



UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM DE SISTEMAS
COMPLEXOS

GUILHERME GIOVANINI

Modelos binários estocásticos para a expressão gênica: uma análise comparativa das propriedades de ruído e uma investigação de respostas a terapias gênicas

São Paulo

2022

GUILHERME GIOVANINI

Modelos binários estocásticos para a expressão gênica: uma análise comparativa das propriedades de ruído e uma investigação de respostas a terapias gênicas

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos.

Área de concentração: Sistemas Complexos

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 26 de janeiro de 2022.

Orientador: Prof. Dr. Alexandre Ferreira Ramos

São Paulo

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha catalográfica elaborada pela Biblioteca da Escola de Artes, Ciências e Humanidades,
com os dados inseridos pelo(a) autor(a)
Brenda Fontes Malheiros de Castro CRB 8-7012; Sandra Tokarevicz CRB 8-4936

Giovanini, Guilherme

Modelos binários estocásticos para a expressão gênica: uma análise comparativa das propriedades de ruído e uma investigação de respostas a terapias gênicas / Guilherme Giovanini; orientador, Alexandre Ferreira Ramos. -- São Paulo, 2022.
118 p: il.

Dissertacao (Mestrado em Ciencias) - Programa de Pós-Graduação em Modelagem de Sistemas Complexos, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, 2022.
Versão corrigida

1. Dinâmica estocástica. 2. Regulação gênica. 3. Terapêutica epigenética. 4. Farmacoterapia multidrogas. I. Ramos, Alexandre Ferreira, orient. II. Título.

Dissertação de autoria de Guilherme Giovanini, sob o título “**Modelos binários estocásticos para a expressão gênica: uma análise comparativa das propriedades de ruído e uma investigação de respostas a terapias gênicas**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos, na área de concentração de Sistemas Complexos, aprovada em 26 de janeiro de 2022 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Alexandre Ferreira Ramos
Universidade de São Paulo
Presidente

Prof. Dr. Paulo Fernando de Arruda Mancera
Universidade Estadual Paulista

Prof. Dr. Marcelo da Silva Reis
Instituto Butantan

Prof. Dr. Humberto Miguel Garay Malpartida
Universidade de São Paulo

*À minha mãe e ao meu pai,
Zilda e Arnaldo,
e à minha noiva
Camila.*

Agradecimentos

A conclusão desta dissertação foi devido ao trabalho, aos esforços e ao envolvimento de muitos, direta ou indiretamente. É com muito prazer e sinceridade que agradeço a todos que contribuíram com esta minha realização.

Ao Prof. Alexandre, agradeço a orientação, paciência, confiança e tempo investido no meu processo de aprendizado e na minha formação acadêmica. Sua experiência, seus incentivos e recomendações contribuíram significativamente para o meu conhecimento científico e cultural.

Agradeço à CAPES e ao PPG-SCX pelos auxílios financeiros recebidos. Agradeço ao IBILCE-UNESP pela minha formação básica. Agradeço à EACH-USP por todos os recursos disponibilizados e aos seus servidores, que mesmo durante a pandemia, mantiveram, com êxito, o funcionamento da universidade e sua comunidade segura. Destaco aqui meus agradecimentos: à Kelly, à Vanessa, ao Jonathan e ao Flávio pelos esclarecimentos e prontidão na realização de trâmites administrativos; aos Profs. M. Lauretto e C. Molina pela disponibilidade e todo o suporte que precisei da coordenação do programa; e aos Profs. Camilo Neto, Diego Gonçalves, Fernando Ferreira, Maria Eliza Bernardes, Masayuki Hase e José Ricardo Mendonça pelas disciplinas ministradas e que foram relevantes na minha formação. Agradeço ao ICESP pelos recursos e o espaço para pesquisa, à Cris, aos Profs. Roger Chammas e Laura Sichero e aos colegas Tharcisio e Luciana. Agradeço à Profa. Marsha Rosner e aos Profs. John Reinitz e Gábor Balázs.

Agradeço à minha noiva Camila pela alegria, cumplicidade, paciência, pelo total apoio e por me cobrir toda vez que não pude dar a devida atenção aos nossos Vincent, Tarsila e Garbis. Agradeço aos meus pais e à minha família: Giovana, Jean, Fátima, Adolfo, Cadú e Baxula. Obrigado pelo amor (que é recíproco!), carinho e zelo.

Agradeço a todos os meus amigos do laboratório AMPhyBio, aqui destaco Alan, Léo, Suzy e Cauê, por toda ajuda que prestaram e bom convívio; aos amigos do programa de mestrado: Luis Amato Neto (*in memoriam*), Fábio e Milton; e aos amigos de longa data: Henrique, Zizo, Peres, XV, Fejão, Kim e Nosso.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*“...with four parameters I can fit an elephant,
and with five I can make him wiggle his trunk.”*

John von Neumann (1903 - 1957)

Resumo

GIOVANINI, Guilherme. **Modelos binários estocásticos para a expressão gênica: uma análise comparativa das propriedades de ruído e uma investigação de respostas a terapias gênicas.** 2022. 118 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2022.

Realizamos uma comparação das propriedades de ruído exibidas por dois modelos de gene binário estocástico no estado estacionário: um gene externamente regulado e um gene autorreprimido. Esses modelos descrevem a dinâmica das distribuições de probabilidade que governam duas variáveis aleatórias: o número de proteínas e o estado do gene (ligado ou desligado). Quantificamos o ruído no número de proteínas por meio de seu fator de Fano e escrevemos essa quantidade como uma função da covariância entre as duas variáveis aleatórias. As distribuições de probabilidade que governam o número de produtos gênicos pode se encontrar em regimes super-Fano, Fano ou sub-Fano, onde as covariâncias são, respectivamente, positiva, nula e negativa. Este último regime é exclusivo do gene autorreprimido e mostramos as condições em que o fator de Fano é suficiente para classificar as flutuações na expressão gênica. Apresentamos também as condições no qual o ruído no número de produtos gênicos gerados por ambos os modelos de gene são quantitativamente similares. Isto é importante para a inferência da regulação gênica a partir do ruído em dados quantitativos na expressão gênica. Os resultados apresentados contribuem para a classificação da função do ruído em sistemas biológicos por demonstrar teoricamente os mecanismos que sustentam a alta precisão na expressão de um gene autorreprimido em comparação de um externamente regulado. Por último, utilizamos um modelo de gene binário estocástico dependente do tempo para investigar a resposta ao tratamento de um gene alvo. Para estabelecer escalas de tempo biologicamente relevantes para os parâmetros do modelo, selecionamos o gene RKIP e duas drogas não específicas conhecidas por alterar os níveis de RKIP em células cancerosas. Demonstramos a importância de nosso método simulando três cenários de tratamento com o objetivo de restabelecer a dinâmica da expressão do gene RKIP em direção a um estado pré-canceroso (1) por aumentar a duração do estado ON do promotor do gene; (2) por aumentar a taxa de síntese de mRNAs; e (3) por aumentar ambas as taxas. Nós mostramos que as taxas cinéticas de pré-tratamento das velocidades de chaveamento do promotor ON e OFF e a síntese e degradação do mRNA afetarão a heterogeneidade e o tempo para a resposta ao tratamento. Portanto, apresentamos uma estratégia para atingir níveis aumentados de número médio de mRNA com heterogeneidade diminuída, que é reduzida a dosagem das drogas que visam simultaneamente as taxas cinéticas que representam efetivamente os processos químicos subjacentes à regulação da expressão gênica. A diminuição da heterogeneidade da resposta ao tratamento por um gene alvo ajuda a diminuir as chances de surgimento de resistência. Nossa abordagem pode ser útil para inferir constantes cinéticas relacionadas à expressão de genes antimetastáticos ou oncogenes e para o desenho de estratégias terapêuticas multidrogas direcionadas aos processos que sustentam a expressão de genes reguladores mestres.

Palavras-chaves: Dinâmica estocástica. Regulação gênica. Terapêutica epigenética. Farmacoterapia multidrogas.

Abstract

GIOVANINI, Guilherme. **Stochastic binary models for gene expression: a comparative analysis of noise properties and an investigation of responses to gene therapies.** 2022. 118 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2022.

We made a comparison of noise properties exhibited by two stochastic binary gene models at the steady state: a repressed or activated externally regulating gene and a self-repressing one. These models describe the dynamics of probability distributions governing two random variables: gene products number and the gene state (ON or OFF). We quantify noise in protein numbers by means of its Fano factor and write this quantity as a function of the covariance between the two random variables. Then we show that distributions governing the number of gene products can be super-Fano, Fano or sub-Fano if the covariance is, respectively, positive, null or negative. The latter condition is exclusive for the self-repressing gene and our analysis shows the conditions for which the Fano factor is a sufficient classifier of fluctuations in gene expression. We also present the conditions for which the noise on the number of gene products generated from a self-repressing gene or an externally regulating one are quantitatively similar. That is important for inference of gene regulation from noise in gene expression quantitative data. Our results contribute to a classification of noise function in biological systems by theoretically demonstrating the mechanisms underpinning the higher precision in expression of a self-repressing gene in comparison with an externally regulated one. Finally, we use a time-dependent stochastic binary gene model to investigate treatment response with a switching target gene. To establish biologically relevant timescales for the parameters of the model, we select the RKIP gene and two non-specific drugs already known for changing RKIP levels in cancer cells. We demonstrate the usefulness of our method simulating three treatment scenarios aiming to reestablish RKIP gene expression dynamics toward a pre-cancerous state: (1) to increase the promoter's ON state duration; (2) to increase the mRNAs' synthesis rate; and (3) to increase both rates. We show that the pre-treatment kinetic rates of ON and OFF promoter switching speeds and mRNA synthesis and degradation will affect the heterogeneity and time for treatment response. Hence, we present a strategy for reaching increased average mRNA levels with diminished heterogeneity while reducing drug dosage by simultaneously targeting multiple kinetic rates that effectively represent the chemical processes underlying the regulation of gene expression. The decrease in heterogeneity of treatment response by a target gene helps to lower the chances of emergence of resistance. Our approach may be useful for inferring kinetic constants related to the expression of antimetastatic genes or oncogenes and for the design of multi-drug therapeutic strategies targeting the processes underpinning the expression of master regulatory genes.

Keywords: Stochastic dynamics. Gene regulation. Epigenetic therapy. Multidrug pharmacotherapy.

Lista de figuras

Figura 1 – Representação dos modelos de expressão gênica: GER e GAR	21
Figura 2 – Correlação em função da probabilidade do gene no estado ON para o GER.	33
Figura 3 – Distribuições quasi-Fano e super-Fano para o GER	34
Figura 4 – Assimetria e curtose das distribuições do GER	36
Figura 5 – Correlação em função da probabilidade do gene no estado ON e do número médio de proteínas para o GAR.	40
Figura 6 – Distribuições Fano e quasi-Fano para o GAR.	42
Figura 7 – Distribuições super-Fano para o GAR.	43
Figura 8 – Distribuições sub-Fano para o GAR.	45
Figura 9 – Fator de Fano em função da probabilidade do gene no estado ON e do número médio de proteínas para o GAR.	45
Figura 10 – Assimetria e curtose das distribuições do GAR.	48
Figura 11 – Esquema das drogas visando taxas cinéticas na transcrição de RKIP.	63
Figura 12 – Distribuições de probabilidade das condições pré e pós-tratamentos.	72
Figura 13 – Dinâmica da resposta ao tratamento com DETANONOato.	74
Figura 14 – Dinâmica da resposta ao tratamento com 5-Azac.	76
Figura 15 – Dinâmica da resposta ao tratamento com DETANONOato e 5-Azac.	77
Figura 16 – Mapeamento da resposta ao tratamento para o efeito do fracionamento das drogas.	80
Figura 17 – Aprimoramento de alguns desenhos de tratamentos não-efetivos.	82

Lista de tabelas

Tabela 1 – Valores aproximados dos parâmetros cinéticos nos gráficos da Fig. 3. . .	35
Tabela 2 – Valores aproximados dos parâmetros cinéticos nos gráficos da Fig. 7. . .	44
Tabela 3 – Valores aproximados dos parâmetros cinéticos no gráfico da Fig. 8. . .	46
Tabela 4 – Regimes de flutuações dados pelos parâmetros cinéticos no GER e no GAR.	50
Tabela 5 – As frações ξ_a e ξ_b da dose máxima tolerada para agendas de tratamento das condições de pré-tratamento da Fig. 15.	78
Tabela 6 – Parâmetros para os tratamentos aprimorados.	117

Sumário

1	Introdução	14
2	Fundamentos teóricos	17
2.1	<i>Cadeias de Markov</i>	17
2.2	<i>Equação mestra</i>	18
3	Um modelo estocástico para o gene binário	20
3.1	<i>Descrição qualitativa</i>	20
3.2	<i>Classificação das flutuações nos modelos de expressão gênica</i>	22
3.2.1	Fator de Fano	22
3.2.2	Covariância	23
3.2.3	Assimetria e Curtose	23
3.3	<i>Modelo estocástico</i>	24
3.4	<i>Técnica das funções geratrizes</i>	26
3.5	<i>Soluções exatas no regime estacionário</i>	26
3.6	<i>Distribuições de probabilidade marginais, totais e seus momentos</i>	28
3.6.1	Covariância aplicada ao modelo do gene binário	29
4	Os genes externamente regulado (GER) e autorreprimido (GAR) 30	
4.1	<i>GER</i>	30
4.1.1	Distribuições de probabilidade marginais, totais e seus momentos	30
4.1.2	Caracterização das flutuações das distribuições de probabilidade	31
4.1.3	Resultados	32
4.2	<i>GAR</i>	37
4.2.1	Distribuições de probabilidade marginais, totais e seus momentos	37
4.2.2	Caracterização das flutuações das distribuições de probabilidade	39
4.2.3	Resultados	40
4.3	<i>Comparações dos resultados entre o GER e o GAR</i>	49
4.4	<i>Discussões</i>	51
5	Um modelo dependente do tempo para investigar respostas a terapias gênicas	53

5.1	<i>Métodos e Modelos</i>	57
5.1.1	Uma breve descrição do papel molecular do RKIP	57
5.1.2	Um modelo efetivo para a regulação da expressão gênica	59
5.1.3	Um modelo efetivo para investigar a regulação da dinâmica da expressão gênica após o tratamento	60
5.1.4	Uma abordagem para investigar os efeitos do tratamento na dinâmica de expressão de RKIP	63
5.1.5	Uma descrição aproximada da dinâmica da expressão gênica binária estocástica com taxas cinéticas dependentes do tempo	65
5.1.6	Um modelo exatamente solúvel para avaliação comparativa do tra- tamento do câncer com o objetivo de modular os níveis de expressão gênica	66
5.1.7	Valores de parâmetros e condições para simulações de tratamento	67
5.2	<i>Resultados</i>	71
5.2.1	Tratamento visando a taxa de chaveamento de estado do gene de OFF para ON	73
5.2.2	Tratamento visando a taxa de síntese de mRNA de RKIP	75
5.2.3	Tratamento com as duas drogas concomitantemente	77
5.2.4	Mapeamento da resposta ao tratamento para o efeito do fraciona- mento ξ_a e ξ_b na redução de drogas	79
5.2.5	Melhorando os tratamentos não-efetivos visando todas as taxas cinéticas	81
5.3	<i>Discussão</i>	84
6	Considerações finais	88
	REFERÊNCIAS	90
	APÊNDICES	100
	Apêndice A – Dedução da Eq. (3.17)	101
	Apêndice B – Dedução da equação de Kummer - Eq. (3.20)	102

Apêndice C – Dedução de $\tilde{\alpha}(\eta)$, $\tilde{\beta}(\eta)$ e $\tilde{\phi}(\eta)$ pelo método de Frobenius	104
Apêndice D – Deduções da covariância ξ e do fator de Fano em função de ξ	107
D.1 Dedução da covariância ξ - Eq. (3.36)	107
D.2 Dedução do fator de Fano em função da covariância ξ - Eq. (3.11)	107
Apêndice E – Dedução de ξ^s - Eq. (4.32)	109
Apêndice F – Aproximação por partes dos parâmetros cinéticos dependentes do tempo	110
Apêndice G – Sistema de EDOs acopladas que regem os momentos da distribuição	111
Apêndice H – Valores dos momentos da distribuição para o estado estacionário	113
Apêndice I – Coeficientes das Eqs. (5.15) e (5.16)	114
Apêndice J – Fórmulas exatas para $A(t)$, $\langle n_\alpha \rangle(t)$ e $\langle n^2 \rangle(t)$	115
Apêndice K – Probabilidades de encontrar n produtos gênicos no estado estacionário	116
Apêndice L – Parâmetros do tratamento e a agenda de doses para desenhos de tratamento aprimorados	117
Anexo A – Códigos para computação dos modelos	118

1 Introdução

A expressão de produtos gênicos em organismos vivos é um processo que ocorre no ambiente intracelular, onde a informação genética codificada na sequência linear de nucleotídeos presentes na molécula de DNA rege a síntese de outras moléculas responsáveis pelo o funcionamento das células: o RNA mensageiro (mRNA) e as proteínas. A partir dos genes, o maquinário celular transforma a informação contida no DNA em moléculas de informação genética temporária, o mRNA, sendo esse processo conhecido como transcrição. A síntese de proteínas ocorre pela leitura dos genes contidos no mRNA por meio de outros conjuntos de máquinas moleculares da célula, tal processo é denominado tradução (ALBERTS *et al.*, 2014). Embora a transcrição de genes e a tradução de mRNA sejam processos fundamentais da expressão gênica, o funcionamento normal de uma célula requer que diferentes genes sejam expressos em instantes e quantidades adequadas.

A regulação da expressão gênica pode ocorrer em múltiplas etapas. No caso das células eucariotas, antes da síntese de mRNA de um determinado gene, a regulação ocorre por controle epigenético na modificação conformacional do DNA estruturado em cromatina, ou por fatores de transcrição que ativam, reprimem, potencializam ou retardam a ação do maquinário de transcrição. Após a transcrição, a regulação pode ocorrer pela ligação de moléculas de microRNAs ao mRNA, inativando o processo de tradução, ou por meio da ubiquitinação da proteína expressa, onde tal mecanismo acelera a degradação da proteína ao sinalizá-la para a ação enzimática do proteossoma. A ubiquitinação é verificada no controle de processos celulares como o ciclo celular (COOPER, 2000). Os processos ao longo da expressão gênica estão submetidos a flutuações aleatórias inevitáveis e, mesmo com isso, ainda é garantida a síntese dos produtos gênicos em porções específicas e adequadamente localizadas no espaço e tempo, com a precisão necessária para a funcionalidade destes organismos.

As flutuações na quantidade de produtos gênicos são causadas pela presença de reagentes em pequenas quantidades de cópias (DELBRÜCK, 1940) e foram detectadas experimentalmente pela aplicação de técnicas de fluorescência e microscopia para investigar a dinâmica dos processos biológicos moleculares que ocorrem em células procarióticas (ELOWITZ *et al.*, 2002) e eucarióticas (BLAKE *et al.*, 2003; RAJ *et al.*, 2006). Essas flutuações permitem determinar a regulação dos genes (MUNSKY; NEUERT; OUDE-

NAARDEN, 2012) e, uma vez que são inevitáveis, é possível lançar a hipótese de que há regulação do ruído na expressão dos genes. Isso evidencia a utilidade de uma classificação das estratégias reguladoras do processo de expressão gênica que permite o controle, tanto da quantidade de produtos gênicos, quanto de seus níveis de ruído.

Os modelos estocásticos mais simples para regulação de genes pressupõem transcrição e tradução como processos acoplados e a regulação da síntese de proteínas realizada pela mudança espontânea do estado do promotor do gene (PECCOUD; YCART, 1995). A autorregulação deste modelo de gene binário estocástico pode ser introduzida se uma das taxas de chaveamento do estado do promotor for escrita em função do número de proteínas. Supomos que o número de proteínas não mude devido à mudança do estado do promotor (HORNOS *et al.*, 2005). Essa condição pode corresponder à proteína expressa atuando como uma enzima que catalisa a produção do verdadeiro repressor ou, se a proteína for diretamente o repressor, seus números são grandes o suficiente para tornar desprezível o efeito da ligação de uma única proteína ao promotor. A regulação gênica por fatores externos é modelada similarmente ao gene autorregulado, mas no caso externamente regulado, consideramos a taxa de chaveamento de estado de ON para OFF independente do número de proteínas sintetizadas.

Apesar das simplificações impostas pelo nosso modelo de gene binário estocástico, ele apresenta uma grande diversidade de regimes de expressão gênica qualitativamente distintos, que possibilita o uso do modelo na interpretação de trabalhos experimentais. Além disso, essa abordagem pode auxiliar: na elaboração de terapia ou diagnóstico do câncer com base na regulação da expressão gênica (RAMOS *et al.*, 2020; GIOVANINI *et al.*, 2021); no estabelecimento de uma classificação da regulação do ruído em expressão gênica de acordo com a função de seus produtos em uma célula ou rede (GAMA *et al.*, 2020); e na construção de modelos cada vez mais complexos ao incorporar efeitos adicionais (RAMOS *et al.*, 2010).

Temos como objetivo a realização de uma análise comparativa do comportamento do ruído no contexto de modelos binários estocásticos exatamente solúveis, no estado estacionário para o gene autorreprimido e o gene externamente regulado (PECCOUD; YCART, 1995; HORNOS *et al.*, 2005; RAMOS; INNOCENTINI; HORNOS, 2011). Encontra-se na literatura a análise do ruído para cada uma das estratégias de regulação gênica (RAMOS; HORNOS, 2007; RAMOS; HORNOS; REINITZ, 2015; RAMOS; REINITZ, 2019) apresentadas aqui, onde contribuímos ao comparar as estratégias de regulação gênica

e ao explorar as formas das distribuições de probabilidades dos números de produtos gênicos por meio dos coeficientes de assimetria e curtose. Também pretendemos investigar a resposta da expressão de um gene alvo submetido a agendas de tratamento multidrogas, baseado num modelo dependente do tempo para a regulação da expressão gênica por meio de suas taxas cinéticas (PRATA; HORNOS; RAMOS, 2016; RAMOS *et al.*, 2020). Os parâmetros da simulação recebem valores experimentais encontrados na literatura. Buscamos a redução das doses nas agendas de tratamentos atuando sobre a transcrição do gene RKIP, um fisiológico supressor de metástase produzido por uma diversidade de organismos e encontrado em níveis deficitários em tumores (YESILKANAL; ROSNER, 2018), de maneira que, ao longo do tratamento, seja garantido o aumento dos transcritos do RKIP em níveis desejáveis e com a menor heterogeneidade possível na expressão.

Esta dissertação está dividido em mais cinco capítulos e organizado da seguinte maneira: no capítulo 2 encontramos a teoria de processos estocásticos para elaboração do modelo de gene binário estocástico que estão desenvolvidos no capítulo 3. Neste último também encontram-se as técnicas para obtenção das soluções exatas do modelo para o estado estacionário e as funções para a classificação do ruído. No capítulo 4 derivamos, a partir das soluções exatas, dois modelos específicos de estratégias da regulação gênica no estado estacionário: o gene externamente regulado (GER) e o gene autorreprimido (GAR). Apresentamos os resultados obtidos em cada modelo e os discutimos de maneira individual e comparativa. O capítulo 5 apresenta uma prova de conceito ao modelar uma terapia gênica multidrogas para um controle epigenético da expressão gênica do RKIP (GIOVANNI *et al.*, 2021). Neste capítulo conceitualizamos a proposta; apresentamos o papel molecular do RKIP e modelamos a terapia via simulação de agendas de tratamento com drogas moduladoras da expressão gênica; apresentamos e discutimos os resultados obtidos. As considerações finais desta dissertação são apresentadas no capítulo 6.

2 Fundamentos teóricos

Neste capítulo são apresentados dois conceitos teóricos utilizados na construção do modelo da regulação do gene binário estocástico, as *cadeias de Markov* e a *equação mestra*, que estão bem estabelecidos na literatura (REICHL, 2016; KAMPEN, 1992; GARDINER, 1983; TOMÉ; OLIVEIRA, 2014). O primeiro trata-se de um modelo estocástico que descreve uma sequência de possíveis eventos cujas probabilidades de cada evento são dependentes apenas do estado do evento anterior. Já o segundo, é uma técnica de modelagem por equações diferenciais que descreve a variação temporal das probabilidades de um sistema assumir cada um dos seus possíveis estados.

2.1 Cadeias de Markov

Num modelo de expressão gênica, o número de produtos gênicos e o estado do promotor do gene podem ser pensados como duas variáveis aleatórias, em que as probabilidades de encontrar uma determinada configuração desse par são regidas por distribuições que variam no tempo, caracterizando um *processo estocástico*. Se nesse processo, a probabilidade de uma configuração ocorrer num determinado instante de tempo depende apenas da configuração do instante anterior, então chamamos esse processo de *cadeia de Markov*.

Formalmente, seja $n_0, n_1, n_2, \dots, n_m$ uma sequência temporal de estados que a variável aleatória X assume entre os instantes 0 e m . Podemos então definir um processo estocástico pela probabilidade conjunta dessa sequência:

$$P_m(n_0, n_1, n_2, \dots, n_m) = P_m(n_m | n_0, n_1, n_2, \dots, n_{m-1}) \dots P_1(n_1 | n_0) P_0(n_0), \quad (2.1)$$

onde $P_m(n_m | n_0, n_1, n_2, \dots, n_{m-1})$ é a probabilidade condicional de X assumir o estado n_m no instante m dado que X assumiu a sequência n_1, n_2, \dots, n_{m-1} nos respectivos instantes anteriores. Uma *cadeia de Markov* é um processo estocástico definido pela seguinte propriedade:

$$P_m(n_m | n_0, n_1, n_2, \dots, n_{m-1}) = P_m(n_m | n_{m-1}), \quad (2.2)$$

ou seja, a probabilidade de n_m ocorrer no instante m depende apenas da ocorrência no instante $m - 1$. Sendo assim, num processo markoviano a Eq. (2.1) torna-se:

$$P_m(n_0, n_1, n_2, \dots, n_m) = P_m(n_m | n_{m-1}) \dots P_2(n_2 | n_1) P_1(n_1 | n_0) P_0(n_0). \quad (2.3)$$

Ao aplicar somatórios sobre n_0, n_1, \dots, n_{m-1} na Eq. (2.3), obtemos

$$\sum_{n_0, n_1, \dots, n_{m-1}} P_m(n_0, n_1, n_2, \dots, n_m) = P_m(n_m), \quad (2.4)$$

que ao substituir a Eq. (2.3) dentro do somatório, resulta em:

$$P_m(n_m) = \sum_{n_0, n_1, \dots, n_{m-1}} P_m(n_m | n_{m-1}) \dots P_2(n_2 | n_1) P_1(n_1 | n_0) P_0(n_0) \quad (2.5)$$

$$= \sum_{n_{m-1}} P_m(n_m | n_{m-1}) P_{m-1}(n_{m-1}). \quad (2.6)$$

A equação de Chapman-Kolmogorov,

$$P_{m+1}(n_{m+1} | n_{m-1}) = \sum_{n_m} P_{m+1}(n_{m+1} | n_m) P_m(n_m | n_{m-1}), \quad (2.7)$$

nos permite aplicar os somatórios em n_1, n_2, \dots, n_{m-1} na Eq. (2.3) de maneira a obter:

$$P_m(n_m) = \sum_{n_0} P_m(n_m | n_0) P_0(n_0), \quad (2.8)$$

onde fica evidente que dado $P_0(n_0)$, é possível obter a probabilidade de n_m no instante m . A probabilidade condicional $P_k(n_k | n_j)$ é interpretada como probabilidade de transição de n_j no instante j , para n_k no instante k . Consideraremos que as probabilidades das transições não variam com o tempo, e sendo assim, podemos escrever $P_k(n_k | n_j) = T(k, j)$, os elementos de uma matriz estocástica T que respeita duas propriedades: $T(k, j) \geq 0$ (fruto da definição de probabilidade) e $\sum_j T(k, j) = 1$ (garantindo que as probabilidades estejam normalizadas). Este resultado do processo markoviano é base para o desenvolvimento da equação mestra, como é mostrado na próxima seção.

2.2 Equação mestra

A partir de uma cadeia de Markov, é possível descrever o processo estocástico por uma equação diferencial denominada *equação mestra*, onde a evolução temporal das probabilidades é dada numa variável contínua. Para o desenvolvimento da equação mestra, considere $P(n, t)$ sendo a probabilidade de ocorrer o estado n no instante t , propriamente:

$$P(n, t + \Delta t) = \sum_{m=1}^M P(n, t + \Delta t | m, t) P(m, t). \quad (2.9)$$

Pela definição de derivada temporal, temos:

$$\frac{dP(n, t)}{dt} \equiv \lim_{\Delta t \rightarrow 0} \left[\frac{P(n, t + \Delta t) - P(n, t)}{\Delta t} \right]. \quad (2.10)$$

Ao substituir a Eq. (2.9) na Eq. (2.10) obtemos:

$$\frac{dP(n, t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \sum_{m=1}^M P(m, t) [P(n, t + \Delta t | m, t) - \delta_{m,n}], \quad (2.11)$$

onde $\delta_{m,n}$ é o *delta de Kronecker*. Agora, considerando que $\Delta t \rightarrow 0$ é possível expandir a probabilidade de transição $P(n, t + \Delta t | m, t)$ em série de potências em Δt até a primeira ordem, que nos dá a forma geral

$$P(n, t + \Delta t | m, t) = w_{m,n}(t)\Delta t + \delta_{m,n} \left[1 - \Delta t \sum_{l=1}^M w_{m,l}(t) \right] + \mathcal{O}(\Delta t^2), \quad (2.12)$$

onde o primeiro termo $w_{m,n}(t)\Delta t$ representa a probabilidade de ocorrência da transição de m para n (onde $m \neq n$) durante o intervalo de tempo entre t e $t + \Delta t$, e o segundo termo representa a probabilidade complementar ($m = n$), ou seja, a probabilidade de não ocorrer a transição neste mesmo intervalo de tempo. Note que $w_{m,n}(t)$ necessariamente tem dimensão de probabilidade por unidade de tempo, sendo assim, é interpretado como a taxa que rege a variação temporal da probabilidade de transição do estado m para n durante o intervalo Δt . Agora, ao substituir a Eq. (2.12) na equação mestra, Eq. (2.11), temos:

$$\frac{dP(n, t)}{dt} = \sum_{m=1}^M P(m, t)w_{m,n}(t) - \sum_{m=1}^M P(n, t)w_{n,m}(t). \quad (2.13)$$

A variação temporal de $P(n, t)$ depende: de acréscimos provenientes de todas possíveis transições para o estado n , explicitado pelo primeiro somatório na Eq. (2.13); e de decréscimos resultantes de todas as transições que abandonam o estado n , representado no segundo somatório. Um modelo de regulação do gene binário pela abordagem da equação mestra está apresentado no próximo capítulo.

3 Um modelo estocástico para o gene binário

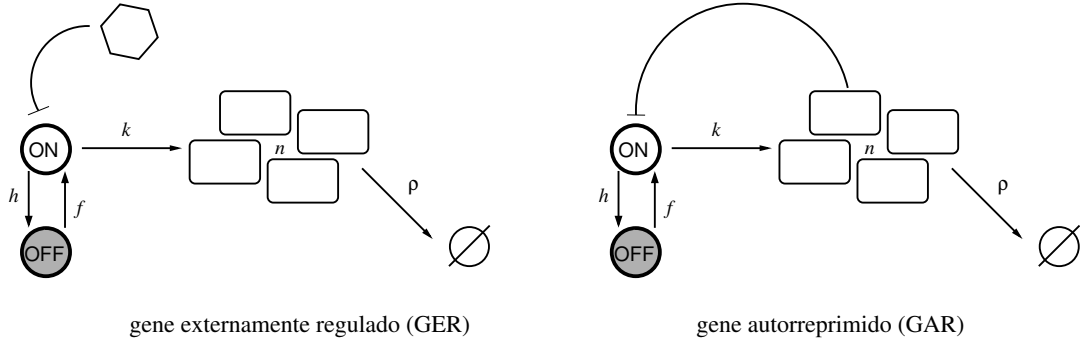
Neste capítulo é apresentada a descrição qualitativa do modelo de expressão gênica para um gene binário. Também são definidas medidas que classificam as flutuações do número de proteínas produzidas pelo gene: o fator de Fano, a covariância e os coeficientes de assimetria e curtose. Em sequência, é formulada a descrição estocástica do modelo por equações mestras e são apresentadas as funções geratrizes de probabilidades para estado estacionário. Por meio de operadores diferenciais aplicados às funções geratrizes, obtivemos as soluções exatas das distribuições de probabilidades marginais, totais e seus momentos. E por fim, descrevemos a covariância entre as variáveis aleatórias do modelo, em função dos momentos obtidos.

3.1 Descrição qualitativa

Numa descrição simplificada, a regulação da expressão gênica pode ser considerada como um chaveamento entre o ligar e o desligar do promotor do gene, onde essa regulação é induzida por proteínas específicas. Consideramos a transcrição e tradução como processos acoplados por assumir que não há regulação pós-transcricional e que há uma variação suficientemente pequena no número de proteínas produzidas a partir de um mRNA. Considerando a aleatoriedade no processo, definimos o estado deste sistema usando duas variáveis aleatórias (m, n) , onde $n = 0, 1, \dots$ indica o número de proteínas e m o estado do promotor do gene, onde o estado ON é representado por $m = \alpha$ e o OFF por $m = \beta$.

As reações químicas que de fato compreendem a produção de proteínas são simplificadas num conjunto mínimo de reações efetivas, onde a síntese de proteínas ocorre numa taxa constante indicada por k (ou 0) quando o promotor do gene está ON (ou OFF), enquanto que ρ indica a taxa de perda de função da proteína. O promotor muda do estado OFF para ON a uma taxa constante indicada por f . A taxa de chaveamento de ON para OFF é escolhida de acordo com a estratégia de regulação gênica: no caso de um gene externamente regulado (GER) a taxa é dada por h_1 , ou se for um gene autorreprimido (GAR) por h_2 . No segundo caso, a transição é uma função do número de proteínas n e para simplificar no nosso modelo consideramos essa dependência sendo linear, $h_2 n$.

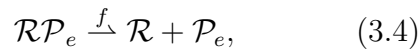
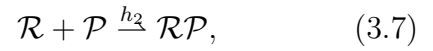
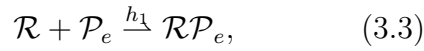
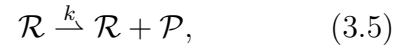
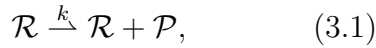
Figura 1 – Representação dos modelos de expressão gênica: GER e GAR


 Fonte – Giovanini *et al.* (2020)

A descrição acima está esquematizada na Fig. 1, que corresponde ao conjunto de reações químicas efetivas representadas pelas Eqs. (3.1)–(3.4) para o GER e pelas Eqs. (3.5)–(3.8) para o GAR. A regulação da proteína do GER é denotada por \mathcal{P}_e . Para o GAR, denotamos uma proteína por \mathcal{P} . A região reguladora do gene é denotado por \mathcal{R} e o estado do gene é determinado pela ligação de \mathcal{P} à região reguladora. Os símbolos para as taxas de reações aparecem no topo das setas indicando os reagentes e produtos das reações efetivas.

Gene externamente regulado (GER)

Gene autorreprimido (GAR)



As Eqs. (3.1) e (3.5) indicam a síntese de proteína quando o estado do promotor está ON, enquanto que a degradação de proteínas é indicada pelas Eqs. (3.2) e (3.6). A mudança do estado do promotor do gene de ON para OFF é indicada por Eqs. (3.3) e (3.7), e a transição oposta é apresentada nas Eqs. (3.4) e (3.8).

A existência de ruído nas quantidades observadas em experimentos que analisam a expressão gênica numa variedade de organismos (ELOWITZ *et al.*, 2002; BLAKE *et al.*, 2003; RAJ *et al.*, 2006; SUTER *et al.*, 2011) nos permite modelar o número de produtos gênicos n e o estado do promotor do gene m como duas variáveis aleatórias, cujas flutuações podem ser analisadas pelo fator de Fano e pela covariância, medidas já estabelecidas na literatura (RAMOS; HORNOS; REINITZ, 2015), como também pelos os coeficientes de assimetria e curtose, duas medidas adicionais para auxiliar na caracterização dos ruídos na

expressão gênica. Uma breve descrição dessas quatro medidas serão apresentadas a seguir na seção 3.2.

3.2 Classificação das flutuações nos modelos de expressão gênica

3.2.1 Fator de Fano

As flutuações nas quantidades de produtos gênicos em um instante de tempo t são quantificadas pelo fator de Fano, denotado por \mathcal{F} , que é definido como a razão da variância com o número médio de proteínas:

$$\mathcal{F} = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle}. \quad (3.9)$$

Podemos calcular \mathcal{F} por meio das probabilidades que governam n , para o estado estacionário ou dependentes do tempo, ao estimar os níveis de expressão em uma amostra de genes, por exemplo, pela medida de intensidade de fluorescência de uma sonda (ROSENFELD *et al.*, 2006). Propriamente, o fator de Fano é definido para uma determinada janela de tempo num processo estocástico, onde são calculados a média e variância de uma variável aleatória. Quando a janela de tempo é infinita, o valor medido é chamado de índice de dispersão ou razão variância-média. Aqui trataremos por fator de Fano ambas as situações.

O fator de Fano é uma medida de como uma distribuição de probabilidade é diferente da distribuição de Poisson. Pode-se nomear uma distribuição de probabilidade de acordo com o valor do seu fator de Fano, de modo que, uma distribuição Fano (ou distribuição poissoniana) tem $\mathcal{F} = 1$, uma quasi-Fano (ou distribuição quasi-poissoniana) tem $\mathcal{F} \approx 1$, uma super-Fano (ou super-poissoniana) tem $\mathcal{F} > 1$, uma sub-Fano (ou sub-poissoniana) tem $\mathcal{F} < 1$ e uma infra-Fano (ou infra-poissoniana) possui $\mathcal{F} \ll 1$ (RAMOS; HORNOS; REINITZ, 2015).

O fator de Fano fornece um critério para avaliar flutuações nos níveis de expressão gênica e os momentos de uma distribuição que deve ser quantificados para sua completa caracterização. Por exemplo, as distribuições poissoniana e sub-poissoniana estabelecem um pico em torno da média, de modo que os momentos de primeira e segunda ordem são suficientes para estimar aproximadamente o número de produtos gênicos que são governados por eles, uma vez que n está principalmente dentro do intervalo $[\langle n \rangle - \sqrt{\mathcal{F}\langle n \rangle}, \langle n \rangle + \sqrt{\mathcal{F}\langle n \rangle}]$.

Esse pode não ser o caso das distribuições super-poissonianas que podem ter uma multiplicidade de formas, por exemplo, a binomial negativa, a bimodal e a em forma de mesa. Essas distribuições são amplamente espalhadas e o cálculo dos momentos de primeira e segunda ordem não é suficiente para uma estimativa aproximada do número de produtos gênicos em um determinado instante. A avaliação do ruído super-poissoniano na expressão gênica e de suas origens é uma difícil tarefa e o uso de modelos matemáticos exatamente solúveis fornecendo distribuições de probabilidade podem produzir mais *insights* sobre a compreensão dos mecanismos de regulação gênica.

3.2.2 Covariância

Ao considerar a fenomenologia do modelo para propor uma variável aleatória ν_m associada aos estados do promotor do gene, podemos calcular a covariância ξ entre ν_m e n . Sendo assim, pela definição de covariância temos:

$$\xi = \langle \nu_m n \rangle - \langle \nu_m \rangle \langle n \rangle. \quad (3.10)$$

Recentemente, foi demonstrado para o GAR, que o fator de Fano do estado estacionário pode ser escrito em termos de seus números médios de proteínas e a covariância entre ν_m e n (RAMOS; HORNOS; REINITZ, 2015). Estendemos essa relação para o GER, de modo que a Eq. (3.9) torna-se:

$$\mathcal{F} = 1 + \frac{\xi}{\langle n \rangle}. \quad (3.11)$$

A forma da Eq. (3.11) permite a classificação de sua distribuição de probabilidade relacionada de acordo com a covariância ξ . Sendo assim, se a covariância entre os números de proteínas e estado do promotor for positiva, o modelo encontra-se num regime super-Fano. Enquanto que um regime sub-Fano ocorre com uma covariância negativa e as distribuições Fano são caracterizadas por covariância nula.

3.2.3 Assimetria e Curtose

Considerando nosso modelo onde uma das variáveis aleatórias é o número de produtos gênicos, podemos analisar as flutuações por meio de momentos de ordens maiores que o número médio e a variância. A assimetria e curtose são, respectivamente, o terceiro

e o quarto momentos padronizados de uma distribuição de probabilidade de uma variável aleatória e relacionam-se com o formato da distribuição. Para o cálculo da assimetria \mathcal{S} (em inglês, *skewness*) são necessários os momentos das distribuições até a terceira ordem, e no caso da curtose \mathcal{K} (em inglês, *kurtosis*), até a quarta ordem, a saber:

$$\mathcal{S} := \mathbb{E} \left[\left(\frac{n - \langle n \rangle}{\sigma} \right)^3 \right] = \frac{\langle n^3 \rangle - \langle n \rangle^3 - 3 \langle n \rangle \cdot [\langle n^2 \rangle - \langle n \rangle^2]}{\sigma^3}, \quad (3.12)$$

$$\mathcal{K} := \mathbb{E} \left[\left(\frac{n - \langle n \rangle}{\sigma} \right)^4 \right] = \frac{\langle n^4 \rangle - 4 \langle n \rangle \langle n^3 \rangle + 6 \langle n \rangle^2 \langle n^2 \rangle - 3 \langle n \rangle^4}{\sigma^4}, \quad (3.13)$$

onde $\sigma = \sqrt{\langle n^2 \rangle - \langle n \rangle^2}$ é o desvio padrão.

A assimetria pode assumir qualquer valor no conjunto dos números reais e a curtose apenas valores positivos. Assimetria é uma medida do quanto a distribuição se espalha em relação ao valor médio, de maneira que se há mais massa de probabilidade a esquerda da média, a assimetria será negativa, enquanto que será positiva se a massa de probabilidade se espalhar mais a direita da média. A curtose é uma medida de ocorrência de valores extremos à média. Quanto menor o valor da curtose, menos *outliers* a distribuição produz, como é o caso da distribuição uniforme, pois os desvios à média são equiparáveis ou menores que o desvio padrão. Quanto maior o valor da curtose, mais contribuições de pontos que distam da média com valores acima do desvio padrão.

A distribuição normal muitas vezes é considerada como referência de simetria e espalhamento dos pontos da distribuição, tendo assimetria igual a 0 e curtose igual a 3. Num processo poissoniano de nascimento e morte com $\langle n \rangle = \langle n^2 \rangle = \frac{k}{\rho}$, temos a assimetria e a curtose como $\mathcal{S} = \langle n \rangle^{-1/2}$ e $\mathcal{K} = 3 + \langle n \rangle^{-1}$. Utilizamos a assimetria e a curtose como medidas auxiliares na distinção entre as distribuições de probabilidade que tem valores coincidentes para o primeiro e o segundo momentos.

3.3 Modelo estocástico

Um modelo estocástico de síntese de proteína bem conhecido é o de um gene com um promotor de dois estados encontrado em Peccoud e Ycart (1995), Hornos *et al.* (2005). As variáveis aleatórias do modelo são: o número de proteínas n e o estado do promotor m . O estado do sistema é caracterizado por uma distribuição de probabilidade

$$\Pi(\alpha_n(t), \beta_n(t)), \quad (3.14)$$

onde a probabilidade de encontrar n proteínas no instante t é indicado por $\alpha_n(t)$ quando o gene está ativado, ou indicado por $\beta_n(t)$ quando desativado. Uma equação mestra que governa a dinâmica das probabilidades é:

$$\frac{d\alpha_n(t)}{dt} = \overbrace{k[\alpha_{n-1}(t) - \alpha_n(t)]}^{\text{síntese}} + \overbrace{\rho[(n+1)\alpha_{n+1}(t) - n\alpha_n(t)]}^{\text{degradação}} - \overbrace{(h_1 + h_2n)\alpha_n(t) + f\beta_n(t)}^{\text{chaveamento}}, \quad (3.15)$$

$$\frac{d\beta_n(t)}{dt} = \overbrace{\rho[(n+1)\beta_{n+1}(t) - n\beta_n(t)]}^{\text{degradação}} + \overbrace{(h_1 + h_2n)\alpha_n(t) - f\beta_n(t)}^{\text{chaveamento}}. \quad (3.16)$$

Assumimos que as taxas k , ρ , f , h_1 , h_2 são constantes durante o intervalo de tempo de integração das Eqs. (3.15) e (3.16). O GER é modelado considerando $h_1 > 0$ e $h_2 = 0$, implicando na mudança aleatoriamente espontânea entre os estados ON e OFF. A condição contrária ($h_1 = 0$ e $h_2 > 0$) resulta num modelo para o GAR de modo que a taxa de chaveamento do estado ON para OFF depende de n . No entanto, pode-se também escrever as taxas de chaveamento como funções das concentrações dos fatores de transcrição reguladores codificados em outro gene para indicar fenomenologicamente a repressão ou ativação. O valor de h_1 depende do valor de montantes instantâneos de \mathcal{P}_e e por aproximação o consideramos constante sob as seguintes premissas: as probabilidades que regem os valores de \mathcal{P}_e atingem o regime de estado estacionário; os valores médios de \mathcal{P}_e são grandes o suficiente para garantir que pequenas flutuações não alterem sua probabilidade de ligação à região reguladora do gene; as flutuações nos valores de \mathcal{P}_e estão dentro de uma faixa suficientemente pequena.

Em caso de mudanças significativas nos valores de \mathcal{P}_e levando a uma paisagem reguladora distinta, podemos usar uma integração por partes das Eqs. (3.15) e (3.16). Portanto, as taxas são assumidas como constantes durante um intervalo de tempo em que as mudanças do ambiente regulatório são insignificantes (PRATA; HORNOS; RAMOS, 2016; RAMOS *et al.*, 2020). Uma outra abordagem possível é a formulação do modelo de expressão gênica por sistemas discretos no tempo, no qual também é possível descrever reações que são integradas dentro de uma pequena janela de tempo e que afetam o processo a longo prazo. Em muitos casos, modelos discretos no tempo podem ser vistos com aproximações de equações mestras (DU; WONG, 2016; TOMSKI; ZAKARCZEMNY, 2021).

3.4 Técnica das funções geratrizes

O modelo representado nas Eqs. (3.15) e (3.16) pode ser resolvido pela técnica das funções geratrizes (RAMOS; HORNOS, 2007), onde se transforma o sistema de infinitas equações diferenciais ordinárias (EDOs) acopladas em um sistema de duas equações diferenciais parciais (EDPs) acopladas:

$$\frac{\partial \alpha}{\partial t} = (z-1) \left(k\alpha - \rho \frac{\partial \alpha}{\partial z} \right) - h_2 z \frac{\partial \alpha}{\partial z} - h_1 \alpha + f\beta, \quad (3.17)$$

$$\frac{\partial \beta}{\partial t} = -\rho(z-1) \frac{\partial \beta}{\partial z} + h_2 z \frac{\partial \alpha}{\partial z} + h_1 \alpha - f\beta, \quad (3.18)$$

que uma vez solucionadas, nos fornecem as funções geratrizes de probabilidades α e β . A técnica consiste em considerar as probabilidades $\alpha_n(t)$ e $\beta_n(t)$ como coeficientes de uma expansão de Taylor de uma função analítica, ou seja,

$$\alpha(z, t) = \sum_{n=0}^{\infty} z^n \alpha_n(t), \quad \beta(z, t) = \sum_{n=0}^{\infty} z^n \beta_n(t). \quad (3.19)$$

A transformação das EDOs em EDPs é obtida ao substituir o par de Eqs. (3.19) nas Eqs. (3.15) e (3.16) e resulta nas Eqs. (3.17) e (3.18), onde o procedimento pode ser visto com detalhes no Apêndice A. As Eqs. (3.17) e (3.18) nos permitem obter as soluções exatas das distribuições de probabilidade do modelo para estado estacionário, que estão mostradas na próxima seção.

3.5 Soluções exatas no regime estacionário

No estado estacionário, onde $\frac{\partial \alpha}{\partial t} = \frac{\partial \beta}{\partial t} = 0$, é possível o desacoplamento das Eqs. (3.17) e (3.18) em duas EDOs solúveis e que tem a forma da conhecida equação hipergeométrica confluyente ou conhecida também como *equação de Kummer* (ABRAMOWITZ; STEGUN, 1972):

$$\eta \frac{d^2 \tilde{\alpha}}{d\eta^2} + (B - \eta) \frac{d\tilde{\alpha}}{d\eta} - A\tilde{\alpha} = 0, \quad (3.20)$$

onde seus parâmetros são dados pelas taxas de reações do modelo:

$$A = 1 + \frac{f}{\rho}, \quad B = 1 + \frac{f + h_1}{\rho + h_2} + \frac{kh_2}{(\rho + h_2)^2}, \quad \eta(z) = \frac{k}{\rho + h_2} \left(z - \frac{\rho}{\rho + h_2} \right). \quad (3.21)$$

As deduções das Eqs. (3.20) e (3.22) estão detalhadas no Apêndice B. Nos utilizaremos apenas da solução regular desta equação: as funções hipergeométricas confluentes

de primeiro tipo (IYER-BISWAS; HAYOT; JAYAPRAKASH, 2009) também conhecidas como funções KummerM (ABRAMOWITZ; STEGUN, 1972). Esta é uma vantagem do uso da técnica das funções geratrizes, a eliminação da solução singular da equação de Kummer, pois algoritmos para obtenção de soluções numéricas da equação de Kummer estão suscetíveis a instabilidades ocasionadas pela solução singular (INNOCENTINI; HORNOS, 2007).

A função KummerM é denotada por $M(A, B, \eta)$ e escrita como uma série de potências:

$$M(A, B, \eta) = \sum_{n=0}^{\infty} \frac{(A)_n \eta^n}{(B)_n n!}, \quad (3.22)$$

onde $(A)_n = A(A+1) \dots (A+n-1)$ é o símbolo de Pochhammer. Note que para $A = B$ a série de potências que define a função KummerM coincide com a função exponencial, portanto:

$$M(A, A, \eta) = e^\eta. \quad (3.23)$$

A Eq. (3.22) pode ser obtida pelo *método de Frobenius*, uma técnica bem conhecida para resolução de EDOs lineares de segunda ordem (BUTKOV, 1968; ARFKEN; WEBER, 2005), onde se supõe uma solução *ansatz* como série de potências diferenciável termo a termo, que nos faz obter $\tilde{\alpha}(z)$, a menos de uma constante de normalização. Esse desenvolvimento é mostrado no Apêndice C, juntamente das funções geratrizes de probabilidades para o estado ON, OFF e no caso independente do estado do gene. Essas funções geratrizes de probabilidades são denotadas respectivamente por $\tilde{\alpha}(z)$, $\tilde{\beta}(z)$, $\tilde{\phi}(z) = \tilde{\alpha}(z) + \tilde{\beta}(z)$:

$$\tilde{\alpha}(z) = \frac{z_0}{c} \frac{A-1}{B-1} M(A, B, \eta(z)), \quad (3.24)$$

$$\tilde{\beta}(z) = \frac{1}{c} \left[\left(1 - \frac{A-1}{B-1}\right) M(A-1, B, \eta(z)) + (1-z_0) \frac{A-1}{B-1} M(A, B, \eta(z)) \right], \quad (3.25)$$

$$\tilde{\phi}(z) = \frac{1}{c} M(A-1, B-1, \eta(z)), \quad (3.26)$$

onde as constantes auxiliares são definidas como $N = \frac{k}{\rho}$, representando o número médio de proteínas sintetizadas quando o gene está sempre ON; $z_0 = \frac{\rho}{\rho+h_2}$ que representa a taxa de degradação de proteínas proporcional às proteínas que tornam-se ausentes no sistema. A constante $c = [M(A-1, B-1, Nz_0(1-z_0))]^{-1}$ é consequente da normalização das distribuições e obtida por $\tilde{\phi}(1) = \tilde{\alpha}(1) + \tilde{\beta}(1) = 1$.

Na próxima seção, apresentamos as técnicas para obtenção das distribuições de probabilidade do modelo e seus momentos até a quarta ordem.

3.6 Distribuições de probabilidade marginais, totais e seus momentos

Uma vez que as soluções do modelo estocástico são obtidas, a partir dos procedimentos mostrados a seguir, que valem tanto para o caso do estado estacionário, quanto para o dependente do tempo, conseguimos obter as distribuições de probabilidade marginais, totais e seus respectivos momentos. As distribuições de probabilidades marginais de $\Pi(\alpha_n, \beta_n)$ são obtidas pela aplicação da derivada de n -ésima ordem no ponto $z = 0$, nas funções geratrizes de probabilidade $\alpha(z, t)$ e $\beta(z, t)$:

$$\alpha_n(t) = \frac{1}{n!} \left. \frac{\partial^n \alpha(z, t)}{\partial z^n} \right|_{z=0}, \quad \beta_n(t) = \frac{1}{n!} \left. \frac{\partial^n \beta(z, t)}{\partial z^n} \right|_{z=0}. \quad (3.27)$$

Note que essas probabilidades coincidem com os coeficientes da expansão de Taylor de uma função analítica. A probabilidade marginal de encontrar n proteínas dentro da célula independentemente do estado do promotor ser ON ou OFF, denotado por $\phi_n(t)$, pode ser encontrada diretamente ao aplicar o mesmo procedimento da n -ésima derivada na função geratriz $\phi(z, t)$,

$$\phi_n(t) = \frac{1}{n!} \left. \frac{\partial^n \phi(z, t)}{\partial z^n} \right|_{z=0}, \quad (3.28)$$

ou também por $\phi_n(t) = \alpha_n(t) + \beta_n(t)$.

As probabilidades marginais de encontrar o promotor no estado ON ou OFF independentemente da quantidade de proteínas dentro da célula são denotadas por $p_\alpha(t)$ e $p_\beta(t)$, respectivamente, de modo que

$$p_\alpha(t) = \sum_{n=0}^{\infty} \alpha_n(t) = \alpha(1, t) \quad \text{e} \quad p_\beta(t) = \sum_{n=0}^{\infty} \beta_n(t) = \beta(1, t). \quad (3.29)$$

O p -ésimo momento da distribuição de probabilidade $\Pi(\alpha_n, \beta_n)$ é obtido pela ação do operador $z \frac{\partial}{\partial z}$ aplicado numa composição p vezes seguidas sobre as funções geratrizes de probabilidades, no caso de $\alpha(z, t)$:

$$\langle n_\alpha^p \rangle(t) = \sum_{n=0}^{\infty} n^p \alpha_n(t) = \left[\left(z \frac{\partial}{\partial z} \right)^p \alpha(z, t) \right] \Big|_{z=1}, \quad (3.30)$$

onde $p \geq 1$. Se o mesmo operador agir em $\beta(z, t)$ ou $\phi(z, t)$, resultará respectivamente em $\langle n_\beta^p \rangle(t)$ e $\langle n^p \rangle(t)$. Ao tomar $\tilde{\phi}(z)$ como exemplo, o desenvolvimento dos seus quatro primeiros momentos resultam em:

$$\langle n \rangle = \left. \frac{d\tilde{\phi}}{dz} \right|_{z=1}; \quad (3.31)$$

$$\langle n^2 \rangle = \left. \frac{d\tilde{\phi}}{dz} \right|_{z=1} + \left. \frac{d^2\tilde{\phi}}{dz^2} \right|_{z=1}; \quad (3.32)$$

$$\langle n^3 \rangle = \left. \frac{d\tilde{\phi}}{dz} \right|_{z=1} + 3 \left. \frac{d^2\tilde{\phi}}{dz^2} \right|_{z=1} + \left. \frac{d^3\tilde{\phi}}{dz^3} \right|_{z=1}; \quad (3.33)$$

$$\langle n^4 \rangle = \left. \frac{d\tilde{\phi}}{dz} \right|_{z=1} + 7 \left. \frac{d^2\tilde{\phi}}{dz^2} \right|_{z=1} + 6 \left. \frac{d^3\tilde{\phi}}{dz^3} \right|_{z=1} + \left. \frac{d^4\tilde{\phi}}{dz^4} \right|_{z=1}, \quad (3.34)$$

onde $\frac{d^p\tilde{\phi}}{dz^p}$ aplicado na Eq. (3.26) resulta em:

$$\frac{d^p\tilde{\phi}}{dz^p} = \frac{(Nz_0)^p}{c} \frac{(A-1)_p}{(B-1)_p} M(A-1+p, B-1+p, Nz_0(z-z_0)), \quad (3.35)$$

e quando substituídos nas Eqs. (3.31)–(3.34), obtemos os momentos em função dos parâmetros do modelo. O mesmo procedimento pode ser aplicado em $\tilde{\alpha}(z)$ ou $\tilde{\beta}(z)$ para a obtenção explícita dos momentos.

Na próxima subseção, mostraremos que o cálculo do fator de Fano e da covariância entre ν_m e n dependem apenas os dois primeiros momentos, Eqs. (3.31) e (3.32).

3.6.1 Covariância aplicada ao modelo do gene binário

O desenvolvimento das Eqs. (3.10) e (3.11) aplicada ao modelo do gene binário encontra-se em Ramos, Hornos e Reintz (2015). A covariância entre o estado do promotor do gene e o número de proteínas sintetizadas na expressão gênica pode ser escrita como:

$$\xi = N (\langle n_\alpha \rangle - \langle n \rangle p_\alpha), \quad (3.36)$$

onde $\langle n_\alpha \rangle$ pode ser calculado a partir das Eqs. (3.30) e (3.31) aplicadas nas soluções exatas das Eqs. (3.15) e (3.16). As deduções das Eqs. (3.36) e (3.11) estão detalhadas no Apêndice D.

4 Os genes externamente regulado (GER) e autorreprimido (GAR)

As duas próximas seções a seguir, explicitam as soluções do regime estacionário para ambos modelos GER (INNOCENTINI; HORNOS, 2007) e GAR (HORNOS *et al.*, 2005), cujas probabilidades e parâmetros são distinguidos por sobrescritos e e s , respectivamente. Os parâmetros que são iguais para ambos os modelos não têm sobrescritos.

4.1 GER

4.1.1 Distribuições de probabilidade marginais, totais e seus momentos

A distribuição $\Pi(\alpha_n^e, \beta_n^e)$ tem três parâmetros a, N, b^e , definidos em termos das taxas de reações como (INNOCENTINI; HORNOS, 2007):

$$a = \frac{f}{\rho}, \quad N = \frac{k}{\rho}, \quad b^e = \frac{f + h_1}{\rho}. \quad (4.1)$$

Aqui, a meia vida da proteína, $\approx 1/\rho$, define a escala de tempo do modelo. O parâmetro a é a razão da taxa de chaveamento de OFF para ON com a taxa de decaimento e N é a constante que indica o número médio de proteínas quando o promotor permanecer exclusivamente ligado. O parâmetro b^e é a razão entre taxa de decaimento do promotor chaveando para estado estacionário com a taxa de degradação da proteína.

As probabilidades α_n^e , β_n^e , e ϕ_n^e são:

$$\alpha_n^e = \frac{a}{b^e} \frac{N^n}{n!} \frac{(a+1)_n}{(b^e+1)_n} M(a+n+1, b^e+n+1, -N), \quad (4.2)$$

$$\beta_n^e = \frac{b^e - a}{b^e} \frac{N^n}{n!} \frac{(a)_n}{(b^e+1)_n} M(a+n, b^e+n+1, -N), \quad (4.3)$$

$$\phi_n^e = \frac{N^n}{n!} \frac{(a)_n}{(b^e)_n} M(a+n, b^e+n, -N), \quad (4.4)$$

foram obtidas usando suas respectivas funções geratrizes encontradas em Peccoud e Ycart (1995), Innocentini e Hornos (2007), Ramos *et al.* (2010) e também apresentadas no capítulo 3. As probabilidades marginais de encontrar o gene ligado ou desligado são

$$p_\alpha^e = \frac{a}{b^e}, \quad p_\beta^e = \frac{b^e - a}{b^e}. \quad (4.5)$$

A probabilidade para o promotor estar ligado é proporcional à taxa de chaveamento do estado OFF para ON e indica a fração de tempo média para o promotor estar ligado durante um intervalo. Note que $0 \leq p_\alpha^e \leq 1$ implica em $0 \leq a \leq b^e$.

Aplicando o resultado $\left. \frac{d^p \tilde{\phi}^e}{dz^p} \right|_{z=1} = N^p \frac{(a)_p}{(b^e)_p}$ nas Eqs. (3.31)–(3.34), obtemos os quatro primeiros momentos para o GER:

$$\langle n \rangle = N \frac{a}{b^e}; \quad (4.6)$$

$$\langle n^2 \rangle = \langle n \rangle \left[1 + N \frac{a+1}{b^e+1} \right]; \quad (4.7)$$

$$\langle n^3 \rangle = \langle n \rangle + [\langle n^2 \rangle - \langle n \rangle] \left(3 + N \frac{a+2}{b^e+2} \right); \quad (4.8)$$

$$\langle n^4 \rangle = \langle n \rangle - [\langle n^2 \rangle - \langle n \rangle] \left(11 + 3N \frac{a+3}{b^e+3} \right) + [\langle n^3 \rangle - \langle n \rangle] \left(6 + N \frac{a+3}{b^e+3} \right), \quad (4.9)$$

nos possibilitando a caracterização das flutuações das distribuições de probabilidade, como segue na próxima subseção.

4.1.2 Caracterização das flutuações das distribuições de probabilidade

Na literatura (INNOCENTINI; HORNOS, 2007) encontra-se o número médio de proteínas para o GER, denotado por $\langle n^e \rangle$, como:

$$\langle n^e \rangle = p_\alpha^e N. \quad (4.10)$$

Note que para $p_\alpha^e = 1$ o número médio de proteínas é N , o qual corrobora para interpretação biológica dada em Eq. (4.1). A Eq. (3.36) é usada para calcular a covariância entre as variáveis (ν_m, n) no GER, de modo que:

$$\xi^e = \frac{N^2}{b^e+1} p_\alpha^e (1 - p_\alpha^e). \quad (4.11)$$

Esta expressão é obtida usando Eq. (4.5) e por

$$\langle n_\alpha^e \rangle = \frac{a}{b^e} \frac{d}{dz} [M(a+1, b^e+1, N(z-1))] \Big|_{z=1} = \frac{a}{b^e} \frac{1+a}{1+b^e} N, \quad (4.12)$$

onde usamos a fórmula 13.4.8 em Abramowitz e Stegun (1972): $\frac{d}{d\eta} M(A, B, \eta) = \frac{A}{B} M(A+1, B+1, \eta)$ para a obtenção da Eq. (4.12).

Conseqüentemente, o fator de Fano para o GER é dado por

$$\mathcal{F}^e = 1 + \frac{N}{b^e+1} (1 - p_\alpha^e), \quad (4.13)$$

o qual é uma função linear da probabilidade do gene estar no estado OFF. A covariância para o GER é sempre não-negativa e suas distribuições de probabilidade que governam n são apenas Fano e super-Fano.

A assimetria e curtose nos parâmetros das distribuições de probabilidade do GER são obtidas pela aplicação das Eqs. (4.6)–(4.9) nas Eqs. (3.12)–(3.13), que tornam-se:

$$\mathcal{S}^e = \frac{Np_\alpha^e}{\sigma^3} \left\{ 1 + \frac{N(1-p_\alpha^e)}{b^e+1} \left[3 - \frac{2N(2p_\alpha^e-1)}{b^e+2} \right] \right\}; \quad (4.14)$$

$$\mathcal{K}^e = \frac{Np_\alpha^e}{\sigma^4} \left\{ 1 + 3N + \frac{N(1-p_\alpha^e)}{b^e+1} \left[4 - 3b^e + \frac{6N[p_\alpha^e(b^e-2)+2]}{b^e+2} + \frac{3N^2[p_\alpha^e(1-p_\alpha^e)(b^e-6)+2]}{(b^e+2)(b^e+3)} \right] \right\}, \quad (4.15)$$

onde $\sigma = \left[Np_\alpha^e \left(1 + \frac{N(1-p_\alpha^e)}{b^e+1} \right) \right]^{1/2}$.

4.1.3 Resultados

Correlação de Pearson entre o número de proteínas e o estado do gene

A correlação de Pearson R entre duas variáveis é definida como a razão entre a covariância e a multiplicação entre os desvios padrão das duas variáveis. No nosso modelo, ξ é a covariância entre n e ν_m , e σ_n e σ_{ν_m} são os desvios padrão das respectivas variáveis. Sendo assim,

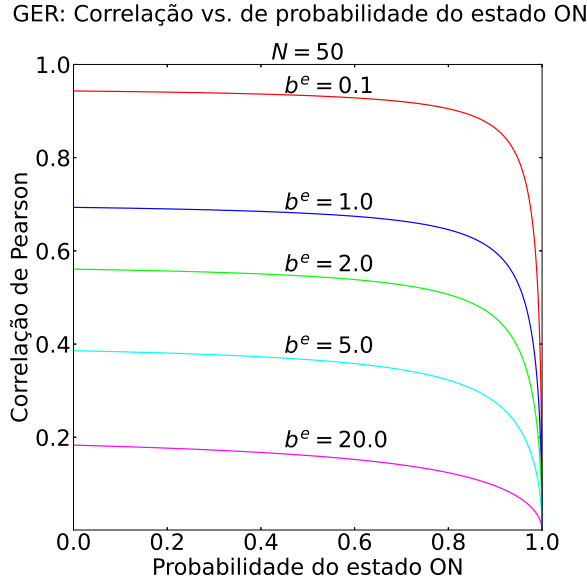
$$R = \frac{\xi}{\sigma_n \sigma_{\nu_m}}. \quad (4.16)$$

A partir dessa equação, calculamos os pontos dos gráficos de correlação tanto para o GER, quanto para o GAR. No GER, a correlação de Pearson $R^e(p_\alpha)$ como função da probabilidade do estado do promotor estar ON, é dada por

$$R^e(p_\alpha) = \frac{1}{\sqrt{1+b^e}} \sqrt{\frac{N(1-p_\alpha^e)}{1+b^e+N(1-p_\alpha^e)}}. \quad (4.17)$$

A Fig. 2 mostra o gráfico dessa função com curvas para diferentes valores do parâmetro b^e :

Figura 2 – Correlação em função da probabilidade do gene no estado ON para o GER.



Para um par fixado (N, b^e) , temos $R^e(p_\alpha)$ se aproximando de 0 quando $p_\alpha^e \rightarrow 1$, enquanto que se aproxima assintoticamente de $\frac{1}{\sqrt{1+b^e}} \sqrt{\frac{N}{1+b^e+N}}$ quando $p_\alpha \rightarrow 0$. A curva $R^e(p_\alpha)$ tem a forma de uma sigmoide invertida com valor de assíntota máxima denotada por R_m^e e sua meia altura ocorrendo em \hat{p}^e :

$$R_m^e = \frac{1}{\sqrt{1+b^e}} \sqrt{\frac{N}{1+b^e+N}}, \quad \text{e} \quad \hat{p}^e = \frac{3}{4 - \frac{N}{1+b^e+N}}. \quad (4.18)$$

Sendo assim, temos

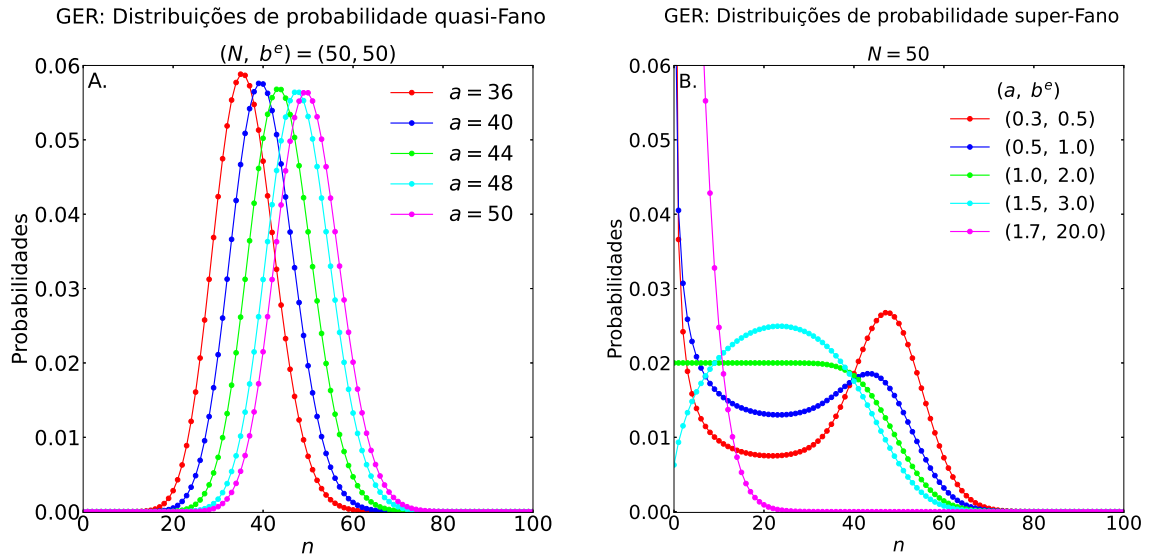
$$\frac{3}{4} < \hat{p}^e < 1 \Rightarrow R^e(\hat{p}^e) = \frac{R_m^e}{2}. \quad (4.19)$$

O gráfico da Fig. 2 mostra que a correlação de Pearson para o GER é $\sim R_m^e$ para a maioria dos valores de p_α . De fato, a correlação é metade de seus valores máximos quando $p_\alpha = \hat{p}^e$ e obedece Eq. (4.19). Conseqüentemente, a forma de sigmoide reversa da correlação começa a decair quando a probabilidade do promotor estar ON for mais próxima de 1. A Eq. (4.18) indica que R_m^e é superiormente limitada por $1/\sqrt{1+b^e}$ de tal modo que $R_m^e \rightarrow 1$ quando $b^e \rightarrow 0$ e $N \gg b^e$, porém quando $b^e \gg 1$ temos $R_m^e \rightarrow 0$ mesmo se $N \gg b^e$, que corresponde a um promotor que alterna rapidamente entre os estados ON e OFF durante a vida útil das proteínas, cuja presença das proteínas não seria uma leitura do estado do promotor do gene. Além disso, também pode corresponder a um regime quasi-poissoniano no qual $\mathcal{F}^e \rightarrow 1$ para o caso de $b^e \gg N(1-p_\alpha^e)$ — veja a

Eq. (4.13). Alternativamente, para o caso da síntese de proteínas de alta eficiência e com probabilidade pequena do gene estar ligado, o gene seria expresso num regime de *burst* (YVINEC *et al.*, 2017). Por outro lado, quando $b^e \ll 1$, temos um promotor que alterna seu estado lentamente, com uma duração de tempo entre do estado ON ou do OFF maior que a vida útil das proteínas. Neste caso, haverá um número finito de proteínas quando o promotor estiver ON, ou, nenhuma proteína, na maioria das vezes, quando o promotor estiver OFF. Este resultado está de acordo com os cálculos usando teoria da informação, onde distribuições binárias permitem a melhora da confiabilidade do fluxo de informação nas redes gênicas (GAMA *et al.*, 2020). Este regime corresponde a um acoplamento fraco entre os dois processos estocásticos do modelo binário para a expressão gênica.

Distribuições de probabilidade quasi-Fano e super-Fano

Figura 3 – Distribuições quasi-Fano e super-Fano para o GER



Em ambos os gráficos $N = 50$. O gráfico **A.** tem distribuições quasi-Fano e **B.** tem distribuições super-Fano. Os momentos das distribuições de ambos os gráficos são apresentados abaixo, em ordem crescente de a como na legenda. No gráfico **A.**: $\mathcal{F}^e \approx (1,3; 1,2; 1,1; 1,0; 1)$; $\xi^e \approx (9,9; 7,8; 5,2; 1,9; 0)$; $\langle n^e \rangle \approx (36; 40; 44; 48; 50)$; $p_\alpha^e \approx (0,72; 0,80; 0,88; 0,96; 1)$. No gráfico **B.**: $\mathcal{F}^e \approx (14,3; 13,5; 9,3; 7,3; 3,2)$; $\xi^e \approx (400; 312,5; 208,3; 156,3; 9,3)$; $\langle n^e \rangle \approx (30; 25; 25; 25; 4,3)$; $p_\alpha^e \approx (0,6; 0,5; 0,5; 0,5; 0,085)$.

Fonte – Giovanini *et al.* (2020)

Distribuições de probabilidade quasi-Fano e super-Fano que governam número de proteínas obtidas no modelo estocástico para um GER são mostrados na Fig. 3. Foi escolhido $N = 50$ para ambos os gráficos. O gráfico **A.** apresenta distribuições quasi-Fano

para grandes b^e e $a \approx b^e$ restrito a $a < b^e$. O caso $a = b^e \Rightarrow h_1 = 0$ também é mostrado pelos círculos sólidos magenta, correspondendo a um promotor que sempre está no estado ON. O restante das distribuições são quasi-Fano com covariância positiva e o número médio de proteínas aumentando linearmente com a . Isso é devido à dependência linear entre o número médio de proteínas produzido a partir de um GER e a probabilidade do promotor estar ligado, fornecido pela Eq. (4.10).

No gráfico **B.**, selecionamos múltiplos valores de a e b^e que resultam em distribuições de probabilidade espalhadas, como aquelas que tem forma de mesa e as bimodais. As distribuições em vermelho e azul são bimodais. A distribuição verde é em forma de mesa. A distribuição em ciano é unimodal e espalhada, semelhante à distribuição em forma de mesa. Por último, temos a distribuição em magenta que é aproximadamente uma binomial negativa.

Os valores dos parâmetros que geram as distribuições de probabilidade dos gráficos **A.** e **B.** podem ser reescritos em termos das constantes cinéticas encontradas nas Eqs. (3.1)–(3.4). A partir de (4.1), obtemos $k = N\rho$, $f = a\rho$ e $h_1 = (b^e - a)\rho$. Considerando $\rho = 0,01 \text{ min}^{-1}$, temos $k = 0,5 \text{ min}^{-1}$ para ambos os conjuntos de valores para cada gráfico e a Tab. 1 nos mostra os valores aproximados de f e h_1 em min^{-1} . As linhas são correspondentes a mesma sequência dos valores na legenda.

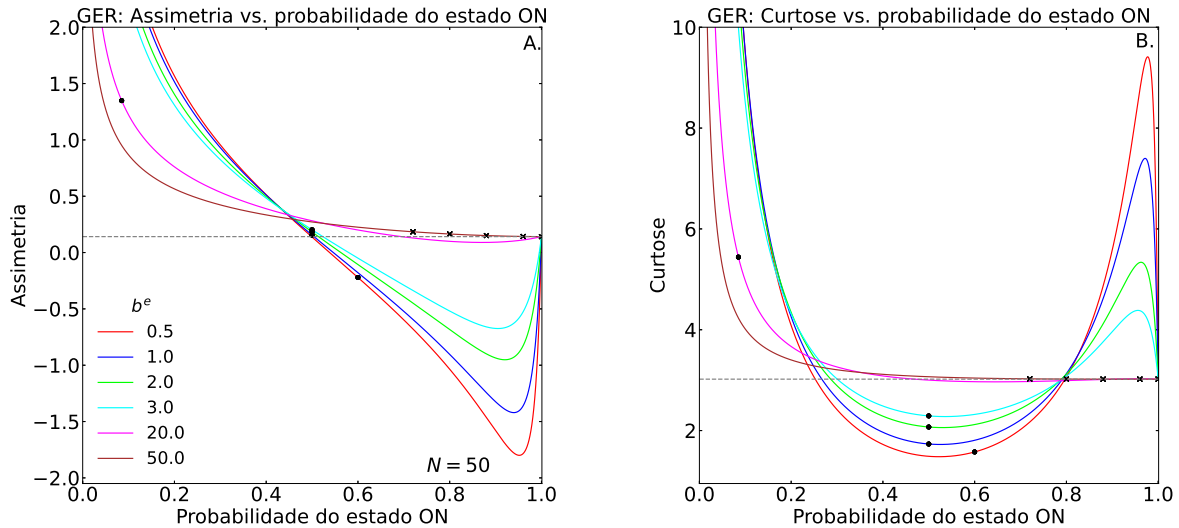
Tabela 1 – Valores aproximados dos parâmetros cinéticos nos gráficos da Fig. 3.

Gráfico A ($f; h_1$)	Gráfico B ($f; h_1$)
$(3,6 \times 10^{-1}; 1,4 \times 10^{-1})$	$(3,0 \times 10^{-3}; 2,0 \times 10^{-3})$
$(4,0 \times 10^{-1}; 1,0 \times 10^{-1})$	$(5,0 \times 10^{-3}; 5,0 \times 10^{-3})$
$(4,4 \times 10^{-1}; 6,0 \times 10^{-2})$	$(1,0 \times 10^{-2}; 1,0 \times 10^{-2})$
$(4,8 \times 10^{-1}; 2,0 \times 10^{-2})$	$(1,5 \times 10^{-2}; 1,5 \times 10^{-2})$
$(5,0 \times 10^{-1}; 0)$	$(1,7 \times 10^{-2}; 1,8 \times 10^{-1})$

Fonte – Adaptado de Giovanini *et al.* (2020)

Assimetria e curtose das distribuições de probabilidade

Figura 4 – Assimetria e curtose das distribuições do GER



Os gráficos **A.** e **B.**, respectivamente, mostram a assimetria e a curtose como função da probabilidade do GER estar no estado ON. Todos tem o mesmo valor fixo $N = 50$ e seguem o mesmo código de cores para os valores de b^e que aparece no gráfico **A.** As linhas tracejadas indicam os valores aproximados da assimetria e curtose no limite $a \rightarrow b^e$, sendo respectivamente: 0,14 e 3,02. Os pontos em forma de cruz marcam as distribuições quasi-Fano da Fig. **3A.** e os círculos marcam as distribuições super-Fano da Fig. **3B.**

Fonte – Guilherme Giovanini, 2021

As Figs. **4A.** e **4B.** mostram para o GER, a assimetria e a curtose como função da probabilidade do estado ON do gene para $N = 50$ fixo. Para $p_\alpha^e \sim 1$ e para valores de $b^e \gg 1$, temos assimetria e curtose próximas dos valores limites para o gene constitutivo ($a = b^e$), indicados pelas retas tracejadas como referência. Os valores de assimetria e curtose das distribuições quasi-Fano da Fig. **3A.** estão indicadas como cruzeiras na curva marrom, tendo características de caudas simétricas e espalhadas próximas de seus números médios de proteínas $\langle n^e \rangle$ semelhante a distribuição poissoniana ($a = b^e$). Mantendo o mesmo regime de b^e alto, a medida que p_α^e torna-se menor, distanciando-se de 1, os valores da assimetria e da curtose são positivos e crescem ao extremo se afastando do valor de referência, o que caracteriza distribuições com pontos que distam da média valores acima do desvio padrão, com viés para direita e bem espalhadas em torno de $\langle n^e \rangle$, alcançando valores mais altos de n com probabilidade não-nula, como é o caso da curva $b^e = 20$ da Fig. **3B.** (indicada pelo círculo na curva magenta na Fig. 4), que se assemelham com distribuições do tipo binomial negativa e aparecem em regime de *burst*.

Agora analisando $b^e \sim 1$, temos para $p_\alpha^e > 0,8$ assimetria negativa e curtose positiva, de maneira que quanto mais $b^e \rightarrow 0$ temos assimetrias com valores cada vez mais negativo e curtoses crescentes acima de 3, caracterizando distribuições de probabilidade com cauda para esquerda e espalhadas de maneira próxima de $\langle n^e \rangle$. A cauda fina para a esquerda indica possibilidades não-nulas de expressão gênica de proteínas com números abaixo e distantes de $\langle n^e \rangle$. Mantendo $b^e \sim 1$ e p_α^e se aproximando de 0,8, o valor da assimetria mantém-se negativo, porém mais próximo de 0, e a curtose sofre transição para valores abaixo da referência, que se torna nítido a medida que $p_\alpha^e \sim 0,5$. Essa região de p_α^e caracteriza a diversidade das distribuições super-Fano, a curtose permite massas de probabilidade bem afastadas da $\langle n^e \rangle$ e a assimetria assume valor negativo para $p_\alpha^e > 0,5$, e positivo no caso contrário, ocasionando distribuições bimodais, que a medida que a assimetria torna-se positiva, ocorre a transferência de massa de probabilidade do pico da direita para o da esquerda. Os pontos em forma de círculos sobre as curvas dos gráficos 4A. e 4B. demonstram as distribuições super-Fano da Fig. 3B. e seguem os mesmos valores de b^e referentes a curva que cada um está sobreposto. A medida que p_α^e se aproxima de 0, a curtose assume valores acima da referência em aproximadamente $p_\alpha^e < 0,3$, de maneira que para $p_\alpha^e \ll 0,3$, a assimetria e a curtose assumem valores positivos extremos, caracterizando distribuições super-Fano com cauda fina para direita e bem espalhadas em torno de $\langle n^e \rangle$, alcançando valores mais altos de n com probabilidade não-nula.

4.2 GAR

4.2.1 Distribuições de probabilidade marginais, totais e seus momentos

A distribuição $\Pi(\alpha_n^s, \beta_n^s)$ para o GAR tem os mesmos parâmetros a, N do GER como na Eq. (4.1) e os parâmetros b^s e z_0 :

$$a = \frac{f}{\rho}, \quad N = \frac{k}{\rho}, \quad z_0 = \frac{\rho}{\rho + h_2}, \quad b^s = \frac{f}{\rho + h_2} + \frac{N\rho h_2}{(\rho + h_2)^2} = az_0 + Nz_0(1 - z_0). \quad (4.20)$$

Esses parâmetros são definidos em termos da taxa de remoção de proteínas do citoplasma por degradação ou por ligar-se à região regulatória do gene, que é $\propto 1/(\rho + h_2)$ (RAMOS; REINITZ, 2019). As probabilidades α_n^s, β_n^s , e ϕ_n^s , são escritas usando as funções

geratrizes apresentadas em [Ramos e Hornos \(2007\)](#) e no capítulo 3 desta dissertação, de tal modo que:

$$\alpha_n^s = \frac{a}{b^s} \frac{z_0}{c} \frac{(a+1)_n}{(b^s+1)_n} \frac{(Nz_0)^n}{n!} M(a+n+1, b^s+n+1, -Nz_0^2), \quad (4.21)$$

$$\phi_n^s = \frac{1}{c} \frac{(a)_n}{(b^s)_n} \frac{(Nz_0)^n}{n!} M(a+n, b^s+n, -Nz_0^2), \quad (4.22)$$

$$\beta_n^s = \phi_n^s - \alpha_n^s, \quad (4.23)$$

onde $c = M(a, b^s, Nz_0(1-z_0))$ é a constante de normalização da distribuição. As probabilidades marginais de encontrar o gene nos estados ON e OFF são:

$$p_\alpha^s = \frac{a}{b^s} \frac{z_0}{c} M(a+1, b^s+1, Nz_0(1-z_0)) \quad \text{e} \quad p_\beta^s = 1 - p_\alpha^s. \quad (4.24)$$

Uma vez que $Nz_0(1-z_0)$ é não-nulo e se aproxima de zero, apenas para os casos limites de $z_0 \rightarrow 1$ ou $z_0 \rightarrow 0$, não é possível escrever p_α^s como uma simples função de a e b^s . Isso acontece porque de ON para OFF, a taxa de chaveamento depende de n . Apesar da conveniência de escrever as Eqs. (4.21)–(4.24) em termos de quatro parâmetros (a, b^s, N, z_0) , essas probabilidades têm apenas três parâmetros livres, como acontece com o GER.

Em ambos os modelos, o quarto parâmetro é obtido como uma taxa de decaimento para o estado estacionário ([IYER-BISWAS; HAYOT; JAYAPRAKASH, 2009](#); [RAMOS; INNOCENTINI; HORNOS, 2011](#)). Isso ocorre porque as probabilidades dos modelos para GAR ou GER são calculadas usando a função KummerM, que tem dois parâmetros, respectivamente, (a, b^s) ou (a, b^e) , e o argumento é dependente de um parâmetro adicional: N no modelo GER e uma combinação de dois parâmetros no GAR, a saber Nz_0^2 . Portanto, é preciso definir os vínculos entre a , b^s , N e z_0 fornecidas pela Eq. (4.20). Escolhemos N para ser escrito como uma função dos parâmetros restantes nas seguintes bases. Na literatura, encontramos as demonstrações das simetrias de Lie subjacentes à existência das soluções analíticas dos modelos estocásticos para o GAR ([RAMOS; HORNOS, 2007](#); [RAMOS; REINITZ, 2019](#)) e para o GER ([RAMOS *et al.*, 2010](#)). Como b^s e b^e são os invariantes da álgebra $\mathfrak{so}(2,1)$ satisfeitos para seus respectivos modelos, portanto são candidatos naturais a serem definidos como constantes. No modelo para o GAR no estado estacionário, o número máximo de proteínas N é dado em termos de b^s .

A mudança do estado do promotor do GAR pode ser caracterizada em termos dos parâmetros a e z_0 sendo o anterior proporcional a f e o último sendo inversamente proporcional a h_1 . Usamos Eq. (4.20) para obter

$$N = \frac{b^s - az_0}{z_0(1 - z_0)} > 0 \Rightarrow 0 < a < b^s/z_0, \quad (4.25)$$

onde a desigualdade de esquerda é imposta pela exigência de positividade de N e as desigualdades do lado direito são obtidas por $0 < z_0 < 1$, conforme a Eq. (4.20).

A p -ésima derivada da função geratriz $\tilde{\phi}^s$ é dada por

$$\left. \frac{d^p \tilde{\phi}^s}{dz^p} \right|_{z=1} = \frac{(Nz_0)^p}{c} \frac{(a)_p}{(b^s)_p} M(a + p, b^s + p, \eta(1)). \quad (4.26)$$

Substituindo as derivadas em Eqs. (3.31)–(3.34) e lembrando que $\eta(1) = Nz_0(1 - z_0)$, é possível obter os quatro primeiros momentos para o GAR:

$$\langle n \rangle = N \frac{a}{b^s} \frac{z_0}{c} M(a + 1, b^s + 1, \eta(1)); \quad (4.27)$$

$$\langle n^2 \rangle = \langle n \rangle \left[1 + Nz_0 \frac{a + 1}{b^s + 1} \frac{M(a + 2, b^s + 2, \eta(1))}{M(a + 1, b^s + 1, \eta(1))} \right]; \quad (4.28)$$

$$\langle n^3 \rangle = \langle n \rangle + [\langle n^2 \rangle - \langle n \rangle] \cdot \left[3 + Nz_0 \frac{a + 2}{b^s + 2} \frac{M(a + 3, b^s + 3, \eta(1))}{M(a + 2, b^s + 2, \eta(1))} \right]; \quad (4.29)$$

$$\begin{aligned} \langle n^4 \rangle = \langle n \rangle - [\langle n^2 \rangle - \langle n \rangle] \cdot \left[11 + 3Nz_0 \frac{a + 3}{b^s + 3} \frac{M(a + 4, b^s + 4, \eta(1))}{M(a + 3, b^s + 3, \eta(1))} \right] + \\ [\langle n^3 \rangle - \langle n \rangle] \cdot \left[6 + Nz_0 \frac{a + 3}{b^s + 3} \frac{M(a + 4, b^s + 4, \eta(1))}{M(a + 3, b^s + 3, \eta(1))} \right]. \end{aligned} \quad (4.30)$$

Note que os momentos das distribuições de probabilidade do GAR são dependentes da função KummerM, uma vez que dependem da Eq. (4.26), tornando inviável expressões explícitas dos momentos padronizados como feito para o GER.

4.2.2 Caracterização das flutuações das distribuições de probabilidade

O número médio de proteínas expressadas por GAR, denotado por $\langle n^s \rangle$, é (RAMOS; HORNOS; REINITZ, 2015):

$$\langle n^s \rangle = p_\alpha^s N, \quad (4.31)$$

como pode ser notada ao empregar a Eq. (4.24) na Eq. (4.27). Para calcular a covariância entre (ν_m, n) , é usada a Eq. (3.36) que resulta:

$$\xi^s = az_0 \frac{N - \langle n^s \rangle}{1 - z_0} - \langle n^s \rangle^2. \quad (4.32)$$

Esta expressão foi apresentada em Ramos, Hornos e Reinitz (2015) e apresentamos uma demonstração no Apêndice E. O fator de Fano para o GAR é então escrito como:

$$\mathcal{F}^s = 1 + \left[\frac{az_0}{\langle n^s \rangle} \frac{N - \langle n^s \rangle}{1 - z_0} - \langle n^s \rangle \right], \quad (4.33)$$

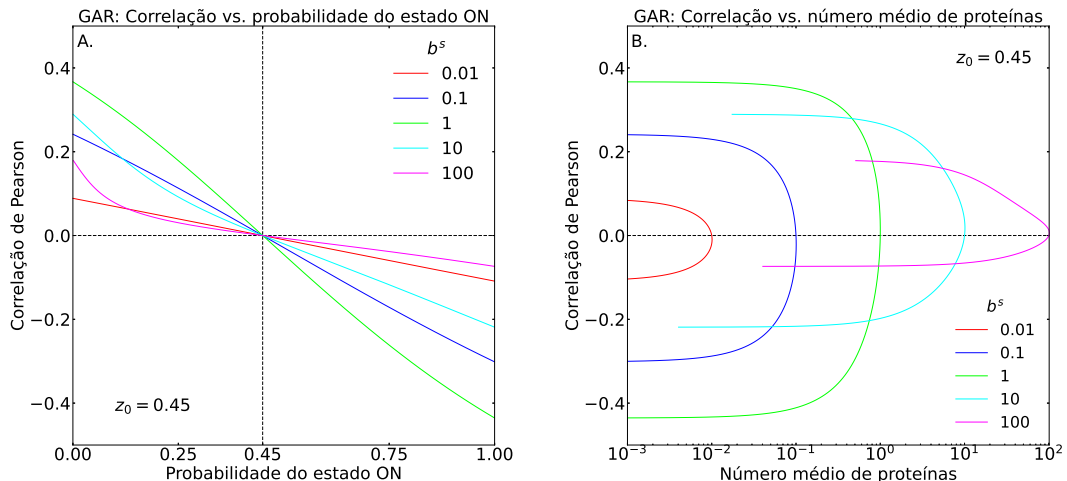
o qual é uma função não-linear dos parâmetros do modelo. Neste modelo, a covariância pode ser positiva, nula e negativa, que resulta respectivamente em distribuições de probabilidade super-Fano, Fano, e sub-Fano (RAMOS; HORNOS; REINITZ, 2015). O último regime mostra por que a autorrepressão permite a redução de ruído na expressão gênica.

A assimetria e a curtose também são funções que podem ser calculadas ao substituir as Eqs. (4.27)–(4.30) nas Eqs. (3.12) e (3.13). Tais substituições acarretam em equações extensas onde não mostramos sua forma explícita, mas calculamos os valores numéricos de interesse por substituições em etapas.

4.2.3 Resultados

Correlação de Pearson entre o número de proteínas e o estado do gene

Figura 5 – Correlação em função da probabilidade do gene no estado ON e do número médio de proteínas para o GAR.



Fonte – Giovanini *et al.* (2020)

Os gráficos **A.** e **B.** mostram, respectivamente, a correlação em função da probabilidade de encontrar o promotor no estado ON e como função do número médio de proteínas para GAR. O valor da correlação depende da relação entre os parâmetros a e b^s : é nulo

para $a = b^s$, positivo para $a < b^s$ e negativo para $a > b^s$. A correlação se comporta quase linearmente como função da probabilidade do estado ON. A condição $a = b^s$ resulta em $p_\alpha^s = z_0$, que é evidenciada pelas retas tracejadas do gráfico **A**. O gráfico **B**. mostra que o GAR permite a existência de duas distribuições de probabilidade tendo o mesmo valor de $\langle n^s \rangle$ e correlações de sinais opostos, implicando na possibilidade da síntese de proteína tanto com pequena ou com grande amplitude de ruído. Para um valor de z_0 fixado, quando $a = b^s$, $\langle n^s \rangle$ assume valor máximo igual à Nz_0 . Para $b^s \geq 1$, o valor absoluto de máximo ou de mínimo na correlação é reduzido e tende a 0. É possível notar que, em ambos gráficos **A**. e **B**., a correlação aproxima-se de um comportamento linear para baixos valores de b^s (curvas vermelha, azul e verde) enquanto que efeitos não-lineares aparecem com o aumento de b^s (curvas ciano e magenta).

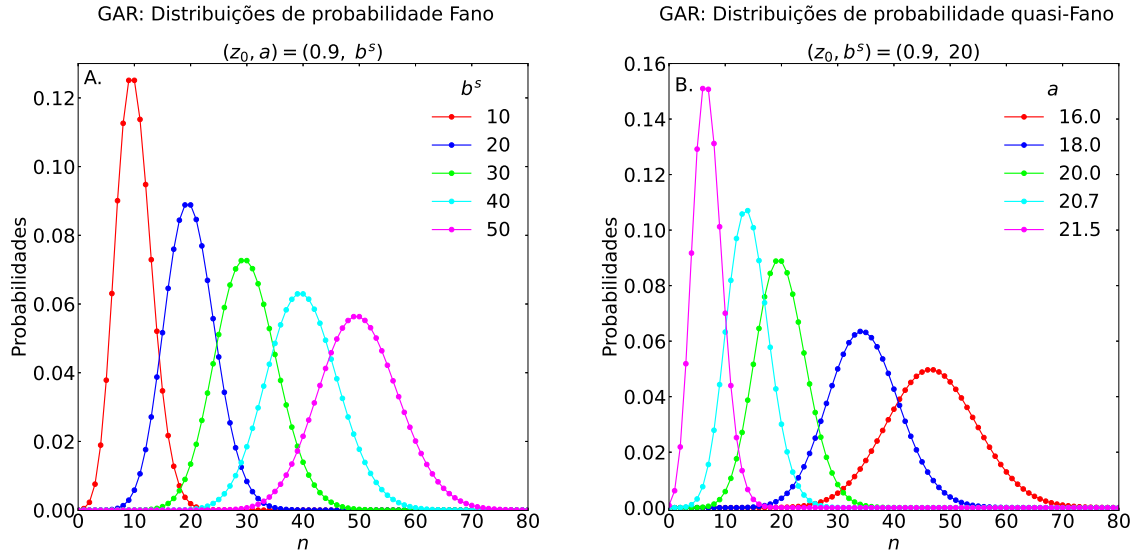
O Gráfico **A**. mostra que para $0 \leq p_\alpha^s \leq z_0$ a correlação para o GAR é positiva e decai até 0 a medida que $p_\alpha^s \rightarrow z_0$ juntamente com $a \rightarrow b^s$. Se $z_0 \leq p_\alpha^s \leq 1$, a medida que p_α^s se aproxima de 1, a correlação continua diminuindo abaixo de 0, caracterizando um regime de baixo ruído, que é exclusivo do GAR.

O Gráfico **B**. revela a degenerescência resultante das simetrias do modelo estocástico para o GAR (RAMOS; HORNOS, 2007; RAMOS; REINITZ, 2019), pois para um par (b^s, z_0) fixo é possível obter o mesmo $\langle n^s \rangle$ para dois distintos valores a , o que não é estabelecido quando o GAR é modelado de modo determinístico (RAMOS; REINITZ, 2019).

Distribuições de probabilidade Fano e quasi-Fano

Distribuições de probabilidade Fano e quasi-Fano que governam número de proteínas obtidas no modelo estocástico para um GAR são mostrados na Fig. 6. O gráfico **A**. mostra as distribuições Fano que governam os números de proteínas expressos, o que acontece quando $a = b^s$. Neste limite, a Eq. (4.24) se torna $p_\alpha^s = z_0$, por causa de $M(A, B, \eta) = M(1 + A, 1 + B, \eta) = e^\eta$ para $A = B$, como pode ser visto na Eq. (3.23). A probabilidade de encontrar n proteínas na Eq. (4.22) torna-se $\phi_n^s = \frac{(Nz_0)^n}{n!} \exp(-Nz_0)$, que corresponde a uma distribuição poissoniana com $\langle n^s \rangle = Nz_0 = b^s$ por causa de $N = \frac{b^s}{z_0}$ como mostrado pela Eq. (4.25). O gráfico **B**. apresenta distribuições quasi-Fano obtidas fixando b^s e tomando $a \approx b^s$. Tomamos valores de a maiores e menores que b^s tal que a

Figura 6 – Distribuições Fano e quasi-Fano para o GAR.



Em ambos os gráficos $z_0 = 0,9$. O gráfico **A.** tem distribuição estritamente Fano ($\mathcal{F} = 1$) com $a = b^s$, onde os momentos da distribuição são $(\mathcal{F}^s; \xi^s; \langle n^s \rangle; p_\alpha^s) = (1; 0; b; z_0)$. O gráfico **B.** tem distribuições quasi-Fano com ($\mathcal{F} \approx 1$) e os momentos são apresentados na ordem crescente de a : $\mathcal{F}^s \approx (1,4; 1,1; 1,0; 0,98; 0,98)$; $\xi^s \approx (17,5; 7,7; 0; -0,3; -0,2)$; $\langle n^s \rangle \approx (47; 35; 20; 14; 7)$; $p_\alpha^s \approx (0,75; 0,82; 0,9; 0,93; 0,97)$.

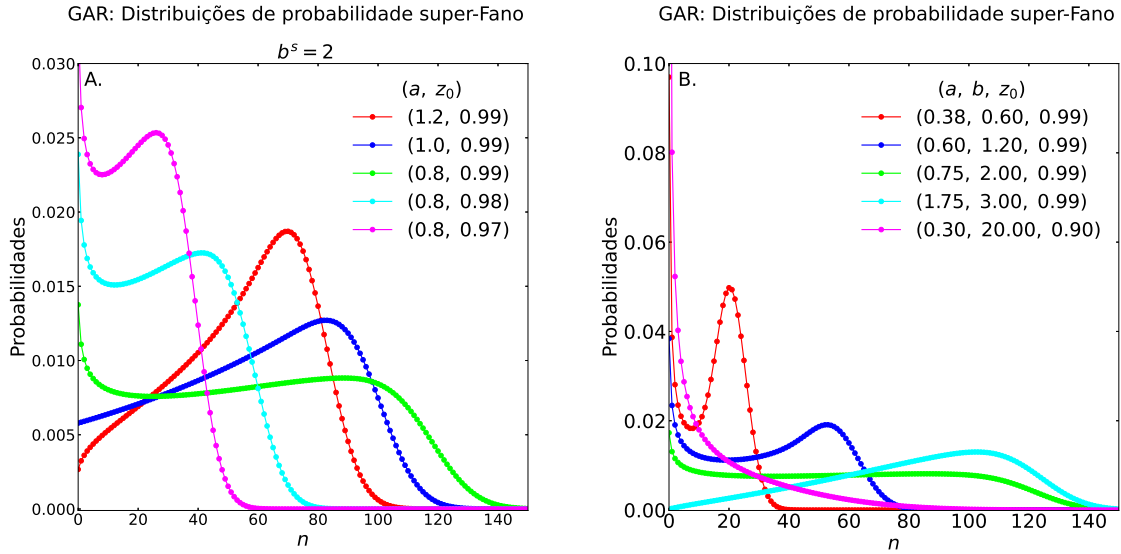
Fonte – Giovanini *et al.* (2020)

covariância será, respectivamente, negativa e positiva. O número médio de proteínas é reduzido quando a aumenta por causa do parâmetro da distribuição poissoniana que é $\frac{b^s - az_0}{z_0(1 - z_0)}$. Uma vez que as distribuições são de pico único, a redução no número médio de proteínas implica num deslocamento do pico para a esquerda.

Distribuições de probabilidade super-Fano

Distribuições de probabilidade super-Fano para ambas estratégias são apresentadas na Fig. 7. Seleccionamos valores de parâmetros que resultam em distribuições de probabilidade espalhadas, como as em forma de mesa e as bimodais. O gráfico **A.** mostra distribuições de probabilidade para o GAR onde tomamos um valor fixo para b^s . As escolhas de a ou z_0 ajudam a entender seus efeitos nas distribuições de probabilidade. Os valores de z_0 são definidos de forma que a taxa de degradação seja a principal causa da remoção de proteínas do citoplasma. O gráfico **B.** mostra uma segunda família de distribuições super-Fano obtidas para o GAR. Tomamos dois valores para z_0 e variamos (a, b) . O gráfico **A.** tem cinco distribuições de probabilidades, onde as exibidas em magenta e ciano são bimodais. A linha verde mostra uma distribuição aproximada à forma de

Figura 7 – Distribuições super-Fano para o GAR.



As distribuições super-Fano ocorrem quando $a < b^s$. No gráfico **A.** há na legenda a indicação dos valores do par de parâmetros (a, z_0) e os momentos das distribuições são dados na mesma ordem que aparecerem esses pares no gráfico: $\mathcal{F}^s \approx (9,8; 14,7; 21,3; 11,2; 7,8)$; $\xi^s \approx (472,3; 808,4; 1225,9; 310,5; 139,8)$; $\langle n^s \rangle \approx (53,7; 58,9; 60,4; 30,5; 20,5)$; $p_\alpha^s \approx (0,66; 0,58; 0,50; 0,49; 0,49)$. No gráfico **B.**, $z_0 = 0,99$ para os quatro menores valores de b^s , e $z_0 = 0,9$ para o valor remanescente de b^s . Os momentos são apresentados na ordem crescente de b^s : $\mathcal{F}^s \approx (5,6; 13,8; 23,5; 11,3; 25,9)$; $\xi^s \approx (69,0; 405,2; 1339,0; 857,9; 352,6)$; $\langle n^s \rangle \approx (15; 34; 60; 83; 14)$; $p_\alpha^s \approx (0,66; 0,56; 0,47; 0,65; 0,065)$.

Fonte – Giovanini *et al.* (2020)

mesa. As linhas azul e vermelha indicam distribuições unimodais espalhadas por causa da probabilidade significativa em encontrar de 0 a 100 proteínas enquanto que suas médias são ≈ 50 . O gráfico **B.** mostra duas distribuições de probabilidade bimodais, mostradas em azul e vermelho. A distribuição em verde é em forma de mesa e a distribuição em ciano é unimodal, ambas com probabilidade significativa de se encontrar de 0 a 130 proteínas. A distribuição em magenta pode ser aproximada a uma distribuição binomial negativa, a qual é uma típica distribuição quando se tem regimes de *burst* transcricional ou translacional (SHAHREZAEI; SWAIN, 2008; YVINEC *et al.*, 2017).

Os valores dos parâmetros das distribuições de probabilidade dos gráficos **A.** e **B.** foram reescritos em termos das constantes cinéticas na Eqs. (3.5)–(3.8). Consideramos $\rho = 0,01 \text{ min}^{-1}$ e conseqüentemente apresentamos os valores aproximados dos parâmetros cinéticos na Tab. 2 seguindo a ordem dos valores do parâmetro em cada legenda dos seus correspondentes gráficos, onde k , f , e h_2 são dados em min^{-1} .

Tabela 2 – Valores aproximados dos parâmetros cinéticos nos gráficos da Fig. 7.

Gráfico A ($N; k; f; h_2$)	Gráfico B ($N; k; f; h_2$)
(80; 0,8; 1×10^{-2} ; 1×10^{-4})	(20; 0,2; 4×10^{-3} ; 1×10^{-4})
(100; 1,0; 1×10^{-2} ; 1×10^{-4})	(60; 0,6; 6×10^{-3} ; 1×10^{-4})
(100; 1,0; 8×10^{-3} ; 1×10^{-4})	(130; 1,3; 8×10^{-3} ; 1×10^{-4})
(60; 0,6; 8×10^{-3} ; 2×10^{-4})	(130; 1,3; 2×10^{-2} ; 1×10^{-4})
(40; 0,4; 8×10^{-3} ; 3×10^{-4})	(220; 2,2; 3×10^{-3} ; 1×10^{-3})

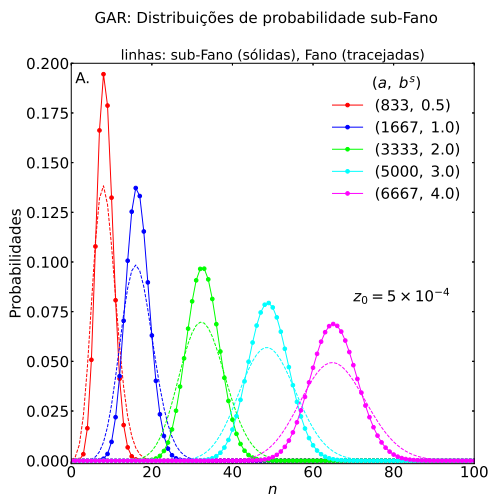
Fonte – Giovanini *et al.* (2020)

No Gráfico **A.**, a taxa de chaveamento do promotor de OFF para ON é da ordem de 10^{-2} min^{-1} , sendo esta a mesma ordem da taxa efetiva de chaveamento de ON para OFF, $\frac{h_2 k}{\rho + h_2}$ (RAMOS; REINITZ, 2019) e da taxa de degradação das proteínas, ocasionando distribuições de dois picos quando os processos bioquímicos são mais lentos. No Gráfico **B.** temos a taxa efetiva de chaveamento de ON para OFF dada em min^{-1} aproximadamente por $(2 \times 10^{-3}; 6 \times 10^{-3}; 10^{-2}; 10^{-2}; 0,2)$ apresentados na ordem crescente de N . Para o valor máximo N , a taxa efetiva de chaveamento de ON para OFF é quase centenas de vezes maior que a taxa de chaveamento do estado de OFF para ON. Perceba que a taxa de síntese é centenas de vezes maior que a degradação, o que pode ser considerado um processo de produção de proteína bastante eficiente. Nesse limite, o promotor permanece principalmente OFF e a distribuição se aproxima de uma binomial negativa como ocorre em síntese de proteínas em regime de *burst* (YVINEC *et al.*, 2017).

Distribuições de probabilidade sub-Fano

As distribuições de probabilidade sub-Fano e Fano para os números de proteínas produzidos a partir do GAR e fatores de Fano *versus* a probabilidade do gene estar ligado, ou *versus* o número médio de proteínas são apresentados nas Figs. 8 e 9. O gráfico da Fig. 8 mostra distribuições de probabilidade sub-Fano com fator de Fano $\approx 0,5$. Para permitir uma comparação, sobrepomos cada sub-Fano com um distribuição Fano com o mesmo número médio de proteínas com linhas mais finas e com as mesmas cores. A diferença no fator de Fano das duas distribuições aumentam com $\frac{\langle n \rangle}{2}$ e tornam as curvas mais distinguíveis. Os gráficos **A.** e **B.** da Fig. 9 foram construídos usando os mesmos parâmetros e seguem o mesmo código de cores. Os gráficos mostram um regime assintótico de ruído para o GAR quando o fator de Fano igual à 0,5, e que pode ser alcançado por um

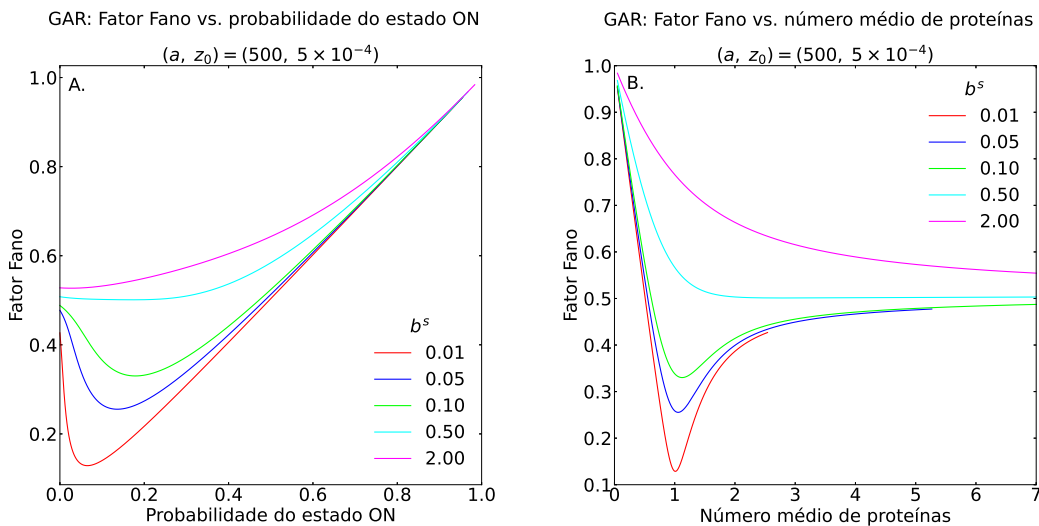
Figura 8 – Distribuições sub-Fano para o GAR.



O gráfico **A.** mostra cinco pares de distribuições Fano e sub-Fano e em cada par temos o mesmo número médio de proteínas. Todas distribuições têm $z_0 = 5 \times 10^{-4}$. As distribuições Fano estão apresentadas com linhas tracejadas e com os momentos iguais a $\mathcal{F}^s = 1$, $\xi^s = 0$, $p_\alpha^s = 5 \times 10^{-4}$. As distribuições sub-Fano estão em linhas sólidas e seus momentos em ordem crescente de b^s são: $\mathcal{F}^s \approx (0,50; 0,51; 0,51; 0,51; 0,52)$; $\xi^s \approx (-4,2; -8,1; -15,9; -23,8; -31,6)$; $\langle n^s \rangle \approx (8,4; 16,5; 32,8; 49,0; 65,3)$; $p_\alpha^s \approx (0,050; 0,049; 0,049; 0,049; 0,048)$.

Fonte – Giovanini *et al.* (2020)

Figura 9 – Fator de Fano em função da probabilidade do gene no estado ON e do número médio de proteínas para o GAR.



Os parâmetros (a, b^s) tem os mesmos valores em ambos os gráficos.

Fonte – Giovanini *et al.* (2020)

número médio de proteínas arbitrário e com p_α^s se aproximando zero. Observe que o fator de Fano se aproxima de 1 conforme p_α^s aumenta, como esperado para um gene que está no estado ON na maior parte do tempo. Temos um regime infra-Fano para um número

médio de proteínas igual a 1, com p_α^s se aproximando de 0 quando o fator de Fano fica arbitrariamente menor (RAMOS; HORNOS; REINITZ, 2015).

Na Tab. 3, os valores dos parâmetros que geram as distribuições de probabilidade do gráfico 8A. são reescritos em termos das constantes cinéticas nas Eqs. (3.5)–(3.8). Com $z_0 = 5 \times 10^{-4}$ fixo e considerando $\rho = 0,01 \text{ min}^{-1}$, temos $h_2 \sim 20 \text{ min}^{-1}$. Para cada par de distribuições sub-Fano e Fano tendo o mesmo valor de $\langle n \rangle$, os pares (a, b^s) resultam nos seguintes valores aproximados dos parâmetros cinéticos, onde k e f são dados em min^{-1} e K indica o equilíbrio da afinidade de ligação da proteína regulatória.

Tabela 3 – Valores aproximados dos parâmetros cinéticos no gráfico da Fig. 8.

b^s	Sub-Fano ($N; k; f; K = h_2/f$)	Fano ($N; k; f; K = h_2/f$)
0,5	(167,1; 1,7; 8,3; 2,4)	(1×10^3 ; 10; 5×10^{-3} ; 4×10^3)
1	(333,2; 3,3; 16,7; 1,2)	(2×10^3 ; 20; 1×10^{-2} ; 2×10^3)
2	(667,3; 6,7; 33,3; 0,6)	(4×10^3 ; 40; 2×10^{-2} ; 1×10^3)
3	(1000,5; 10,0; 50,0; 0,4)	(6×10^3 ; 60; 3×10^{-2} ; 7×10^2)
4	(1333,7; 13,3; 66,7; 0,3)	(8×10^3 ; 80; 4×10^{-2} ; 5×10^2)

Fonte – Giovanini *et al.* (2020)

A ordem de K para as distribuições sub-Fano são ~ 1 , enquanto que para as distribuições Fano são da ordem de $\sim 10^3$. As taxas de chaveamento de estado do gene de OFF para ON nas distribuições sub-Fano são da ordem de centenas de vezes mais rápidas que aquelas das distribuições Fano. Por outro lado, as taxas efetivas de chaveamento de estado ON para OFF, $\frac{h_2 k}{\rho + h}$, dadas em min^{-1} , são aproximadamente: para distribuição sub-Fano (1,7; 3,3; 6,7; 10; 13,3) e para Fano (10; 20; 40; 60; 80), seguindo a ordem para valores de b^s dados na legenda. Portanto, essas taxas efetivas de desligamento são quase 10 vezes mais rápidas nas distribuições Fano. Ao comparar as taxas efetivas entre os dois estados do promotor, temos que processos sub-Fano tem uma transição de OFF para ON ~ 10 vezes mais rápida, enquanto que nos processos Fano é $\sim 10^{-3}$ vezes mais devagar. Isso faz com que as probabilidades do estado ON no regime sub-Fano e, portanto, a duração do estado ON, seja em cerca de cem vezes maior do que no regime Fano. Os valores da taxa de síntese fornecem o efeito compensatório para assegurar que os dois regimes de ruídos terão o mesmo número médio de produtos gênicos.

Assimetria e curtose das distribuições de probabilidade

Os gráficos de assimetria e da curtose para o GAR foram obtidos fixando o valor de b^s e N , variando o valor de a de maneira a respeitar a condição $0 < z_0 < 1$, intervalo dotado de interpretação biológica (RAMOS; REINITZ, 2019). Sendo assim, o valor de z_0 para cada ponto (a, b^s, N) foi calculado por

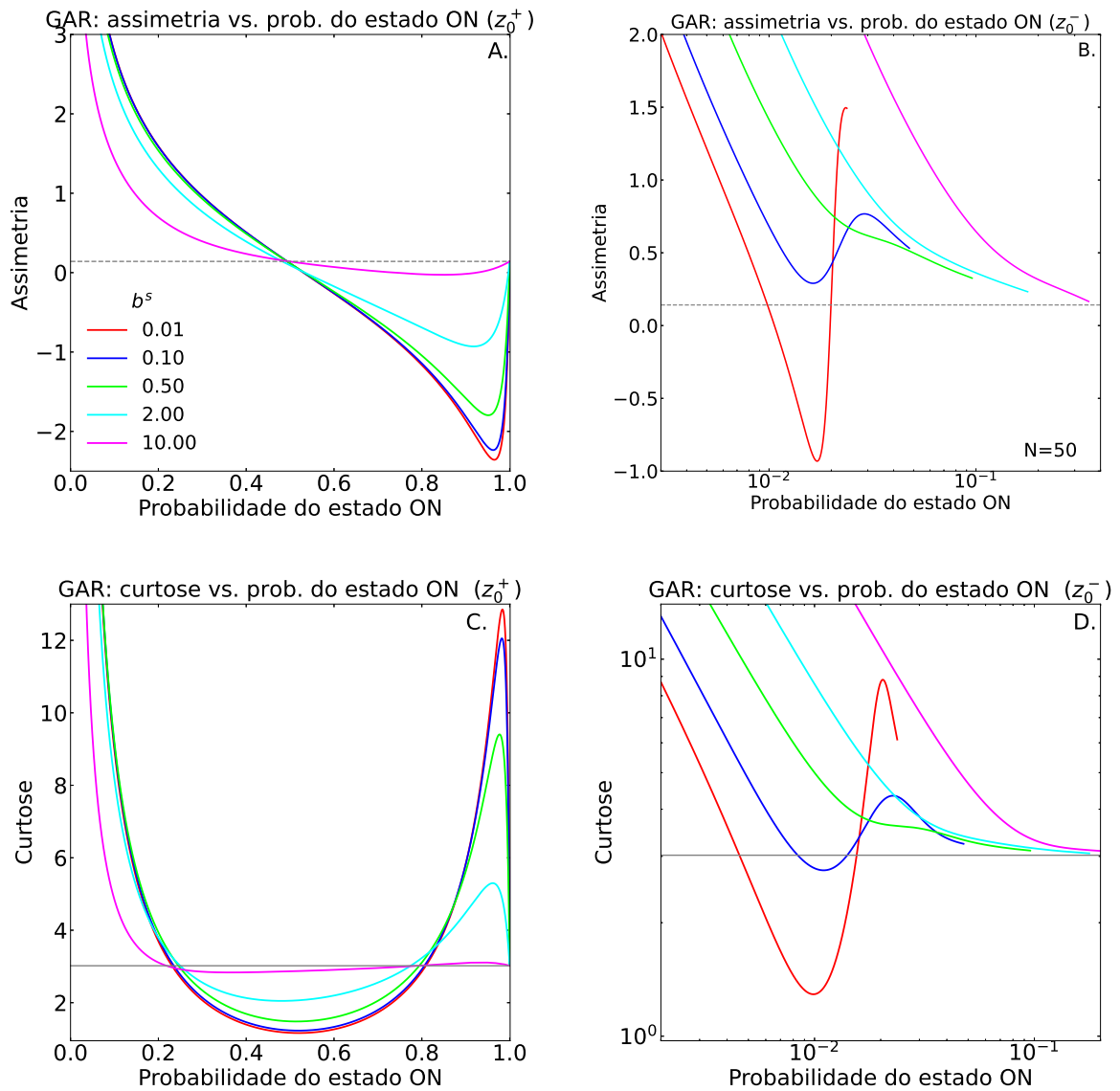
$$z_0^\pm = \frac{1}{2} \left(1 + \frac{a}{N} \right) \left[1 \pm \sqrt{1 - \frac{4Nb^s}{(N+a)^2}} \right], \quad (4.34)$$

onde para z_0^+ é apenas aplicado na situação $a \leq b^s$, que se limita a regimes Fano e super-Fano no modelo de expressão gênica.

O comportamento da assimetria e da curtose no GAR para valores em z_0^+ , mostrados gráficos **A.** e **C.** da Fig. 10, são muito similares ao comportamento encontrado para o GER, nos gráficos da Fig. 4. Nota-se apenas uma diferença na escala, onde as amplitudes de assimetria e curtose para $p_\alpha^s \rightarrow 1$ são maiores no GAR em relação ao GER. Assim, a análise do comportamento desses momentos padronizados para o z_0^+ do GAR explica as mesmas formas das distribuições de probabilidades descritas na análise qualitativa feita para o GER.

Para z_0^- , temos a possibilidade do regime sub-Fano, o que não ocorre no GER. Quando $p_\alpha^s \rightarrow 0$, temos $\langle n^s \rangle \sim 1$ e extremos valores positivos de assimetria e de curtose, caracterizando distribuições de probabilidade com pontos que distam da média com valores maior que o desvio padrão e com viés para a direita. A medida que p_α^s aumenta se aproximando de 10^{-2} , os valores de $\langle n^s \rangle$ também aumentam e temos duas situações distintas com relação ao valores de b^s . Quando b^s é baixo, a assimetria e a curtose caem abruptamente de maneira atingir seus valores mínimos, ficando abaixo da referência do gene constitutivo e tais mínimos da assimetria assumem valor negativo. Isso indica distribuições bem espalhadas e com maior massa de probabilidade localizada à esquerda da média. No caso, a distribuição de probabilidade se apresenta na forma de mesa, abrangendo um intervalo de n entre 0 e 5. Essa região é sensível à variação de probabilidade do estado do gene, que se aumentada, os valores da média, da assimetria e da curtose aumentam rapidamente acima do valor de referência atingindo valores máximos locais, caracterizando distribuições com picos e alguns pontos que desviam para a direita do valor médio. Quanto menor for o valor b^s , menor é o valor máximo limite para p_α^s , que fica evidente para os valores de b^s igual a 0,01 e 0,1. Perceba que esse aumento de b^s atenua os valores nos

Figura 10 – Assimetria e curtose das distribuições do GAR.



Os gráficos **A.** e **B.**, respectivamente, mostram a assimetria e para z_0^+ e z_0^- como função da probabilidade do GAR estar no estado ON. Os gráficos **C.** e **D.** apresentam esta mesma relação para curtose. Todos tem o mesmo valor fixo $N = 50$ e seguem o mesmo código de cores para os valores de b^s e que aparecem no gráfico **A.** As linhas tracejadas indicam os valores aproximados da assimetria e curtose para $a = b^s$, sendo respectivamente: 0,14 e 3,02.

Fonte – Guilherme Giovanini, 2021

extremos locais para a assimetria e curtose. Em regimes de $b^s \gg 0,1$, a medida que $p_\alpha^s \gg 10^{-2}$ as distribuições que assumem a forma de mesa vão tornando-se cada vez mais próximas da distribuição do gene constitutivo, com um pico cada vez menor e sempre se deslocando para direita, pois os altos valores de assimetria e curtose se aproximam pelo lado de cima dos valores de referência.

4.3 Comparações dos resultados entre o GER e o GAR

Ao comparar o gráfico da Fig. 2 para o GER, com os gráficos da Fig. 5 para o GAR, nota-se que o valor máximo da correlação para o GAR é menor que para o GER e isso ocorre para $b^s = 1$. Os parâmetros b^s e b^e tem a mesma interpretação bioquímica (RAMOS; REINITZ, 2019), e portanto, quando as duas escalas de tempos do modelo binário para o GAR são as mesmas ($b^s = 1$), a correlação entre o estado do promotor e o número de proteínas conseguem alcançar os maiores valores absolutos. No GAR, a correlação é, mais intensa quando $b^e \rightarrow 1$ e mais fraca em comparação com o modelo para o GER por causa do acoplamento mais forte da taxa de desligamento, h_1 , causado pela dependência do número de proteínas n .

O gráfico A. da Fig. 6 mostra que o ruído no número de proteínas sintetizadas por um GAR é o mesmo que na síntese de proteínas por um gene constitutivo não regulado. Este regime que é possível no GAR, mas não é possível no GER, e ocorre quando $a = Nz_0 = \langle n^s \rangle$. Esse resultado é derivado das Eqs. (4.20), (4.24) e (4.31), onde os termos dos parâmetros bioquímicos tornam-se

$$\langle n^s \rangle = \frac{f}{\rho} = \frac{k}{\rho + h_2}. \quad (4.35)$$

Consequentemente, a taxa de chaveamento entre os estados do GAR comporta-se como um gene constitutivo com taxas de síntese e degradação dadas por k e $\rho + h_2$, respectivamente. Isso corresponde a um regime "ressonante", no qual a taxa média do chaveamento ON-OFF-ON do GAR, que é definida como $f + kh_1/(\rho + h_2)$ (RAMOS; REINITZ, 2019), é igual à taxa de síntese k . Isso pode ser verificado depois de usar a relação entre as taxas de reação dadas pela Eq. (4.35).

O Gráfico A. da Fig. 3 e o gráfico B. da Fig. 6 mostram distribuições quasi-Fano ($a \sim b$). Também existe para o GER, uma condição no qual as taxas de sínteses e de chaveamento do gene são as mesmas, $f + h_1 = k$, o que implica em $b^e = N$ — veja a Eq. (4.1). Portanto, o fator de Fano para a distribuição de probabilidade neste limite é $\mathcal{F}^e = 1 + \frac{N}{1+N}(1 - p_\alpha^e)$, onde $\mathcal{F}^e \rightarrow 2 - p_\alpha^e$ para $N \gg 1$. Sendo assim, o GER não tem limite ressonante que resulta numa expressão gênica efetivamente constitutiva como ocorre para o GAR.

O gráfico B. da Fig. 3 e o gráfico B. da Fig. 7 mostram que o modelo binário estocástico gera distribuições de probabilidade parecidas para o GER e para o GAR,

quando tem-se parâmetros cinéticos de mesma ordem gerando as distribuições. Isso pode vir a ser um elemento de confusão na inferência da regulação gênica baseada no ruído contido no processo de expressão dos genes. Este problema pode ser apenas de cunho teórico, se as diferenças nas distribuições geram flutuações no número de proteínas situadas dentro dos limites de detecção dos sistemas biológicos. Cerca de 40% dos genes da *E. coli* são autorreprimidos (HORNOS *et al.*, 2005), o qual torna interessante determinar se este fato é devido a possibilidade mais ampla de intervalos de regimes da expressão do gene estocástico permitido pelo GAR, ou se um gene regulado localizado em *downstream* (sentido 3') pode detectar as particularidades nas flutuações do número de proteínas gerados por GER ou GAR.

Na Fig. 9, o gráfico B mostra que o fator de Fano assintótico para o regime sub-Fano ocorre para uma ampla gama de valores de $\langle n \rangle$. O gráfico A mostra que o fator de Fano assintótico ocorre para pequenos valores de p_α^s combinados com grandes valores para a taxa de síntese. Este regime de parâmetros do GAR é qualitativamente distinto do GER, pois o último estaria no limite de *burst* da expressão gênica (YVINEC *et al.*, 2017).

Tabela 4 – Regimes de flutuações dados pelos parâmetros cinéticos no GER e no GAR.

	GER	GAR
$a = b$ (Fano)	-	$z_0 = \frac{f}{k}$ e $h_2 = \rho \left(\frac{k}{f} - 1 \right)$
$a \sim b$ (quasi-Fano)	$h_1 \rightarrow 0$	$z_0 \sim \frac{f}{k}$ e $h_2 \sim \rho \left(\frac{k}{f} - 1 \right)$
$a < b$ (super-Fano)	$h_1 > 0$	$0 < z_0 < \frac{f}{k}$ e $h_2 > \rho \left(\frac{k}{f} - 1 \right)$
$a > b$ (sub-Fano)	-	$\frac{f}{k} < z_0 < 1$ e $0 < h_2 < \rho \left(\frac{k}{f} - 1 \right)$

Fonte – Guilherme Giovanini, 2020

Na Tab. 4 estão discriminados os intervalos dos parâmetros cinéticos em relação aos tipos de ruídos sofrido pela distribuição de probabilidade dos modelos GER e GAR. Para o GER, o regime de ruído é definido por h_1 , a taxa de chaveamento do estado ON para o estado OFF do promotor do gene. Quanto maior for o valor de h_1 , menor será a probabilidade p_α^e do GER estar ativado e para $h_1 \rightarrow 0$, o valor de $p_\alpha^e \rightarrow 1$. Pela formulação de $p_\alpha^e = \frac{a}{b^e} = \frac{f}{f+h_1}$, temos que a taxa de ativação do gene f define a escala dos valores para h_1 para a determinação dos regimes de ruídos.

No GAR, os regimes de ruídos são delimitados por $\frac{f}{k}$, a razão entre taxa de chaveamento do estado do gene de OFF para ON e a taxa de síntese de proteínas. Quando a fração entre a taxa de degradação de proteínas e a taxa que as proteínas que

tornam-se indisponíveis no citoplasma, $z_0 = \frac{\rho}{\rho+h_2}$, for igual ou em torno de $\frac{f}{k}$ caracteriza respectivamente os regimes Fano e quasi-Fano. Quando $z_0 < \frac{f}{k}$, temos o regime super-Fano e no caso contrário, temos o regime sub-Fano. Como z_0 é função de h_2 , o valor do último também caracteriza o regime de ruídos de acordo com qual intervalo de valores que h_2 pertence, como é mostrado na Tab. 4.

4.4 Discussões

A modelagem matemática de sistemas biomoleculares tem mostrado um grande potencial ao auxiliar na escolha de abordagens experimentais sobre sistemas que costumam ter um alto grau de complexidade, como é o caso da expressão gênica. Modelos simplificados para a regulação gênica, como o modelo do gene binário estocástico que analisamos aqui (PECCOUD; YCART, 1995; HORNOS *et al.*, 2005; RAMOS *et al.*, 2010), conseguem ilustrar mecanismos como regimes de expressão gênica que resultam em distribuições de probabilidades bimodais para o número de produtos gênicos e distribuições em *burst*, onde ocorre transcrição gênica com eficiência durante o curto estado ON e decaimento desses produtos durante o longo estado OFF (YVINEC *et al.*, 2017).

A análise do ruído no processo de expressão gênica via o sinal de fluorescência dos marcadores introduzidos nos genes de interesse, é chave para validação de modelos, uma vez que a informação obtida experimentalmente é a atividade gênica e a concentração de produtos gênicos representados pela intensidade de fluorescência do experimento. A caracterização das flutuações no número dos produtos gênicos aqui apresentada, nos mostrou regimes distintos de ruídos e uma variedade de formatos para as distribuições de probabilidades resultantes destes regimes.

O GER tem dois regimes de ruídos: aquele que é próximo a situação do gene constitutivo (quasi-Fano) e o regime com ruídos altos que fogem da situação do processo poissoniano (super-Fano). O quasi-Fano é caracterizado por covariância $\xi \sim 0$, assimetria e curtose com valores próximos da distribuição de Poisson, sendo uma distribuição unimodal, que quanto maior for o valor de N , mais as caudas tornam-se simétricas e com densidades aproximadas às das caudas da distribuição normal. O regime super-Fano tem $\xi > 0$, e gera distribuições de probabilidades com formatos diversos, sendo possíveis valores de assimetria negativa, nula e positiva, e a curtose com valores tanto abaixo, quanto acima do valor

da distribuição de Poisson. Os valores supremos para assimetria e para curtose ocorrem quando $p_\alpha \rightarrow 0$ e $b^e \gg 1$ caracterizando distribuições tipo binomial negativa. Os valores ínfimos são dependentes de $b^e \rightarrow 0$, situação em que há diversidade de formatos para as distribuições, dependendo do valor de $p_\alpha \sim 0,5$.

O GAR tem os mesmos regimes de ruído que o GER, adicionados de dois regimes não existentes no GER: o regime ressonante que ocorre no gene constitutivo (Fano) e o regime de baixos ruídos (sub-Fano). Nos regimes de ruído quasi-Fano e super-Fano, o GAR apresenta as mesmas características qualitativas das distribuições do GER. No regime Fano, onde $\xi = 0$, apresenta a condição ressonante, em que a taxa de síntese de proteínas e a taxa média de chaveamento do estado do promotor do gene são as mesmas. O regime sub-Fano é caracterizado por $\xi < 0$ e os valores de assimetria e curtose são qualitativamente diferentes da situação do GER, pois mesmo para $p_\alpha \rightarrow 1$, o valor de assimetria e de curtose aparece acima do valor para o gene constitutivo.

5 Um modelo dependente do tempo para investigar respostas a terapias gênicas

Avanços recentes em tecnologias de edição de genes trouxeram a promessa de um ponto de inflexão para a terapia gênica (BULAKLAK; GERSBACH, 2020) em direção a desenhos terapêuticos mais complexos com o objetivo de orquestrar a expressão de redes de genes para reprogramação de fenótipos celulares (THAKORE *et al.*, 2016). Uma possibilidade é desenvolver estratégias de tratamento do câncer para reverter a metástase, visando genes reguladores mestres (CAI *et al.*, 2020). Modelos matemáticos que descrevem a regulação da expressão gênica podem ser úteis para a engenharia da dinâmica das redes gênicas que governam o comportamento celular.

Considerando o caso ideal, em que a edição afeta exclusivamente seu alvo epigenético (THAKORE *et al.*, 2016) dentro das células tumorais. A tarefa pode ser formulada como um problema de controle para permitir que o número de transcritos de um gene regulador mestre tenha seu valor médio em um determinado nível e flutuações aleatórias dentro de uma faixa suficientemente pequena. O controle pode ser feito por agentes externos, como uma combinação de drogas que gostaríamos de manter em dosagem minimamente efetivas devido à eventual toxicidade.

Apesar do nosso conhecimento aprofundado da biologia do câncer devido aos avanços nas técnicas de biologia molecular, o uso de métodos quantitativos para integrar a abundância de dados gerados para desenhar tratamentos visando a regressão de metástases ainda está em suas fases iniciais (YESILKANAL *et al.*, 2021; YESILKANAL *et al.*, 2021). A variabilidade genotípica e a aleatoriedade intrínseca das reações bioquímicas (DELBRÜCK, 1940) que regem a epigenética sustentam a multiplicidade de fenótipos de células cancerosas comumente denominados como heterogeneidade tumoral (MARUSYK; ALMENDRO; POLYAK, 2012; MICHOR; BEAL, 2015; ALIZADEH *et al.*, 2015; BROCK; KRAUSE; INGBER, 2015; WELCH, 2016; GUINN *et al.*, 2020).

Além disso, os processos celulares são controlados por várias redes de reações químicas caracterizadas por topologias mutáveis e redundâncias funcionais que equipam as células com capacidades de adaptação e robustez (BHALLA; IYENGAR, 1999; BALÁZSI; OUDENAARDEN; COLLINS, 2011; FERRELL, 2016; GÉRARD; GONZE; GOLDBETER, 2019), tornando desafiante o desenho de tratamentos efetivos. As numerosas escalas de

tempo características dos processos químicos que ocorrem dentro da célula adicionam outra camada de dificuldade (PURVIS; LAHAV, 2013; LEVINE; LIN; ELOWITZ, 2013).

Assim, o aprimoramento das estratégias terapêuticas requer a análise e a engenharia da dinâmica da resposta ao tratamento em um sistema complexo composto por vários componentes em interação com uma multiplicidade de escalas de tempo características e sujeito à aleatoriedade. Encontrar blocos de construção com essas características pode fornecer ideias úteis sobre como modular a dinâmica de um sistema tão complexo (RAMOS *et al.*, 2010). Para tanto, o modelo estocástico exatamente solúvel para a transcrição de um gene binário (PECCOUD; YCART, 1995; IYER-BISWAS; HAYOT; JAYAPRAKASH, 2009; RAMOS *et al.*, 2020; GIOVANINI *et al.*, 2020) é um bom candidato para ser usado como um protótipo para simular estratégias aprimoradas de tratamento.

Propomos uma *prova de conceito* de uma ferramenta quantitativa para investigar a resposta à aplicação de múltiplas drogas para o controle epigenético de genes reguladores mestres. Esses desenhos de tratamento envolvem um grande número de compostos químicos com diversas meias-vidas e afinidades de interação com os componentes do gene. Um gene pode ter sua transcrição guiada por um promotor com múltiplos estados, e a duração desses estados pode ser regulada por múltiplos fatores de transcrição com várias afinidades aos sítios regulatórios do gene.

Para o promotor no estado ON, ainda haveria variação na afinidade entre RNA Polimerase II (RNAPolII) e o promotor. Além disso, é necessário selecionar os agentes externos típicos, aqui as drogas, que afetarão os processos envolvidos na modulação da expressão gênica de acordo com seus parâmetros farmacocinéticos. Portanto, a modelagem matemática do tratamento visando genes específicos requer a seleção de um sistema de modelo de gene adequado para permitir a determinação dos valores dos parâmetros associados à regulação de sua expressão. É preciso acoplar esse modelo à farmacocinética da interação das drogas com seus alvos.

Isso pode ser formulado como um problema de controle no qual pretendemos manter o nível médio de expressão de um gene alvo e a heterogeneidade de sua resposta ao tratamento dentro de intervalos específicos, minimizando as quantidades de drogas usadas para prevenir a toxicidade. Para conceber uma estratégia quantitativa para superar o desafio acima, começamos com o modelo matemático mais simples possível para a expressão gênica regulada sob influência da aleatoriedade para estabelecer uma *prova de conceito*. Este modelo binário estocástico para expressão gênica tem quatro parâmetros

e duas escalas de tempo características. Idealmente, cada uma das múltiplas drogas que compõem um tratamento afetaria um processo específico (ou matematicamente falando, parâmetro) envolvido na expressão do gene alvo.

Isso implica na adição de pelo menos dois parâmetros por droga: a meia-vida da droga e a intensidade de seu efeito em cada parâmetro específico do modelo de expressão gênica. O próximo desafio é orquestrar essa multiplicidade de parâmetros para regular adequadamente a expressão do gene alvo. Aqui, analisamos a dose da droga e a resposta do gene usando uma “otimização do pedestre”, já que a aplicação da teoria de controle ideal é uma técnica altamente elaborada com requisitos específicos e a formulação está além do escopo do estudo atual (JARRETT *et al.*, 2020). Uma vantagem é que várias características qualitativas do modelo estocástico para a expressão do gene são preservadas independentemente dos valores numéricos que são usados.

Nosso próximo desafio é a seleção de um candidato apropriado para o sistema modelo. Aqui, usamos o gene que codifica a proteína inibitória da quinase Raf (RKIP) porque desempenha um papel importante na regulação da dinâmica de vários componentes de uma célula (BONAVIDA; BARITAKI, 2020). Uma vez que a letalidade do câncer é causada principalmente por metástases, a escolha de RKIP é promissora, pois suas concentrações são tipicamente reduzidas em cânceres metastáticos (MARTINHO *et al.*, 2012; MARTINHO *et al.*, 2013; LAMIMAN *et al.*, 2014). Esta correlação negativa transforma as assinaturas do RKIP e os genes relacionados em biomarcadores úteis de risco metastático em pacientes com câncer (YESILKANAL; ROSNER, 2018).

A caracterização do RKIP como um gene antimetastático é reforçada por experimentos que demonstram que sua superexpressão bloqueia a invasão *in vivo* e a progressão metastática (DANGI-GARIMELLA *et al.*, 2009; FU *et al.*, 2003). Portanto, nos concentramos em possíveis estratégias para aumentar as quantidades de RKIP em células cancerosas, remodulando seus perfis de expressão para os regimes pré-câncer. Em alguns cânceres, o número médio de transcritos de RKIP é semelhante ao de células de um tecido normal, mas a variabilidade é significativamente aumentada (ZARAVINOS *et al.*, 2018). Nossa abordagem também é potencialmente útil para esses casos porque é baseada em processos estocásticos.

A seleção do RKIP como sistema modelo nos fornece sua taxa de degradação, ou seja, um dos tempos característicos de nosso modelo estocástico para regulação da expressão gênica. O tempo característico adicional é a taxa de chaveamento dos genes, que

geralmente não está disponível por causa da necessidade de sua inferência. Uma vez que selecionamos o RKIP, também podemos considerar um tratamento multidrogas baseado em 5-AzaC e DETANONOato, pois foram testados anteriormente, seus mecanismos de ação são suficientemente conhecidos e porque poderíamos lançar mão de seus tempos de meia-vida.

Assim, no caso de um desenho de tratamento com duas drogas, teremos três dos quatro tempos de meia-vida para o sistema composto pelo gene e pelo tratamento. Isso nos leva a investigar as características qualitativas de nosso modelo prototípico como uma ferramenta adicional para simular desenhos de tratamento de câncer visando genes específicos ¹.

Para esse objetivo, simulamos o efeito da aplicação de várias drogas, cada uma visando uma taxa cinética diferente de um modelo estocástico de dois estados para a transcrição de genes. Devido ao tratamento, consideraremos as taxas cinéticas do modelo como dependentes do tempo em resposta à aplicação da droga. Isso permite estimar a velocidade e a heterogeneidade da resposta ao tratamento, respectivamente, calculando a dinâmica do número médio de transcrições e sua variância. As soluções exatas do modelo em taxas cinéticas constantes têm duas escalas de tempo características, uma relacionada com o chaveamento do estado do promotor e a segunda sendo o tempo de vida do mRNA.

Recentemente, foi demonstrado que a razão entre essas duas escalas de tempo pode ser usada para classificar os regimes de ruído qualitativamente distinguíveis da transcrição do gene no modelo binário (RAMOS; REINITZ, 2019; GIOVANINI *et al.*, 2020) e a confiabilidade das informações sobre o estado do promotor que é transmitido pelos produtos gênicos (GAMA *et al.*, 2020). Aqui, assumimos que a adição de uma droga faz com que as taxas cinéticas do modelo se tornem dependentes do tempo e que esse efeito decai exponencialmente de tal forma que tem até quatro escalas de tempo adicionais para o problema. Em seguida, mostramos que o tratamento direcionado a um gene com um promotor de chaveamento rápido, e expresso como um processo quasi-poissoniano, permite a resposta mais rápida e menos heterogênea ao tratamento.

Se o gene tiver um promotor de chaveamento lento, a resposta será mais lenta e terá maior heterogeneidade. Um gene expresso de forma intermitente permitirá uma resposta rápida com heterogeneidade máxima. Em seguida, construímos a análise anterior

¹ Observe, no entanto, que ajustes de valor de parâmetro para desenhos experimentais específicos podem ser realizados se julgarmos que nosso modelo pode fornecer resultados suficientemente úteis.

para desenhar um tratamento aprimorado, permitindo uma resposta mais rápida com heterogeneidade reduzida, que é independente das escalas de tempo pré-tratamento. O tratamento aprimorado é baseado em dosagens reduzidas da droga para diminuir as chances de toxicidade e surgimento de resistência causada por efeitos compensatórios (YESILKANAL *et al.*, 2021; YESILKANAL *et al.*, 2021).

5.1 Métodos e Modelos

Modelaremos aqui a transcrição do RKIP, onde a ligação de RNAPolIII ao sítio do promotor pode ser regulada por sua interação com uma região de DNA, chamada de *enhancer*. Os *enhancers* são regiões do DNA que interagem com os fatores de transcrição que podem fornecer uma regulação positiva ou negativa da ligação de RNAPolIII ao sítio promotor.

5.1.1 Uma breve descrição do papel molecular do RKIP

A proteína RKIP é um regulador de quinases que se liga diretamente a quinase Raf (YEUNG *et al.*, 1999; TRAKUL *et al.*, 2005; YESILKANAL; ROSNER, 2014; ZARAVINOS *et al.*, 2018). O RKIP está envolvido na regulação das vias de sinalização, como a cascata Raf–Mek–Erk e as vias relacionadas com NF- κ B (YEUNG *et al.*, 2001; ZHAO; WENZEL, 2014), ambas participando da regulação dos processos anti-apoptose. Além disso, essas vias modulam a proliferação celular com a cascata Raf–Mek–Erk participando da diferenciação e as vias relacionadas ao NF- κ B atuando na inflamação. Na cascata Raf–Mek–Erk, o RKIP inibe a via de sinalização *downstream* pela ligação direta da proteína RKIP desfosforilada ao Raf-1.

Este complexo molecular impede a fosforilação de Raf-1 e a associação Raf-1/Mek, que, por sua vez, causa a interrupção da sinalização de Erk. Esta inibição pode ser revertida pela ação da Proteína Quinase C (PKC) (LORENZ; LOHSE; QUITTERER, 2003; CORBIT *et al.*, 2003; SHVARTSUR *et al.*, 2017), um regulador pós-tradução que fosforila RKIP. Este último se dissocia do Raf-1 por causa de sua mudança estrutural, e a via Erk é ativada. RKIP regula negativamente a via de sinalização NF- κ B indiretamente

(YEUNG *et al.*, 2001; ZARAVINOS *et al.*, 2018) quando interage com o complexo de quinase IKK reduzindo sua atividade.

Isso causa uma redução na fosforilação e degradação das proteínas inibitórias I κ B, que, por sua vez, inativam NF- κ B. Além disso, o RKIP modula outras vias de sinalização celular fundamentais envolvendo proteínas G heterotriméricas, keap1/nrf2, STAT3 e GSK (DATAR *et al.*, 2014). Devido à sua interação com múltiplas vias, consideramos o RKIP como um modulador mestre dos processos celulares.

As quantidades de proteínas RKIP dentro da célula podem ser reguladas em várias etapas de expressão, desde a pré-transcrição do gene até a pós-tradução (GALAL; ZARAVINOS; BONAVIDA, 2021). Uma grande quantidade de tumores sólidos metastáticos tem RKIP regulado em baixos níveis ou perdido, e os dados experimentais sugerem que isso acontece por causa da regulação transcricional ou pós-transcricional (YESILKANAL; ROSNER, 2018). Aqui, limitamos nossa atenção aos casos em que ocorre uma redução no número de proteínas RKIP devido à repressão da transcrição.

A transcrição de RKIP pode ser silenciada por metilação de seu promotor. De fato, a análise de PCR específica para metilação (MSP) mostrou uma correlação suficientemente forte entre a metilação do promotor RKIP e baixos níveis de expressão de RKIP em tumores cancerosos, (YESILKANAL; ROSNER, 2018), incluindo o esofágico e o gástrico (LI *et al.*, 2014; WEI *et al.*, 2017). As modificações das histonas também são encontradas como mecanismos epigenéticos para regular os níveis de RKIP.

Os inibidores da histona desacetilase podem aumentar as transcrições de RKIP (BEACH *et al.*, 2007; LABBOZZETTA *et al.*, 2015). Fatores de transcrição Snail e BACH1 podem diminuir a transcrição de RKIP por histonas metiltransferases (REN *et al.*, 2012; LEE *et al.*, 2014), e ambos estão associados com a transição epitelial-mesenquimal. Snail é um repressor transcricional direto do gene que codifica a proteína de adesão celular E-caderina (BARITAKI *et al.*, 2010), e BACH1 é a proteína zipper de leucina básica expressa em tecido de mamífero, que regula positivamente genes relacionados à motilidade que promovem metástases em cânceres de mama (YUN *et al.*, 2011). A expressão dos genes Snail e BACH1 é autorreprimida e reprimida pelas proteínas RKIP. Além disso, BACH1 e RKIP se combinam em um circuito gênico biestável que descreve uma mudança para o fenótipo metastático na população de células tumorais, como mostrado em Lee *et al.* (2014).

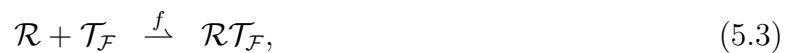
Os mRNAs de RKIP também podem ser negativamente regulados na pós-transcrição devido à interação com miRNAs específicos. De fato, em vários cânceres, o gene RKIP é silenciado pela ação de miRNAs (-224, -27a, -23a e -543) direcionados a mRNAs de RKIP (DU *et al.*, 2017b; CASTRO *et al.*, 2020). Isso sugere uma alternativa terapêutica baseada no direcionamento de miRNAs que regulam negativamente a expressão de RKIP, como lncRNA XIST, que estabiliza a expressão de RKIP suprimindo miR-23a (DU *et al.*, 2017a).

Os dados acima mencionados sugerem uma associação entre os níveis de expressão de RKIP e fenótipos de células tumorais. Isso leva à possibilidade de usar RKIP como um marcador prognóstico para probabilidades de sobrevivência (FU *et al.*, 2006; AL-MULLA *et al.*, 2011). Além disso, como os níveis de RKIP são reduzidos principalmente em vários cânceres (ZARAVINOS *et al.*, 2018), compreender seus mecanismos regulatórios (BEACH *et al.*, 2007; BARITAKI *et al.*, 2010; YUN *et al.*, 2011; REN *et al.*, 2012; LEE *et al.*, 2014; LI *et al.*, 2014; LABBOZZETTA *et al.*, 2015; WEI *et al.*, 2017) pode permitir o desenho de tratamentos antimetastáticos.

5.1.2 Um modelo efetivo para a regulação da expressão gênica

Interpretamos o gene como uma fonte de produtos gênicos alternando aleatoriamente entre os estados ON e OFF. A síntese de produtos gênicos ocorre quando o gene está ON à taxa k , enquanto, no estado OFF, não há síntese. A taxa de degradação dos produtos gênicos é denotada por ρ . O gene muda do estado ON para OFF e de OFF para ON, com taxas h e f , respectivamente.

Os processos mencionados acima podem ser representados como um sistema de reações químicas efetivas conforme apresentado nas Eqs. (5.1)-(5.4). Denotamos um produto gênico por \mathcal{P} e seu sítio regulatório por \mathcal{R} . Consideramos aqui, o caso particular de regulação positiva do gene por um fator de transcrição denotado por $\mathcal{T}_{\mathcal{F}}$.



As Eqs. (5.1) e (5.2) indicam, respectivamente, a síntese e degradação do produto gênico. A mudança do estado OFF para ON por causa da ligação da proteína ativadora, e a transição inversa causada por sua desvinculação, são, respectivamente, indicadas em Eqs. (5.3) e (5.4). Pode-se também considerar o caso de um fator de transcrição como repressor, e neste caso, as reações efetivas nas Eqs. (5.3) e (5.4) denotariam as transições de estado ON para OFF e OFF para ON com taxas as transformadas de $f \rightarrow h$ e $h \rightarrow f$.

Este sistema de reações efetivas é muito simples se considerarmos a complexidade da regulação da expressão gênica em mamíferos. No entanto, tal simplificação permite a construção de modelos quantitativos exatamente solúveis com um número menor de parâmetros que podemos usar para investigar estratégias de tratamento hipotéticas antes de realizar experimentos. Um exemplo é o uso de tratamento multidrogas com doses reduzidas visando uma rede funcional. Uma determinada droga tem uma meia-vida específica, e os várias drogas que atuam em um sistema se combinam como processos em uma escala de tempo múltipla.

Combinar as agendas de doses e aplicação dessas várias drogas para garantir efetividade e não toxicidade pode ser uma tarefa difícil, e um modelo quantitativo pode fornecer *insights* inestimáveis sobre o desenho de terapias. Aqui, usamos o modelo binário estocástico para a regulação da expressão gênica para investigar como a combinação de duas drogas modula a dinâmica da produção de mRNA.

5.1.3 Um modelo efetivo para investigar a regulação da dinâmica da expressão gênica após o tratamento

Uma interpretação biológica do modelo efetivo apresentado nas Eqs. (5.1)-(5.4) é fornecida a seguir. A taxa k é proporcional ao inverso do intervalo de tempo entre duas ligações consecutivas de RNAPolIII ao sítio promotor. Isso implica assumir que a transcrição começa após um intervalo insignificante depois da ligação de RNAPolIII. Alternativamente, pode ser interpretado como o inverso do intervalo médio entre o início de dois processos de transcrição subsequentes.

O primeiro caso implica assumir a disponibilidade de grandes quantidades de RNAPolIII e interpretar k como consequência da afinidade entre o promotor e RNAPolIII. O segundo pode ser interpretado como a eficiência do promotor no início da transcrição. Essas

duas interpretações não são exclusivas e, para simplificar, nos referimos a um aumento em k como um aumento na eficiência do promotor.

As taxas de chaveamento h e f são o inverso do tempo médio de disponibilidade e indisponibilidade do promotor para a ligação do RNAPolIII, respectivamente. O valor dessas taxas será determinado pela ligação de fatores de transcrição às regiões regulatórias do gene. Apesar da quantidade de fatores de transcrição que mudam com o tempo, é correto supor que eles permanecerão constantes durante intervalos suficientemente curtos. Portanto, durante esses curtos intervalos de tempo, podemos empregar o modelo efetivo de Eqs. (5.1)-(5.4). Os valores das constantes de chaveamento refletirão o equilíbrio resultante da ligação de ativadores, repressores, inibidores, fatores pioneiros e outros elementos reguladores que interagem com as regiões regulatórias do gene.

Nosso modelo para regulação da expressão gênica sugere o desenho de múltiplas estratégias de tratamento com o objetivo de aumentar a expressão do gene RKIP. O modelo fornece quatro taxas cinéticas que podem ser visadas por drogas. No caso específico do gene RKIP, pode-se tentar aumentar k e f ou diminuir ρ e h . Essas taxas podem ser afetadas individual ou coletivamente, e de forma coordenada, no caso de uma poliquimioterapia. No esquema mostrado na Fig. 11, consideramos os casos particulares em que o tratamento visa: (1) aumentar f ; (2) aumentar k ; e (3) aumentar f e k concomitantemente. Assumimos que um determinado tratamento visa uma taxa de maneira exclusiva — ou seja, as taxas não direcionadas permanecerão constantes durante o tratamento. Dois mecanismos podem ser usados para aumentar os níveis de expressão de RKIP:

(1) **Desmetilação do promotor.** A regulação negativa de RKIP foi associada à metilação do promotor em muitos tipos de câncer (GUO *et al.*, 2012; KIM *et al.*, 2017; LI *et al.*, 2014; MINOO *et al.*, 2006; WEI *et al.*, 2017). Supomos o uso de um agente de desmetilação, a saber, 5-Azacitidina (5-AzaC), para reverter isso, como pode ser visto na Fig. 2A em Ramos *et al.* (2020). Conseqüentemente, os níveis de expressão de RKIP seriam aumentados conforme testado anteriormente na linhagem de células de câncer de mama triplo-negativo (TNBC) SUM159 e linhas de células de esôfago TE-1 e TE-13 (GUO *et al.*, 2012; LABBOZZETTA *et al.*, 2015).

(2) **Regulação dos fatores de transcrição.** O ambiente tumoral é uma fonte inesgotável de sinais que induzem uma mudança nas quantidades dos fatores de transcrição que regulam a expressão gênica e, conseqüentemente, os fenótipos das células tumorais. Um possível tratamento é usar doadores de óxido nítrico (NO) ou NO atuando como agen-

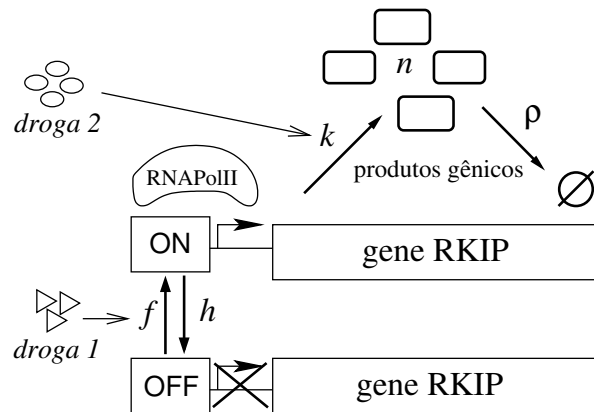
tes anticâncer diretos ou quimiossensibilizadores (BONAVIDA, 2021). Por exemplo, NO (FUKUMURA; KASHIWAGI; JAIN, 2006) afeta o crescimento do tumor, regulando negativamente as quantidades funcionais de NF- κ B e SNAIL, que, em quantidades reduzidas, estão associadas a um aumento na expressão de RKIP.

Assim, pode-se propor uma estratégia terapêutica usando doadores de NO, como (Z)-1-[2-(2-aminoetil)-N-(2-amônio-etil)amino]diazén-1-ium-1,2-diolato (DETANONOato) (HUERTA-YEPEZ *et al.*, 2004; BARITAKI *et al.*, 2010; BONAVIDA, 2014; ZARAVINOS *et al.*, 2018), para manter uma inibição constante do *loop* NF- κ B/SNAIL. De fato, RKIP foi aumentado em tratamentos com DETANONOato desenhado para inibir a transição epitelial-mesenquimal (EMT) e invasão em linhagens de células de carcinoma de próstata humano metastático, o que foi corroborado em camundongos com xenoinxertos tumorais (BARITAKI *et al.*, 2010).

Doadores de NO foram usados em combinação com fotossensibilizadores para aumentar a eficácia da terapia fotodinâmica na inibição da proliferação de células de melanoma murino (RAPOZZI *et al.*, 2013). Note, entretanto, que o NO desempenha um papel duplo (BONAVIDA; BARITAKI, 2011) já que, em baixas doses, ele promove a carcinogênese (PERVIN *et al.*, 2007; PERVIN *et al.*, 2013). De fato, baixos níveis de terapia fotodinâmica induzem baixos níveis de NO, o que contribui para uma resposta antiapoptótica pelo *loop* NF- κ B/Snail/RKIP (RAPOZZI; PIETRA; BONAVIDA, 2015). Aqui, assumimos que as concentrações de NO nas células cancerosas são suficientemente grandes, em níveis micromolares (WINK *et al.*, 2008), para garantir seu papel na promoção de um aumento na quantidade de transcritos de RKIP, um esquema é mostrado na Fig. 2B em Ramos *et al.* (2020).

A escolha do 5-AzaC e do DETANONOato foi motivada por uma caracterização suficientemente boa de seus mecanismos de ação nas vias de sinalização que participam da expressão de RKIP regulado positivamente. No entanto, desenhos de tratamento alternativos podem ser usados para promover a expressão de RKIP, permitindo uma associação sinérgica com terapias contra o câncer, como Sorafenib associado a Gemcitabina (PASQUALETTI *et al.*, 2011) ou Erlotinib (GIOVANNETTI *et al.*, 2012) em cânceres de pulmão e Gemcitabina com Sorafenib em câncer de pâncreas (RICCIARDI *et al.*, 2010). O inibidor da topoisomerase I 9NC (CHATTERJEE *et al.*, 2004) e os agentes antimetabólicos ENMD-1198 e MKC-1 também demonstraram os efeitos de regulação positiva de RKIP (YOUSUF *et al.*, 2014).

Figura 11 – Esquema das drogas visando taxas cinéticas na transcrição de RKIP.



A RNAPolIII liga-se à região promotora (quando está ON) do gene RKIP para sintetizar mRNA (produtos gênicos). O chaveamento entre os estados ON e OFF depende dos componentes regulatórios (proteínas) que cercam o gene. *Droga 1* tem como alvo uma proteína ativadora e visa aumentar o tempo de exposição do promotor para ligação de RNAPolIII. *Droga 2* afeta apenas o promotor, aumentando sua eficiência ou a afinidade do promotor para o RNAPolIII. Um tratamento consiste em administrar uma única droga ou uma combinação de drogas seguindo uma agenda de doses.

Fonte – Giovanini *et al.* (2021)

No entanto, encontrar drogas que visam especificamente a desmetilação do promotor do gene RKIP ou fatores de transcrição que regulam os níveis de expressão de RKIP é um desafio que está além do escopo da nossa proposta. Portanto, como uma estratégia para apresentar nossa metodologia, para definir escalas de tempo clinicamente relevantes e para mostrar as dificuldades de combinar várias drogas com vários alvos e escalas de tempo, consideramos drogas não específicos. Nesse caso, estamos considerando apenas o efeito da droga no gene que descrevemos. Os efeitos do nível celular não serão considerados aqui, pois exigiriam uma abordagem mais complexa, em que a dinâmica dos níveis de expressão de vários genes potencialmente interagentes precisaria ser considerada.

5.1.4 Uma abordagem para investigar os efeitos do tratamento na dinâmica de expressão de RKIP

Nossas estratégias de tratamento visam aumentar f e k . Assumimos que a efetividade da droga nessas quantidades decai exponencialmente e que as taxas cinéticas retornarão aos valores anteriores ao tratamento. Portanto, uma vez que o tratamento começa, as taxas f e k tornam-se dependentes do tempo com f_0 e k_0 sendo de OFF para ON e as taxas de síntese, respectivamente, antes do tratamento. Denotamos o efeito do fracionamento de uma dose da droga sobre o valor de uma constante cinética por ξ , onde $0 < \xi \leq 1$.

Assumimos que, na dose máxima tolerada, o efeito da droga é dado por $\xi = 1$ e que essa dose aumenta as taxas desejadas para f_1 e k_1 .

Isso não implica a suposição de eficiência máxima da efetividade da droga. Na verdade, consideramos que ξ é o efeito líquido de uma determinada dose em sua taxa alvo. Portanto, uma dose menor que a máxima tolerada mudará instantaneamente sua taxa alvo para $\xi_a f_1$ e $\xi_b k_1$, onde ξ_a e ξ_b são funções não lineares da dose necessária a ser formulada de acordo com os dados experimentais. A taxa de chaveamento de OFF para ON dependente do tempo, $f(t)$, é

$$f(t) = \begin{cases} f_0, & 0 \leq t < \tau_1 \\ f_0 + [f_s(\tau_j) - f_0] e^{-\lambda_a(t-\tau_j)}, & \tau_j \leq t < \tau_{j+1} \end{cases} \quad (5.5)$$

e $k(t)$

$$k(t) = \begin{cases} k_0, & 0 \leq t < \tau_1 \\ k_0 + [k_s(\tau_j) - k_0] e^{-\lambda_b(t-\tau_j)}, & \tau_j \leq t < \tau_{j+1} \end{cases} \quad (5.6)$$

onde $j = 1, \dots, J - 1$ denota a j -ésima aplicação da droga, J é a quantidade de doses da droga e τ_j é o tempo da j -ésima aplicação da droga. A cada aplicação da droga, a condição de estado estacionário é aquela referente ao sistema não tratado, a saber f_0 e k_0 , e (λ_a, λ_b) denotam as taxas de decaimento exponencial do efeito das drogas nas taxas (f, k) .

Conforme mencionado anteriormente, estamos considerando que o efeito das drogas em suas taxas alvo é rápido o suficiente para ser considerado instantâneo. Portanto, os valores das taxas f e k imediatamente após a chegada da j -ésima dose, respectivamente denotados por $f_s(\tau_j)$ e $k_s(\tau_j)$, serão considerados como as condições iniciais durante o intervalo $\tau_j \leq t < \tau_{j+1}$. Observe que a j -ésima dose se soma às quantidades de drogas remanescentes de aplicações anteriores. Portanto, a condição inicial no momento τ_j é escrita como:

$$f_s(\tau_j) = f_1 \sum_{i=1}^j \xi_a(\tau_i) e^{-\lambda_a(\tau_j - \tau_i)}, \quad (5.7)$$

$$k_s(\tau_j) = k_1 \sum_{i=1}^j \xi_b(\tau_i) e^{-\lambda_b(\tau_j - \tau_i)}, \quad (5.8)$$

onde a dose pode ser calibrada para gerar um efeito diferencial a cada instante e evitar o acúmulo tóxico de quantidades de drogas.

5.1.5 Uma descrição aproximada da dinâmica da expressão gênica binária estocástica com taxas cinéticas dependentes do tempo

A aleatoriedade dos fenômenos intracelulares sugere uma descrição dos efeitos do tratamento na expressão do gene RKIP a ser construída em termos de um processo estocástico. A Fig. 11 sugere a existência de duas variáveis aleatórias que determinam o estado do sistema: o estado do gene (sendo ON ou OFF) e o número de produtos gênicos, denotados por n . A descrição da dinâmica do estado do sistema é dada em termos de uma distribuição de probabilidade

$$\Pi(\alpha_n(t), \beta_n(t)), \quad (5.9)$$

onde $\alpha_n(t)$, ou $\beta_n(t)$, denotam a probabilidade de encontrar n proteínas no tempo t quando o gene está ON ou OFF, respectivamente. Uma equação mestra que rege a distribuição de probabilidade para um gene regulado externamente pode ser escrita como:

$$\frac{d\alpha_n}{dt} = k(t)(\alpha_{n-1} - \alpha_n) + \rho[(n+1)\alpha_{n+1} - n\alpha_n] - h\alpha_n + f(t)\beta_n, \quad (5.10)$$

$$\frac{d\beta_n}{dt} = \rho[(n+1)\beta_{n+1} - n\beta_n] + h\alpha_n - f(t)\beta_n, \quad (5.11)$$

onde h e ρ são constantes e Eqs. (5.5) e (5.6) dão $f(t)$ e $k(t)$. A existência de coeficientes dependentes do tempo torna difícil solucionar analiticamente as Eqs. (5.10) e (5.11), e alguns métodos numéricos precisam ser empregados. A interpretação da equação mestra é construída em termos dos coeficientes $k(t)$, ρ , $f(t)$ e h conforme apresentado na descrição das Eqs. (5.1)-(5.4) e o esquema mostrado na Fig. 11.

Aqui, propomos aproximar a dinâmica das taxas cinéticas $f(t)$ e $k(t)$ das Eqs. (5.5) e (5.6) como funções por partes que assumem valores constantes durante intervalos de tempo suficientemente pequenos, garantindo que a diferença entre os valores exatos e aproximados esteja dentro de um determinado erro, onde os detalhes do cálculo está descrito no Apêndice F. Então, podemos considerar o modelo para taxas cinéticas constantes durante um intervalo, o que é exatamente solucionável. Nesse caso, a condição inicial de um intervalo é a condição final em seu vizinho anterior.

5.1.6 Um modelo exatamente solúvel para avaliação comparativa do tratamento do câncer com o objetivo de modular os níveis de expressão gênica

As equações mestras, Eqs. (5.10) e (5.11), com coeficientes constantes já foram propostas, e são totalmente solucionáveis nos regimes estacionários (PECCOUD; YCART, 1995) e dependentes do tempo (IYER-BISWAS; HAYOT; JAYAPRAKASH, 2009). A existência de soluções exatas permite calcular as funções dependentes do tempo que regem o primeiro e o segundo momento do número de produtos gênicos (RAMOS *et al.*, 2020). Aqui, escrevemos as expressões explícitas que governam a dinâmica do número médio de produtos gênicos, $\langle n \rangle(t)$ e o desvio padrão, $\sigma(t) = \sqrt{\langle n^2 \rangle(t) - \langle n \rangle^2(t)}$. Usamos Eqs. (5.1)-(5.4) e definimos as seguintes constantes:

$$N = \frac{k}{\rho}; \quad A_s = \frac{f}{f+h}; \quad \epsilon = \frac{f+h}{\rho}, \quad (5.12)$$

que são, respectivamente, o número esperado de produtos gênicos em estado estacionário no caso de um gene estar totalmente ON (chamamos isso de *número máximo de mRNA* N); a probabilidade de estado estacionário para o gene estar ON (A_s); e a proporção da taxa de chaveamento de genes entre os estados ON e OFF e a taxa de degradação dos produtos gênicos (chamamos isso de *velocidade de chaveamento* ϵ).

O número médio de mRNAs e o desvio padrão no regime de estado estacionário são, respectivamente, denotados por $\langle n \rangle_s$ e σ_s . Nós os escrevemos como funções dos parâmetros da Eq. (5.12):

$$\langle n \rangle_s = A_s N, \quad (5.13)$$

$$\sigma_s^2 = \langle n \rangle_s \left(1 + N \frac{1 - A_s}{1 + \epsilon} \right). \quad (5.14)$$

A dinâmica da média e do desvio padrão é denotada por $\langle n \rangle(t)$ e $\sigma(t)$, respectivamente, e escrevemos:

$$\langle n \rangle(t) = \langle n \rangle_s + Y e^{-\epsilon \rho t} + V e^{-\rho t}, \quad (5.15)$$

$$\sigma^2(t) = \sigma_s^2 + U_1 e^{-\epsilon \rho t} + V e^{-\rho t} + W_1 e^{-(1+\epsilon)\rho t} + X_1 e^{-2\rho t} - Y^2 e^{-2\epsilon \rho t}. \quad (5.16)$$

Os coeficientes das exponenciais são constantes de integração fornecidas no Apêndice I. Essas soluções são obtidas a partir de um sistema de equações diferenciais ordinárias acoplando os momentos A , $\langle n_\alpha \rangle$ e $\langle n^2 \rangle$, onde as formas exatas são fornecidas no Apêndice J.

As Eqs. (5.15) e (5.16) nos permitem calcular a evolução do número médio de produtos do gene RKIP e seu desvio padrão usando uma representação por partes de $f(t)$ e $k(t)$, como está mostrado na próxima seção.

A taxa de decaimento para o estado estacionário de $\langle n \rangle(t)$, Eq. (5.15), e $\sigma(t)$, Eq. (5.16), pode ser estabelecida em termos de ϵ . Para $\epsilon \gg 1$, os termos $e^{-\epsilon\rho t}$, $e^{-(\epsilon+1)\rho t}$ tornam-se nulos muito mais rápido do que o termo $e^{-\rho t}$, que determina a aproximação do estado estacionário. Alternativamente, para $\epsilon \ll 1$, o termo $e^{-\rho t}$ governará o tempo de vida do regime dinâmico, ou seja, o regime durante o qual $\langle n \rangle(t)$ e $\sigma(t)$ estão variando com o tempo.

Além disso, pode-se notar que ϵ também reflete a razão entre a frequência de chaveamento do gene e a taxa de degradação, os tempos característicos dos dois processos sendo acoplados. A Eq. (5.14) indica que $\sigma_s^2 \rightarrow \langle n \rangle_s$ quando $\epsilon \gg N > 1$. Isso coincide com o decaimento para o estado estacionário sendo determinado pela taxa de degradação do produto gênico, como ocorre em um processo de Poisson. Por outro lado, quando a taxa de decaimento do chaveamento do gene para o estado estacionário é predominante, para $\epsilon \ll 1$, σ_s^2 terá valores maiores. Então, tal chaveamento terá um efeito mais forte nas flutuações do número de produtos gênicos.

5.1.7 Valores de parâmetros e condições para simulações de tratamento

O modelo exatamente solúvel para a expressão do gene estocástico binário permite a simulação da dinâmica da resposta ao tratamento, medida em termos de expressão média e flutuações aleatórias em torno da média de um gene hipotético (abstrato). No entanto, tal escolha pode dificultar o fornecimento de *insights* para alguém que deseja desenhar um tratamento para modular a expressão de um gene regulador mestre. Um dos objetivos é o desenho de uma agenda ótima de aplicação de tratamento para controlar os desvios médio e padrão dos produtos gênicos para permanecer dentro de faixas específicas.

Os instantes de aplicação de cada droga, denotados por τ , estão relevantemente relacionados aos tempos característicos do sistema composto por um gene e suas drogas em interação. Entre esses tempos característicos, as meias-vidas das drogas, denotadas por λ_i , onde i rotula uma dada droga, e a taxa de degradação dos produtos do gene, denotada por ρ , são amplamente utilizadas. No entanto, conforme ilustrado na subseção anterior,

a introdução do chaveamento do promotor fornece um tempo característico adicional, denotado por $\epsilon\rho$, que terá efeitos muito importantes na resposta ao tratamento, conforme está mostrado a seguir. Portanto, a escolha do sistema RKIP, 5-AzaC e DETANONOato nos ajuda a definir alguns valores de parâmetro biologicamente razoáveis para realizar nossa análise. Adaptar nossa estrutura para sistemas diferentes, no entanto, é uma tarefa totalmente viável.

Aqui, as taxas relacionadas ao DETANONOato e 5-AzaC são, respectivamente, denotadas por um índice a e b . Portanto, as taxas de decaimento de DETANONOato (atuando em f) e 5-AzaC (atuando em k) são denotadas por λ_a e λ_b , respectivamente. Seus valores são $\lambda_a = 0,05 \text{ h}^{-1}$ (HAMPL *et al.*, 1996) e $\lambda_b = 0,25 \text{ h}^{-1}$ (entrada DB00928 no DrugBank (WISHART *et al.*, 2017)).

Um desafio à parte no desenho de uma terapia visando um gene específico é a formulação de um modelo efetivo para o efeito do tratamento em um parâmetro do gene em questão. Como isso está fora do escopo do presente estudo, definimos uma quantidade denotada por ξ , que indicará o efeito de uma dose de tratamento em seu parâmetro alvo. Para a dose máxima tolerada, o efeito será considerado máximo e indicado por $\xi = 1$. Quando temos $0 < \xi < 1$, estamos indicando que a dose é menor que a máxima. Observe, entretanto, que não temos um modelo matemático relacionando ξ à dose da droga, e sua construção requer experimentos específicos.

Denotamos, a resposta gerada nas taxas cinéticas pelo tratamento na dose máxima tolerada, por $\xi_a = \xi_b = 1$. Para uma única dose, assumimos que as drogas alteram os valores de suas taxas alvo instantaneamente após a aplicação, e para a dose máxima tolerada, $f \rightarrow f_1$ e $k \rightarrow k_1$, conforme indicado pelas Eqs. (5.7) e (5.8), respectivamente.

Os valores específicos das doses de ambas as drogas aqui consideradas foram definidos em estudos anteriores. Os valores da dose e os efeitos nos níveis de expressão de RKIP podem variar em até uma ordem de magnitude. Por exemplo, a análise experimental de células metastáticas da próstata humana (DU145 e PC-3) tratadas com $1000\mu\text{M}$ DETANONOato mostrou a regulação positiva do mRNA de RKIP por 4 e 12h pós-tratamento (BARITAKI *et al.*, 2010). Células de câncer de mama triplo-negativas (SUM159) também aumentam em RKIP mRNA 1,4 vezes quando tratadas com $500\text{-}2000\mu\text{M}$ 5-AzaC por 72h (LABBOZZETTA *et al.*, 2015) e célula de câncer de esôfago humano (células TE-1 e TE-13) tratadas com $2\mu\text{g}/\text{mL}$ 5-AzaC mostraram um aumento de 1,5-10 vezes na expressão de RKIP (GUO *et al.*, 2012).

O tratamento induz uma mudança nos parâmetros cinéticos do modelo. Como consequência, todas as quantidades que dependem desses parâmetros serão alteradas. Portanto, $f_0 \rightarrow f_1$ e $k_0 \rightarrow k_1$ causa uma mudança de $\epsilon_0 \rightarrow \epsilon_1$ e uma mudança nos valores de estado estacionário das quantidades estatísticas $\langle n \rangle_{s,0} \rightarrow \langle n \rangle_{s,1}$, e $\sigma_{s,0} \rightarrow \sigma_{s,1}$ de acordo com as Eqs. (5.12)-(5.14). As distribuições de probabilidade de estado estacionário esperadas a governar o número de transcritos de RNA instantaneamente após a aplicação do tratamento estão mostradas na Fig. 12. No entanto, como os efeitos da droga decaem exponencialmente com uma taxa determinada por sua meia-vida, o sistema não atinge esses novos valores de estado estacionário e tende a retornar às condições anteriores ao tratamento.

Aqui, assumimos que o produto gênico de RKIP são mRNAs, cuja taxa de degradação é $\rho = 0,17 \text{ h}^{-1}$ (WALKER *et al.*, 2011). A condição de pré-tratamento, ocorrendo para $0 \leq t < \tau_1$, não é mostrada porque a assumimos como um estado estacionário e, portanto, definimos $\tau_1 = 0$. A condição de pré-tratamento é caracterizada por um baixo número médio de cópias de mRNAs de RKIP (aqui, escolhido como $\langle n \rangle_0 = 10$). Assumimos arbitrariamente que, em condições de pré-tratamento, os níveis de mRNAs estão oito vezes abaixo do que seria esperado ser encontrado em uma célula não metastática. Consequentemente, o tratamento é considerado bem-sucedido se conduzir a probabilidade de encontrar menos de $\langle n \rangle_T \approx 80$ mRNAs para um valor insignificante. Este é um requisito importante para minimizar a heterogeneidade em uma resposta ao tratamento.

A aleatoriedade dos processos intracelulares causa uma variabilidade na resposta ao tratamento, mesmo em condições hipotéticas em que todos os indivíduos de uma população de células geneticamente idênticas absorvem a mesma dosagem da droga. A heterogeneidade da resposta pode ser quantificada pelo desvio padrão do número de produtos gênicos. Portanto, primeiro definimos $\langle n \rangle_1 = 100$ como o nível de expressão pretendido pelo tratamento. Este valor é escolhido assumindo que um gene que expressa uma média de 100 mRNAs em um regime de Poisson tem o valor limiar $80 = \langle n \rangle_1 - 2\sigma_1$. Aqui, $\sigma_1 = \sqrt{\langle n \rangle_1}$ é o desvio padrão de uma distribuição de Poisson com $\langle n \rangle_1$ como sua média.

As dinâmicas de f e k são descritas pelas Eqs. (5.5) e (5.6). Nós aproximamos esses parâmetros como funções por partes e calculamos o erro usando integrais dos decaimentos exponenciais ao longo de cada subintervalo. Em seguida, definimos o comprimento de

cada subintervalo corrigindo seu erro. A aproximação por partes permite o uso de $\langle n \rangle(t)$, Eq. (5.15), e $\sigma(t)$, Eq. (5.16), para descrever a expressão de RKIP.

Uma única dose não será suficiente para manter $\langle n \rangle_1 \approx 100$ por causa do decaimento exponencial do efeito da droga nas constantes cinéticas (como será mostrado nos gráficos A-E das Figs. 13-15). Portanto, múltiplas doses são necessárias, e determinamos os intervalos entre as aplicações de DETANONOato e 5-AzaC como, respectivamente, 10h e 4h. Esses números foram escolhidos para garantir $\langle n \rangle(t) \approx 100$ durante um intervalo de tempo suficientemente longo, e sua escolha é baseada nas taxas de decaimento de cada droga.

O tratamento altera as propriedades dinâmicas do chaveamento de genes. Quando consideramos o *DETANONOato*, a probabilidade de pré-tratamento de encontrar o gene ON é $A_0 = \frac{f_0}{f_0+h}$. A dose de droga administrada em τ_j causa $f_0 \rightarrow f_s(\tau_j)$ instantaneamente conforme definido na Eq. (5.7). Então, a probabilidade no estado estacionário almejada para o gene estar ON é $A_1 = \frac{f_s(\tau_j)}{f_s(\tau_j)+h}$. Os valores pré e instantaneamente pós-tratamento da frequência de chaveamento gênico são, respectivamente, $\epsilon_0 = \frac{f_0+h}{\rho}$ e $\epsilon_1 = \frac{f_s(\tau_j)+h}{\rho}$. Quando consideramos a *5-AzaC* a probabilidade de encontrar o gene ON, dada por $A_0 = \frac{f_0}{f_0+h}$, e frequência de chaveamento de gene $\epsilon_0 = \frac{f_0+h}{\rho}$, permanecerão ambos constantes, enquanto o valor de $k_0 \rightarrow k_s(\tau_j)$ instantaneamente após a aplicação da droga em τ_j — veja a Eq. (5.8) — de modo que $N_0 = \frac{k}{\rho} \rightarrow N_1 = \frac{k_s(\tau_j)}{\rho}$.

Como estamos usando um modelo estocástico exatamente solúvel, podemos mapear suas características qualitativas em termos das relações entre seus parâmetros cinéticos. Por exemplo, o valor de ϵ na Eq. (5.12) é um parâmetro chave que determina a forma das distribuições de probabilidade de estado estacionário mostradas na Fig. 12. Portanto, a interpretação de nossos resultados permanecerá útil na análise de desenhos de experimentos específicos, apesar da necessidade de valores específicos para os parâmetros do modelo.

Portanto, nossa escolha para os valores médios de mRNAs pré e pós-tratamento são arbitrários e selecionados para a clareza da apresentação de nossos resultados e previsões. Um aumento de quatro a oito vezes na mediana dos números de mRNA foi relatado em PCPG (feocromocitoma e paraganglioma), CHOL (colangiocarcinoma) e SARC (sarcoma), conforme mostrado na Fig. 1 em [Zaravinos et al. \(2018\)](#).

Os níveis de expressão de RKIP em LIHC (carcinoma hepatocelular do fígado) ([WU et al., 2016](#)), GBMLGG (glioma) ([LEI et al., 2015](#)) e STES (carcinoma de estômago e esôfago) ([ZHAO et al., 2013](#)) aumentaram em até duas vezes.

Outras comparações de níveis de expressão de RKIP em câncer de mama (BRCA e TNBC) (HAO *et al.*, 2012; ZOU *et al.*, 2016), câncer de pele (SKCM) (LIN *et al.*, 2010; SCHUIERER *et al.*, 2004) e cânceres colorretais (COADREAD) (MINOO *et al.*, 2007; AL-MULLA *et al.*, 2011; LEE *et al.*, 2016) indicou uma mudança insignificante na mediana. Em todos os cânceres mencionados, a variabilidade no número de transcritos de RKIP foi maior nas células tumorais do que nas células normais. Uma vantagem da utilização de um modelo estocástico é a possibilidade de levantar possíveis explicações para tais características.

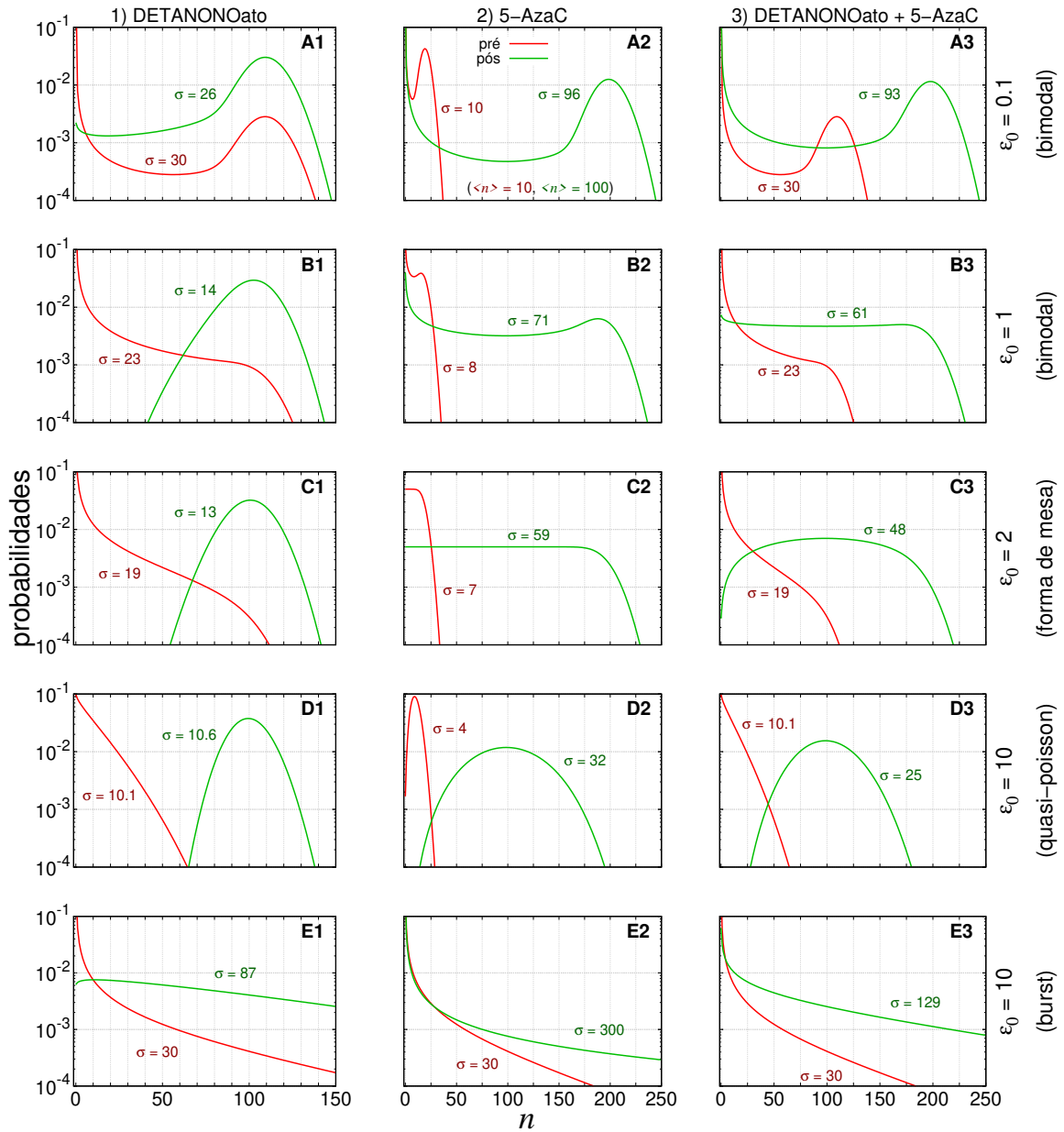
5.2 Resultados

Simulamos três cenários de tratamento: (1) denotando o efeito da dose máxima tolerada de DETANONOato; (2) denotando o efeito da dose máxima tolerada de 5-AzaC; e (3) denotando o efeito de doses fracionadas das duas drogas juntas. A resposta ao tratamento é quantificada em termos do tempo para o valor médio atingir o limiar e o desvio padrão. Essas duas quantidades são relevantemente dependentes dos valores de ϵ_0 , como mostraremos nas próximas subseções. Os valores de ρ , λ_a e λ_b são considerados constantes durante o tratamento. Todas as taxas (k , f e h) são fornecidas em h^{-1} .

Os valores dos parâmetros foram selecionados para simular condições de tratamento cujas distribuições de probabilidade governando a expressão de células pré-tratadas indicam regimes de estado estacionário qualitativamente distinguíveis como recentemente classificado em Giovanini *et al.* (2020). A Fig. 12 mostra as distribuições de probabilidade de estado estacionário pré-tratamento (e pós-tratamento pretendido) que governam os números de mRNAs de RKIP em vermelho (e verde).

Os valores dos parâmetros em cada linha são definidos para garantir que os gráficos A indiquem distribuições bimodais, os gráficos B sejam distribuições próximas ao limite do regime bimodal, os gráficos C indicam o regime em que as probabilidades são em forma de mesa para $A_0 = A_1 = 0,5$ (conforme indicado no gráfico C2), os gráficos D indicam as distribuições quasi-poissonianas e os gráficos E denotam o limite de *burst*. As curvas das distribuições de probabilidade do estado estacionário de encontrar n mRNAs, dentro da Fig. 12, são calculadas pela Eq. (K.1) e são denotadas por $\tilde{\phi}_n$ no Apêndice K.

Figura 12 – Distribuições de probabilidade das condições pré e pós-tratamentos.



As linhas vermelhas indicam probabilidades de pré-tratamento, enquanto as linhas verdes indicam as probabilidades pretendidas que regem o número de produtos gênicos pós-tratamento. Os parâmetros das distribuição pós-tratamento foram definidos com o objetivo de aumentar o número médio de transcrições para ≈ 100 . Desconsideramos o efeito dos novos valores dos parâmetros nas variâncias, ou seja, a heterogeneidade da resposta ao tratamento. Consideramos cinco condições iniciais para a velocidade de chaveamento do gene, $\epsilon_0 = (0,1; 1; 2; 10; 10)$, arranjados em linhas e indicados por (A1-A3), (B1-B3), (C1-C3), (D1-D3) e (E1-E3) nas etiquetas dos respectivos gráficos. Os gráficos D, E têm o mesmo valor para ϵ_0 , mas representam diferentes regimes por causa dos valores diferentes de A_0 . As distribuições para os desenhos de tratamento são organizadas em colunas: (1) DETANONOato visando a taxa de f , (2) 5-AzaC visando a taxa de k e (3) ambas as drogas visando as taxas f e k , simultaneamente.

Fonte – Giovanini *et al.* (2021)

Os parâmetros das distribuições de probabilidade pré e pós-tratamento são fixos de modo que o número médio de estado estacionário de mRNAs seja ~ 10 e ~ 100 , res-

pectivamente. As distribuições bimodais indicam que os mRNAs sintetizados enquanto o gene estiver ON se degradarão rapidamente após passar para o estado OFF. Para as distribuições em forma de mesa e quasi-poissonianas, o chaveamento entre os estados ON-OFF é, respectivamente, lento e rápido o suficiente para garantir que a degradação do mRNA seja compensada por sua síntese, gerando as distribuições específicas. O limite de *burst* é caracterizado pelo estado ON de duração muito curta, no qual a síntese é muito eficiente, enquanto que a duração do estado OFF é proporcionalmente muito longa.

Nas próximas subseções, apresentamos os resultados das simulações da dinâmica da resposta ao tratamento considerando as cinco condições de pré-tratamento mostradas na Fig. 12. As trajetórias de $\langle n \rangle(t)$ e $\sigma(t)$ foram obtidas, respectivamente, usando as Eqs. (5.15) e (5.16) dentro dos intervalos discretos usados para aproximar as taxas cinéticas após a injeção de tratamento dada pelas Eqs. (5.7) e (5.8).

As Figs. 13-15 mostram a resposta após a aplicação de uma única (ou múltiplas) dose(s) nos gráficos rotulados como A-E (ou F-J) seguido pelo número que indica o cenário de tratamento (1, 2, ou 3). As condições de pré-tratamento têm $\epsilon_0 = (0,1; 1; 2; 10; 10)$. As linhas pretas tracejadas indicam o valor médio pretendido após o tratamento. As linhas verdes indicam $\langle n \rangle(t)$, e as linhas vermelhas indicam os valores de $\langle n \rangle(t) \pm \sigma(t)$. Os parâmetros do tratamento são definidos para permitir que o número médio de produtos gênicos atinja o regime de pós-tratamento mostrado na Fig. 12 para cada conjunto de valores de parâmetro.

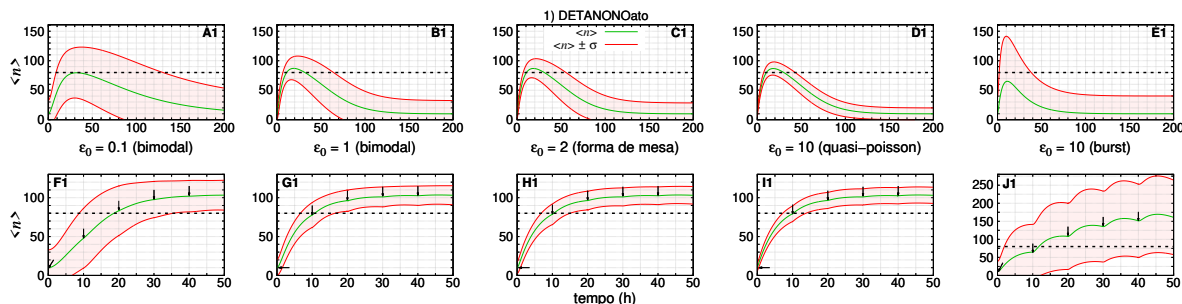
5.2.1 Tratamento visando a taxa de chaveamento de estado do gene de OFF para ON

A Fig. 13 mostra uma simulação da dinâmica do número médio de mRNAs e seu desvio padrão resultante de uma mudança em f após a introdução de DETANONOato. O erro absoluto de cada subintervalo da aproximação por partes para $f(t)$ é 1×10^{-4} . Assumimos que as taxas ($k; h; \rho$) são constantes durante o tratamento. Para obter os gráficos A1-D1 e F1-I1, definimos $(k_0; N_0) = (k_1; N_1) = (18,5; 110)$ e $(A_0; A_1) = (0,09; 0,9)$. Os gráficos E1 e J1 têm $(k_0; N_0) = (k_1; N_1) = (166,7; 1000)$ e $(A_0; A_1) = (0,01; 0,1)$ para garantir o regime de *burst* da expressão do gene durante a fase de pré-tratamento.

Nos gráficos A1-E1 (ou F1-J1), definimos $f_0 = (0,0015; 0,015; 0,03; 0,15; 0,017)$ e $f_1 = (0,14; 1,37; 2,73; 13,65; 0,18)$, de modo que $h = (0,015; 0,15; 0,3; 1,52; 1,65)$ e

$\epsilon_1 = (0,9; 9,1; 18,2; 91; 11)$ com os valores dados de A_0 , A_1 e ρ . Considera-se que a dose máxima tolerada faz com que a probabilidade de estado estacionário ON seja multiplicada por dez, conforme indicado pelos valores de A_0 e A_1 .

Figura 13 – Dinâmica da resposta ao tratamento com DETANONOato.



A dinâmica do número de RKIP mRNA (linha sólida verde) e seu desvio padrão (linha sólida vermelha) ao longo do tempo de tratamento visando a taxa cinética f . A linha tracejada preta em 80 indica um limiar arbitrário no número de mRNA de RKIP. Os gráficos A1-E1 mostram doses únicas da droga com valores progressivos diferentes para ϵ_0 (mostrado no respectivo gráfico). Os gráficos F1-J1 mostram as mesmas condições iniciais que os gráficos A1-E1 para 5 doses com intervalo de tempo de 10h (indicado por setas). Todas as doses estão na máxima tolerância ($\xi_a = 1$).

Fonte – Giovanini *et al.* (2021)

A primeira e a segunda linhas de gráficos da Fig. 13 mostram a simulação da dinâmica de expressão após uma e cinco doses da droga, respectivamente, em cinco condições de pré-tratamento.

Os gráficos A1, B1 mostram a resposta da dinâmica do gene com chaveamento lento, um regime de síntese de mRNAs cujos números são governados por uma distribuição bimodal no estado estacionário. Para $\epsilon_0 = 0,1$, temos a resposta mais lenta ao tratamento e o retorno às condições anteriores ao tratamento.

O número médio de mRNAs atinge um máximo de ~ 80 após ~ 30 h. O desvio padrão é o segundo maior. Os gráficos A1-D1 mostram o número médio de mRNA alcançando ou cruzando o limiar em nossas simulações, o tempo para o número médio atingir um máximo quando $\epsilon_0 \geq 1$ é ~ 20 h, e o ruído da resposta diminui à medida que aumentamos ϵ_0 . A exceção é mostrada no gráfico E1 onde o ruído é maior e o tempo para que o número médio alcance o máximo é ~ 10 h. Nesse caso, o número médio não ultrapassa o limiar e a linha inferior $\langle n \rangle(t) - \sigma(t)$ não aparece.

A resposta ao tratamento com múltiplas doses é mostrada nos gráficos F1-J1. O intervalo entre as doses foi escolhido para permitir uma resposta do tipo sigmoidal, caracterizando dois níveis distintos de expressão com o número médio aumentando de 10

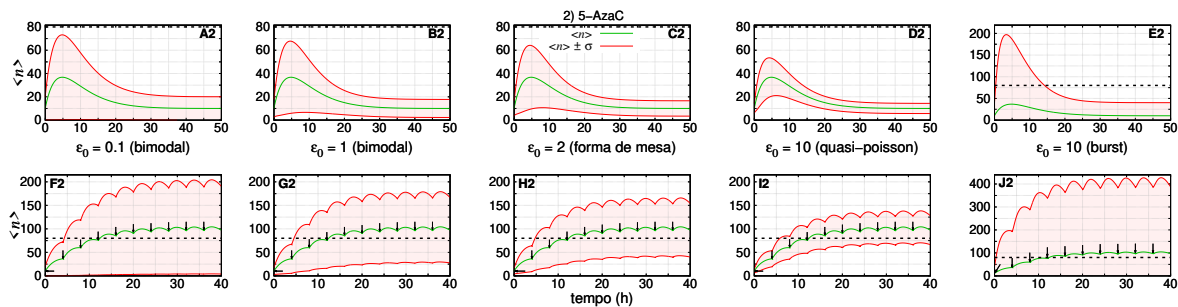
para pelo menos ~ 100 . O gráfico F1 mostra que o gene de chaveamento lento também causa a resposta mais lenta quando $\langle n \rangle(t)$ cruza o limiar após ~ 18 h. A curva para $\langle n \rangle(t) - \sigma(t)$ também cruza o limiar após ~ 35 h, o que garante uma resposta menos heterogênea ao tratamento. Do gráfico G1 ao I1 são mostradas as simulações para valores crescentes da velocidade do chaveamento gênico.

Essas simulações mostram que $\langle n \rangle(t)$ cruza o limiar após ~ 10 h, e as curvas $\langle n \rangle(t) - \sigma(t)$ alcançam valores mais altos, o que estabelece a resposta ao tratamento com o mínimo de heterogeneidade. O gráfico J1 mostra a resposta de um gene em *burst*, onde $\langle n \rangle(t)$ cruza o limiar após ~ 12 h e atinge um valor máximo de ~ 170 . No entanto, este regime causa a resposta mais ruidosa, conforme indicado por $\langle n \rangle(t) - \sigma(t)$ não ultrapassando o limiar. Observe também que há alguns "solavancos", pois o efeito da dose única está próximo do máximo no momento da próxima aplicação da droga. É possível prevenir os solavancos, no entanto, optamos por deixá-los evidentes para demonstrar que o intervalo de tempo entre as doses da droga também requer consideração na dinâmica do gene alvo.

5.2.2 Tratamento visando a taxa de síntese de mRNA de RKIP

A Fig. 14 mostra a dinâmica do número médio de mRNAs de RKIP e os desvios padrão sob tratamento com 5-AzaC. O erro absoluto de cada subintervalo da aproximação por partes para $k(t)$ é 1×10^{-4} . Assumimos que as taxas $(f; h; \rho)$ permanecem constantes durante o tratamento, o que implica que $A_0 = A_1$ e $\epsilon_0 = \epsilon_1$ também permanecem constantes. Para os gráficos A2-D2 e F2-I2, definimos $(k_0; N_0; k_1; N_1) = (3,3; 20; 33; 200)$, $A_0 = 0,5$ e $f = h = (0,008; 0,08; 0,17; 0,83)$. Para os gráficos E2 e J2, definimos $(k_0; N_0; k_1; N_1) = (166,7; 1000; 1667; 10000)$, $A_0 = 0,01$ e $(f; h) = (0,017; 1,65)$.

Figura 14 – Dinâmica da resposta ao tratamento com 5-Azac.



A dinâmica do número de mRNA de RKIP (linha sólida verde) e seu desvio padrão (linha sólida vermelha) ao longo do tempo de tratamento visando k taxa cinética. A linha tracejada preta em 80 indica um limiar arbitrário no número de mRNA de RKIP. Os gráficos A2-E2 mostram doses únicas de drogas com valores progressivos diferentes para ϵ_0 (mostrado dentro do respectivo gráfico). Os gráficos F2-J2 mostram as mesmas condições iniciais que os gráficos A2-E2 para 10 doses com intervalo de tempo de 4h (indicado por setas). Todas as doses estão na tolerância máxima ($\xi_b = 1$).

Fonte – Giovanini *et al.* (2021)

As primeira e segunda linhas de gráficos da Fig. 14 mostram a simulação da dinâmica de expressão após uma e dez doses da droga, respectivamente, em cinco condições de pré-tratamento.

Os tratamentos não afetam as taxas de decaimento do sistema, ϵ ou ρ . Sendo assim, os gráficos A2-E2 mostram que o número médio de mRNAs atinge o máximo (~ 35 moléculas) após intervalos semelhantes de ~ 5 h e atinge as condições de pré-tratamento após ~ 40 h. O gráfico A2 mostra a condição com o segundo maior desvio padrão. Os gráficos A2-D2, mostram que o ruído da resposta diminui com o aumento de ϵ_0 . O regime de *burst* mostrado no gráfico E2 leva à resposta mais ruidosa. De fato, $\langle n \rangle(t) - \sigma(t)$ permanece negativo, enquanto $\langle n \rangle(t) + \sigma(t)$ excede o limiar em mais de 2 vezes.

A resposta ao tratamento com múltiplas doses é mostrada nos gráficos F2-J2. O intervalo entre as doses foi escolhido para permitir uma resposta do tipo sigmoidal e com o número médio atingindo pelo menos 100 transcritos. Todos os gráficos mostram que $\langle n \rangle(t)$ cruza o limiar após ~ 10 h. Isso ocorre porque a resposta aos tratamentos e o intervalo entre as aplicações das drogas são os mesmos em todos os cenários. Uma vez que ϵ_0 e ρ não são afetados, as taxas de decaimento para as condições de pré-tratamento também não são afetadas.

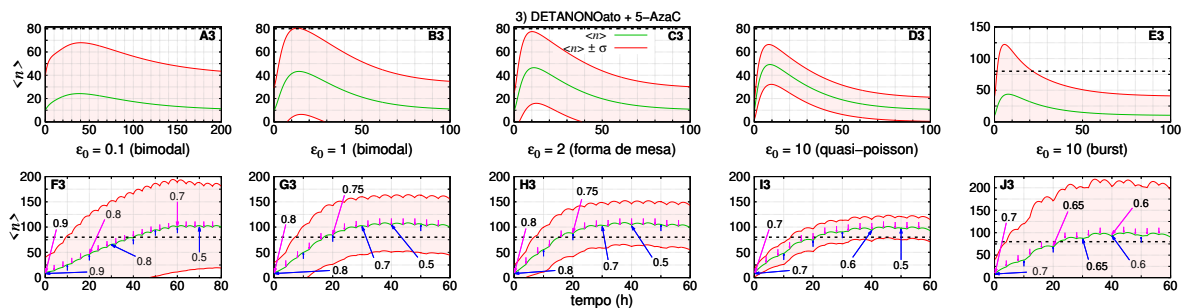
Em todos os cenários, as curvas para $\langle n \rangle(t) - \sigma(t)$ não ultrapassam o limiar porque o tratamento não aumenta ϵ_0 , o que causaria uma redução no ruído na síntese de mRNA em um determinado cenário. No entanto, os gráficos F2-I2 mostram que $\langle n \rangle(t) - \sigma(t)$ atinge

valores máximos mais altos à medida que ϵ_0 aumenta. O gráfico J2 mostra a resposta no caso de *bursts* transcricionais, o que leva à resposta mais heterogênea ao tratamento, conforme indicado pela curva $\langle n \rangle(t) + \sigma(t)$, que atinge o maior valor entre os máximos.

5.2.3 Tratamento com as duas drogas concomitantemente

A Fig. 15 mostra a dinâmica do número médio de mRNAs e seu desvio padrão sob tratamento com ambas as drogas: DETANONOato e 5-AzaC. Os erros absolutos de cada subintervalo da aproximação por partes para $f(t)$ e $k(t)$ são 1×10^{-4} para ambos. As taxas ($h; \rho$) permanecem constantes durante o tratamento. Para gráficos A3-D3 (ou F3-I3), definimos $(k_0; N_0; k_1; N_1) = (18,5; 110; 33; 200)$, $(A_0; A_1) = (0,09; 0,5)$, $f_0 = (0,0015; 0,015; 0,03; 0,15)$ e $f_1 = h = (0,015; 0,15; 0,3; 1,5)$. Para os gráficos E3 e J3, definimos $(k_0; N_0; k_1; N_1) = (167; 1000; 333; 2000)$, $(A_0; A_1) = (0,01; 0,05)$, $f_0 = 0,017$, $f_1 = 0,087$ e $h = 1,65$. Os valores de A_0 , A_1 e ρ resultam em $\epsilon_1 = (0,18; 1,8; 3,6; 18; 10,4)$ nos gráficos A3-E3 (ou F3-J3).

Figura 15 – Dinâmica da resposta ao tratamento com DETANONOato e 5-Azac.



A dinâmica do número de mRNAs de RKIP ao longo do tempo para tratamento com duas drogas que almejam as taxas cinéticas de f e k é mostrada em linhas contínuas verdes e seus desvios padrão em linhas contínuas vermelhas. O valor limiar para o RKIP é 80 e aparecem nos gráficos como uma linhas tracejada preta. Os gráficos A3-E3 mostram as doses únicas de ambos as drogas para diferentes valores progressivos de ϵ_0 (indicados em cada gráfico). Os gráficos F3-J3 mostram as mesmas condições iniciais que os gráficos A3-E3 para múltiplas doses fracionadas da droga que visam f (ou k) com intervalo de tempo de 10h (ou 4h), na ordem, com 8 (ou 20) doses no gráfico F3 e 6 (ou 15) doses nos gráficos G3-J3. As setas indicam a agenda com os instantes da aplicação e frações de doses (ξ_a e ξ_b) para cada droga, DETANONOato (visando f) em azul e 5-AzaC (visando k) em magenta. Os números nas setas maiores indicam a fração da dose aplicada naquele momento e nas seguintes (setas sem números). A agenda também é mostrada na Tab. 5.

Fonte – Giovanini *et al.* (2021)

A Fig. 15 mostra as simulações da dinâmica da resposta ao tratamento. Os gráficos da primeira linha indicam os resultados da aplicação simultânea de uma única dose das

duas drogas. A segunda linha mostra os resultados das aplicações de doses múltiplas com a droga almejando f (ou k) sendo aplicada a cada 10h (ou 4h) de acordo com a agenda na Tab. 5.

O gráfico A3 mostra a dinâmica da resposta ao tratamento quando as condições iniciais são definidas para um gene de chaveamento lento. O número médio atinge um máximo de ~ 25 em $t \sim 35$ h. O máximo de $\langle n \rangle(t) + \sigma(t)$ é ~ 70 , sendo esta, uma resposta altamente heterogênea se considerarmos que o desvio padrão máximo é igual à média máxima — isto é, um regime super-poissoniano de transcrição gênica. O retorno ao pré-tratamento é lento, conforme indicado pela queda da média após ~ 200 h.

Nos gráficos B3, C3 e D3, os números médios máximos de mRNAs são ~ 40 , ~ 45 e ~ 50 , respectivamente, alcançados após 15, 12 e 8h. As curvas $\langle n \rangle(t) - \sigma(t)$ e $\langle n \rangle(t) + \sigma(t)$ tornam-se mais próximas com o aumento de ϵ_0 como mostrado nos gráficos A3-D3. O limite de *burst* é diferente e, apesar do valor alto de ϵ_0 , possui ruído maior conforme mostra o gráfico E3. Neste regime, $\langle n \rangle(t)$ máximo é ~ 45 alcançado após ~ 8 h. $\langle n \rangle(t) + \sigma(t)$ atinge um máximo de ~ 120 , indicando a resposta mais ruidosa. Para todas as condições de pré-tratamento, o número médio não ultrapassa o limiar.

Tabela 5 – As frações ξ_a e ξ_b da dose máxima tolerada para agendas de tratamento das condições de pré-tratamento da Fig. 15.

Gráfico	Sequência do número de doses \times fração		Redução acumulada em	
	ξ_a	ξ_b	ξ_a	ξ_b
F3	$3 \times 0,9; 4 \times 0,8; 1 \times 0,5$	$5 \times 0,9; 10 \times 0,8; 5 \times 0,7$	20%	20%
G3	$3 \times 0,8; 1 \times 0,7; 2 \times 0,5$	$5 \times 0,8; 10 \times 0,75$	32%	23%
H3	$3 \times 0,8; 1 \times 0,7; 2 \times 0,5$	$5 \times 0,8; 10 \times 0,75$	32%	23%
I3	$4 \times 0,7; 1 \times 0,6; 1 \times 0,5$	$15 \times 0,7$	35%	30%
J3	$3 \times 0,7; 1 \times 0,65; 2 \times 0,6$	$5 \times 0,7; 5 \times 0,65; 5 \times 0,6$	29%	35%

Fonte – Giovanini *et al.* (2021)

A resposta ao tratamento com múltiplas doses é mostrada nos gráficos F3-J3. A Tab. 5 mostra as sequências de aplicações de uma determinada droga em frações de dose máxima tolerada, ξ_a e ξ_b . Os valores foram escolhidos visando que a quantidade total de drogas seja reduzida em comparação ao tratamento com uma única droga. Também tentamos garantir uma resposta do tipo sigmoidal de forma que o número médio alcance pelo menos ~ 100 . O gráfico F3 mostra a dinâmica da resposta para 8 (ou 20) doses de drogas almejando f (ou k) com ξ_a e ξ_b variando de 0,9 a 0,5 (detalhes na Tab. 5).

A agenda permite uma redução de 20% em comparação com a aplicação de doses completas ($\xi_a = \xi_b = 1$). Como o gene em pré-tratamento está em um regime de chaveamento lento, isso leva à resposta mais lenta, pois $\langle n \rangle(t)$ cruza o limiar após ~ 40 h. Também possui alto ruído, como pode ser notado pelos valores de $\langle n \rangle(t) \pm \sigma(t)$. Os gráficos G3-J3 mostram as simulações da resposta para 6 (ou 15) doses de drogas a (ou b) com ξ_a e ξ_b variando de 0,8 a 0,5 (ver a Tab. 5).

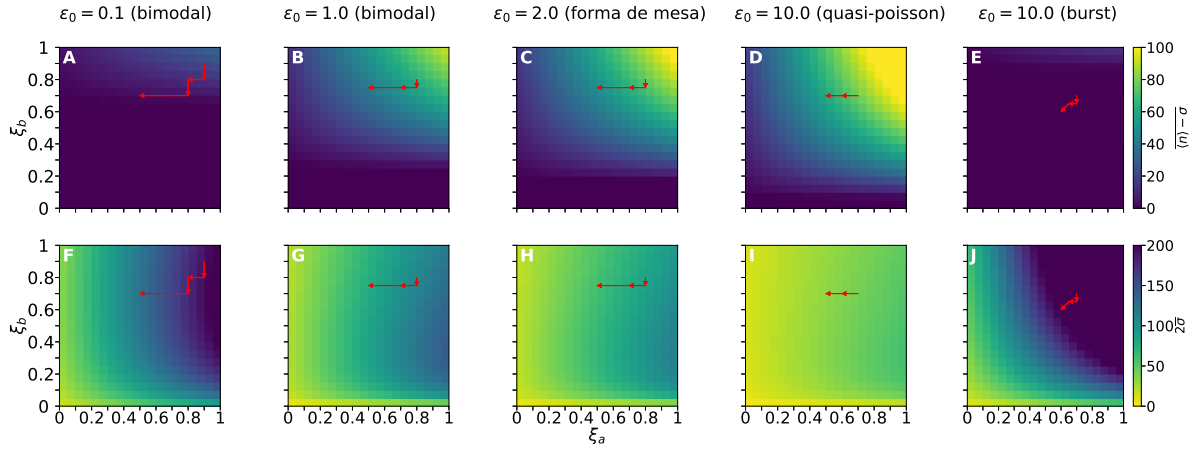
A redução cumulativa em doses completas varia de 29% a 35% (ou de 23% a 35%) em ξ_a (ou ξ_b). Os gráficos G3-I3 mostram que $\langle n \rangle(t)$ cruza o limiar após ~ 15 h. As curvas $\langle n \rangle(t) \pm \sigma(t)$ tornam-se mais próximas à medida que o valor de ϵ_0 aumenta, o que indica a direção de como desenhar estratégias de tratamento com heterogeneidade da resposta reduzida. O gráfico J3 mostra a condição de pré-tratamento do *burst*. Aqui, $\langle n \rangle(t)$ cruza o limiar após ~ 22 h e $\langle n \rangle(t) + \sigma(t)$ atinge ~ 200 , que indica a resposta mais ruidosa.

5.2.4 Mapeamento da resposta ao tratamento para o efeito do fracionamento ξ_a e ξ_b na redução de drogas

A Fig. 16 mostra a resposta ao tratamento medida em termos dos valores médios (não ponderados) $\overline{\langle n \rangle} - \sigma$ (gráficos A-E) e 2σ (gráficos F-J) para o intervalo entre 60 a 80h após a primeira dose. Cinco regimes de pré-tratamento marcados com valores ϵ_0 são considerados. A efeito da fração de DETANONOato e 5-AzaC em f e k é indicada por ξ_a e ξ_b , respectivamente. Calculamos o valor médio da função $x(t)$ entre os instantes t_i e t_f , denotado por $\overline{x(t)}$, usando a integral $\overline{x(t)} = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} x(t) dt$.

O erro absoluto da aproximação por partes para o decaimento exponencial do efeito da droga é definido como 1×10^{-4} . Para obter cada ponto dos mapas de calor na Fig. 16, calculamos a dinâmica da resposta do tratamento a uma agenda de doses com frações constantes ξ_a e ξ_b . O DETANONOato foi aplicado em 8 doses com intervalo de tempo de 10h e o 5-AzaC em 20 doses com intervalo de 4h. A grade é construída calculando as trajetórias para ξ_a e ξ_b variando de 0 a 1 em incrementos de 0,05.

Figura 16 – Mapeamento da resposta ao tratamento para o efeito do fracionamento das drogas.



A efetividade das agendas de tratamento medida pelos valores médios $\overline{\langle n \rangle} - \sigma$ (linha superior) e $2\bar{\sigma}$ (linha inferior) durante o intervalo de 60 a 80h após a aplicação da primeira dose é apresentada. As barras verticais à direita fornecem um código de cores que denota os valores médios. Em cada quadrado do mapa de calor, indicamos o valor de $\overline{\langle n \rangle} - \sigma$ e $2\bar{\sigma}$ como resultado do efeito do fracionamento das doses ξ_a e ξ_b nas taxas f e k , respectivamente. A análise foi realizada para as cinco condições de pré-tratamento identificadas pelo valor da velocidade de chaveamento ϵ_0 , rotulado em cada coluna. Para cada par (ξ_a, ξ_b) , simulamos a aplicação de DETANONOato (e 5-AzaC) a cada 10h (e 4h). Para um tratamento que visa o gene RKIP, a redução da dose para minimizar os efeitos citotóxicos para cada ϵ_0 consiste em manter (ξ_a, ξ_b) dentro de regiões com altos valores de $\overline{\langle n \rangle} - \sigma$ nos gráficos A-E e baixos valores de $2\bar{\sigma}$ nos gráficos F-J. As regiões amarelas dos mapas de calor indicam os domínios mais desejáveis avaliados em termos de $\overline{\langle n \rangle} - \sigma$ e $2\bar{\sigma}$. A sequência de setas vermelhas em cada gráfico corresponde ao efeito das combinações de fração de dose nas agendas da Fig. 15 e Tab. 5, para cada condição de pré-tratamento.

Fonte – Giovanini *et al.* (2021)

A redução dos efeitos do fracionamento das doses ξ_a e ξ_b para as agendas mostradas na Tab. 5 e na Fig. 15 foram baseadas nos mapas de calor da Fig. 16. O melhor cenário está configurado para os pontos que se localizam na região mais amarela dos mapas de calor. As mudanças no efeito do fracionamento para reduzir as doses ao longo do tempo são indicadas pela sequência de setas vermelhas nos gráficos que indicam cada regime de pré-tratamento. Os regimes bimodal (gráficos A e F) e de *burst* (gráficos E e J) são as condições de pré-tratamento mais difíceis para redução da dose porque quase não há $(\xi_a, \xi_b) \rightarrow (0, 0)$ que garanta valores suficientemente altos de $\overline{\langle n \rangle} - \sigma$ (gráficos A e E) e $2\bar{\sigma}$ baixos (gráficos F e J).

As condições adicionais de pré-tratamento, propriamente, a bimodal remanescente, a em forma de mesa e a quasi-poissoniana (gráficos B-D e G-I), permitem melhores cenários para redução da dose devido aos maiores valores de ϵ_0 . No geral, todos os regimes de

pré-tratamento permitem uma redução maior de ξ_a do que de ξ_b , como pode ser observado na agenda proposta na Tab. 5 (veja as setas vermelhas na Fig. 16), por causa da ação de 5-AzaC em k (dado em termos de ξ_b). Valores maiores de ξ_b permitem aumentos maiores nos níveis médios de mRNA de RKIP.

No entanto, os regimes bimodal e de *burst* permitem reduzir ainda mais ξ_b , pois o desvio padrão diminui ou permanece constante, enquanto os números médios de mRNA de RKIP permanecem quase constantes. No entanto, embora seja possível reduzir a variância, os valores médios não atingem valores suficientemente grandes. Logo, é necessário entender como o desenho do tratamento pode ter como alvo as outras taxas cinéticas, h e ρ , para permitir uma resposta eficaz independente das condições pré-tratamento.

5.2.5 Melhorando os tratamentos não-efetivos visando todas as taxas cinéticas

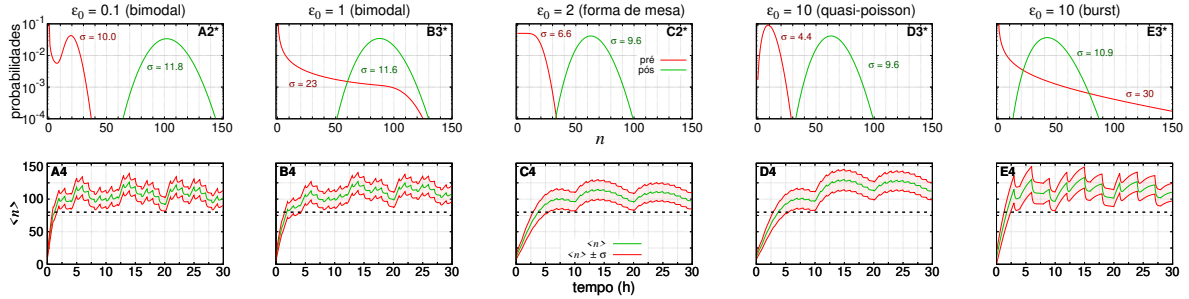
A resposta ao tratamento aprimorado é apresentado na Fig. 17 para cinco condições iniciais qualitativamente diferentes visando uma distribuição ideal. O coquetel de drogas hipotético visa todas as taxas do modelo e, assim, alcançamos um tempo da resposta ideal e a redução de heterogeneidade para todas as condições iniciais qualitativamente distinguíveis, o que, nas subseções anteriores, levou a resultados insatisfatórios. As distribuições de pré-tratamento são mostradas pelas curvas vermelhas dentro dos gráficos na Fig. 12: A2, B3, C2, D3, E3, que, respectivamente, descrevem os seguintes regimes de transcrição gênica no estado estacionário: bimodal, limite bimodal, em forma de mesa, quasi-poissoniano, e de *burst*.

Nosso objetivo era manter o número médio de mRNAs em torno de 100 e suas flutuações acima do limiar. As taxas de decaimento das drogas que afetam k , ρ , f e h são denotadas por λ_i , onde i indica a taxa visada pela droga. Essas taxas estão em h^{-1} e fixadas nos valores $(\lambda_k; \lambda_\rho; \lambda_f; \lambda_h) = (0,25; 6; 0,05; 0,053)$, e a dose máxima tolerada de cada droga i é indicada por $\xi_i = 1$.

Os valores de λ_f e λ_k são os mesmos usados anteriormente para o DETANONOato e o 5-AzaC, respectivamente, enquanto que os valores de λ_ρ e λ_h foram definidos com base no tempo de duração da ligação mRNA-microRNA (MOROZOVA *et al.*, 2012) e h na meia-vida de BACH1 (HOU *et al.*, 2008). A dependência do tempo de $h(t)$ e $\rho(t)$ é descrita seguindo a mesma estrutura usada para $k(t)$ e $h(t)$ — veja Eqs. (5.5)-(5.8). Para obter

a dinâmica do número médio de mRNAs e do desvio padrão, estendemos a formulação anterior da equação mestra, Eqs. (5.10) e (5.11), para todas as taxas cinéticas.

Figura 17 – Aprimoramento de alguns desenhos de tratamentos não-efetivos.



Os desenhos de tratamento aprimorados resultantes das distribuições dos gráficos A2, B3, C2, D3 e E3 da Fig. 12 são apresentados. Simulamos um coquetel hipotético de drogas alterando todas as taxas (f, k, h, ρ). A primeira linha de gráficos mostra as distribuições de probabilidade de pré (linhas vermelhas) e pós-tratamento (linhas verdes) que governam os números de mRNA. O asterisco adicionado ao rótulo do gráfico indica que estamos simulando o desenho do tratamento aprimorado para o respectivo tratamento não-efetivo de mesmo rótulo na Fig. 12. O número médio $\langle n \rangle$ para todas as distribuições pré-tratamento é 10 e para cada distribuição pós-tratamento é (103, 88, 63, 63, 44), respectivos à sequência dos gráficos da esquerda para a direita. A segunda linha mostra a respectiva dinâmica das respostas ao desenho de tratamento aprimorado nos gráficos A4-E4. As linhas verdes indicam o número médio de mRNAs de RKIP, $\langle n \rangle$. As linhas vermelhas mostram um desvio padrão em torno da média, $\langle n \rangle \pm \sigma$. Linhas pretas tracejadas em $\langle n \rangle = 80$ indicam o limiar que separa os níveis de expressão de mRNA em uma célula saudável (linha acima) e no câncer (abaixo). Os parâmetros das distribuições pós-tratamento e a agenda de doses de cada desenho de tratamento aprimorado estão apresentados no Apêndice L.

Fonte – Giovanini *et al.* (2021)

O desvio padrão da distribuição de probabilidade pós-tratamento é definido sendo

$$\sigma_1 = \sqrt{N_1 A_1 \left[1 + N_1 \frac{(1 - A_1)}{1 + \epsilon_1} \right]}, \quad (5.17)$$

que tem um mínimo local em $N_1 \rightarrow 0$ e um máximo local em

$$A_1 = \frac{1}{2} \left(1 + \frac{1 + \epsilon_1}{N_1} \right), \quad (5.18)$$

onde N é o número máximo de mRNA, A é a probabilidade de estado estacionário para o gene estar ON, ϵ é a velocidade de chaveamento do gene — veja a Eq. (5.12), onde o subscrito 1 indica os valores dos parâmetros pós-tratamento pretendidos. Uma estratégia para reduzir σ_1 é manter $N_1 \gtrsim 1 + \epsilon_1$ (ou $k_1 \gtrsim f_1 + h_1 + \rho_1$), o que garante que σ_1 se aproxime do máximo local como $A_1 \rightarrow 1$. A redução dos valores de σ_1 depende do número médio $\langle n \rangle_1$ desejado, de forma que $\langle n \rangle_1 \approx A_1(1 + \epsilon_1)$.

Para aprimorar os desenhos de tratamentos mostrados anteriormente, usamos a abordagem de otimização que acabamos de descrever acima. A dose da droga e o intervalo

de aplicação são definidos para permitir $\epsilon_1 = 91$, que verificamos ser o valor que levou ao desenho de tratamento bem-sucedido mostrado nos gráficos D1 da Fig. 12. Na verdade, isso resultou na melhor dinâmica da resposta do número médio de mRNAs de RKIP, conforme mostrado no gráfico I1 da Fig. 13.

Os gráficos A4 e B4 mostram a dinâmica da resposta ao tratamento aprimorado quando a transcrição do gene RKIP em pré-tratamento é governada por um promotor de chaveamento lento dando origem a distribuições de probabilidade bimodal ($\epsilon_0 \leq 1$). O desenho de tratamento não-efetivo voltado para os regimes bimodal e em forma de mesa é, respectivamente, mostrado pelas linhas verdes dos gráficos A2 e B3 na Fig. 12. A dinâmica da resposta aos desenhos de tratamento aprimorados é mostrada nos gráficos A2* e B3* da Fig. 17. Ambas as respostas otimizadas resultaram de mudanças semelhantes nos parâmetros cinéticos que causaram um aumento em A_1 e uma redução em N_1 e têm agendas semelhantes (detalhadas no Apêndice L).

O desenho de tratamento aprimorado com o objetivo de um gene de chaveamento rápido em pré-tratamento é mostrado nos gráficos C4 e D4 da Fig. 17. As distribuições de probabilidade iniciais são em forma de mesa e quasi-poissonianas, e têm $\epsilon_0 > 1$. A dinâmica da resposta a desenhos de tratamento não-efetivos é mostrada nos gráficos C2 e D3 da Fig. 12. A resposta ao tratamento aprimorado com heterogeneidade reduzida e velocidade aumentada é mostrada nos gráficos C2* e D3* da Fig. 17. As agendas aprimoradas de tratamento foram as mesmas em ambos os casos (ver parâmetros e agendas no Apêndice L). O valor de $\langle n \rangle_1$ foi reduzido em comparação com o desenho de tratamento não-efetivo devido à diminuição de N_1 não ser compensada pelo aumento de A_1 .

O desenho de tratamento não-efetivo para um regime inicialmente em *burst* de transcrição de RKIP mostrado no gráfico E3 da Fig. 12 foi aprimorado. A distribuição pós-tratamento resultante é quasi-poissoniana. A dinâmica da resposta ao desenho de tratamento aprimorado é mostrada, respectivamente, nos gráficos E3* e E4 da Fig. 17. Este aprimoramento exigiu as maiores reduções em $\langle n \rangle_1$ em comparação com o desenho não-efetivo, aumentando A_1 e diminuindo N_1 . Observe, no entanto, a clara redução nos desvios padrão (heterogeneidade) da resposta. Os valores dos parâmetros e agenda aplicados são detalhados no Apêndice L.

5.3 Discussão

Quando consideramos o modelo binário para regulação da expressão gênica, é importante caracterizar o regime de expressão gênica pré-tratamento em termos de A_0 , ϵ_0 e N_0 — ver a Eq. (5.12). Essas taxas afetarão o ruído pré-existente (e pós-tratamento) no número de mRNAs — ver as Eqs. (5.13) e (5.14) e Figs. 13-15 e 17. Além disso, o tempo da resposta ao tratamento também depende de ϵ_0 , isto é devido ao regime de decaimento para estado estacionário do número médio de mRNAs e sua variância terem ρ e $\epsilon\rho$ como suas menores taxas — veja a Eq. (5.15). Para $\epsilon > 1$ e $\epsilon < 1$, o tempo para o sistema atingir o estado estacionário é, respectivamente, $\propto \rho^{-1}$ e $\propto (\epsilon\rho)^{-1}$.

Todas as trajetórias dos valores médios mostrados nos gráficos da Fig. 14 atingem um máximo (gráficos A2-E2) ou cruzam o limiar (gráficos F2-J2) após um intervalo aproximadamente fixo. Isso ocorre porque o componente $\propto e^{-\epsilon\rho t}$ de $\langle n \rangle(t)$ é nulo, pois o tratamento não altera a probabilidade de estado estacionário do gene para estar ON — ver Eq. (I.1). Além disso, o ruído na resposta ao tratamento, medido em termos das trajetórias de $\langle n \rangle(t) \pm \sigma(t)$, é maior quando temos os menores valores da velocidade de chaveamento relativa. A inspeção das trajetórias para única dose (múltiplas doses) mostradas nos gráficos A-D (F-I) das Figs. 13-15 ajuda a identificar essas características.

A condição de pré-tratamento, que permite a resposta mais rápida e menos heterogênea, ocorre para valores mais altos da velocidade de chaveamento relativa e para a probabilidade do gene estar ON sendo $> 0,1$. Os gráficos D e I das Figs. 13-15 mostram que essas são as condições que permitem as menores diferenças entre as trajetórias $\langle n \rangle(t) \pm \sigma(t)$. Nesse regime, o gene está se comportando como uma fonte quasi-poissoniana de transcrições (GIOVANINI *et al.*, 2020). Como o valor da velocidade de chaveamento do pré-tratamento é reduzido para dois, temos um aumento no ruído da resposta do tratamento, embora a velocidade da resposta não aumente significativamente, conforme mostrado nos gráficos C e H das Figs. 13-15.

O pré-tratamento em regime de *burst* ($\epsilon_0 = 10$) leva às respostas mais ruidosas ao tratamento, embora esteja entre as mais rápidas. A resposta ruidosa ao tratamento ocorre parcialmente porque quando simulamos os cenários de tratamento mostrados nos gráficos E e J das Figs. 13-15, o valor de $A(t)$ permaneceu pequeno. Nesse caso, a dinâmica ainda permaneceria em regime de *burst* transcricional. O pequeno aumento em $A(t)$ ocorre

porque assumimos que a dose máxima tolerada da droga responsável pelo aumento de f faria com que a probabilidade do estado ON fosse 10 vezes maior, independente de seu valor no pré-tratamento. Tal suposição precisa ser confirmada por estudos experimentais, e uma formulação alternativa pode ser proposta devido à falta de confirmação. Os gráficos B-E e G-J das Figs. 13-15 mostram que os tempos da resposta ao tratamento são todos semelhantes, uma vez que as taxas de decaimento para o estado estacionário são $\leq \rho$ para $\epsilon \leq 1$.

A condição de pré-tratamento de baixa velocidade de chaveamento relativa ($\epsilon_0 = 0,1$ e $\epsilon_0 = 1$) estabelece um regime desafiador a ser abordado. A resposta é mais lenta e ruidosa se seguirmos as estratégias consideradas aqui. Os gráficos A e F das Figs. 13 e 15 indicam que o aumento no número médio de mRNA é mais lento quando temos o mínimo de ϵ , apesar do aumento no tratamento f causar um aumento da velocidade de chaveamento relativa.

Nas respostas de alto ruído, as trajetórias $\langle n \rangle(t) - \sigma(t)$ não cruzam o limiar. Isso indica que, em uma população de células com as mesmas condições regulatórias de expressão do gene RKIP, a resposta ao tratamento é heterogênea e pode ser insuficiente. Portanto, a dosagem e o alvo precisam ser construídos para garantir uma redução na heterogeneidade da resposta. Isso pode ser realizado aumentando a velocidade de chaveamento, de modo que a expressão do gene se aproxime de um regime quasi-poissoniano. O aumento da velocidade de chaveamento também tem o benefício de reduzir o tempo da resposta à terapia.

O efeito do fracionamento para redução da dose de DETANONOato e 5-AzaC mostrado na Fig. 16 indica as dificuldades na otimização de agendas para regimes de pré-tratamento longe de ser quasi-poissoniano. Destacamos os regimes bimodal e de *burst*, onde as respostas precisam ter aumentos maiores em relação aos níveis médios de mRNA. Para isso, é necessário almejar todas as taxas cinéticas efetivas do modelo binário estocástico para a expressão gênica. Para o caso de RKIP, pode-se também direcionar os repressores Snail, BACH1, Sp1, CREB e p300 para alterar h , enquanto uma variedade de miRNA e seus possíveis inativadores modulariam ρ (ZARAVINOS *et al.*, 2018).

As respostas aos tratamentos aprimoradas mostrados na Fig. 17 (gráficos A4-E4) indicam o potencial de nossa abordagem para ajudar no desenho de tratamentos multi-drogas de baixa dosagem administradas em pulsos. Os três regimes de pré-tratamento qualitativamente diferentes, ou seja, bimodal lento, quasi-poissoniano rápido e de *burst*,

têm curvas de resposta distinguíveis. Esses são resultantes das especificidades do desenho do tratamento aprimorado com o objetivo de manter os números médios de mRNA e desvios (heterogeneidade) acima do limiar.

A síntese da proteína RKIP pode ter uma dependência não linear dos números do mRNA porque a tradução pode ser regulada por mecanismos, como estabilidade do mRNA, controle da tradução e degradação proteossômica (WALKER *et al.*, 2011). Portanto, uma abordagem teórica considerando a tradução pode exigir a adição de duas ou mais escalas de tempo características, dependendo da síntese e degradação da proteína (SHAHREZAEI; SWAIN, 2008). Enfatizamos que as quantidades de mRNAs de RKIP em nosso modelo são o efeito líquido do tratamento aprimorado, uma vez que sua expressão resulta da interação com uma rede complexa de genes. Nossa abordagem, no entanto, fornece os blocos de construção de redes de genes (RAMOS *et al.*, 2010) que apresentam-se úteis para entender o funcionamento de módulos maiores (BALÁZSI; OUDENAARDEN; COLLINS, 2011; FERRELL, 2016; GÉRARD; GONZE; GOLDBETER, 2019) e, portanto, para aprimorar ainda mais os desenhos de tratamento de câncer em estágio de metástase.

5-AzaC e DETANONOato são drogas não específicas para aumentar a RKIP, que também podem interferir com processos biológicos adicionais que ocorrem em células tumorais ou normais. O presente estudo não fornece uma visão sobre esses problemas e desenhos experimentais específicos são necessários para abordá-los. Inicialmente, pode-se considerar simulações usando condições de pré-tratamento correspondentes à condição normal. Nesse caso, é justo considerar que a fração de efeito da dose reduzida é menor, se assumirmos que as drogas têm maior absorção nas células cancerosas. Um modelo mais complexo para a dinâmica de vários genes é necessário para investigar o efeito geral de desenhos de tratamento não específicos. Alternativamente, pode-se desenhar tratamentos com maior especificidade para investigar a validade de nossa abordagem na simulação de desenhos terapêuticos em genes.

Outra possível aplicação de nossa abordagem é na investigação de desenhos terapêuticos baseados em 5-AzaC e DETANONOato. Por exemplo, a promoção de desmetilação do DNA por 5-AzaC pode ser usada para regular a expressão de genes envolvidos na reversão do envelhecimento das células-tronco adultas (YAN *et al.*, 2014), enquanto suas propriedades inibidoras da DNA metiltransferase tem o potencial para regenerar células ciliadas do ouvido interno de mamíferos maduros (DENG *et al.*, 2019). Doadores de NO como o DETANONOato estão envolvidos em muitos processos celulares, como vasodilata-

ção, neurotransmissão, imunidade mediada por macrófagos e respostas anti-inflamatórias (HAYS; BONAVIDA, 2019). Uma aplicação de nossa abordagem é ajudar a elucidar os mecanismos desses desenhos de terapias.

Para melhorar ainda mais nossa abordagem, é necessário realizar uma validação do nosso modelo baseada em dados. Um possível experimento envolveria a diminuição dos níveis de RKIP usando uma abordagem de siRNA para simular diferentes níveis de ativação do gene RKIP. Esta seria uma forma mais específica e precisa de interferir na via RKIP em comparação com o uso de 5-AzaC. Poderíamos fazer diferentes níveis de inibição de RKIP por siRNA e observar se a modelagem consegue prever os níveis de fosforilação de MEK e ERK, a expressão de genes induzidos por MAPK e a consequência dessas ativações na migração, proliferação e invasão celular. A abordagem de siRNA não é capaz de verificar o estado OFF do gene RKIP; entretanto, é possível usar um sistema CRISPR/CAS9 para eliminar o gene RKIP, e então ter a informação do estado OFF. Métodos de inferência seriam empregados na estimativa de valores de parâmetros e números de mRNA de RKIP.

Tais experimentos ajudariam a entender como a informação ambiental é processada pelas células, investigando como a regulação da expressão do gene RKIP afeta suas relativas quinases em *downstream*. Recentemente, demonstrou-se que o regime de baixa velocidade de chaveamento maximiza a informação mútua entre o número de transcrições e a atividade do promotor (GAMA *et al.*, 2020). Este regime também coincide com condições de resposta lenta ao tratamento e altamente heterogênea. Desta forma, podemos perguntar se é relevante para um gene regulador mestre transmitir informações sobre o estado do seu próprio promotor e em que contexto celular, propriamente, numa célula cancerosa ou saudável. A resposta para esta questão é importante pois nos ajuda a compreender o papel do ruído e as condições em que sua redução é desejável (RAMOS; REINITZ, 2019) como ocorre quando temos um gene autorregulado negativamente.

6 Considerações finais

Neste trabalho apresentamos dois modelos binários estocásticos de expressão gênica para o estado estacionário: um externamente reprimido (o GER) e outro autorreprimido (o GAR). O Fator de Fano permite a caracterização do ruído nas distribuições de probabilidade que governam o número de produtos gênicos em três possíveis regimes: sub-Fano, Fano e super-Fano, sendo a covariância entre o número de produtos gênicos e o estado do promotor uma medida capaz de fornecer esta mesma informação. Mostramos que apenas o GAR é dotado de sub-Fano, onde a redução de ruído permite o controle mais preciso no número de produtos gênicos. Para o regime Fano, tanto o GER quanto o GAR apresentam distribuições quasi-poissonianas, porém apenas o GAR apresenta um regime poissoniano como na expressão gênica constitutiva. O regime super-Fano apresenta distribuições com diversidade nos seus formatos, tanto para o GER, quanto para o GAR e há grande similaridade entre os formatos dos dois modelos, podendo trazer desafios para inferir a estratégia de regulação gênica apenas pelo ruído na expressão.

Como prova de conceito, apresentamos um modelo binário estocástico para a transcrição do gene RKIP sob tratamento induzido pelas taxas cinéticas dependentes do tempo. As soluções exatas do modelo de terapia gênica são aproximadas usando as soluções exatas do modelo equivalente com taxas cinéticas estacionárias. Este é o modelo mais simples que pode ser resolvido exatamente para descrever a transcrição regulada do gene RKIP. Isso nos permite simular os efeitos da aplicação de uma droga que altera um (ou vários) parâmetros cinéticos que participam da regulação do gene RKIP.

Para demonstrar a utilidade de nossa abordagem, simulamos três cenários nos quais pretendíamos elevar o número de mRNAs de RKIP por aumentar: (1) a taxa de chaveamento de OFF para ON usando DETANONOato; (2) a taxa de síntese usando 5-AzaC; e (3) ambas as taxas do gene usando as duas drogas concomitantemente. Mostramos que a velocidade da resposta ao tratamento e a heterogeneidade dependiam do estado pré-tratamento do gene. E então, apresentamos um desenho de tratamento aprimorado que garantiu a redução da heterogeneidade e do tempo de resposta na dinâmica. Além de ser útil para inspecionar desenhos de tratamento, a resposta ao tratamento pode ser usada para inferência das constantes cinéticas de um determinado gene em um sistema sintético.

Numa perspectiva teórica, um próximo passo para tornar a nossa prova de conceito robusta é descrevê-la como um sistema de controle que tem como meta a otimização de agendas de tratamento com respostas efetivas dos níveis de expressão gênica e que, ao mesmo tempo, evitam a toxicidade excessiva às células saudáveis por reduzir dosagens ao longo da agenda administrada. As teorias do controle ótimo e do controle proporcional-integral-derivativo (PID) são duas possíveis aplicações para a resolução deste problema proposto que pretendemos realizar num futuro próximo, bem como a realização de experimentos *in vitro* e *in vivo* para validar o modelo de terapia e para elaborar técnicas de calibragem dos parâmetros relativos a agenda de tratamento, sendo estes passos necessários para o avanço da prova de conceito como ferramenta para desenhar tratamentos médicos.

Referências

- ABRAMOWITZ, M.; STEGUN, I. A. (Ed.). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. 10th. ed. Washington, DC, USA: U.S. Department of Commerce, National Bureau of Standards, 1972. v. 55. (National Bureau of Standards Applied Mathematics Series, v. 55). Citado 6 vezes nas páginas 26, 27, 31, 105, 109 e 116.
- AL-MULLA, F.; BITAR, M. S.; AL-MAGHREBI, M.; BEHBEHANI, A. I.; AL-ALI, W.; RATH, O.; DOYLE, B.; TAN, K. Y.; PITT, A.; KOLCH, W. Raf kinase inhibitor protein RKIP enhances signaling by glycogen synthase kinase-3 β . *Cancer Res.*, v. 71, n. 4, p. 1334–1343, 2011. Citado 2 vezes nas páginas 59 e 71.
- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; MORGAN, D.; RAFF, M.; ROBERTS, K.; WALTER, P. *Molecular Biology of the Cell*. 6th. ed. New York: Garland Science, 2014. Citado na página 14.
- ALIZADEH, A. A.; ARANDA, V.; BARDELLI, A.; BLANPAIN, C.; BOCK, C.; BOROWSKI, C.; CALDAS, C.; CALIFANO, A.; DOHERTY, M.; ELSNER, M.; ESTELLER, M.; FITZGERALD, R.; KORBEL, J. O.; LICHTER, P.; MASON, C. E.; NAVIN, N.; PE'ER, D.; POLYAK, K.; ROBERTS, C. W. M.; SIU, L.; SNYDER, A.; STOWER, H.; SWANTON, C.; VERHAAK, R. G. W.; ZENKLUSEN, J. C.; ZUBER, J.; ZUCMAN-ROSSI, J. Toward understanding and exploiting tumor heterogeneity. *Nat. Med.*, v. 21, n. 8, p. 846–853, 2015. Citado na página 53.
- ARFKEN, G. B.; WEBER, H. J. *Mathematical Methods for Physicists*. 6th. ed. Amsterdam: Elsevier, 2005. Citado 3 vezes nas páginas 27, 104 e 115.
- BALÁZSI, G.; OUDENAARDEN, A. van; COLLINS, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell*, v. 144, n. 6, p. 910–925, 2011. Citado 2 vezes nas páginas 53 e 86.
- BARITAKI, S.; HUERTA-YEPEZ, S.; SAHAKYAN, A.; KARAGIANNIDES, I.; BAKIRTZI, K.; JAZIREHI, A.; BONAVIDA, B. Mechanisms of nitric oxide-mediated inhibition of EMT in cancer. *Cell Cycle*, v. 9, n. 24, p. 4931–4940, 2010. Citado 4 vezes nas páginas 58, 59, 62 e 68.
- BEACH, S.; TANG, H.; PARK, S.; DHILLON, A. S.; KELLER, E. T.; KOLCH, W.; YEUNG, K. C. Snail is a repressor of RKIP transcription in metastatic prostate cancer cells. *Oncogene*, v. 27, n. 15, p. 2243–2248, 2007. Citado 2 vezes nas páginas 58 e 59.
- BHALLA, U. S.; IYENGAR, R. Emergent properties of networks of biological signaling pathways. *Science*, v. 283, n. 5400, p. 381–387, 1999. Citado na página 53.
- BLAKE, W. J.; KÆRN, M.; CANTOR, C. R.; COLLINS, J. J. Noise in eukaryotic gene expression. *Nature*, v. 422, p. 633–637, 2003. Citado 2 vezes nas páginas 14 e 21.
- BONAVIDA, B. RKIP-mediated chemo-immunosensitization of resistant cancer cells via disruption of the NF- κ B/Snail/YY1/RKIP resistance-driver loop. *Crit. Rev. Oncog.*, v. 19, n. 6, p. 431–445, 2014. Citado na página 62.

- BONAVIDA, B. Historical perspectives of the role of NO/NO donors in anti-tumor activities: Acknowledging Dr. Keefer's pioneering research. *Crit. Rev. Oncog.*, 2021. Citado na página 62.
- BONAVIDA, B.; BARITAKI, S. The novel role of yin yang 1 in the regulation of epithelial to mesenchymal transition in cancer via the dysregulated NF- κ B/Snail/YY1/RKIP/PTEN circuitry. *Crit. Rev. Oncog.*, v. 16, n. 3-4, p. 211–226, 2011. Citado na página 62.
- BONAVIDA, B.; BARITAKI, S. (Ed.). *Prognostic and Therapeutic Applications of RKIP in Cancer*. Cambridge, Massachusetts: Academic Press, 2020. Citado na página 55.
- BROCK, A.; KRAUSE, S.; INGBER, D. E. Control of cancer formation by intrinsic genetic noise and microenvironmental cues. *Nat. Rev. Cancer*, v. 15, n. 8, p. 499–509, 2015. Citado na página 53.
- BULAKLAK, K.; GERSBACH, C. A. The once and future gene therapy. *Nat. Commun.*, v. 11, n. 1, 2020. Citado na página 53.
- BUTKOV, E. *Mathematical Physics*. Reading, Massachusetts: Addison-Wesley, 1968. Citado 2 vezes nas páginas 27 e 104.
- CAI, W.; ZHOU, W.; HAN, Z.; LEI, J.; ZHUANG, J.; ZHU, P.; WU, X.; YUAN, W. Master regulator genes and their impact on major diseases. *PeerJ*, v. 8, p. e9952, 2020. Citado na página 53.
- CASTRO, J. de; ODEH, H. N.; FIGY, C.; YEUNG, M. L.; TRUMBLY, R.; YEUNG, K. C. Regulation of RKIP expression in breast cancer cells by miRNAs. In: *Prognostic and Therapeutic Applications of RKIP in Cancer*. Cambridge, Massachusetts: Academic Press, 2020. p. 139–146. Citado na página 59.
- CHATTERJEE, D.; BAI, Y.; WANG, Z.; BEACH, S.; MOTT, S.; ROY, R.; BRAASTAD, C.; SUN, Y.; MUKHOPADHYAY, A.; AGGARWAL, B. B.; DARNOWSKI, J.; PANTAZIS, P.; WYCHE, J.; FU, Z.; KITAGWA, Y.; KELLER, E. T.; SEDIVY, J. M.; YEUNG, K. C. RKIP sensitizes prostate and breast cancer cells to drug-induced apoptosis. *J. Biol. Chem.*, v. 279, n. 17, p. 17515–17523, 2004. Citado na página 62.
- COOPER, G. M. *The Cell: A Molecular Approach*. Sunderland, Massachusetts: Sinauer Associates Inc, 2000. Citado na página 14.
- CORBIT, K. C.; TRAKUL, N.; EVES, E. M.; DIAZ, B.; MARSHALL, M.; ROSNER, M. R. Activation of Raf-1 signaling by protein kinase C through a mechanism involving Raf kinase inhibitory protein. *J. Biol. Chem.*, v. 278, n. 15, p. 13061–13068, 2003. Citado na página 57.
- DANGI-GARIMELLA, S.; YUN, J.; EVES, E. M.; NEWMAN, M.; ERKELAND, S. J.; HAMMOND, S. M.; MINN, A. J.; ROSNER, M. R. Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. *EMBO J.*, v. 28, n. 4, p. 347–358, 2009. Citado na página 55.
- DATAR, I.; TEGEGNE, H.; QIN, K.; AL-MULLA, F.; BITAR, M. S.; TRUMBLY, R. J.; YEUNG, K. C. Genetic and epigenetic control of RKIP transcription. *Crit. Rev. Oncog.*, v. 19, n. 6, p. 417–430, 2014. Citado na página 58.

DELBRÜCK, M. Statistical fluctuations in autocatalytic reactions. *J. Chem. Phys.*, v. 8, n. 1, p. 120–124, 1940. Citado 2 vezes nas páginas 14 e 53.

DENG, X.; LIU, Z.; LI, X.; ZHOU, Y.; HU, Z. Generation of new hair cells by DNA methyltransferase (Dnmt) inhibitor 5-azacytidine in a chemically-deafened mouse model. *Sci. Rep.*, v. 9, n. 1, 2019. Citado na página 86.

DU, C.; WONG, W. H. Stochastic modeling and statistical inference of intrinsic noise in gene regulation system via chemical master equation. *arXiv*, 2016. Disponível em: <<https://arxiv.org/abs/1610.07213>>. Citado na página 25.

DU, Y.; WENG, X.-D.; WANG, L.; LIU, X.-H.; ZHU, H.-C.; GUO, J.; NING, J.-Z.; XIAO, C.-C. LncRNA XIST acts as a tumor suppressor in prostate cancer through sponging miR-23a to modulate RKIP expression. *Oncotarget.*, v. 8, n. 55, p. 94358–94370, 2017. Citado na página 59.

DU, Y.; ZHU, H.-C.; LIU, X.-H.; WANG, L.; NING, J.-Z.; XIAO, C.-C. MiR-543 promotes proliferation and epithelial-mesenchymal transition in prostate cancer via targeting RKIP. *Cell. Physiol. Biochem.*, v. 41, n. 3, p. 1135–1146, 2017. Citado na página 59.

ELOWITZ, M. B.; LEVINE, A. J.; SIGGIA, E. D.; SWAIN, P. S. Stochastic gene expression in a single cell. *Science*, v. 297, n. 5584, p. 1183–1186, 2002. Citado 2 vezes nas páginas 14 e 21.

FERRELL, J. E. Perfect and near-perfect adaptation in cell signaling. *Cell Systems*, v. 2, n. 2, p. 62–67, 2016. Citado 2 vezes nas páginas 53 e 86.

FU, Z.; KITAGAWA, Y.; SHEN, R.; SHAH, R.; MEHRA, R.; RHODES, D.; KELLER, P. J.; MIZOKAMI, A.; DUNN, R.; CHINNAIYAN, A. M.; YAO, Z.; KELLER, E. T. Metastasis suppressor gene raf kinase inhibitor protein (RKIP) is a novel prognostic marker in prostate cancer. *The Prostate*, v. 66, n. 3, p. 248–256, 2006. Citado na página 59.

FU, Z.; SMITH, P. C.; ZHANG, L.; RUBIN, M. A.; DUNN, R. L.; YAO, Z.; KELLER, E. T. Effects of raf kinase inhibitor protein expression on suppression of prostate cancer metastasis. *J. Natl. Cancer Inst.*, v. 95, n. 12, p. 878–889, 2003. Citado na página 55.

FUKUMURA, D.; KASHIWAGI, S.; JAIN, R. K. The role of nitric oxide in tumour progression. *Nat. Rev. Cancer*, v. 6, n. 7, p. 521–534, 2006. Citado na página 62.

GALAL, Y.; ZARAVINOS, A.; BONAVIDA, B. Regulation of NKG2D by RKIP: Implications on NK-mediated cytotoxicity and cytokine production. In: *Successes and Challenges of NK Immunotherapy*. Cambridge, Massachusetts: American Press, 2021. p. 233–265. Citado na página 58.

GAMA, L. R.; GIOVANINI, G.; BALÁZSI, G.; RAMOS, A. F. Binary expression enhances reliability of messaging in gene networks. *Entropy*, v. 22, n. 4, p. 479, 2020. Citado 4 vezes nas páginas 15, 34, 56 e 87.

GARDINER, C. W. *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Berlin: Springer Verlag, 1983. Citado na página 17.

GÉRARD, C.; GONZE, D.; GOLDBETER, A. Revisiting a skeleton model for the mammalian cell cycle: From bistability to cdk oscillations and cellular heterogeneity. *J. Theor. Biol.*, v. 461, p. 276–290, 2019. Citado 2 vezes nas páginas 53 e 86.

GIOVANINI, G.; BARROS, L. R. C.; GAMA, L. R.; TORTELLI JR, T. C.; RAMOS, A. F. A stochastic binary model for regulation of gene expression to investigate treatment effects targeting RKIP, *Preprints*, 2021. Disponível em: <<https://www.preprints.org/manuscript/202109.0403/v1>>. Citado 11 vezes nas páginas 15, 16, 63, 72, 74, 76, 77, 78, 80, 82 e 117.

GIOVANINI, G.; SABINO, A. U.; BARROS, L. R. C.; RAMOS, A. F. A comparative analysis of noise properties of stochastic binary models for a self-repressing and for an externally regulating gene. *Math. Biosci. Eng.*, v. 17, n. 5, p. 5477–5503, 2020. Citado 14 vezes nas páginas 21, 33, 34, 35, 40, 42, 43, 44, 45, 46, 54, 56, 71 e 84.

GIOVANNETTI, E.; LABOTS, M.; DEKKER, H.; GALVANI, E.; LIND, J.; SCIARRILLO, R.; HONEYWELL, R.; SMIT, E.; VERHEUL, H.; PETERS, G. Molecular mechanisms and modulation of key pathways underlying the synergistic interaction of sorafenib with erlotinib in non-small-cell-lung cancer (NSCLC) cells. *Curr. Pharm. Des.*, v. 19, n. 5, p. 927–939, 2012. Citado na página 62.

GUINN, M. T.; WAN, Y.; LEVOVITZ, S.; YANG, D.; ROSNER, M. R.; BALÁZSI, G. Observation and control of gene expression noise: Barrier crossing analogies between drug resistance and metastasis. *Front. Genet.*, v. 11, 2020. Citado na página 53.

GUO, W.; DONG, Z.; LIN, X.; ZHANG, M.; KUANG, G.; ZHU, T. Decreased expression and aberrant methylation of raf kinase inhibitory protein gene in esophageal squamous cell carcinoma. *Cancer Invest.*, v. 30, n. 10, p. 703–711, 2012. Citado 2 vezes nas páginas 61 e 68.

HAMPL, V.; TRISTANI-FIROUZI, M.; HUTSELL, T. C.; ARCHER, S. L. Nebulized nitric oxide/nucleophile adduct reduces chronic pulmonary hypertension. *Cardiovasc. Res.*, v. 31, n. 1, p. 55–62, 1996. Citado na página 68.

HAO, C.; WEI, S.; TONG, Z.; LI, S.; SHI, Y.; WANG, X.; ZHU, Z. hua. The effects of RKIP gene expression on the biological characteristics of human triple-negative breast cancer cells in vitro. *Tumour Biol.*, v. 33, n. 4, p. 1159–1167, 2012. Citado na página 71.

HAYS, E.; BONAVIDA, B. Nitric oxide-mediated enhancement and reversal of resistance of anticancer therapies. *Antioxidants*, v. 8, n. 9, p. 407, 2019. Citado na página 87.

HORNOS, J. E. M.; SCHULTZ, D.; INNOCENTINI, G. C. P.; WANG, J.; WALCZAK, A. M.; ONUCHIC, J. N.; WOLYNES, P. G. Self-regulating gene: an exact solution. *Phys. Rev. E*, v. 72, p. 051907, 2005. Citado 5 vezes nas páginas 15, 24, 30, 50 e 51.

HOU, W.; SHAN, Y.; ZHENG, J.; LAMBRECHT, R. W.; DONOHUE, S. E.; BONKOVSKY, H. L. Zinc mesoporphyrin induces rapid and marked degradation of the transcription factor bach1 and up-regulates HO-1. *Biochim. Biophys. Acta, Gene Regul. Mech.*, v. 1779, n. 3, p. 195–203, 2008. Citado na página 81.

HUERTA-YEPEZ, S.; VEGA, M.; JAZIREHI, A.; GARBAN, H.; HONGO, F.; CHENG, G.; BONAVIDA, B. Nitric oxide sensitizes prostate carcinoma cell lines to

- TRAIL-mediated apoptosis via inactivation of NF- κ b and inhibition of Bcl-xL expression. *Oncogene*, v. 23, n. 29, p. 4993–5003, 2004. Citado na página 62.
- INNOCENTINI, G. C. P.; HORNOS, J. E. M. Modeling stochastic gene expression under repression. *J. Math. Biol.*, v. 55, n. 3, p. 413–431, 2007. Citado 4 vezes nas páginas 27, 30, 31 e 116.
- IYER-BISWAS, S.; HAYOT, F.; JAYAPRAKASH, C. Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E*, v. 79, n. 3, p. 031911, 2009. Citado 4 vezes nas páginas 27, 38, 54 e 66.
- JARRETT, A. M.; FAGHIHI, D.; HORMUTH, D. A.; LIMA, E. A. B. F.; VIROSTKO, J.; BIROS, G.; PATT, D.; YANKEELOV, T. E. Optimal control theory for personalized therapeutic regimens in oncology: Background, history, challenges, and opportunities. *J. Clin. Med.*, v. 9, n. 5, p. 1314, 2020. Citado na página 55.
- KAMPEN van. *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland, 1992. Citado na página 17.
- KIM, G.-E.; KIM, N. I.; LEE, J. S.; PARK, M. H.; YOON, J. H. Reduced RKIP expression is associated with breast neoplastic progression and is correlated with poor outcomes and aberrant methylation in breast carcinoma. *Appl. Immunohistochem. Mol. Morphol.*, v. 25, n. 7, p. 467–474, 2017. Citado na página 61.
- LABBOZZETTA, M.; POMA, P.; VIVONA, N.; GULINO, A.; D’ALESSANDRO, N.; NOTARBARTOLO, M. Epigenetic changes and nuclear factor- κ B activation, but not microRNA-224, downregulate Raf-1 kinase inhibitor protein in triple-negative breast cancer SUM 159 cells. *Oncol. Lett.*, v. 10, n. 6, p. 3807–3815, 2015. Citado 4 vezes nas páginas 58, 59, 61 e 68.
- LAMIMAN, K.; KELLER, J. M.; MIZOKAMI, A.; ZHANG, J.; KELLER, E. T. Survey of raf kinase inhibitor protein (RKIP) in multiple cancer types. *Crit. Rev. Oncog.*, v. 19, n. 6, p. 455–468, 2014. Citado na página 55.
- LEE, J.; LEE, J.; FARQUHAR, K. S.; YUN, J.; FRANKENBERGER, C. A.; BEVILACQUA, E.; YEUNG, K.; KIM, E.-J.; BALÁZSI, G.; ROSNER, M. R. Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proc. Natl. Acad. Sci. U.S.A.*, v. 111, n. 3, p. E364–E373, 2014. Citado 2 vezes nas páginas 58 e 59.
- LEE, T.-Y.; LIU, C.-L.; CHANG, Y.-C.; NIEH, S.; LIN, Y.-S.; JAO, S.-W.; CHEN, S.-F.; LIU, T.-Y. Increased chemoresistance via Snail-Raf kinase inhibitor protein signaling in colorectal cancer in response to a nicotine derivative. *Oncotarget*, v. 7, n. 17, p. 23512–23520, 2016. Citado na página 71.
- LEI, X.; CHANG, L.; YE, W.; JIANG, C.; ZHANG, Z. Raf kinase inhibitor protein (RKIP) inhibits the cell migration and invasion in human glioma cell lines in vitro. *Int. J. Clin. Exp. Pathol.*, v. 8, n. 11, p. 14214–14220, 2015. Citado na página 70.
- LEVINE, J. H.; LIN, Y.; ELOWITZ, M. B. Functional roles of pulsing in genetic circuits. *Science*, v. 342, n. 6163, p. 1193–1200, 2013. Citado na página 54.

LI, D.-X.; CAI, H.-Y.; WANG, X.; FENG, Y.-L.; CAI, S.-W. Promoter methylation of raf kinase inhibitory protein: A significant prognostic indicator for patients with gastric adenocarcinoma. *Exp. Ther. Med.*, v. 8, n. 3, p. 844–850, 2014. Citado 3 vezes nas páginas 58, 59 e 61.

LIN, K.; BARITAKI, S.; MILITELLO, L.; MALAPONTE, G.; BEVELACQUA, Y.; BONAVIDA, B. The role of B-RAF mutations in melanoma and the induction of EMT via dysregulation of the NF- κ B/Snail/RKIP/PTEN circuit. *Genes Cancer*, v. 1, n. 5, p. 409–420, 2010. Citado na página 71.

LORENZ, K.; LOHSE, M. J.; QUITTERER, U. Protein kinase C switches the Raf kinase inhibitor from Raf-1 to GRK-2. *Nature*, v. 426, n. 6966, p. 574–579, 2003. Citado na página 57.

MARTINHO, O.; GRANJA, S.; JARAQUEMADA, T.; CAEIRO, C.; MIRANDA-GONÇALVES, V.; HONAVAR, M.; COSTA, P.; DAMASCENO, M.; ROSNER, M. R.; LOPES, J. M.; REIS, R. M. Downregulation of RKIP is associated with poor outcome and malignant progression in gliomas. *PLoS One*, v. 7, n. 1, p. e30769, 2012. Citado na página 55.

MARTINHO, O.; PINTO, F.; GRANJA, S.; MIRANDA-GONÇALVES, V.; MOREIRA, M. A. R.; RIBEIRO, L. F. J.; LORETO, C. di; ROSNER, M. R.; LONGATTO-FILHO, A.; REIS, R. M. RKIP inhibition in cervical cancer is associated with higher tumor aggressive behavior and resistance to cisplatin therapy. *PLoS One*, v. 8, n. 3, p. e59104, 2013. Citado na página 55.

MARUSYK, A.; ALMENDRO, V.; POLYAK, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, v. 12, n. 5, p. 323–334, 2012. Citado na página 53.

MICHOR, F.; BEAL, K. Improving cancer treatment via mathematical modeling: Surmounting the challenges is worth the effort. *Cell*, v. 163, n. 5, p. 1059–1063, 2015. Citado na página 53.

MINOO, P.; BAKER, K.; GOSWAMI, R.; CHONG, G.; FOULKES, W. D.; RUSZKIEWICZ, A. R.; BARKER, M.; BUCHANAN, D.; YOUNG, J.; JASS, J. R. Extensive DNA methylation in normal colorectal mucosa in hyperplastic polyposis. *Gut*, v. 55, n. 10, p. 1467–1474, 2006. Citado na página 61.

MINOO, P.; ZLOBEC, I.; BAKER, K.; TORNILLO, L.; TERRACCIANO, L.; JASS, J. R.; LUGLI, A. Loss of Raf-1 kinase inhibitor protein expression is associated with tumor progression and metastasis in colorectal cancer. *Am. J. Clin. Pathol.*, v. 127, n. 5, p. 820–827, 2007. Citado na página 71.

MOROZOVA, N.; ZINOVYEV, A.; NONNE, N.; PRITCHARD, L.-L.; GORBAN, A. N.; HAREL-BELLAN, A. Kinetic signatures of microRNA modes of action. *RNA*, v. 18, n. 9, p. 1635–1655, 2012. Citado na página 81.

MUNSKY, B.; NEUERT, G.; OUDENAARDEN, A. van. Using gene expression noise to understand gene regulation. *Science*, v. 336, n. 6078, p. 183–187, 2012. Citado na página 15.

- PASQUALETTI, G.; RICCIARDI, S.; MEY, V.; TACCA, M. D.; DANESI, R. Synergistic cytotoxicity, inhibition of signal transduction pathways and pharmacogenetics of sorafenib and gemcitabine in human NSCLC cell lines. *Lung Cancer*, v. 74, n. 2, p. 197–205, 2011. Citado na página 62.
- PECCOUD, J.; YCART, B. Markovian modelling of gene product synthesis. *Theor. Popul. Biol.*, v. 48, n. 2, p. 222–234, 1995. Citado 6 vezes nas páginas 15, 24, 30, 51, 54 e 66.
- PERVIN, S.; SINGH, R.; HERNANDEZ, E.; WU, G.; CHAUDHURI, G. Nitric oxide in physiologic concentrations targets the translational machinery to increase the proliferation of human breast cancer cells: Involvement of mammalian target of rapamycin/eIF4E pathway. *Cancer Res.*, v. 67, n. 1, p. 289–299, 2007. Citado na página 62.
- PERVIN, S.; TRAN, L.; URMAN, R.; BRAGA, M.; PARVEEN, M.; LI, S. A.; CHAUDHURI, G.; SINGH, R. Oxidative stress specifically downregulates survivin to promote breast tumour formation. *Br. J. Cancer*, v. 108, n. 4, p. 848–858, 2013. Citado na página 62.
- PRATA, G. N.; HORNOS, J. E. M.; RAMOS, A. F. Stochastic model for gene transcription on *Drosophila melanogaster* embryos. *Phys. Rev. E*, v. 93, n. 2, p. 022403, 2016. Citado 2 vezes nas páginas 16 e 25.
- PURVIS, J. E.; LAHAV, G. Encoding and decoding cellular information through signaling dynamics. *Cell*, v. 152, n. 5, p. 945–956, 2013. Citado na página 54.
- RAJ, A.; PESKIN, C. S.; TRANCHINA, D.; VARGAS, D. Y.; TYAGI, S. Stochastic mRNA synthesis in mammalian cells. *PLOS Biology*, v. 4, p. 1707–1719, 2006. Citado 2 vezes nas páginas 14 e 21.
- RAMOS, A. F.; GAMA, L. R.; MORAIS, M. C. C.; MARTINS, P. C. M. Stochastic modeling for investigation of the regulation of transcription of the RKIP gene. In: BONAVIDA, B.; BARITAKI, S. (Ed.). *Prognostic and Therapeutic Applications of RKIP in Cancer*. Cambridge, Massachusetts: Academic Press, 2020. p. 257–276. Citado 7 vezes nas páginas 15, 16, 25, 54, 61, 62 e 66.
- RAMOS, A. F.; HORNOS, J. E. M. Symmetry and stochastic gene regulation. *Phys. Rev. Lett.*, v. 99, n. 10, p. 108103, 2007. Citado 4 vezes nas páginas 15, 26, 38 e 41.
- RAMOS, A. F.; HORNOS, J. E. M.; REINITZ, J. Gene regulation and noise reduction by coupling of stochastic processes. *Phys. Rev. E*, v. 91, p. 020701(R), 2015. Citado 8 vezes nas páginas 15, 21, 22, 23, 29, 39, 40 e 46.
- RAMOS, A. F.; INNOCENTINI, G. C. P.; FORGER, F. M.; HORNOS, J. E. M. Symmetry in biology: from genetic code to stochastic gene regulation. *IET Syst. Biol.*, v. 4, n. 5, p. 311–329, 2010. Citado 7 vezes nas páginas 15, 30, 38, 51, 54, 86 e 116.
- RAMOS, A. F.; INNOCENTINI, G. C. P.; HORNOS, J. E. M. Exact time-dependent solutions for a self-regulating gene. *Phys. Rev. E*, v. 83, n. 6, p. 062902, 2011. Citado 2 vezes nas páginas 15 e 38.
- RAMOS, A. F.; REINITZ, J. Physical implications of $so(2, 1)$ symmetry in exact solutions for a self-repressing gene. *J. Chem. Phys.*, v. 151, n. 4, p. 041101, 2019. Citado 9 vezes nas páginas 15, 37, 38, 41, 44, 47, 49, 56 e 87.

RAPOZZI, V.; PIETRA, E. D.; BONAVIDA, B. Dual roles of nitric oxide in the regulation of tumor cell response and resistance to photodynamic therapy. *Redox Biol.*, v. 6, p. 311–317, 2015. Citado na página 62.

RAPOZZI, V.; PIETRA, E. D.; ZORZET, S.; ZACCHIGNA, M.; BONAVIDA, B.; XODO, L. E. Nitric oxide-mediated activity in anti-cancer photodynamic therapy. *Nitric Oxide*, v. 30, p. 26–35, 2013. Citado na página 62.

REICHL, L. E. *A Modern Course in Statistical Physics*. Weinheim, Germany: Wiley-VCH GmbH, 2016. Citado na página 17.

REN, G.; BARITAKI, S.; MARATHE, H.; FENG, J.; PARK, S.; BEACH, S.; BAZELEY, P. S.; BESHIR, A. B.; FENTEANY, G.; MEHRA, R.; DAIGNAULT, S.; AL-MULLA, F.; KELLER, E.; BONAVIDA, B.; SERNA, I. de la; YEUNG, K. C. Polycomb protein EZH2 regulates tumor invasion via the transcriptional repression of the metastasis suppressor RKIP in breast and prostate cancer. *Cancer Res.*, v. 72, n. 12, p. 3091–3104, 2012. Citado 2 vezes nas páginas 58 e 59.

RICCIARDI, S.; MEY, V.; NANNIZZI, S.; PASQUALETTI, G.; CREA, F.; TACCA, M. D.; DANESI, R. Synergistic cytotoxicity and molecular interaction on drug targets of sorafenib and gemcitabine in human pancreas cancer cells. *Chemotherapy*, v. 56, n. 4, p. 303–312, 2010. Citado na página 62.

ROSENFELD, N.; PERKINS, T. J.; ALON, U.; ELOWITZ, M. B.; SWAIN, P. S. A fluctuation method to quantify in vivo fluorescence data. *Biophys. J.*, v. 91, p. 759–766, 2006. Citado na página 22.

SCHUIERER, M. M.; BATAILLE, F.; HAGAN, S.; KOLCH, W.; BOSSERHOFF, A.-K. Reduction in Raf kinase inhibitor protein expression is associated with increased Ras-extracellular signal-regulated kinase signaling in melanoma cell lines. *Cancer Res.*, v. 64, n. 15, p. 5186–5192, 2004. Citado na página 71.

SHAHREZAEI, V.; SWAIN, P. S. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, v. 105, n. 45, p. 17256–17261, 2008. Citado 2 vezes nas páginas 43 e 86.

SHVARTSUR, A.; GIVECHIAN, K. B.; GARBAN, H.; BONAVIDA, B. Overexpression of RKIP and its cross-talk with several regulatory gene products in multiple myeloma. *J. Exp. Clin. Cancer Res.*, v. 36, n. 1, 2017. Citado na página 57.

SUTER, D. M.; MOLINA, N.; GATFIELD, D.; SCHNEIDER, K.; SCHIBLER, U.; NAEF, F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*, v. 332, n. 6028, p. 472–474, 2011. Citado na página 21.

THAKORE, P. I.; BLACK, J. B.; HILTON, I. B.; GERSBACH, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods*, v. 13, n. 2, p. 127–137, 2016. Citado na página 53.

TOMÉ, T.; OLIVEIRA, M. J. *Dinâmica Estocástica e Irreversibilidade*. São Paulo: EDUSP, 2014. Citado na página 17.

TOMSKI, A.; ZAKARCZEMNY, M. Stochastic gene expression revisited. *Genes*, v. 12, n. 5, p. 648, 2021. Citado na página 25.

TRAKUL, N.; MENARD, R. E.; SCHADE, G. R.; QIAN, Z.; ROSNER, M. R. Raf kinase inhibitory protein regulates Raf-1 but not B-raf kinase activation. *J. Biol. Chem.*, v. 280, n. 26, p. 24931–24940, 2005. Citado na página 57.

WALKER, E. J.; ROSENBERG, S. A.; WANDS, J. R.; KIM, M. Role of Raf kinase inhibitor protein in hepatocellular carcinoma. *For. Immunopathol. Dis. Therap.*, v. 2, n. 2, p. 195–204, 2011. Citado 2 vezes nas páginas 69 e 86.

WEI, H.; LIU, Z.; SHE, H.; LIU, B.; GU, J.; WEI, D.; ZHANG, X.; WANG, J.; QI, S.; PING, F. Promoter methylation and expression of Raf kinase inhibitory protein in esophageal squamous cell carcinoma. *Oncol. Lett.*, v. 13, n. 3, p. 1866–1872, 2017. Citado 3 vezes nas páginas 58, 59 e 61.

WELCH, D. R. Tumor heterogeneity—a ‘contemporary concept’ founded on historical insights and predictions. *Cancer Res.*, v. 76, n. 1, p. 4–6, 2016. Citado na página 53.

WINK, D. A.; RIDNOUR, L. A.; HUSSAIN, S. P.; HARRIS, C. C. The reemergence of nitric oxide and cancer. *Nitric Oxide*, v. 19, n. 2, p. 65–67, 2008. Citado na página 62.

WISHART, D. S.; FEUNANG, Y. D.; GUO, A. C.; LO, E. J.; MARCU, A.; GRANT, J. R.; SAJED, T.; JOHNSON, D.; LI, C.; SAYEEDA, Z.; ASSEMPOUR, N.; IYNNKARAN, I.; LIU, Y.; MACIEJEWSKI, A.; GALE, N.; WILSON, A.; CHIN, L.; CUMMINGS, R.; LE, D.; PON, A.; KNOX, C.; WILSON, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, v. 46, n. D1, p. D1074–D1082, 2017. Citado na página 68.

WU, X.; YANG, Y.; XU, Z.; LI, J.; YANG, B.; FENG, N.; ZHANG, Y.; WANG, S. Raf kinase inhibitor protein mediated signaling inhibits invasion and metastasis of hepatocellular carcinoma. *Biochim. Biophys. Acta*, v. 1860, n. 2, p. 384–391, 2016. Citado na página 70.

YAN, X.; EHNERT, S.; CULMES, M.; BACHMANN, A.; SEELIGER, C.; SCHYSCHKA, L.; WANG, Z.; RAHMANIAN-SCHWARZ, A.; STÖCKLE, U.; SOUSA, P. A. D.; PELISEK, J.; NUSSLER, A. K. 5-azacytidine improves the osteogenic differentiation potential of aged human adipose-derived mesenchymal stem cells by DNA demethylation. *PLoS One*, v. 9, n. 3, p. e90846, 2014. Citado na página 86.

YESILKANAL, A.; ROSNER, M. Targeting Raf kinase inhibitory protein regulation and function. *Cancers*, v. 10, n. 9, p. 306, 2018. Citado 3 vezes nas páginas 16, 55 e 58.

YESILKANAL, A. E.; JOHNSON, G. L.; RAMOS, A. F.; ROSNER, M. R. New strategies for targeting kinase networks in cancer. *J. Biol. Chem.*, p. 101128, 2021. Citado 2 vezes nas páginas 53 e 57.

YESILKANAL, A. E.; ROSNER, M. R. Raf kinase inhibitory protein (RKIP) as a metastasis suppressor: Regulation of signaling networks in cancer. *Crit. Rev. Oncog.*, v. 19, n. 6, p. 447–454, 2014. Citado na página 57.

YESILKANAL, A. E.; YANG, D.; VALDESPINO, A.; TIWARI, P.; SABINO, A. U.; NGUYEN, L. C.; LEE, J.; XIE, X.-H.; SUN, S.; DANN, C.; ROBINSON-MAILMAN, L.; STEINBERG, E.; STUHLMILLER, T.; FRANKENBERGER, C.; GOLDSMITH, E.; JOHNSON, G. L.; RAMOS, A. F.; ROSNER, M. R. Limited inhibition of multiple nodes

in a driver network blocks metastasis. *eLife*, Ltd, v. 10, 2021. Citado 2 vezes nas páginas 53 e 57.

YEUNG, K.; SEITZ, T.; LI, S.; JANOSCH, P.; MCFERRAN, B.; KAISER, C.; FEE, F.; KATSANAKIS, K. D.; ROSE, D. W.; MISCHAK, H.; SEDIVY, J. M.; KOLCH, W. Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature*, v. 401, n. 6749, p. 173–177, 1999. Citado na página 57.

YEUNG, K. C.; ROSE, D. W.; DHILLON, A. S.; YAROS, D.; GUSTAFSSON, M.; CHATTERJEE, D.; MCFERRAN, B.; WYCHE, J.; KOLCH, W.; SEDIVY, J. M. Raf kinase inhibitor protein interacts with NF- κ B-inducing kinase and TAK1 and inhibits NF- κ B activation. *Mol. Cell. Biol.*, v. 21, n. 21, p. 7207–7217, 2001. Citado 2 vezes nas páginas 57 e 58.

YOUSUF, S.; DUAN, M.; MOEN, E. L.; CROSS-KNORR, S.; BRILLIANT, K.; BONAVIDA, B.; LAVALLE, T.; YEUNG, K. C.; AL-MULLA, F.; CHIN, E.; CHATTERJEE, D. Raf kinase inhibitor protein (RKIP) blocks signal transducer and activator of transcription 3 (STAT3) activation in breast and prostate cancer. *PLoS One*, v. 9, n. 3, p. e92478, 2014. Citado na página 62.

YUN, J.; FRANKENBERGER, C. A.; KUO, W.-L.; BOELEN, M. C.; EVES, E. M.; CHENG, N.; LIANG, H.; LI, W.-H.; ISHWARAN, H.; MINN, A. J.; ROSNER, M. R. Signalling pathway for RKIP and Let-7 regulates and predicts metastatic breast cancer. *EMBO J.*, v. 30, n. 21, p. 4500–4514, 2011. Citado 2 vezes nas páginas 58 e 59.

YVINEC, R.; SILVA, L. G. S.; PRATA, G. N.; REINITZ, J.; RAMOS, A. F. Bursting on a two state stochastic model for gene transcription in drosophila embryos. *bioRxiv*, 2017. Disponível em: <<https://www.biorxiv.org/content/early/2017/02/13/107979>>. Citado 5 vezes nas páginas 34, 43, 44, 50 e 51.

ZARAVINOS, A.; BONAVIDA, B.; CHATZAKI, E.; BARITAKI, S. RKIP: A key regulator in tumor metastasis initiation and resistance to apoptosis: Therapeutic targeting and impact. *Cancers*, v. 10, n. 9, p. 287, 2018. Citado 7 vezes nas páginas 55, 57, 58, 59, 62, 70 e 85.

ZHAO, D.; MA, J.; SHI, J.; CHENG, L.; LI, F.; JIANG, X.; JIANG, H. Raf kinase inhibitor protein inhibits esophageal cancer cell invasion through downregulation of matrix metalloproteinase expression. *Oncol. Rep.*, v. 30, n. 1, p. 304–312, 2013. Citado na página 70.

ZHAO, J.; WENZEL, S. Interactions of RKIP with inflammatory signaling pathways. *Crit. Rev. Oncog.*, v. 19, n. 6, p. 497–504, 2014. Citado na página 57.

ZOU, Q.; WU, H.; FU, F.; YI, W.; PEI, L.; ZHOU, M. RKIP suppresses the proliferation and metastasis of breast cancer cell lines through up-regulation of miR-185 targeting HMGA2. *Arch. Biochem. Biophys.*, v. 610, p. 25–32, 2016. Citado na página 71.

Apêndices

Apêndice A – Dedução da Eq. (3.17)

As EDPs (3.17) e (3.18) são obtidas ao substituir o par de Eqs. (3.19) nas Eqs. (3.15) e (3.16). Segue abaixo as etapas do desenvolvimento algébrico para obtê-las.

Ao multiplicar todos os termos da equação mestra por z^n e, juntamente, aplicar o somatório em n para todos os valores possíveis do número de proteínas, o par de Eqs. (3.15) e (3.16) se tornam

$$\begin{aligned} \frac{d}{dt} \sum_{n=0}^{\infty} z^n \alpha_n = k \left[\sum_{n=0}^{\infty} z^n \alpha_{n-1} - \sum_{n=0}^{\infty} z^n \alpha_n \right] + \rho \left[\sum_{n=0}^{\infty} (n+1) z^n \alpha_{n+1} - \sum_{n=0}^{\infty} n z^n \alpha_n \right] \\ - \sum_{n=0}^{\infty} (h_1 + h_2 n) z^n \alpha_n + f \sum_{n=0}^{\infty} z^n \beta_n, \end{aligned} \quad (\text{A.1})$$

$$\frac{d}{dt} \sum_{n=0}^{\infty} z^n \beta_n = \rho \left[\sum_{n=0}^{\infty} (n+1) z^n \beta_{n+1} - \sum_{n=0}^{\infty} n z^n \beta_n \right] + \sum_{n=0}^{\infty} (h_1 + h_2 n) z^n \alpha_n - f \sum_{n=0}^{\infty} z^n \beta_n. \quad (\text{A.2})$$

Note que para evitar uma notação carregada, omitimos a variável t , propriamente, $\alpha_n := \alpha_n(t)$ e $\beta_n := \beta_n(t)$. Os somatórios das Eqs. (A.1) e (A.2) podem ser reescritos como

$$\sum_{n=0}^{\infty} z^n \alpha_{n-1} = z \sum_{n=0}^{\infty} z^n \alpha_n = z \alpha(z, t),$$

$$\sum_{n=0}^{\infty} (n+1) z^n \alpha_{n+1} = \frac{\partial}{\partial z} \sum_{n=0}^{\infty} z^n \alpha_n = \frac{\partial \alpha(z, t)}{\partial z}, \quad (\text{A.3})$$

$$\sum_{n=0}^{\infty} n z^n \alpha_n = z \frac{\partial}{\partial z} \sum_{n=0}^{\infty} z^n \alpha_n = z \frac{\partial \alpha(z, t)}{\partial z},$$

sendo esses resultados igualmente válidos ao substituir α_n e $\alpha(z, t)$ por β_n e $\beta(z, t)$, respectivamente. As EDPs do modelo do gene estocástico binário são obtidas ao substituir as Eqs. (3.19) e (A.3) nas Eqs. (A.1) e (A.2). As substituições resultam nas Eqs. (3.17) e (3.18), o par de EDPs acopladas que, uma vez resolvido, nos fornece as soluções exatas das distribuições de probabilidade do modelo.

Apêndice B – Dedução da equação de Kummer - Eq. (3.20)

Considerando as soluções $\tilde{\alpha}$ e $\tilde{\beta}$ no regime estacionário do sistema de EDPs acopladas (3.17) e (3.18), temos:

$$0 = (z - 1) \left(k\tilde{\alpha} - \rho \frac{d\tilde{\alpha}}{dz} \right) - h_2 z \frac{d\tilde{\alpha}}{dz} - h_1 \tilde{\alpha} + f\tilde{\beta}, \quad (\text{B.1})$$

$$0 = -\rho(z - 1) \frac{d\tilde{\beta}}{dz} + h_2 z \frac{d\tilde{\alpha}}{dz} + h_1 \tilde{\alpha} - f\tilde{\beta}, \quad (\text{B.2})$$

e ao isolar $\tilde{\beta}$ na Eq. (B.1) e derivar essa mesma equação em relação a z , são obtidas:

$$\tilde{\beta} = \frac{1}{f} \left\{ [(\rho + h_2)z - \rho] \frac{d\tilde{\alpha}}{dz} + [k(1 - z) + h_1] \tilde{\alpha} \right\}, \quad (\text{B.3})$$

$$\frac{d\tilde{\beta}}{dz} = \frac{1}{f} \left\{ [(\rho + h_2)z - \rho] \frac{d^2\tilde{\alpha}}{dz^2} + [k(1 - z) + \rho + h_1 + h_2] \frac{d\tilde{\alpha}}{dz} - k\tilde{\alpha} \right\}.$$

Agora, substituindo o par de Eqs. (B.3) na Eq. (B.2) e fatorando os coeficientes de acordo com a ordem das derivadas, temos:

$$\begin{aligned} -\frac{\rho}{f}(z - 1) [(\rho + h_2)z - \rho] \frac{d^2\tilde{\alpha}}{dz^2} + \\ \left\{ -\frac{\rho}{f}(z - 1) [-k(z - 1) + \rho + h_1 + h_2] - \rho(z - 1) \right\} \frac{d\tilde{\alpha}}{dz} + \\ \left[k\frac{\rho}{f}(z - 1) + k(z - 1) \right] \tilde{\alpha} = 0, \end{aligned} \quad (\text{B.4})$$

que é uma equação que não depende de $\tilde{\beta}$. Ao fatorar $-\frac{\rho}{f}(z - 1)(\rho + h_2)$, e desprezando a solução trivial $z = 1$, obtemos:

$$\left[z - \frac{\rho}{\rho + h_2} \right] \frac{d^2\tilde{\alpha}}{dz^2} + \left[1 + \frac{f + h_1}{\rho + h_2} - \frac{k}{\rho + h_2}(z - 1) \right] \frac{d\tilde{\alpha}}{dz} - \left[\frac{k}{\rho + h_2} \left(1 + \frac{f}{\rho} \right) \right] \tilde{\alpha} = 0. \quad (\text{B.5})$$

Convenientemente, definimos as constantes auxiliares $N = \frac{k}{\rho}$, $z_0 = \frac{\rho}{\rho + h_2}$, que são respectivamente, o número médio de proteínas sintetizadas pelo gene constitutivo no estado estacionário e a taxa de degradação de proteínas proporcional às proteínas que tornam-se ausentes no citoplasma. Também aplicamos a manipulação algébrica $(z - 1) = (z - z_0) - (1 - z_0)$ no segundo termo da Eq. (B.5), que ao reescrevê-la torna-se:

$$[z - z_0] \frac{d^2\tilde{\alpha}}{dz^2} + \left[1 + \frac{f + h_1}{\rho} z_0 + N z_0 (1 - z_0) - N z_0 (z - z_0) \right] \frac{d\tilde{\alpha}}{dz} - \left[N z_0 \left(1 + \frac{f}{\rho} \right) \right] \tilde{\alpha} = 0. \quad (\text{B.6})$$

Os parâmetros B e η da função KummerM são dados pela Eq. (3.21) e ao serem reescritos em função de z_0 ficam:

$$B = 1 + \frac{f + h_1}{\rho} z_0 + Nz_0(1 - z_0), \quad \eta(z) = Nz_0(z - z_0), \quad (\text{B.7})$$

e ao substituir na Eq. (B.6) obtemos:

$$\frac{\eta(z)}{Nz_0} \frac{d^2 \tilde{\alpha}}{dz^2} + [B - \eta(z)] \frac{d\tilde{\alpha}}{dz} - Nz_0 A \tilde{\alpha} = 0. \quad (\text{B.8})$$

Ao dividir todos os termos pelo coeficiente da derivada de segunda ordem, obtemos

$$\frac{d^2 \tilde{\alpha}}{dz^2} + Nz_0 \frac{B - \eta(z)}{\eta(z)} \frac{d\tilde{\alpha}}{dz} - (Nz_0)^2 \frac{A}{\eta(z)} \tilde{\alpha} = 0, \quad (\text{B.9})$$

ficando evidente que a EDO tem uma singularidade regular em $z = z_0$ e uma singularidade irregular em $z \rightarrow \infty$. Se utilizando da mudança de variável $\frac{d^n \tilde{\alpha}}{d\eta^n} = (Nz_0)^n \frac{d^n \tilde{\alpha}}{dz^n}$ e fatorando $(Nz_0)^2$, que quando nulo é uma solução trivial descartada, temos:

$$\frac{d^2 \tilde{\alpha}}{d\eta^2} + \frac{B - \eta}{\eta} \frac{d\tilde{\alpha}}{d\eta} - \frac{A}{\eta} \tilde{\alpha} = 0. \quad (\text{B.10})$$

A Eq. (B.10) é conhecida como equação hipergeométrica confluyente ou *equação de Kummer*. Pelo método de Frobenius, obtém-se a Eq. (3.22) como uma solução da Eq. (B.10), onde tal solução é conhecida como função hipergeométrica confluyente de primeiro tipo ou função KummerM.

Apêndice C – Dedução de $\tilde{\alpha}(\eta)$, $\tilde{\beta}(\eta)$ e $\tilde{\phi}(\eta)$ pelo método de Frobenius

A Eq. de Kummer (3.20) é uma EDO linear de segunda ordem que tem como solução a Eq. (3.22), e tal solução pode ser obtida pelo *método de Frobenius* (BUTKOV, 1968; ARFKEN; WEBER, 2005), onde se supõe uma solução como série de potências diferenciável termo a termo:

$$\tilde{\alpha}(\eta) = \eta^s \sum_{m=0}^{\infty} C_m \eta^m = \sum_{m=0}^{\infty} C_m \eta^{m+s}, \quad (\text{C.1})$$

onde s é uma incógnita a ser determinada. Isto tornar a solução analítica no ponto de singularidade regular, no nosso caso $\eta = 0$. A técnica consiste substituir a Eq. (C.1) e suas derivadas de primeira e segunda ordem na Eq. (3.20), que nos faz obter:

$$\sum_{m=0}^{\infty} C_m (m+s)(m+s-1) \eta^{m+s-1} + (B-\eta) \sum_{m=0}^{\infty} C_m (m+s) \eta^{m+s-1} - A \sum_{m=0}^{\infty} C_m \eta^{m+s} = 0, \quad (\text{C.2})$$

que após manipular os índices de maneira a fatorar z^{m+s-1} nos termos dos somatórios, temos

$$\sum_{m=0}^{\infty} C_m (m+s)(m+s-1+B) \eta^{m+s-1} - \sum_{m=1}^{\infty} C_{m-1} (m+s-1+A) \eta^{m+s-1} = 0. \quad (\text{C.3})$$

Para agrupar os dois somatórios em um único, expandimos o primeiro termo do somatório que começa em $m = 0$:

$$C_0 s(s-1+B) \eta^{s-1} + \sum_{m=0}^{\infty} C_m (m+s)(m+s-1+B) \eta^{m+s-1} - \sum_{m=1}^{\infty} C_{m-1} (m+s-1+A) \eta^{m+s-1} = 0. \quad (\text{C.4})$$

Uma solução possível implica em anular todos os fatores das potências de z da Eq. (C.4) de maneira que $C_0 \neq 0$, ou seja:

$$C_0 s(s-1+B) = 0, \quad (\text{C.5})$$

$$C_m (m+s)(m+s-1+B) - C_{m-1} (m+s-1+A) = 0, \quad (\text{C.6})$$

onde da Eq. (C.5) é conhecida como equação indicial e dela resulta $s = 0$ ou $s = 1 - B$. Na Eq. (C.6) temos uma fórmula recursiva para C_m . Atentaremos para a solução onde $s = 0$, resultando uma solução em série de Taylor. Sendo assim, a recursão torna-se:

$$C_m = C_{m-1} \frac{(m-1+A)}{m(m-1+B)}, \quad (\text{C.7})$$

onde ao transformar $m - 1 \mapsto m$ e pelo processo indutivo obtemos C_m :

$$C_1 = C_0 \frac{A}{1(B)}, \quad C_2 = \frac{C_0}{1 \times 2} \frac{A \times (A + 1)}{B \times (B + 1)}, \quad \dots, \quad C_m = \frac{C_0}{m!} \frac{(A)_m}{(B)_m}. \quad (\text{C.8})$$

Ao substituir esse resultado correspondente a $s = 0$ na Eq. (C.1) obtemos a Eq. (3.22). A constante C_0 é a normalização de $\tilde{\alpha}$ a se determinar:

$$\tilde{\alpha}(\eta) = C_0 \sum_{m=0}^{\infty} \frac{(A)_m \eta^m}{(B)_m m!} = C_0 M(A, B, \eta(z)). \quad (\text{C.9})$$

Conseqüentemente, $\tilde{\beta}(\eta)$ é obtida ao substituirmos na Eq. (B.3) os parâmetros A , B , η — ver Eqs. (3.21), considerarmos $N = \frac{k}{\rho}$, $z_0 = \frac{\rho}{\rho + h_2}$, $(1 - z) = (1 - z_0) - (z - z_0)$ e fatorarmos $\frac{\rho}{f z_0}$:

$$\begin{aligned} \tilde{\beta}(z) &= \frac{\rho}{f z_0} \left\{ (z - z_0) \frac{d\tilde{\alpha}}{dz} + \left[\frac{k}{\rho} z_0 (1 - z_0) - \frac{k}{\rho} z_0 (z - z_0) + \frac{h_1 + f - f}{\rho} z_0 \right] \tilde{\alpha}(z) \right\}, \\ \tilde{\beta}(z) &= \frac{\rho}{f z_0} \left\{ \frac{\eta}{N z_0} \frac{d\tilde{\alpha}}{dz} + \left[\frac{k}{\rho} z_0 (1 - z_0) + \frac{h_1 + f}{\rho} z_0 - \eta - \frac{f}{\rho} z_0 \right] \tilde{\alpha}(z) \right\}, \\ \tilde{\beta}(z) &= \frac{1}{(A - 1) z_0} \left\{ \frac{\eta}{N z_0} \frac{d\tilde{\alpha}}{dz} + [(B - 1) - (A - 1) z_0 - \eta] \tilde{\alpha}(z) \right\}. \end{aligned} \quad (\text{C.10})$$

Por fim, considerando a Eq. (C.9) e sua derivada em relação a η , obtida em Abramowitz e Stegun (1972), (ver relação 13.4.11 na p. 507),

$$\eta \frac{d}{d\eta} M(A, B, \eta) = (B - A) M(A - 1, B, \eta) - (B - A - \eta) M(A, B, \eta(z)), \quad (\text{C.11})$$

e a mudança de variável para $\frac{d}{dz} = N z_0 \frac{d}{d\eta}$, obtemos:

$$\tilde{\beta}(z) = \frac{C}{(A - 1) z_0} [(B - A) M(A - 1, B, \eta(z)) + (A - 1)(1 - z_0) M(A, B, \eta(z))]. \quad (\text{C.12})$$

A função geratriz $\tilde{\phi}$ independente do estado do promotor do gene é dado por $\tilde{\phi} = \tilde{\alpha} + \tilde{\beta}$, que ao lançar mão das Eqs. (C.9), (C.12) e a relação de recorrência 13.4.3 em Abramowitz e Stegun (1972),

$$(B - A) M(A - 1, B, \eta) = (B - 1) M(A - 1, B - 1, \eta) - (A - 1) M(A, B, \eta), \quad (\text{C.13})$$

conseguimos obter:

$$\tilde{\phi}(z) = \frac{C}{z_0} \frac{B - 1}{A - 1} M(A - 1, B - 1, \eta(z)). \quad (\text{C.14})$$

A constante de normalização C é obtida por $\tilde{\phi}(1) = 1$, sendo assim:

$$C = z_0 \frac{A - 1}{B - 1} [M(A - 1, B - 1, N z_0 (1 - z_0))]^{-1}, \quad (\text{C.15})$$

que ao substituir nas Eqs. (C.9), (C.12) e (C.14), obtemos:

$$\tilde{\alpha}(z) = \frac{z_0}{c} \frac{A-1}{B-1} M(A, B, \eta(z)), \quad (\text{C.16})$$

$$\tilde{\beta}(z) = \frac{1}{c} \left[\left(1 - \frac{A-1}{B-1}\right) M(A-1, B, \eta(z)) + (1-z_0) \frac{A-1}{B-1} M(A, B, \eta(z)) \right], \quad (\text{C.17})$$

$$\tilde{\phi}(z) = \frac{1}{c} M(A-1, B-1, \eta(z)), \quad (\text{C.18})$$

onde $c = [M(A-1, B-1, Nz_0(1-z_0))]^{-1}$.

Apêndice D – Deduções da covariância ξ e do fator de Fano em função de ξ

D.1 Dedução da covariância ξ - Eq. (3.36)

Descrevemos os estados ON e OFF do promotor por meio das taxas médias de síntese de proteínas no estado estacionário e pela fenomenologia do modelo, escrevemos nossa variável aleatória auxiliar como $\nu_m = N \mathbb{I}_\alpha(m)$, onde $\mathbb{I}_\alpha(m)$ é a função indicadora definida como $\mathbb{I}_\alpha(m) = \begin{cases} 1, & \text{se } m=\alpha, \\ 0, & \text{se } m \neq \alpha. \end{cases}$. As probabilidades de $\nu_\alpha = N$ e $\nu_\beta = 0$ são, respectivamente, p_α e p_β . Com isso, escrevemos a covariância entre ν_m e n como

$$\xi = \langle \nu_m n \rangle - \langle \nu_m \rangle \langle n \rangle, \quad (\text{D.1})$$

onde

$$\langle \nu_m \rangle = \nu_\alpha p_\alpha + \nu_\beta p_\beta = N p_\alpha, \quad (\text{D.2})$$

e

$$\langle \nu_m n \rangle = \sum_{n=0}^{\infty} n (\nu_\alpha \alpha_n + \nu_\beta \beta_n) = N \sum_{n=0}^{\infty} n \alpha_n = N \langle n_\alpha \rangle, \quad (\text{D.3})$$

que ao substituí-las na Eq. (D.1), resulta na Eq. (3.36).

D.2 Dedução do fator de Fano em função da covariância ξ - Eq. (3.11)

Para deduzir a Eq. (3.11), basta mostrar que

$$\langle n^2 \rangle - \langle n \rangle^2 = \langle n \rangle + \xi. \quad (\text{D.4})$$

Sendo assim, ao somar as Eqs. (3.15) e (3.16) no limite do estado estacionário, e dividindo a expressão resultante por ρ , temos:

$$0 = N(\tilde{\alpha}_{n-1} - \tilde{\alpha}_n) + [(n+1)\tilde{\phi}_{n+1} - n\tilde{\phi}_n], \quad (\text{D.5})$$

onde $N = k/\rho$ e $\phi_n = \tilde{\alpha}_n + \tilde{\beta}_n$. Agora, multiplicando a Eq. (D.5) por n , ou por n^2 , e aplicando o operador $\sum_{n=0}^{\infty}$, obtemos respectivamente

$$\langle n \rangle = p_\alpha N, \quad (\text{D.6})$$

$$\langle n^2 \rangle = \langle n \rangle + N \langle n_\alpha \rangle, \quad (\text{D.7})$$

que foram obtidas usando Eqs. (3.29) e (3.30) para a função $\tilde{\phi}$ e com $p = 1$ para Eq. (D.6), e $p = 2$ para Eq. (D.7). Substituindo as Eqs. (D.6) e (D.7) no lado esquerdo da igualdade de Eq. (D.4), temos

$$\langle n \rangle + N \langle n_\alpha \rangle - \langle n \rangle^2 = \langle n \rangle + \underbrace{N[\langle n_\alpha \rangle - \langle n \rangle p_\alpha]}_{\text{Eq. (3.36)}} = \langle n \rangle + \xi, \quad (\text{D.8})$$

resultando na expressão esperada para a validade da Eq. (3.11).

Apêndice E – Dedução de ξ^s - Eq. (4.32)

Adotando o mesmo procedimento que foi utilizado para o GER, o cálculo do momento $\langle n_\alpha^s \rangle$ é obtido por $\langle n_\alpha^s \rangle = \left. \frac{d\alpha^s(z)}{dz} \right|_{z=1}$. Sendo assim,

$$\langle n_\alpha^s \rangle = \left. \frac{a}{b^s} \frac{z_0}{c} \frac{d}{dz} [M(a+1, b+1, Nz_0(z-z_0))] \right|_{z=1}. \quad (\text{E.1})$$

Lançando mão da identidade

$$\frac{d}{d\eta} M(a+1, a+1, \eta) = \frac{1}{\eta} [bM(a, b, \eta) + (\eta - b)M(a+1, b+1, \eta)], \quad (\text{E.2})$$

fórmula 13.4.14 em Abramowitz e Stegun (1972), com $\eta = Nz_0(z-z_0)$, reescrevemos a Eq. (E.1) como

$$\begin{aligned} \langle n_\alpha^s \rangle = \frac{a}{b^s} \frac{z_0}{c} \frac{1}{1-z_0} \{ & b^s M(a, b^s, Nz_0(1-z_0)) \\ & - [b^s - Nz_0(1-z_0)] M(a+1, b^s+1, Nz_0(1-z_0)) \}, \quad (\text{E.3}) \end{aligned}$$

e ao usarmos as Eqs. (4.20) e (4.24) é possível obter:

$$\langle n_\alpha^s \rangle = \frac{az_0}{1-z_0} (1 - p_\alpha^s). \quad (\text{E.4})$$

Conseqüentemente, a covariância, Eq. (3.36), pode ser escrita para o GAR como $\xi^s = N \left[\frac{az_0}{1-z_0} (1 - p_\alpha^s) - \langle n \rangle p_\alpha^s \right]$ que torna-se a Eq. (4.32) após rearranjá-la.

Apêndice F – Aproximação por partes dos parâmetros cinéticos dependentes do tempo

Denotamos por E_i o erro da aproximação por partes para o decaimento do efeito da droga em uma dada taxa, dentro de um intervalo $[t_i, t_i + \delta_i)$. Calculamos o erro como o valor absoluto da diferença entre a área do retângulo com base $\delta_i > 0$ e a altura $e^{-\lambda t_i}$, e a área sob a função exponencial $e^{-\lambda t}$ dentro do intervalo $[t_i, t_i + \delta_i]$, a saber:

$$E_i = \delta_i e^{-\lambda t_i} - \int_{t_i}^{t_i + \delta_i} e^{-\lambda t'} dt' = \frac{e^{-\lambda t_i}}{\lambda} (\lambda \delta_i - 1 + e^{-\lambda \delta_i}), \quad (\text{F.1})$$

onde $E_i > 0$ e uma vez estabelecido num determinado intervalo, resolvemos numericamente a equação transcendental $-\lambda \delta_i + 1 + E_i \lambda e^{\lambda t_i} = e^{-\lambda \delta_i}$ para δ_i . Usamos $E_i = 10^{-4}$ para todos os subintervalos ao longo de $[0, t_{\max}]$.

Apêndice G – Sistema de EDOs acopladas que regem os momentos da distribuição

Aqui, estamos interessados em entender a dinâmica de $\langle n \rangle(t)$ e $\sigma^2(t)$. Para isso, precisamos recuperar o sistema de EDOs que regem os momentos de p -ésima ordem da distribuição $\Pi(\alpha_n, \beta_n)$, onde $p = 1, 2, \dots$. Isso é feito aplicando-se $\left(z \frac{\partial}{\partial z}\right)^p$ nas Eqs. (3.17) e (3.18), e definindo $z = 1$ na expressão resultante. Este procedimento é proposto porque o momento de ordem p pode ser recuperado das funções geratrizes da seguinte forma:

$$\langle n_\alpha^p \rangle = \sum_{n=0}^{\infty} \alpha_n n^p = \left[\left(z \frac{\partial}{\partial z} \right)^p \alpha(z, t) \right] \Big|_{z=1}, \quad (\text{G.1})$$

$$\langle n_\beta^p \rangle = \sum_{n=0}^{\infty} \beta_n n^p = \left[\left(z \frac{\partial}{\partial z} \right)^p \beta(z, t) \right] \Big|_{z=1}, \quad (\text{G.2})$$

$$\langle n^p \rangle = \sum_{n=0}^{\infty} \phi_n n^p = \left[\left(z \frac{\partial}{\partial z} \right)^p \phi(z, t) \right] \Big|_{z=1}. \quad (\text{G.3})$$

Aqui, $\phi(z, t) = \alpha(z, t) + \beta(z, t)$ é a função geratriz da probabilidade marginal ϕ_n de encontrar n mRNAs. Portanto, $\langle n^p \rangle$ é o momento de p -ésima ordem no número de mRNAs.

As expressões de $\langle n \rangle(t)$ e $\sigma^2(t)$ são obtidas após a resolução do seguinte sistema composto de EDOs lineares não homogêneas acopladas:

$$\frac{dA}{dt} + (f + h)A = f, \quad (\text{G.4})$$

$$\frac{d\langle n \rangle}{dt} + \rho\langle n \rangle = kA, \quad (\text{G.5})$$

$$\frac{d\langle n_\alpha \rangle}{dt} + (f + h + \rho)\langle n_\alpha \rangle = kA + f\langle n \rangle, \quad (\text{G.6})$$

$$\frac{d\langle n^2 \rangle}{dt} + 2\rho\langle n^2 \rangle = kA + \rho\langle n \rangle + 2k\langle n_\alpha \rangle, \quad (\text{G.7})$$

onde omitimos a dependência do tempo de A , $\langle n \rangle$, $\langle n_\alpha \rangle$ e $\langle n^2 \rangle$.

O sistema de EDOs acima foi obtido da seguinte forma:

1) Eq. (G.4): na Eq. (3.17) definimos, na seguinte ordem, $\beta = \phi - \alpha$, $z = 1$, $\alpha(1, t) = A(t)$ e $\phi(1, t) = 1$;

2) Eq. (G.5): na seguinte ordem, somamos as Eqs. (3.17) e (3.18), aplicamos $\beta = \phi - \alpha$ e o operador $\frac{\partial}{\partial z}$, avaliamos a equação resultante em $z = 1$, e usamos as identidades $\alpha(1, t) = A(t)$ e $\frac{\partial \phi}{\partial z} \Big|_{z=1} = \langle n \rangle$;

3) Eq. (G.6): na Eq. (3.17) definimos, na seguinte ordem, $\beta = \phi - \alpha$, aplicamos o operador $\frac{\partial}{\partial z}$, avaliamos a equação resultante em $z = 1$, e usamos as identidades

$$\alpha(1, t) = A(t), \quad \left. \frac{\partial \phi}{\partial z} \right|_{z=1} = \langle n \rangle \text{ e } \left. \frac{\partial \alpha}{\partial z} \right|_{z=1} = \langle n_\alpha \rangle;$$

4) Eq. (G.7): na seguinte ordem, somamos as Eqs. (3.17) e (3.18), aplicamos $\beta = \phi - \alpha$ e o operador $\frac{\partial^2}{\partial z^2}$, avaliamos a equação resultante em $z = 1$, e usamos as

$$\text{identidades } \alpha(1, t) = A(t), \quad \left. \frac{\partial \phi}{\partial z} \right|_{z=1} = \langle n \rangle, \quad \left. \frac{\partial \alpha}{\partial z} \right|_{z=1} = \langle n_\alpha \rangle \text{ e } \left. \frac{\partial^2 \phi}{\partial z^2} \right|_{z=1} = \langle n^2 \rangle - \langle n \rangle.$$

Apêndice H – Valores dos momentos da distribuição para o estado estacionário

Para obter as soluções de estado estacionário do sistema de EDOs das Eqs. (G.4)-(G.7) pode-se definir $A \rightarrow A_s$, $\langle n \rangle \rightarrow \langle n \rangle_s$, $\langle n_\alpha \rangle \rightarrow \langle n_\alpha \rangle_s$ e $\langle n^2 \rangle \rightarrow \langle n^2 \rangle_s$, onde o índice s indica essas quantidades no limite do estado estacionário, quando se tornam constantes. Portanto, as derivadas são nulas e pode-se resolver o sistema de equações resultante. A Eq. (G.4) é resolvida para A_s , que é substituído no próximo, e assim por diante. Este procedimento resultará em

$$A_s = \frac{f}{f+h}, \quad (\text{H.1})$$

$$\langle n \rangle_s = A_s N \quad (\text{H.2})$$

$$\langle n_\alpha \rangle_s = A_s N \frac{1 + A_s \epsilon}{1 + \epsilon}, \quad (\text{H.3})$$

$$\langle n^2 \rangle_s = A_s N \left(1 + N \frac{1 + A_s \epsilon}{1 + \epsilon} \right). \quad (\text{H.4})$$

Apêndice I – Coeficientes das Eqs. (5.15) e (5.16)

$$Y = N \frac{(A_0 - A_s)}{1 - \epsilon}, \quad (\text{I.1})$$

$$V = \langle n \rangle_0 - Y - \langle n \rangle_s, \quad (\text{I.2})$$

$$U_1 = U - 2Y \langle n \rangle_s, \quad W_1 = W - 2YV, \quad X_1 = X - V^2, \quad (\text{I.3})$$

$$U = Y \left[1 + 2N \frac{1 - \epsilon(1 - A_s)}{2 - \epsilon} \right], \quad (\text{I.4})$$

$$W = \frac{2N}{1 - \epsilon} \left[N(1 - A_s) \frac{\epsilon(A_s - A_0) - A_0}{\epsilon + 1} - A_s \langle n \rangle_0 + \langle n_\alpha \rangle_0 \right], \quad (\text{I.5})$$

$$X = \langle n^2 \rangle_0 - \langle n^2 \rangle_s - U - (1 + 2 \langle n \rangle_s) V - W. \quad (\text{I.6})$$

Apêndice J – Fórmulas exatas para $A(t)$, $\langle n_\alpha \rangle(t)$ e $\langle n^2 \rangle(t)$

As soluções exatas dependentes do tempo do sistema de EDOs das Eqs. (G.4)-(G.7) podem ser obtidas usando o método da variação da constante¹. Para isso, pode-se resolver a Eq. (G.4) para $A(t)$, inseri-lá na próxima equação, e assim por diante. As soluções resultantes são:

$$A(t) = A_s + (A_0 - A_s) e^{-(h+f)t}, \quad (\text{J.1})$$

$$\langle n \rangle(t) = \langle n \rangle_s + Y e^{-\epsilon \rho t} + V e^{-\rho t}, \quad (\text{J.2})$$

$$\langle n_\alpha \rangle(t) = \langle n_\alpha \rangle_s + R e^{-(h+f)t} + S e^{-\rho t} + T e^{-(f+h+\rho)t}, \quad (\text{J.3})$$

$$\langle n^2 \rangle(t) = \langle n^2 \rangle_s + U e^{-(h+f)t} + (1 + 2\langle n \rangle_s) V e^{-\rho t} + W e^{-(f+h+\rho)t} + X e^{-2\rho t}, \quad (\text{J.4})$$

onde

$$R = Y [1 - \epsilon (1 - A_s)], \quad (\text{J.5})$$

$$S = A_s \langle n \rangle_0 + \langle n \rangle_s \frac{A_s \epsilon - A_0}{1 - \epsilon}, \quad (\text{J.6})$$

$$T = \langle n_\alpha \rangle_0 - \langle n_\alpha \rangle_s - R - S. \quad (\text{J.7})$$

¹ Ver [Arfken e Weber \(2005\)](#), Seção 9.2.

Apêndice K – Probabilidades de encontrar n produtos gênicos no estado estacionário

As distribuições de probabilidade do estado estacionário apresentadas aqui foram obtidas a partir das soluções exatas do modelo binário estocástico para regulação da expressão gênica apresentado em [Innocentini e Hornos \(2007\)](#) e [Ramos *et al.* \(2010\)](#). Denotamos as probabilidades marginais, para estado estacionário, de encontrar n produtos gênicos por $\tilde{\phi}_n$. É definido em termos das probabilidades de encontrar o gene ON ou OFF quando n produtos do gene são encontrados dentro da célula, que são denotados por $\tilde{\alpha}_n$ ou $\tilde{\beta}_n$. Portanto, escrevemos $\tilde{\phi}_n = \tilde{\alpha}_n + \tilde{\beta}_n$, que é dado por:

$$\tilde{\phi}_n(A_s, \epsilon, N) = \frac{N^n}{n!} \frac{(A_s \epsilon)_n}{(\epsilon)_n} M(A_s \epsilon + n, \epsilon + n, -N), \quad (\text{K.1})$$

onde $(a)_n \equiv a(a+1) \cdots (a+n-1)$ é o símbolo de Pochhammer e $M(a, b, z) = \sum_{m=0}^{\infty} \frac{(a)_m}{(b)_m} \frac{z^m}{m!}$ é a função Kummer¹.

¹ Ver em [Abramowitz e Stegun \(1972\)](#), p. 503.

Apêndice L – Parâmetros do tratamento e a agenda de doses para desenhos de tratamento aprimorados

A Tab. 6 a seguir mostra os parâmetros do tratamento usados para obter os gráficos contidos na Fig. 17. As taxas cinéticas pós-tratamento, $(k_1; \rho_1; f_1; h_1)$, e os parâmetros da distribuição, $(N_1; A_1; \langle n \rangle_1; \sigma_1)$ são configurados para habilitar os desenhos dos tratamentos aprimorados. O número de doses é representado por n_i e o intervalo de tempo entre elas por τ_i , onde i indica a taxa cinética desejada. Para todos os tratamentos, $\epsilon_1 = 91$, $n_f = 3$ e $\tau_f = 10$.

Tabela 6 – Parâmetros para os tratamentos aprimorados.

Gráficos	$(k_1; \rho_1; f_1; h_1)$	$(N_1; A_1; \langle n \rangle_1; \sigma_1)$	$(n_k; n_\rho; n_h)$	$(\tau_k; \tau_\rho; \tau_h)$
A4	(112; 0,83; 58; 18)	(135; 0,76; 103; 11,8)	(8; 30; 30)	(4; 1; 1)
B4	(91; 0,67; 40; 21)	(136; 0,65; 88; 11,6)	(8; 30; 15)	(4; 1; 2)
C4 e D4	(44; 0,42; 23; 15)	(106; 0,6; 63,6; 9,6)	(30; 30; 30)	(1; 1; 1)
E4	(235; 1,17; 23; 83)	(201; 0,22; 44; 10,9)	(8; 10; 10)	(4; 3; 3)

Fonte – Giovanini *et al.* (2021)

Anexo A – Códigos para computação dos modelos

Os cálculos para a obtenção das distribuições de probabilidades e seus momentos, dos valores de correlação de Pearson e dos valores do fator de Fano para ambos os modelos GER e GAR foram realizados em Maple 18. O código está disponível no repositório GitHub do nosso laboratório <https://github.com/amphybio/Giovanini2020_ComparativeAnalysis>. O código que calcula as simulações dos tratamentos que almejam as taxas cinéticas do modelo dependente do tempo para expressão gênica podem ser obtidos por solicitação ao autor.