt}$

  $= \sum_{j=1}^{m} [a_j(x) P(x - v_j, t|x_0, t_0) - a_j(x) P(x, t|x_0, t_0)]$

- Typically cannot be solved analytically since it represents a set of equations as large as number of molecules in the system.
Stochastic Simulation

- Trajectories for $X(t)$ can be generated using stochastic simulation.
- Could pick a small time step $dt$ and at each step update the system state by selecting a reaction to occur or doing nothing.
- For a sufficiently small $dt$, however, the vast majority of time steps result in no reaction.

Gillespie’s Stochastic Simulation Algorithm

- Gillespie’s stochastic simulation algorithm (SSA) improves the efficiency of simulation by stepping over useless time steps.
- Not based directly on CME, but rather equivalent formulation that uses $p(\tau, \mu|x, t)$.
- Defined such that $p(\tau, \mu|x, t)d\tau$ is probability that the next reaction is $R_\mu$ which occurs in $[t + \tau, t + \tau + d\tau]$ assuming current state is $X(t) = x$.
- This is a joint PDF for two random variables, $\tau$ and $\mu$ given that the system is in state $x$ at time $t$.
- Simulation advances from one reaction to the next skipping over time points in which no reaction occurs.

Derivation of Gillespie’s SSA

- Introduce $P_0(\tau|x, t)$ that represents probability that there is no reaction in the time interval $[t, t + \tau]$.
- $p(\tau, \mu|x, t)$ defined as follows:
  \[
p(\tau, \mu|x, t)d\tau = P_0(\tau|x, t) \times (a_\mu(x)d\tau).
  \]
- No reactions occur in the interval $[t, t + \tau]$ and the $R_\mu$ reaction occurs in the interval $[t + \tau, t + \tau + d\tau]$.

Derivation of Gillespie’s SSA (cont)

- The function $P_0(\tau|x, t)$ must satisfy the following:
  \[
P_0(\tau + d\tau|x, t) = P_0(\tau|x, t) \times \left[ 1 - \sum_{j=1}^{m} (a_j(x)d\tau) \right].
  \]
- Using this formula, get following differential equation:
  \[
  \frac{dP_0(\tau|x, t)}{d\tau} = -a_0(x)P_0(\tau|x, t) \quad \text{where} \quad a_0(x) = \sum_{j=1}^{m} a_j(x).
  \]
- With $P_0(\tau = 0|x, t) = 1$, has following solution:
  \[
  P_0(\tau|x, t) = \exp(-a_0(x)\tau).
  \]
Derivation of Gillespie’s SSA (cont)

- Inserting Equation 2 into Equation 1 and canceling $dt$ yields:
  
  $$ p(\tau, \mu|\mathbf{x}, t) = \exp(-a_0(\mathbf{x})\tau) \times a_\mu(\mathbf{x}), $$

  which can be rewritten as:
  
  $$ p(\tau, \mu|\mathbf{x}, t) = a_0(\mathbf{x})\exp(-a_0(\mathbf{x})\tau) \times \frac{a_\mu(\mathbf{x})}{a_0(\mathbf{x})}. $$

- $p(\tau, \mu|\mathbf{x}, t)$ can be divided into a PDFs for $\tau$ and $\mu$.
- $\tau$ is exponential RV with mean and std dev of $\frac{1}{a_0(\mathbf{x})}$.
- $\mu$ is integer random variable with point probabilities $\frac{a_\mu(\mathbf{x})}{a_0(\mathbf{x})}$.

Gillespie’s SSA

1. Initialize: $t = t_0$ and $\mathbf{x} = \mathbf{x}_0$.
2. Evaluate $a_j(\mathbf{x})$ and $a_0(\mathbf{x}) = \sum_{j=1}^{m} a_j(\mathbf{x})$.
3. Draw two unit uniform random numbers, $r_1$ and $r_2$.
4. Determine the time, $\tau$, until the next reaction:
   
   $$ \tau = \frac{1}{a_0(\mathbf{x})} \ln \left( \frac{1}{r_1} \right). $$

5. Determine the next reaction, $\mu$:
   
   $$ \mu = \text{the smallest integer satisfying } \sum_{j=1}^{\mu} a_j(\mathbf{x}) > r_2a_0(\mathbf{x}). $$

6. Determine the new state: $t = t + \tau$ and $\mathbf{x} = \mathbf{x} + v_\mu$.
7. If $t$ is greater than the desired simulation time then halt.
8. Record $(\mathbf{x}, t)$ and goto step 2.

Simulation of $P_{RE}$ Promoter

$(P_{RE} = 1$, $RNAP = 30$, and $cII = 30)$

Discussion

- Using SSA to compute a single trajectory is no more complex than numerical simulation of RREs.
- Provides a closer approximation of molecular reality for systems with small molecule counts such as genetic circuits.
- Unfortunately, SSA has a substantial computational cost:
  - Must be run many times (1000s) to produce reasonable statistics while simulations of RREs only run once.
  - Very slow since $\tau$ is equal to $1/a_0(\mathbf{x})$ and can be very large when any molecule counts become large.
- When molecule counts increase, relative difference between deterministic and stochastic trajectories decrease.
Simulation of $P_{RE}$ Promoter
($P_{RE} = 100$, $RNAP = 3000$, and $cII = 3000$)

Gillespie’s First Reaction Method
1. Initialize: $t = t_0$ and $x = x_0$.
2. Evaluate propensity functions $a_j(x)$ at state $x$.
3. For each $j$, determine the time, $\tau_j$, until the next $R_j$ reaction:
   $$\tau = \frac{1}{a_j(x)} \ln \left( \frac{1}{\tau_j} \right).$$
   where each $\tau_j$ is a unit uniform random number.
4. Let $\mu$ be the reaction whose $\tau_\mu$ is the smallest.
5. Let $\tau$ equal $\tau_\mu$.
6. Determine the new state: $t = t + \tau$ and $x = x + v_\mu$.
7. If $t$ is greater than the desired simulation time then halt.
8. Record $(x, t)$ and goto step 2.

Observations
- FRM requires $m$ random variables per simulation!
- OBSERVATION: not all propensities change after a reaction.
- Following three steps are taken during every iteration and take a time proportional to the number of reactions, $m$.
  1. Update all $m$ propensity functions, $a_j(x)$.
  2. Generate $m$ random numbers and next reaction times.
  3. Find the smallest reaction time, $\tau_\mu$.
- Must eliminate each of these performance bottlenecks.

Gibson/Bruck’s Improvements
- $\tau_j$ and $a_j(x)$ stored for use in future iterations.
- $\tau_j$ use absolute time to make useful for multiple iterations.
- Dependency graph used to indicate relations between reactions.
  - Has vertex for each $R_j$ and edge from $R_j$ to other reaction that has as reactant either reactant or product of $R_j$.
- Reuse every $\tau_j$ except the one for $\tau_\mu$, renormalizing $\tau_j$ when its propensity has changed.
- Indexed priority queue used to organize $a_j(x)$ and $\tau_j$ data to make easy to update and to find smallest entry.
  - An indexed priority queue is a tree structure in which the parent always has a lower $\tau_j$ value than both its children.
  - This means the top node always has the smallest $\tau_j$ value.
  - Can be updated in $O(\log(m))$ time.
Gibson/Bruck’s Next Reaction Method
1. Initialize:
   (a) \( t = t_0 \) and \( x = x_0 \).
   (b) Generate a dependency graph, \( G \).
   (c) Evaluate propensity functions \( a_j(x) \) at state \( x \).
   (d) For each \( j \), determine time, \( \tau_j \), until next \( R_j \) reaction:
      \[
      \tau = t + \frac{1}{a_j(x)} \ln \left( \frac{1}{r_j} \right).
      \]
      where each \( r_j \) is a unit uniform random number.
   (e) Store the \( \tau_j \) values in an indexed priority queue \( Q \).

Gibson/Bruck’s Next Reaction Method (cont)
1. Let \( \mu \) be the reaction whose \( \tau_\mu \) is the smallest stored in \( Q \).
2. Let \( \tau \) equal \( \tau_\mu \).
3. Determine the new state: \( t = \tau \) and \( x = x + v_\mu \).
4. For each edge \((\mu, \alpha)\) in the dependency graph \( G \),
   (a) Update \( a_\alpha \).
   (b) If \( \alpha \neq \mu \), set \( \tau_\alpha = (a_{\alpha,old}/a_{\alpha,new})(\tau_\alpha - t) + t \)
   (c) If \( \alpha = \mu \), generate a random number, \( r_\mu \), and
   \[
   \tau_\mu = t + \frac{1}{a_\mu(x)} \ln \left( \frac{1}{r_\mu} \right).
   \]
5. If \( t \) is greater than the desired simulation time then halt.
6. Record \((x, t)\) and goto step 2.

Tau Leaping
- NRM still simulates every reaction event one at a time which is not practical for many interesting systems.
- **Tau-leaping** gives up exactness to improve simulation speed.
- Many reactions are fired at once in the time interval \([t, t + \tau]\).
- Introduce \( m \) random functions, \( K_j(\tau; x, t) \), where each returns number of times that \( R_j \) fires in \([t, t + \tau]\) in state \( X(t) = x \).
- New state after \( \tau \)-leap is:
  \[
  X(t + \tau) = x + \sum_{j=1}^{m} K_j(\tau, x, t)v_j.
  \]

Tau Leaping (cont)
- Unfortunately, these functions are dependent on each other.
- Number of \( R_j \) reactions depends on \( a_j(x) \) which depends on \( x \) which depends on number of all other reactions.
- Even if joint PDF can be computed, likely as expensive as doing the full simulation.
Leap Condition

- States that $\tau$ be chosen to be small enough such that no propensity function changes by a significant amount.
- If satisfied, $K_j(\tau, x, t)$ can be approximated to be a statistically independent Poisson random variable:

$$K_j(\tau, x, t) \approx \mathcal{P}_j(a_j(x), \tau) \quad (j = 1, \ldots, m)$$

where $\mathcal{P}_j(a_j(x), \tau)$ returns the number of events $k$ in the interval $[t, t + \tau]$ such that:

$$P[k \text{ events}] = \frac{e^{-a_j(x)\tau} (a_j(x)\tau)^k}{k!}$$

- How to find $\tau$ small enough to satisfy leap condition, but large enough to fire enough events to speedup simulation?

Determineing Tau

- While it is still an open area of research, one method for selecting $\tau$ proposed by Gillespie and Petzold uses Equation 3.
- Goal is to ensure that no propensity changes by more than $\epsilon a_0(x)$, where $0 < \epsilon << 1$.

$$\tau = \min_{j\in[1,m]} \left\{ \frac{\epsilon a_0(x)}{|\mu_j(x)|}, \frac{\epsilon^2 a_0^2(x)}{\sigma_j^2(x)} \right\}$$

(3)

$$\mu_j(x) = \sum_{j' = 1}^m f_{j'j}(x)a_{j'}(x) \quad (j = 1, \ldots, m)$$

$$\sigma_j^2(x) = \sum_{j' = 1}^m f_{j'j}^2(x)a_{j'}(x) \quad (j = 1, \ldots, m)$$

$$f_{j'j}(x) = \sum_{i=1}^n \frac{\partial a_j(x)}{\partial x_i} v_{ij'} \quad (j, j' = 1, \ldots, m).$$

Tau Leaping Algorithm

1. Initialize: $t = t_0$ and $x = x_0$.
2. Evaluate $a_j(x)$ and $a_0(x) = \sum_{j=1}^m a_j(x)$.
3. Determine value for $\tau$ using Equation 3 for a chosen $\epsilon$.
4. For $j = 1, \ldots, m$, determine the number of firings, $k_j$ of $R_j$ in time $\tau$ using Poisson random variable $\mathcal{P}_j(a_j(x), \tau)$.
5. Leap, by replacing $t = t + \tau$ and $x = x + \sum_{j=1}^m k_j v_j$.
6. If $t$ is greater than the desired simulation time then halt.
7. Record $(x, t)$ and goto step 2.

Discussion

- $\epsilon$ provides means of trading off accuracy for runtime.
- For large $\epsilon$, a significant runtime improvement can be achieved at the cost of some accuracy.
- Care has be taken though as large jumps can cause bad things such as species counts being made negative.
- As $\epsilon$ is made smaller, tau-leaping gradually reduces to the SSA.
- For a very small $\epsilon$, not as efficient as SSA as it takes many $\tau$ leaps that produce no events.
- If $\tau$ much less than a few multiples of $1/a_0(x)$, revert to SSA.
The Chemical Langevin Equation

- If $\Delta t = \tau$ is small enough that no $a_j(x)$ changes significantly,
  \[ X(t + \Delta t) \approx x + \sum_{j=1}^{m} P_j(a_j(x), \Delta t)v_j. \]

- If $\Delta t$ is large enough that there are many firings of each $R_j$, Poisson can be approximated with a Normal random variable:
  \[ X(t + \Delta t) \approx x + \sum_{j=1}^{m} \mathcal{N}(a_j(x)\Delta t, a_j(x)\Delta t)v_j \]
  \[ = x + \sum_{j=1}^{m} v_j a_j(x) \Delta t + \sum_{j=1}^{m} v_j \sqrt{a_j(x)\mathcal{N}_j(0,1)} \sqrt{\Delta t} \]
  using $\mathcal{N}(m, \sigma^2) = m + \sigma \mathcal{N}(0,1)$.

The Reaction Rate Equation

- CLE has two parts:
  - A deterministic part that grows linearly with $a_j(x)$.
  - A stochastic part that grows proportional to $\sqrt{a_j(x)}$.
- $a_j(x)$ grow in direct proportion system size.
- Stochastic part scales relative to deterministic part as the inverse square root of the system size.

The Chemical Langevin Equation (cont)

- If $\Delta t$ is a macroscopically infinitesimal time increment $dt$,
  \[ X(t + dt) \approx X(t) + \sum_{j=1}^{m} v_j a_j(X(t)) dt \]
  \[ + \sum_{j=1}^{m} v_j \sqrt{a_j(X(t)) \mathcal{N}_j(t)} \sqrt{dt} \]
  where $\mathcal{N}_j(t)$ are $m$ statistically independent and temporally uncorrelated Normal RVs w/mean 0 and variance 1.
- This equation is the chemical Langevin equation.

The Reaction Rate Equation (cont)

- As size increases, magnitude of fluctuations diminish until CLE can be reduced to:
  \[ X(t + dt) \approx X(t) + \sum_{j=1}^{m} v_j a_j(X(t)) dt \]
  - This is simply the reaction rate equation, but it has been derived from stochastic chemical kinetics.
  - RREs valid when system large enough such that no propensity changes significantly in $dt$ and every $R_j$ fires many within $dt$. 
**Stochastic Petri Nets**

- One interesting model that has been applied to biological systems is *stochastic Petri nets* (SPNs).
- SPNs are a graphical representation which is quite similar to representations used in biochemistry.
- SPNs are also isomorphic to jump Markov processes.
- Using SPNs has the advantage that several tools have been developed for their analysis.

**Stochastic Petri Net Model**

- Composed of:
  - A set of places $P$ (molecular species),
  - A set of transitions $T$ (reactions),
  - An input function $I$ (stoichiometry of reactants),
  - An output function $O$ (stoichiometry of products),
  - A weight function $W$ (rate of reaction), and
  - An initial marking $M_0$ (initial molecule counts).

- For elementary reactions, transition labeled with rate constant and assumed that rate function includes product of reactants.
- The state of an SPN is its marking which is an assignment of a number of tokens to each place in the net.
- Corresponds to current molecule counts.

**Simple Stochastic Petri Net**

- Formal description of the simple stochastic petri net.

**SPN for Part of the $P_{RE}$ w/ $cI$ Dimer. and Deg.**

- Diagram representing a portion of the $P_{RE}$ with $cI$ dimer and degradation processes.

---

*Slide 41*

*Slide 42*

*Slide 43*

*Slide 44*
Shea/Ackers Statistical Thermodynamical Model

- Stochastic sim. time depends on number of reaction events.
- Models with many fast reactions simulate very slowly.
- Phage λ model has many reactions involved in binding and unbinding of transcription factors to operator sites.
- These are typically very fast in relation to other rates.
- Can use Shea/Ackers statistical thermodynamical model.
- Assumes occupancy of operator sites can be determined by equilibrium statistical thermodynamic probabilities.
- Probability of each potential configuration of transcription factors and RNAP bound to operator sites can be determined.
- Does not include reactions for operator site binding, but determines configuration during each simulation cycle.

Assumptions

- Occupancy determined by equilibrium statistical thermodynamic probabilities.
- Repressor bound to adjacent operator sites interact.
- Cooperative interaction between $O_{R2}$ and $O_{R3}$ only happens when $O_{R1}$ is vacant.
- $P_R$, cro gene, is off when $O_{R1}$ or $O_{R2}$ are occupied.
- $P_{RM}$, cl gene, is off when $O_{R3}$ is occupied.
- In mutants with one operator damaged, others work the same.
Free Energies for Configurations

<table>
<thead>
<tr>
<th>s</th>
<th>$O_R3$</th>
<th>$O_R2$</th>
<th>$O_R1$</th>
<th>Free energy contributions</th>
<th>$\Delta G_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>Reference</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>O</td>
<td>O</td>
<td>R</td>
<td>$\Delta G_1$</td>
<td>-11.7</td>
</tr>
<tr>
<td>3</td>
<td>O</td>
<td>R</td>
<td>O</td>
<td>$\Delta G_2$</td>
<td>-10.1</td>
</tr>
<tr>
<td>4</td>
<td>R</td>
<td>O</td>
<td>O</td>
<td>$\Delta G_3$</td>
<td>-10.1</td>
</tr>
<tr>
<td>5</td>
<td>O</td>
<td>R</td>
<td>R</td>
<td>$\Delta G_1 + \Delta G_2 + \Delta G_{12}$</td>
<td>-23.8</td>
</tr>
<tr>
<td>6</td>
<td>R</td>
<td>O</td>
<td>R</td>
<td>$\Delta G_1 + \Delta G_3$</td>
<td>-21.8</td>
</tr>
<tr>
<td>7</td>
<td>R</td>
<td>R</td>
<td>O</td>
<td>$\Delta G_2 + \Delta G_3 + \Delta G_{23}$</td>
<td>-22.2</td>
</tr>
<tr>
<td>8</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>$\Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_{12}$</td>
<td>-33.9</td>
</tr>
</tbody>
</table>

$\Delta G_i = -RT \ln K_i$.

Values for $[cI2]$ for Half Occupation (units of 3nM)

<table>
<thead>
<tr>
<th>DNA</th>
<th>Template</th>
<th>$O_R3$</th>
<th>$O_R2$</th>
<th>$O_R1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>$O_R^R$ (wild type)</td>
<td>25</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>DNA</td>
<td>$O_R1^+$</td>
<td>5</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>DNA</td>
<td>$O_R2^-$</td>
<td>25</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>DNA</td>
<td>$O_R1^+, O_R2^-$</td>
<td>25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DNA</td>
<td>$O_R1^-, O_R3^-$</td>
<td>-</td>
<td>25</td>
<td>-</td>
</tr>
<tr>
<td>DNA</td>
<td>$O_R3^-$</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Mathematical Relationships of the Model

\[
\begin{align*}
  f_s &= \frac{\exp(-\Delta G_s/RT)[cI2]^j}{\sum_j \exp(-\Delta G_j/RT)[cI2]^j} \\
  f_{O_R1} &= f_2 + f_5 + f_6 + f_8 \\
  f_{O_R2} &= f_3 + f_5 + f_7 + f_8 \\
  f_{O_R3} &= f_4 + f_6 + f_7 + f_8 = f_{P_R} \\
  f_{P_R} &= f_2 + f_3 + f_5 + f_6 + f_7 + f_8 \\
  [cI] &= [cI] + 2[cI2] + 2[O_J] \sum f_s \\
  [cI2] &= K_u[cI]^2
\end{align*}
\]

Resolved Interaction Free Energies for $O_R$

<table>
<thead>
<tr>
<th>Energy, kcal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta G_1$</td>
</tr>
<tr>
<td>$\Delta G_2$</td>
</tr>
<tr>
<td>$\Delta G_3$</td>
</tr>
</tbody>
</table>

Cooperative interaction

| $\Delta G_{12}$ | $-1.99 \pm 0.06$ |
| $\Delta G_{23}$ | $-1.94 \pm 0.06$ |
**Phage λ Model: Motivation**

- One of the first stochastic models of a biological network was Arkin/Ross/McAdams’s phage λ model.
- Earlier stochastic model for gene expression found fluctuations could produce erratic pattern for protein production.
- Interesting case when two independently produced regulatory proteins compete for a switch.
- Result is switch behavior is nondeterministic.
- Initially homogeneous population can follow different pathways with different phenotypes.
- Two *E. Coli* in same environment and infected with the same number of phages, one may lyse while other is lysogenized.
- Random variations can enhance virulence.

**Lysis versus Lysogeny**

- Goal is to predict probability of lysogeny.
- Shown experimentally to depend on *multiplicity of infection* (MOI), and on nutritional state of the cell.
- Well-fed cells tend to go into lysis.
  - Higher Hfl-related proteolytic activity shortens lifetimes of cII and cIII.
- Cells with higher MOI tend to lysogeny.
  - Hfl concentration is constant in MOI.
  - cII and cIII genes are proportional to MOI.
- The model developed should reproduce these effects.

**Discussion**

- 25-fold more repressor is needed to half repress $P_{RM}$ than $P_R$.
- Shape is different because $P_R$ controlled by two operator sites to which repressor binds cooperatively.
- In a lysogen, $P_R$ is tightly repressed while $P_{RM}$ is down $\approx 20\%$.
- Cooperativity maintains stable lysogen, yet allows induction.
- See Shea/Ackers 1985 for much more detail.
Arkin/Ross/McAdam’s Phage $\lambda$ Model

- Developed a chemical reaction network that they simulated with Gillespie’s SSA.
- Simplified using statistical thermodynamic approach.
- Shea/Ackers found free energies for all 40 potential configurations of the $O_R$ operator sites.

Arkin/Ross/McAdam’s Phage $\lambda$ Model (cont)

- Using the literature, proposed free energies for 4 configurations for $P_{RE}$ and 10 configurations for $O_L$.
- At each step of the simulation, determine probability of each configuration for each set of operator sites.
- Using this distribution, select one configuration that the operator sites would be in for that simulation cycle.
- Result is their model does not include reactions for binding of transcription factors and RNAP to the operator sites.

Activation Contours for $P_R$ and $P_{RM}$

Phage $\lambda$ Decision Circuit
Some of the Assumptions

- Cell cycle is 35 minutes, and cell grows in a linear fashion during this time from $1 \times 10^{-15}$ to $2 \times 10^{-15}$ liters.
  \[
  \Omega(t) = (1 + k_0 \cdot t) \times 10^{-15} \text{ liters}
  \]
  where $k_0$ equals $4.76 \times 10^{-18}$ liters sec$^{-1}$.
- All housekeeping molecules needed for gene expression and protein degradation are at constant levels.
- Phage gene expression is stochastic.

Phage Gene Expression Models

- Movement of transcribing RNAP uses an exponential distribution of interstep times.
- At termination sites, RNAP slows down and has a probability of termination.
- RNAP can be anti-terminated at NUT sites resulting in RNAP moving through terminators.
- Result of transcription is an mRNA that must be go to a ribosome to be turned into a protein.
- Translation modeled as series of reactions with the ribosome moving along with exponentially distributed interstep times.
- The mRNA can also bind to RNase which degrades the mRNA.
- Rate constant selected such that an mRNA typically yields 10 proteins before it is degraded.

Some of the Assumptions (cont)

- Infection occurs early in the cell cycle.
  - Neglecting cell division may lead to overestimate of lysogenic fraction.
- Target cells are infected simultaneously.
- Cell is homogenous and well-stirred medium.
- Lysogeny results if:
  - Sufficient cII concentration to activate $P_{RE}$.
  - $[cI] > [cro2]$ at end of 35-min cell cycle.

Dimerization/Degradation Models

- cI and cro reaction rates selected to match measured lifetimes.
- Half-life of unprotected cII from 5m to 30s.
- cIII protects cII from proteolysis by either:
  - Competitive binding with Hfl.
  - Or, direct binding with cII.
- Best model uses proteolytic system in which they compete for two independent proteases.
Kourilsky’s Measurements

- Measured lysogenic fraction vs. API.
- Included measurements of $O^-$ and $P^-$ strains incapable of phage chromosome replication.
- Shape of curves same for starved vs. unstarved cells with starved shifted 50-100 times higher.
- Arkin used starved data as number of simulation runs goes like $1/f$ where $f$ is fraction of lysogens.

Stochastic Kinetic Model

- Model includes genetic mechanisms and coupled protein dimerization/degradation.
- Model system with set of coupled stochastic equations and solve using Monte Carlo algorithm.
- $4(1-p)/f^2p$ runs with same initial conditions.
- Predict lysogenic fraction for various MOI, $F(M)$.
- Use Poisson distribution to obtain lysogenic fraction vs. API:

$$ P(M, A) = \frac{A^M}{M!} e^{-A} $$

$$ F_{\text{lyso}}(A) = \sum_{M} P(M, A) \cdot F(M) $$
Discussion

- Molecular level thermal fluctuations can produce different phenotypic outcomes.
- Many organisms with bistable switching phenomena.
- Switching dynamics of λ likely found elsewhere.

Spatial Gillespie

- As the volume increases, well-stirred assumption is less valid.
- Several methods proposed to add spatial considerations.
- Stundza and Lumsden proposed spatial Gillespie method.
- System divided into several discrete subvolumes.
- Size of subvolumes selected such that within them well-stirred assumption is reasonable.
- Diffusion within subvolume faster than rate of reactions.
Spatial Gillespie (cont)

- State is now number of each species within each subvolume.
- During simulation cycle, molecule either reacts with others within the subvolume or diffuses to adjacent subvolume.
- Beginning with \((S_1, \ldots, S_n)\), spatial considerations added as follows assuming volume divided into \(p \times q \times r\) subvolumes:
  \[
  (S_1^{(i,j,k)}, \ldots, S_n^{(i,j,k)})
  \]
  where \(i = 1, \ldots, p\), \(j = 1, \ldots, q\), and \(k = 1, \ldots, r\).
  
  \[
  S_\mu^{(i,j,k)} \overset{\mathcal{K}}{\rightarrow} S_\mu^{(i+1,j,k)} \\
  S_\mu^{(i,j,k)} \overset{\mathcal{K}}{\rightarrow} S_\mu^{(i,j+1,k)} \\
  S_\mu^{(i,j,k)} \overset{\mathcal{K}}{\rightarrow} S_\mu^{(i,j,k+1)}.
  \]

- Now any stochastic simulation algorithm can be used.

Sources

- Gibson and Bruck papers.
- Arkin’s lambda model in Genetic 1998.