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

LUIZ GUILHERME SOARES DA SILVA

Modelo binário para expressão gênica em um embrião de
Drosophila melanogaster

São Paulo
2017

LUIZ GUILHERME SOARES DA SILVA

Modelo binário para expressão gênica em um embrião de
Drosophila melanogaster

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.

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 05 de abril de 2017. A versão original encontra-se em acervo reservado na Biblioteca da EACH/USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Área de Concentração: Fundamentos de Sistemas Complexos

Orientador: Prof. Dr. Alexandre Ferreira Ramos

São Paulo

2017

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.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

Silva, Luiz Guilherme Soares da
Modelo binário para a expressão gênica em um embrião de
Drosophila melanogaster / Luiz Guilherme Soares da Silva ; orientador,
Alexandre Ferreira Ramos. – São Paulo, 2017.
74 f. : il

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

1. Biologia molecular - Simulação computacional. 2. Biologia
molecular - Modelos matemáticos. 3. Expressão gênica. 4.
Dinâmica estocástica. 5. *Drosophila melanogaster*. I. Ramos,
Alexandre Ferreira, orient. II. Título

CDD 22.ed.— 572.80285

Dissertação de autoria de Luiz Guilherme Soares da Silva, sob o título "**Modelo binário para expressão gênica em um embrião de *Drosophila melanogaster***", 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 Fundamentos de Sistemas Complexos, aprovada em 05 de abril de 2017 pela comissão julgadora constituída pelos doutores:

Prof. Dr. José Ricardo Gonçalves de Mendonça
Universidade de São Paulo

Prof. Dr. Luiz Paulo Moura Andrioli
Universidade de São Paulo

Prof. Dr. Vincent Olivier Jean-Marie Noël
Instituto Butantan

Dedico este trabalho:

*Ao meu pai e minha mãe, Dermeval (in memoriam) e Edna, que sempre me apoiaram,
incentivaram e acreditaram na minha capacidade de atingir os objetivos propostos.
À minha esposa, Irene, e meu filho Luiz Felipe, pelo apoio e compreensão durante todo o
período de elaboração deste trabalho.*

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Dr. Alexandre F. Ramos, por todos os ensinamentos passados ao longo desses mais de dois anos de estudos e todo o apoio na elaboração desta dissertação. O desafio de fazer um mestrado em uma área do conhecimento que reúne Biologia do Desenvolvimento, Biologia Molecular, modelos matemáticos e simulações computacionais somente pôde ser vencido devido à sua orientação e seus ensinamentos.

Agradeço a minha esposa, Irene, e meu filho, Luiz Felipe, por terem compreendido o desafio a que me propus e apoiado neste período de estudos e de elaboração desta dissertação. Foi um período de muito trabalho que reduziu bastante nossos momentos de lazer.

Agradeço aos meus pais, Dermeval (*in memorian*) e Maria Edna, por terem sempre valorizado a aprendizagem e dado a mim e a meu irmão, dentre outras coisas, as melhores condições de estudo. Os jornais lidos diariamente, os livros, as enciclopédias que eles traziam para casa e o conhecimento que eles nos passaram (e minha mãe continua passando) através dos seus exemplos de vida sempre nos incentivaram a querer aprender e aplicar aquilo que aprendemos. Sobre esse alicerce de conhecimento construímos nossas vidas.

Agradeço aos professores e funcionários da Universidade de São Paulo (USP), particularmente aqueles eu que tive o prazer de conhecer e de ter sido aluno nesse período: Prof. Dr. Camilo Rodrigues Neto, Prof. Dr. Fernando Fagundes Ferreira, Prof. Dr. José Ricardo Gonçalves de Mendonça, Prof. Dr. Carlos de Brito Pereira, Prof. Dr. Masayuki Oka Hase, Profa. Dra. Flávia Mori Sarti e Prof. Dr. Paulo Sinisgalli.

Agradeço também ao Prof. Dr. Luiz Paulo Moura Andrioli pelo apoio em assuntos referentes à biologia do desenvolvimento e à Profa. Dra. Ana Amélia Benedito-Silva pelo apoio e confiança no estágio PAE.

Agradeço também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Agradeço também aos colegas estudantes do curso, pela troca de informações e ajuda na aprendizagem de cada disciplina.

Finalizando, agradeço a toda a sociedade que, através do pagamento dos impostos e taxas, permitiu meus estudos em uma das mais conceituadas universidades públicas do Brasil.

“Não se gerencia o que não se mede, não se mede o que não se define, não se define o que não se entende, não há sucesso no que não se gerencia”

(William Edwards Deming)

RESUMO

SILVA, Luiz Guilherme Soares. **Modelo binário para Expressão Gênica em um Embrião de *Drosophila melanogaster***. 2017. 74 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2017. Versão Corrigida

Esta dissertação faz uma análise dos resultados de diversas simulações de transcrição usando o modelo binário da expressão gênica regulado externamente. Foram consideradas como variáveis estocásticas o número de moléculas de mRNA e o estado do gene. Neste modelo o gene pode estar no estado ativo (ON) ou reprimido (OFF) e a mudança de estado do gene ocorre a taxas determinadas para a ativação e para a repressão em cada simulação. O número de moléculas de mRNA é alterado pela transcrição de novas moléculas e pela degradação das moléculas existentes, sendo que esses processos também ocorrem a taxas determinadas. Nas simulações foi utilizado o algoritmo de Gillespie para simulação estocástica exata de reações químicas acopladas.

Palavras-chave: Expressão gênica. Ruído da expressão gênica. Gene binário. *Drosophila melanogaster*.

ABSTRACT

SILVA, Luiz Guilherme Soares. **Binary model for gene expression on *Drosophila melanogaster* embryos**. 2017. 74p. Dissertation (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2017. Corrected version

This dissertation analyzes the results of several transcription simulations using the binary model of externally regulated gene expression. The number of mRNA molecules and the state of the gene were considered as stochastic variables. In this model the gene may be in the active state (ON) or repressed state (OFF) and the change of gene state occurs at determined rates in each simulation for activation and for repression. The number of mRNA molecules is changed by the transcription of new molecules and by the degradation of existing molecules at determined rates. These simulations used the Gillespie algorithm for exact stochastic simulation of coupled chemical reactions was used.

Palavras-chave: Gene expression. Gene expression noise. *Drosophila melanogaster*. Binary gene

LISTA DE FIGURAS

Figura 1 - Expressão dos genes <i>pair-rule</i>	13
Figura 2 – Estrutura típica do gene.....	18
Figura 3 - Expressão gênica.	20
Figura 4 - – Hierarquia da expressão dos genes na segmentação da <i>D. melanogaster</i>	23
Figura 5 - Regulação gênica	25
Figura 6 - Ruídos extrínsecos e intrínsecos.	26
Figura 7 - Desenvolvimento da <i>Drosóphila melanogaster</i>	29
Figura 8 - Representação do modelo binário e modelo cinético.	33

SIMULAÇÕES

SIMULAÇÃO 1 - Modelo estocástico de gene binário com $\varepsilon \ll 1$ e $A \sim 0,0$	47
SIMULAÇÃO 2 - Modelo estocástico de gene binário com $\varepsilon \ll 1$ e $A \sim 0,5$	48
SIMULAÇÃO 3 - Modelo estocástico de gene binário com $\varepsilon \ll 1$ e $A \sim 1,0$	49
SIMULAÇÃO 4 - Modelo estocástico de gene binário com $\varepsilon < 1$ e $A \sim 0,0$	50
SIMULAÇÃO 5 - Modelo estocástico de gene binário com $\varepsilon < 1$ e $A \sim 0,5$	51
SIMULAÇÃO 6 - Modelo estocástico de gene binário com $\varepsilon < 1$ e $A \sim 1,0$	52
SIMULAÇÃO 7 - Modelo estocástico de gene binário com $\varepsilon > 1$ e $A \sim 0,0$	53
SIMULAÇÃO 8 - Modelo estocástico de gene binário com $\varepsilon > 1$ e $A \sim 0,5$	54
SIMULAÇÃO 9 - Modelo estocástico de gene binário com $\varepsilon > 1$ e $A \sim 1,0$	55
SIMULAÇÃO 10 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 0,0$	56
SIMULAÇÃO 11 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 0,5$	57
SIMULAÇÃO 12 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 1,0$	58
SIMULAÇÃO 13 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 0,0$	59
SIMULAÇÃO 14 - Gene <i>even-skipped</i> da <i>D. melanogaster</i>	60

SUMÁRIO

1	INTRODUÇÃO.....	12
1.1	BREVE HISTÓRICO	14
1.2	DNA, RNA E PROTEÍNAS	15
1.3	GENES.....	18
1.4	EXPRESSÃO GÊNICA.....	19
1.5	TRANSCRIÇÃO E TRADUÇÃO NA EXPRESSÃO GÊNICA	20
1.6	REGULAÇÃO DA EXPRESSÃO GÊNICA	22
1.7	O RUÍDO NA EXPRESSÃO GÊNICA	25
1.7.1	Análise quantitativa do ruído na expressão gênica	27
1.8	ORGANISMOS MODELO - <i>DROSOPHILA MELANOGASTER</i>	28
2	UM MODELO ESTOCÁSTICO PARA A TRANSCRIÇÃO GÊNICA.....	32
2.1	SOLUÇÃO EXATA DAS EQUAÇÕES MESTRAS.....	35
2.2	DINÂMICA DAS PROBABILIDADES DO GENE ATIVO OU REPRIMIDO	35
	Dinâmica das probabilidades de o gene ficar ativado	36
2.3	DINÂMICA DO NÚMERO MÉDIO DE MOLÉCULAS DE MRNA	37
2.4	REGIME ESTACIONÁRIO	38
3	APLICAÇÃO AO PROCESSO DE DESENVOLVIMENTO DA <i>D.</i>	
	<i>MELANOGASTER</i>	40
3.1	SIMULAÇÕES	40
3.1.1	A influência dos parâmetros na simulação da expressão gênica.....	42
3.1.2	Análise dos resultados das simulações.....	44
3.1.3	Simulação da transcrição do gene <i>even-skipped</i> da <i>D. melanogaster</i>	45
4	DISCUSSÃO E CONSIDERAÇÕES FINAIS	61
5	REFERÊNCIAS.....	63
	APÊNDICE A - FLUXOGRAMA DA IMPLEMENTAÇÃO DO ALGORITMO DE GILLESPIE.....	65
	APÊNDICE B - GERAÇÃO DAS EQUAÇÕES DIFERENCIAIS PARCIAIS ACOPLADAS	67
	APÊNDICE C - OBTENÇÃO DAS DERIVATIVAS	68
	APÊNDICE D - OBTENÇÃO DOS MOMENTOS.....	69
	APÊNDICE E - OBTENÇÃO DA EQUAÇÃO DE KUMMER	70
	APÊNDICE F - OBTENÇÃO DA EQUAÇÃO MESTRA	71
	APÊNDICE G - PROBABILIDADE DE ENCONTRAR <i>N</i> MRNA	72
	ANEXO I.....	73

1 INTRODUÇÃO

A despeito da diversidade de organismos vivos habitando a Terra, eles apresentam a característica comum de serem formados por células que carregam maquinário biológico semelhante e possuem propriedades universais. Dentre essas propriedades é possível citar a capacidade de se reproduzirem e transferirem suas características aos descendentes, o armazenamento de informações hereditárias em um mesmo tipo de código químico linear denominado ácido desoxirribonucleico (DNA – *deoxyribonucleic acid* em inglês), o uso de um formato padrão para replicar suas informações hereditárias, uso de uma mesma forma intermediária de armazenamento temporário das informações hereditárias denominada ácido ribonucleico (RNA – *ribonucleic acid* em inglês); o uso de proteínas na execução de diversas funções celulares como catalisadores, metabolização, estrutura celular, o mesmo processo para tradução do RNA em proteínas e o envolvimento das células por uma membrana denominada membrana plasmática (1).

A capacidade de se reproduzir, transferindo de forma minuciosa as características que o descendente deve ter, é o diferencial entre a vida e demais processos existentes na natureza. Esse processo de armazenamento e transferência das características aos descendentes é denominado hereditariedade e é estudado pelo ramo da biologia denominado Genética (1). Embora importante, aqui pretendemos compreender como a informação genética é expressa em um organismo vivo, isto é, como a informação armazenada no DNA é transformada em diversas funções celulares. Essas funções são executadas tal que uma determinada célula, numa determinada localização de um organismo e em determinado estágio de seu desenvolvimento, tenha sua função estabelecida com precisão. Essa precisão ocorre pela presença de moléculas de RNA mensageiro (mRNA) e proteínas em quantidades apropriadas e se torna impressionante se consideramos que as moléculas do ambiente intracelular estão presentes em quantidades aleatórias. A quantidade apropriada de moléculas de mRNA e proteínas decorre do fato de essas moléculas serem produtos de reações cujas moléculas reagentes estão presentes em quantidade aleatória e baixa de cópias, que leva à aleatoriedade dos intervalos de tempo entre as reações tanto de criação como de degradação das moléculas. Embora as quantidades de moléculas reagentes sejam aleatórias, suas quantidades flutuam em torno de valores bioquimicamente apropriados (2). Portanto, caracterizar as amplitudes dessas flutuações que distinguem o funcionamento celular normal do anômalo em seres vivos é uma questão premente. Esse trabalho visa contribuir nessa direção pela análise das flutuações aleatórias no número de moléculas de mRNA de um embrião da mosca da fruta (*Drosophila*

melanogaster) e em como a regulação dessa produção afeta a quantidade de mRNA no interior celular.

Esta dissertação tem o objetivo de analisar a estocasticidade no processo de transcrição do gene *even-skipped* durante a formação da faixa 2 em um embrião de *Drosophila melanogaster*. A escolha deste gene se deve à importância de sua expressão no desenvolvimento do embrião, particularmente na determinação dos segmentos corporais do indivíduo adulto. A repressão definitiva desse gene leva a não formação de diversas partes do embrião e à consequente morte nos primeiros estágios do desenvolvimento (1). Além dessa importância, no estágio do desenvolvimento em que ocorre a expressão desse gene, o embrião é formado por uma única grande célula com diversos núcleos em um único citoplasma. A expressão do gene *even-skipped* é regulada pela presença de diferentes combinações de proteínas em diferentes concentrações ao longo do eixo ântero-posterior e essas diferenças fazem com que o gene seja expresso em sete faixas ao longo desse eixo, com 5 a 6 núcleos em cada faixa (1) (2). A expressão do gene *even-skipped* está representada na figura 1.

Para essa análise foi adotado o modelo estocástico para a regulação da transcrição considerando um gene que opera em dois estados - ativo e reprimido, com simulações realizadas utilizando-se o algoritmo de Gillespie para simulação estocástica exata de reações químicas acopladas.

Figura 1 - Expressão dos genes *pair-rule*. Os genes estabelecem o plano de segmentação anteroposterior do corpo do embrião. Em azul está representada a expressão do gene *even-skipped* e em verde a expressão do gene *fushi tarazu*. (3)

Esta dissertação está organizada em quatro capítulos.

No Capítulo 1 é apresentado um histórico da genética e da expressão gênica com o intuito de compreender o processo de segmentação em embriões de *Drosophila melanogaster*. Em seguida revisamos como as flutuações aleatórias das quantidades de moléculas têm sido investigadas em sistemas biológicos teórica e experimentalmente.

No Capítulo 2 será apresentado o modelo estocástico de regulação gênica que considera o gene binário, sua dinâmica de transição de fase, as equações de cálculo da probabilidade do gene estar em cada estado, o número médio de moléculas de mRNA dependente do tempo e posição espacial no eixo AP.

No Capítulo 3 será estudada a aplicação do modelo no embrião da *Drosophila melanogaster* e a representação gráfica dos resultados.

No Capítulo 4 serão feitas considerações finais referentes aos resultados do trabalho e algumas proposições de avanços no estudo do gene binário.

1.1 BREVE HISTÓRICO

A preocupação com as características herdadas pelos descendentes de um organismo vivo existe desde a pré-história, quando agricultores e criadores de animais buscavam fazer cruzamentos seletivos de animais e plantas com o objetivo de melhorar as espécies. (4) Esse processo prático começa a ganhar explicações teóricas com os trabalhos de Gregor Mendel (1822-1884), que estudou a hereditariedade a partir da análise da transmissão de cada característica isoladamente, ao invés de analisá-las em bloco, como faziam os demais pesquisadores. Os estudos de Mendel são considerados como ponto inicial da genética (4).

Apesar de estudar a hereditariedade, Mendel não investigou a forma como as características hereditárias estão armazenadas e são processadas no interior das células. Foi Theodor Boveri que, em 1902, executou uma série de experimentos com ovos de ouriço-do-mar e percebeu que tipos específicos de cromossomos são responsáveis pelo desenvolvimento normal. Ele concluiu que os cromossomos individuais possuem diferentes características e afirmou que as características relatadas nos experimentos de Mendel estão relacionadas a cromossomos específicos. Nessa mesma época, Walter Sutton chegou de forma independente às mesmas conclusões e a teoria foi denominada *Teoria dos Cromossomos de Boveri-Sutton*. (5)

A Teoria dos Cromossomos viria a se consolidar com os trabalhos do geneticista Thomas Hunt Morgan sobre a hereditariedade. Morgan concluiu que mutações em larga escala podem dar origem a novas espécies, confirmou as leis de Mendel e o conceito do cromossomo como portador dos genes. Em seus trabalhos, nos quais utilizou milhões de moscas da espécie *Drosophila melanogaster*, Morgan identificou que uma mutação na cor dos olhos da mosca (cor branca ao invés da vermelha) reaparecia apenas na segunda geração e apenas no sexo masculino. A partir dessa descoberta publicou, em conjunto com outros

pesquisadores, *The Mechanism of Mendelian Heredity* (6), bem como tornou a *Drosophila melanogaster* um dos principais organismos-modelo em experimentos genéticos. Posteriormente a associação entre a hereditariedade genética e a teoria da seleção natural de Darwin seria definida através da *Teoria Sintética da Evolução* desenvolvida por cientistas como Ronald Fisher, John Haldane, Sewall Wright, Willian Bateson, Hugo de Vries, Wilhelm Johannsen e outros (2).

Na década de 1920, Hermann Muller descobre a atuação dos raios-X como agentes mutagênicos e em 1935 Nikolai Timofeeff, Karl Zimmer e Max Delbruck propõem um modelo de mutagênese causadas pelos raios-X e raios- γ . O modelo chama a atenção de Erwin Schrödinger, que associa a estrutura do material genético a moléculas. Tal associação levou outros cientistas a analisarem a formação molecular do material genético, como o físico Francis Crick e o biólogo James Watson. Em 1953 ambos determinariam o formato de dupla-hélice do ácido desoxirribonucleico (DNA), um marco no surgimento da biologia molecular (2).

1.2 DNA, RNA E PROTEÍNAS

As moléculas de DNA de fita dupla formam o repositório de informações hereditárias de todas as células da Terra. As moléculas de DNA são constituídas por duas longas cadeias poliméricas¹ pareadas e não ramificadas formadas por quatro tipos de monômeros. Cada monômero no DNA é denominado nucleotídeo e é composto por um açúcar (desoxirribose) associado a um fosfato e uma base nitrogenada que pode ser Adenina (A), Timina (T), Citosina (C) ou Guanina (G). Cada nucleotídeo se une ao próximo através de ligações covalentes entre o fosfato de um nucleotídeo e o açúcar do outro, gerando sequências repetitivas que formam a cadeia (fita). A base de cada nucleotídeo se unirá a uma base de nucleotídeo de outra fita, seguindo regras de associação bem definidas e formando a fita dupla. As informações são codificadas em sequências lineares desses nucleotídeos e a assimetria das unidades de açúcar fosfato entre as duas fitas guia os processos moleculares que realizam a interpretação das informações e as copiam para as células (1).

A transferência das informações contidas no DNA é feita através da cópia do DNA, em um processo idêntico em todas as células vivas conhecidas. Esse processo é baseado na

¹ Formadas por polímeros, isto é, macromoléculas constituídas a partir de unidades estruturais menores denominadas monômeros. A mudança da unidade estrutural do polímero ou da forma com que se arranjam gera materiais com características diferentes (40)

possibilidade de separação das duas fitas de DNA sem danos a ambas devido à pouca força da ligação por pontes de hidrogênio entre as bases em comparação com a força da ligação fosfodiéster (um tipo de ligação covalente) entre os fosfatos. A partir dessa separação, cada fita pode ser usada como molde para a construção de novas fitas complementares a ela (1).

Além da possibilidade de cópia da informação armazenada no DNA, também existe o processo de expressão dessa informação, isto é, a síntese de outros tipos de moléculas a partir da informação contida no DNA. O processo se inicia com a **transcrição** da informação, isto é, a criação de um polímero de ácido ribonucleico (RNA) a partir de um segmento específico de uma das fitas do DNA. Após a transcrição, segue a etapa de **tradução**, onde a informação contida no RNA direciona a criação de outros tipos de polímeros denominados proteínas (1).

O RNA tem uma constituição similar a uma fita de DNA e fica contido no citoplasma e no núcleo. As diferenças entre o RNA e o DNA estão relacionadas ao tipo de açúcar, que no caso do RNA é a ribose ao invés da desoxirribose do DNA, e a presença da base uracila (U) ao invés da base timina (T). As moléculas de RNA são classificadas conforme o Quadro 1 abaixo:

Quadro 1 - Tipos de RNA (1)

Tipo de RNA	Função
RNA mensageiro (mRNA)	Atua no transporte da informação que guiará a produção de proteínas do DNA aos ribossomos.
RNA ribossômico (rRNA)	Constitui componente majoritário do ribossomo.
RNA transportador (tRNA)	Faz a conexão entre o códon ² e o aminoácido, transportando os resíduos de aminoácidos para os ribossomos a fim de sintetizar as proteínas.
Small nuclear RNA (snRNA)	Atuam em diversos processos, por exemplo o <i>splicing</i> do pré-mRNA.
Small nucleolar RNA(snoRNA)	Atuam no processamento e modificação química dos rRNA.
Small Cajal body RNA(scaRNA)	Atuam na modificação dos snoRNA e snRNA.
Micro RNA (miRNA)	Atuam na regulação da expressão gênica através do bloqueio da tradução de determinados mRNA.
Small interfering RNA(siRNA)	Desativam a expressão gênica pela degradação de determinados mRNA e pelo uso de estruturas de cromatina compacta
Outros tipos de RNA não codificantes	Atuam em diversos processos celulares. Exemplo: Síntese de telômeros, inativação do cromossomo X e transporte de proteínas para o retículo endoplasmático

² Unidades hereditárias que contém o código de informação para um aminoácido e são compostos por três nucleotídeos

As proteínas são polímeros que catalisam reações químicas e atuam em praticamente todas as funções biológicas. As proteínas participam do transporte de moléculas no interior das células e entre células, do controle da permeabilidade das membranas celulares, da sustentação das células, órgãos e estruturas corporais, da proteção contra toxinas e infecções e da regulação da produção de produtos gênicos (7). As moléculas de proteínas são cadeias longas e não ramificadas formadas por sequências de monômeros que carregam informações codificadas de forma semelhante ao DNA e RNA. Entretanto, a proteína possui aminoácidos completamente diferentes, sendo ao todo 20 tipos de aminoácidos (Quadro 2) (1).

Quadro 2 - Tipos de aminoácidos (1) .

Aminoácidos			Códons
Símbolo	Abreviatura	Aminoácido	
A	Ala	Alanina	GCA GCC GCG GCU
C	Cys	Cisteína	UGC UGU
D	Asp	Ácido Aspártico	GAC GAU
E	Glu	Ácido Glutâmico	GAA GAG
F	Phe	Fenilalanina	UUC UUU
G	Gly	Glicina	GGA GGC GGG GGU
H	His	Histidina	CAC CAU
I	Ile	Isoleucina	AUA AUC AUU
K	Lys	Lisina	AAA AAG
L	Leu	Leucina	UUA UUG CUA CUC CUG CUU
M	Met	Metionina	AUG
N	Asn	Asparagina	AAC AAU
P	Pro	Prolina	CCA CCC CCG CCU
Q	Gln	Glutamina	CAA CAG
R	Arg	Arginina	AGA AGG CGA CGC CGG CGU
S	Ser	Serina	AGC AGU UCA UCC UCG UCU
T	Thr	Treonina	ACA ACC ACG ACU
V	Val	Valina	GUA GUC GUG GUU
W	Trp	Triptofano	UGG
Y	Tyr	Tirosina	UAC UAU

1.3 GENES

Genes são segmentos do DNA correspondente a uma unidade básica física e funcional de hereditariedade que contém as informações que caracterizam as espécies de forma geral e os indivíduos de forma particular.

Os genes são compostos por sequências de nucleotídeos que serão traduzidos na sequência de aminoácidos que dará origem às proteínas. A estrutura típica do gene contém os seguintes elementos: Sítio promotor, região do DNA ao qual uma molécula de RNA polimerase se liga antes de iniciar a transcrição e que contém informações que determinam qual das cadeias de DNA será transcrita e onde começa a transcrição; Sítio de início da transcrição, região onde inicia transcrição; Região não traduzida; Sítio de início da tradução, Sequência de exon e introns, Sítio de fim da tradução, Região não traduzida, Sítio de término da transcrição e Sítio terminador, onde a RNA polimerase se desconecta da cadeia de DNA (1) (8).

A área a ser codificada do gene, isto é, a área do gene que será transcrita, é dividida em regiões denominadas exons e íntrons. Os exons contém o código para geração de proteínas e estão intercalados por íntrons. Ambas as regiões são transcritas na síntese do RNA, mas os íntrons são posteriormente removidos das moléculas de mRNA sintetizadas. Assim sendo apenas os exons atuam na codificação das proteínas. A estrutura típica do gene está ilustrada na Figura 2.

Figura 2 – Estrutura típica do gene. Genes são segmentos de DNA divididos em exons e íntrons estruturados nos seguintes sítios: promotor, região do DNA ao qual uma molécula de RNA polimerase se liga antes de iniciar a transcrição e que contém informações que determinam qual das cadeias de DNA será transcrita e onde começa

a transcrição; sítio de início da transcrição, sítio onde inicia a transcrição; região não traduzida; Sítio de início da tradução; sequência de exon e introns; sítio onde finaliza a tradução; região não traduzida, sítio onde ocorre o término da transcrição e sítio terminador, onde a molécula de mRNA polimerase se desconecta da cadeia do DNA. Imagem adaptada de (9)

O conjunto de informações genéticas armazenadas nos genes do DNA das células dos organismos vivos e que os caracteriza é denominado genoma. (1) (2) (10) (11)

1.4 EXPRESSÃO GÊNICA

O processo de expressão das informações armazenadas no DNA é denominado expressão gênica e ocorre em todas as etapas da vida de seres uni ou pluricelulares. A expressão gênica é um fenômeno fundamental para Biologia do Desenvolvimento, área da biologia que estuda ao desenvolvimento dos seres vivos em suas diversas fases, incluindo o crescimento, diferenciação celular e desenvolvimento, a partir do embrião, das formas e estruturas características da espécie, devido à sua importância tanto no processo de desenvolvimento embrionário como em processos de diferenciação celular e resposta a influência do meio.

A expressão gênica é base do processo de desenvolvimento de um organismo, isto é, a sequência que leva uma célula embrionária transformar-se num indivíduo adulto estruturado em múltiplos órgãos com diferentes funções. O processo apresenta grande precisão espaço-temporal, de maneira que, além de conhecer o DNA de um organismo, isto é, quais proteínas ele codifica, é essencial saber em que quantidades, locais do organismo e momentos de suas vidas essas proteínas serão produzidas. Portanto, compreender os mecanismos de produção de proteínas e a sua regulação é uma questão essencial da biologia contemporânea. (12) (13).

O processo da expressão gênica não é direto. O DNA genômico utiliza o RNA como uma molécula intermediária para a produção de proteínas específicas. Na primeira etapa é executada a transcrição de uma sequência de nucleotídeos da região apropriada de uma molécula de DNA inserida em um cromossomo para a forma de RNA. Posteriormente o RNA será usado diretamente como molde para direcionar a síntese da proteína em um processo denominado tradução. Dessa forma, o fluxo de informações, que ocorre da mesma forma em todas as células conhecidas, inicia-se do DNA para o RNA e posteriormente do RNA para a proteína. Esse é um princípio tão fundamental e universal³ que foi denominado por Francis

³ Apesar da universalidade do dogma central, há variações importantes no fluxo que podem modificar o significado de uma molécula de RNA, bem como no caso de alguns genes onde o RNA é o produto final (1), entretanto essas situações não serão abordadas no presente trabalho.

Crick como o dogma central da biologia molecular (1) (14). O processo de expressão gênica de uma célula eucariota está representado na Figura 3.

Figura 3 - Expressão gênica. A transcrição ocorre no núcleo celular e é realizada por RNA-polimerases, que se movem lentamente ao longo do DNA para catalisar a formação uma cadeia linear denominada mRNA. O mRNA para o citoplasma onde ocorrerá a tradução cujo produto são proteínas. Adaptado de (15)

1.5 TRANSCRIÇÃO E TRADUÇÃO NA EXPRESSÃO GÊNICA

Transcrição⁴ é o processo de síntese do RNA através da cópia de uma parte específica da sequência de nucleotídeos do DNA (ou seja, de um gene) em uma sequência de nucleotídeos de RNA. A transcrição é realizada por enzimas processivas denominadas RNA-polimerases, que se movem lentamente ao longo do DNA para catalisar a formação de pontes fosfodiéster entre ribonucleotídeos, formando uma cadeia linear.

A transcrição é dividida em três etapas: 1 - Início, onde ocorre a abertura e desespiralização de uma pequena parte da dupla-hélice do DNA, expondo as fitas e onde sequências específicas do DNA indicam o local de início da transcrição, 2 - Fase de alongamento da cadeia, em que é determinada uma sequência de nucleotídeos da cadeia de

⁴ Este processo é denominado transcrição porque a linguagem de armazenamento no RNA é basicamente a mesma da linguagem de armazenamento no DNA. Dessa forma, o que ocorre é a transcrição da informação do formato do DNA para o formato do RNA.

RNA de forma complementar a uma das fitas do DNA-molde. Esta fase pode ser dividida em subprocessos: cada pareamento adequado formando o ribonucleotídeo é ligado de forma covalente à cadeia de RNA em formação, fazendo com que a cadeia de RNA formada (denominada transcrito) seja aumentada um nucleotídeo por vez. Imediatamente após a inclusão de um nucleotídeo, o RNA se desloca e a hélice de DNA se reassocia, 3 – Terminação, onde a síntese é finalizada a partir de sinais de terminação (terminadores). Ao final do processo, obtém-se uma fita de RNA complementar ao gene (1) (14) (8).

A ligação inicial do RNA com o DNA na etapa 1 acontece em uma região do gene denominada *promotor* e reconhecida pela holoenzima⁵ RNA-polimerase. É o promotor que possui a informação que determina qual das duas fitas de DNA será transcrita e o ponto onde começará a transcrição (8).

A transcrição pode ser considerada a primeira etapa da regulação da expressão gênica. Neste processo são definidos quais genes serão transcritos (e, conseqüentemente, expressos), quando e em que quantidade os produtos serão produzidos (16).

No caso dos eucariontes, o mRNA gerado, denominando transcrito primário de mRNA, é processado dando origem a moléculas de mRNA maduras. Nesse processamento, denominado *splicing*, serão retiradas as sequências de íntrons e as sequências de éxons são reunidas para formar um mRNA final (11). As moléculas de mRNA finais formadas no processo de transcrição são usadas como molde para sintetizar proteínas no processo de tradução⁶ (ou síntese proteica), que ocorre no ribossomo⁷

Na **tradução** ocorre a formação das proteínas e, para tanto, o processo demanda um dicionário denominado **código genético** que estabelece a correspondência entre um **códon** e um aminoácido. Os quatro nucleotídeos (ATCG no DNA e AUCG no RNA) são combinados três a três para formar um códon, podendo formar 64 diferentes códons. Todavia, há apenas 20 tipos de aminoácidos. As moléculas de cada um dos 20 aminoácidos que irão compor a proteína são unidas às moléculas adaptadoras denominadas tRNAs, as quais podem reconhecer e se ligarem ao aminoácido e ao códon no mRNA, que contém a informação necessária para o sequenciamento correto do aminoácidos e, conseqüentemente, torna possível traduzir a sequência de códons em uma sequência de aminoácidos (molécula de proteína) (1) (17).

⁵ Enzima cataliticamente ativa formada por uma proteína e um cofator que pode ser um íon ou uma molécula orgânica completa (ou não).

⁶ Este processo é denominado tradução porque a informação passa a ser codificada em outra linguagem com 20 aminoácidos diferentes na proteína ao invés dos 4 aminoácidos diferentes no RNA

⁷ Estrutura citoplasmática composta por dois terços de rRNA localizados na posição central da estrutura e por um terço de proteínas posicionadas na região superficial, nas frestas e enrugamentos do rRNA.

1.6 REGULAÇÃO DA EXPRESSÃO GÊNICA

A expressão gênica é um processo regulado que ocorre apenas na presença de condições específicas. Dessa forma, apesar de todas as células de um organismo possuírem o mesmo DNA, os genes desse DNA são expressos no interior celular de acordo com sua função no tecido em que se encontram. A síntese proteica e o acúmulo desses produtos em diferentes quantidades permite que as células de organismos multicelulares se tornem diferentes uma das outras. (1)

O início dos estudos da regulação da expressão gênica, isto é, do controle na produção de proteínas, se deu nas décadas de 1950 e 1960, com a identificação da molécula de RNA, do tRNA e da relação existente entre DNA, RNA e as proteínas. O primeiro trabalho sobre expressão gênica foi realizado por André Lwoff (1902-1994) e abordava o comportamento da bactéria *E. coli* que carrega o vírus bacteriófago λ –fago de forma dormente. Este vírus se prolifera invadindo outras bactérias, as quais podem sofrer ruptura da membrana celular liberando cópias do λ -fago no ambiente ou podem carregar o vírus em forma dormente (2).

No experimento realizado por André Lwoff, as bactérias tinham o crescimento interrompido ao serem irradiadas por uma dose moderada de luz ultravioleta. Decorridos 90 minutos da interrupção as bactérias sofriam *lyse*, rompendo-se e espalhando cópias do vírus bacteriófagos. Esses vírus libertados infectavam outras bactérias e o experimento era reiniciado com nova irradiação de luz ultravioleta. Neste experimento percebeu-se que a *lyse* das bactérias hospedeiras ocorria apenas após a irradiação por luz ultravioleta, levando os pesquisadores a concluir a existência de dois estados de virulência – inativo e ativo. No estado inativo a bactéria cresce e se divide carregando o vírus, enquanto no estado ativo, logo após a irradiação, o vírus se difunde no ambiente após o rompimento da bactéria. Esses dois estados estão relacionados a processos biológicos denominados ativação e inativação.

Na segunda metade da década de 1980 e na década de 1990 identificou-se que os padrões de expressão gênica em embriões de *Drosophila melanogaster* derivam de sequências de interações cujos genes são transcritos em uma ordem hierárquica temporal e espacial de atividades gênicas. Nessa hierarquia, as proteínas expressas de um grupo de genes em um estágio do desenvolvimento são essenciais para que o próximo estágio seja bem sucedido.

Um exemplo da transcrição em ordem temporal e espacial é a transcrição dos genes de segmentação *gap*, *pair-rule*, *segment-polarity* e *homeóticos* da *D. melanogaster*.

Os genes *gap* definem uma divisão inicial e não refinada do corpo do embrião, determinando a posição, identidade e delimitação de cada segmento corporal. Os genes *gap*

são *hunchback (hb)*, *Krüppel (Kr)*, *knirps (kni)*, *giant (gt)*, *tailless (tll)* e *huckebein (hkb)*. A expressão desse grupo de genes é regulada pelos produtos da expressão dos genes maternos *bicoid (bcd)*, *caudal (cad)* e *hunchback (hb)*, que já se encontram no citoplasma do óvulo antes da fertilização (18).

As proteínas originadas da expressão dos genes maternos e *gap* regulam a expressão dos genes *pair-rule*, também responsáveis pela segmentação do corpo do embrião. O grupo de genes *pair-rule* é composto pelos genes *even-skipped (eve)*, *hairy (h)*, *odd-skipped (odd)*, *paired (prd)*, *runt (run)*, *fushi tarazu (ftz)*, *odd-paired (opa)* e *sloppy paired (slp)* (19).

Já as proteínas originadas da expressão dos genes *pair-rule* regulam a expressão dos genes *segment-polarity*, responsáveis pela integridade dos segmentos. Neste grupo de genes encontram-se os genes *engrailed (en)*, *wingless (wg)*, *cubitus interruptus (ci)*, *hedgehog (hh)*, *fused (fu)*, *armadillo (arm)*, *patched (ptc)*, *gooseberry (gsb)* e *pangolin (pan)*.

As proteínas originadas da expressão dos genes dos grupos *gap*, *pair-rule*, e *segment-polarity* regulam a expressão dos genes homeóticos, que definem e preservam as diferenças entre segmentos vizinhos. Na Figura 4 está ilustrado um exemplo da hierarquia da expressão dos genes na segmentação da *D. melanogaster*. (1) (2) (20)

Figura 4 - – Hierarquia da expressão dos genes na segmentação da *D. melanogaster*. A expressão dos genes maternos regula a transcrição dos genes zigóticos *gap* e a expressão de ambos regula a transcrição dos genes *pair-rule*. A expressão dos genes *pair-rule* regula a expressão dos genes de *polaridade de segmento*. A expressão dos genes *gap*, *pair-rule*, e de *polaridade de segmento* regulam a expressão dos genes homeóticos que irão determinar as estruturas de cada segmento do corpo do embrião/larva. Imagem adaptada de (2)

Os processos de ativação e inativação podem ser iniciados por moléculas contidas na própria célula a partir de um subconjunto de genes que orientam a produção de outras moléculas ativando ou inativando genes que as codificam. Em cada momento as células utilizam apenas um subgrupo de genes para produzir proteínas, os quais são denominados

genes ativados (ou ativos), enquanto genes inativados (ou inativos) não estão sendo expressos. Esta alternância controlada é denominada regulação da expressão gênica.

A regulação gênica decorre da ação do sítio promotor, regiões da dupla hélice do DNA que possuem uma sequência de nucleotídeos específica para indicar o ponto inicial da síntese de RNA e onde a RNA-polimerase se conectará ao DNA. Dentro do sítio promotor existe um elemento regulador denominado sítio operador nos procariotos e sítio de ligação nos eucariotos (“*enhancers*”). A partir da forma com que as proteínas regulatórias atuam nos pontos de conexão do sítio operador (ou de ligação, nos eucariotos), é possível identificar duas formas de regulação gênica, denominadas regulatória negativa e regulatória positiva.

A regulação negativa da expressão gênica é característica dos procariotos e se dá pelo bloqueio e desbloqueio do sítio promotor por uma proteína regulatória repressora que reconhece a sequência de nucleotídeos e se conecta ou desconecta do elemento operador. Dessa forma, quando o repressor ocupa o operador, ela impede a conexão da RNA-polimerase e, conseqüentemente, impede a fase de transcrição da expressão gênica.

No processo de controle positivo, característico dos eucariotos, outras proteínas regulatórias atuam exatamente de forma inversa, isto é, ao estarem ligadas ao DNA podem estimular a conexão da RNA-polimerase levando a execução da transcrição. Neste caso as proteínas são denominadas ativadoras e o processo chamado de controle positivo (ou regulatória positiva) (1) (2) (10).

Tanto o processo de controle positivo como o controle negativo podem fazer com que a transcrição de mRNA ocorra na forma de *bursts* com frequência e intensidade aleatórias, as quais determinam a taxa de transcrição de mRNA (21).

Também é possível classificar a regulação gênica a partir da origem da proteína que a exerce. Por este critério a regulação pode ser classificada como regulação externa, autorregulação e regulação mista. Na regulação externa a proteína que regula a expressão de um gene X é codificada na expressão de um gene Y. Já na autorregulação a proteína reguladora é codificada na expressão do próprio gene que está sendo regulado. Na regulação mista, o gene é regulado por proteínas codificadas por outros genes e também é autorregulado. Dois genes são interagentes entre si se as proteínas codificadas por um desses genes regulam a expressão do outro gene⁸. As interações que ocorrem entre os genes, isto é, a forma como os genes se relacionam para a execução das funções biológicas celulares, é denominada **rede gênica**.

⁸ Esta definição pode ser estendida para um número de genes maior que dois.

Figura 5 - Regulação gênica – Autorregulação, regulação externa e regulação mista. As linhas tracejadas indicam as regulações exercidas na expressão dos genes e as linhas contínuas indicam as etapas da expressão gênica (transcrição e tradução). Na figura, o gene A é autorregulado, o gene B tem regulação mista e o gene C é regulado externamente.

A Figura 5 representa a codificação das proteínas A, B, C respectivamente a partir dos genes A, B, C e mostra os três tipos de regulações⁹ - O gene A é autorregulado, pois sofre regulação a partir da proteína A expressada pelo próprio gene A. O gene B tem regulação mista, pois é regulado externamente pela proteína A e autorregulado pela proteína B expressada pelo próprio gene B. Já o gene C é regulado apenas externamente pela proteína A. As linhas contínuas representam as etapas do processo de expressão gênica (transcrição e tradução) e as linhas tracejadas indicam as regulações exercidas na expressão gênica de cada gene.

1.7 O RUÍDO NA EXPRESSÃO GÊNICA

As reações químicas ocorridas no interior celular são intrinsecamente aleatórias no seu número de produtos. Isto ocorre por conta de as moléculas regentes estarem presentes em pequenos números de cópias e as colisões entre reagentes que causam as reações ocorrem a intervalos de tempo irregulares. A expressão gênica consiste de uma grande cadeia de reações químicas e reflete essa aleatoriedade. Em virtude de a expressão gênica ser um processo estocástico de interações e reações moleculares e envolvendo DNA, RNA e proteínas que ocorre em *bursts*¹⁰, existe grande variabilidade no número de seus produtos, periodicidade de ocorrência e momentos em que ocorre. Essa variabilidade é denominada ruído da expressão gênica. A variabilidade referente à flutuação aleatória do número de produtos gênicos (mRNAs ou proteínas) é denominada variabilidade de amplitude, enquanto que a variabilidade de frequência refere-se à periodicidade da ocorrência da expressão gênica e a variabilidade de momento refere-se a variação dos instantes em que ela ocorre. (22) (2)

⁹ Esta classificação pode ser usada tanto para regulação positiva quanto negativa.

¹⁰ pulsos

O ruído pode eventualmente ser suficiente para criar uma mudança de um estado estável para outro, gerando diferenciação entre organismos multicelulares ou a sobrevivência em ambientes instáveis (23), bem como afetar a evolução das células-tronco, provocar o desenvolvimento de tumores cancerígenos e alterar a resposta a fatores indutores de apoptose¹¹ (24).

O ruído é classificado a partir de suas fontes geradoras como extrínseco e intrínseco (22). Ruídos extrínsecos são originados de fontes extrínsecas, isto é, fontes que afetam a expressão de diversos genes e podem ser descritos como a principal fonte de variabilidade na expressão gênica devido ao maior tempo de ocorrência se comparado ao tempo dos fatores intrínsecos (23) (25).

Dentre os fatores que podem causar ruído extrínseco é possível citar a variação na disponibilidade de maquinário biológico para o processo de expressão gênica¹², a propagação de ruídos intrínsecos na expressão gênica de proteínas regulatórias¹³, microflutuações no ambiente celular e divisão celular assimétrica com eventual separação do mRNA e material proteico (22).

Figura 6 - Ruídos extrínsecos e intrínsecos – A presença de ruídos foi demonstrada através da quantificação da variabilidade da expressão gênica de células de *E. coli* contendo duas cópias do mesmo gene. Um dos genes é associado a um promotor que codifica uma proteína fluorescente de cor vermelha enquanto o outro gene é associado a um promotor que codifica a luz fluorescente de cor verde. A ausência de ruído faria com que os genes produzissem quantidades iguais de proteínas com fluorescência verde e vermelha resultando em uma célula amarela. Mas o que se observa no experimento é a presença de células vermelhas e verdes, indicando que o nível de expressão não é idêntico nas células e que em algumas o gene marcado com verde é mais ativamente expresso enquanto em outras o mesmo ocorre com o gene marcado com vermelho. Ruídos extrínsecos afetam a expressão de ambas as cópias dos genes em uma célula (Gráfico B superior), já ruídos intrínsecos afetam cada cópia do gene de forma isolada (gráfico B inferior). Figura adaptada de (26).

Já ruídos intrínsecos são originados de fontes relacionadas ao gene individualmente e têm sua origem em eventos estocásticos no processo de expressão gênica associados à baixa

¹¹ Morte celular programada

¹² Por exemplo, variação na quantidade de ribossomos.

¹³ A organização dos genes em rede leva à propagação do ruído para os genes seguintes, dessa forma, ruídos intrínsecos em um gene podem causar ruído extrínseco nos genes seguintes.

quantidade de moléculas reguladoras. Um exemplo de fonte de ruído intrínseco é a estrutura do sítio promotor, sua localização no cromossomo e a arquitetura do núcleo celular. (23) (25).

A Figura 6 ilustra o experimento realizado por (27) que demonstra a presença de ruídos extrínsecos e intrínsecos através da quantificação da variabilidade da expressão gênica de células de *E. coli*. contendo duas cópias do mesmo gene.

1.7.1 Análise quantitativa do ruído na expressão gênica

O ruído na expressão gênica necessita ser analisado quantitativamente utilizando ferramentas da teoria das probabilidades. Representaremos por n o número de um determinado produto gênico, por exemplo, o número de moléculas de mRNA. O número médio de moléculas de mRNA será representado por $\langle n \rangle$ e a variância em n por σ^2 em que:

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 \quad (1)$$

e o desvio padrão é dado por:

$$\sigma = \sqrt{\langle n^2 \rangle - \langle n \rangle^2} \quad (2)$$

A variância e o desvio padrão permitem quantificar a magnitude das flutuações aleatórias.

Para que possamos avaliar a magnitude das flutuações aleatórias em relação a $\langle n \rangle$ podemos utilizar duas quantidades: O fator de Fano, denotado por F e o coeficiente de variação indicado por C_v .

O fator de Fano possibilita identificar o quanto uma distribuição de probabilidades em uma variável aleatória difere da distribuição de Poisson e é definido como:

$$F = \frac{\sigma^2}{\langle n \rangle} \quad (3)$$

Se $F=1$ temos uma distribuição de Poisson, se $F>1$ ($F<1$), temos uma distribuição Super-Poisson (ou sub-Poisson), que é mais larga (ou mais estreita) que a distribuição de Poisson (28).

O ruído da expressão gênica também pode ser obtido em função do coeficiente de variação (2), isto é: que mede a percentagem do desvio padrão em relação ao número médio.

$$C_v = \frac{\sigma}{\langle n \rangle} \quad (4)$$

1.8 ORGANISMOS MODELO - *DROSOPHILA MELANOGASTER*

Estudos relativos à expressão gênica em processos de desenvolvimento utilizam organismos modelos escolhidos principalmente por questões históricas e pela facilidade de estudo e interesse biológico. Alguns exemplos de organismos modelos são 1 – As rãs do gênero *Xenopus*, cujas espécies estão relacionadas entre si pela duplicação ou triplicação do genoma. 2 - O camundongo, cuja preferência para estudos relativos à mamíferos vertebrados ocorre por ser pequeno, forte e de rápida geração, 3 - A mosca da fruta *Drosophila melanogaster*, que alcança a fase adulta nove dias após o início do desenvolvimento, tem baixo custo de criação, tem mecanismos genéticos mais facilmente identificáveis do que os vertebrados e possui cromossomos gigantes visíveis em muitas células. (1) (20)

A *D. melanogaster* é um dos organismos modelo mais bem compreendido em nível genético devido à grande quantidade de estudos genéticos aliados à facilidade de realizar manipulações genéticas e microcirúrgicas. A similaridade dos genes de controle do seu desenvolvimento com os genes de controle do desenvolvimento dos vertebrados faz com que essa compreensão possa ser estendida também aos vertebrados (20).

No seu processo de desenvolvimento, a *Drosophila* eclode do ovo como uma larva que posteriormente sofre metamorfose transformando-se no inseto adulto. Tanto os óvulos como o embrião da *Drosophila* possuem formato elipsoide com dois eixos independentes, o eixo antero-posterior (AP) e o eixo dorsoventral (DV). A larva possui segmentação regular ao longo do eixo antero-posterior constituído por uma extremidade contendo a região que dará origem à cabeça, seguida pela região que dará origem aos segmentos torácicos e posteriormente à região que dará origem aos segmentos abdominais. No eixo dorsoventral existe quatro regiões que darão origem à organização dorsoventral do corpo da larva. (20). Esse processo está representado na figura 7.

A larva é formada por um processo de embriogênese cuja duração desde o início, na fecundação, até o primeiro estágio larval tem a duração de 24 horas. Após a fertilização e a fusão dos núcleos do espermatozoide e do óvulo, ocorrem 13 divisões mitóticas, mas sem a clivagem do citoplasma, gerando um embrião que, no início do processo de desenvolvimento, é formado por com uma célula única contendo diversos núcleos difusos em todo o embrião¹⁴.

¹⁴ Deve-se considerar que, apesar de ter ocorrido apenas a Cariocinese e não a Citocinese, o citoplasma não é totalmente uniforme e cada núcleo está contido dentro de seu próprio território determinado por um citoesqueleto de proteínas (30)

Figura 7 - Desenvolvimento da *Drosophila melanogaster*. Adaptado de (29)

A especificação dos tipos de células ao longo do eixo AP e DV ocorre pela interação dos materiais contidos no citoplasma dentro de uma única célula e a inicialização das diferenças ao longo dos eixos é controlada pela posição do ovo dentro do ovário da mãe. Após os nove primeiros ciclos, que duram cerca de 10 minutos, ocorre a movimentação dos núcleos para a periferia do embrião, formando a blastoderme sincicial. Posteriormente as membranas crescem a partir dessa superfície, envolvendo os núcleos e formando as células e tornando a blastoderme verdadeiramente celular após 13 ciclos. Dos núcleos existentes no citoplasma, cerca de quinze não dão origem a células no blastoderme e sim a células polares que irão formar as células germinativas. (20) (30)

Existe cerca de 6000 núcleos no último ciclo (14A) antes da gastrulação, processo que se inicia cerca de 3 horas após a fertilização e no qual os tecidos endodérmicos e mesodérmicos movem-se para as posições que ocuparão dentro do embrião, mantendo a ectoderme como a camada externa. Esse movimento é caracterizado pela invaginação da mesoderme para formar uma fenda denominada arquênteron (2) (20).

É no ciclo 14A que se definem os domínios relacionados à formação e desenvolvimento dos 14 segmentos do corpo do indivíduo adulto (3 segmentos na cabeça, 3 no tórax e 8 no abdômen) (2).

O primeiro grupo de genes da hierarquia de atividades gênicas do processo de desenvolvimento são os genes herdados da mãe e que determinarão diferentes características ao longo do eixo AP do óvulo. A partir dessas diferentes características ocorrerá a formação

das extremidades anterior (cabeça) e posterior da larva. Neste grupo de genes maternos é possível destacar a expressão de 4 genes de um total de 50. São eles: *bicoid*, *hunchback*, *nanos* e *cauda*. (20).

O mRNA do gene *bicoid* fica fixado ao citoesqueleto na extremidade anterior do óvulo não fecundado e após a fecundação é expressado como proteína *bicoid*, a qual espalha em forma de gradiente por todo o eixo AP do ovo. O gene *bicoid* atua como um morfógeno, ativando certos genes zigóticos em diferentes níveis de concentração ao longo do eixo AP. Dessa forma, a gradiente de sua distribuição determinará as informações necessárias para a padronização do eixo AP e possui maior presença na extremidade anterior do eixo.

No extremo posterior do óvulo está fixado o mRNA materno do gene *nanos*, que sintetizará a proteína *nanos* após a fecundação e também se distribuirá em forma de gradiente ao longo do eixo AP, mas com maior concentração na extremidade posterior. A função da proteína *nanos* é reprimir de forma gradual a expressão do gene materno *hunchbacks*, cuja proteína resultante também é sintetizada a partir do gene zigótico *hunchback*. A distribuição do gene também ocorre em forma de gradiente a partir da região anterior do óvulo. Como o gene *hunchback* materno está distribuído igualmente por todo o eixo AP, sua inibição pela proteína *nano* permite que a distribuição da proteína *hunchback* não tenha o gradiente afetado pela presença do gene materno.

A proteína caudal, sintetizada a partir do mRNA materno do gene caudal, tem seu gradiente posterior-anterior definido pela ação inibidora da proteína *bicoid*, cuja distribuição em gradiente faz com que a proteína caudal tenha também uma distribuição em gradiente, mas com a concentração inversamente proporcional a concentração da proteína *bicoid*.

As estruturas terminais anterior (ácron) e posterior (telso) são determinadas pela expressão do gene materno *torso*, que só ocorre nessas extremidades, pois a proteína ligante está presente em pequena quantidade e apenas nessas regiões.

No eixo dorso ventral a padronização ocorre regida pela proteína materna *dorsal*, a qual é distribuída inicialmente de forma uniforme ao longo do eixo e posteriormente, sob a influência do receptor *toll* ventralmente ativado, passa a ter uma distribuição em gradiente com maior concentração nos núcleos ventrais e quase nenhuma concentração nos núcleos dorsais.

A segmentação da larva da *D. melanogaster* se dá inicialmente pela definição dos 14 módulos básicos e inicialmente similares denominados parassegmentos, os quais são delimitados pela expressão dos genes *pair-rule* e darão origem aos segmentos corporais da larva. Os genes *pair-rule* são expressos em uma série de sete faixas transversais ao eixo AP e

de forma alternada, sendo que alguns definem os parassegmentos de número ímpar, como o gene *even-skipped*, enquanto outros definem os parassegmentos de número par, como o gene *fushi tarazu* (20).

Seguindo a hierarquia da atividade gênica, o desenvolvimento prossegue com a atuação dos genes de *polaridade do segmento*, os quais terão suas atividades sucedidas pelos genes *homeóticos* responsáveis pela identidade de cada segmento do embrião (2).

Nesse trabalho estamos interessados em compreender como a aleatoriedade afeta a expressão do gene *even-skipped* durante a formação da faixa 2. Em particular, mostramos nesse trabalho que o modelo binário pode ser uma importante ferramenta para a compreensão da regulação desse gene.

2 UM MODELO ESTOCÁSTICO PARA A TRANSCRIÇÃO GÊNICA

A transcrição gênica é um conjunto de reações químicas envolvendo uma baixa quantidade de moléculas reagentes presentes na célula. Essa condição dá origem a flutuações nas taxas das reações moleculares, fazendo com que elas ocorram em forma de *bursting* de intensidades variadas e a intervalos de tempo também variados. Essas condições tornam inadequada uma abordagem determinística do processo e indicam o uso de uma abordagem estocástica como a melhor estratégia para o estudo de fenômenos de regulação gênica, particularmente o modelo de Sasai e Wolynes, o qual considera o gene binário (2) (12) (13) (25) (31) (32) (33) (34) (35).

Um gene pode ser denominado binário se forem considerados dois estados de ativação – ligado e desligado – relacionados respectivamente a um estado favorável ou desfavorável à transcrição. Estes estados são determinados por um tipo de proteína que pode se ligar ou se desligar do sítio operador dos procariotos ou sítio de ligação nos eucariotos (2). O estado favorável é definido como **ativo** e o estado desfavorável como **reprimido**. Quando o gene está no estado ativo, a taxa de transcrição ocorre à taxa máxima. Já quando o gene está no estado reprimido, a taxa de transcrição ocorre à taxa mínima

Neste trabalho apresentamos um modelo binário para a transcrição gênica com regulação externa onde foi considerado um sistema composto por um gene binário e duas variáveis: 1 – A quantidade de moléculas de mRNA, indicada por n , com $n = 0, 1, 2, \dots$, e 2 - O estado do gene que é denotado por $m = \text{ON}$ ou OFF (0 ou 1). O modelo também considera o gene como externamente regulado e que ocorre a degradação das moléculas de mRNA.

Neste modelo binário, a transição do estado do sítio promotor do gene binário de ativo para reprimido ocorre aleatoriamente e com taxas de transição f (do estado reprimido para ativo) e h (do estado ativo para reprimido). A transcrição ocorre a uma taxa máxima κ quando o gene está ativo e a uma fração χ dessa taxa máxima ($\chi\kappa \mid 0 \leq \chi < 1$) quando o gene está reprimido, gerando uma quantidade n de moléculas de mRNA, as quais degradarão a uma taxa ρ . Neste trabalho o valor de χ foi considerado 0.

O modelo do gene binário regulado externamente é representado na Figura 8A. A dinâmica do processo de transcrição do gene binário externamente regulado também pode ser representada utilizando o modelo cinético da Figura 8B. A partir da leitura deste modelo, identifica-se que em cada instante de tempo t existe um número n de moléculas de mRNA e o gene encontra-se no estado **ativo** ou **reprimido**

Figura 8 - Representação do modelo binário e modelo cinético. Na representação do modelo binário em A, os círculos indicam um gene que pode estar no estado ativo (disco branco) ou reprimido (disco laranja). A taxa de transição do estado ativo para o reprimido é indicada por h enquanto a taxa de transição do estado reprimido para o estado ativo é indicada por f . O processo de transcrição gera n moléculas de mRNA, indicadas pelos retângulos com cantos arredondados, e ocorre a uma taxa máxima k quando o gene está ativo e a uma fração χ de κ quando o gene está no estado reprimido. As moléculas de mRNA por sua vez degradam a uma taxa ρ . No modelo cinético representado em B, $\emptyset$ representa a inexistência de mRNA, M representa o mRNA produzido pelo gene em estado ativo a uma taxa κ e pelo gene em estado reprimido a uma taxa $\chi\kappa$. A taxa de execução do processo inverso, de aniquilação de mRNA, é representada por ρ . R indica o estado ativo do sítio promotor do gene e o estado reprimido é indicado por A . As taxas de transição de estado de reprimido para ativo e do estado ativo para reprimido são representadas respectivamente por h e f .

Deve-se ressaltar que este esquema é uma supersimplificação do conjunto de reações que ocorrem durante o processo de transcrição gênica. Essa abordagem de selecionar os eventos-chave do fenômeno investigado é denominada abordagem efetiva.

Dado que o modelo apresentado possui dinâmica estocástica, pode-se afirmar que no instante de tempo t e com uma quantidade n de moléculas de mRNA no interior celular, existe uma probabilidade $\alpha_n(t)$ de encontrar o gene no estado ativo e uma probabilidade $\beta_n(t)$ do gene estar no estado reprimido.

Para simplificar omitiremos a dependência temporal das probabilidades, isto é, $\alpha_n \equiv \alpha_n(t)$ e $\beta_n \equiv \beta_n(t)$. Isto é possível se considerarmos que os fatores que determinam a probabilidade do gene estar no estado ativo ou reprimido não são modificados ao longo do tempo, observando-se apenas variação dessas probabilidades apenas quando ocorre variação no número de moléculas de mRNA. Dessa forma, a probabilidade de que um gene esteja no estado ativo quando estiverem presentes n moléculas de mRNA no citoplasma celular é indicada por α_n enquanto β_n indicará a probabilidade de que o gene esteja em estado reprimido, não importando os instantes em que n ocorra.

A dinâmica das probabilidades α_n e β_n é governada por equações mestras que incluem termos relacionados à alteração do número de moléculas e termos relacionados à alteração do nível de transcrição. Nas equações, n representa o número de moléculas de mRNA presentes

no citoplasma celular, κ é a taxa(máxima) de transcrição de moléculas de mRNA com o gene no estado ativo, h é a taxa de transição do estado ativo para o reprimido, f é a taxa de transição do estado reprimido para o ativo, ρ é a taxa de degradação das moléculas de mRNA e χ é a fração da taxa máxima de transcrição de moléculas de mRNA que indica a taxa de transcrição com o gene reprimido.

As equações mestras podem ser representadas da seguinte forma:

$$\frac{d}{dx}\alpha_n = \kappa(\alpha_{n-1} - \alpha_n) + \rho[(n+1)\alpha_{n+1} - n\alpha_n] - h\alpha_n + f\beta_n \quad (5)$$

$$\frac{d}{dx}\beta_n = \chi\kappa(\beta_{n-1} - \beta_n) + \rho[(n+1)\beta_{n+1} - n\beta_n] + h\alpha_n - f\beta_n \quad (6)$$

Considerando-se que o número de moléculas de mRNA varia de $n-1$ para n a uma taxa k ($\chi\kappa$) quando o gene está no estado ativo (reprimido) devido à transcrição e também variam de n para $n-1$ a uma taxa ρ devido ao decaimento, podemos analisar as equações termo a termo.

Os termos $\kappa(\alpha_{n-1} - \alpha_n)$ e $\chi\kappa(\beta_{n-1} - \beta_n)$ referem-se à síntese de moléculas de mRNA quando o gene encontra-se respectivamente no estado ativo ou reprimido. A variação da probabilidade do gene estar no estado ativo (reprimido) quando o número de moléculas de mRNA varia de $n-1$ para n em decorrência da síntese é representada por $\alpha_{n-1} - \alpha_n$ ($\beta_{n-1} - \beta_n$) e a taxa de síntese de moléculas de mRNA quando o gene está no estado ativo (reprimido) é representada por κ ($\chi\kappa$).

Os termos $\rho[(n+1)\alpha_{n+1} - n\alpha_n]$ e $\rho[(n+1)\beta_{n+1} - n\beta_n]$ referem-se à degradação das moléculas de mRNA. A variação da probabilidade do gene estar no estado ativo (reprimido) quando o número de moléculas de mRNA varia de $n+1$ para n devido à degradação é representada por $(n+1)\alpha_{n+1} - n\alpha_n$ ($(n+1)\beta_{n+1} - n\beta_n$) e a taxa de degradação de moléculas de mRNA quando o gene esta no estado ativo ou reprimido é representada por ρ .

Os termos $h\alpha_n$ e $f\beta_n$ referem-se à probabilidade do gene estar no estado ativo (reprimido) quando estão presentes n moléculas de mRNA é representada por α_n (β_n) e a taxa de mudança do gene do estado reprimido para o estado ativo (ativo para reprimido) é representada por h (f).

A conservação da probabilidade nas equações (5) e (6) é representada da seguinte forma:

$$\sum_{n=0}^{\infty} (\alpha_n(t) + \beta_n(t)) = \sum_{n=0}^{\infty} \alpha_n(t) + \sum_{n=0}^{\infty} \beta_n(t) = 1 \quad (7)$$

2.1 SOLUÇÃO EXATA DAS EQUAÇÕES MESTRAS

É possível calcular a solução exata das equações mestras (5) e (6). Para isso utilizamos o método das funções geratrizes das probabilidades α_n e β_n , que permitem reescrever as funções na forma de equações diferenciais. As distribuições marginais¹⁵ das variáveis n e m podem ser obtidas utilizando as funções geratrizes $\alpha(z, t)$ e $\beta(z, t)$:

$$\alpha(z, t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n \quad ; \quad \beta(z, t) = \sum_{n=0}^{\infty} \beta_n(t) z^n \quad (8)$$

Para reformularmos as equações mestras iniciamos multiplicando seus dois lados por z^n e em seguida aplicamos a somatória de 0 à infinito. Após manipulações e a utilização da definição das funções geratrizes obtemos um par de equações diferenciais parciais acopladas, em que omitimos a dependência em (z, t) , isto é, $\alpha = \alpha(z, t)$ e $\beta = \beta(z, t)$. Assim, temos:

$$\frac{\partial \alpha}{\partial t} = (z - 1) \left(\kappa \alpha - \rho \frac{\partial \alpha}{\partial z} \right) - h \alpha + f \beta \quad (9)$$

$$\frac{\partial \beta}{\partial t} = (z - 1) \left(\chi \kappa \beta - \rho \frac{\partial \beta}{\partial z} \right) + h \alpha - f \beta \quad (10)$$

Consideraremos que não haverá transcrição de moléculas de mRNA quando o gene está no estado reprimido, o que implica em considerar $\chi = 0$ nas equações (6) e (10).

2.2 DINÂMICA DAS PROBABILIDADES DO GENE ATIVO OU REPRIMIDO

Para recuperar as probabilidades de serem encontradas n moléculas de mRNA com o gene em estado ativo ou reprimido utiliza-se as n -ésimas derivativas das funções geratrizes das probabilidades em (8), conforme as equações em (11) abaixo¹⁶.

$$\alpha_n(t) = \frac{1}{n!} \left. \frac{\partial^n \alpha}{\partial z^n} \right|_{z=0} \quad ; \quad \beta_n(t) = \frac{1}{n!} \left. \frac{\partial^n \beta}{\partial z^n} \right|_{z=0} \quad (11)$$

¹⁵ Distribuição da probabilidade de vários valores de variáveis. Denomina-se “marginal” porque são obtidas a partir da soma de valores ao longo de linhas ou colunas de uma tabela, com resultados escritos nas margens da tabela. Ex.:

	Condição A	Condição B	Probabilidade Marginal
Probabilidade de ocorrer X	0.2	0.5	0.7
Probabilidade de ocorrer Y	0.1	0.2	0.3
Total	0.3	0.7	1

¹⁶ No apêndice 3 está o desenvolvimento das equações em (11) a partir das equações de (8):

A distribuição marginal da probabilidade de encontrar o gene em estado ativo ou reprimido, independente do número de moléculas de mRNAs no interior celular, indicado por A e B, é obtida a partir das equações (8), considerando $z = 1$:

$$A(t) = \alpha(1, t) = \sum_{n=0}^{\infty} \alpha_n(t) \quad ; \quad B(t) = \beta(1, t) = \sum_{n=0}^{\infty} \beta_n(t) \quad (12)$$

A distribuição marginal da probabilidade de encontrar n moléculas de mRNA no interior celular, independente do estado do gene estar ativo ou reprimido, é indicada por $\varphi_n(t)$ e obtida por:

$$\varphi_n(t) = \alpha_n(t) + \beta_n(t) \quad (13)$$

Também podemos obter a função geratriz das probabilidades de encontrar n moléculas de mRNA no interior celular, que fica definida como:

$$\varphi(z, t) = \sum_{n=0}^{\infty} \varphi_n(t) z^n \quad (14)$$

em que $\varphi(z, t)$ é a função geratriz.

Para recuperar as probabilidades de encontrar n moléculas de mRNA independentemente do estado do gene utiliza-se as n -ésimas derivativas das funções geratrizes das probabilidades:

$$\varphi_n(t) = \frac{1}{n!} \left. \frac{\partial^n \varphi}{\partial z^n} \right|_{z=0} \quad (15)$$

A conservação das probabilidades representada em (7) também pode ser representada no formalismo das funções geratrizes como segue:

$$\varphi(1, t) = \alpha(1, t) + \beta(1, t) = A(t) + B(t) = 1 \quad (16)$$

Em que utilizamos a definição das equações de (8).

Dinâmica das probabilidades de o gene ficar ativado

Podemos efetuar as probabilidades de o gene estar em estado ativo ou reprimido aplicando $z=1$ nas equações (9) e (10). Utilizando a conservação de probabilidades temos que $B(t) = 1 - A(t)$ e que a probabilidade de o gene estar ativado será governada pela seguinte EDO:

$$\frac{dA}{dt} = -(h + f)A + f \quad (17)$$

A solução dessa equação é dada por:

$$A(t) = \frac{f}{f+h} + \left(A(0) - \frac{f}{f+h} \right) e^{-(f+h)t} \quad (18)$$

em que $A(0)$ é a condição inicial.

Note que quando o tempo vai a infinito a probabilidade de o gene estar ativado assume um valor constante dado por:

$$A = \frac{f}{f+h} \quad (19)$$

que é a probabilidade estacionária de o gene estar ativado

2.3 DINÂMICA DO NÚMERO MÉDIO DE MOLÉCULAS DE mRNA

A tendência central de uma distribuição de probabilidades é caracterizada pelo primeiro momento dessa distribuição.

É possível obter os p -ésimos momentos¹⁷ das distribuições das probabilidades marginais α_n e β_n através das p -ésimas derivativas das suas respectivas equações geratrizes, isto é:

$$\langle n_\alpha^p \rangle = \left(z \frac{\partial}{\partial z} \right)^p \alpha \Big|_{z=1} \quad ; \quad \langle n_\beta^p \rangle = \left(z \frac{\partial}{\partial z} \right)^p \beta \Big|_{z=1} \quad (20)$$

É possível obter então o valor médio do número de moléculas de mRNA quando o gene está no estado ativo ou no estado reprimido através da obtenção do primeiro momento nas equações (13).

Considerando a distribuição marginal das probabilidades em (13), é possível obter também o valor médio do número de moléculas de mRNA, independentemente do estado do gene, obtendo o primeiro momento da equação em (21):

$$\langle n^p \rangle = \left(z \frac{\partial}{\partial z} \right)^p \varphi \Big|_{z=1} \quad (21)$$

Sendo assim, podemos aplicar a primeira derivada em relação a z às equações (9) e (10) e, após somar os resultados, considerando $z=1$ e $\chi=0$ teremos:

$$\frac{d}{dx} \langle n \rangle_t = \kappa A(t) - \rho \langle n \rangle(t) \quad (22)$$

¹⁷ Os primeiro, segundo, terceiro e quarto momentos caracterizam a tendência central, dispersão, assimetria e curtose, respectivamente, de uma distribuição de probabilidade. O desenvolvimento está no Apêndice D.

onde foi considerado $p=1$ na equação (21). E solução dessa EDO pode ser obtida se consideramos a solução exata para $A(t)$, tal que $\langle n \rangle(t)$ fica:

$$\langle n \rangle(t) = \mu + Ue^{-\rho t} + Ve^{-\rho \varepsilon t} \quad (23)$$

em que

$$\mu = NA \quad ; \quad U = \langle n \rangle(0) - \mu - V \quad ; \quad V = N \frac{(A(0) - A)}{1 - \varepsilon} \quad (24)$$

sendo os parâmetros N e ε definidos como:

$$N = \frac{\kappa}{\rho} \quad ; \quad \varepsilon = \frac{f + h}{f} \quad (25)$$

e $\langle n \rangle(0)$ e $A(0)$ são, respectivamente, os valores iniciais do número médio de moléculas de mRNA e da probabilidade de o gene estar ativado.

O parâmetro N indica o número de moléculas de mRNA em regime estacionário quando a probabilidade de o gene estar ativado é 1. Já o parâmetro ε indica a rapidez com que o gene chaveia em relação à frequência de degradação do mRNA. Se $\varepsilon \ll 1$ o número de moléculas de mRNA em um dado instante indica o estado do gene enquanto que se $\varepsilon \gg 1$ o gene realiza múltiplos chaveamentos durante a meia vida do mRNA e não é possível distinguir o estado do gene apenas pela observação do número de moléculas de mRNA.

2.4 REGIME ESTACIONÁRIO

No regime estacionário temos que as derivadas temporais das probabilidades vão à zero. Isso implica de as derivadas temporais das funções geratrizes serem nulas, isto é:

$$\frac{\partial \alpha}{\partial t} = 0 \quad ; \quad \frac{\partial \beta}{\partial t} = 0 \quad (26)$$

Dessa forma, as equações (9) e (10) passam a assumir as seguintes formas:

$$0 = (z - 1) \left(\kappa \alpha - \rho \frac{d\alpha}{dz} \right) - h\alpha + f\beta \quad (27)$$

$$0 = (z - 1) \left(-\rho \frac{d\beta}{dz} \right) + h\alpha - f\beta \quad (28)$$

em que $\alpha = \alpha(z)$ e $\chi = 0$.

Para calcularmos as soluções exatas desse par de EDO acopladas, iniciamos solucionando a equação (27) para β , o que resulta em:

$$\beta = -\frac{1}{f} (z - 1) \left(\kappa \alpha - \rho \frac{d\alpha}{dz} \right) + \frac{h}{f} \alpha \quad (29)$$

Em seguida substituímos a expressão para β da equação (29) na equação (28), o que resulta na seguinte equação diferencial ordinária de ordem 2:

$$(z-1)\frac{d^2\alpha}{dz^2} + [1 + \varepsilon - N(z-1)]\frac{d\alpha}{dz} - N(1 + \varepsilon A)\alpha = 0 \quad (30)$$

em que $\varepsilon = \frac{f+h}{\rho}$ e A e N foram definidos acima.

Essa equação é caracterizada pela existência de duas singularidades, uma regular em $z = 1$ e outra irregular em $z = \infty$. Assim sendo, ela é uma equação hipergeométrica confluyente cuja solução é dada em termos da função de Kummer¹⁸. Explicitamente para:

$$\alpha(z) = AM(1 + \varepsilon A, 1 + \varepsilon, N(z-1)) \quad (31)$$

em que M indica a função de Kummer, que é definida como

$$M(a, b, \eta) = \sum_{n=0}^{+\infty} \frac{(a)_n}{(b)_n} \frac{\eta^n}{n!} \quad (32)$$

em que o termo $(a)_n$ simboliza a função de Pochhammer¹⁹.

A função geratriz das probabilidades estacionárias de o gene ficar em estado reprimido é obtida a partir da aplicação da equação (29) sobre a função geratriz da equação (31), resultando em:

$$\beta(z) = (1 - A)M(\varepsilon A, 1 + \varepsilon, N(z-1)) \quad (33)$$

e a soma das duas equações resulta na probabilidade de encontrarmos n moléculas de mRNA no interior celular, independente do estado do gene. Nominalmente temos:

$$\varphi(z) = M(\varepsilon A, \varepsilon, N(z-1)) \quad (34)$$

¹⁸ No Apêndice E está o desenvolvimento da equação (31) a partir da equação (27):

¹⁹ Também denominada “fatorial crescente”. É dada por $(a)_0 = 1$ e $(a)_n = a(a+1)(a+2) \dots (a+n-1)$.

3 APLICAÇÃO AO PROCESSO DE DESENVOLVIMENTO DA *D. MELANOGASTER*

Um exemplo de processo biológico de desenvolvimento que ocorre de forma precisa apesar das flutuações aleatórias é o processo de segmentação e formação de faixas de produtos gênicos da *D. melanogaster*. A identificação de reagentes que regulam a formação da faixa 2 do mRNA *even-skipped (eve)* da *D. melanogaster* pode ser realizada por técnicas experimentais, como a análise de dados de imunofluorescência²⁰. Entretanto, é fundamental uma abordagem teórica na determinação das interações entre proteínas reguladoras e sítios específicos do DNA que controlam a expressão do gene da *eve*, dentre outros motivos, para pesquisar os efeitos do ruído ou de supressão de ruído em processos de desenvolvimento (36).

A aplicação do modelo estocástico binário descrito no capítulo 3 permite a reprodução do comportamento do processo de desenvolvimento da faixa 2 em termos de quantidade n de moléculas de mRNA ao longo do tempo de execução do processo. Também permite a identificação de valores médios $\langle n \rangle$, da mediana e da moda da quantidade de mRNA, bem como quantificar as flutuações em n . O uso do modelo binário também permite estabelecer critérios experimentais dedicados à investigação da regulação da expressão do *eve 2*. (37)

3.1 SIMULAÇÕES

Neste trabalho foram executadas simulações utilizando o algoritmo de Gillespie para simulação estocástica exata de reações químicas acopladas. Dessa forma é possível identificar o nível populacional de moléculas resultantes de uma reação química presente em um volume V a partir das correlações existentes entre os reagentes, as flutuações dessas correlações e o valor inicial do nível populacional. (38)

O algoritmo de Gillespie é derivado do método de Monte Carlo (2) e está baseado na obtenção de dois números aleatórios em cada iteração. Um dos números é utilizado para definir o instante em que a iteração ocorre e o outro para determinar qual tipo de reação ocorrerá naquela iteração.

Aplicado ao modelo do gene binário, o algoritmo de Gillespie pode ser implementado para identificar a quantidade de moléculas de mRNA em cada instante durante um período de tempo, considerando-se o gene autorregulado, regulado externamente ou o gene com

²⁰ A análise de dados de imunofluorescência converte a intensidade da imunofluorescência em número de mRNA (36)

regulação mista (representado na figura 1). Para as simulações do presente trabalho foi considerado apenas o gene regulado externamente, cujo modelo binário da expressão gênica está representado na figura 2.

A implementação do algoritmo de Gillespie neste trabalho é descrita detalhadamente no fluxograma representado no Apêndice A,

O algoritmo de Gillespie destina-se à simulação de um processo estocástico dado em termos de uma cadeia de Markov discreta a tempo discreto. A cada iteração, ou transição, ocorrerá uma transformação do sistema em que podem ocorrer uma síntese ou degradação de mRNA e o chaveamento de gene entre os estados ativado e reprimido. Indicaremos as iterações por i , o número de moléculas de mRNA na i -ésima iteração é dado por n , com $n=0,1,\dots$ e o estado do gene é dado por ON ou OFF. O algoritmo de Gillespie também possibilita calcular o intervalo de tempo contínuo entre duas iterações, obtido como uma variável aleatória governada por uma distribuição exponencial. Na tabela abaixo resumimos o algoritmo de Gillespie:

1. Definição de um número pseudo-aleatório a .
2. Cálculo do instante t_i (onde $i = 0,1,\dots,500$) em que ocorre a iteração:

$$\begin{cases} \text{Se } i = 0 \rightarrow t_i = 0 \\ \text{Se } i \neq 0 \rightarrow \begin{cases} \text{Se gene ativo} \rightarrow t_i = t_{i-1} + \frac{1}{\kappa + \rho n + h} \log \frac{1}{a} \\ \text{Se gene reprimido} \rightarrow t_i = t_{i-1} + \frac{1}{\chi \kappa + \rho n + f} \log \frac{1}{a} \end{cases} \end{cases}$$

3. Definição da reação que ocorre no instante t_i de acordo com a tabela 1 abaixo:

TABELA 1 - REAÇÃO NO INSTANTE t_i

Se estado do gene = ativo	Se estado do gene = Reprimido
$a(\kappa + \rho n + h) > \kappa \Rightarrow n + 1$	$a(\chi \kappa + \rho n + h) > \chi \kappa \Rightarrow n + 1$
$\left. \begin{array}{l} a(\kappa + \rho n + h) \leq \kappa \\ e \\ a(\kappa + \rho n + h) > \kappa + \rho n \end{array} \right\} \Rightarrow n - 1$	$\left. \begin{array}{l} a(\chi \kappa + \rho n + h) \leq \chi \kappa \\ e \\ a(\chi \kappa + \rho n + h) > \chi \kappa + \rho n \end{array} \right\} \Rightarrow n - 1$
$\left. \begin{array}{l} a(\kappa + \rho n + h) \leq \kappa \\ e \\ a(\kappa + \rho n + h) \leq \kappa + \rho n \\ e \\ a(\kappa + \rho n + h) > \kappa + \rho n + h \end{array} \right\} \Rightarrow \text{Muda estado do gene para Reprimido}$	$\left. \begin{array}{l} a(\chi \kappa + \rho n + f) \leq \chi \kappa \\ e \\ a(\chi \kappa + \rho n + f) \leq \chi \kappa + \rho n \\ e \\ a(\chi \kappa + \rho n + f) > \chi \kappa + \rho n + f \end{array} \right\} \Rightarrow \text{Muda estado do gene para Ativo}$

Para a definição dos valores dos parâmetros κ, χ, ρ, f e h foram inicialmente definidos o número N de moléculas de mRNA caso o gene estivesse exclusivamente ON e a taxa de degradação ρ . Foram então definidos os valores de f e h para obter valores de probabilidade (A) de o gene estar no estado ON quando o sistema estiver em equilíbrio e para obter valores

para a razão ε entre as taxas de mudança de estado (h e f) e a taxa de degradação das moléculas de mRNA (ρ) conforme a tabela 1

Os valores de A e ε foram obtidos usando as equações abaixo:

$$A = \frac{f}{f + h} \quad N = \frac{\kappa}{\rho} \quad \varepsilon = \frac{f + h}{\rho} \quad (35)$$

conforme definido em (19).

3.1.1 A influência dos parâmetros na simulação da expressão gênica

Considerando o sistema em equilíbrio, onde a regulação não é alterada durante o processo de transcrição, as taxas f e h permanecem inalteradas durante todo o processo. Dessa forma, a probabilidade de o gene estar no estado ativo em um determinado instante t e com uma determinada quantidade n de moléculas de mRNA, $A(t, n)$, é igual em qualquer instante t e para qualquer quantidade n . Assim sendo, podemos representar probabilidade do gene estar no estado ON pelo parâmetro A . Considerando também que uma fração das n moléculas de mRNA serão traduzidas em proteínas, o fato de a alteração de n não modificar A permite concluir que neste modelo não ocorre autorregulação.

Se $h \gg f$ ($h \ll f$), então $A \rightarrow 0$ ($A \rightarrow 1$) e a maior parte de um intervalo de tempo onde ocorre um ciclo completo será com o gene no estado reprimido (OFF) (ativo (ON)). Se for considerado $h \gg f$ e $\kappa > h$, então é possível interpretar que uma grande quantidade de moléculas de mRNA é transcrita nos pequenos intervalos de tempo em que o gene está ativo (ON).

A taxa de mudança de estado do gene em relação ao tempo de vida médio das moléculas de mRNA transcritas é indicada pelo parâmetro ε . Um parâmetro $\varepsilon < 1$ ($\varepsilon \ll 1$) indica que a variação do estado do gene ocorre a uma taxa inferior (muito inferior) à taxa de degradação das moléculas de mRNA, indicando um ciclo completo²¹ mais demorado, com o gene permanecendo a maior parte do tempo no estado ativo (ON) quando $\varepsilon \ll 1$.

Já um valor de $\varepsilon > 1$ ($\varepsilon \gg 1$) indica que a variação do estado do gene ocorre a uma taxa superior (muito superior) à taxa de degradação das moléculas de mRNA, refletindo em um ciclo completo realizado em menor tempo.

Para demonstrar essas condições foram executadas simulações utilizando as diferentes combinações de A e ε , o que implicava em diferentes parâmetros obtidos a partir das

²¹ Um ciclo completo significa o gene transitar de ON $\rightarrow$ OFF $\rightarrow$ ON ou de OFF $\rightarrow$ ON $\rightarrow$ OFF.

equações em (35). Em cada grupo foram realizadas 1000 simulações com 10000 iterações cada simulação. Os parâmetros estão na TABELA 2 e os resultados estão exibidos na figura

O parâmetro χ foi considerado de valor 0 em todas as simulações, pois desejamos simular um gene com dois modos de transcrição bem distintos, sendo um com uma alta taxa de transcrição (ON) e outro com taxa de transcrição 0 (OFF).

A partir das simulações de cada grupo, foram obtidos os valores médios $\langle n \rangle$ de cada iteração e o respectivo desvio padrão σ , a moda e a mediana da quantidade de moléculas de mRNA de cada iteração e a partir desses valores traçadas as trajetórias dos gráficos.

Também foram obtidas a frequência, a frequência e a intensidade do *bursting* e a frequência dos valores de n .

Neste trabalho foram executadas simulações do processo de transcrição para cada conjunto de parâmetros listados na tabela 2. Cada simulação é composta por iterações que podem adicionar ou remover uma molécula de mRNA ou então mudar o estado do gene sem alterar a quantidade de mRNA, de acordo com as regras definidas na tabela 1.

Para a implementação do algoritmo de Gillespie foram desenvolvidos programas utilizando linguagem C++ para reproduzir a execução do modelo e o aplicativo Gnuplot para gerar os gráficos.

TABELA 2 - PARÂMETROS DE SIMULAÇÃO

	f	h	κ	ρ	A	N	ϵ
1	0,00001	0,00009	0,0289	0,001	0,01	28,9	0,1
2	0,00005	0,00005	0,0289	0,001	0,50	28,9	0,1
3	0,000099	0,000001	0,0289	0,001	0,99	28,9	0,1
4	0,000075	0,000675	0,0289	0,001	0,01	28,9	0,75
5	0,000375	0,000375	0,0289	0,001	0,50	28,9	0,75
6	0,0007425	0,0000075	0,0289	0,001	0,99	28,9	0,75
7	0,000125	0,001125	0,0289	0,001	0,01	28,9	1,25
8	0,000625	0,000625	0,0289	0,001	0,50	28,9	1,25
9	0,0012375	0,0000125	0,0289	0,001	0,99	28,9	1,25
10	0,001	0,009	0,0289	0,001	0,01	28,9	10,0
11	0,005	0,005	0,0289	0,001	0,50	28,9	10,0
12	0,0099	0,0001	0,0289	0,001	0,99	28,9	10,0
13	10,0	990,0	1000,0	0,001	0,01	10000,0	10000,0
14	0,0034428	0,000346	0,0701998	0,00242665	0,9900498	28,9286877	1,43301

As simulações de 1 a 13 da tabela 2 buscam demonstrar o comportamento de genes a partir das possíveis combinações de valores referentes à probabilidade do gene estar no estado

ativo (ON) e do valor da relação entre a taxa de mudança de estado do gene e o tempo médio de vida da molécula de mRNA.

Já o conjunto 14 de parâmetros da tabela 2 utiliza valores do gene *even-skipped* (*eve*) da *D. melanogaster*. tais parâmetros foram obtidos por (36). No presente trabalho foram considerados os parâmetros do gene *eve*, entretanto realizou-se a simulação considerando o gene com apenas um.

3.1.2 Análise dos resultados das simulações

Nas simulações realizadas, o gráfico A contém a trajetória de $\langle n \rangle$ obtida na simulação e a trajetória obtida através do valor obtido pela resolução das equações geratrizes. A variação da expressão gênica em termos de amplitude, intensidade e frequência, característica perceptível nos gráficos E das simulações, faz com que diferentes combinações de parâmetros necessitem de diferentes quantidades de execuções para que a trajetória de $\langle n \rangle$ possa ser aproximar da trajetória do valor médio previsto.

Nas trajetórias dos gráficos E se pode perceber das simulações com $A \rightarrow 0$ que o tempo do gene no estado ON é reduzido e o gene passa a maior parte do tempo no estado OFF. Entretanto, enquanto o gene está no estado ON ocorrem *bursts* de transcrição de alta intensidade. Por outro lado, as moléculas de mRNA transcritas degradam a uma velocidade inversamente proporcional ao valor de ε . Essas características são perceptíveis nas trajetórias dos gráficos E das simulações 1, 2, 4, 5, 7, 8 e 10.

Simulações com $A \rightarrow 1$ indicam que o gene passa a maior parte do tempo no estado ON, fazendo com que a transcrição de moléculas de mRNA seja constante. Essa característica é perceptível nos gráficos E das simulações 3, 6, 9 e 11.

Nas trajetórias do gráfico E também é perceptível que, quando o parâmetro $\varepsilon \ll 1$ o número de moléculas de mRNA reduz rapidamente assim que o gene entra em estado OFF, característica que pode ser percebida nos gráficos 1, 2, 4, 5, 7, 8 e 10. Tal característica não ocorre nos gráficos 3, 6, 9 devido ao fato de $A \rightarrow 1$ e o gene permanecer em estado ON.

Dessa forma, em termos biológicos, a combinação desses fatores indica a magnitude da quantidade de genes que devem estar presentes para que ocorra um número constante de moléculas de mRNA disponíveis na célula.

Essa característica pode ser percebida também na trajetória dos gráficos A das simulações. Neles, a quantidade de simulações necessárias para gerar uma curva sem grandes oscilações em relação à curva do valor médio varia conforme os parâmetros A e ε utilizados.

Outro reflexo dos valores de A e ε pode ser percebido nos gráficos B das simulações, os quais contêm a frequência relativa do número de moléculas de mRNA. Os gráficos das simulações com $A \sim 0,0$ possuem formato de curva de dois picos, devido ao fato do número de moléculas de mRNA diminuir e frequentemente chegar a zero. Nos gráficos B das simulações deste trabalho foi retirada a coluna referente ao número zero de moléculas de mRNA, já que a escala necessária para representá-la dificultava a visualização das colunas de frequência dos demais números de mRNA. Como exemplo, na simulação 1 a frequência de ocorrência de zero moléculas de mRNA foi 0,5484, enquanto os valores das frequência de todas as outras quantidades de moléculas de mRNA foram abaixo de 0,06.

O formato do gráfico B se altera para uma distribuição binomial na medida em que o parâmetro A assume valores próximos de 1, indicando que o número de moléculas de mRNA não atinge zero ao longo do tempo de vida da célula.

Na simulação 13 foram usados dados publicados em (37) onde $A \sim 0,0$ e $\varepsilon \gg 1$

3.1.3 Simulação da transcrição do gene *even-skipped* da *D. melanogaster*

A simulação do gene *even-skipped* (*eve*) da *D. melanogaster* considerou dados referentes à expressão gênica obtidos por (39) apud (36). No estudo citado, foi considerada a expressão do gene *eve* no intervalo das posições equivalentes a 40,5% e 41,5% do comprimento do eixo antero-posterior do embrião da *D. melanogaster*. É nesta posição que o experimento detectou, usando técnicas de imunofluorescência, uma quantidade próxima de 28,75 moléculas de mRNA. Para a posição 41,5% do embrião, foi estimada uma taxa de transcrição (κ) de 0,07 moléculas de mRNA por segundo, uma taxa de degradação (ρ) de 0,0024 moléculas de mRNA por segundo e a constante $\varepsilon = 1,43$ por segundo.

A partir desses valores e usando as equações de (35) foi possível calcular o valor da probabilidade do gene estar no estado ativo quando o sistema atinge a estabilidade (A) em 0,99 e a razão entre a taxa de transcrição e de degradação de moléculas de mRNA (N) em 28,93.

O referido estudo também indicou que após o instante $t=2059$ segundos o estado do gene *eve* muda definitivamente para reprimido (OFF). Esta condição não foi representada nesta simulação.

Com relação à frequência da intensidade de *bursting*, os gráficos indicam que os *bursts* ocorrem com grande duração (em relação ao tempo de vida transcritiva do gene, limitado a 2059 segundos). Nas quatro execuções mostradas no gráfico E é possível perceber que os genes permaneceram no estado ON por toda a execução.

Não foram exibidos os gráficos de frequência de intensidade do *bursting* e de frequência de tempos de *bursting* para apenas uma execução porque, dentro do intervalo de tempo considerado, o simulador indica a existência de poucos e longos períodos com o gene no estado ON.

SIMULAÇÃO 1 - Modelo estocástico de gene binário com $\varepsilon \ll 1$ e $A \sim 0,0$

Os parâmetros usados nesta simulação foram $\varepsilon=0,1$; $A=0,01$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $p = 0,001$; $f = 0,00001$ e $h = 0,00009$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 6000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 2 - Modelo estocástico de gene binário com $\varepsilon \ll 1$ e $A \sim 0,5$

Os parâmetros usados nesta simulação foram $\varepsilon=0,1$; $A=0,50$, $N=28,90$, $n_0=0$, $\kappa=0,0289$, $\rho=0,001$, $f=0,00005$ e $h=0,00005$. Para a elaboração dos gráficos A, B, C e D foram realizadas 2000 simulações com 6000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 3 - Modelo estocástico de gene binário com $\varepsilon \ll 1$ e $A \sim 1,0$

Os parâmetros usados nesta simulação foram $\varepsilon=0,1$; $A=0,99$, $N=28,90$, $n_0=0$, $\kappa = 0,0289$, $\rho = 0,001$, $f = 0,00000995$ e $h = 0,00001$. Para a elaboração dos gráficos A, B, C e D foram realizadas 1000 simulações com 6000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E e F referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. O gráfico F apresenta a frequência relativa de n . Não foram exibidos os gráficos de frequência de intensidade e de tempo de *bursting* porque a execução permaneceu todo o tempo em estado ON. Dessa forma, ambos os gráficos reproduziam uma única coluna de valor.

SIMULAÇÃO 4 - Modelo estocástico de gene binário com $\varepsilon < 1$ e $A \sim 0,0$

Os parâmetros usados nesta simulação foram $\varepsilon=0,75$; $A=0,0$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,000075$ e $h = 0,000675$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 5 - Modelo estocástico de gene binário com $\varepsilon < 1$ e $A \sim 0,5$

Os parâmetros usados nesta simulação foram $\varepsilon=0,75$; $A=0,5$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,000375$ e $h = 0,000375$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_t \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 6 - Modelo estocástico de gene binário com $\varepsilon < 1$ e $A \sim 1,0$

Os parâmetros usados nesta simulação foram $\varepsilon=0,75$; $A=0,99$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,0007425$ e $h = 0,001125$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 7 - Modelo estocástico de gene binário com $\varepsilon > 1$ e $A \sim 0,0$

Os parâmetros usados nesta simulação foram $\varepsilon=1,25$; $A=0,01$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,000125$ e $h = 0,001125$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 8 - Modelo estocástico de gene binário com $\varepsilon > 1$ e $A \sim 0,5$

Os parâmetros usados nesta simulação foram $\varepsilon=1,25$; $A=0,5$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,000625$ e $h = 0,000625$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_t \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 9 - Modelo estocástico de gene binário com $\varepsilon > 1$ e $A \sim 1,0$

Os parâmetros usados nesta simulação foram $\varepsilon=1,25$; $A=0,99$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,0012375$ e $h = 0,0000125$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 10 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 0,0$

Os parâmetros usados nesta simulação foram $\varepsilon=10$; $A=0,01$, $N=28,90$, $n_0=0$, $\kappa=0,0289$, $\rho=0,001$, $f=0,001$ e $h=0,009$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 11 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 0,5$

Os parâmetros usados nesta simulação foram $\varepsilon=10$; $A=0,5$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,005$ e $h = 0,005$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 12 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 1,0$

Os parâmetros usados nesta simulação foram $\varepsilon=10$; $A=0,99$; $N=28,90$; $n_0=0$; $\kappa = 0,0289$; $\rho = 0,001$; $f = 0,0099$ e $h = 0,0001$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_t \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 13 - Modelo estocástico de gene binário com $\varepsilon \gg 1$ e $A \sim 0,0$

Os parâmetros usados nesta simulação foram $\varepsilon=10000$; $A=0,005$, $N=10000$, $n_0=50$, $\kappa=1000$, $\rho=0,1$, $f=5,0$ e $h=995,0$. Para a elaboração dos gráficos A, B, C e D foram realizadas 20000 simulações com 3000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. Os gráficos E, F, G e H referem-se a apenas uma execução com 100000 iterações. O gráfico E apresenta a trajetória de n . O gráfico E (*Detalhe*) apresenta um pequeno trecho do gráfico E em destaque. Os gráficos F, G e H apresentam respectivamente a frequência relativa de n , a frequência relativa da intensidade do *bursting* e a frequência relativa do tempo de duração do *bursting*.

SIMULAÇÃO 14 - Gene *even-skipped* da *D. melanogaster*.

Os parâmetros usados nesta simulação foram $\varepsilon=1,43$; $A=0,99$, $N=28,93$, $n_0=0$, $\kappa=0,0$; $\rho=0,0024$, $f=0,0034$ e $h=0,00034$. Para a elaboração dos gráficos A, B, C e D foram realizadas 10000 simulações com 1000 iterações cada uma. Os gráficos A, B e C mostram respectivamente a trajetória de $\langle n_i \rangle$, a frequência relativa de mRNA em todas as execuções da simulação, a frequência relativa da intensidade do *bursting* e a frequência relativa dos intervalos de *bursting*. O gráfico E mostra a trajetória de 4 execuções e a trajetória gerada pela equação da solução exata.

4 DISCUSSÃO E CONSIDERAÇÕES FINAIS

O modelo do gene binário permite determinar o comportamento da expressão gênica a partir do estado do gene, ativo ou reprimido, e da quantidade de moléculas de mRNA existentes na célula. Conhecendo a taxa de mudança de estado do gene, de ativo para reprimido e vice-versa, bem como a taxa de transcrição e de degradação das moléculas de mRNA, é possível reproduzir os valores de expressão gênica a partir de equações mestras, que geram os valores exatos do número de moléculas de mRNA ao longo do tempo.

Também é possível utilizar simuladores de reações químicas acopladas baseadas no algoritmo de Gillespie para simular o comportamento de gene. Neste caso, trajetória da média dos valores obtidos a partir de um grande número de execuções, que seria equivalente a um grande número de um determinado gene, se aproxima e se sobrepõe à trajetória gerada pelos resultados analíticos. Da mesma forma, valores de frequência obtidos pelas simulações utilizando o algoritmo de Gillespie, se mostraram compatíveis com os valores de frequência obtidos a partir de equações mestras.

A aproximação das trajetórias do valor médio $\langle n \rangle$ do número de moléculas de mRNA da simulação e da solução analítica é influenciada pela quantidade de execuções realizada na simulação, de forma que, quanto maior a razão ε entre as taxas de chaveamento de estado do gene e a taxa de degradação das moléculas de mRNA, maior será a quantidade necessária de execuções para aproximar a trajetória de $\langle n \rangle$ obtida na simulação à trajetória de $\langle n \rangle$ obtida na solução analítica. Na execução da simulação, o aumento da taxa de chaveamento implica também em um maior número de eventos dentro de um intervalo de tempo. Como consequência, será necessário um maior número de iterações em cada execução para que a trajetória de $\langle n \rangle$ inicie do estado inicial com 0 moléculas de mRNA e atinja o estado estacionário. A associação de uma alta quantidade de execuções com um alto número de iterações em cada execução torna necessária uma grande capacidade de processamento para a simulação quando são usados parâmetros de genes com tempo de vida longo associado à alta taxa de chaveamento de estado, sendo interessante então definir o período que se deseja simular, obtendo um valor médio inicial da quantidade de moléculas de mRNA para iniciar a simulação. O instante onde uma simulação parcial iniciará somente será necessário caso as taxas de transcrição e de degradação de moléculas de mRNA, bem como a taxa de chaveamento de estado do gene forem dependentes do tempo.

Como exemplo do crescimento da capacidade de processamento necessária atrelada ao aumento do número de execuções e de iterações de cada execução, temos que simulação 14,

referente à transcrição do gene *eve* e com $\varepsilon > 1$ e $A \sim 1$, exigiu 10 mil execuções de 1000 iterações cada, totalizando o processamento de 10 milhões de iterações. Já a simulação 13, onde $\varepsilon \gg 1$ e $A \sim 0,0$ exigiu 20 mil execuções de 3 mil iterações cada, totalizando 60 milhões de iterações.

As características do gene *even-skipped* do embrião *D. melanogaster* permitem simulação com baixa capacidade de processamento, já que possuem a probabilidade próxima de 1 do gene estar no estado ativo (ON). Dessa forma o valor de $\langle n \rangle$ atinge o estado estacionário rapidamente. Considerando que após 2059 segundos o gene encerra definitivamente a transcrição de mRNA (36), as simulações da transcrição do gene *eve* precisam ser executadas apenas até instante 2059s, o que permite fazer cada execução com cerca de 300 iterações.

Dessa forma, os resultados de simulações realizadas usando parâmetros de transcrição do gene *eve* podem servir tanto para auxiliar na definição de requisitos não funcionais da implementação como para calibrar o modelo implementado.

Com relação à obtenção de $\langle n \rangle$ deve-se ressaltar que, é necessário que o valor de $\langle n \rangle$ seja calculado em cada instante de tempo e não apenas nos instantes em que ocorreram iterações, pois uma determinada iteração pode assumir diferentes instantes nas diferentes execuções da simulação.

Através da construção do simulador baseado no algoritmo de Gillespie, foi feita a análise da estocasticidade da transcrição de moléculas de mRNA, identificando as variáveis (estado do gene e número de moléculas de mRNA) que atuam nesse processo a partir do modelo de gene binário. Com base nas variáveis, foram feitas simulações que permitiram analisar como as diferentes combinações das taxas de mudança dos valores das variáveis (mudança de estado do gene e alteração no número de moléculas de mRNA através da transcrição e degradação) podem interferir na definição das características de desempenho do simulador.

Como perspectiva futura, o simulador será redesenhado para ter um melhor desempenho em simulações que exigem um alto número de execuções associadas a um alto número de iterações, permitindo obter além dos valores de tendência central de n também valores de dispersão, que poderão ser comparados com os respectivos valores analíticos.

5 REFERÊNCIAS

1. ALBERTS, B. et al. **Biologia Molecular da Célula**. 5ª. ed. Porto Alegre: Artmed, 2010.
2. PRATA, G. N. **Um Modelo Estocástico para a transcrição do Gene even-skipped de Drosophila melanogaster**. São Carlos: [s.n.], 2013. Tese de doutorado.
3. ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY. Visualization Group. **Drosophila Gene Expression Data Exploration and Visualization**. Disponível em: <<http://vis.lbl.gov/Vignettes/Drosophila/#Genetic%20Fundamental>>. Acesso em: 08 maio 2017.
4. GRIFFITHS, A. J. F. et al. **An Introduction to Genetic Analysis**. 8ª. ed. New York: W. H. Freeman, 2000.
5. J. CRAIG VENTER INSTITUTE. Genetics and Genomics Timeline - 1902 - Theodor Boveri (1862-1915) and Walter Sutton (1877-1916) propose that chromosomes bear hereditary factors in accordance with Mendelian laws. **GNN - Genome News Network**. Disponível em: <http://www.genomenewsnetwork.org/resources/timeline/1902_Boveri_Sutton.php>. Acesso em: 12 jan. 2016.
6. J. CRAIG VENTER INSTITUTE. Genetics and Genomics Timeline - 1910 - Thomas Hunt Morgan (1866-1945) establishes the chromosomal theory of heredity. **GNN - Genome News Network**. Disponível em: <http://www.genomenewsnetwork.org/resources/timeline/1910_Morgan.php>. Acesso em: 12 jan. 2016.
7. PASTERNAK, J. J. **Genética Molecular Humana**. 1ª. ed. Barueri: Manole Ltda, 2002. 89 p.
8. KARP, G. **Biologia celular e molecular: conceitos e experimentos**. 3. ed. Barueri: Manole, 2005. 443 p.
9. NATIONAL HUMAN GENOME RESEARCH INSTITUTE. **Wikimedia commons**, 2014. Disponível em: <<https://commons.wikimedia.org/wiki/File:Gene.png>>. Acesso em: 05 jan. 2017.
10. PTASHNE, M. **A genetic switch: phage lambda revisited**. 3ª. ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 2004.
11. BERG, J. M.; TYMOCZKO, J. T.; STRYER, L. **Bioquímica**. Tradução de Antonio José Magalhães da Silva Moreira; João Paulo de Campos e Paulo A. Motta. 6ª. ed. Rio de Janeiro: Guanabara Koogan S.A, 2008.
12. RAMOS, A. F.; HORNOS, J. E. M. Symmetry and Stochastic Gene Regulation. **PHYSICAL REVIEW LETTERS**, v. 99, n. 10, p. 108103, 07 set. 2007.
13. RAMOS, A. F. et al. Symmetry in biology: from genetic code to tochastic gene regulation. **IET Syst. Biol**, v. 4, n. 5, 2010.
14. CRICK, F. Central Dogma of Molecular Biology. **Nature**, v. 227, 08 ago. 1970. Disponível em <http://www.nature.com/nature/focus/crick/pdf/crick227.pdf>.
15. NATIONAL HUMAN GENOME RESEARCH INSTITUTE. Encyclopedia of Science. **The Worlds of David Darling**. Disponível em: <http://www.daviddarling.info/encyclopedia/G/gene_expression.html>. Acesso em: 06 jan. 2017.
16. SCRANK, A. Transcrição. In: ZAHA, A.; FERRIERA, H. B.; PASSAGLIA, L. M. P. **Biologia Molecular Básica**. 5ª. ed. Porto Alegre: Artmed, 2014. Cap. 10.
17. GUYTON, A. C. **Tratado de Fisiologia Médica**. 11ª. ed. Rio de Janeiro: Elsevier, 2006.
18. JAEGER, J. The gap gene network. **Cellular and Molecular Life Sciences**, Barcelona, n. 68, p. 243-274, 2011.
19. CLARK, E.; AKAM, M. Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network. **eLife**, ago. 2016. Disponível em <https://elifesciences.org/content/5/e18215>.
20. WOLPERT, L. et al. **Biologia do Desenvolvimento**. Tradução de Henrique Bunselmeyer Ferreira. Porto Alegre: Artmed - Artes Médicas Sul, 2000.
21. SINGH, A.; SOLTANI, M. Quantifying Intrinsic and Extrinsic Variability in Stochastic Gene Expression Models. **PLOS ONE**, v. 8, n. 12, dez. 2013.
22. CHALANCON, G. et al. Interplay between gene expression noise and regulatory network architecture. **Trends in Genetics**, v. 28, n. 5, p. 221-223, Maio 2012. Disponível em: http://ccsb.dfci.harvard.edu/web/export/sites/default/ccsb/publications/papers/2012/Chalancon-Balaji_TrendsGenet_May2012c.pdf.
23. RASER, J. M.; O'SHEA, E. K. Noise in Gene Expression: Origins, Consequences, and Control. **NIH Public Access**, p. 309, 23 Setembro 2005.
24. JOHNSTON, I. G. et al. Mitochondrial Variability as a Source of Extrinsic Cellular Noise. **PLoS**

- Computational Biology**, v. 8, n. 3, mar. 2012. disponível em <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002416>.
25. WATSON, J. D. et al. **Biologia molecular do gene**. 7ª. ed. Porto Alegre: Artmed, 2015.
 26. RAJ, A.; OUDENAARDEN, A. V. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. **CELL**, v. 135, n. 2, p. 216-226, 17 out. 2008. Disponível em [http://www.cell.com/fulltext/S0092-8674\(08\)01243-9](http://www.cell.com/fulltext/S0092-8674(08)01243-9).
 27. ELOWITZ, M. B. et al. Stochastic gene expression in a single cell. **Science**, v. 297, p. 1183–1186, 2002.
 28. RAMOS, A.; HORNOS, J. E. M.; REINITZ, J. Gene regulation and noise reduction by coupling of stochastic process. **Physical Review**, n. 92, 2015.
 29. BDTNP - THE BERKELEY DROSOPHILA TRANSCRIPTION NETWORK PROJECT. BDTNP - The Berkeley Drosophila Transcription Network Project. **Centre for Image Analysis**, 2010. Disponível em: http://www.cb.uu.se/~cris/BDTNP_Imaging.html. Acesso em: 31 mar. 2017.
 30. GILBERT, S. F. **Developmental Biology**. 9. ed. Sunderland: Sinauer, 2010.
 31. MCADAMS, H. H.; ARKIN, A. It's a noisy business! Genetic regulation at the nanomolar scale. **Trends in Genetics**, v. 15, n. 2, p. 65-69, Fevereiro 1999.
 32. HORNOS, J. E. M. et al. Self-regulating gene: An exact solution. **PHYSICAL REVIEW**, 75, n. 5, 2005. 1 - 5.
 33. INNOCENTINI, G. C. P.; HORNOS, J. E. M. Modeling stochastic gene expression under repression. **Journal of Mathematical Biology**, 55, 2007. 413-431.
 34. INNOCENTINI, G. C. P. **Modelamento estocástico para a expressão gênica**. Universidade de São Paulo. São Carlos. 2008. Dissertação (Mestrado em Ciências).
 35. RAMOS, A. F.; INNOCENTINI, G. C. P.; HORNOS, J. E. M. Exact time-dependent solutions for a self-regulating gene. **Physical Review E**, v. 83, n. 6, 2011.
 36. PRATA, G. N.; HORNOS, J. E. M.; RAMOS, A. F. A stochastic model for gene transcription on Drosophila melanogaster embryos, jun. 2015. a publicar.
 37. YVINEC, R. et al. Bursting on a two states stochastic model for gene transcription in Drosophila embryos, 2017.
 38. GILLESPIE, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. **The Journal of Physical Chemistry**, v. 81, n. 25, 1977.
 39. JANSSENS, H. et al. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. **Nature Genetics**, v. 38, p. 1159, 2006.
 40. MATEUS, A. L. **Química na Cabeça**. Belo Horizonte: UFMG, 2001. 81 p.
 41. SCHROEDER, M. D. et al. Transcriptional Control in the Segmentation Gene Network of Drosophila. **PLoS Biology**, v. 2, n. 271, p. 1396 - 1410, set. 2004. www.plosbiology.org.

APÊNDICE A - FLUXOGRAMA DA IMPLEMENTAÇÃO DO ALGORITMO DE GILLESPIE

APÊNDICE A.2 - CÁLCULO DO INSTANTE E TIPO DE REAÇÃO

APÊNDICE B - GERAÇÃO DAS EQUAÇÕES DIFERENCIAIS PARCIAIS ACOPLADAS

A resolução das equações 1 e 2 se dá através da seguinte execução:

$$\frac{d}{dt} \alpha_n = k(\alpha_{n-1} - \alpha_n) + \rho[(n+1)\alpha_{n+1} - n\alpha_n] - h\alpha_n + f\beta_n \quad (2.1)$$

$$\frac{d}{dt} \beta_n = \chi k(\beta_{n-1} - \beta_n) + \rho[(n+1)\beta_{n+1} - n\beta_n] + h\alpha_n - f\beta_n \quad (2.2)$$

Multiplicando ambos os lados por z^n , obtém-se uma série de potência:

$$\frac{d\alpha_n}{dt} z^n = z^n k(\alpha_{n-1} - \alpha_n) + z^n \rho[(n+1)\alpha_{n+1} - n\alpha_n] - z^n h\alpha_n + z^n f\beta_n \quad (2.3)$$

$$\frac{d\beta_n}{dt} z^n = z^n \chi k(\beta_{n-1} - \beta_n) + z^n \rho[(n+1)\beta_{n+1} - n\beta_n] + z^n h\alpha_n - z^n f\beta_n \quad (2.4)$$

Considerando a somatória das probabilidades obtém-se

$$\frac{d\alpha_n}{dt} \sum_{n=1}^{\infty} z^n = \sum_{n=1}^{\infty} z^n k(\alpha_{n-1} - \alpha_n) + \sum_{n=1}^{\infty} z^n \rho[(n+1)\alpha_{n+1} - n\alpha_n] - \sum_{n=1}^{\infty} z^n h\alpha_n + \sum_{n=1}^{\infty} z^n f\beta_n \quad (2.5)$$

$$\frac{d\beta_n}{dt} \sum_{n=1}^{\infty} z^n = \sum_{n=1}^{\infty} z^n \chi k(\beta_{n-1} - \beta_n) + \sum_{n=1}^{\infty} z^n \rho[(n+1)\beta_{n+1} - n\beta_n] + \sum_{n=1}^{\infty} z^n h\alpha_n - \sum_{n=1}^{\infty} z^n f\beta_n \quad (2.6)$$

Distribuindo:

$$\frac{d\alpha_n}{dt} \sum_{n=1}^{\infty} z^n = k(\sum_{n=1}^{\infty} z^n \alpha_{n-1} - \sum_{n=1}^{\infty} z^n \alpha_n) + \rho[\sum_{n=1}^{\infty} z^n (n+1)\alpha_{n+1} - \sum_{n=1}^{\infty} z^n n\alpha_n] - h\sum_{n=1}^{\infty} z^n \alpha_n + f\sum_{n=1}^{\infty} z^n \beta_n \quad (2.7)$$

$$\frac{d\beta_n}{dt} \sum_{n=1}^{\infty} z^n = \chi k(\sum_{n=1}^{\infty} z^n \beta_{n-1} - \sum_{n=1}^{\infty} z^n \beta_n) + \rho[\sum_{n=1}^{\infty} z^n (n+1)\beta_{n+1} - \sum_{n=1}^{\infty} z^n n\beta_n] + h\sum_{n=1}^{\infty} z^n \alpha_n - f\sum_{n=1}^{\infty} z^n \beta_n \quad (2.8)$$

Considerando $\alpha_{(z,t)} = \sum_{n=1}^{\infty} z^n \alpha_n$ e $\beta_{(z,t)} = \sum_{n=1}^{\infty} z^n \beta_n$

$$\frac{\partial}{\partial t} \alpha_{(z,t)} = k(\sum_{n=1}^{\infty} z^n \alpha_{n-1} - \alpha_{(z,t)}) + \rho[\sum_{n=1}^{\infty} z^n (n+1)\alpha_{n+1} - \sum_{n=1}^{\infty} z^n n\alpha_n] - h\alpha_{(z,t)} + f\beta_{(z,t)} \quad (2.9)$$

$$\frac{\partial}{\partial t} \beta_{(z,t)} = \chi k(\sum_{n=1}^{\infty} z^n \beta_{n-1} - \beta_{(z,t)}) + \rho[\sum_{n=1}^{\infty} z^n (n+1)\beta_{n+1} - \sum_{n=1}^{\infty} z^n n\beta_n] + h\alpha_{(z,t)} - f\beta_{(z,t)} \quad (2.10)$$

Considerando $m = n - 1 \Rightarrow n = m + 1 \Rightarrow \sum_{n=1}^{\infty} z^n \alpha_{n-1} = \sum_{m=1}^{\infty} z^{m+1} \alpha_m = z \sum_{m=1}^{\infty} z^m \alpha_m = z\alpha_{(z,t)}$
 $\sum_{n=1}^{\infty} z^n \beta_{n-1} = \sum_{m=1}^{\infty} z^{m+1} \beta_m = z \sum_{m=1}^{\infty} z^m \beta_m = z\beta_{(z,t)}$

$$\frac{\partial}{\partial t} \alpha_{(z,t)} = k(z\alpha_{(z,t)} - \alpha_{(z,t)}) + \rho[\sum_{n=1}^{\infty} z^n (n+1)\alpha_{n+1} - \sum_{n=1}^{\infty} z^n n\alpha_n] - h\alpha_{(z,t)} + f\beta_{(z,t)} \quad (2.11)$$

$$\frac{\partial}{\partial t} \beta_{(z,t)} = \chi k(z\beta_{(z,t)} - \beta_{(z,t)}) + \rho[\sum_{n=1}^{\infty} z^n (n+1)\beta_{n+1} - \sum_{n=1}^{\infty} z^n n\beta_n] + h\alpha_{(z,t)} - f\beta_{(z,t)} \quad (2.12)$$

Considerando $(n+1)z^n = \frac{dz^{n+1}}{dz}$ e $m = n + 1$ tem-se $n = m - 1 \Rightarrow (n+1)z^n \alpha_{n+1} = mz^{(m-1)} \alpha_m = \frac{dz^m}{dz} \alpha_m$
 $(n+1)z^n \beta_{n+1} = mz^{(m-1)} \beta_m = \frac{dz^m}{dz} \beta_m$

$$\frac{\partial}{\partial t} \alpha_{(z,t)} = k(z\alpha_{(z,t)} - \alpha_{(z,t)}) + \rho\left[\frac{\partial}{\partial z} \alpha_{(z,t)} - \sum_{n=1}^{\infty} z^n n\alpha_n\right] - h\alpha_{(z,t)} + f\beta_{(z,t)} \quad (2.13)$$

$$\frac{\partial}{\partial t} \beta_{(z,t)} = \chi k(z\beta_{(z,t)} - \beta_{(z,t)}) + \rho\left[\frac{\partial}{\partial z} \beta_{(z,t)} - \sum_{n=1}^{\infty} z^n n\beta_n\right] + h\alpha_{(z,t)} - f\beta_{(z,t)} \quad (2.14)$$

Considerando $nz^n = znz^{(n-1)} = z \frac{dz^n}{dz} \Rightarrow \sum_{n=1}^{\infty} z^n n\alpha_n = \sum_{n=1}^{\infty} z \frac{dz^n}{dz} \alpha_n = z \frac{d}{dz} \alpha_{(z,t)}$
 $\sum_{n=1}^{\infty} z^n n\beta_n = \sum_{n=1}^{\infty} z \frac{dz^n}{dz} \beta_n = z \frac{d}{dz} \beta_{(z,t)}$

$$\frac{\partial}{\partial t} \alpha_{(z,t)} = k(z\alpha_{(z,t)} - \alpha_{(z,t)}) + \rho\left[\frac{\partial}{\partial z} \alpha_{(z,t)} - z \frac{\partial}{\partial z} \alpha_{(z,t)}\right] - h\alpha_{(z,t)} + f\beta_{(z,t)} \quad (2.15)$$

$$\frac{\partial}{\partial t} \beta_{(z,t)} = \chi k(z\beta_{(z,t)} - \beta_{(z,t)}) + \rho\left[\frac{\partial}{\partial z} \beta_{(z,t)} - z \frac{\partial}{\partial z} \beta_{(z,t)}\right] + h\alpha_{(z,t)} - f\beta_{(z,t)} \quad (2.16)$$

Redistribuindo:

$$\frac{\partial}{\partial t} \alpha_{(z,t)} = (z-1) \left(k\alpha_{(z,t)} - \rho \frac{\partial}{\partial z} \alpha_{(z,t)} \right) - h\alpha_{(z,t)} + f\beta_{(z,t)} \quad (2.17)$$

$$\frac{\partial}{\partial t} \beta_{(z,t)} = (z-1) \left(\chi k\beta_{(z,t)} - \rho \frac{\partial}{\partial z} \beta_{(z,t)} \right) + h\alpha_{(z,t)} - f\beta_{(z,t)} \quad (2.18)$$

APÊNDICE C - OBTENÇÃO DAS DERIVATIVAS

$\alpha(z, t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n$	$\beta(z, t) = \sum_{n=0}^{\infty} \beta_n(t) z^n \quad (3.1)$
--	--

Podemos considerar que:

$\alpha(z, t) = \alpha_0(t) z^0 + \alpha_1(t) z + \alpha_2(t) z^2 + \dots$	$\beta(z, t) = \beta_0(t) z^0 + \beta_1(t) z + \beta_2(t) z^2 + \dots \quad (3.2)$
--	--

Então:

$\frac{\partial \alpha}{\partial z} = \frac{\partial}{\partial z} (\alpha_0(t) z^0 + \alpha_1(t) z + \alpha_2(t) z^2 + \dots)$	$\frac{\partial \beta}{\partial z} = \frac{\partial}{\partial z} (\beta_0(t) z^0 + \beta_1(t) z + \beta_2(t) z^2 + \dots) \quad (3.3)$
--	--

Assim sendo, temos as derivadas:

$\begin{aligned} \frac{d\alpha}{dz} &= \alpha_1(t) + 2\alpha_2(t)z + 3\alpha_3(t)z^2 + 4\alpha_4(t)z^3 + \dots \\ \frac{d^2\alpha}{dz^2} &= 2\alpha_2(t) + 6\alpha_3(t)z + 12\alpha_4(t)z^2 + 20\alpha_5(t)z^3 + \dots \\ \frac{d^3\alpha}{dz^3} &= 6\alpha_3(t) + 24\alpha_4(t)z + 60\alpha_5(t)z^2 + \dots \end{aligned}$	$\begin{aligned} \frac{d\beta}{dz} &= \beta_1(t) + 2\beta_2(t)z + 3\beta_3(t)z^2 + 4\beta_4(t)z^3 + \dots \\ \frac{d^2\beta}{dz^2} &= 2\beta_2(t) + 6\beta_3(t)z + 12\beta_4(t)z^2 + 20\beta_5(t)z^3 + \dots \\ \frac{d^3\beta}{dz^3} &= 6\beta_3(t) + 24\beta_4(t)z + 60\beta_5(t)z^2 + \dots \end{aligned} \quad (3.4)$
---	---

De forma geral:

$\frac{d^n \alpha}{dz^n} = n! \alpha_n(t) + \frac{(n+1)!}{1!} \alpha_{n+1}(t) z + \frac{(n+2)!}{2!} \alpha_{n+2}(t) z^2 + \dots$	$\frac{d^n \beta}{dz^n} = n! \beta_n(t) + \frac{(n+1)!}{1!} \beta_{n+1}(t) z + \frac{(n+2)!}{2!} \beta_{n+2}(t) z^2 + \dots \quad (3.5)$
--	--

Considerando a derivada parcial para $z = 0$ tem-se:

$\left. \frac{\partial^n \alpha}{\partial z^n} \right _{z=0} = n! \alpha_n(t)$	$\left. \frac{\partial^n \beta}{\partial z^n} \right _{z=0} = n! \beta_n(t) \quad (3.6)$
--	--

O que implica em:

$\alpha_n(t) = \frac{1}{n!} \left. \frac{\partial^n \alpha}{\partial z^n} \right _{z=0}$	$\beta_n(t) = \frac{1}{n!} \left. \frac{\partial^n \beta}{\partial z^n} \right _{z=0} \quad (3.7)$
--	--

APÊNDICE D - OBTENÇÃO DOS MOMENTOS

1º Momento – Valor médio

A partir da 1ª derivada definida em (3.4), para $z=1$:

$\left. \frac{d\alpha}{dz} \right _{z=1} = \alpha_1(t) + 2\alpha_2(t)z + 3\alpha_3(t)z^2 + \dots \Big _{z=1}$	$\left. \frac{d\beta}{dz} \right _{z=1} = \beta_1(t) + 2\beta_2(t)z + 3\beta_3(t)z^2 + \dots \Big _{z=1} \quad (4.1)$
---	---

Temos que:

$\left. \frac{d\alpha}{dz} \right _{z=1} = \alpha_1(t) + 2\alpha_2(t) + 3\alpha_3(t) + \dots$	$\left. \frac{d\beta}{dz} \right _{z=1} = \beta_1(t) + 2\beta_2(t) + 3\beta_3(t) + \dots \quad (4.2)$
---	---

Sabendo que $\langle n_\alpha \rangle = \sum_{n=0}^{\infty} n\alpha_n(t)$ e $\langle n_\beta \rangle = \sum_{n=0}^{\infty} n\beta_n(t)$ temos que

$\langle n_\alpha \rangle = \left. \frac{d\alpha}{dz} \right _{z=1} = \sum_{n=0}^{\infty} n\alpha_n(t)$	$\langle n_\beta \rangle = \left. \frac{d\beta}{dz} \right _{z=1} = \sum_{n=0}^{\infty} n\beta_n(t) \quad (4.3)$
---	--

2º Momento – Dispersão

A partir da 1ª derivada definida em (3.4), para $z=1$:

$\left. \frac{d\alpha}{dz} \right _{z=1} = \alpha_1(t) + 2\alpha_2(t)z + 3\alpha_3(t)z^2 + \dots \Big _{z=1}$	$\left. \frac{d\beta}{dz} \right _{z=1} = \beta_1(t) + 2\beta_2(t)z + 3\beta_3(t)z^2 + \dots \Big _{z=1} \quad (4.4)$
---	---

Multiplica ambos os lados da igualdade por z e deriva:

$\left. \frac{d}{dz} z \frac{d\alpha}{dz} \right _{z=1} = \alpha_1(t) + 4\alpha_2(t)z + 9\alpha_3(t)z^2 + \dots \Big _{z=1}$	$\left. \frac{d}{dz} z \frac{d\beta}{dz} \right _{z=1} = \beta_1(t) + 4\beta_2(t)z + 9\beta_3(t)z^2 + \dots \Big _{z=1} \quad (4.5)$
--	--

Como $z=1$ temos que:

$\left. \frac{d}{dz} z \frac{d\alpha}{dz} \right _{z=1} = \alpha_1(t) + 4\alpha_2(t) + 9\alpha_3(t) + \dots$	$\left. \frac{d}{dz} z \frac{d\beta}{dz} \right _{z=1} = \beta_1(t) + 4\beta_2(t) + 9\beta_3(t) + \dots \quad (4.6)$
--	--

De forma geral temos:

$\left. \frac{d^2\alpha}{dz^2} \right _{z=1} = \sum_{n=0}^{\infty} n^2\alpha_n(t)$	$\left. \frac{d^2\beta}{dz^2} \right _{z=1} = \sum_{n=0}^{\infty} n^2\beta_n(t) \quad (4.7)$
--	--

Sabendo que $\langle n_\alpha^2 \rangle = \sum_{n=0}^{\infty} n^2\alpha_n(t)$ e $\langle n_\beta^2 \rangle = \sum_{n=0}^{\infty} n^2\beta_n(t)$ temos que

$\langle n_\alpha^2 \rangle = \left. \frac{d^2\alpha}{dz^2} \right _{z=1} = \sum_{n=0}^{\infty} n^2\alpha_n(t)$	$\langle n_\beta^2 \rangle = \left. \frac{d^2\beta}{dz^2} \right _{z=1} = \sum_{n=0}^{\infty} n^2\beta_n(t) \quad (4.8)$
---	--

P-ésimo momento

Os demais momentos podem ser obtidos usando o mesmo processo de derivação. Dessa forma temos que:

$\langle n_\alpha^p \rangle = \left. \frac{d^p\alpha}{dz^p} \right _{z=1} = \sum_{n=0}^{\infty} n^p\alpha_n(t)$	$\langle n_\beta^p \rangle = \left. \frac{d^p\beta}{dz^p} \right _{z=1} = \sum_{n=0}^{\infty} n^p\beta_n(t) \quad (4.9)$
---	--

APÊNDICE E - OBTENÇÃO DA EQUAÇÃO DE KUMMER

No estado estacionário, considerando $A = \sum_{n=0}^{\infty} \alpha_n$, $B = \sum_{n=0}^{\infty} \beta_n$ e $\chi = 0$ tem-se que:

$(z-1) \left(\kappa \alpha - \rho \frac{dA}{dz} \right) - hA + fB = 0 \quad (5.1)$	$(z-1) \left(-\rho \frac{dB}{dz} \right) + hA - fB = 0 \quad (5.2)$
---	--

Resolvendo a eq. 5.1 para B e substituímos B na eq. 5.2

$\begin{aligned} & (z-1) \left(-\rho \frac{d}{dz} \left(\frac{1}{f} \left(hA - (z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) \right) \right) \right) + hA - f \left(\frac{1}{f} \left(hA - (z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) \right) \right) = 0 \\ & -\frac{\rho}{f} (z-1) \frac{d}{dz} \left(hA - (z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) \right) + hA - hA + (z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) = 0 \\ & -\frac{\rho}{f} (z-1) \frac{d}{dz} \left(hA - (z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) \right) + (z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) = 0 \\ & (z-1) \left(-\frac{\rho}{f} \frac{d}{dz} \left(hA - (z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) \right) + \left(\kappa A - \rho \frac{dA}{dz} \right) \right) = 0 \\ & (z-1) \left(-\frac{\rho}{f} h \frac{dA}{dz} + \frac{\rho}{f} \frac{d}{dz} \left((z-1) \left(\kappa A - \rho \frac{dA}{dz} \right) \right) + \left(\kappa A - \rho \frac{dA}{dz} \right) \right) = 0 \\ & (z-1) \left(-\frac{\rho}{f} h \frac{dA}{dz} + \frac{\rho}{f} \left(\kappa A - \rho \frac{dA}{dz} \right) + \frac{\rho}{f} (z-1) \kappa \frac{dA}{dz} - \frac{\rho}{f} (z-1) \rho \frac{d^2 A}{dz^2} + \kappa A - \rho \frac{dA}{dz} \right) = 0 \end{aligned}$	(5.3)
---	---------

Considerando $(z-1) \neq 0$:

$\begin{aligned} & -\frac{\rho}{f} h \frac{dA}{dz} + \frac{\rho}{f} \kappa A - \frac{\rho^2}{f} \frac{dA}{dz} + \frac{\rho}{f} (z-1) \kappa \frac{dA}{dz} - \frac{\rho^2}{f} (z-1) \frac{d^2 A}{dz^2} + \kappa A - \rho \frac{dA}{dz} = 0 \\ & -\frac{\rho^2}{f} (z-1) \frac{d^2 A}{dz^2} + \left[-\rho - \frac{\rho}{f} h - \frac{\rho^2}{f} + \frac{\rho}{f} (z-1) \kappa \right] \frac{dA}{dz} + \kappa \left(1 + \frac{\rho}{f} \right) A = 0 \end{aligned}$	(5.4)
--	---------

Multiplicando ambos os lados da igualdade por $-\frac{f}{\rho^2}$ temos:

$\begin{aligned} & (z-1) \frac{d^2 A}{dz^2} + \left[1 + \frac{f}{\rho} + \frac{h}{\rho} - \frac{\kappa}{\rho} (z-1) \right] \frac{dA}{dz} - \frac{f}{\rho^2} \kappa \left(1 + \frac{\rho}{f} \right) A = 0 \\ & (z-1) \frac{d^2 A}{dz^2} + \left[1 + \frac{f+h}{\rho} - \frac{\kappa}{\rho} (z-1) \right] \frac{dA}{dz} - \frac{f\kappa}{\rho^2} A - \frac{\kappa}{\rho} A = 0 \\ & (z-1) \frac{d^2 A}{dz^2} + \left[1 + \frac{f+h}{\rho} - \frac{\kappa}{\rho} (z-1) \right] \frac{dA}{dz} - \frac{\kappa}{\rho} \left(1 + \frac{f}{\rho} \right) A = 0 \end{aligned}$	(5.5)
--	---------

Considerando $a = \frac{f}{\rho}$; $b = \frac{f+h}{\rho}$ e $N = \frac{\kappa}{\rho}$ temos:

$(z-1) \frac{d^2 A}{dz^2} + [1 + b - N(z-1)] \frac{dA}{dz} - N(1+a)A = 0$	(5.6)
---	---------

Considerando $\eta = N(z-1)$ e as regras de cadeia $\frac{d}{dz} = \frac{d}{d\eta} \frac{d\eta}{dz} = N \frac{d}{d\eta}$ e $\frac{dA}{dz} = \frac{dA}{d\eta} \frac{d\eta}{dz} = N \frac{dA}{d\eta}$ temos:

$\begin{aligned} & \frac{\eta}{N} \frac{d^2 A}{dz^2} + [1 + b - \eta] \frac{dA}{dz} - N(1+a)A = 0 \\ & \frac{\eta}{N} N^2 \frac{d^2 A}{d\eta^2} + [1 + b - \eta] N \frac{dA}{d\eta} - N(1+a)A = 0 \\ & \eta N \frac{d^2 A}{d\eta^2} + [1 + b - \eta] N \frac{dA}{d\eta} - N(1+a)A = 0 \\ & N \left(\eta \frac{d^2 A}{d\eta^2} + [1 + b - \eta] \frac{dA}{d\eta} - (1+a)A \right) = 0 \end{aligned}$	(5.7)
--	---------

Considerando $N \neq 0$:

$\eta \frac{d^2 A}{d\eta^2} + [1 + b - \eta] \frac{dA}{d\eta} - (1+a)A = 0$	(5.8)
---	---------

De forma similar, encontramos:

$\eta \frac{d^2 B}{d\eta^2} + [1 + b - \eta] \frac{dB}{d\eta} - aB = 0$	(5.9)
---	---------

APÊNDICE F- OBTENÇÃO DA EQUAÇÃO MESTRA

Considerando que

Temos que:

$$\frac{d}{dt} \alpha_n = [k\alpha_{n-1} + \rho(n+1)\alpha_{n+1}] - [(k + \rho n)\alpha_n]$$

$$\frac{d}{dt} \alpha_n = k\alpha_{n-1} + \rho(n+1)\alpha_{n+1} - k\alpha_n - \rho n\alpha_n$$

$$\frac{d}{dt} \alpha_n = k\alpha_{n-1} - k\alpha_n - \rho n\alpha_n + \rho(n+1)\alpha_{n+1}$$

$$\frac{d}{dt} \alpha_n = k(\alpha_{n-1} - \alpha_n) + \rho[(n+1)\alpha_{n+1} - n\alpha_n]$$

Considerando que

Temos que:

$$\frac{d}{dt} \alpha_n = -h\alpha_n + f\beta_n$$

Somando as equações temos que

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

Considerando que

Temos que:

$$\frac{d}{dt} \beta_n = [\chi k\beta_{n-1} + \rho(n+1)\beta_{n+1}] - [(\chi k + \rho n)\beta_n]$$

$$\frac{d}{dt} \beta_n = \chi k\beta_{n-1} + \rho(n+1)\beta_{n+1} - \chi k\beta_n - \rho n\beta_n$$

$$\frac{d}{dt} \beta_n = \chi k\beta_{n-1} - \chi k\beta_n - \rho n\beta_n + \rho(n+1)\beta_{n+1}$$

$$\frac{d}{dt} \beta_n = \chi k(\beta_{n-1} - \beta_n) + \rho[(n+1)\beta_{n+1} - n\beta_n]$$

Considerando que

Temos que:

$$\frac{d}{dt} \beta_n = h\alpha_n - f\beta_n$$

Somando as equações temos que

$$\frac{d}{dt} \beta_n = \chi k(\beta_{n-1} - \beta_n) + \rho[(n+1)\beta_{n+1} - n\beta_n] + h\alpha_n - f\beta_n$$

APÊNDICE G - PROBABILIDADE DE ENCONTRAR n mRNA

Considerando que

Temos que:

$\alpha_1 = \frac{k}{\rho} \alpha_0$	$\alpha_2 = \frac{k}{2\rho} \alpha_1 \Rightarrow \alpha_2 = \frac{k}{2\rho} \frac{k}{\rho} \alpha_0 \Rightarrow \alpha_2 = \left(\frac{k}{\rho}\right)^2 \frac{1}{2} \alpha_0$	$\alpha_3 = \frac{k}{3\rho} \alpha_2 \Rightarrow \alpha_3 = \frac{k}{3\rho} \frac{k}{2\rho} \frac{k}{\rho} \alpha_0 \Rightarrow \alpha_3 = \left(\frac{k}{\rho}\right)^3 \frac{1}{3*2} \alpha_0$
--------------------------------------	--	--

O que resulta em :

$\alpha_n = \left(\frac{k}{\rho}\right)^n \frac{1}{n!} \alpha_0$	Considerando $\left(\frac{k}{\rho}\right)^n = \lambda^n$ então:	$\alpha_n = \frac{\lambda^n}{n!} \alpha_0$
--	---	--

Considerando $\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^\lambda$ e que $e^x = 1 + x + \frac{x^2}{2} +$ temos:

$\sum_{n=0}^{\infty} \alpha_n = 1 \Rightarrow \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \alpha_0 = 1 \Rightarrow \alpha_0 \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = 1 \Rightarrow \alpha_0 e^\lambda = 1 \Rightarrow \alpha_0 = e^{-\lambda}$

Considerando que

Temos que:

$\beta_1 = \frac{\chi k}{\rho} \beta_0$	$\beta_2 = \frac{\chi k}{2\rho} \beta_1 \Rightarrow \beta_2 = \frac{\chi k}{2\rho} \frac{\chi k}{\rho} \beta_0 \Rightarrow \beta_2 = \left(\frac{\chi k}{\rho}\right)^2 \frac{1}{2} \beta_0$	$\beta_3 = \frac{\chi k}{3\rho} \beta_2 \Rightarrow \beta_3 = \frac{\chi k}{3\rho} \frac{\chi k}{2\rho} \frac{\chi k}{\rho} \beta_0 \Rightarrow \beta_3 = \left(\frac{\chi k}{\rho}\right)^3 \frac{1}{3*2} \beta_0$
---	--	---

O que resulta em :

$\beta_n = \left(\frac{\chi k}{\rho}\right)^n \frac{1}{n!} \beta_0$	Considerando $\left(\frac{\chi k}{\rho}\right)^n = \gamma^n$ então:	$\beta_n = \frac{\gamma^n}{n!} \beta_0$
---	---	---

Considerando $\sum_{n=0}^{\infty} \frac{\gamma^n}{n!} = e^\gamma$ e que $e^x = 1 + x + \frac{x^2}{2} +$ temos:

$\sum_{n=0}^{\infty} \beta_n = 1 \Rightarrow \sum_{n=0}^{\infty} \frac{\gamma^n}{n!} \beta_0 = 1 \Rightarrow \beta_0 \sum_{n=0}^{\infty} \frac{\gamma^n}{n!} = 1 \Rightarrow \beta_0 e^\gamma = 1 \Rightarrow \beta_0 = e^{-\gamma}$
--

ANEXO I

Bursting on a two states stochastic model for gene transcription in *Drosophila* embryos

Romain Yvinec¹, Luiz Guilherme Soares², Guilherme N. Prata², John Reinitz³, and Alexandre F. Ramos^{2,4*}

¹ *Physiologie de la Reproduction et des Comportements,*

Institut National de la Recherche Agronomique (INRA) UMR85,

CNRS-Université François-Rabelais UMR7247, IFCE, Nouzilly, 37380 France

² *Escola de Artes, Ciências e Humanidades & Núcleo de Estudos Interdisciplinares em Sistemas Complexos, Universidade de São Paulo, Av. Arlindo Bettio, 1000 CEP 03828-000, São Paulo, SP, Brazil*

³ *Department of Statistics, Department of Ecology and Evolution,*

Department of Molecular Genetics and Cell Biology,

University of Chicago 5734 S. University Avenue Eckhart 134 Chicago, IL 60637, USA and

⁴ *Departamento de Radiologia - Faculdade de Medicina,*

Universidade de São Paulo & Instituto do Câncer do Estado de São Paulo, São Paulo, SP CEP 05403-911

In this manuscript we approach a stochastic two states model for gene transcription in *Drosophila melanogaster* embryos and derive its bursting limit analytically. Simulations based on the Gillespie algorithm at the bursting limit show that burst size and time duration varies at each burst event. That characteristic have been observed recently in experiments with *eve* gene transcription *Drosophila* embryos and was used for proposing that its promoter site has multiple activated states. Based on our theoretical approach, we interpret that experimental data as a demonstration of the intrinsic stochasticity of the transcriptional processes in fruit fly embryos. Then, we conceive the experimental arrangement to determine when gene transcription has multiple ON promoter states in a noisy environment.

The control of transcription in cells is characterized by intrinsic fluctuations which result from the small number of molecules of proteins which control this process. In prokaryotes, the stochastic theory for this process is relatively well developed because of good understanding of the fundamental reaction mechanisms, but in eukaryotes the detailed reaction and control mechanisms are poorly understood, and a detailed stochastic theory does not exist. Nevertheless, high resolution experimental studies clearly reveal intrinsic fluctuations in the form of bursts of transcription of variable size and duration [1–7]. The number of underlying transcriptional states implied by this bursting behavior remain unclear. Previous investigations have explained the variable size and duration of bursts in terms of the random initiation of multiple underlying states [8], while variation in the duration of bursts has been explained by a two state stochastic model [9]. In the latter study an *ansatz* required by the experimental data prevented estimation of burst size, while the former used a deterministic model of the binding and elongation of RNA PolII. In this manuscript we consider the exact solutions for the two states stochastic model to calculate its burst limit approximation by rigorous methods and demonstrate the occurrence of variable burst size using Gillespie's exact simulation algorithm. This provides an estimate of the average burst size in one previous study [9] and demonstrates that the data reported in the other study [8] is fully compatible with a two state model.

Here we use a two state master equation model of transcription which has exact solutions given by confluent hypergeometric functions [10, 11] which has been previously applied to the control of the transcription of *even-skipped*

stripe 2 in *Drosophila melanogaster* [12]. The stochastic variables of the model are the number of mRNA molecules, denoted by n , and the state of the gene's promoter being ON or OFF. The probability of finding the promoter state as ON (or OFF) at time t when n mRNA molecules are found within a nucleus is denoted by $\alpha_n(t)$ (or $\beta_n(t)$). mRNA synthesis occurs at rate k when the gene is ON and is zero when the gene is OFF. mRNA degradation occurs at rate ρ . The promoter transition from the ON to the OFF (or from the OFF to the ON) state has rate h (or f). We omit the temporal dependence of the probabilities such that $\alpha_n(t) \equiv \alpha_n$ and $\beta_n(t) \equiv \beta_n$ and write the master equation governing their dynamics as

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

$$\frac{d\beta_n}{dt} = \rho[(n+1)\beta_{n+1} - n\beta_n] + h\alpha_n - f\beta_n. \quad (2)$$

The steady state probabilities of finding n mRNA molecules independently of the gene state are denoted by ϕ_n , with $\phi_n = \alpha_n + \beta_n$, given by

$$\phi_n = \frac{N^n (A\epsilon)_n}{n! (\epsilon)_n} M(A\epsilon + n, \epsilon + n, -N). \quad (3)$$

$M(a, b, z)$ denotes the KummerM function and $(x)_n = x(x+1)\dots(x+n-1)$ denotes the Pochhammer function, where $(x)_0 = 1$. Eq. (3) is written in terms of the parameters $\{N, \epsilon, A\}$ defined by

$$N = \frac{k}{\rho}, \quad A = \frac{f}{f+h}, \quad \epsilon = \frac{f+h}{\rho}. \quad (4)$$