

Stochastic Simulation of Biochemical Reactions

Jorge Júlvez
University of Zaragoza

Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation
- 4 Stochastic Simulation Algorithms
- 5 Comparing Models
- 6 Improved SSA
- 7 Stochastic Differential Equations

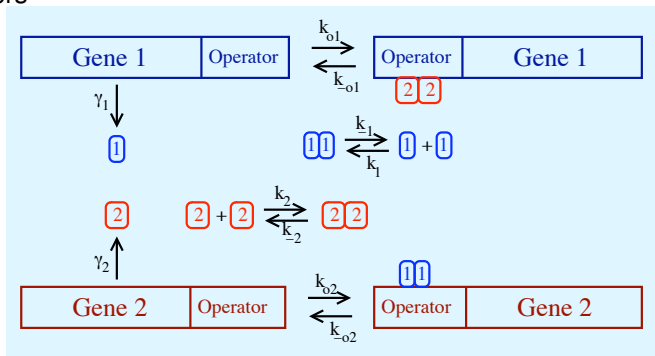
Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation
- 4 Stochastic Simulation Algorithms
- 5 Comparing Models
- 6 Improved SSA
- 7 Stochastic Differential Equations

Motivation: Gene regulatory network

Simple system of two mutually repressing genes

System described by states of genes (on/off) and protein molecule numbers¹

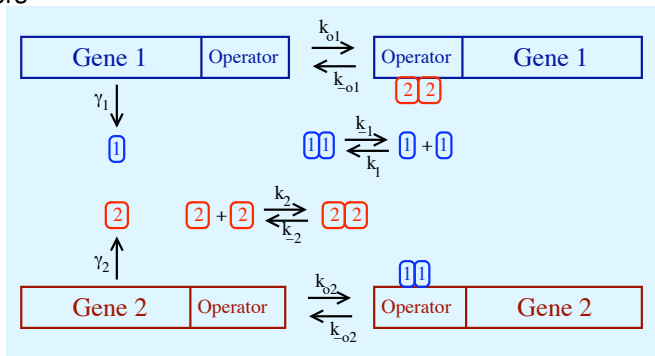


¹ T. Kepler and T. Elston, Biophysical Journal, 2001; T. Gardner et al., Nature, 2000.

Motivation: Gene regulatory network

Simple system of two mutually repressing genes

System described by states of genes (on/off) and protein molecule numbers¹



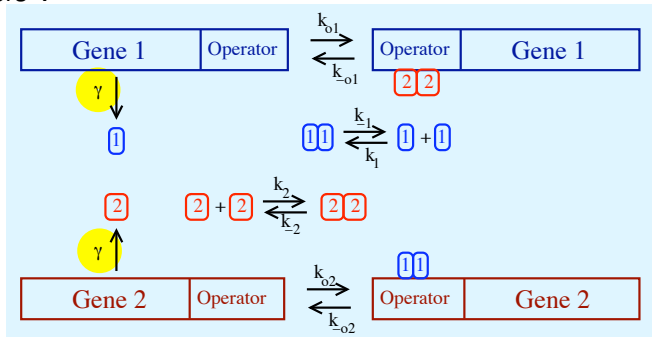
What is the time evolution of the system?

¹ T. Kepler and T. Elston, Biophysical Journal, 2001; T. Gardner et al., Nature, 2000.

Motivation: Gene regulatory network

Simple system of two mutually repressing genes

System described by states of genes (on/off) and protein molecule numbers².

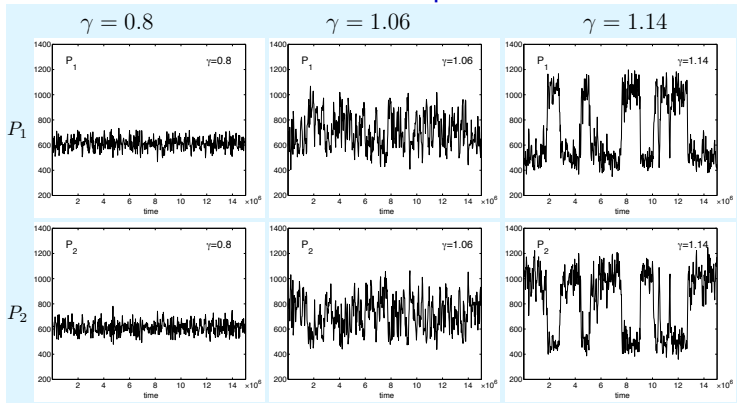


What is the time evolution of the system?

²T. Kepler and T. Elston, Biophysical Journal, 2001; T. Gardner et al., Nature, 2000

Motivation: Gene regulatory network

Evolution of the number of protein molecules



Observation: Switching between two favourable states of the system for $\gamma > 1.06$

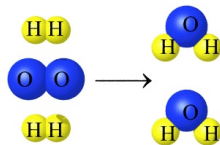
Law of Mass Action

- The modelling of chemical reactions using **deterministic** rate laws has proven extremely successful in both chemistry and biochemistry for many years.
- This deterministic approach has at its core the **law of mass action**: an empirical law giving a simple relation between reaction **rates** and molecular component concentrations.
- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.

Kinetics of biochemical reactions

Biochemical kinetics:

- It studies the **rate** at which a biochemical process occurs.
- It also sheds light on the reaction mechanism (exactly *how* the reaction occurs).



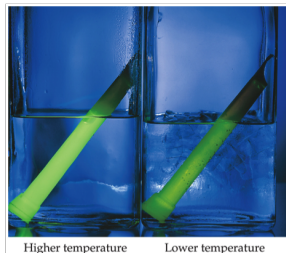
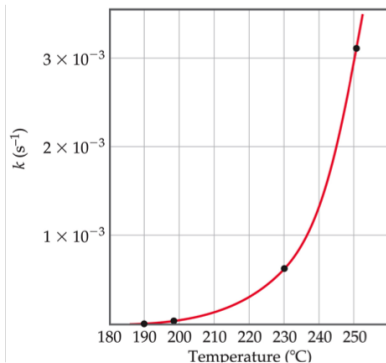
Factors that affect reaction rates

Biochemical kinetics:

- **Temperature.** At higher temperatures, reactant molecules have more kinetic energy, move faster, and collide more often and with greater energy.
- **Catalysts.** Speed reaction by changing mechanism. Enzymes (or biocatalysts) are proteins that catalyze (i.e., increase the rates of) chemical reactions
- **Concentration of reactants.** As the concentration of reactants increases, so does the likelihood that reactant molecules collide.

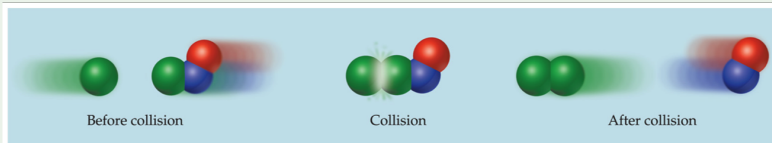


Temperature and Rate



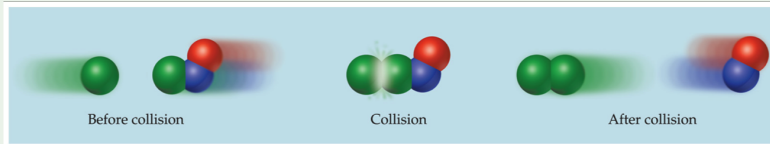
- Generally, as the temperature increases, so does the reaction rate.
- That is, the rate is temperature dependent.

The Collision Model



- In a chemical reaction, bonds are broken and new bonds are formed.
- Molecules can only react if they collide with each other.

The Collision Model



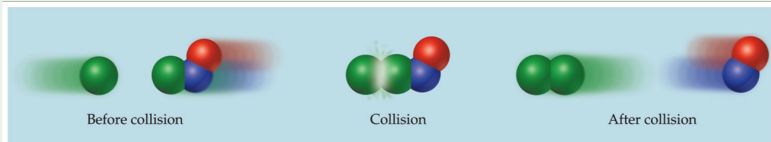
Furthermore, molecules must collide with the correct **orientation** and with enough **energy** to cause bond breakage and formation.

Activation Energy

- In other words, there is a minimum amount of energy required for reaction: the **activation energy**.
- Just as a ball cannot get over a hill if it does not roll up the hill with enough energy, a reaction cannot occur unless the molecules possess sufficient energy to get over the activation energy barrier.



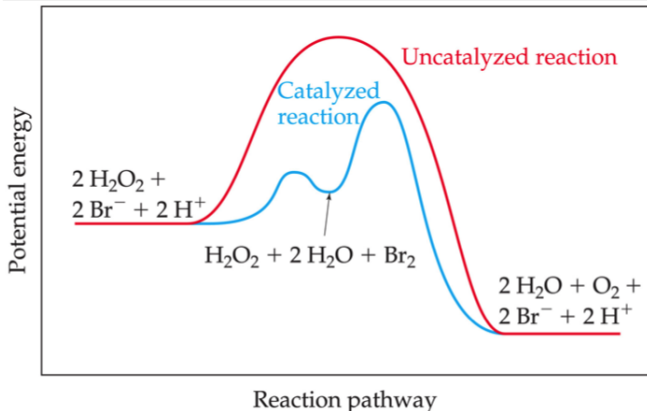
Activation Energy



- **Temperature** is defined as a measure of the average kinetic energy of the molecules in a sample.
- Thus at higher temperatures, a larger population of molecules has higher energy.
- As a result, the reaction rate increases.

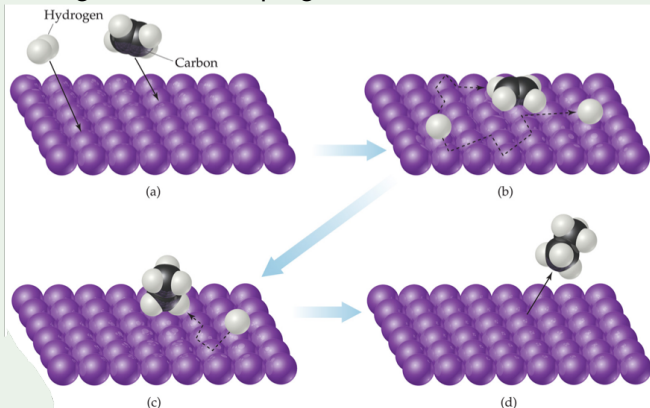
Catalysts

- Catalysts increase the rate of a reaction by decreasing the activation energy of the reaction.
- Catalysts change the mechanism by which the process occurs.



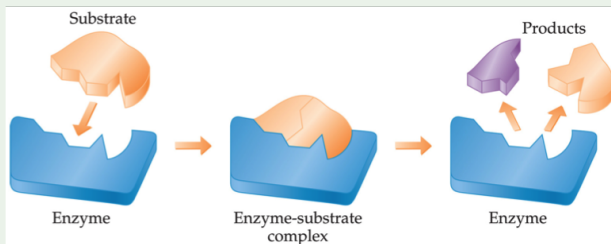
Catalysts

One way a catalyst can speed up a reaction is by holding the reactants together and helping bonds to break.

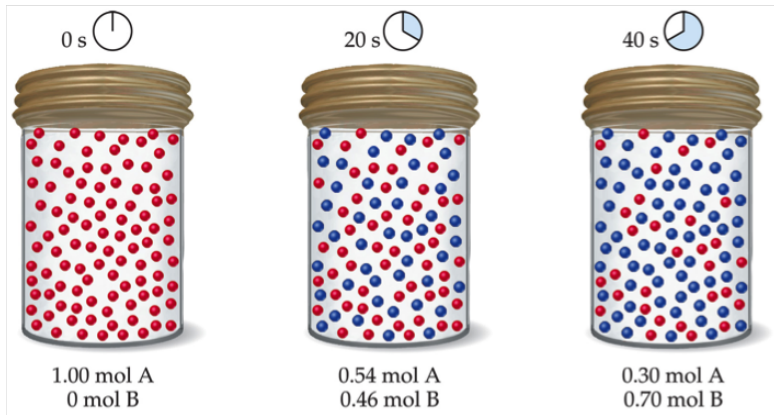


Enzymes

- Enzymes are catalysts in biological systems.
- The substrate fits into the active site of the enzyme much like a key fits into a lock.

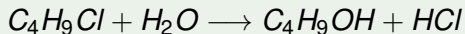


Reaction rates



Rates can be determined by monitoring the change in concentration of either reactants or products as a function of time.

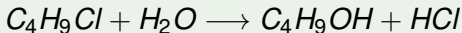
Reaction rates



Time, t (s)	$[C_4H_9Cl]$ (M)
0.0	0.1000
50.0	0.0905
100.0	0.0820
150.0	0.0741
200.0	0.0671
300.0	0.0549
400.0	0.0448
500.0	0.0368
800.0	0.0200
10,000	0

The concentration of butyl chloride C_4H_9Cl is measured at various times t .

Reaction rates

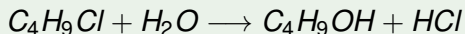


Time, $t(s)$	$[C_4H_9Cl] (M)$	Average Rate, M/s
0.0	0.1000	
50.0	0.0905	1.9×10^{-4}
100.0	0.0820	1.7×10^{-4}
150.0	0.0741	1.6×10^{-4}
200.0	0.0671	1.4×10^{-4}
300.0	0.0549	1.22×10^{-4}
400.0	0.0448	1.01×10^{-4}
500.0	0.0368	0.80×10^{-4}
800.0	0.0200	0.560×10^{-4}
10,000	0	

The average rate of the reaction over each interval is the change of concentration divided by the change in time.

$$average\ rate = \frac{\Delta[C_4H_9]}{\Delta t} = \frac{0.1000 - 0.0905\ M}{50.00 - 0.0\ s} = 1.9 \cdot 10^{-4} M/s$$

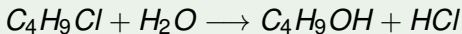
Reaction rates



Time, t (s)	$[C_4H_9Cl]$ (M)	Average Rate (M/s)
0.0	0.1000	
50.0	0.0905	1.9×10^{-4}
100.0	0.0820	1.7×10^{-4}
150.0	0.0741	1.6×10^{-4}
200.0	0.0671	1.4×10^{-4}
300.0	0.0549	1.22×10^{-4}
400.0	0.0448	1.01×10^{-4}
500.0	0.0368	0.80×10^{-4}
800.0	0.0200	0.560×10^{-4}
10,000	0	

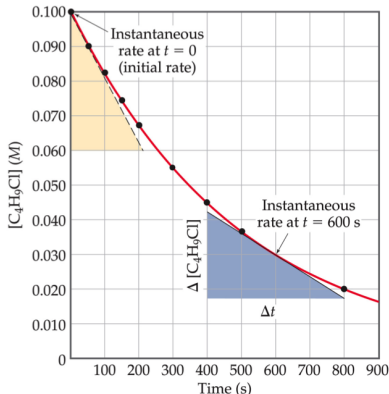
- The average rate decreases as the reaction proceeds.
- This is because as the reaction goes forward, there are fewer collisions among reactant molecules.

Reaction rates

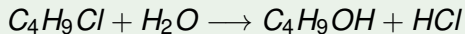


- A plot of concentration vs. time yields a curve like this.
- The slope of a line tangent to the curve at any point is the instantaneous rate at that time.

$$\frac{\Delta[A]}{\Delta t} \Rightarrow \frac{d[A]}{dt}$$

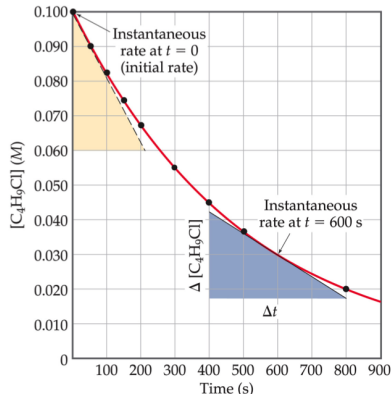


Reaction rates

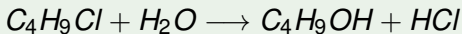


- The reaction slows down because the concentration of the reactants decreases.

$$\frac{\Delta[A]}{\Delta t} \Rightarrow \frac{d[A]}{dt}$$

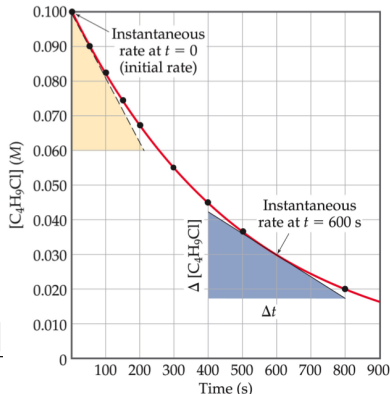


Reaction rates and Stoichiometry



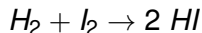
- In this reaction, the ratio of C_4H_9Cl to C_4H_9OH is 1 : 1
- Thus, the rate of disappearance of C_4H_9Cl is the same as the rate of appearance of C_4H_9OH .

$$Rate = \frac{-\Delta[C_4H_9Cl]}{\Delta t} = \frac{\Delta[C_4H_9OH]}{\Delta t}$$



Reaction rates and Stoichiometry

- What if the ratio is not 1 : 1?

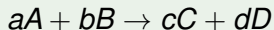


- Two HI are made for each H_2 used.

$$rate = -\frac{\Delta[H_2]}{\Delta t} = \frac{1}{2} \frac{\Delta[HI]}{\Delta t}$$

Reaction rates and Stoichiometry

Generalization for the reaction



$$rate = -\frac{1}{a} \frac{\Delta[A]}{\Delta t} = -\frac{1}{b} \frac{\Delta[B]}{\Delta t} = \frac{1}{c} \frac{\Delta[C]}{\Delta t} = \frac{1}{d} \frac{\Delta[D]}{\Delta t}$$

Reactants (decrease)
Products (increase)

Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation
- 4 Stochastic Simulation Algorithms
- 5 Comparing Models
- 6 Improved SSA
- 7 Stochastic Differential Equations

Concentration and Rate

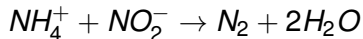
- Each reaction has its own equation that gives its rate as a function of reactant concentrations.

This is called its **Rate Law**

- To determine the rate law, we measure the rate at different starting concentrations.

Concentration and Rate

Experiment Number	Initial NH_4^+ Concentration (M)	Initial NO_2^- Concentration (M)	Observed Initial Rate (M/s)
1	0.0100	0.200	5.4×10^{-7}
2	0.0200	0.200	10.8×10^{-7}
3	0.0400	0.200	21.5×10^{-7}
4	0.0600	0.200	32.3×10^{-7}
5	0.200	0.0202	10.8×10^{-7}
6	0.200	0.0404	21.6×10^{-7}
7	0.200	0.0606	32.4×10^{-7}
8	0.200	0.0808	43.3×10^{-7}

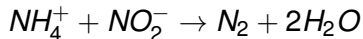


Compare experiments 1 and 2

When $[\text{NH}_4^+]$ doubles, the initial rate doubles.

Concentration and Rate

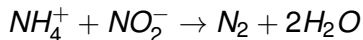
Experiment Number	Initial NH_4^+ Concentration (M)	Initial NO_2^- Concentration (M)	Observed Initial Rate (M/s)
1	0.0100	0.200	5.4×10^{-7}
2	0.0200	0.200	10.8×10^{-7}
3	0.0400	0.200	21.5×10^{-7}
4	0.0600	0.200	32.3×10^{-7}
5	0.200	0.0202	10.8×10^{-7}
6	0.200	0.0404	21.6×10^{-7}
7	0.200	0.0606	32.4×10^{-7}
8	0.200	0.0808	43.3×10^{-7}



Compare experiments 5 and 6

When $[\text{NO}_2^-]$ doubles, the initial rate doubles.

Concentration and Rate



- $\text{rate} \propto [\text{NH}_4^+]$
- $\text{rate} \propto [\text{NO}_2^-]$
- $\text{rate} \propto [\text{NH}_4^+][\text{NO}_2^-]$

$$\text{rate} = k[\text{NH}_4^+][\text{NO}_2^-]$$

This equation is called the **rate law** and k is the **rate constant**.

Rate Laws

- The *rate law* shows the relationship between the reaction rate and the concentrations of reactants.
- k is a constant that has a specific value for each reaction.
- The value of k is determined experimentally.



"Constant" is relative here:

- k is unique for each reaction.
- k changes with respect to the temperature.

Rate Laws

$$rate = k[NH_4^+][NO_2^-]$$

- Exponents tell the **order** of the reaction with respect to each reactant.
- This reaction is:
 - First-order in $[NH_4^+]$
 - First-order in $[NO_2^-]$
- The **overall reaction order** is found by adding the exponents on the reactants in the rate law.
- This reaction is **second-order overall**.

Integrated Rate Laws

Consider a simple 1st order reaction: $A \rightarrow B$

The rate of the reaction is: $rate = k[A]$

Differential form: $-\frac{d[A]}{dt} = k[A]$

How much A is left at time t ?

Integrated Rate Laws

Consider a simple 1st order reaction: $A \rightarrow B$

The rate of the reaction is: $rate = k[A]$

Differential form: $-\frac{d[A]}{dt} = k[A]$

How much A is left at time t ?

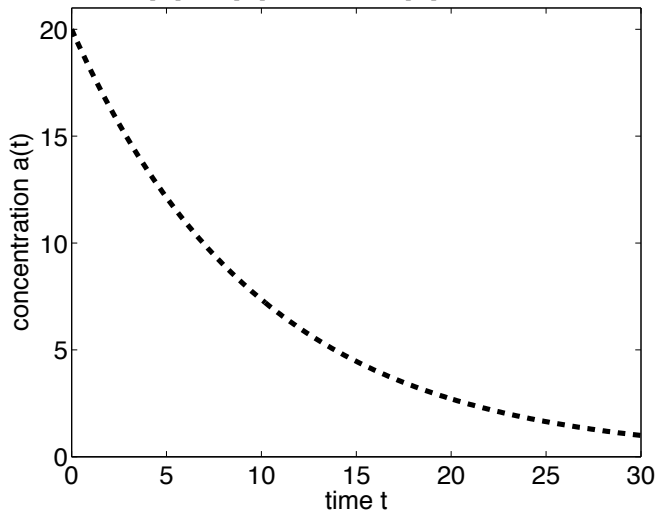
By integrating, the equation for the time evolution of $[A]$ is obtained:

$$[A]_t = [A]_0 e^{-kt}$$

where $[A]_0$ is the initial concentration of $A(t = 0)$, and $[A]_t$ is the concentration of A at some time t during the course of the reaction.

Concentration of $[A]$ as a function of time

$A \rightarrow B$. Solution: $[A]_t = [A]_0 e^{-kt}$ with $[A]_0 = 20$:



Integrated Rate Laws

The integrated form of first order rate law: $[A]_t = [A]_0 e^{-kt}$ can be rearranged to give:

$$\ln \frac{[A]_t}{[A]_0} = -kt$$

what can be manipulated this way:

$$\ln \frac{[A]_t}{[A]_0} = -kt$$

$$\ln[A]_t - \ln[A]_0 = -kt$$

$$\ln[A]_t = -kt + \ln[A]_0$$

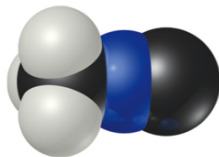
which is in the form: $y = mx + b$

First-Order Processes

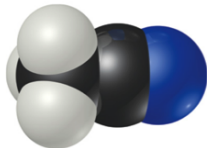
$$\ln[A]_t = -kt + \ln[A]_0$$

- If a reaction is first-order, a plot of $\ln[A]_t$ vs. t will yield a straight line with a slope of $-k$.
- Thus, plots can be used to determine the reaction order.

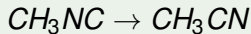
First-Order Processes



Methyl isonitrile

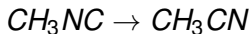


Consider the process in which methyl isonitrile is converted to acetonitrile.

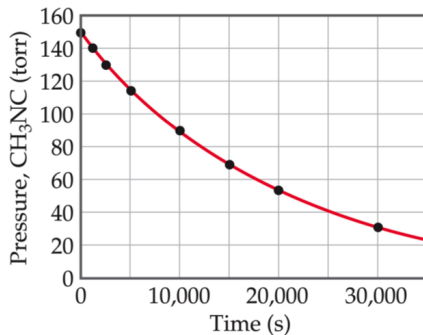


How do we know this is a first order reaction?

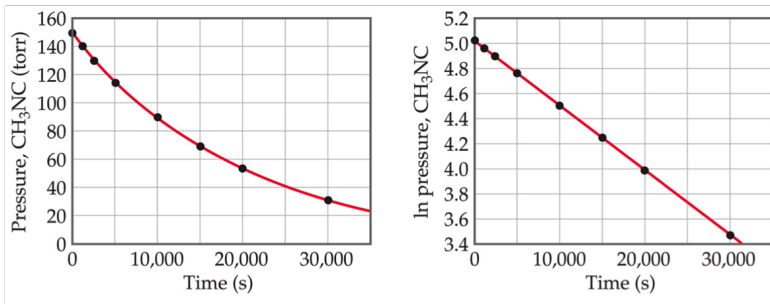
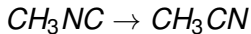
First-Order Processes



- This data was collected for this reaction at 198.9°C .
- Does $\text{rate} = k[\text{CH}_3\text{NC}]$ for all time intervals?



First-Order Processes



$$[A]_t = [A]_0 e^{-kt} \quad \ln[A]_t = -kt + \ln[A]_0$$

Second-Order Processes

Similarly, we can integrate the rate law for a process that is second-order in reactant A :

$$\text{rate} = -\frac{d[A]}{dt} = k[A]^2$$

By integrating, we obtain: $[A]_t = \frac{1}{kt + \frac{1}{[A]_0}}$

After rearranging: $\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$

what is also in the form:

$$y = mx + b$$

Second-Order Processes

$$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$$

If a process is second-order in A , then a plot of $\frac{1}{[A]}$ vs. t will yield a straight line with a slope of k .

Kinetics outline



	First order	Second order
Rate Laws	$rate = -k[A]$	$rate = -k[A]^2$
Integrated Rate Laws	$\ln \frac{[A]_t}{[A]_0} = -kt$	$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$

Kinetics outline



	First order	Second order
Rate Laws	$rate = -k[A]$	$rate = -k[A]^2$
Integrated Rate Laws	$\ln \frac{[A]_t}{[A]_0} = -kt$	$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$



	Second order
Rate Laws	$rate = -k[A][B]$
Integrated Rate Laws	complicated

Reaction Mechanisms

Molecularity	Elementary Reaction	Rate Law
<i>Unimolecular</i>	$A \longrightarrow \text{products}$	$\text{Rate} = k[A]$
<i>Bimolecular</i>	$A + A \longrightarrow \text{products}$	$\text{Rate} = k[A]^2$
<i>Bimolecular</i>	$A + B \longrightarrow \text{products}$	$\text{Rate} = k[A][B]$
<i>Termolecular</i>	$A + A + A \longrightarrow \text{products}$	$\text{Rate} = k[A]^3$
<i>Termolecular</i>	$A + A + B \longrightarrow \text{products}$	$\text{Rate} = k[A]^2[B]$
<i>Termolecular</i>	$A + B + C \longrightarrow \text{products}$	$\text{Rate} = k[A][B][C]$

The molecularity of a process tells how many molecules are involved in the process.

Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation**
- 4 Stochastic Simulation Algorithms
- 5 Comparing Models
- 6 Improved SSA
- 7 Stochastic Differential Equations

Deterministic: The law of mass action

- The fundamental empirical law governing reaction rates in biochemistry is the law of mass action.
- This states that for a reaction in a homogeneous, free medium, the reaction rate will be proportional to the concentrations of the individual reactants involved.

Criticism: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, **continuous** and **deterministic**.
- These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules.
- As we consider smaller and smaller systems, the validity of a continuous approach becomes ever more tenuous.
- As such, the adequacy of the law of mass action has been questioned for describing biological systems formed by single living cells where the small population numbers of some reactant species can result in dynamical behaviour that is noticeably **discrete rather than continuous** and noticeable **stochastic rather than deterministic**.

Stochastic Models

Arguments for the use of stochastic models for biochemical reactions (1/2):

- The quantity of the components is discrete:
 - Molecules come in *whole numbers*.
 - Molecular populations change only by *integer* amounts.
- The character of the phenomena is inherently random:
 - Only if we were to define the system's states as the positions and the velocities of all the molecules (and assume Newtonian mechanics) could we regard the system as being *deterministic*.
 - But even then, the extreme sensitivity to initial conditions will render the system effectively stochastic - like a tossed coin.
 - Chemical reactions occur as discrete events, as a result of molecular collisions that cannot be precisely predicted.
 - At best, we can predict only the *probability* that a reaction event will occur.

Stochastic Models

Arguments for the use of stochastic models for biochemical reactions (2/2):

- Stochastic models are in accordance with the theory of thermodynamics.
- Stochastic models are appropriate to describe small systems and instability phenomena.
- Whereas the deterministic approach outlined above is essentially an empirical law, derived from in vitro experiments, the stochastic approach is far more physically rigorous.
- Fundamental to the principle of stochastic modelling is the idea that molecular reactions are essentially **random processes**; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur.

Stochastic: Predictability of macroscopic states

- In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable.
- It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in in vivo conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Stochastic: Propensity function

As explicitly derived by Gillespie, the stochastic model uses basic Newtonian physics and thermodynamics to arrive at a form termed the *propensity function*.

The *propensity function* gives the probability a_j of reaction j occurring in time interval $(t, t + dt)$.

$$a_j dt = h_j c_j dt$$

where the M reaction mechanisms are given arbitrary index j ($1 \leq j \leq M$) and h_j denotes the number of possible combinations of reactant molecules involved in reaction j .

For example, if reaction l involves two species S_1 and S_2 , with X_i molecules of species S_i , we have $h_l = X_1 X_2$.

Stochastic: Propensity function

Propensity function

$$a_j dt = h_j c_j dt$$

- The rate constant c_j is dependent on the radii of the molecules involved in the reaction and their average velocities.
- These quantities are basic chemical properties which for most systems are either well known or easily measurable.
- The above equation for a_j constitutes the fundamental hypothesis of the stochastic formulation of chemical kinetics.

Stochastic: Grand probability function

The stochastic formulation proceeds by considering the *grand probability function*:

$P(\mathbf{X}; t) \equiv$ probability that there will be present in volume V at time t , X_i of species S_i , where $\mathbf{X} \equiv (X_1, X_2, \dots, X_N)$ is a vector of molecular species populations.

Evidently, knowledge of this function provides a complete understanding of the probability distribution of all possible states at all times.

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + dt)$ in which either 0 or 1 reactions occur, we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state \mathbf{X} at time $t + dt$:

$$P(\mathbf{X}; t + dt) = P(\mathbf{X}; t)P(\text{no state change over } dt) \\ + P(\text{state change to } \mathbf{X} \text{ over } dt)$$

where ν_j is a stoichiometric vector defining the result of reaction j on state vector \mathbf{X} , i.e., $\mathbf{X} \rightarrow \mathbf{X} + \nu_j$ after an occurrence of reaction j .

The probability of more than one reaction occurring in time interval $(t, t + dt)$ is $o(dt)$, and hence vanishes in the limit $dt \rightarrow 0$

Stochastic: State change probabilities

It is straightforward to show that:

$$P(\text{no state change over } dt) = 1 - \sum_{j=1}^M a_j(\mathbf{X})dt$$

$$P(\text{state change to } \mathbf{X} \text{ over } dt) = \sum_{j=1}^M P(\mathbf{X} - \nu_j; t) a_j(\mathbf{X} - \nu_j)dt$$

Then:

$$\begin{aligned} P(\mathbf{X}; t + dt) = & P(\mathbf{X}; t) \left(1 - \sum_{j=1}^M a_j(\mathbf{X})dt \right) \\ & + \sum_{j=1}^M P(\mathbf{X} - \nu_j; t) a_j(\mathbf{X} - \nu_j)dt \end{aligned}$$

Stochastic: Chemical Master Equation

If we note that:

$$\lim_{dt \rightarrow 0} \frac{P(\mathbf{X}; t + dt) - P(\mathbf{X}; t)}{dt} = \frac{\partial P(\mathbf{X}; t)}{\partial t}$$

we arrive at the

Chemical Master Equation (CME)

that describes the stochastic dynamics of the system:

$$\frac{\partial P(\mathbf{X}; t)}{\partial t} = \sum_{j=1}^M (P(\mathbf{X} - \nu_j; t) a_j(\mathbf{X} - \nu_j) - P(\mathbf{X}; t) a_j(\mathbf{X}))$$

Stochastic: Chemical Master Equation

- The CME is really a set of nearly as many coupled ordinary differential equations as there are combinations of molecules that can exist in the system!
- The CME can be solved analytically for only a very few simple systems, and numerical solutions are usually prohibitively difficult.

Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation
- 4 Stochastic Simulation Algorithms**
- 5 Comparing Models
- 6 Improved SSA
- 7 Stochastic Differential Equations

Stochastic Simulation Algorithms (SSA)

Gillespie algorithm

- Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.
- It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.
- As with the chemical master equation, the SSA converges, in the *limit* of large numbers of reactants, to the same solution as the law of mass action.

Stochastic Simulation Algorithms (SSA)

Gillespie's exact SSA

Key points:

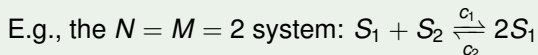
- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.
- The probability of one reaction occurring relative to another is dictated by their relative propensity functions.
- According to the correct probability distribution derived from the statistical thermodynamics theory, a random variable is then used to choose which reaction will occur, and another random variable determines how long the step will last.
- The chemical populations are altered according to the stoichiometry of the reaction and the process is repeated.

Gillespie's exact SSA

More formally:

Problem statement:

- We have N chemical species S_1, \dots, S_N and M reaction channels R_1, \dots, R_M .



- Assume the system has constant volume Ω , and constant temperature T .
- Assume the system is **well-stirred**, i.e., spatially homogeneous
- Let $X_i(t)$ = number of S_i molecules in the system at time t .
- $(X_1(t), \dots, X_N(t)) = \mathbf{X}(t)$, the **state** of the system at time t .

The Problem: Given $\mathbf{X}_0 = \mathbf{x}_0$, find $\mathbf{X}(t)$ for $t > t_0$.

Gillespie's exact SSA

Recall that traditional approach...

- Asserts that $\mathbf{X}(t)$ evolves in time according to a set of coupled, first-order, *ordinary differential equations* of the form:

$$\frac{dX_i}{dt} = f_i(X_1, \dots, X_N) \quad i = 1 \dots N$$

where the f_i are determined by the forms of the M reaction channels.

- Called the **reaction rate equations** (RRE).
- Usually written in terms of the concentrations, $Z_i = X_i/\Omega$
- According to RRE, $\mathbf{X}(t)$ is a **continuous, deterministic** process.

Gillespie's exact SSA

But in fact.... $\mathbf{X}(t)$ is *not continuous*; it is **discrete**:

- Molecules come in *whole numbers*.
- Molecular populations change only by *integer* amounts.

$\mathbf{X}(t)$ is *not deterministic*; it is **stochastic**:

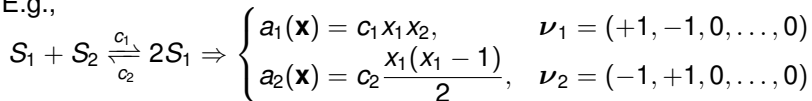
- Only if we were to define the system's state as the position and the velocities of all the molecules (and assume Newtonian mechanics) could we regard the system as being "deterministic".
- But even then, the extreme *sensitivity to initial conditions* will render the system *effectively* stochastic - like a tossed coin.
- Chemical reactions occur as discrete events, as a result of molecular collisions that cannot be precisely predicted.
- At best, we can predict only the *probability* that a reaction event will occur.

Gillespie's exact SSA

Each elemental reaction channel R_j is defined by two quantities:

- a **propensity function** $a_j(\mathbf{x})$, where $a_j(\mathbf{x})dt = \text{probability}$, given $\mathbf{X}(t) = \mathbf{x}$, that one R_j reaction event will occur in $[t, t + dt)$
- a **state change vector** $\boldsymbol{\nu}_j = (\nu_{1j}, \dots, \nu_{Nj})$, where $\nu_{ij} = \text{change in the } S_i \text{ population caused by one } R_j \text{ reaction event.}$

E.g.,



Implication: $\mathbf{X}(t)$ is a **jump Markov process**, i.e., a continuous-time, discrete-state, past-forgetting, stochastic process.

Gillespie's exact SSA

Two Approaches to Stochastic Chemical Kinetics

- An **analytical** approach, and a **simulation** approach.
- They are *logically equivalent* - both follow rigorously from the same stochastic premise.
- They are *exact* for **well-stirred** systems.

Gillespie's exact SSA. The analytical approach.

It focuses on the *probability density function* of the random variable $\mathbf{X}(t)$, namely:

$$P(\mathbf{x}, t | \mathbf{x}_0, t_0) = \text{Prob}\{\mathbf{X}(t) = \mathbf{x}, \text{ given that } \mathbf{X}(t_0) = \mathbf{x}_0\}$$

We have proved that P obeys the time-evolution equation

$$\frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} = \sum_{j=1}^M (P(\mathbf{x} - \boldsymbol{\nu}_j; t | \mathbf{x}_0, t_0) a_j(\mathbf{x} - \boldsymbol{\nu}_j) - P(\mathbf{x}, t | \mathbf{x}_0, t_0) a_j(\mathbf{x}))$$

- It is called **the chemical master equation** (CME).
- In principle, it completely determines $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$, and hence $\mathbf{X}(t)$.
- In practice, it is impossible to solve for all but the simplest of systems.

Gillespie's exact SSA. The simulation approach.

It focuses on the function $p(\tau, j | \mathbf{x}, t)$, defined by

$p(\tau, j | \mathbf{x}, t) d\tau$ = probability, given $\mathbf{X}(t) = \mathbf{x}$, that the next reaction in the system will occur in the infinitesimal time interval $[t + \tau, t + \tau + d\tau)$, and will be an R_j reaction.

It can be proved that p is given by:

$$p(\tau, j | \mathbf{x}, t) = a_j(\mathbf{x}) e^{-a_0(\mathbf{x})\tau}, \text{ where } a_0(\mathbf{x}) = \sum_{j'=1}^M a_{j'}(\mathbf{x})$$

Gillespie's exact SSA. The simulation approach.

It can be proved that p is given by:

$$p(\tau, j | \mathbf{x}, t) = a_j(\mathbf{x}) e^{-a_0(\mathbf{x})\tau}, \text{ where } a_0(\mathbf{x}) = \sum_{j'=1}^M a_{j'}(\mathbf{x})$$

This implies:

- The time τ to the next reaction event is an exponentially distributed random variable with mean $1/a_0(\mathbf{x})$
- The channel index j of that reaction is an integer random variable with probability $a_j(\mathbf{x})/a_0(\mathbf{x})$

It's easy to generate on a computer (pseudo)random samples of τ and j according to this prescription.

Gillespie's exact SSA

Derivation of the Next-Reaction Probability Density Function

Let $P_0(\tau|\mathbf{x}, t)$ = the probability, given $\mathbf{X}(t) = \mathbf{x}$, that no reactions will occur in the time interval $[t, t + \tau)$. Then,

$$p(\tau, j|\mathbf{x}, t)d\tau = P_0(\tau|\mathbf{x}, t)(a_j(\mathbf{x})d\tau)$$

so

$$p(\tau, j|\mathbf{x}, t) = a_j(\mathbf{x})P_0(\tau|\mathbf{x}, t)$$

To calculate $P_0(\tau|\mathbf{x}, t)$, observe that it must satisfy:

$$\begin{aligned} P_0(\tau + d\tau|\mathbf{x}, t) &= P_0(\tau|\mathbf{x}, t) \left(1 - \sum_{j'=1}^M (a_{j'}(\mathbf{x})d\tau) \right) \\ &= P_0(\tau|\mathbf{x}, t)(1 - a_0(\mathbf{x})d\tau) \end{aligned}$$

Gillespie's exact SSA

whence

$$\frac{dP_0(\tau|\mathbf{x}, t)}{d\tau} = -a_0(\mathbf{x})P_0(\tau|\mathbf{x}, t)$$

The solution to this ODE for the initial condition $P_0(\tau = 0|\mathbf{x}, t) = 1$ is:

$$P_0(\tau|\mathbf{x}, t) = e^{-a_0(\mathbf{x})\tau}$$

Then:

$$p(\tau, j|\mathbf{x}, t) = a_j(\mathbf{x})e^{-a_0(\mathbf{x})\tau}, \text{ where } a_0(\mathbf{x}) = \sum_{j'=1}^M a_{j'}(\mathbf{x})$$

Gillespie's exact SSA. The Stochastic Simulation Algorithm

An explicit, exact procedure for constructing a *numerical realization* of the stochastic process $\mathbf{X}(t)$.

The **direct** version of the SSA is:

- 1 With the system in state \mathbf{x} at time t , **compute**

$$a_0(\mathbf{x}) = \sum_{j'=1}^M a_{j'}(\mathbf{x})$$

- 2 **Draw** two *unit-interval uniform* random numbers r_1 and r_2 , and compute τ and j according to:

- $\tau = \frac{1}{a_0(\mathbf{x})} \ln \left(\frac{1}{r_1} \right),$

- $j = \text{the smallest integer satisfying } \sum_{j'=1}^j a_{j'}(\mathbf{x}) > r_2 a_0(\mathbf{x}).$

- 3 **Effect** the next reaction: $t := t + \tau$ and $\mathbf{x} := \mathbf{x} + \nu_j$.
- 4 **Record** (\mathbf{x}, t) . Then **return** to Step 1, or else **end** the simulation.

Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation
- 4 Stochastic Simulation Algorithms
- 5 Comparing Models**
- 6 Improved SSA
- 7 Stochastic Differential Equations

Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation
- 4 Stochastic Simulation Algorithms
- 5 Comparing Models
- 6 Improved SSA**
- 7 Stochastic Differential Equations

Outline

- 1 Biochemical Kinetics
- 2 Reaction Rate Equation
- 3 Chemical Master Equation
- 4 Stochastic Simulation Algorithms
- 5 Comparing Models
- 6 Improved SSA
- 7 Stochastic Differential Equations

Acknowledgments

Much of the material in the course is based on the following courses:

- Stochastic Chemical Kinetics
Dan T. Gillespie
- Introduction to Stochastic Modelling of Reaction, Diffusion and Taxis Processes
Radek Erban
Mathematical Institute
University of Oxford
- Stochastic Simulation for Systems Biology
Stephen Gilmore
Laboratory for Foundations of Computer Science, University of Edinburgh, Galileo workshop, Pisa, October 2007
- Sustained Oscillations in stochastic systems
J. P. Aparicio, H. G. Solari
Mathematical Biosciences 169 (2001) 15-15
- Stochastic approaches for modelling in vivo reactions
T.E. Turner, S. Schnell, K. Burrage
Computational Biology and Chemistry, 28:165-178, 2004