

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM MATEMÁTICA APLICADA

**Um modelo estocástico da regulação
gênica no operon *lac***

Rayssa Oliveira Santos

MONOGRAFIA FINAL
MAP2010 — TRABALHO DE FORMATURA

Orientador: Prof. Dr. Nelson Mugayar Kuhl

São Paulo
2024

Um modelo estocástico da regulação gênica no operon *lac*

Rayssa Oliveira Santos

Esta versão da monografia contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 12 de Dezembro de 2024.

Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão julgadora:

Prof. Dr. Nelson Mugayar Kuhl (orientador) – IME-USP

Prof.^a Dr.^a Cláudia Monteiro Peixoto – IME-USP

Prof.^a Dr.^a Maria Magdalena Rossi – IB-USP

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0
(Creative Commons Attribution 4.0 International License)*

Resumo

Rayssa Oliveira Santos. **Um modelo estocástico da regulação gênica no operon *lac***.
Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São
Paulo, São Paulo, 2024.

Este trabalho aplica um modelo estocástico da regulação gênica no operon *lac* de *E. coli* e busca verificar sua capacidade em descrever o comportamento do sistema em comparação com a abordagem determinística de cinética clássica. O modelo estocástico utilizado foi formulado por Stamatakis e Mantzaris (2009) e simulado com uma implementação própria do algoritmo de Gillespie. A modelagem determinística foi realizada com base nas equações de taxa de reação. Ambas as abordagens consideraram apenas o mecanismo de repressão mediado pelo repressor LacI, com o IPTG utilizado como indutor. A comparação entre as abordagens determinística e estocástica destacou a importância de se considerar a estocasticidade no operon *lac*. A modelagem determinística fornece uma descrição contínua da dinâmica do sistema e não leva em conta as flutuações moleculares, que podem ser significativas quando o número de moléculas é pequeno. Por outro lado, a modelagem estocástica considera essas flutuações e é capaz de capturar ativações transientes do operon *lac*, em que o repressor LacI se dissocia momentaneamente do operador, permitindo a transcrição de LacY em intervalos não regulares. Conclui-se que a modelagem estocástica pode fornecer uma visão adicional e mais detalhada do comportamento do operon *lac*, especialmente em condições de baixa concentração de IPTG, onde as flutuações moleculares podem influenciar significativamente o sistema.

Palavras-chave: Operon *lac*. Modelagem Estocástica. Algoritmo de Gillespie.

Abstract

Rayssa Oliveira Santos. **A stochastic model of gene regulation in the *lac* operon.**
Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University
of São Paulo, São Paulo, 2024.

This work applies a stochastic model of gene regulation in the *lac* operon of *E. coli* and aims to verify its ability to describe the system's behavior in comparison with the deterministic approach of classical kinetics. The stochastic model used was formulated by Stamatakis and Mantzaris (2009) and simulated with a self-developed implementation of the Gillespie Algorithm. The deterministic modeling was carried out using reaction rate equations. Both approaches considered only the repression mechanism mediated by the LacI repressor, with IPTG used as an inducer. The comparison between the deterministic and stochastic approaches highlighted the importance of considering stochasticity in the *lac* operon. The deterministic model provides a continuous description of the system's dynamics and does not account for molecular fluctuations, which can be significant when the number of molecules is small. On the other hand, the stochastic model takes these fluctuations into account and is able to capture transient activations of the *lac* operon, where the LacI repressor momentarily dissociates from the operator, allowing the transcription of LacY at irregular intervals. It is concluded that the stochastic model can provide an additional and more detailed view of the *lac* operon behavior, especially under conditions of low IPTG concentration, where molecular fluctuations can significantly influence the system.

Keywords: *lac* operon. Stochastic Modelling. Gillespie's Algorithm.

Lista de figuras

1.1	Estrutura do operon <i>lac</i> (baseada na figura 18-5 de [20]).	5
1.2	Expressão dos genes do operon <i>lac</i> . (1) Nível de transcrição basal; (2) Sem expressão; (3) Nível ativado de transcrição.	6
3.1	Simulação do modelo de Michaelis-Menten via SSA (implementação própria).	28
3.2	Simulação do modelo de Michaelis-Menten via SSA (retirado de Wilkinson [21]).	28
3.3	Simulação de 100 trajetórias do modelo de Michaelis-Menten via SSA.	29
3.4	Soluções das RREs - Espécies M_Y, Y, YI_{ex}	29
3.5	Simulação via SSA - Y	30
3.6	Soluções das RREs - Espécies M_R, R, R_2	30
3.7	Soluções das RREs - Espécies O, I e I_2R_2	31
3.8	Simulação via SSA - Espécies M_R, R e R_2	31

Lista de tabelas

2.1	Espécies químicas do operon <i>lac</i>	10
2.2	Reações químicas do operon <i>lac</i>	14
2.3	Taxas de reação. As unidades diferentes das taxas de reação são escolhidas para que, no final, todas as equações de taxa de reação resultem em uma variação de concentração expressa em $\text{nM} \cdot \text{min}^{-1}$	15

Sumário

Introdução	1
1 Operon <i>lac</i>: fundamentos biológicos	3
1.1 Expressão gênica e o fluxo de informação genética	3
1.2 Regulação da expressão gênica	4
1.3 Estrutura do operon <i>lac</i>	5
2 Modelos matemáticos	9
2.1 Equações de taxa de reação	13
2.2 Equação química mestra	17
2.2.1 Algoritmo de Gillespie	23
3 Resultados e discussão	27
Referências	33

Introdução

A expressão da informação genética codificada pelo DNA envolve diferentes etapas como a transcrição de suas sequências para RNA e a subsequente tradução desse RNA em proteínas. Contudo, nem todos os genes são expressos continuamente em todas as células e a complexidade da vida depende, em grande medida, da capacidade das células de regular a expressão gênica em diferentes combinações, momentos e locais [20]. Todos os processos que ocorrem em diferentes níveis para resultar na expressão de um gene podem ser regulados, sendo a iniciação da transcrição do RNA o principal ponto de controle [1].

A ideia de que a expressão de um gene pode ser controlada pelo produto de outro gene foi proposta por François Jacob e Jacques Monod nas décadas de 1950 e 1960, com base em seus estudos sobre o operon *lac*. Esse trabalho culminou na publicação do modelo do operon em 1961 e lhes rendeu, em 1965, o Prêmio Nobel de Fisiologia ou Medicina, compartilhado com Andre Lwoff. O operon *lac*, um dos sistemas mais emblemáticos para o estudo da regulação gênica, é um sistema regulatório em *E. coli* responsável pela metabolização da lactose. Ele é composto pelos genes *lacZ*, *lacY* e *lacA*, cuja expressão é controlada por elementos regulatórios que ajustam sua atividade de acordo com a disponibilidade de lactose e glicose no ambiente [20].

Considerando a relevância do operon *lac* como modelo para a regulação gênica, a modelagem matemática pode ser útil para compreender e descrever os processos envolvidos. Um dos principais objetivos da modelagem de sistemas biológicos é descrever com precisão seus componentes e interações, o que contribui para aprimorar o entendimento atual do sistema.

O comportamento do sistema descrito pelo operon *lac* pode ser estudado a partir da abordagem determinística de cinética clássica, caracterizada pelas equações de taxa de reação. Porém, essa abordagem não leva em conta as flutuações moleculares que ocorrem em sistemas biológicos, especialmente quando o número de moléculas de uma espécie é pequeno [21].

A regulação da expressão gênica é um dos diversos processos bioquímicos intracelulares de natureza estocástica, em que a aleatoriedade se manifesta na transcrição e tradução mesmo em condições ambientais constantes, desempenhando um papel fundamental na dinâmica dos organismos vivos [11]. Portanto, num modelo de regulação gênica as flutuações estocásticas podem ter papel significativo suficiente no sistema, para que uma análise determinística não consiga prever seu comportamento [21].

Nesse sentido, para compreender tais sistemas é interessante que a formulação utilizada seja capaz de capturar fenômenos estocásticos. Apesar da existência de teorias robustas que

seguem essa linha, os modelos realistas podem ser difíceis de serem tratados analiticamente, o que motiva a utilização de métodos computacionais [19].

Portanto, este trabalho tem como objetivo aplicar um modelo estocástico, que caracteriza o sistema como um processo de salto, para estudar a regulação gênica do operon *lac*, e observar as diferenças no comportamento do sistema em comparação com a abordagem determinística clássica. O modelo utilizado, formulado por Stamatakis e Mantzaris [17], considera apenas o mecanismo de repressão mediado pelo repressor LacI, cujo funcionamento será descrito na próxima seção. As simulações foram realizadas com uma implementação própria do algoritmo de Gillespie. Além disso, ao modelar o sistema como um processo de salto, o trabalho também visa estudar a teoria de processos estocásticos que fundamenta essa abordagem.

O Capítulo 1 apresenta os conceitos biológicos principais para a compreensão do operon *lac*. O Capítulo 2 descreve as abordagens de modelagem utilizadas e apresenta uma visão geral da teoria que as sustenta. Por fim, o último capítulo apresenta os resultados das simulações e uma breve discussão.

Capítulo 1

Operon *lac*: fundamentos biológicos

1.1 Expressão gênica e o fluxo de informação genética

O ácido desoxirribonucleico (DNA) e o ácido ribonucleico (RNA) são polímeros formados por longas cadeias de nucleotídeos, suas unidades estruturais básicas. Cada nucleotídeo é composto por três partes: um grupo fosfato, um açúcar e uma base nitrogenada. O DNA é formado por duas cadeias complementares de nucleotídeos, contendo o açúcar desoxirribose e quatro tipos de bases nitrogenadas: adenina, timina, guanina e citosina. Já o RNA é composto por uma única cadeia de nucleotídeos, possui o açúcar ribose e também quatro tipos de bases nitrogenadas, substituindo a timina pela uracila [2, 1, 20]. A sequência de bases nitrogenadas ao longo da cadeia do polímero constitui uma forma linear de informação, que caracteriza de maneira única cada ácido nucleico [2].

A expressão gênica refere-se ao processo pelo qual a informação codificada no DNA é convertida em moléculas funcionais. Um gene é definido como uma sequência específica de nucleotídeos em uma molécula de DNA que atua como unidade funcional, sendo responsável pela síntese de proteínas, RNAs estruturais ou RNAs com funções catalíticas ou reguladoras. Esse processo inicia-se com a transcrição do DNA em RNA, seguida, na maioria dos casos, pela tradução do RNA em proteínas. [2, 1].

A transcrição, primeira etapa desse processo, consiste na síntese do RNA mensageiro (mRNA) a partir do DNA do gene. Nesse contexto, a estrutura do DNA permite que ele seja utilizado como um molde para a síntese dessa nova fita complementar, que atua como portador intermediário da informação necessária para a síntese de proteínas [2, 1].

Esse fluxo de informação depende do código genético, isto é, da correspondência entre as sequências das bases nitrogenadas do DNA (e do mRNA transcrito) e a sequência de aminoácidos numa proteína [2]. A sequência linear de nucleotídeos de um gene está, portanto, associada à sequência linear de aminoácidos em uma proteína, cujas propriedades – e consequentemente sua função biológica – variam a depender de sua estrutura tridimensional [1].

A tradução do mRNA produto da transcrição em proteína é feita a partir da decodificação da sequência de nucleotídeos do mRNA em conjuntos de três nucleotídeos, denominados códons, que especificam um aminoácido ou determinam a finalização do processo de tradução. O número de combinações de trinca possíveis a partir de quatro bases é 64, todos os quais ocorrem na natureza [1]. No entanto, o código genético é degenerado, ou seja, existem códons diferentes que correspondem a um mesmo aminoácido. De fato, 64 códons codificam um total de 20 aminoácidos, em que 3 desses códons possíveis indicam o fim da tradução [2].

Em suma, dentro do contexto do dogma central da biologia molecular (ver [5, 4]), a expressão gênica pode ser representada em geral como:



1.2 Regulação da expressão gênica

Um único gene pode originar múltiplas cópias de RNA, e cada molécula de RNA pode servir como molde para a síntese de diversas proteínas idênticas. Entretanto, a taxa de expressão gênica em uma célula varia em função de suas necessidades, podendo ocorrer de forma mais ou menos eficiente e em diferentes níveis, conforme as demandas específicas da célula [1]. Além disso, devido ao alto custo energético da síntese de proteínas, a regulação da expressão gênica é fundamental para otimizar o uso da energia disponível [15].

A concentração celular de uma proteína depende de vários processos, que possuem diversos pontos passíveis de regulação, incluindo a síntese do transcrito primário de RNA (transcrição), modificações pós-transcricionais do mRNA, degradação do mRNA, síntese da proteína (tradução), modificações pós-traducionais das proteínas, direcionamento e transporte das proteínas, além da degradação proteica [15, 13]. Dentre esses processos, a iniciação da transcrição do mRNA se destaca como o principal ponto de controle [1, 13, 20]. Uma das vantagens de a transcrição ser o principal ponto de controle da expressão gênica é que, ao regular esse processo no início, evita-se o gasto de energia e recursos em etapas subsequentes, como a tradução, caso a expressão da proteína não seja necessária [20].

Nesse sentido, a transcrição de um gene depende, em grande parte, da interação entre sequências específicas de nucleotídeos no DNA e proteínas que se ligam a essas regiões. Quando ligadas a esses sítios, conhecidos como sítios regulatórios, essas proteínas podem estimular (regulação positiva) ou reprimir (regulação negativa) a expressão gênica [2]. As proteínas que promovem a transcrição são chamadas de *ativadores*, enquanto as que a inibem são conhecidas como *repressores* [20]. No caso de bactérias, essas proteínas regulatórias atuam comunicando sinais emitidos por moléculas presentes no meio externo e, tipicamente, estão localizadas nas proximidades dos genes que regulam [13].

A transcrição inicia-se quando a enzima RNA polimerase se liga ao *promotor*, uma região no início do gene que define o ponto de partida para a síntese do mRNA. Esse processo pode ser inibido se um repressor se ligar a uma região específica próxima ao promotor – que em bactérias é chamada de *operador* [15] –, sobrepondo-o e bloqueando o acesso da RNA polimerase ao sítio de ligação. Por outro lado, ativadores podem se ligar a

regiões próximas ao promotor de modo a auxiliar a ligação entre a RNA polimerase e a região promotora, contribuindo para o início da transcrição. Este mecanismo é conhecido como recrutamento ou ligação cooperativa [20].

Sítios regulatórios foram inicialmente identificados em estudos com *Escherichia coli*, que investigaram alterações na expressão de genes associados ao metabolismo de lactose para utilização como fonte de energia [2, 15]. Tais genes de *E. coli* são transcritos a partir de um promotor que é regulado por um ativador e um repressor, conforme os mecanismos descritos previamente, que serão detalhados na próxima seção.

1.3 Estrutura do operon *lac*

O *operon* é uma unidade regulatória comum em procariontes [2] que descreve um conjunto de genes que são controlados pelo mesmo operador [10]. Os genes *lacZ*, *lacY* e *lacA* constituem o *operon lac* (lactose) e estão envolvidos na metabolização da lactose em *E. coli* [20].

O gene *lacZ* codifica a enzima β -galactosidase, que atua na transformação da lactose no indutor alolactose, um isômero da lactose. O gene *lacY*, por sua vez, codifica a proteína *lactose permease*, que atua no transporte de lactose para o meio intracelular. Por fim, o gene *lacA* codifica a enzima *galactoside transacetilase*, que transfere um grupo acetil para β -galactosidase [14]. A Figura 1.1 apresenta um esquema da estrutura operon *lac*; os genes *lacZ*, *lacY* e *lacA* são transcritos como um único mRNA a partir do promotor, conforme indicado pela seta.

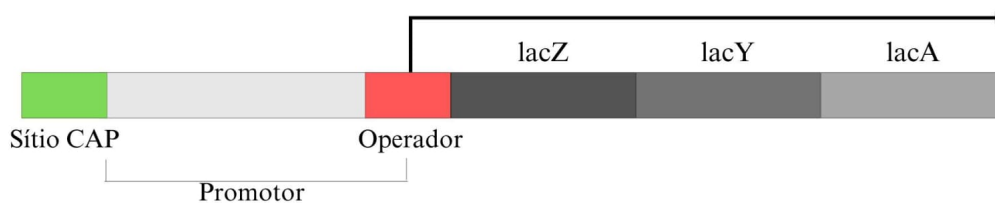


Figura 1.1: Estrutura do operon *lac* (baseada na figura 18-5 de [20]).

A glicose é a principal fonte de energia para *E. coli*, mas outros açúcares, como a lactose, também podem ser utilizados. No entanto, a metabolização da lactose requer a ativação dos genes do operon *lac*. Logo, quando a glicose está disponível em abundância, a expressão desses genes é suprimida, evitando o gasto desnecessário de recursos [15].

Portanto, os genes do operon *lac* são expressos em níveis elevados somente quando a lactose está disponível e a glicose, a fonte de energia preferencial, não está presente. A regulação dessa expressão é controlada por duas proteínas reguladoras principais: uma delas é o repressor *LacI* e a outra é o ativador conhecido como *CAP* (*Catabolite Activator Protein*). Ressalta-se que, mesmo sem a ação do ativador, a RNA polimerase é capaz de se ligar ao promotor do operon *lac*, porém essa ligação é relativamente fraca, o que limita a eficiência da transcrição, mesmo na ausência do repressor *LacI* [20]. A Figura 1.2 ilustra o mecanismo de regulação da expressão dos genes do operon *lac*.

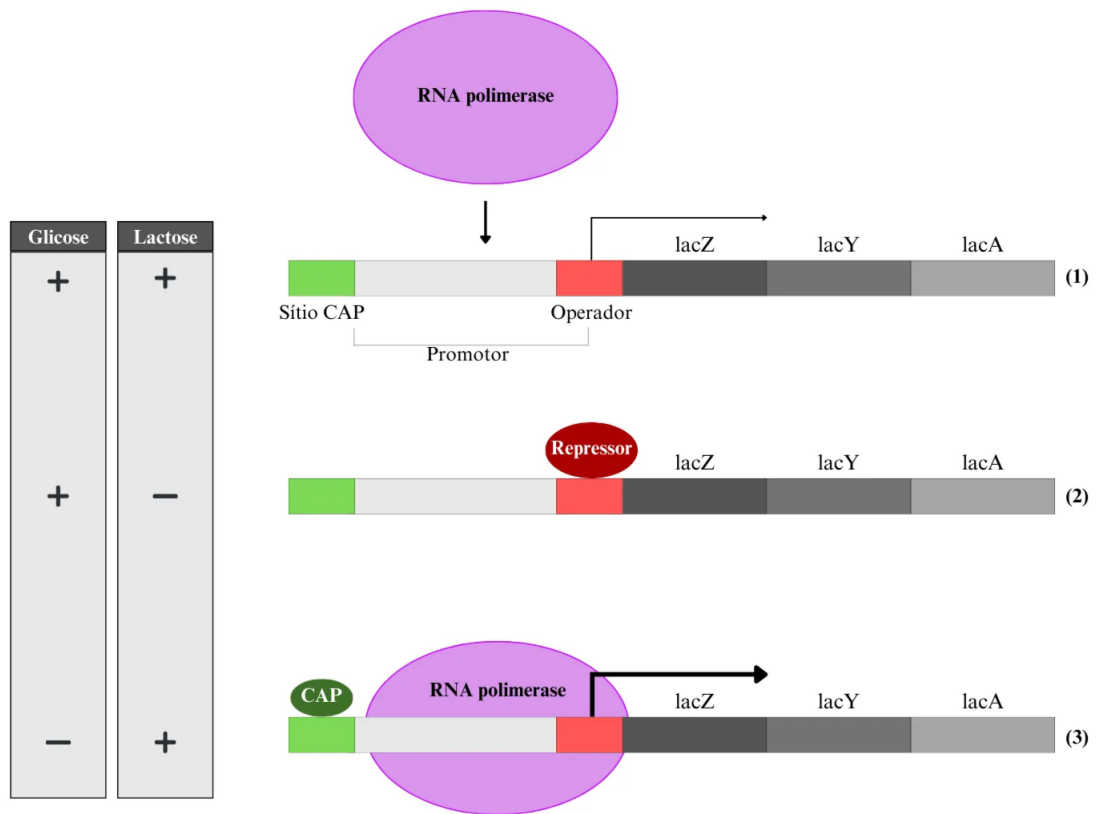


Figura 1.2: Expressão dos genes do operon *lac*. (1) Nível de transcrição basal; (2) Sem expressão; (3) Nível ativado de transcrição.

Um dímero de LacI é suficiente para que ocorra a ligação à sequência do operador e para que haja o efeito repressivo, enquanto um monômero de LacI é incapaz de se ligar ao operador ou de inibir a transcrição [17]. Além disso, o gene codificador do LacI é transcrito a partir de seu próprio promotor e é expresso constitutivamente, ou seja, é expresso de forma constante [20]. Ademais, mesmo que o operon *lac* esteja no estado reprimido, algumas moléculas de β -galactosidase e lactose permease estão presentes na célula, provavelmente sintetizadas nas raras ocasiões em que o repressor se dissocia transitoriamente do operador [15].

Na ausência de lactose no ambiente, o LacI se liga ao operador – com uma afinidade 4 milhões de vezes maior do que a que ele tem por regiões aleatórias do genoma de *E. coli* [2] – e impede que a RNA polimerase se ligue ao promotor, bloqueando a transcrição dos genes (figura 1.2 (2)). O operon *lac* possui três operadores: o principal, que se localiza entre o promotor e o primeiro gene estrutural, e os outros dois, que são também chamados de pseudo-operadores, localizados em regiões mais distantes, mas ainda assim influenciando a regulação da transcrição [15].

Por outro lado, se há lactose no meio extracelular, ela é transportada para dentro da célula e uma parte é hidrolizada – resultando em glicose e galactose –, enquanto a outra parte é transformada em alolactose pela β -galactosidase. Então, a alolactose se liga ao repressor LacI, causando uma mudança de conformação que resulta na dissociação do repressor e operador, tornando este último disponível para que a RNA polimerase se ligue

e inicie a transcrição [15].

A lactose, que é um β -galactosídeo,¹ é o substrato da enzima β -galactosidase que atua na formação da alolactose, indutor do operon *lac*. Outros β -galactosídeos, como o IPTG (isopropiltiogalactosídeo), também podem induzir o operon *lac*, mas não são metabolizados pela β -galactosidase. O IPTG é um indutor não metabolizável, amplamente utilizado em experimentos, pois ativa o operon *lac* sem ser degradado pela enzima, garantindo uma indução contínua [15].

Em condições de baixa glicose, o CAP, também conhecido como CRP (*cAMP Receptor Protein*), atua como ativador. Quando os níveis de glicose são baixos, a concentração de *cAMP* aumenta, e este se liga à proteína CRP, formando o *complexo cAMP-CRP*. Este complexo se liga ao sítio CAP (Figura 1.2 (3)), localizado acima do promotor, o que aumenta a afinidade da RNA polimerase pelo promotor e promove a transcrição dos genes do operon *lac* [20].

Ademais, a expressão inicial dos genes promove um loop de feedback positivo, pois um aumento na expressão de *lacY* facilita o transporte de lactose para o interior da célula, o que resulta em uma maior quantidade de alolactose que, por sua vez, inativa o repressor LacI, promovendo um ciclo de amplificação na expressão gênica.

Em conclusão, os genes *lacY*, *lacZ* e *lacA* são expressos somente quando, simultaneamente, a concentração de lactose é alta e a de glicose é baixa; nesse sentido, a expressão atinge seu nível máximo – isto é, o maior número de moléculas de RNA mensageiro produzido a partir do DNA, resultando na ativação máxima do gene – quando a ligação da alolactose remove a inibição do repressor, enquanto o complexo CAP-*cAMP* estimula a ligação da RNA polimerase.

¹ β -galactosídeos são compostos químicos que contêm uma molécula de galactose ligada a outro grupo, geralmente um açúcar, por uma ligação glicosídica [15].

Capítulo 2

Modelos matemáticos

A modelagem do operon *lac* foi realizada tratando as componentes descritas na seção anterior como *espécies químicas*, que interagem entre si por meio de *reações químicas*. Além disso, considerou-se que *E. coli* permanece com volume constante e que não ocorre divisão celular. Especificamente, o modelo inclui $M = 25$ reações químicas ocorrendo entre $N = 10$ espécies químicas.

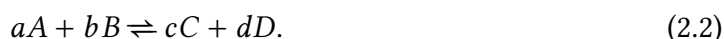
Ressalta-se que, no presente trabalho, a modelagem é focada no mecanismo regulatório exercido pelo repressor LacI, responsável por inibir a expressão dos genes do operon *lac* na ausência de lactose. O papel do ativador CAP, que facilita a transcrição em condições de baixa glicose, foi desconsiderado para simplificar o modelo. Além disso, assume-se que a transcrição dos genes de LacY – assim como dos demais genes do operon *lac* – é conduzida por um promotor mutante que é insensível à glicose. Dessa forma, os efeitos da regulação pela glicose foram deixados de lado na modelagem, permitindo concentrar o estudo apenas no papel do LacI e na interação com o indutor.

Adicionalmente, embora o LacI possa formar um tetrâmero, intensificando a repressão devido à cooperatividade entre os pseudo-operadores [15, 16], essa interação foi omitida. O IPTG (isopropiltiogalactosídeo) será utilizado como indutor do operon *lac*, já que, ao contrário da lactose, ele não é metabolizado pela β -galactosidase, permitindo uma indução contínua. Assim, o modelo se limita à análise do mecanismo de repressão mediado pelo operador principal e ao *feedback* positivo gerado pela produção de LacY.

A mudança decorrente de qualquer reação química pode ser representada por uma *equação estequiométrica* como



em que a e b representam a quantidade em mols dos reagentes A e B , respectivamente, enquanto c e d denotam a quantidade em mols dos produtos C e D , respectivamente. Os termos a , b , c e d são chamados de *coeficientes estequiométricos*. Uma reação também pode ser *reversível*, nesse caso pode ser representada com uma seta dupla, como no exemplo abaixo:



A variação na concentração dos reagentes de uma reação ao longo do tempo define a *taxa de reação*. A maioria das reações estudadas experimentalmente tem a taxa de reação dependente da concentração de um ou mais reagentes, portanto, de maneira geral, a taxa de reação pode ser expressa como uma função dessas concentrações. Entretanto, em alguns casos a taxa de reação também pode depender da concentração de outras espécies químicas intermediárias – como em reações que envolvem enzimas –, ou de espécies que não aparecem na equação estequiométrica [18].

Na maioria dos casos, a taxa de reação é proporcional ao produto das concentrações dos reagentes elevadas a potências [18]. No caso da reação 2.1, isso pode ser descrito pela expressão:

$$k \propto [A]^m[B]^n, \quad (2.3)$$

em que $[\cdot]$ denota, de forma geral, a concentração molar de uma espécie química. O expoente m indica a ordem da reação com respeito ao reagente A , enquanto o expoente n indica a ordem da reação com respeito ao reagente B . A *ordem geral da reação* (p) é a soma das ordens dos reagentes envolvidos, nesse caso $p = m + n$. Nesse sentido, uma reação de ordem zero, por exemplo, é uma reação cuja taxa de reação é independente da concentração dos reagentes.

As Tabelas 2.1, 2.2 e 2.3 apresentam, respectivamente, as espécies químicas, as reações químicas e as taxas de reação consideradas na modelagem, que foi baseada no artigo de Stamatakis e Mantzaris [17].

Símbolo	Espécie química
M_R	mRNA do repressor LacI
R	Repressor LacI
R_2	Dímero do repressor LacI
O	Operador
R_2O	Complexo repressor-operador
I	Indutor IPTG intracelular
I_{ex}	Indutor IPTG extracelular
I_2R_2	Complexo indutor-repressor
M_Y	mRNA de LacY
Y	LacY
YI_{ex}	Complexo LacY-indutor
\emptyset	Reagente ou produto genérico

Tabela 2.1: Espécies químicas do operon lac.

Expressão do gene do repressor LacI

Como mencionado anteriormente, a expressão do gene do repressor LacI (espécie química representada por R) ocorre de forma constitutiva a partir de seu próprio promotor. A transcrição do mRNA do repressor (espécie química representada por M_R) é modelada

como uma reação de ordem zero com taxa de reação k_{M_R} , conforme descrito pela equação:



A tradução de M_R foi considerada como uma reação de primeira ordem com taxa k_R :



Note que o mRNA do repressor LacI (M_R) não é consumido na tradução.

Por fim, a reação de dimerização do repressor e a correspondente reação de dissociação:



Expressão do gene de LacY

Para modelar a transcrição do mRNA de LacY (espécie M_Y), foi considerada uma reação de primeira ordem, em que a taxa de reação depende da concentração do operador livre:



A transcrição basal, que ocorre mesmo quando o gene deveria estar reprimido, como em casos de dissociação transitória do repressor, foi representada pela seguinte reação:



Para a síntese de LacY (espécie Y) a partir de seu mRNA (M_Y) foi considerada uma reação de primeira ordem:



Repressão e seus mecanismos de inibição

Assume-se que a ligação do dímero de LacI ao operador (espécie química O) ocorre conforme a estequiometria descrita pelas seguintes equações:



Com base em estudos experimentais, Stamatakis e Mantzaris [17] sugerem que a interação entre um dímero de LacI e duas moléculas de IPTG intracelular (espécie I) é suficiente para descrever o mecanismo de regulação do operon *lac*. Assim, para representar

o primeiro mecanismo de inibição da repressão, adotou-se a abordagem de reações de terceira ordem, descritas pelas equações:



O segundo mecanismo de inibição da repressão é baseado na ligação do IPTG ao complexo formado pelo repressor LacI e o operador, o que resulta na liberação do operador, permitindo a transcrição. Esse mecanismo é descrito pelas seguintes reações:



Transporte do indutor (IPTG)

Considerou-se que o transporte de IPTG do espaço extracelular para o intracelular segue a cinética de Michaelis-Menten.

O modelo de Michaelis-Menten descreve a cinética de enzimas que catalisam reações bioquímicas. Ele propõe que a enzima (E) se liga a um substrato (S) para formar um complexo enzima-substrato (SE), que posteriormente se converte em produto (P) e a enzima é liberada. Esse modelo pode ser descrito pelas reações:



Nesse contexto, LacY desempenha o papel de enzima e o IPTG extracelular (espécie I_{ex}) o papel de substrato que compõem o complexo representado pela espécie YI_{ex} :



Além disso, o transporte espontâneo de IPTG entre o meio extracelular e intracelular foi representado no modelo por reações de primeira ordem, descritas da seguinte forma:



Reações de degradação

Para todas as reações de degradação foram consideradas reações de primeira ordem:



Note que nem o operador (O) nem o complexo que ele forma com repressor (R_2O) se degradam. Ademais, a concentração total do operador ($[O_T]$) é constante e igual à soma da concentração de quando está liberado ($[O]$) e quando está ligado ao repressor ($[R_2O]$). Além disso, o indutor IPTG (I e I_{ex}) não se degrada – e foi escolhido justamente por não ser metabolizável –, porém, considerou-se que o complexo I_2R_2 pode se degradar em duas moléculas de I :



Analogamente,



Nas próximas seções, serão descritas as abordagens utilizadas para realizar as simulações. Inicialmente, são apresentadas as equações de taxa de reação, que descrevem a evolução contínua das concentrações das espécies químicas envolvidas, baseando-se na cinética clássica derivada da lei de ação das massas. Em seguida, é discutida a Equação Química Mestra, que fornece uma base teórica para descrever as variações discretas e aleatórias nas quantidades moleculares ao longo do tempo. Por fim, o Algoritmo de Gillespie é apresentado como a ferramenta principal para simulação estocástica.

As duas abordagens possuem as seguintes hipóteses em comum: considera-se que as moléculas das espécies químicas estão dispostas no espaço de modo homogêneo, num sistema de volume constante – no caso, o volume de *E. coli* considerado foi de $8 \cdot 10^{-16}$ L, como em [17] – e em equilíbrio térmico.

2.1 Equações de taxa de reação

Essa abordagem clássica de cinética química emerge da lei de ação massas, que considera que cada reação afeta a taxa instantânea de variação das espécies envolvidas de modo

Reação	Descrição
$\emptyset \xrightarrow{k_{MR}} M_R$	Transcrição do mRNA de LacI
$M_R \xrightarrow{k_R} M_R + R$	Tradução de LacI
$2R \xrightarrow{k_{R_2}} R_2$	Formação do dímero de LacI
$R_2 \xrightarrow{k_{-R_2}} 2R$	Dissociação do dímero de LacI
$O \xrightarrow{k_{s1M_Y}} O + M_Y$	Transcrição do mRNA de LacY (operador livre)
$R_2O \xrightarrow{k_{s0M_Y}} R_2O + M_Y$	Transcrição basal do mRNA de LacY
$M_Y \xrightarrow{k_Y} M_Y + Y$	Tradução de LacY
$R_2 + O \xrightarrow{k_r} R_2O$	Formação do complexo repressor-operador
$R_2O \xrightarrow{k_{-r}} R_2 + O$	Dissociação do complexo repressor-operador
$2I + R_2 \xrightarrow{k_{dr1}} I_2R_2$	Formação do complexo indutor-repressor
$I_2R_2 \xrightarrow{k_{-dr1}} 2I + R_2$	Dissociação do complexo indutor-repressor
$2I + R_2O \xrightarrow{k_{dr2}} I_2R_2 + O$	Liberação do operador pelo indutor
$I_2R_2 + O \xrightarrow{k_{-dr2}} 2I + R_2O$	Reassociação do repressor ao operador
$Y + I_{ex} \xrightarrow{k_p} YI_{ex}$	Formação do complexo LacY-IPTG
$YI_{ex} \xrightarrow{k_{-p}} Y + I_{ex}$	Dissociação do complexo LacY-IPTG (extracelular)
$YI_{ex} \xrightarrow{k_{ft}} Y + I$	Dissociação do complexo LacY-IPTG (intracelular)
$I_{ex} \xrightarrow{k_i} I$	Transporte de IPTG para o meio intracelular
$I \xrightarrow{k_e} I_{ex}$	Transporte de IPTG para o meio extracelular
$M_R \xrightarrow{\lambda_{MR}} \emptyset$	Degradação do mRNA de LacI
$R \xrightarrow{\lambda_R} \emptyset$	Degradação do monômero de LacI
$R_2 \xrightarrow{\lambda_{R_2}} \emptyset$	Degradação do dímero de LacI
$M_Y \xrightarrow{\lambda_{M_Y}} \emptyset$	Degradação do mRNA de LacY
$Y \xrightarrow{\lambda_Y} \emptyset$	Degradação de LacY
$YI_{ex} \xrightarrow{\lambda_{YI_{ex}}} I$	Degradação do complexo LacY-IPTG
$I_2R_2 \xrightarrow{\lambda_{I_2R_2}} 2I$	Degradação do repressor ligado ao indutor

Tabela 2.2: Reações químicas do operon *lac*.

proporcional ao produto da concentração das espécies reagentes [18]. Nesse contexto, um estado do sistema é definido como um vetor $\mathbf{y}(t) \in \mathbb{R}_+^N$ cujas componentes $y_i(t)$, com $i = 1, 2, \dots, N$, expressam a concentração da espécie i no tempo $t \geq 0$.

Considerando a modelagem do operon *lac*, cada componente do vetor $\mathbf{y} = [y_1, y_2, \dots, y_9]^T$ corresponde a uma das espécies envolvidas, representadas por M_R , R , R_2 , O , I , I_2R_2 , M_Y , Y e YI_{ex} , respectivamente. Note que não foi incluída a espécie I_{ex} pois sua concentração será considerada constante, uma vez que ele não é metabolizável.

Símbolo	Valor	Unidade
k_{M_R}	0.23	$\text{nM} \cdot \text{min}^{-1}$
k_R	15	min^{-1}
k_{R_2}	50	$\text{nM}^{-1} \cdot \text{min}^{-1}$
k_{-R_2}	10^{-3}	min^{-1}
k_{s1M_Y}	0.5	min^{-1}
k_{s0M_Y}	0.01	min^{-1}
k_Y	30	min^{-1}
k_r	960	$\text{nM}^{-1} \cdot \text{min}^{-1}$
k_{-r}	2.4	min^{-1}
k_{dr1}	$3 \cdot 10^{-7}$	$\text{nM}^{-2} \cdot \text{min}^{-1}$
k_{-dr1}	12	min^{-1}
k_{dr2}	$3 \cdot 10^{-7}$	$\text{nM}^{-2} \cdot \text{min}^{-1}$
k_{-dr2}	$4.8 \cdot 10^3$	$\text{nM}^{-1} \cdot \text{min}^{-1}$
k_p	0.12	$\text{nM}^{-1} \cdot \text{min}^{-1}$
k_{-p}	0.1	min^{-1}
k_{ft}	$6 \cdot 10^4$	min^{-1}
k_t	0.92	min^{-1}
λ_{M_R}	0.462	min^{-1}
λ_R	0.2	min^{-1}
λ_{R_2}	0.2	min^{-1}
λ_{M_Y}	0.462	min^{-1}
λ_Y	0.2	min^{-1}
$\lambda_{Y_{I_{ex}}}$	0.2	min^{-1}
$\lambda_{I_2R_2}$	0.2	min^{-1}

Tabela 2.3: Taxas de reação. As unidades diferentes das taxas de reação são escolhidas para que, no final, todas as equações de taxa de reação resultem em uma variação de concentração expressa em $\text{nM} \cdot \text{min}^{-1}$.

Para descrever a variação nas concentrações das diferentes espécies químicas são determinadas as equações de taxa de reação, ou *Reaction Rate Equations* (RRE), que compõem um sistema de equações diferenciais não lineares e autônomas. As equações do sistema são formuladas considerando a influência de cada reação na concentração de cada espécie química. A contribuição de uma reação para a RRE de uma espécie química depende principalmente de dois fatores: se a espécie é produzida ou consumida na reação e o tipo de reação envolvida [3, 8].

Uma reação de primeira ordem como



em que o reagente é uma espécie química genérica S_m , contribui para a equação de taxa de reação com $\alpha k[S_m]$, em que α é o coeficiente estequiométrico de S_m na reação.

Analogamente, para reações de segunda ordem que envolvem duas espécies distintas,

como



a reação contribui com $\alpha_m \alpha_n k [S_m][S_n]$ para a equação de taxa de reação.

No caso de reações envolvendo duas moléculas da mesma espécie, como na dimerização



a taxa de reação é proporcional ao quadrado da concentração do reagente, resultando na contribuição $2k[S_m]^2$. A contribuição de uma reação terá sinal positivo na RRE caso a espécie esteja sendo produzida, e sinal negativo caso esteja sendo consumida.

A espécie M_R , por exemplo, participa das reações 2.4, 2.5 e 2.23. Na primeira reação, M_R é o produto, com uma taxa de reação dada por k_{M_R} . Na segunda reação, a concentração de M_R não é afetada, pois ele aparece tanto nos reagentes quanto nos produtos com os mesmos coeficientes estequiométricos, ou seja, a quantidade de M_R permanece constante durante a reação. Na terceira reação, entretanto, M_R é consumido (degradado), com uma taxa de λ_{M_R} , o que afeta negativamente sua concentração. Como a ocorrência da primeira reação não depende das concentrações de outras espécies químicas do sistema e está relacionada apenas à taxa de reação k_{M_R} , a primeira parcela da equação de taxa de reação (RRE) será dada por k_{M_R} . A terceira reação, por sua vez, depende tanto da taxa de degradação λ_{M_R} quanto da concentração de M_R . Assim, a RRE associada à espécie M_R é dada por:

$$\frac{d[M_R]}{dt} = k_{M_R} - \lambda_{M_R}[M_R]. \quad (2.34)$$

Em relação às unidades das taxas de reação, o termo k_{M_R} na equação acima representa a taxa de produção de M_R , adicionando diretamente à variação da concentração e, por isso, possui unidade $\text{nM} \cdot \text{min}^{-1}$. Já o termo $\lambda_{M_R}[M_R]$ descreve a taxa de degradação, proporcional à concentração M_R . Para que a contribuição desta última parcela também seja expressa em $\text{nM} \cdot \text{min}^{-1}$, o parâmetro λ_{M_R} é definido com unidade min^{-1} .

A espécie química R participa das reações 2.5, 2.6, 2.7 e 2.24. Na reação 2.5, R é produzido a depender da concentração de M_R disponível e da taxa de reação k_R , o que adiciona a parcela $+k_R[M_R]$ à sua RRE. Na reação 2.6, no entanto, R é reagente com coeficiente estequiométrico igual a 2, logo, essa reação contribui para a RRE de R com a parcela $-2k_{R_2}[R]^2$. Já na reação 2.7, 2 moléculas de R são produzidas a partir de 1 de R_2 , com taxa k_{-R_2} . Portanto, a contribuição dessa reação para a RRE de R é de $+2k_{-R_2}[R_2]$. Por fim, R é degradada na reação 2.24 com taxa λ_R e a RRE associada a esta espécie química é

$$\frac{d[R]}{dt} = k_R[M_R] - 2k_{R_2}[R]^2 + 2k_{-R_2}[R_2] - \lambda_R \cdot [R]. \quad (2.35)$$

Analogamente, é possível determinar as demais equações:

$$\begin{aligned} \frac{d[R_2]}{dt} = & k_{R_2}[R]^2 - k_{-R_2}[R_2] - k_r[R_2][O] + k_{-r}(O_T - [O]) \\ & - k_{dr1}[R_2][I]^2 + k_{-dr1}[I_2R_2] - \lambda_{R_2}[R_2], \end{aligned} \quad (2.36)$$

$$\begin{aligned} \frac{d[O]}{dt} = & -k_r[R_2][O] + k_{-r}(O_T - [O]) \\ & + k_{dr2}(O_T - [O])[I]^2 - k_{-dr2}[O][I_2R_2], \end{aligned} \quad (2.37)$$

$$\begin{aligned} \frac{d[I]}{dt} = & -2k_{dr1}[R_2][I]^2 + 2k_{-dr1}[I_2R_2] - 2k_{dr2}([O_T] - [O])[I]^2 \\ & + 2k_{-dr2}[O][I_2R_2] + k_{ft}[YI_{ex}] + k_t([I_{ex}] - [I]) \\ & + 2\lambda_{I_2R_2}[I_2R_2] + \lambda_{YI_{ex}}[YI_{ex}], \end{aligned} \quad (2.38)$$

$$\begin{aligned} \frac{d[I_2R_2]}{dt} = & k_{dr1}[R_2][I]^2 - k_{-dr1}[I_2R_2] \\ & + k_{dr2}([O_T] - [O])[I]^2 - k_{-dr2}[O][I_2R_2] \\ & - \lambda_{I_2R_2}[I_2R_2], \end{aligned} \quad (2.39)$$

$$\frac{d[M_Y]}{dt} = k_{s0M_Y}([O_T] - [O]) + k_{s1M_Y}[O] - \lambda_{M_Y}[M_Y], \quad (2.40)$$

$$\frac{d[Y]}{dt} = k_Y[M_Y] + (k_{ft} + k_{-p})[YI_{ex}] - k_p[Y][I_{ex}] - \lambda_Y[Y], \quad (2.41)$$

$$\frac{d[YI_{ex}]}{dt} = -(k_{ft} + k_{-p})[YI_{ex}] + k_p[Y][I_{ex}] - \lambda_{YI_{ex}}[YI_{ex}]. \quad (2.42)$$

Esse tipo de modelagem é macroscópica em comparação com a abordagem na qual são levadas em conta as interações entre cada molécula individualmente [12], que será detalhada na próxima seção.

2.2 Equação química mestra

Essa abordagem baseia-se na probabilidade de uma reação ocorrer e na probabilidade de o sistema se encontrar em determinado estado, considerando um sistema em equilíbrio térmico e com volume constante. Neste caso, a hipótese de ter as moléculas dispostas de modo homogêneo no espaço se traduz em considerá-las distribuídas aleatoriamente de modo uniforme no volume estabelecido.

Assim como no método anterior, considera-se como objeto de estudo um sistema com N espécies químicas que participam de ao menos uma dentre M reações químicas. Porém, no caso dessa abordagem estocástica, o *estado* do sistema é um vetor $\mathbf{X} \in \mathbb{Z}_+^N$,

$$\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_N(t)]^T,$$

em que $X_i(t)$ indica a quantidade de moléculas da espécie química i no instante $t \geq 0$, com $i = 1, 2, \dots, N$.

Além disso, é conveniente definir o *vetor estequiométrico* associado à reação j , denotado

por v_j , que em cada componente indica a mudança na componente correspondente do vetor de estados após a ocorrência da reação j , isto é, após a reação, o sistema passa do estado $\mathbf{X}(t)$ para o estado $\mathbf{X}(t) + v_j$, $j = 1, 2, \dots, M$. Os vetores estequiométricos correspondem às colunas da chamada *matriz estequiométrica*, que possui dimensões $N \times M$, e descreve as mudanças nos estados das espécies químicas do sistema causadas por cada reação.

Função de propensidade e parâmetro de reação

A cada reação j no sistema, também é associada uma *função de propensidade* a_j , que representa a probabilidade de a j -ésima reação ocorrer em um intervalo de tempo infinitesimal $[t, t + dt]$. Essa probabilidade é dada por $a_j(\mathbf{X}(t))dt$, onde $dt > 0$ e sua forma depende da natureza da reação.

Para uma reação de primeira ordem, em que uma única espécie química S_m é consumida, como



a função de propensidade é dada por

$$a_j(\mathbf{X}(t)) = c_j X_m(t), \quad (2.44)$$

em que X_m indica a quantidade de moléculas da espécie S_m e c_j é o *parâmetro de reação* associado à reação j .

Uma das principais hipóteses a serem consideradas é a de que o parâmetro de reação c_j pode ser definido da seguinte maneira: $c_j dt$ é igual à probabilidade de que uma *combinação particular* de moléculas dos reagentes da reação j irão reagir no intervalo de tempo $[t, t + dt]$. Numa reação de segunda ordem envolvendo duas espécies distintas como



$c_j dt$ indica a probabilidade de uma molécula de S_m reagir com uma molécula de S_n . No entanto, sistemas de interesse possuem muito mais do que uma molécula de cada espécie; nesse sentido, se no sistema houver, no instante t , $X_1(t)$ moléculas da espécie S_m e $X_2(t)$ moléculas da espécie S_n , a probabilidade de a reação 2.45 ocorrer no intervalo $[t, t + dt]$ é dada por

$$a_j(\mathbf{X}(t))dt = c_j X_m(t) X_n(t) dt. \quad (2.46)$$

Logo, a função de propensidade para uma reação de segunda ordem entre duas espécies distintas é da forma

$$a_j(\mathbf{X}(t)) = c_j X_m(t) X_n(t). \quad (2.47)$$

No caso de reações de dimerização, em que duas moléculas da mesma espécie S_m

reagem entre si, como



a forma da função de propensidade considera o número de pares distintos de moléculas que podem ser formados. Nesse caso, a função de propensidade é dada por

$$a_j(\mathbf{X}(t)) = c_j \frac{1}{2} X_m(t)(X_m(t) - 1). \quad (2.49)$$

A função de propensidade depende da natureza da reação e o parâmetro de reação a ela associado se baseia em fundamentos físicos legítimos quando o sistema é tal que os encontros moleculares não reativos – ou elásticos, que servem para randomizar e uniformizar as posições das moléculas – ocorram muito mais frequentemente do que os encontros moleculares reativos – ou inelásticos, que alteram os níveis de população das várias espécies moleculares [6].

Dedução da Equação Química Mestra

Pode-se chegar até a expressão da Equação Química Mestra ou *Chemical Master Equation* (CME) utilizando a definição de função de propensidade e a lei das probabilidades totais. Assumindo conhecido um estado inicial $\mathbf{X}(0) = x_0$, seja $P(x, t)$ a probabilidade de $\mathbf{X}(t) = x$, ou seja, $P(x, t)$ expressa a probabilidade de o sistema se encontrar no estado x no instante t . Agora, considere os eventos:

- H_0 : o sistema se encontra no estado x no instante t , i.e., $\mathbf{X}(t) = x$;
- H_j : o sistema se encontra no estado $x - v_j$, $j = 1, 2, \dots, M$, no instante t , i.e., $\mathbf{X}(t) = x - v_j$;
- H_{M+1} : o sistema se encontra no estado em qualquer estado diferente de $x - v_j$, $j = 1, 2, \dots, M$, e de x no instante t ;
- A : o sistema se encontra no estado x no instante $t + dt$, i.e., $\mathbf{X}(t + dt) = x$.

Pela lei das probabilidades totais,

$$P(x, t + dt) = \mathbb{P}(A) = \sum_{j=0}^M \mathbb{P}(A|H_j)\mathbb{P}(H_j). \quad (2.50)$$

A definição de função de propensidade diz que

$$\mathbb{P}(A|H_j) = a_j(x - v_j)dt, \quad j = 1, 2, \dots, M. \quad (2.51)$$

Além disso,

$$\mathbb{P}(A|H_0) = 1 - \sum_{j=1}^M a_j(x)dt, \quad (2.52)$$

pois essa é a probabilidade de não ter ocorrido nenhuma reação no intervalo de tempo $[t, t + dt]$, que é o evento complementar a ter ocorrido alguma delas.

Ademais, como as únicas possibilidades para que $\mathbf{X}(t + dt) = x$ são: o sistema já estava no estado x e não ocorreu nenhuma reação no intervalo $[t, t + dt]$ ou o sistema se encontrava no estado $x - \nu_j$ e a j -ésima reação ocorreu em $[t, t + dt]$, conclui-se que

$$\mathbb{P}(A|H_{M+1}) = 0. \quad (2.53)$$

Portanto, substituindo 2.51, 2.52 e 2.53 em 2.50, vem

$$P(x, t + dt) = \mathbb{P}(A | H_0)\mathbb{P}(H_0) + \sum_{j=1}^{M+1} \mathbb{P}(A | H_j)\mathbb{P}(H_j) \quad (2.54)$$

$$= \left(1 - \sum_{j=1}^M a_j(x)dt\right)P(x, t) + \sum_{j=1}^M a_j(x - \nu_j)dt P(x - \nu_j, t) \quad (2.55)$$

$$= P(x, t) - \sum_{j=1}^M a_j(x)dt P(x, t) + \sum_{j=1}^M a_j(x - \nu_j)dt P(x - \nu_j, t). \quad (2.56)$$

Reorganizando os termos e dividindo por dt , obtém-se

$$P(x, t + dt) - P(x, t) = \sum_{j=1}^M a_j(x - \nu_j)dt P(x - \nu_j, t) - \sum_{j=1}^M a_j(x)dt P(x, t) \quad (2.57)$$

$$\Rightarrow \frac{P(x, t + dt) - P(x, t)}{dt} = \sum_{j=1}^M a_j(x - \nu_j) P(x - \nu_j, t) - a_j(x) P(x, t). \quad (2.58)$$

Por fim, tomando o limite quando $dt \rightarrow 0$, obtém-se a equação química mestra:

$$\frac{dP(x, t)}{dt} = \sum_{j=1}^M a_j(x - \nu_j)P(x - \nu_j, t) - a_j(x)P(x, t). \quad (2.59)$$

Note que isso corresponde a um sistema de equações diferenciais, uma para cada estado possível do sistema, e a solução da k -ésima equação no instante t é a probabilidade do sistema estar nesse estado particular k no tempo t .

A Equação Química Mestra no contexto de processos estocásticos

É possível caracterizar as mudanças no vetor de estados como um processo de salto e a Equação Química Mestra como uma *forward equation* no contexto de processos estocásticos.

O conjunto de todos os estados possíveis que um sistema pode assumir é o *espaço de estados* desse sistema. Considere um sistema em que seu espaço de estados S é finito ou infinito enumerável e que $\mathbf{X}(t)$ denota o estado do sistema no instante $t \geq 0$. Se o sistema se inicia em $t = 0$ estando no estado $x_0 \in S$ e permanece nesse estado até $\tau_1 > t$, quando salta

para o estado $x_1 \in S$, com $x_1 \neq x_0$, e permanece até o tempo τ_2 , quando muda novamente de estado, ele pode ser representado por

$$\mathbf{X}(t) = \begin{cases} x_0, & 0 \leq t < \tau_1, \\ x_1, & \tau_1 \leq t < \tau_2, \\ \vdots & \end{cases} \quad (2.60)$$

e 2.60 define um *processo de salto*. Além disso, quando

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} \tau_n = +\infty\right) = 1, \quad (2.61)$$

diz-se que o processo de salto é *não explosivo*. Intuitivamente, isso significa que o processo de salto só muda de estado um número infinito de vezes se o tempo total para essas mudanças também for infinito. Portanto, a partir deste ponto, consideraremos que o processo de salto em discussão é não explosivo.

Um estado é considerado *absorvente* quando, uma vez que o sistema o atinge, ele não pode mais sair desse estado. Por outro lado, se um estado $x \in S$ não é absorvente, então existe uma função $F_x(t)$, que representa a distribuição do tempo que o sistema permanece no estado x . Além disso, para cada estado x que não é absorvente, existem probabilidades de transição Q_{xy} que descrevem a chance de o sistema ir de x para outro estado $y \in S$ em um único passo. Essas probabilidades de transição possuem as seguintes propriedades: $Q(x, x) = 0$, indicando que não há transições do estado x para ele mesmo, e $\sum_{y \in S, y \neq x} Q(x, y) = 1$, garantindo que, ao sair de x , o sistema transite para algum outro estado $y \in S$.

Seja $P_{x,y} = \mathbb{P}(\mathbf{X}(t) = y \mid \mathbf{X}(0) = x)$ a *função de transição*, que descreve a probabilidade de o sistema estar no estado y no tempo t , dado que estava no estado x no tempo inicial $t = 0$. Num processo de salto, a escolha do tempo de permanência num estado e do próximo estado para o qual o sistema irá saltar é feita de modo independente. Isto é, a probabilidade de o sistema permanecer no estado x por um tempo t e depois transitar para o estado y é dada pelo produto da distribuição de tempo $F_x(t)$ e da probabilidade de transição $Q(x, y)$:

$$\mathbb{P}(\tau_1 \leq t, X(\tau_1) = y \mid \mathbf{X}(0) = x) = F_x(t)Q(x, y). \quad (2.62)$$

A *propriedade de Markov* afirma que o comportamento futuro de um processo depende apenas do seu estado atual, e não de como o processo chegou a esse estado. Isso pode ser expresso como:

$$\mathbb{P}(\mathbf{X}(t) = y \mid \mathbf{X}(s_1) = x_1, \dots, \mathbf{X}(s_n) = x_n, \mathbf{X}(s) = x) = \mathbb{P}(\mathbf{X}(t) = y \mid \mathbf{X}(s) = x), \quad (2.63)$$

em que $s_1, \dots, s_n, s < t$ e $x_1, \dots, x_n, x, y \in S$.

Um processo de salto satisfaz a propriedade de Markov se, e somente se, o tempo entre dois saltos consecutivos tiver distribuição exponencial (Ver [9]).

A evolução das probabilidades dadas pela função de transição é governada pelos *parâmetros infinitesimais* q_{xy} , que podem ser organizados em uma matriz de taxas de

transição Q . Esses parâmetros são definidos como:

$$q_{xy} = P'_{x,y}(0), \quad x, y \in S, \quad x \neq y, \quad (2.64)$$

onde $P'_{x,y}(0)$ denota a derivada da função de transição $P_{x,y}(t)$ avaliada no tempo $t = 0$. O parâmetro q_{xy} descreve a taxa de transição do estado x para o estado y em um intervalo de tempo infinitesimal.

A diagonal da matriz de parâmetros infinitesimais é dada por:

$$q_{xx} = - \sum_{y \neq x} q_{xy}, \quad (2.65)$$

o que garante que, para cada estado $x \in S$, a soma das taxas de transição de x para todos os outros estados $y \neq x$ é igual a $-q_{xx}$. Em outras palavras, a soma das taxas de transição de x para todos os outros estados precisa ser a negativa da taxa de transição de x para ele mesmo (ou seja, q_{xx}), para garantir que, a cada transição, a soma das probabilidades de todos os estados seja sempre igual a 1.

A partir da definição dos parâmetros infinitesimais, podemos derivar a *equação de Chapman-Kolmogorov*, que descreve como as probabilidades de transição se comportam ao longo do tempo.

Considere um processo de salto $\mathbf{X}(t)$ que satisfaz a propriedade de Markov (2.63), possui espaço de estados S finito e $x, y, z \in S$. Então,

$$\mathbb{P}(\mathbf{X}(t) = z, \mathbf{X}(t+s) = y \mid \mathbf{X}(0) = x) = P_{xz}(t)P_{zy}(s), \quad s, t \geq 0. \quad (2.66)$$

O primeiro fator, $P_{xz}(t)$, representa a probabilidade de transição do estado inicial x para o estado intermediário z em um tempo t . O segundo fator, $P_{zy}(s)$, é a probabilidade de transição do estado z para o estado final y no intervalo de tempo subsequente s . A decomposição dessa probabilidade conjunta expressa a independência das transições futuras em relação aos estados anteriores, característica dos processos de Markov.

Além disso, a probabilidade de transição de x para y no tempo $t+s$ pode ser expressa como a soma das probabilidades de transição de x para todos os estados intermediários z no tempo t , seguida pela transição de z para y no tempo s . Ou seja, temos a seguinte relação:

$$P_{xy}(t+s) = \sum_z \mathbb{P}(\mathbf{X}(t) = z, \mathbf{X}(t+s) = y), \quad (2.67)$$

e, portanto, chega-se à *equação de Chapman-Kolmogorov*:

$$P_{xy}(t+s) = \sum_z P_{xz}(t)P_{zy}(s). \quad (2.68)$$

Derivando 2.68 com respeito a s , obtém-se

$$P'_{x,y}(t+s) = \sum_z P_{xz}(t)P'_{z,y}(s), \quad t \geq 0. \quad (2.69)$$

Em particular, ao avaliar a expressão acima em $s = 0$ e expressá-la em termos dos parâmetros infinitesimais, obtemos a chamada *forward equation*:

$$P'_{xy}(t) = \sum_z P_{xz}(t)P'_{zy}(0) = \sum_z P_{xz}(t)q_{zy}, \quad t \geq 0. \quad (2.70)$$

A expressão $P'_{xy}(t)$ representa a derivada da função de transição de x para y em relação ao tempo, enquanto q_{zy} denota as taxas de transição entre os estados z e y em um intervalo infinitesimal, conforme definido anteriormente. Esta equação descreve a evolução temporal das probabilidades de transição no processo de salto.

Considere agora que x é um estado inicial para a Equação Química Mestra. Reescrevendo 2.59 na notação apresentada, obtém-se

$$P'_{x,y} = \sum_{j=1}^M P_{x,y-v_j}(t)a_j(y - v_j) + P_{x,y} \left(- \sum_{k=1}^M a_k(y) \right) \quad (2.71)$$

$$= \sum_{z \in S} P_{x,z}(t)q_{z,y} \quad (2.72)$$

com

$$q_{z,y} = \begin{cases} a_j(z), & s = y - v_j, \quad 1 \leq j \leq M \\ - \sum_{k=1}^M a_k(y), & z = y \\ 0, & \text{caso contrário.} \end{cases} \quad (2.73)$$

Desse modo, a Equação Química Mestra tem a forma de uma equação *forward* e o sistema é descrito como um processo de salto que verifica a propriedade de Markov.

As abordagens vistas – via equações de taxa de reação e pela Equação Química Mestra – estão inter-relacionadas no limite termodinâmico. À medida que o número de moléculas de cada espécie e o volume que as contém tendem ao infinito de maneira que as concentrações moleculares se aproximam de valores finitos, a modelagem estocástica converge para a modelagem determinística [6, p. 407].

Enfim, resolver todas as equações diferenciais dadas pela Equação Química Mestra para determinar uma distribuição de probabilidade para todos os estados possíveis pode ser extremamente complicado dado o grande número de equações. Nesse sentido, surge o *Stochastic Simulation Algorithm* (SSA), ou Algoritmo de Gillespie [6, 7], que computa realizações do vetor de estados do sistema $\{t, \mathbf{X}(t)\}$ de modo que a probabilidade dessa amostra ter sido sorteada reflita a probabilidade correspondente dada pela Equação Química Mestra.

2.2.1 Algoritmo de Gillespie

O método numérico proposto por Gillespie [6, 7] é equivalente à Equação Química Mestra, embora esta última nunca seja utilizada explicitamente. De acordo com o autor, com exceção da necessidade de depender de alguma sub-rotina computacional para a geração de números aleatórios que seguem uma distribuição uniforme, o procedimento computacional não introduz aproximações e incorpora todas as correlações estatísticas inerentes e flutuações que são desconsideradas na formulação determinística.

O algoritmo opera com base em dois componentes principais: a escolha de qual reação ocorrerá e a determinação de quando essa reação ocorrerá. O processo se dá em duas etapas:

1. *Cálculo das funções de propensidade*: O algoritmo calcula a funções de propensidade de cada uma das reações do sistema com base na quantidade de moléculas das espécies químicas envolvidas.
2. *Escolha da reação e do tempo de transição*: A escolha da reação é feita com base nas probabilidades associadas às funções de propensidade. O tempo até a ocorrência da reação é determinado por uma distribuição exponencial, com parâmetro igual à soma das funções de propensidade de todas as reações.

Primeiramente, introduz-se a função $P_0(\tau|x, t)$, que denota a probabilidade de que, dado que o sistema está no estado $X(t) = x$, nenhuma reação ocorra no intervalo de tempo $[t, t + \tau)$.

O comportamento do sistema ao longo do intervalo $[t, t + \tau + d\tau)$ pode ser modelado de forma independente nos intervalos $[t, t + \tau)$ e $[t + \tau, t + \tau + d\tau)$ devido à natureza do processo estocástico em questão. Como explicado na seção anterior, trata-se de um processo de salto que verifica a propriedade de Markov. Desse modo, a probabilidade de nenhuma reação ocorrer no intervalo $[t, t + \tau + d\tau)$ é o produto das probabilidades de nenhum evento ocorrer nos intervalos $[t, t + \tau)$ e $[t + \tau, t + \tau + d\tau)$. A expressão para isso é:

$$P_0(\tau + d\tau|x, t) = P_0(\tau|x, t) \left(1 - \sum_{k=1}^M a_k(x) d\tau \right) \quad (2.74)$$

em que $\sum_{k=1}^M a_k(x)$ é a soma das funções de propensidade de todas as reações. Chegamos então na seguinte equação diferencial linear:

$$\frac{P_0(\tau + d\tau|x, t) - P_0(\tau|x, t)}{d\tau} = -a_{\text{sum}}(x)P_0(\tau|x, t), \quad (2.75)$$

em que $a_{\text{sum}}(x) = \sum_{k=1}^M a_k(x)$ é a soma total das funções de propensidade. Portanto, a probabilidade de que nenhuma reação tenha ocorrido até o tempo τ é dada pela solução dessa equação diferencial, que é

$$P_0(\tau|x, t) = e^{-a_{\text{sum}}(x)\tau}. \quad (2.76)$$

Com base nisso, é possível determinar a função de densidade conjunta das variáveis aleatórias que representam o tempo até a próxima reação e qual reação ocorrerá. Denotada por $p(\tau, j|x, t)$, essa função é definida da seguinte forma: dado que o sistema está no estado x no instante t , isto é, $\mathbf{X}(t) = x$, então $p(\tau, j|x, t)d\tau$ representa a probabilidade de que a próxima reação será a reação j e que ela ocorrerá no intervalo de tempo $[t + \tau, t + \tau + d\tau)$.

Do mesmo modo, a probabilidade de que nenhuma reação ocorra no intervalo $[t, t + \tau)$ e de que a j -ésima reação ocorra no intervalo $[t + \tau, t + \tau + d\tau)$ é calculada como o produto das probabilidades desses dois eventos. Assume-se que o intervalo $d\tau$ é suficientemente pequeno para que, no máximo, apenas uma reação possa ocorrer durante

esse período. Então,

$$p(\tau, j | x, t) d\tau = P_0(\tau | x, t) a_j(x) d\tau. \quad (2.77)$$

Tem-se que $p(\tau, j | x, t)$ é a função de densidade conjunta de duas variáveis aleatórias: o tempo até a próxima reação e o índice da próxima reação. A Equação 2.77 demonstra que essa densidade conjunta pode ser escrita como o produto de duas funções de densidade individuais. A variável aleatória associada ao índice da próxima reação é uma variável discreta que toma valores em $1, 2, \dots, M$, com a probabilidade de o índice ser j proporcional a $a_j(x)$. Já o tempo até a próxima reação é uma variável aleatória contínua com densidade exponencial descrita por $a_{\text{sum}}(x)e^{-a_{\text{sum}}(x)\tau}$.

O algoritmo de Gillespie está apresentado no algoritmo 1. A próxima seção discute como foram realizadas as simulações utilizando o algoritmo de Gillespie, além de apresentar o solucionador empregado para as equações de taxa de reação.

Algoritmo 1 Algoritmo de Gillespie ou *Stochastic Simulation Algorithm (SSA)*

- 1: Compute $\{a_k(\mathbf{X}(t))\}_{k=1}^M$ e $a_{\text{sum}} := \sum_{k=1}^M a_k(\mathbf{X}(t))$.
 - 2: Gere dois números aleatórios ξ_1 e ξ_2 independentes com distribuição uniforme no intervalo $(0,1)$.
 - 3: Tome j como o menor número inteiro que satisfaz $\sum_{k=1}^j a_k(\mathbf{X}(t)) > \xi_1 a_{\text{sum}}(\mathbf{X}(t))$.
 - 4: Tome $\tau = \ln(1/\xi_2)/a_{\text{sum}}(\mathbf{X}(t))$.
 - 5: Atualize: $\mathbf{X}(t + \tau) = \mathbf{X}(t) + \mathbf{v}_j$ e $t = t + \tau$.
 - 6: Retorne ao passo 1.
-

Capítulo 3

Resultados e discussão

As simulações foram feitas utilizando uma implementação própria do algoritmo de Gillespie (SSA) (1) em C++, que se encontra disponível [neste repositório](#). Utilizou-se o servidor de processamento da Rede IME, mais especificamente a máquina *brucutu*, que possui dois processadores Intel Xeon Gold 6148, com 20 núcleos, 40 threads e 512GB de memória RAM.

Para resolver o sistema de equações de taxa de reação do operon *lac*, utilizou-se a função `solve_ivp` da biblioteca *SciPy* do Python, configurada para o método numérico Radau. Este método de Runge-Kutta implícito é recomendado para sistemas de equações diferenciais ordinárias *stiff*, como o sistema em questão. Os parâmetros de tolerância relativa e absoluta foram configurados como 10^{-6} e 10^{-9} , respectivamente.

Simulações iniciais

Inicialmente, o SSA foi aplicado a estudos de caso apresentados no livro de Wilkinson [21]. Esses estudos consistem em sistemas mais simples, que possuem gráficos de referência disponíveis, tornando-os adequados para validar a implementação realizada. As condições iniciais e as taxas das reações utilizadas para as simulações foram as mesmas desta referência.

Um dos estudos de caso considerados foi o do modelo de Michaelis-Menten, representado pelas reações já apresentadas em 2.17. A Figura 3.1 apresenta o resultado de uma simulação de 50 segundos desse sistema feita com a implementação própria, enquanto o gráfico de uma simulação presente em Wilkinson [21] é exibido na Figura 3.2. Para analisar o comportamento do sistema em múltiplas simulações, a Figura 3.3 exibe os resultados de 100 execuções do modelo de Michaelis-Menten.

Além desse modelo, foram realizadas simulações dos estudos de caso *Dimerisation Kinetics* e *An Autoregulatory Genetic Network*, que, assim como o modelo de Michaelis-Menten, apresentaram resultados consistentes com os gráficos disponíveis na referência [21].

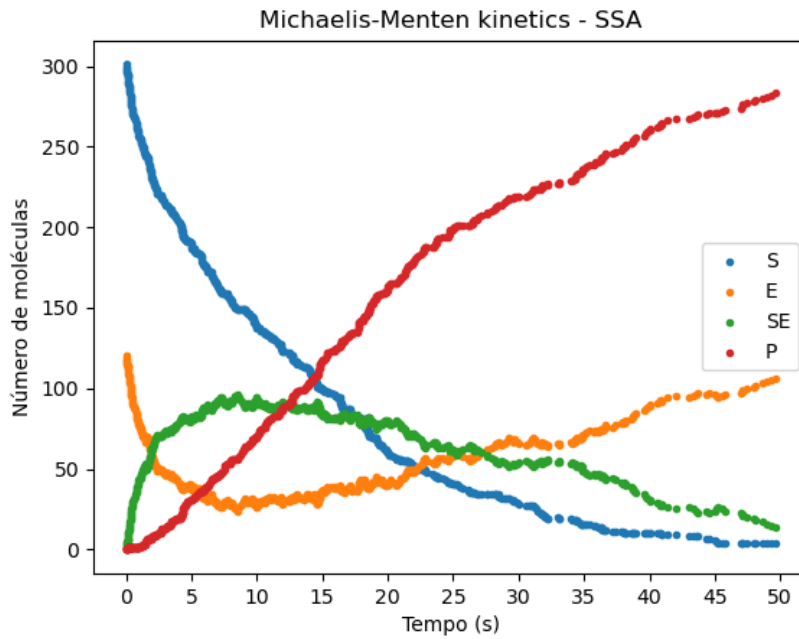


Figura 3.1: Simulação do modelo de Michaelis-Menten via SSA (implementação própria).

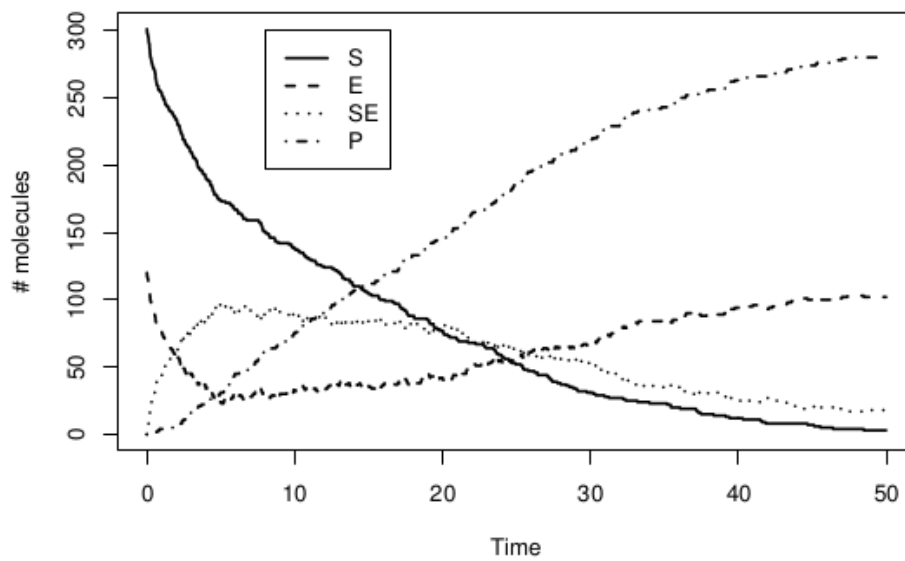


Figura 3.2: Simulação do modelo de Michaelis-Menten via SSA (retirado de Wilkinson [21]).

Simulações do operon *lac*

Nesta seção, são apresentados os resultados das simulações realizadas para o operon *lac*, considerando a concentração de IPTG como $I_{ex} = 20 \mu M$, tanto para a abordagem determinística quanto para a estocástica. Para possibilitar uma comparação entre os dois métodos, as taxas de reação foram convertidas em parâmetros de reação, e as concentrações foram traduzidas para número de moléculas, utilizando o número de Avogadro (para mais

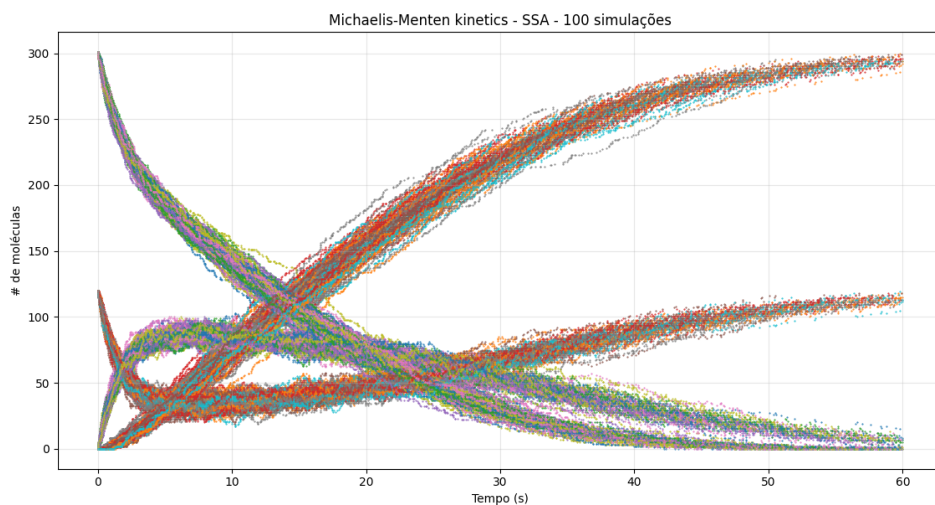


Figura 3.3: Simulação de 100 trajetórias do modelo de Michaelis-Menten via SSA.

detalhes ver [8]).

Na abordagem determinística, as concentrações das espécies químicas variam de maneira suave e contínua ao longo do tempo. Os resultados fornecidos pelas RREs mostram um aumento na quantidade de LacY (Y) ao longo do tempo (Figura 3.4). No entanto, após atingir um pico, a concentração de LacY diminui gradualmente, devido à compensação entre os processos de síntese e degradação de Y .

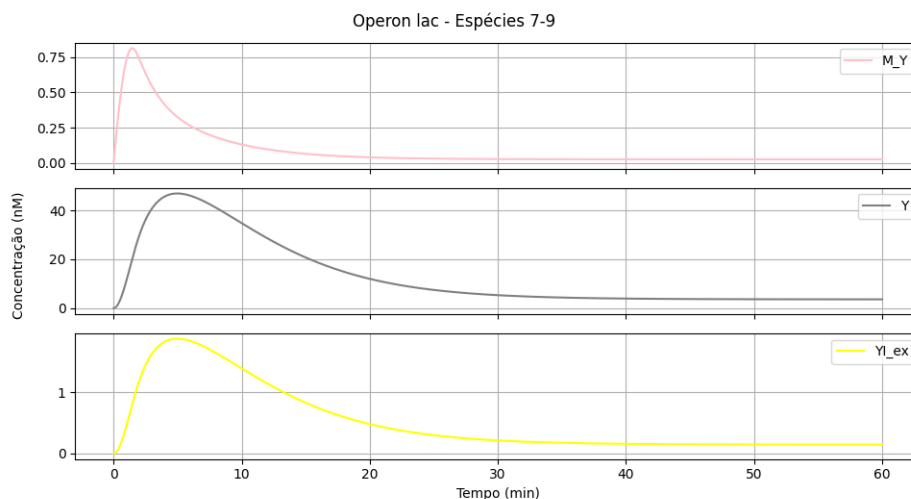


Figura 3.4: Soluções das RREs - Espécies M_Y , Y , YI_{ex}

Diferentemente do que foi visto na modelagem por RREs, uma simulação do modelo estocástico pode apresentar mais de um pico na concentração de LacY, como mostrado na Figura 3.5, indicando que o operon pode ser ativado mais de uma vez durante a simulação. Esse é um exemplo de como a estocasticidade não apenas introduz variações temporárias,

mas também pode levar a dinâmicas qualitativamente diferentes, como oscilações ou ativações.

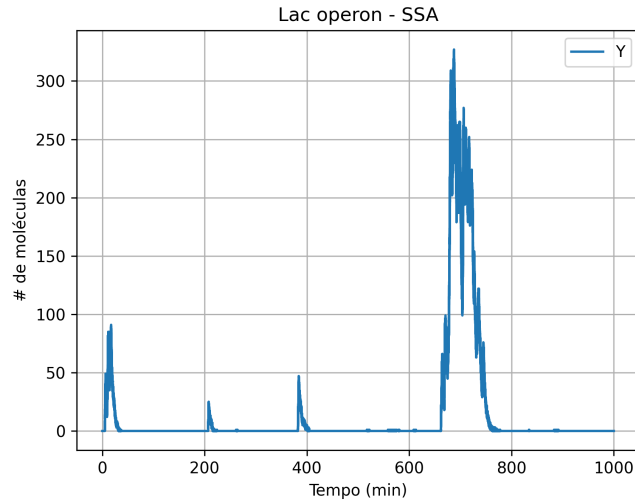


Figura 3.5: Simulação via SSA - Y

Além da concentração de LacY, o comportamento das outras espécies químicas também segue padrões distintos nas duas abordagens. Na modelagem determinística, as concentrações de M_R , R e R_2 variam suavemente ao longo do tempo, com a concentração de R_2 aumentando à medida que o sistema atinge um equilíbrio, refletindo a repressão do operon *lac* (Figura 3.6). A espécie O se mantém em níveis baixos, uma vez que a ligação do repressor ao operador impede a transcrição de *LacY* (Figura 3.7).

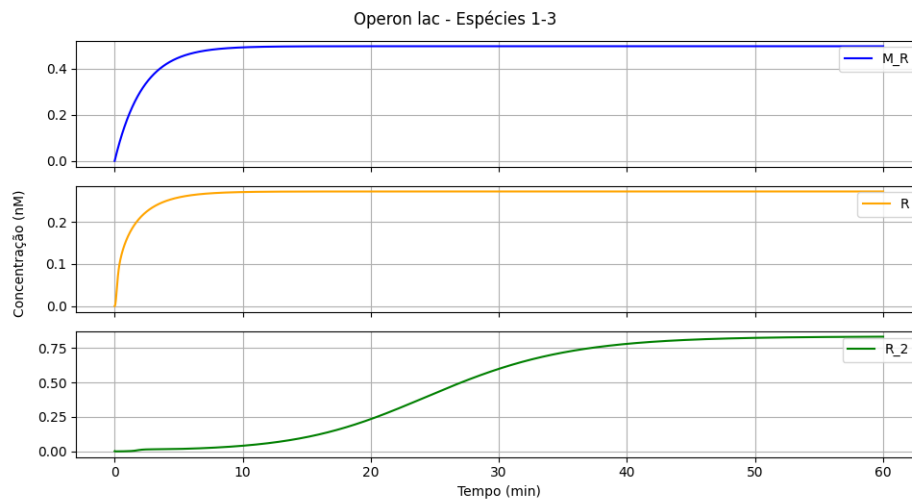


Figura 3.6: Soluções das RREs - Espécies M_R , R , R_2

Na modelagem estocástica, por outro lado, o comportamento das espécies é influenciado pelas flutuações moleculares. Diferentemente da abordagem determinística, as concentrações de R_2 (figura 3.8) e O oscilam ao longo do tempo devido à natureza probabilística

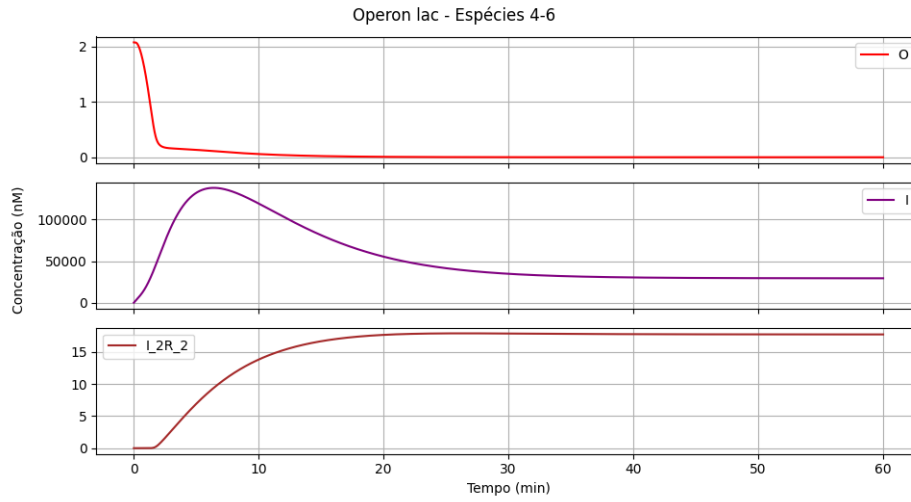


Figura 3.7: Soluções das RREs - Espécies O , I e I_2R_2

das interações entre as espécies químicas. Isso implica que o operador pode alternar entre estados reprimidos e ativados, mesmo sob condições fixas de IPTG.

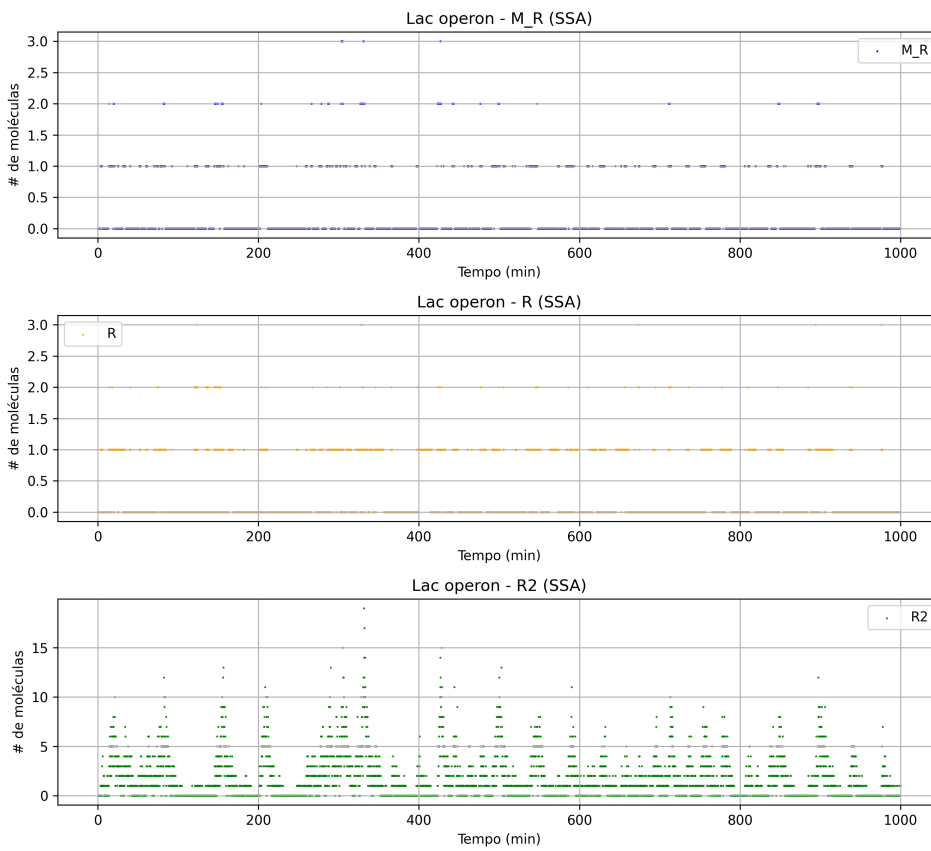


Figura 3.8: Simulação via SSA - Espécies M_R , R e R_2

Na comparação entre os métodos determinístico e estocástico, observou-se que o

modelo dado pelas RREs não captura as flutuações estocásticas que surgem em sistemas com números reduzidos de moléculas, em que as interações entre as espécies podem ser influenciadas por eventos aleatórios.

Por outro lado, a modelagem estocástica leva em consideração essas flutuações moleculares e, como resultado, é capaz de capturar picos e quedas nas concentrações das espécies, refletindo o comportamento probabilístico dos processos de transição entre os estados. Essa diferença é especialmente importante para sistemas como o operon *lac*, em que a quantidade de moléculas de determinadas espécies, como o repressor LacI e o indutor IPTG, pode ser baixa e as flutuações moleculares podem ter um impacto significativo no comportamento global do sistema.

Em resumo, os resultados obtidos pelas abordagens determinística e estocástica destacam a importância de considerar a estocasticidade no estudo do operon *lac*. Para aprofundar a compreensão do sistema, futuras simulações poderiam investigar a influência de diferentes indutores, explorar variações nas configurações do operon e considerar outros mecanismos regulatórios, a fim de expandir a compreensão do sistema.

Referências

- [1] Bruce Alberts et al. *Biologia Molecular da Célula*. Artmed, 2017. ISBN: 9780815344322 (ver pp. 1, 3, 4).
- [2] Jeremy M. Berg, John L. Tymoczko e Lubert Stryer. *Biochemistry*. W. H. Freeman e Company, 2007. ISBN: 0-7167-8724-5 (ver pp. 3–6).
- [3] J. M. Bower e H. Bolouri, ed. *Computational Modeling of Genetic and Biochemical Networks*. Cambridge, Massachusetts: MIT Press, 2000 (ver p. 15).
- [4] Francis H. C. Crick. “Central Dogma of Molecular Biology”. Em: *Nature* 227.5258 (1970), pp. 561–563. DOI: [10.1038/227561a0](https://doi.org/10.1038/227561a0) (ver p. 4).
- [5] Francis H. C. Crick. “On Protein Synthesis”. Em: *Symposia of the Society for Experimental Biology*. Vol. 12. Cambridge University Press, 1958, pp. 138–163 (ver p. 4).
- [6] Daniel T. Gillespie. “A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions”. Em: *Journal of Computational Physics* 22 (1976), pp. 403–434 (ver pp. 19, 23).
- [7] Daniel T. Gillespie. “Exact Stochastic Simulation of Coupled Chemical Reactions”. Em: *Journal of Physical Chemistry* 81 (1977), pp. 2340–2361 (ver p. 23).
- [8] Desmond J. Higham. “Modelling and Simulating Chemical Reactions”. Em: *SIAM Review* 50.2 (2008), pp. 347–368 (ver pp. 15, 29).
- [9] Paul G. Hoel, Sidney C. Port e Charles J. Stone. *Introduction to Stochastic Processes*. 9th. Boston: Houghton Mifflin Company, 1972 (ver p. 21).
- [10] F. Jacob et al. “L’operon: groupe de gènes à expression coordonnée par un opérateur”. Em: *Comptes Rendus Hebdomadaires des Séances de L’Académie des Sciences. Série D: Sciences naturelles* 250 (1960), pp. 1727–1729 (ver p. 5).
- [11] Mads Kærn et al. “Stochasticity in gene expression: from theories to phenotypes”. Em: *Nature Reviews: Genetics* (2005). DOI: [10.1038/nrg1615](https://doi.org/10.1038/nrg1615) (ver p. 1).
- [12] E. Klipp et al. *Systems Biology in Practice: Concepts, Implementation and Application*. New York: Wiley-VCH, 2005 (ver p. 17).
- [13] David S. Latchman. *Gene Control*. 2th. London: Taylor & Francis Group, 2015. ISBN: 978-0-8153-4503-9 (ver p. 4).
- [14] Benno Müller-Hill. *The lac operon: a Short History of the Genetic Paradigm*. 1th. New York: Walter de Gruyter, 1996. ISBN: 3-11-014830-7 (ver p. 5).

- [15] David L. Nelson e Michael M. Cox. *Lehninger Principles of Biochemistry*. 5th. New York: W.H. Freeman e Company, 2013. ISBN: 978-0-7167-7108-1 (ver pp. 4–7, 9).
- [16] S. Oehler et al. “Quality and position of the three lac operators of *E. coli* define efficiency of repression”. Em: *EMBO Journal* 13 (1994), pp. 3384–3385 (ver p. 9).
- [17] Michail Stamatakis e Nikos V. Mantzaris. “Comparison of Deterministic and Stochastic Models of the lac Operon Genetic Network”. Em: *Biophysical Journal* 96 (fev. de 2009), pp. 887–906. DOI: [10.1016/j.bpj.2008.10.028](https://doi.org/10.1016/j.bpj.2008.10.028) (ver pp. 2, 6, 10, 11, 13).
- [18] Jeffrey I. Steinfeld, Joseph S. Francisco e William L. Hase. *Chemical Kinetics and Dynamics: Second Edition*. Prentice-Hall, 1998. ISBN: 0-13-737123-2 (ver pp. 10, 14).
- [19] David J. Warne, Ruth E. Baker e Matthew J. Simpson. “Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art”. Em: *The Royal Society* (2019) (ver p. 2).
- [20] James D. Watson et al. *Molecular Biology of the Gene*. Benjamin Cummings, 2004. ISBN: 0-321-22368-3 (ver pp. 1, 3–7).
- [21] Darren J. Wilkinson. *Stochastic Modelling for Systems Biology*. CRC Press, 2019. ISBN: 9781138549289 (ver pp. v, 1, 27, 28).