1 Stochastic Processes in Biological Systems

1.1 Introduction

As organized and controlled complex biological systems might seem, noise plays a large role within nature. Reaching from the random paths of evolution to the controlled employment of variation that underlies the immune system we can find that stochasticity is a general treat of nature as it exists. It is only the combination and repetition of many different individually random events that make up the deterministic behaviour we might observe. On a large scale, those approximations with mean values and average behaviour in a population of molecules or organisms might be appropriate, but if we look at biological systems we can also find many examples were stochasticity plays a major role.

Random processes are numerous within biological systems, the most popular surely being evolution where random mutations accumulate to create new phenotypes which are than consecutively subject to selection events which are also somewhat random. Apart from that, we can find many more examples where random processes play a role such as the adaptive immune response, the detection of external signals in noisy environments and gene expression (for an example see Fig. 1).

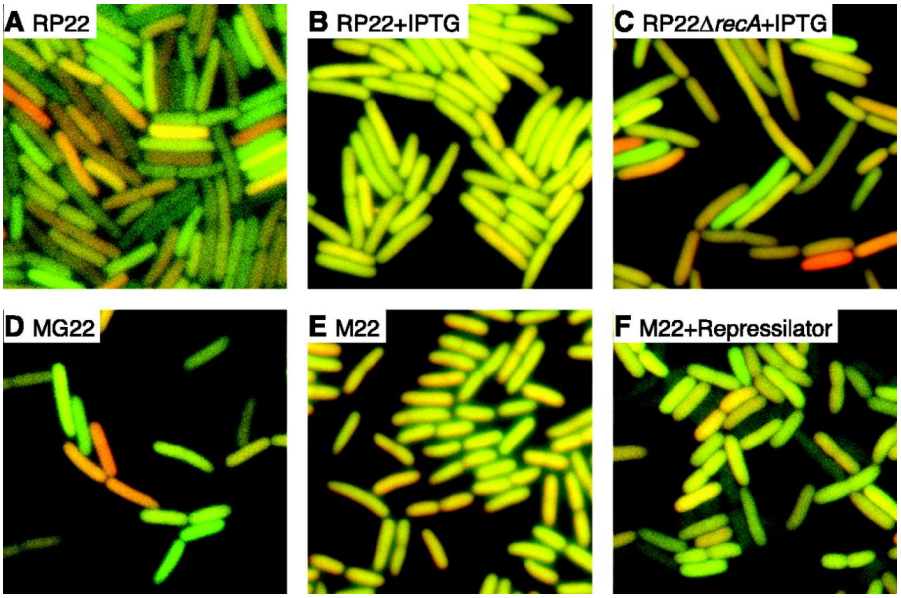


Figure 1: Noise within the gene expression of *E. coli* lac1 genes. A and D denote the wild type. The two genes for green and red flourescent protein have been expressed under two independent promoters. One can observe large fluctuations within their expression levels. The other images denote several mutants with the noise level being reduced or artificially raised. (taken from Elowitz et al., 2002)

When it comes down to modeling biological systems, the inherent stochasticity is often ignored with the argumentation that the mean value is a valid measure for the behaviour of a system or even populations of systems (such as bacterial colonies). However, it is easy to see that the existence of a mean behaviour does not imply that we will ever observe this behaviour in the system. The most popular illustration for this problem surely is "Schroedinger's cat". In average the cat will be 50% alive and 50% dead, still we are quite sure never to encounter

a cat that is alive and dead at the same time, not even in a physics lab (Schroedinger, 1935). Additionally, the random nature of a stochastic process can actually lead to serious deviations from what is usually regarded as the average in deterministic modeling. Imagine, for instance, a gene being transcribed to mRNA where we can only have one or three mRNAs being transcribed from the gene. Following that, there is the translation on the ribosome with a positive feedback that will create $mRNA^2$ proteins. In a deterministic model that would yield in average 2 mRNAs in the first step, which are now squared to yield 4 proteins in average. However, in reality, the first step will either create 1 mRNA yielding 1 protein or 3 mRNAs yielding 9 proteins, which will give us in average $0.5 \cdot (1+9) = 5$ proteins (illustrated in Fig. 2). Errors like that can be arbitrarily amplified, leading to wrong characterizations of the system. Fortunately, this modeling error is usually confined to systems inheriting small molecule numbers or large external variations combined with non-linearities. This is usually *not* the case in metabolism but becomes apparent in signaling pathways or gene expression.

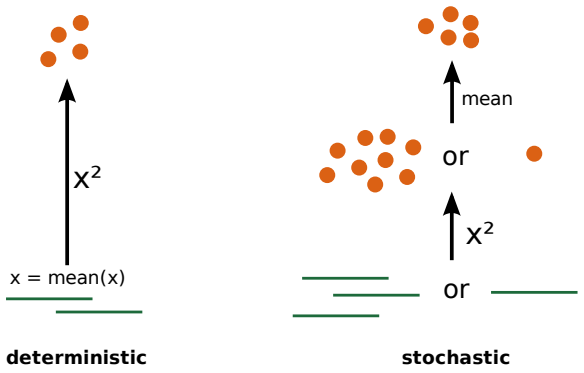


Figure 2: An exemplary example of modeling errors due to deterministic approximations

Here we have to get back to a stochastic interpretation of the system where the state space is usually discrete (numbers of molecules) and can only be characterized by probabilities. In the end we will aim at characterizing the system by generating samples, thus "real" possible outcomes of the system and accurate descriptive quantities such as estimators for the expectation ("real" mean) and variance.

1.2 Stochastic Processes and the Master Equation

When modeling with ODEs or any other discrete modeling environment the state is a unique property of the system at a certain time point. In biochemical systems the state is usually the concentration or number of each molecule, thus a vector of real numbers. In stochastic modeling we only consider states that can really occur in the system, which reduces us to the number of molecules as elements of the state. Since there are no real numbers of molecules in the system our state space will consist of vectors of non-negative integers. Furthermore, in a probabilistic setting we can not assign a unique state to a certain time point anymore, but we can assign a probability to be in a certain state at some time t. If we do that for all the states we end up with a (possibly multidimensional) probability distribution for the states $\mathbb{P}(\mathbf{S}, t)$[1]. In order to characterize this distribution we will have to make some basic assumptions about the nature of the stochastic process. The major assumption underlying the stochastic processes

[1]From here on all variables in bold face will be vectors or matrices, whereas the ones in roman will be scalars.

we will deal with is the *Markov property*. The assumption simply states that the system has no memory of its past. So after we once are in a certain state the system will not be able to remember in what state it was before; it is memory-less. Thus, if we knew the events that change the state, we could derive the probability for a specific state from the probability of another state and the probablity of the event happening.

In biochemical reaction systems we know that the state can only be changed by reactions in the system (we will regard all in- and effluxes as reactions too). So what do we know about the reactions? The time, as we consider it, is continuous. Analogous to continuous probability distributions we will not be able to assign a probability for a reaction happening at a certain time t, but only for a very small interval $(t, t + \Delta t]$. If the minimum number of substrates required for the reaction is present we know that we have a positive probability for the reaction happening in some interval $(t, t + \Delta t]$. If we also assume that the system is well-stirred, which means that we can ignore the movement of molecules as a contributing factor to the probabilities, we can derive those probabilities quite simple. By common sense we expect a reaction happening once every second to happen once if we wait one second and twice if we wait two seconds. So we would expect the probability for a reaction to scale with the length of the time interval. As such the probability for a reaction will be linear in Δt for very small intervals. So if we had the minimum number of substrates required for a single reaction the probability $\mathbb{P}(j, \Delta t)$ for a reaction of type j within some $(t, t + \Delta t]$ would be

$$\mathbb{P}(j, \Delta t) = k_j \cdot \Delta t, \tag{1}$$

with some constant k_j measured in s^{-1}. However, if we had more available molecules of one substrate than we need we could have several possible reactions for any subset of molecules. So, for a reaction only taking a single molecule as a subtrate but having, for instance, two possible molecules of the substrate available, we could have the first molecule reacting *or* the second, which leads to an addition of the two probabilities. Thus, the probability for a reaction also depends on the state $\mathbf{S}$ of the system and we have to multiply the probability shown above by the number of possibilities to choose the required molecules from the actual available molecules. Defining $\mathbf{a}_j = (a_k)_k$ as the vector of required molecules of type k to react in reaction j and $\mathbf{S} = (S_k)_k$ as the state (number of molecules of type k) of the system we can now derive

$$\mathbb{P}(j, \mathbf{S}, \Delta t) = k_j \cdot \prod_k \binom{S_k}{a_k} \cdot \Delta t =: r_j(\mathbf{S}, t) \cdot \Delta t. \tag{2}$$

The terms $r_j(\mathbf{S}, t)$ are often called *propensities* or *infinitesimal characterization*. They are the probabilistic equivalent of the reaction rates in ODE modeling.

Given that, we can now construct the probability for being in a given state at a certain time point surprisingly easy if we know the initial time t_0 and the initial probability distribution of the state $\mathbb{P}(\mathbf{S}_0, t_0)$. We only have to realize that there are only four basic ways to actually influence the probability for a state: (i) we stay in the state, (ii) a reaction within $(t, t + \Delta t]$ gets us into the state, (iii) we leave the state due to a reaction in $(t, t + \Delta t]$ or (iv) the state changes due to several reactions within $(t, t + \Delta t]$. Thus, the probability to be in a certain state after Δt, $\mathbb{P}(\mathbf{S}, t + \Delta t | \mathbf{S}_0, t_0)$, is the sum of the mentioned four probabilities.

The probability of (i) is the probability of being in a state exactly one reaction j away from entering the desired state *and* having that reaction actually hapening within the next Δt. Defining $\boldsymbol{\phi}_j$ as the state change propagated by the reaction (equivalent to the j^{th} column of

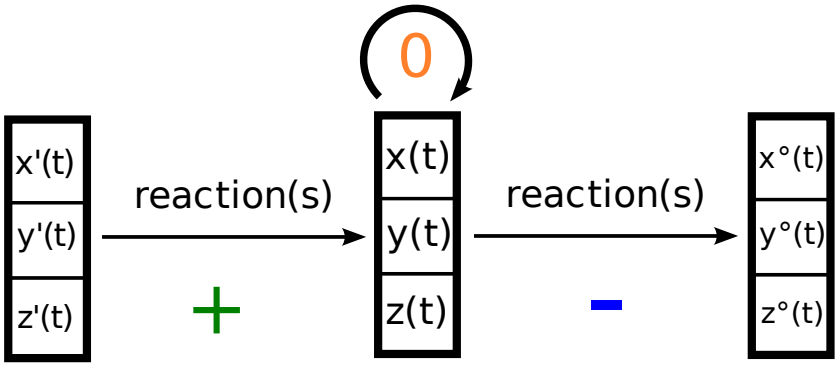


Figure 3: The events contributing to changes of the state space and summarized in the Chemical Master Equation

the stochiometric matrix), this is given by

$$\sum_j \mathbb{P}(\mathbf{S} - \boldsymbol{\phi}_j, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S} - \boldsymbol{\phi}_j, t) \cdot \Delta t, \tag{3}$$

with probabilities of negative states being equal to zero. Analogous the probability for leaving the state in Δt (ii) is given by

$$\sum_j \mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S}, t) \cdot \Delta t \tag{4}$$

The probability for being in the state simply is $\mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0)$ and since the probability for having more than one reaction within Δt will at some point include the term $(\Delta t)^k$ for $k > 1$, it will be a term in $o(\Delta t)$. That means that dividing it by Δt and having $\Delta t \to 0$ the entire probability for (iv) will become zero[2]. Since either of the events can change the probability we have to sum them to get the joint probability (with probability contributions leaving the state being negative), leading to

$$\begin{aligned}\mathbb{P}(\mathbf{S}, t + \Delta t|\mathbf{S}_0, t_0) = \mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0) \\ + \sum_j \mathbb{P}(\mathbf{S} - \boldsymbol{\phi}_j, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S} - \boldsymbol{\phi}_j, t) \cdot \Delta t \\ - \sum_j \mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S}, t) \cdot \Delta t \\ + o(\Delta t)\end{aligned} \tag{5}$$

We still have the problem of the Δt. To actually fulfill the assumptions we have to eliminate all secondary effects on that equation which will only hold for a Δt smaller than the minimum time to the next reaction. Since there is a positive probability for all positive Δt we have to consider that equation in the limit of $\Delta t \to 0$. We will do that by first substracting the $\mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0)$ from both sides of the equation, dividing by Δt and letting $\Delta t \to 0$ which leads

[2]You can show the fact that there is indeed a $(\Delta t)^k$ and that it is in $o(\Delta t)$ in the exercise section.

to the definition of the derivative on the left side

$$\lim_{\Delta t \to 0} \frac{\mathbb{P}(\mathbf{S}, t+\Delta t|\mathbf{S}_0, t_0) - \mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0)}{\Delta t} =$$
$$\lim_{\Delta t \to 0} \sum_j \left[\mathbb{P}(\mathbf{S} - \boldsymbol{\phi}_j, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S} - \boldsymbol{\phi}_j, t) - \mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S}, t)\right] + \frac{o(\Delta t)}{\Delta t}. \quad (6)$$

This gives us the so-called *Chemical Master Equation*, with

$$\frac{d\,\mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0)}{dt} = \sum_j \left[\mathbb{P}(\mathbf{S} - \boldsymbol{\phi}_j, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S} - \boldsymbol{\phi}_j, t) - \mathbb{P}(\mathbf{S}, t|\mathbf{S}_0, t_0) \cdot r_j(\mathbf{S}, t)\right]. \quad (7)$$

As we can see, on the left side we have the probabilities for all possible states, which are in theory all vectors of non-negative integers. So solving the Chemical Master Equations means solving a (possibly infinite and non-linear) set of differential equations. Thus, except for very simple systems, we can solve the Chemical Master Equation neither analytically nor numerically.

Well, this was quite an effort to formulate a problem we can not solve, wasn't it? But, if we think about it, we didn't want to have the real probability distribution anyways. We would be quite content with having some approximation for the mean and variance and to generate possible outcomes of the process. As we will see now, that is by far easier than dealing with the Master Equation.

1.3 Markov Jumping and Gillespie's Direct Method

If we go back to our basic assumptions we see that we have defined a process on a continuous time but with finite propensities. This means the "rate" by which reactions appear in our system might be stochastic but it is not unbounded. This makes sense as we would not expect an infinite number of reactions happening in real life either. So the process characterized by the Chemical Master Equation takes a quite easy form, where the state is only changed by reactions with a waiting time between consecutive reactions (see Fig. 4). Thus, the process waits for some random time before it suddenly jumps and changes the state space. Because of that, processes defined as the one here are also referred to as *Markov Jump Processes*.

So can we say something about the waiting times then? Well, to regard you for getting through the hard part of formulating the Master Equation let us just omit the proof[3] and be happy that the waiting times simply obey the exponential distribution, where the rate is also given quite easily by the sum of the propensities. Thus, the probability $\mathbb{P}(\tau, 0)$ that we have to wait a time τ before the next reaction happens is

$$\mathbb{P}(\tau, 0) = r_0(\mathbf{S}, t) \exp\left(-r_0(\mathbf{S}, t)\tau\right), \text{ with } r_0(\mathbf{S}, t) := \sum_{j=1}^{n} r_j(\mathbf{S}, t). \quad (8)$$

The probability $\mathbb{P}(j|\tau)$ that the next reaction will be reaction j is even simpler and has the form (again without proof)

$$\mathbb{P}(j|\tau) = \frac{r_j(\mathbf{S}, t)}{r_0(\mathbf{S}, t)}. \quad (9)$$

[3] If you actually think that the first part was too easy you can of course attempt to proof the following theorems. It's actually not that hard. Simply state the probability for any reaction occuring and solve it for $\Delta t \to 0$...

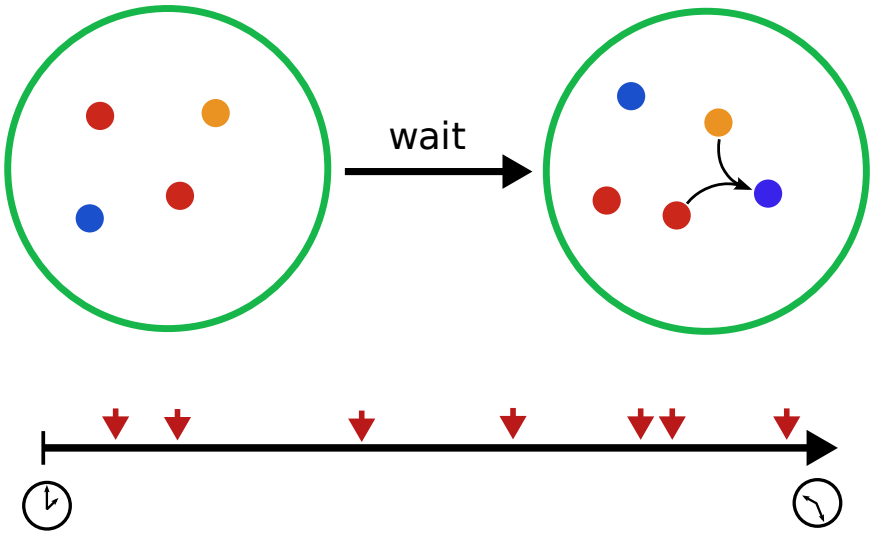


Figure 4: Markov Jump Processes are characterized by random waiting times followed by some kind of event changing the state space.

Summarizing those two results we can now obtain the probability $\mathbb{P}(\tau, j|\mathbf{S}, t)$ of waiting time τ before the next reaction and that reaction being of type j, given we are in time t and in state $\mathbf{S}$, as

$$\mathbb{P}(\tau, j|\mathbf{S}, t) = \frac{r_j(\mathbf{S}, t)}{r_0(\mathbf{S}, t)} \cdot r_0(\mathbf{S}, t) \exp\left(-r_0(\mathbf{S}, t)\tau\right) = r_j(\mathbf{S}, t) \exp\left(-r_0(\mathbf{S}, t)\tau\right). \tag{10}$$

This concise solution regarding the jump characterization of the process was introduced to biochemical systems by Daniel T. Gillespie (reviewed in Gillespie, 2007). He is also reponsible for almost all of the following theory and algorithms, the first of them being his *Direct Method*, which uses the results stated above to simulate a stochastic system and thus generate an exact sample. The algorithms is straight forward and looks like this:

1. pick a state from the initial distribution, set $t = t_0$
2. calculate the reaction propensities and r_0
3. draw a random τ from $\mathbb{P}(\tau, 0) = r_0(\mathbf{S}, t) \exp\left(-r_0(\mathbf{S}, t)\tau\right)$
4. draw a random reaction according to $\mathbb{P}(j|\tau) = \frac{r_j(\mathbf{S},t)}{r_0(\mathbf{S},t)}$
5. execute the reaction and set $t = t + \tau$
6. repeat step 1-5 till we reached or overshot the end time.

So we see that, even if the formulation of the probability distribution is quite complicated, we can derive a very simple exact procedure to generate samples from it. If we would now generate several samples we could simply calculate the population mean and variance and would get a good approximation for the expectation and variance of the distribution[4].

[4] ...because population mean and variance are both unbiased estimators

Gillespie's Direct Method is exact for arbitrary small numbers of molecules and therefore the most accurate system sampling you can get apart from Molecular Dynamics. But it also has some drawbacks. The most obvious is the necessity of repeating the sampling quite often in order to get good approximations for the mean and variance. Additionally, the algorithm strongly depends on the reaction propensities. As we can see, the number of times we have to update is given by the relation between the time interval simulated and the frequency of the reactions taking place. For that reason the algorithm can be quite slow for systems that have either fast reactions or some molecules in high abundances. Therefore, we will now direct our attention towards finding suitable approximations to the Chemical Master Equation.

1.4 Biased and Unbiased Tau-leaping

1.4.1 Biased Tau-leaping

As we have seen in the previous section, the Gillespie Method will update the system every time a single reaction takes place. This makes us quite dependent on the rates in the method, since we know that the mean of an exponential distribution is given by λ^{-1}, with λ as the rate parameter. So to simulate 100 seconds of a reaction which, in average, has 1 substrate available and happens with a rate of $1s^{-1}$ we would need in average 100 updates, whereas a reaction with 1000 substrates available and a rate of $10s^{-1}$ takes in average $10 \cdot 1000 \cdot 100 = 1000000$ updates. Thus, in particular if at least one substrate is high-abundant or the reaction rates are large, the Gillespie algorithm becomes infeasibly slow. Even though there are many variations of the Gillespie method in order to deal with a high number of reactions[5], all of them essentially inherit the strong rate-dependence (see Gibson and Bruck, 2000; Slepoy et al., 2008).

However, to hold the unfortunate situation of constantly high substrate numbers and/or fast reactions it is easy to see that a single reaction may only slightly change the state of the system. In fact, since the state only changes slightly the reaction rates will also differ only slightly from the ones calculated in the previous step. This basically means that our reaction rates are now close to being constant. So instead of merely executing a single reaction, we could also execute several reactions at once. We can use the following "tactic": first we choose a small time step τ[6]. We now try to find out how many reactions of each type might happen in this time step. In the following step we execute those possible reactions and advance by the whole time step τ. If our requirements from above are met, this would allow us to execute many reactions at once without sacrificing too much accuracy. From the Gillespie Method we already know the probability distribution of the waiting time for a single reaction, but how do we get this distribution for several reactions? Since our rates are close to being constant the Markov Jump Process now becomes a simple Poisson Process, and the distribution of several occurences of a reaction is (what a surprise) Poisson distributed. Thus we can update our

[5] In fact, you can even design the algorithm completely independent of the *number* of different reactions in the system.

[6] thus the name tau-leaping

system state the following way[7]:

$$\begin{aligned} \mathbf{S}(t+\tau) &= \mathbf{S}(t) + \sum_j \rho_j \cdot \boldsymbol{\phi}_j, \\ \text{with } \mathbb{P}(\rho_j = k, \mathbf{S}, t) &= \exp(-r_j(\mathbf{S},t)\cdot\tau)\cdot\frac{(r_j(\mathbf{S},t)\tau)^k}{k!}. \end{aligned} \tag{11}$$

This looks quite simple. However, we are still left with one problem: what about the τ? Up to now we just assumed that there is a τ so that the reaction rates stay constant. But for what τ do they actually do that? Our reaction rates can of course still be non-linear so we can not expect to answer that question in general. Furthermore, the change of the system still depends on the realizations of the Poisson variables we are using. However, what we can do is linearize the *average* change in the reaction rates taking place in $(t, t+\tau]$. If we also do that for the variance of the change in the reaction rates we get safe boundaries for that change. After some fiddling around with the Taylor series[8] we end up with the following results for the mean and variance of the rate changes [9]:

$$\left\langle \Delta r_j(\mathbf{S},\tau)\right\rangle \approx \sum_k \frac{\partial r_j(\mathbf{S},t)}{\partial S_k}\sum_l \phi_{kl}\cdot r_j(\mathbf{S},t)\tau = \sum_l \mathcal{D}_{jl}\cdot r_j(\mathbf{S},t)\tau \tag{12}$$

$$\mathbf{Var}\left(\Delta r_j(\mathbf{S},\tau)\right) \approx \sum_k \left(\frac{\partial r_j(\mathbf{S},t)}{\partial S_k}\right)^2 \sum_l \phi_{kl}^2\cdot r_j(\mathbf{S},t)\tau = \sum_l \mathcal{D}_{jl}^2\cdot r_j(\mathbf{S},t)\tau \tag{13}$$

$$\text{with } \mathcal{D}_{jl} := \sum_k \frac{\partial r_j(\mathbf{S},t)}{\partial S_k}\cdot\phi_{kl}. \tag{14}$$

With that approximation we can now choose our τ in such a way that the reaction rates only slightly deviate from its prior values and as such stay close to constant. We will do that by requiring the tau to be small enough that individual reaction rates will not change by more than a fraction ϵ. We will assume that this is is valid if neither the mean nor the standard deviation of the rate changes deviate by more than $\epsilon\cdot r_j(\mathbf{S},t)$, therefore fulfilling the so called leap condition

$$\forall j : \left\langle \Delta r_j(\mathbf{S},\tau)\right\rangle \le \epsilon\cdot r_0(\mathbf{S},t) \text{ and } \mathbf{Var}(\Delta r_j(\mathbf{S},\tau)) \le \epsilon^2\cdot r_0(\mathbf{S},t)^2. \tag{15}$$

The maximum τ fulfilling that is than given by

$$\tau = \min_j \left\{ \frac{\epsilon\cdot r_j(\mathbf{S},t)}{|\sum_l \mathcal{D}_{jl}\cdot r_j(\mathbf{S},t)|}, \frac{\epsilon^2\cdot r_j(\mathbf{S},t)^2}{\sum_l \mathcal{D}_{jl}^2\cdot r_j(\mathbf{S},t)} \right\}. \tag{16}$$

Usually, one will furthermore preselect the set of reactions which are close to consuming all of their substrates and restrict them to maximally one reaction. This will effectively avoid negative molecule numbers. Another approach is to draw the number of reactions from a

[7] $\boldsymbol{\phi_j}$ is again the j^{th} column of the stochiometrix matrix

[8] If you don't believe me, do that yourself and calculate the first order Taylor expansions for the mean and variance of $\Delta r_j(\mathbf{S},t) := r_j(\mathbf{S},t+\tau) - r_j(\mathbf{S},t)$.

[9] $\langle X\rangle$ denotes the expectation, $\mathbf{Var}(X) = \left\langle \left(X - \langle X\rangle\right)^2\right\rangle$ the variance and ϕ_{kl} denotes the l^{th} element of $\boldsymbol{\phi}_k$.

Multinomial distribution consistent with the Poisson distribution, since in a Multinomial distribution you can not have more reactions than substrates which is, in theory, possible in a Poisson-based tau-leaping.

In practice the ϵ has to be small to provide solutions consistent with Gillespie's Direct Method (usually $\epsilon < 0.01$). Otherwise you will often observe that the results obtained by the tau-leaping have somewhat smaller molecule numbers. This effect follows from the structure of the tau-leaping algorithm. If we assume a simple degradation reaction we know that the number of reactions in a fixed time interval would be decreasing, because there is less and less substrate available. However, the tau-leap will always approximate the number of reactions happening from the reaction rate at the beginning of the interval. Since those rates are a little higher than the actual ones within the interval, tau-leaping will constantly perform more reactions than is actually correct which skews the resulting molecule numbers in the direction of lower values. Due to that, this kind of tau-leaping is *biased* towards the reaction rates at the extremes of the $(t, t + \tau]$ interval.

In order to overcome that problem Gillespie also proposed the midpoint tau-leaping where you try to evaluate the reaction rates on the time point $t + \tau/2$ (Gillespie, 2001). We will not spend too much time on that one here, since even the midpoint method still incorporates a small bias. We will see that we can actually construct a nearly unbiased tau-leaping.

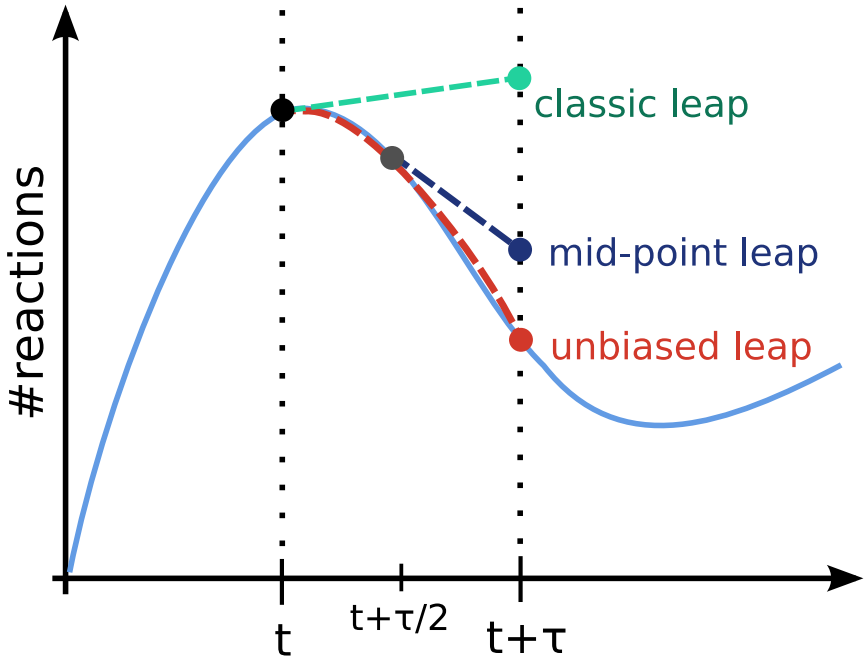


Figure 5: Comparison of the three leap procedures. The light blue line denotes the number of reactions in some fixed time interval. The dashed lines show the approximations of the leaps. The biased and midpoint tau-leaping still make a large approximation error.

1.4.2 Unbiased Tau-leaping

In order to remove the bias from the tau-leaping Xu. et. al. developed a method which goes a somewhat different way (Xu and Cai, 2008). Here they did not assume a certain distribution

of the number of the reactions *per se*, but rather tried to compute an accurate approximation for the number of reactions taking place in $(t, t+\tau]$. The main idea is to somehow derive the mean and variance of the number of reactions in advance and than choose a distribution which resembles their structure. We will not attempt to calculate the entire derivation here, since it is neither short nor simple. The basic idea is to formulate a Master Equation for the *number of reactions* (and not for the states as we have done before) and than linearize the propensities. If you additionally assume independence of the different posible numbers of reactions taking place in $(t, t+\tau]$ you end up with the following ODE system for the mean and variance of the number of reactions, $\boldsymbol{\rho}(\tau)$,

$$\frac{d\langle\rho(\tau)\rangle}{d\tau} \approx \boldsymbol{\mathcal{D}}\rho(\tau) + r_0(\mathbf{S}, t) \tag{17}$$

$$\frac{d\,\mathbf{Var}(\rho(\tau))}{d\tau} \approx 2\,\mathbf{diag}(\boldsymbol{\mathcal{D}})\,\mathbf{Var}(\rho(\tau)) + \frac{d\langle\rho(\tau)\rangle}{d\tau}, \tag{18}$$

where $\boldsymbol{\mathcal{D}}$ is the matrix for $\mathcal{D}_{ij}$ as defined earlier[10] and $\mathbf{diag}(\cdot)$ is the matrix being identical to the argument on the diagonal and zero elsewhere.

This equation can be solved analytically. But in practice one usually uses a simple ODE solver such as the Euler-Method or a low-order Runge-Kutta scheme as they are much faster than calculating the eigenvalues directly. This now gives us the approximate mean and variance in the number of reactions happening within $(t, t+\tau]$ *without* bias. Depending on the values and their relative scaling the number of reactions is than sampled from a Normal, Poisson or Multinomial distribution having the same mean and variance.

The unbiased tau-leaping generally allows for higher ϵ values, depending on the quality of the linearizations. In fact, for linear problems the unbiased tau-leaping delivers the exact number of reactions for every positive tau. This actually leads to the problem that the selection for the τ as proposed above is too "careful" since it forbids the propensities to change too much, even though our approximation might still make no error at all. The problem of selecting the optimal τ for the unbiased tau-leaping method ist still unsolved.

The tau-leaping methods are a generally valid approximation for the underlying stochastic process and are, as such, still valid for systems in which you can observe small molecule numbers. Their strength lies within their accuracy in many situations. They perform close to exact if we have only small molecule numbers and provide a valid approximation if the molecule numbers go up, without sacrificing too much computation time. This makes them the ideal choice for systems where the two cases are mixed, thus systems with some molecules in low abundance and others in high abundance[11]. But what if the molecule numbers are generally high? We will now see that this gives us even faster approximations.

1.5 The Chemical Langevin Equation and the Fokker-Planck-Equation

If our molecule numbers are generally high we can simplify our previous results even more. The Poisson distribution as well as the Binomial distribution can be well approximated by a Normal distribution in that case. Let us now assume we would always be able to choose a τ

[10] $\mathcal{D}_{jl} := \sum_k \frac{\partial r_j(\mathbf{S},t)}{\partial S_k} \cdot \phi_{kl}$

[11] This situation is often found in signaling networks or models connecting gene expression to protein synthesis.

that fulfills the leap condition and still allows for a high number of reactions within $(t, t+\tau]$. We can now formulate our tau-leaping update in the following way:

$$\begin{aligned} \mathbf{S}(t+\tau) &= \mathbf{S}(t) + \sum_j \rho_j \cdot \boldsymbol{\phi}_j, \\ \text{with } \mathbb{P}(\rho_j = k, \mathbf{S}, t) &\propto \mathcal{N}\big(r_j(\mathbf{S},t)\tau,\, r_j(\mathbf{S},t)\tau\big), \end{aligned} \tag{19}$$

with $\mathbf{S}(t)$ being the current state at time t.

If we now use the scaling properties of the Normal distribution[12] we can further decompose our distribution of reaction numbers by

$$\mathcal{N}\big(r_j(\mathbf{S},t)\tau,\, r_j(\mathbf{S},t)\tau\big) = r_j(\mathbf{S},t)\tau + \sqrt{r_j(\mathbf{S},t)\tau} \cdot \mathcal{N}\big(0,\, 1\big). \tag{20}$$

Plugging that into the update rule gives

$$\mathbf{S}(t+\tau) = \mathbf{S}(t) + \sum_j \left(r_j(\mathbf{S},t)\tau + \sqrt{r_j(\mathbf{S},t)\tau} \cdot \eta_j \right) \cdot \boldsymbol{\phi}_j \tag{21}$$

$$= \mathbf{S}(t) + \sum_j r_j(\mathbf{S},t)\tau \cdot \boldsymbol{\phi}_j + \sum_j \sqrt{r_j(\mathbf{S},t)\tau} \cdot \eta_j \cdot \boldsymbol{\phi}_j, \tag{22}$$

$$\tag{23}$$

where η_j are standard normal i.i.d.[13] variables. This leaping formula is now a much faster version than the ones regarded above, since Normal distributed random variables can be generated much faster than Poisson or Multinomial random variables. Due to our assumptions, we will also leap over several reactions in each step and can still use a large τ.

Subtracting $\mathbf{S}(t)$ and dividing both sides of the equation by τ now yields

$$\frac{\mathbf{S}(t+\tau) - \mathbf{S}(t)}{\tau} = \sum_j \boldsymbol{\phi}_j \cdot r_j(\mathbf{S},t) + \sum_j \boldsymbol{\phi}_j \cdot \sqrt{r_j(\mathbf{S},t)} \cdot \frac{\eta_j}{\sqrt{\tau}}. \tag{24}$$

After taking the limit [14] for $\tau \to 0$ we end up with the *Chemical Langevin Equation*

$$\frac{d\mathbf{S}(t)}{dt} = \sum_j \boldsymbol{\phi}_j \cdot r_j(\mathbf{S},t) + \sum_j \boldsymbol{\phi}_j \cdot \sqrt{r_j(\mathbf{S},t)} \cdot \Gamma_j(t), \tag{25}$$

The term $\Gamma_j(t)$ is the so-called *Gaussian White Noise* with

$$\Gamma_j(t) := \lim_{\sigma^2 \to 0^+} \frac{\eta_j}{\sqrt{\sigma^2}}. \tag{26}$$

It describes what would happen to a Normal distribution with mean zero if you make the variance smaller and smaller. The form it takes for the limit case is somewhat similar to the Dirac δ function since all the mass of the distribution is now concentrated into a single infinitely high peak. Without worrying too much about that weird noise term let us just note that it is

[12] $\mathcal{N}(\mu, \sigma^2) = \mu + \sigma \cdot \mathcal{N}(0,1)$

[13] i.i.d. $\triangleq$ identically independently distributed

[14] We will ignore all issues regarding the existence of that wacky differential over random variables here.

dependent of the time and is uncorrelated for all j. The important thing to see here is that the first part of the Chemical Langevin Equation before the stochastic term is completely equal to the reaction rate equations used in ODE modeling. If we would consider the thermodynamic limit[15] the second term would grow with the system size whereas the stochastic term only grows with the square root of the system size. Thus, relative to the mean of the state, the relative fluctuations die off with the square root of the system size, finally converging towards the reaction rate equations. Thus, as a rule of thumb, fluctuations in biological systems die off with the square root of the volume of the considered system.

Additionally, the characterization provided by the Chemical Langevin Equation also allows us to deduce the temporal change of the probability distribution. This is due to the fact that the Normal distribution has some nice analytical properties such as vanishing moments. Since we argued that the number of reactions happening in the system is Normal distributed we can use that property to reduce the complexity of the Chemical Master Equation. In the end, the probability distribution for the states will only be influenced by a drift and a diffusion term, that determine how the distribution changes in time. The equation summarizing those results is called the *Chemical Fokker-Planck Equation*, with

$$\frac{d\,\mathbb{P}(\mathbf{S},t)}{dt} = -\sum_i \frac{\partial}{\partial S_i}\Phi_i(\mathbf{S},t)\,\mathbb{P}(\mathbf{S},t) + \sum_i \sum_j \frac{\partial}{\partial S_i \partial S_j}\Psi_{ij}\,\mathbb{P}(\mathbf{S},t). \tag{27}$$

Where the drift term Φ is given by

$$\Phi_i(\mathbf{S},t) := \sum_k \phi_{ki} \cdot r_k(\mathbf{S},t) \tag{28}$$

and the diffusion term Ψ by

$$\Psi_{ij}(\mathbf{S},t) := \frac{1}{2}\sum_k \phi_{ki}\phi_{kj} \cdot r_k(\mathbf{S},t). \tag{29}$$

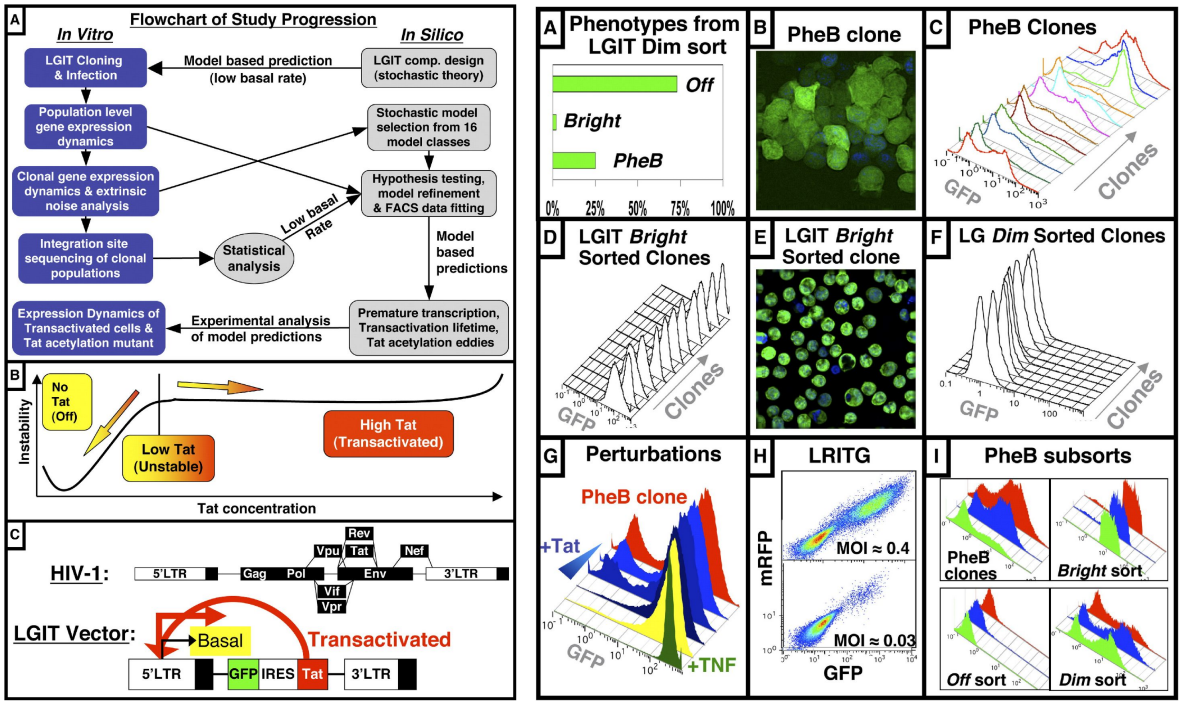
It is somehow complicated to interpret the terms in general. One could say that the drift term is responsible for the position of the distribution and the diffusion term for the shape. If the resulting distribution would, for instance, be the Normal distribution the drift term would govern how the the mean of the distribution acts in time whereas the diffusion part would govern the standard deviation. The Fokker-Planck Equation is usually much friendlier than the original Chemical Master Equation and sometimes even allows for analytical solutions where the Master Equation would not.

[15] Molecule numbers and system size (volume) go towards infinity, whereas the concentration stays constant.

1.6 An Example: Phenotypic Variation of HIV1

Having the theory, there remains the question whether we can actually observe biological systems where stochastic variations may play a role *in vivo*. One important example where this has been shown to be valid is the phenotypic variation observed within the host interation of the virus HIV1.

One of the reasons for the imminent danger caused by the HI virus is its ability to transfer part of its own genome into the genome of our $CD4^+$ T helper cells and than enter latency cycles. That means that, apart from immediately multiplying inside of the host cell, the virus also has the capability of remaining in a quiescient state. This effectively keeps infected cells secured from the immune system, allowing them to enter back into virus production when the immune system turned back to a low activity level. Due to that, the HI virus uses those repeated latency cycles to finally infect the major part of the $CD4^+$ helper cells, which leads to the symptoms of AIDS.



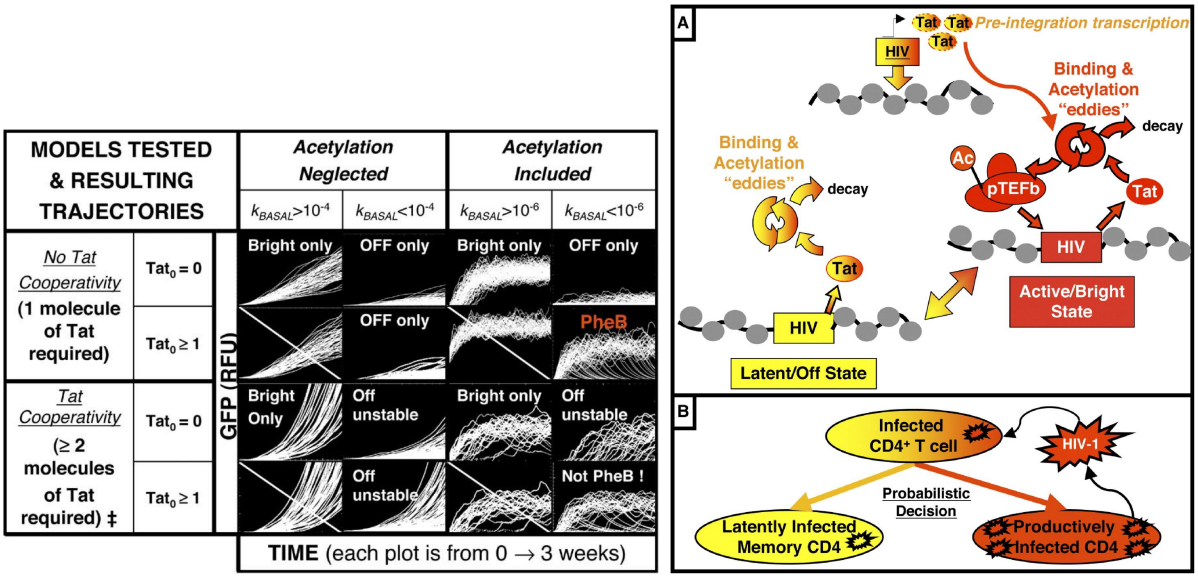
(a) (b)

Figure 6: (a) Design of the study. The study was performed by a vector (LGIT) mimicking the Tat feedbacks of the original HIV1 and a control vector without the feedback (LG). (b) Some results of the study. Shown are the sorted clones and their dynamics during the course of infection. Clones that entered a latency cycle could always be returned to an active state (for a detailed description see Weinberger et al., 2005).

Even though the mentioned behaviour of the virus has long been known, the underlying biological mechanism remained unclear. In 2005 Weinberger et. al. made a major step in the field by identifying stochastic fluctuations within a positive feedback of the transactivation regulatory protein Tat[16] as the major factor for proviral latency (Weinberger et al., 2005).

[16] or p14

They did that by comparing the known behaviour of the original HIV1 strain to mutants which only had the Tat activation cycle including its positive feedback coupled to the gene for a flourescent protein marker and mutants lacking the positive feedback (see 6). This was accompanied by a stochastic model of the underlying dynamics. They showed that the fluctuations in activated Tat (about 5-50 molecules) combined with the positive feedback were able to completely explain the observed phenotypic bistability between highly active or latent infected cells *in vivo*. By testing different versions of the corresponding stochastic model for their ability to reproduce phenotypic bistability they were able to select a single version that showed dynamics in accordance with the experiments and thus deduced the underlying mechanisms. The simulations they performed used the Gillespie Method as described earlier.



(a) (b)

Figure 7: (a) The different model version tested with only a single one showing clear phenotypic bistability. (b)The resulting mechanism of Tat activation and phenotypic bistability caused by random fluctuations in Tat activity

1.7 Exercises

The exercises here are **not** mandatory. They serve as a platform to evaluate your own understanding of the script and are a little trickier than what I would give you in a normal exercise or an exam. If you manage to solve them, you can feel quite sure that you have understood stochastic modeling. I therefore encourage you to give them at least a try.

1.7.1 Everything at once...

(a) Show that the probability for *any* two reactions happening within the interval $(t, t + \Delta t]$ takes the form

$$p(\mathbf{S}, t) \cdot (\Delta t)^2$$

with some function p independent of Δt, where $\mathbf{S}$ is the corresponding state vector.

(b) Using that result also show that the probability for more than one reaction happening in $(t, t + \Delta t]$ takes the form

$$\sum_{k=2}^{n} p_k(\mathbf{S}, t) \cdot (\Delta t)^k$$

with n reaction types in total and a set of functions p_k, each independent of Δt.

(c) Finally show that the probability for more than one reaction taking place in $(t, t + \Delta t]$ is in $o(\Delta t)$.

1.7.2 Small scale gene transcription

Consider a system where a gene with G copy numbers is transcribed into its corresponding mRNA (R) with a rate $k = 0.1s^{-1}$, thus

$$G \xrightarrow{k} R.$$

The gene copy number is a constant thus the transcription rate $k_G = k \cdot G$ is constant, the initial number of mRNAs is $R(0) := R_0$.

(a) Show that the mean and variance of the corresponding Markov process are given by

$$\langle R \rangle_t = \sum_{R=0}^{\infty} R \cdot \mathbb{P}(R, t) = k_G \cdot t + R_0 \quad \text{and}$$

$$\mathbf{Var}(R)_t = \sum_{R=0}^{\infty} \left(R - \langle R \rangle\right)^2 \cdot \mathbb{P}(R, t) = k_G \cdot t + R_0.$$

Hint: Use the Chemical Master Equation to express the derivative of $\mathbb{P}(R, t)$. Plug that into the equations above and solve them.

(b) Using a tool of your choice plot the relative deviation (or Fano factor)

$$F_t = \frac{\sqrt{\mathbf{Var}(R)_t}}{\langle R \rangle_t}$$

Do that with the time on the x-axis and for a small set of gene copy numbers ranging from 1 to 100 (just a few exemplary ones). Interpret your observation.

(c) Without explicitly calculating it, what distribution do you expect $\mathbb{P}(R, t)$ to take?

1.7.3 Tiny tau...

Consider the following system containing a single degradation of the molecule X with degradation rate $k = 0.01s^{-1}$ and initial molecule number $X(0) = X_0$. The mean and standard deviation is given by:

$$\begin{aligned} \langle X \rangle_t &= X_0 \cdot \exp(-k \cdot t) \quad \text{and} \\ \mathbf{Var}(X)_t &= X_0 \cdot \exp(-k \cdot t) \cdot (1 - \exp(-k \cdot t)). \end{aligned}$$

(a) What distribution do you expect $\mathbb{P}(X, t)$ to have?

(b) At which time point t can we observe the maximum variance?

(c) Let $X_0 = 10$. Depending on t calculate the length of the tau-leap, τ, using the leap condition for $\epsilon = 0.1$. What do you observe for small and large t?

1.7.4 Who to choose...

In your opinion, what is the appropriate method for the following problems? Choose from the following options:

- Gillespie Method
- Tau-leaping
- Langevin-leaping
- Reaction Rate Equations (ODE model).

(a) Dynamic Model of Protein synthesis for a small cell. All molecules only existent in very low numbers between 0 and 10. Slow reactions.

(b) Signal transduction model for a human cell (relatively large). Only a few activated receptors but a high number of effector molecules.

(c) Large model of a metabolism network. Cell has a large volume and all metabolite and enzymes are present in high concentrations.

(d) Model of a small cell (small volume). Concentrations are medium sized, but fast reactions.

(e) Model of a large organism. Volume very big, but concentrations low. All reactions have a linear form.

References

Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, Aug 2002. doi: 10.1126/science.1070919. URL http://dx.doi.org/10.1126/science.1070919.

Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104 (9):1876–1889, 2000. doi: 10.1021/jp993732q. URL http://pubs.acs.org/doi/abs/10.1021/jp993732q.

Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001. doi: 10.1063/1.1378322. URL http://link.aip.org/link/?JCP/115/1716/1.

Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem*, 58: 35–55, 2007. doi: 10.1146/annurev.physchem.58.032806.104637. URL http://dx.doi.org/10.1146/annurev.physchem.58.032806.104637.

Erwin Schroedinger. Die gegenwaertige situation in der quantenmechanik. *Naturwissenschaften*, 23(48):807–812, Nov 1935. doi: 10.1007/BF01491891. URL http://www.springerlink.com/content/l50v426520375016.

Alexander Slepoy, Aidan P. Thompson, and Steven J. Plimpton. A constant-time kinetic monte carlo algorithm for simulation of large biochemical reaction networks. *The Journal of Chemical Physics*, 128(20):205101, 2008. doi: 10.1063/1.2919546. URL http://link.aip.org/link/?JCP/128/205101/1.

Leor S Weinberger, John C Burnett, Jared E Toettcher, Adam P Arkin, and David V Schaffer. Stochastic gene expression in a lentiviral positive-feedback loop: Hiv-1 tat fluctuations drive phenotypic diversity. *Cell*, 122(2):169–182, Jul 2005. doi: 10.1016/j.cell.2005.06.006. URL http://dx.doi.org/10.1016/j.cell.2005.06.006.

Zhouyi Xu and Xiaodong Cai. Unbiased tau-leap methods for stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 128(15):154112, 2008. doi: 10.1063/1.2894479. URL http://link.aip.org/link/?JCP/128/154112/1.