

Fundação Getúlio Vargas
Escola de Matemática Aplicada
Curso de Graduação em Matemática
Aplicada

Bruna Fernanda Fistarol

Modelagem Estocástica e Aplicações em
Cinética Enzimática

Rio de Janeiro - Brasil

2020

Bruna Fernanda Fistarol

**Modelagem Estocástica e Aplicações em
Cinética Enzimática**

Monografia apresentada à Escola de Matemática Aplicada da Fundação Getulio Vargas para a obtenção do grau de bacharelado.

Orientador: Hugo A. de La Cruz Cancino

Rio de Janeiro - Brasil

2020

BRUNA FERNANDA FISTAROL

“MODELAGEM ESTOCÁSTICA E APLICAÇÕES EM CINÉTICA ENZIMÁTICA”


Trabalho de Conclusão de Curso - TCC apresentado ao Curso de Graduação em Matemática Aplicada da Escola de Matemática Aplicada para obtenção do grau de Bacharel (a) em Matemática Aplicada.

Data da Defesa: 08/12/2020

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA



Hugo Alexander de la Cruz Cancino
Orientador



Paulo Cezar Pinto Carvalho
Membro

DocuSigned by:



6812AE5B8AEA41A...

Alexandre Madureira
Membro

Nos termos da Lei nº 13.979 de 06/02/20 - DOU nº 27 de 07/02/20 e Portaria MEC nº 544 de 16/06/20 - DOU nº 114 de 17/06/20 que dispõem sobre a suspensão temporária das atividades acadêmicas presenciais e a utilização de recursos tecnológicos face ao COVID-19, as apresentações dos Trabalhos de Conclusão de Curso, de forma excepcional, serão realizadas de forma remota e síncrona, incluindo-se nessa modalidade membros da banca e discente.

Agradecimentos

Aos meus pais, Silvani e Renildo, por me apoiarem sempre em todas as decisões, principalmente quando escolhi estudar na EMap, precisando me mudar de uma cidade com menos de 20 mil habitantes do interior de Santa Catarina para a segunda maior metrópole do país aos 17 anos. Obrigada por terem confiado em mim. Ao Leandro, nunca irei esquecer de quando disse, com uns 9 anos de idade, que tinha orgulho de si mesmo porque tinha a mim como irmã. Obrigada por reconhecer meu esforço.

Breno, pelo amor, carinho e (muita) paciência que tem comigo, por ter permanecido ao meu lado durante toda a graduação mesmo nos piores momentos da depressão, quando era muito difícil lidar comigo. Você e a Lily sempre foram minhas alegrias diárias, e não teria chegado até o final sem isso. Eu te amo.

Ao Team M. Alex, por terem me acolhido como família no INCA e pela oportunidade de poder conhecer o universo da Biologia Molecular, da ciência e todo conhecimento que compartilharam comigo ao longo de dois anos, que foi essencial para a realização desse trabalho. Marcelo, Vanessa, Anna, Babi, Thales e Thiago, sem vocês eu não teria conseguido persistir.

Meus amigos que, por maior que seja a distância e mesmo que passemos algum tempo sem conversar, toda vez que nos encontramos a essência da amizade permanece a mesma. Obrigada por sempre torcerem por mim e por serem pacientes comigo durante esses anos de amadurecimento.

Ao CDMC, por ter me dado essa grande oportunidade de estudar na EMap e por não medir esforços para dar suporte aos alunos. Vocês são uma equipe incrível.

Ao meu professor e orientador Hugo, pela atenção durante o desenvolvimento do trabalho e compreensão com os imprevistos que tive nesse atípico final de graduação.

E a mim mesma, por não ter desistido durante o percurso mesmo com tantos motivos que surgiram.

Resumo

Em geral, a incorporação da aleatoriedade na modelagem matemática de sistemas biológicos permite obter uma realização mais realista do fenômeno estudado, já que a abordagem determinística nem sempre é capaz de descrever detalhes importantes do sistema. A modelagem estocástica compreende modelos em que, tanto o tempo quanto o espaço de estados, podem ser discretos ou contínuos. Nesse aspecto, esse trabalho realiza um estudo de modelos em tempo contínuo, sobretudo Cadeias de Markov a Tempo Contínuo aplicadas ao estudo e simulação da cinética enzimática, que é adequadamente modelada por essa abordagem, já que se trata de um sistema molecular com estados de transição e suas probabilidades.

Palavras-chave: Modelagem Estocástica. Cadeias de Markov a Tempo Contínuo. Cinética Enzimática.

Abstract

Generally, randomness incorporation in mathematical modeling of biological systems allows more accurate realizations of studied phenom, since deterministic approach may not always describes essentials system details. Stochastic modeling includes models in either time and state-space may be discrete or continuous. In this way, this work conducts a continuous-time models study, mainly Continuous-Time Markov Chains applied on enzyme kinetics study and simulation, which adequately modeled by this approach since it is a molecular system with transition states and its probabilities.

Keywords: Stochastic Modeling. Continuous-Time Markov Chains. Enzyme Kinetics.

Sumário

1	Introdução	7
2	Objetivo Geral	9
2.1	Objetivos Específicos	9
3	Cadeias de Markov a Tempo Contínuo	10
3.1	Definição	10
3.2	Homogeneidade	10
3.3	Obtenção das Probabilidades de Transição	10
3.4	Algoritmo de Gillespie	13
3.4.1	Implementação em MATLAB	14
4	Aplicação	16
4.1	<i>Chemical Master Equation</i> (CME)	16
4.2	Simulação da Cinética Enzimática	20
4.2.1	Michaelis-Menten	20
4.2.2	Reação com duas enzimas	23
4.2.3	Dados de atividade prática	26
4.2.4	Inibidores Enzimáticos	27
5	Discussão	32
6	Referências	34

1 Introdução

A modelagem matemática existe com o objetivo de simular sistemas a fim de obter resultados que são, sobretudo, relacionados à previsão desses sistemas. Nesse sentido, podemos falar sobre duas abordagens: determinística e estocástica, que em relação ao tempo também podem ser discretas ou contínuas. A escolha da forma como o sistema será modelado depende das condições e do fenômeno em si.

A abordagem determinística no processo de modelagem de sistemas biológicos nem sempre consegue capturar detalhes importantes do processo em questão ou se ajustar devidamente aos dados experimentais relacionados a ele. Por este motivo, a incorporação de aleatoriedade no modelo é fundamental para obter uma representação mais realista do fenômeno em estudo.

No projeto de iniciação científica realizado no período de 2017.2 a 2019.1, foram adquiridos conhecimentos nas áreas de Bioquímica e Biologia Molecular. E dentro desse contexto, a modelagem estocástica permite, por exemplo, simular sistemas relacionados, como reações químicas e cinética enzimática (Mohseni-Salehi *et al.*, 2019) e a dinâmica de células cancerígenas (Oroji *et al.*, 2016). A modelagem da cinética enzimática é motivada por esse contexto.

Enzimas são catalisadores biológicos altamente específicos que fazem parte da classe de proteínas globulares (com exceção das ribozimas). São produtos naturais biológicos, que dependem primariamente da sua estabilidade estrutural para sua atividade. Elas realizam o controle preciso de vias metabólicas, sendo extremamente eficientes e acelerando o tempo das reações. Uma das características das enzimas é a sua estabilidade ao longo da reação, pois elas não são consumidas, permanecendo inalteradas. A região da enzima responsável pela reação é denominada de sítio-ativo, sendo o local onde se liga o substrato.

As enzimas atuam diminuindo a energia de ativação (energia mínima necessária para que a reação ocorra) e facilitam a formação do complexo ativado, que é o estado de transição entre os reagentes e o produto onde

as ligações dos reagentes estão enfraquecidas e as do produto estão sendo formadas.

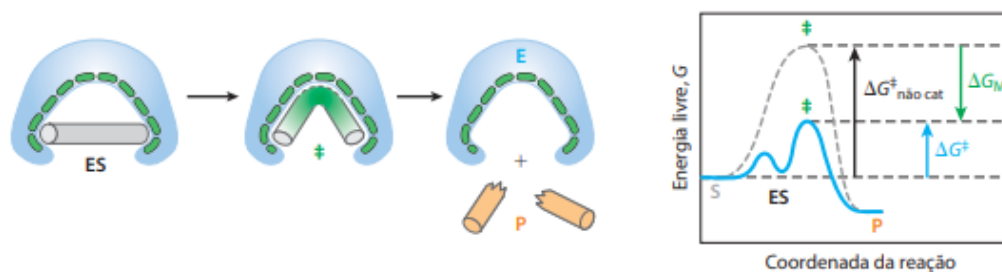


Figura 1: esquema representativo da catálise mediada por uma enzima E, formando o complexo enzima substrato ES e, posteriormente, o produto P. À direita, as consequências energéticas do sistema. Fonte: adaptado de Nelson D.L, Cox, M.M.

Quanto menor for a energia de ativação, maior é a probabilidade de ocorrência da reação espontaneamente e a velocidade de formação de produto. Para cada reação, inicialmente, observa-se a enzima e o substrato livres no meio. Posteriormente, a enzima (E) e o substrato (S) se ligam. Nesse instante os componentes sofrem pequenas modificações em sua conformação e formam o complexo enzima substrato (ES). Após essa etapa, há a formação do estado de transição onde ocorre a reação: o substrato é transformado em produto e é possível verificar a formação do complexo enzima-produto (EP). Por último, o produto é liberado no meio e o sítio ativo da enzima encontra-se livre para a conversão de uma nova molécula de substrato.

Nesse sentido, a dinâmica da cinética enzimática pode ser modelada por processos estocásticos em tempo contínuo, já que se trata de um sistema molecular com estados de transição e suas probabilidades.

2 Objetivo Geral

O estudo da modelagem e dos fenômenos relacionados a ela permite criar pontes entre áreas diferentes do conhecimento. Assim, o trabalho realizado ao longo do semestre tem por finalidade introduzir o papel do matemático aplicado não só trabalhando com a matemática em si, mas também compreendendo a fundo os fenômenos biológicos modelados, inclusive podendo colaborar com o estudo e pesquisa desses fenômenos.

2.1 Objetivos Específicos

- Estudar e compreender as principais abordagens estocásticas a tempo contínuo, principalmente as baseadas em modelos com espaços de estados discretos, isto é, Cadeias de Markov a tempo contínuo (CTMC);
- Estudar e compreender as principais técnicas que são utilizadas para construir os modelos estocásticos via CTMC.
- Estudar e compreender os métodos e algoritmos necessários para poder analisar o comportamento dos modelos, tanto analiticamente como computacionalmente.
- Por fim, como parte da aplicação dos conhecimentos adquiridos com esses objetivos, pretende-se desenvolver um estudo relativo à construção, simulação e análise de modelos no contexto da Cinética Enzimática.

3 Cadeias de Markov a Tempo Contínuo

3.1 Definição

Para certos modelos, a ideia de considerar uma mudança de estado a qualquer momento é mais interessante, ao invés de tomar intervalos de tempo determinados, como é o caso de processos estocásticos a tempo discreto. Neste caso, uma coleção de variáveis aleatórias discretas $\{X(t) : t \in [0, \infty)\}$ com espaço de estado enumerável S é uma Cadeia de Markov a Tempo Contínuo (CMTC) se satisfaz a seguinte propriedade:

$$\begin{aligned} Pr(X(t+s) = j | X(s) = i, X(s_n) = i_n, X(s_{n-1}) = i_{n-1}, \dots, X(s_1) = i_1) = \\ Pr(X(t+s) = j | X(s) = i) \\ \text{para } (0 \leq s_1 \leq \dots \leq s_n \leq s) \text{ (Allen, L.J.S, 2010).} \end{aligned}$$

3.2 Homogeneidade

Em muitas aplicações, também se considera que, em relação às probabilidades de transição, a cadeia é homogênea no tempo. Ou seja, satisfaz o seguinte:

$$Pr(X(t) = j | X(s) = i) = Pr(X(t-s) = j | X(0) = i).$$

Isso significa que as probabilidades de transição dependem apenas do tempo corrente entre os estados, e não do tempo em si (Allen, L.J.S, 2010).

3.3 Obtenção das Probabilidades de Transição

O primeiro passo na modelagem de uma CMTC antes do estudo é a construção das probabilidades de transição, denotadas por:

$$p_{ij}(\Delta t) = Pr(X(\Delta t) = j | X(0) = i).$$

O tempo de espera T_i , que é o tempo em que se permanece no estado i é tal que $T_i \sim \exp(\lambda_i)$, onde λ_i é definido de acordo com o fenômeno modelado (Dobrow R.P, 2016). Demonstração: basta mostrar que T_i “não tem memória”, propriedade de uma variável com distribuição exponencial. Sejam $s, t \geq 0$. Note que o evento $T_i > s$ é equivalente a ter $X(u) = i$, para todo $0 \leq u \leq s$, supondo i o estado inicial. Daí,

$$\begin{aligned}
 Pr(T_i > s + t | X(0) = i) &= Pr(T_i > s + t, T_i > s | X(0) = i) \\
 &= Pr(T_i > s + t | T_i > s, X(0) = i) Pr(T_i > s | X(0) = i) \\
 &= Pr(T_i > s + t | X(u) = i \ \forall \ 0 \leq u \leq s) Pr(T_i > s | X(0) = i) \\
 &= Pr(T_i > s + t | X(s) = i) Pr(T_i > s | X(0) = i) \\
 &= Pr(T_i > t | X(0) = i) Pr(T_i > s | X(0) = i).
 \end{aligned}$$

Além disso, a probabilidade de ir do estado i ao estado j é dada por $p_{ij} \geq 0$, que satisfaz

$$\sum_{j \in S} p_{ij} = 1.$$

O processo $X(t)$ fica determinado ao definir os valores de λ_i e p_{ij} . Esses valores são estabelecidos pelo estudo do fenômeno em si, ou seja, nesse caso, é necessário compreender o teórico biológico. A partir desse ponto, podemos trabalhar na resolução do modelo de maneira analítica ou computacional e realizar simulações da cadeia com algoritmo de Gillespie, que será mencionado posteriormente. Esse algoritmo foi desenvolvido baseado na ideia de simular a cinética química.

Queremos determinar, a partir destes parâmetros, $p_{ij}(\Delta t)$. Sabemos que

$$Pr(T_i > \Delta t) = e^{-\lambda_i \Delta t}.$$

Utilizando a Fórmula de Taylor:

$$\begin{aligned} Pr(T_i > \Delta t) &= e^{-\lambda_i \Delta t} \\ &= 1 - \lambda_i \Delta t + \frac{(\lambda_i \Delta t)^2}{2!} - \frac{(\lambda_i \Delta t)^3}{3!} + \dots \\ &= 1 - \lambda_i \Delta t + o(\Delta t). \end{aligned}$$

Isso implica que

$$Pr(T_i \leq \Delta t) = \lambda_i \Delta t + o(\Delta t).$$

Além disso,

$$Pr(T_i + T_j \leq \Delta t) \leq Pr(T_i \leq \Delta t)Pr(T_j \leq \Delta t).$$

Logo,

$$\begin{aligned} Pr(T_i + T_j \leq \Delta t) &\leq Pr(T_i \leq \Delta t)Pr(T_j \leq \Delta t) \\ &= (\lambda_i \Delta t + o(\Delta t))(\lambda_j \Delta t + o(\Delta t)) \\ &= \lambda_i \lambda_j (\Delta t)^2 + o(\Delta t) \\ &= o(\Delta t). \end{aligned}$$

Daí, a probabilidade de visitar 2 ou mais estados num intervalo de tempo Δt , estando no estado i , é dada por:

$$\sum_{i \neq j} Pr(T_i + T_j \leq \Delta t) p_{ij} = o(\Delta t) p_{ij} = o(\Delta t).$$

Assim, a probabilidade de visitar exatamente um estado $j \neq i$ estando no estado i equivale a:

$$\begin{aligned} Pr(\text{visitar um estado } j | X(0) = i) &= 1 - Pr(T_i > \Delta t) - o(\Delta t) \\ &= 1 - (1 - \lambda_i \Delta t + o(\Delta t)) - o(\Delta t) \end{aligned}$$

$$= \lambda_i \Delta t + o(\Delta t).$$

Finalmente, com $i \neq j$:

$$\begin{aligned} p_{ij}(\Delta t) &= Pr(X(\Delta t) = j | X(0) = i) \\ &= (\lambda_i \Delta t + o(\Delta t)) p_{ij} + o(\Delta t) \\ &= \lambda_i p_{ij} \Delta t + o(\Delta t). \end{aligned}$$

$$\begin{aligned} p_{ii}(\Delta t) &= Pr(X(\Delta t) = i | X(0) = i) \\ &= 1 - \lambda_i p_{ij} \Delta t + o(\Delta t). \end{aligned}$$

3.4 Algoritmo de Gillespie

O Algoritmo de Gillespie fornece uma maneira simples de simular processos associados aos modelos estocásticos. Originalmente, foi formulado por Daniel T. Gillespie em 1976¹ para analisar numericamente a modelagem estocástica da cinética química, a fim de simular a dinâmica de certas reações dadas as suas características.

A ideia por trás do algoritmo é gerar duas variáveis aleatórias: uma para decidir o tempo de permanência em no estado atual e outra para decidir o próximo estado. No caso da formulação original, estamos interessados em saber **qual** a próxima reação que deve o correr e **quando** ela deve ocorrer. Sendo r_1 e r_2 as variáveis aleatórias geradas com distribuição uniforme no intervalo $[0, 1]$ e i o estado atual, temos:

1. Seja $Pr(X(\Delta t) = j | X(0) = i) = q_{ij} \Delta t + o(\Delta t)$. Estimamos λ_i como $\lambda_i = \sum_{j \in S} q_{ij}$. Sabendo que $T_i \sim exp(\lambda_i)$, o tempo de permanência no estado i é dado por

$$T_i = -\frac{1}{\lambda_i} \log(r_1).$$

Isso porque a função acumulada de uma variável aleatória com distribuição exponencial de parâmetro λ é dada por $F(x) = 1 - e^{-\lambda x}$, para

¹Ver referências.

$x \geq 0$. Calculando a inversa de F , obtemos que $x = -\frac{1}{\lambda} \log(1 - F(x))$. Como $(1 - F(x)) \in [0, 1]$, isso justifica o modo como T_i é definido.

2. Dados os valores de p_{ij} por $p_{ij} = \frac{q_{ij}}{\lambda_i}$, suponha que o espaço dos N estados é dado por

$$S = (j_1, j_2, \dots, j_{N-1}, j_N).$$

Daí, construímos um intervalo da seguinte forma:

$$[0, p_{ij_1}] \cup [p_{ij_1}, p_{ij_1} + p_{ij_2}] \cup \dots \cup \left[\sum_{k=1}^{N-1} p_{ij_k}, \sum_{k=1}^N p_{ij_k} \right].$$

O próximo estado j_s é decidido de tal forma que

$$\sum_{k=1}^{s-1} p_{ij_k} < r_2 \leq \sum_{k=1}^s p_{ij_k}.$$

3.4.1 Implementação em MATLAB

O algoritmo implementado recebe como entrada a matriz dos coeficientes de transição, estado inicial e o tempo até o qual se quer simular a cadeia. O código pode ser facilmente modificado a fim de passar como entrada o número de simulações ao invés de determinar o tempo de simulação, podendo também receber diretamente as probabilidades de transição, já que elas são calculadas dentro do código a partir da matriz dos coeficientes de transição.

`% entrada:`

`% Q: coeficientes de transição (N x N), onde N é o número de estados`

`% t: tempo até o qual se quer simular a cadeia`

`% i: estado inicial (1 <= i <= N)`

`% saída:`

```
% jumps: tempo entre transições
% states: percurso da cadeia

function [jumps, states] = gillespie(Q, t, i)

time = 0;
jumps = [];
states = i;

while time < t
    lambda = sum(Q(i, :));
    r_1 = rand;
    T = -log(r_1)/lambda;
    jumps = [jumps, T];
    r_2 = rand;
    i = sum(cumsum(Q(i, :)/lambda)<r_2)+1;
    time = time + T;
    states = [states, i];
end
end
```

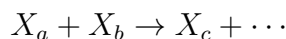

4 Aplicação

4.1 *Chemical Master Equation* (CME)

As equações mestre são utilizadas na química para descrever, ao longo do tempo, o comportamento de um sistema de moléculas, as quais podem ser contadas e determinadas com exatidão em qualquer momento.

Considere um sistema de N compostos moleculares X_1, X_2, \dots, X_N , que reagem entre si através de M vias de reações químicas R_1, R_2, \dots, R_M , com temperatura T e volume Ω constantes. O objetivo é descrever o comportamento de $X_i(t)$, definido pelo número de moléculas da espécie X_i no tempo t , para $i = 1, 2, \dots, N$.

A forma abaixo denota uma via bimolecular colisão-iniciada R_μ : e



Na reação, os compostos X_a e X_b - chamados de reagentes - são transformados no produto X_c , e assim por diante. As suposições sobre o sistema permitem deduzir o comportamento das moléculas com relação à colisão, o que reflete diretamente na definição de probabilidade.

Dessa forma, o seguinte fato pode ser definido: para cada reação R_μ ($\mu = 1, 2, \dots, M$) num sistema termicamente equilibrado, existe c_μ independente de Δt , o qual

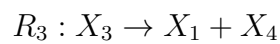
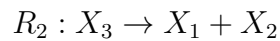
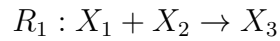
$c_\mu \Delta t =$ probabilidade de um conjunto aleatório de reagentes de R_μ no tempo t reagirem no intervalo infinitesimal $[t, t + \Delta t)$.

Na prática, essa constante representa a propensão de colisão dos reagentes no meio. Além disso, para caracterizar completamente a via R_μ , é necessário definir outros dois elementos:

$h_\mu(n_1, \dots, n_N) \equiv$ número de combinações distintas de reagentes de R_μ quando há exatamente n_i moléculas de X_i ($\mu = 1, \dots, M$)

$v_{\mu i} \equiv$ mudança na população de X_i causada por uma reação R_μ .

Para fins do estudo em questão, vamos definir as vias de reação da cinética enzimática de Michaelis-Menten:



Onde:

- X_1 é enzima
- X_2 é substrato
- X_3 é o complexo enzima-substrato
- X_4 é o produto da reação

Para simplificar a notação, tome $n = (n_1, n_2, \dots, n_N)$. A função $h_\mu(n)$ de cada uma das vias é dada por:

$$h_1(n_1, \dots, n_4) = n_1 n_2$$

$$h_2(n_1, \dots, n_4) = n_3$$

$$h_3(n_1, \dots, n_4) = n_3$$

E os valores de $v_{\mu i}$ são:

$$V(\mu, i) = \begin{pmatrix} -1 & -1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix}$$

Agora, iremos estabelecer a relação dessa dinâmica com o descrito na seção 3.

Estamos interessados em calcular as probabilidades de transição entre os estados num intervalo de tempo Δt dada uma configuração

$$X(t) = (n_1, n_2, \dots, n_N) = n.$$

Como definido, existe c_μ tal que $c_\mu \Delta t$ é a probabilidade de um conjunto aleatório de reagentes de R_μ no tempo t reagirem no intervalo infinitesimal $[t, t + \Delta t)$. Logo, a probabilidade de que ocorra uma única reação nesse intervalo é dada por

$$\begin{aligned} Pr(\text{um evento} \mid X(t) = n) &= \binom{h_\mu(n)}{1} (c_\mu \Delta t) (1 - c_\mu \Delta t)^{(h_\mu(n)-1)} \\ &= h_\mu(n) c_\mu \Delta t + o(\Delta t). \end{aligned}$$

Daqui, segue que a probabilidade de que não ocorra nenhuma reação no mesmo intervalo de tempo é dada por

$$\begin{aligned} Pr(\text{nenhum evento} \mid X(t) = n) &= \prod_\mu (1 - c_\mu h_\mu(n) \Delta t + o(\Delta t)) \\ &= 1 - \sum_\mu c_\mu h_\mu(n) \Delta t + o(\Delta t). \end{aligned}$$

Por último, a probabilidade de que ocorra mais de uma reação no intervalo $[t, t + \Delta t)$ é da ordem $(\Delta t)^k$ com $k \geq 2$ inteiro. Assim,

$$Pr(\text{mais de um evento} \mid X(t) = n) = o(\Delta t).$$

Agora, o problema já pode ser descrito analiticamente, uma vez que o processo relacionado a ele está caracterizado. Dadas configurações iniciais n_0 no tempo t_0 , a probabilidade de ir de $X(t_0) = n_0$ para $X(t + \Delta t) = n$ é expressa como:

$$\begin{aligned}
Pr(n, t + \Delta t \mid n_0, t_0) &= Pr(n, t \mid n_0, t_0) \left(1 - \sum_{\mu} c_{\mu} h_{\mu}(n) \Delta t + o(\Delta t) \right) \\
&\quad + \sum_{\mu} Pr(n - v_{\mu}, t \mid n_0, t_0) (c_{\mu} h_{\mu}(n - v_{\mu}) \Delta t + o(\Delta t)) \\
&\quad + o(\Delta t).
\end{aligned}$$

A expressão acima justifica-se com as equações que caracterizam o processo:

1. Se nenhuma reação ocorrer no intervalo $[t, t + \Delta t)$, devemos ter $X(t) = n$, expresso como probabilidade no primeiro termo multiplicado pela probabilidade de não haver reação.
2. Caso ocorra exatamente uma reação R_{μ} no intervalo $[t, t + \Delta t)$, $X(t)$ é tal que seu estado é imediatamente anterior a n , ou seja, $n - v_{\mu}$. O somatório em μ representa cada uma das M vias de reação.
3. O termo $o(\Delta t)$ surge da probabilidade de ocorrer mais de uma reação no intervalo.

Subtraindo $Pr(n, t \mid n_0, t_0)$ de ambos os lados da equação, dividindo por Δt e tomando $\Delta t \rightarrow 0$, obtemos a CME de interesse:

$$\frac{\partial}{\partial t} Pr(n, t \mid n_0, t_0) = \sum_{\mu} [c_{\mu} h_{\mu}(n - v_{\mu}) Pr(n - v_{\mu}, t \mid n_0, t_0) - c_{\mu} h_{\mu}(n) Pr(n, t \mid n_0, t_0)].$$

4.2 Simulação da Cinética Enzimática

4.2.1 Michaelis-Menten

A probabilidade de que ocorra uma única reação no intervalo $[t, t + \Delta t)$ é dada por

$$Pr(\text{um evento} \mid X(t) = n) = h_\mu(n)c_\mu\Delta t + o(\Delta t).$$

Essa probabilidade é suficiente para caracterizar a simulação do modelo, extremamente mais simples que o estudo analítico da cinética enzimática. O problema de simular os dados está em definir quem são os coeficientes em questão, o que requer entendimento teórico do fenômeno.

Na via simplificada de cinética enzimática em questão, há reações de primeira e segunda ordem.

Reações de primeira ordem: nesse tipo de reação, há um único reagente. Nesse caso, temos

$$c_\mu = k_\mu$$

onde k_μ é uma constante de taxa determinística da reação específica R_μ .

Reações de segunda ordem: há dois elementos reagindo nesse tipo de reação, e nesse caso

$$c_\mu = \frac{k_\mu}{n_A\Omega}$$

onde n_A é o número de Avogadro (Lecca P, 2013).

Em particular, temos que R_1 é uma reação de segunda ordem e R_2 e R_3 são reações de primeira ordem.

Baseado em valores de c_μ , volume e número de moléculas estabelecidos por Darren J. Wilkinson², iremos definir no MATLAB os seguintes parâmetros:

²Ver referências.

```
nA = 6.023e23; %número de Avogadro
t = 50; %tempo da simulação
omega = 1e-15; %volume da reação
X = zeros(1,4); %vetor de estados
X(1) = round(2e-7*nA*omega); % moléculas de enzima
X(2) = round(5e-7*nA*omega); % moléculas de substrato
c(1) = 1e6/(nA*omega); c(2) = 1e-4; c(3) = 0.1; % constantes
V = [-1 -1 1 0; 1 1 -1 0; 1 0 -1 1]; % matriz de mudança na qtd de moléculas
```

Algumas modificações precisam ser feitas no código genérico de Gillespie. É necessário incluir a matriz V de mudança na população de moléculas. Além disso, a matriz dos coeficientes de transição é uma função da população de moléculas a cada instante, logo, precisa ser atualizada a cada iteração.

```
% entrada:
    % c: vetor com taxas de reação estocásticas
    % t: tempo até o qual se quer simular a cadeia
    % i: quantidade inicial de cada molécula em M
    % V: matriz estoquiométrica
```

```
% saída:
    % jumps: tempo entre transições
    % states: percurso da cadeia
```

```
function [jumps, states] = gillespie(c, t, i, V)
```

```
    time = 0;
    jumps = [0];
    states = i;

    while time < t
```

```
Q = zeros(1, size(c,2));
Q(1) = c(1)*i(1)*i(2);
Q(2) = c(2)*i(3);
Q(3) = c(3)*i(3);
lambda = sum(Q);
r_1 = rand;
T = -log(r_1)/lambda;
jumps = [jumps; T];
r_2 = rand;
j = sum(cumsum(Q./lambda)<r_2)+1;
time = time + T;
i = i + V(j,:);
states = [states; i];
end
end
```

Ao realizar 10 simulações, obtemos o seguinte perfil de reação:

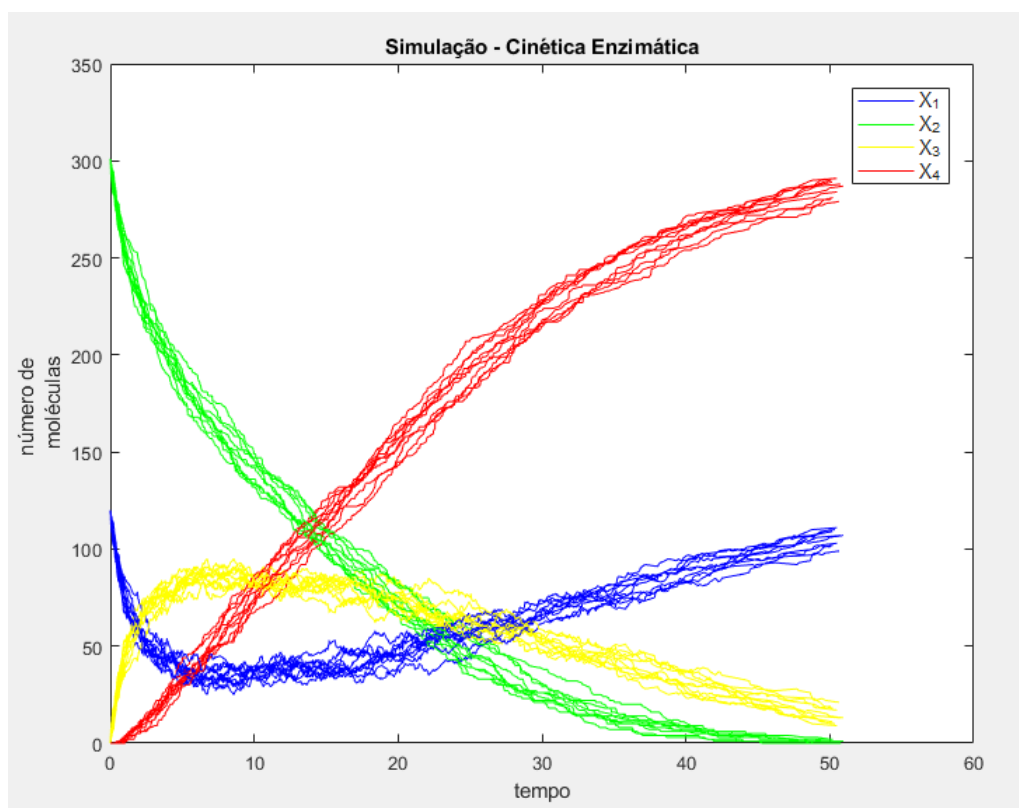
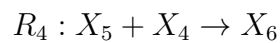
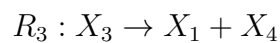
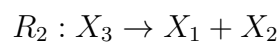
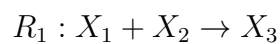


Figura 2: simulações de cinética enzimática utilizando o algoritmo de Gillespie.

4.2.2 Reação com duas enzimas

Podemos tornar a simulação um pouco mais interessante. Imagine que, no mesmo meio reacional da reação anterior, há uma segunda enzima que tem como substrato o produto da primeira reação. O novo modelo é dado por:



$$R_5 : X_6 \rightarrow X_5 + X_4$$

$$R_6 : X_6 \rightarrow X_5 + X_7$$

Onde:

- X_1 é a enzima A
- X_2 é substrato de A
- X_3 é o complexo enzima A/substrato de A
- X_4 é o produto A
- X_5 é a enzima B
- X_6 é o complexo enzima B/produto A
- X_7 é o produto B

A função h_μ de cada uma das vias é dada por:

$$h_1(n_1, \dots, n_N) = n_1 n_2$$

$$h_2(n_1, \dots, n_N) = n_3$$

$$h_3(n_1, \dots, n_N) = n_3$$

$$h_4(n_1, \dots, n_N) = n_4 n_5$$

$$h_5(n_1, \dots, n_N) = n_6$$

$$h_6(n_1, \dots, n_N) = n_6$$

E os valores de $v_{\mu i}$ são:

$$V(\mu, i) = \begin{pmatrix} -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}$$

Podemos refletir nas constantes das novas reações uma propensão maior ou menor da reação acontecer. Também é possível alterar o volume da reação, o que modifica a dinâmica das reações bimoleculares (um volume maior diminui a probabilidade de dois reagentes se encontrarem).

A simulação a seguir assume que a enzima B é menos eficiente que a enzima A. Isso significa que o complexo enzima/substrato se dissocia mais facilmente e se forma com mais dificuldade. Considere os novos parâmetros definidos em MATLAB:

```
t = 80;
X = zeros(1,7); %vetor de estados
X(1) = round(2e-7*nA*omega); % moléculas de enzima A
X(2) = round(5e-7*nA*omega); % moléculas de substrato de A
X(5) = round(2e-7*nA*omega); % moléculas de enzima B
c(1) = 1e6/(nA*omega); c(2) = 1e-4; c(3) = 0.1; % constantes
c(4) = 1e5/(nA*omega); c(5) = 1e-3; c(6) = 0.05;
V = [-1 -1 1 0 0 0 0;
      1 1 -1 0 0 0 0;
      1 0 -1 1 0 0 0;
      0 0 0 -1 -1 1 0;
      0 0 0 1 1 -1 0;
      0 0 0 0 1 -1 1]; %matriz de mudança na qtd de moléculas
```

O código do algoritmo precisa ser adaptado para lidar com os novos valores $h_\mu(n)$. A simulação desse meio reacional retorna o seguinte perfil:

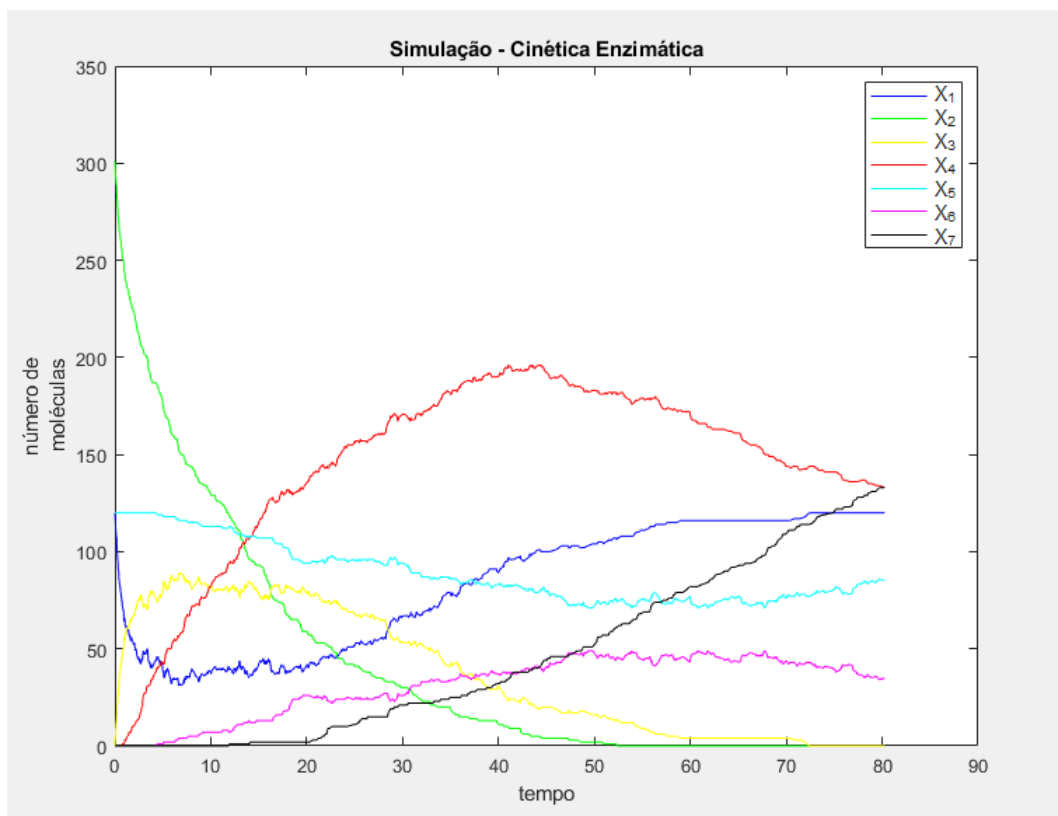


Figura 3: simulação de um sistema com duas enzimas utilizando o algoritmo de Gillespie.

4.2.3 Dados de atividade prática

No primeiro semestre de 2019, em uma aula prática da disciplina de Bioquímica I, foram gerados dados sobre dinâmica enzimática, dentre eles, o curso temporal da fosfatase alcalina em condições ideais e constantes de temperatura. Um meio reacional foi preparado em solução tampão com pH 9,5, $0,125\mu M$ do substrato p-nitro-fenil-fosfato (PNPP) e $62,5\mu g$ de fosfatase

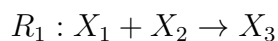
alcalina, com um volume final de $200\mu l$. A absorvância da reação foi lida a cada três minutos, durante 60 minutos num leitor de ELISA. Originalmente, pretendia-se simular a cinética dessa enzima utilizando o algoritmo com os mesmos parâmetros do experimento, dadas as constantes de taxa determinística. No entanto, ao calcular a quantidade de moléculas da reação, encontramos uma quantidade de substrato da ordem de 10^{19} e enzima da ordem de 10^{15} . Isso inviabiliza a utilização dessa versão do algoritmo de Gillespie para a simulação de experimentos reais, tópico abordado na discussão.

4.2.4 Inibidores Enzimáticos

Existe uma série de fatores que podem reduzir a atividade enzimática. Elementos químicos que se associam à enzima diminuindo a sua eficiência recebem o nome de **inibidor enzimático**. Considere os seguintes componentes:

- X_1 é a enzima
- X_2 é o substrato
- X_3 é o complexo enzima/substrato
- X_4 é o produto
- X_5 é o inibidor
- X_6 é o complexo enzima/inibidor
- X_7 é o complexo enzima/substrato/inibidor

1) Inibidor Irreversível: é quando um elemento químico liga-se covalentemente e de forma irreversível à enzima, levando à sua inativação. Nesse caso, as vias de reação seriam:



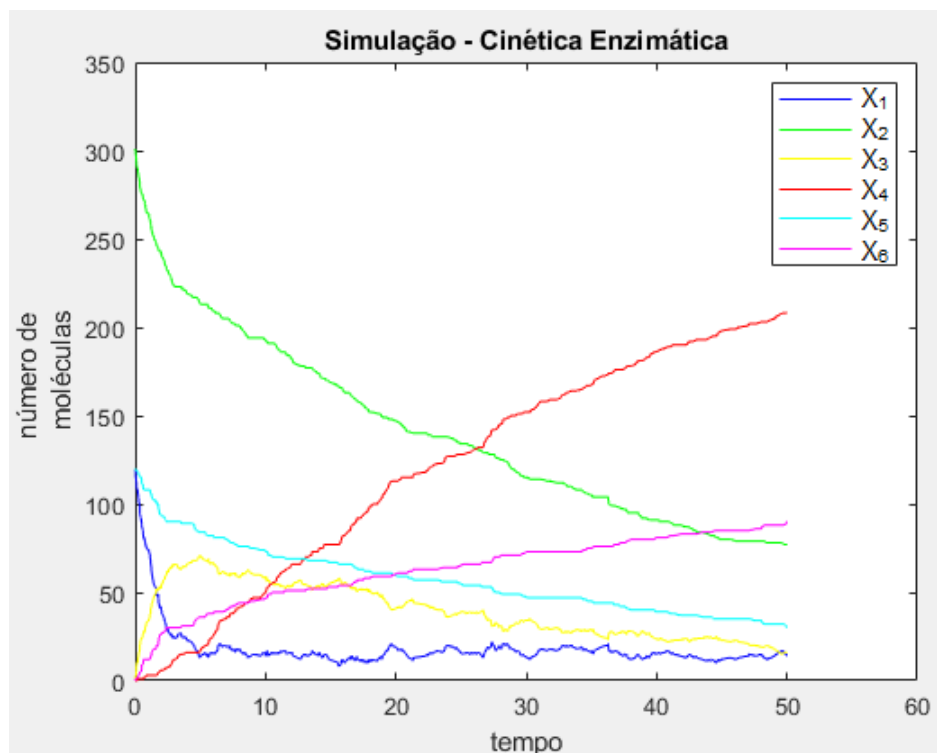
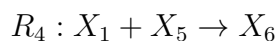
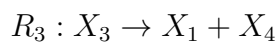
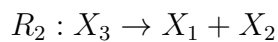


Figura 4: Simulação de cinética enzimática com inibidor irreversível com constantes $c_1 = \frac{10^6}{nA\Omega}$; $c_2 = 10^{-4}$; $c_3 = 0, 1$; $c_4 = \frac{10^6}{nA\Omega}$; $c_5 = 10^{-4}$ e quantidade de inibidor igual à quantidade de enzima.

Note o comportamento dos elementos com relação à figura 2: a simulação dura o mesmo tempo e termina com uma quantidade menor de produto formado.

Inibidor Reversível Competitivo: liga-se à enzima por meio de ligações iônicas, podendo se dissociar dela. Esse inibidor compete com o substrato pelo sítio ativo da enzima, formando um complexo que não é levado à catálise.

O perfil da simulação desse tipo de dinâmica é muito semelhante ao anterior, mas dependendo da taxa de dissociação do inibidor, a quantidade de produto varia. As vias são:

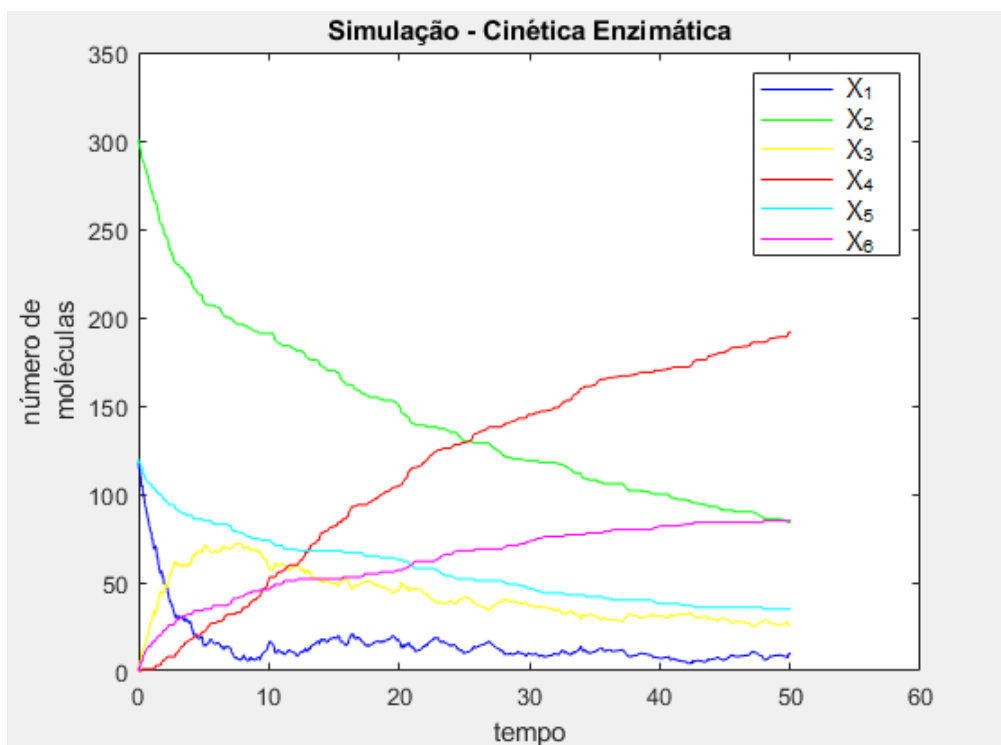
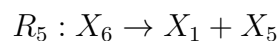
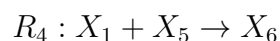
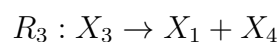
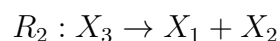
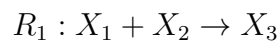
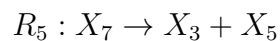
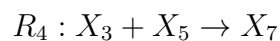
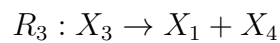
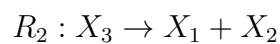
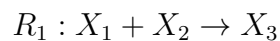


Figura 5: Simulação de cinética enzimática com inibidor reversível competitivo, utilizando as constantes $c_1 = \frac{10^6}{nA\Omega}$; $c_2 = 10^{-4}$; $c_3 = 0, 1$; $c_4 = \frac{10^6}{nA\Omega}$; $c_5 = 10^{-2}$ e quantidade de inibidor igual à quantidade de enzima.

De acordo com simulações realizadas, aumentar o valor de c_5 (que signi-

fica dissociar o inibidor mais facilmente) aumenta a quantidade de produto.

Inibidor Reversível Incompetitivo: liga-se ionicamente a cadeia polipeptídica da enzima fora da região do sítio ativo associando-se ao complexo enzima-substrato. Tem-se as vias:



Neste caso, como há menor quantidade de complexo enzima-substrato que enzima disponível no meio, o rendimento da reação acaba sendo menor com relação ao inibidor competitivo.

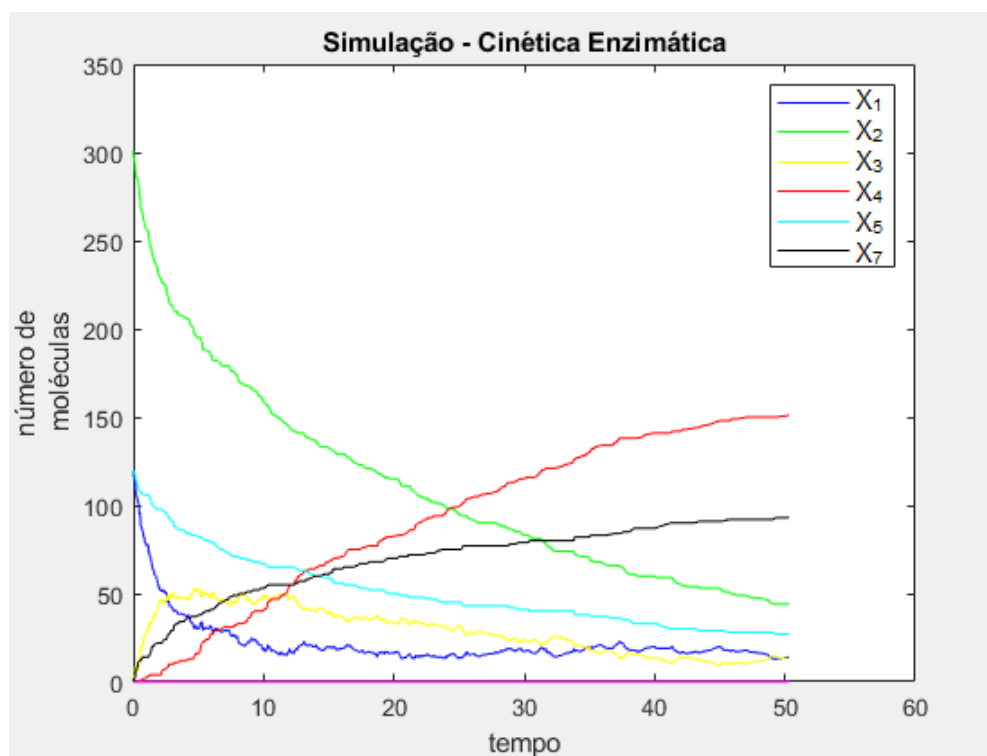


Figura 6: Simulação de cinética enzimática com inibidor reversível competitivo, utilizando as constantes $c_1 = \frac{10^6}{nA\Omega}$; $c_2 = 10^{-4}$; $c_3 = 0, 1$; $c_4 = \frac{10^6}{nA\Omega}$; $c_5 = 10^{-2}$ e quantidade de inibidor igual à quantidade de enzima.

5 Discussão

Claramente, o Algoritmo de Gillespie é um método direto, fornecendo a simulação exata do sistema com histórico completo. Por outro lado, essa acurácia é computacionalmente cara se quisermos trabalhar com um número significativamente maior de moléculas, já que o algoritmo guarda todos os estados do processo. A exatidão do algoritmo também não leva em consideração erros experimentais, comuns em bancadas de laboratório.

As simulações fornecem ótimas referências para analisar o comportamento de uma enzima qualitativamente, mas não podem ser comparados quantitativamente com dados práticos, o que leva à necessidade de adaptar o método para trabalhar com valores maiores, embora isso custe a precisão da simulação. Daí, surge uma variação do algoritmo denominada *tau-leaping* que, ao invés de calcular individualmente cada estado da cadeia, faz saltos de tamanho pré determinado onde várias reações ocorrem numa única atualização do vetor de estados. Esse salto é restringido pela chamada *leap condition*, que obriga que o tamanho do salto seja suficientemente pequeno para que o conjunto de reações nesse intervalo não altere significativamente o valor das constantes de taxa estocástica. O valor esperado do passo τ é tal que

$$E(\tau) < \epsilon \sum_{\mu} c_{\mu} h_{\mu}(n)$$

onde $0 < \epsilon \leq 1$ (Gillespie D.T, 2001).

Os meios intracelulares são extremamente heterogêneos, e para utilizar o algoritmo nesse contexto, esse problema precisa ser contornado, uma vez que o algoritmo considera meios homogêneos. Para lidar com isso, uma ideia seria dividir um espaço em um *grid* e considerar dois modelos com o próprio algoritmo: um para modelar cada fração do grid como um sistema isolado e outro para modelar transição de moléculas entre frações do grid.

Meios reacionais de elementos químicos podem ter sua temperatura al-

terada, seja pelas próprias reações (chamadas de reações exotérmicas e endotérmicas), pelo meio externo ou qualquer outro fator. Poderia ser possível incorporar uma mudança de temperatura no modelo refletindo nas constantes de propensão reacional. Por exemplo, considere uma simulação em que a temperatura é ideal para a enzima em questão. Um aumento de temperatura reduziria a eficiência da enzima, o que requer diminuir o coeficiente da enzima transformar substrato em produto e aumentar o coeficiente de dissociação do complexo enzima/substrato. Por outro lado, aumento de temperatura também aumenta a agitação das moléculas, o que reflete em um aumento de coeficientes de propensão de encontro entre dois reagentes. Dessa forma, a temperatura poderia virar uma variável do sistema que altera coeficientes de propensão de acordo com a dinâmica específica de cada reação.

Em uma reação enzimática, embora a concentração de produto na reação seja crescente, existem fatores que diminuem a velocidade desse crescimento, como a quantidade crescente de produto na reação, que pode saturar o meio e fatores externos que podem alterar as condições ideais da enzima. Assim, a enzima tende a perder eficiência ao longo da reação e a sua propensão de reação poderia ser diminuída em função do aumento da quantidade de produto na reação, por exemplo.

Todos esses fatores mencionados e outros inúmeros motivos dependentes do modelo específico podem ser utilizados para aproximar ainda mais o fenômeno em estudo daquilo que é observado na prática. Conhecer mais a fundo a área na qual o modelo é proposto facilita ao matemático aplicado a ajustar seu modelo conforme suas necessidades.

6 Referências

Allen, L.J.S. *An Introduction to Stochastic Processes with Application to Biology*. Second Edition, 2010.

Dobrow R.P. *Introduction to Stochastic Processes with R*. First Edition, 2016.

Gillespie D.T. *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*. *Journal of Computational Physics*, vol 22, pág 403-434, dezembro de 1976.

Gillespie D.T. *Approximate accelerated stochastic simulation of chemically reacting systems*. *Journal of Chemical Physics* vol 115 n^o 4, pág 1716-1733, julho de 2001.

Lecca P. *Stochastic chemical kinetics: a review of the modelling and simulation approaches*. *Springer, Biophysical Reviews*, 5(4), pág 323–345, dezembro de 2013.

Mohseni-Salehi F.S, Zare-Mirakabad F, Sadeghi M, Ghafouri-Fard M. *A Stochastic Model of DNA Double-Strand Breaks Repair Throughout the Cell Cycle*. *Society for Mathematical Biology*, 82:11, 2020.

Nelson D.L, Cox, M.M *Princípios de Bioquímica de Lehninger*. Editora Artmed, 6^a edição, 2014.

Oroji A, Omar M, Yarahmadian S. *An Itô stochastic differential equations model for the dynamics of the MCF-7 breast cancer cell line treated by radiotherapy*. *Journal of Theoretical Biology* 407, pág 128-137, 2016.