Universidade de São Paulo
Instituto de Física de São Carlos

Guilherme Nery Prata

Um modelo estocástico para a transcrição do
gene *even-skipped* de *Drosophila*
melanogaster

São Carlos

2013

Guilherme Nery Prata

Um modelo estocástico para a transcrição do
gene *even-skipped* de *Drosophila*
melanogaster

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

Área de concentração: Física Básica.

Orientador: Prof. Dr. José Eduardo Martinho Hornos.

Versão corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos

2013

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Prata, Guilherme Nery

Um modelo estocástico para a transcrição do gene even-skipped de *Drosophila melanogaster* / Guilherme Nery Prata; orientador José Eduardo Martinho Hornos - versão corrigida -- São Carlos, 2013.
188 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Básica) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2013.

1. Expressão gênica. 2. Regulação transcricional. 3. Processos estocásticos. 4. Soluções exatas. 5. *Drosophila*. I. Hornos, José Eduardo Martinho, orient. II. Título.

AGRADECIMENTOS

Agradeço ao Prof. José Eduardo Martinho Hornos, pela orientação e por todos os ensinamentos e conselhos. Agradeço ao Prof. Alexandre Ramos (EACH - USP) pela ajuda, pelas ideias e discussões bastante frutíferas. Agradeço também aos demais professores e funcionários do Instituto de Física de São Carlos onde realizei este trabalho. E agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Um agradecimento muito mais do que especial aos meus pais, José Nery Prata (*in memoriam*) e Nilce Helena Francisco Gomes Pereti, pelo amor, educação e apoio. Sem dúvida, esta é mais uma conquista que se deve a cada gota de suor dedicada à família e à educação dos filhos. Este trabalho é dedicado a vocês! Agradeço também aos meus irmãos, Gabriel Nery Prata e Silvana Prata Maineri, e ao querido Antonio Gomes Pereti (Toninho) pelo carinho e apoio. Agradeço também aos meus avós, tios e tias, primos e primas, sobrinhos e sobrinhas e demais familiares.

Agradeço à Luciane Akemi Sato e ao Lucas Xavier por todos os bons sentimentos, os bons momentos, as alegrias, as gargalhadas, as músicas, o carinho, o respeito, a compreensão, o companheirismo e, sobretudo, a paciência nesses últimos anos.

Aos meus amigos e colegas de trabalho Marcos Felipe Sampaio, Luis Fernando Fontana Campos, José Eduardo (“Garçom”), Jonathas Siqueira, Carlos Brasil, Diego Paiva Pires, Renato Pimentel, Mariana Odashima, Gabriel Luchini, Jorge Massayuki, Paulo Moriya, Ismael Infante, Bruno Barbagallo, Rafael Ribeiro, Roberto “Mantena”, Edmundo Macha, Henrique Rodrigues e Leandro Carrierio, obrigado pelas conversas, pelas idéias, pelas gargalhadas, pelos desabafos, pelos conselhos, pela boa vontade em ajudar e pela amizade.

Agradeço também a cada cidadão brasileiro que, direta ou indiretamente, consciente ou inconscientemente, contribui para que uma instituição pública de excelência, como a Universidade de São Paulo, ofereça condições de estudo e pesquisa de notável qualidade e atingisse o atual grau de reconhecimento. Tentarei retribuir o máximo possível.

Por fim, agradeço a Deus, pela dádiva de viver que traz consigo o fardo desafiante de crescer e a saborosa sede de saber.

RESUMO

PRATA G.N. **Um modelo estocástico para a transcrição do gene *even-skipped* de *Drosophila melanogaster***. 2013. 188 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

Nesta tese desenvolvemos um modelo estocástico para a transcrição do gene *even-skipped* de *Drosophila melanogaster* no qual a variável estocástica é o número de moléculas de mRNA transcritas. Nesse modelo, consideramos um gene com dois níveis de ativação sendo regulado externamente. As probabilidades de se encontrar o gene ligado ou desligado e com determinado número de moléculas de mRNA transcritas obedecem equações lineares dadas por processos markovianos de nascimento-e-morte (taxas de produção e degradação) e termos de chaveamento entre os níveis cujas dependências temporais são dadas por funções de Heaviside. Notamos que tal dependência é suficiente para garantir uma atividade transcricional inicial intensa seguida de uma súbita interrupção, conforme sugerem os dados experimentais. Desconsiderando efeitos difusivos e fenômenos de transporte, estendemos esse constructo às outras regiões do embrião, permitindo dependências espaciais apenas às termos de chaveamento, e o resultado gerado descreve os dados experimentais com boa concordância, indicando também que o aspecto binário do gene é suficiente para uma descrição semiquantitativa do fenômeno. Notavelmente, na região onde a listra se forma e concomitantemente a sua formação, o modelo prevê a redução do desvio quadrático (flutuação) e do ruído. Calculando a distribuição de probabilidade, verificamos que o regime estacionário é atingido antes da listra começar a desaparecer. Também estudamos uma conexão entre parâmetros do modelo e as proteínas envolvidas na regulação e, baseado em resultados da literatura, obtemos uma função com aproximadamente o mesmo efeito regulatório considerando gradientes de seis fatores de transcrição (Bcd, Hb, Gt, Kr, Kni e Tll) e apenas quatro sítios de ligação, o que sugere que a informação transcricional pode estar concentrada na regulação de poucos sítios.

Palavras-chave: Expressão gênica. Ruído. Regulação transcricional. Processos estocásticos. *Drosophila*.

ABSTRACT

PRATA G. **A stochastic model to transcription of *Drosophila melanogaster even-skipped* gene.** 2013. 188 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

In this thesis we develop a stochastic model to transcription of *Drosophila melanogaster even-skipped* gene in which the stochastic variable is the number of mRNA molecules transcribed. In this model we considered a gene with two activation levels being regulated externally. The probabilities of gene being *on* or *off* when there is a certain number of transcripts obey linear equations given by Markovian birth-and-death processes (production and degradation rates) and terms of switch between levels whose time-dependence is given by Heaviside functions. We note that is sufficient to ensure a strong transcriptional activity followed by a sudden disruption, as suggested by the experimental data. Disregarding diffusion effects and transport phenomena, we extend this construct to the others regions of the embryo, allowing space-dependence only to terms of switch, and the results describe the experimental data with good agreement, indicating also that binary character of gene is sufficient to a semiquantitative description of the phenomenon. Notably, in the region where the stripe 2 is formed and simultaneously with its formation, the model predicts the reduction in standard deviation (fluctuation) and noise. By calculating the probability distribution, we find that stationary state is reached before stripe 2 starts to fade. We also study a connection between the parameters of the model and proteins involved in regulation and, based on results from the literature, we obtain a function with approximately the same regulatory effect considering six transcription factors (Bcd, Hb, Gt, Kr, Kni e Tll) and only four binding sites, suggesting that transcriptional information may be concentrated in regulation of few sites.

Keywords: Gene expression. Noise. Transcriptional regulation. Stochastic process. *Drosophila*

SUMÁRIO

1	Introdução	13
2	A expressão do gene <i>even-skipped</i> no desenvolvimento embrionário de <i>Drosophila melanogaster</i>	25
2.1	Transcrição, Tradução e Regulação Gênica	25
2.2	Ativadores, Repressores e <i>Enhancers</i>	31
2.3	Embriogênese de <i>Drosophila melanogaster</i>	35
2.4	Dados experimentais da expressão de <i>even-skipped</i>	42
3	A teoria de spins e bósons para expressão estocástica de um gene	53
3.1	O modelo	54
3.2	Probabilidade, valor médio e variância	59
3.3	Distribuição de probabilidade	70
3.3.1	Regime estacionário	70
3.3.2	Regime dependente do tempo	72
4	Análise da segmentação embrionária em <i>Drosophila melanogaster</i>	79
4.1	Dependência temporal nas taxas h e f	80
4.2	Dependência espacial em f	93
4.2.1	Condições iniciais	95
4.2.2	Dependência espacial: tipo “caixa”	99
4.2.3	Refinando dependências espaciais no modelo: encontrando uma solução	103
4.3	Otimizando um modelo para f com dependências espacial e temporal . . .	110

4.3.1	Descrevendo o número de moléculas	113
5	Flutuação, ruído, distribuição de probabilidade e um <i>ansatz</i> para a função q_α	125
5.1	Flutuação e ruído	132
5.2	Distribuição de probabilidade	157
5.3	A função q_α e os gradientes de proteínas: um <i>ansatz</i>	162
6	Conclusões e perspectivas	169
	REFERÊNCIAS	173

Capítulo 1

Introdução

A segunda metade do século XIX é caracterizada por grandes transformações científicas. Particularmente, em biologia, Gregor Mendel (1822-1884) publica seus estudos sobre hereditariedade baseados em experimentos de hibridização de plantas e enuncia as leis de hereditariedade que, surpreendentemente, não tiveram muita repercussão. A redescoberta dos trabalhos de Mendel, 40 anos após sua publicação, por Hugo de Vries (1848-1935), Erich von Tschermak (1871-1962) e Carl Correns (1864-1933) instigou novas pesquisas sobre o fenômeno da hereditariedade. De fato, os mecanismos intracelulares que explicariam como heranças eram transmitidas eram desconhecidos uma vez que Mendel havia trabalhado apenas com fenótipos dos organismos.

Em 1902, Boveri (1862-1915) e Sutton (1877-1916) propõem independentemente que os cromossomos seriam os portadores do material hereditário e que seriam segregados de acordo com as leis Mendel. Estava proposta a primeira formulação da Teoria dos Cromossomos(1) que só se consolidou a partir de 1910 com os trabalhos teóricos e experimentais do geneticista americano Thomas Morgan. Ele propõe que os genes estão contidos nos cromossomos e dispostos nestes como pérolas em um colar. Morgan é responsável por tornar a *Drosophila melanogaster*, espécie popularmente conhecida como *mosca-do-vinagre* ou *mosca-da-banana*, num dos principais organismos-modelo e seus estudos também confirmaram a 1ª Lei de Mendel(2–4).

Além da consolidação dos conhecimentos sobre hereditariedade, os trabalhos de Morgan vislumbram uma relação entre genética e evolução, explorado em trabalhos posteriores

como *Evolution and Genetics*(1925) e *The Theory of the Gene*(1926)(2).

Paralelamente, no intervalo entre as décadas de 1920 e 1950, é desenvolvida o que é chamada hoje de *Teoria Sintética da Evolução*. Nomes como Ronald Fisher, John Haldane, Sewall Wright, teóricos em genética de populações, foram os principais nomes que inicialmente propuseram, como explicação para a origem da diversidade biológica, unificar a seleção natural de Darwin — cuja natureza não-aleatória orienta a evolução incrementando mudanças gradualmente e de modo recursivo (produto de um processo de seleção serve de *input* para uma seleção subsequente) — com a genética mendeliana da hereditariedade — cuja natureza randômica, sem direcionalidade e sem predictibilidade proporciona mudanças súbitas — considerada por geneticistas como William Bateson (1861 – 1926), Hugo de Vries (1848 - 1935) e Wilhelm Johannsen (1857 - 1927). De fato, o desenvolvimento dessa teoria, que representa a união da Genética de Mendel com a Evolução de Darwin, contou com a participação de mais cientistas e de diferentes áreas como, por exemplo, o geneticista Theodosius Dobzhansky (1900-1975), os zoólogos Julian Huxley (1887 – 1975) e Bernhard Rensch (1900 – 1990), o ornitólogo Ernst Mayr (1904 – 2005), o paleontólogo George Simpson (1902 – 1984) e o botânico George Stebbins (1906 – 2000)(5,6).

Também paralelamente, ocorrem a descoberta da radioatividade, a detecção do elétron, do núcleo e a formulação da estrutura da matéria na Inglaterra, o desenvolvimento da mecânica quântica que tinha como um de seus objetivos explicar a interação entre radiação e matéria no continente europeu e cientistas importantes como Niels Bohr (1885-1962), George Gamow (1904-1968) e Leó Szilárd (1898-1964) passam a se pronunciar a respeito da natureza de fenômenos e sistemas biológicos.

Em 1926, Hermann Muller descobre a mutagênese induzida por raios-X em seus estudos sobre *Drosophila*. Em 1935, um modelo para mutações genéticas causadas por raios-X e raios- γ baseado em física atômica é proposto Nikolai Timofeeff, Karl Zimmer e Max Delbrück — um físico cujo interesse em biologia foi despertado por um discurso de Bohr em 1932(7). O artigo ganha destaque na comunidade científica e, em 1937, Delbrück viaja aos EUA para trabalhar com Morgan. Em 1944, em seu livro *What's life?*, o físico

Erwin Schrödinger (1887-1961) analisa tópicos do artigo de Delbrück, Timofeeff e Zimmer bem como discute diversas questões acerca da física em processos biológicos como, por exemplo, a relação entre o número de moléculas presentes em um processo biológico e a precisão deste, e formula a hipótese de que o material genético seria caracterizado estruturalmente por um “cristal aperiódico”(7,8).

O impacto da obra de Schrödinger na comunidade científica é tal que vários cientistas passam a dar atenção à questão molecular do material genético. Entre eles estão o físico Francis Crick (1916–2004) e o biólogo James Watson (1928-) que, em 1953, determinariam a estrutura de dupla-hélice do DNA. É o nascimento da biologia molecular. No ano seguinte, Watson e Gamow fundam o *RNA Tie Club*, um grupo voltado para a discussão sobre como DNA e proteína estariam relacionados. Em 1953, Gamow sugere que um códon seria formado por três bases, o que é verificado no experimento de Crick e Sidney Brenner (1927-) em 1961. Em 1955, Crick, também membro do grupo, propõe a existência de “adaptadores” (hoje chamados de tRNA) para explicar como a informação codificada no DNA especificaria uma sequência de aminoácidos em proteínas. Em 1958, esses “adaptadores” são descobertos por Mahlon Hoagland (1921-2009) e Paul Zamecnik (1912–2009). Em 1964, Har Khorana (1922-2011) e Marshall Nirenberg (1927-2010) descrevem a síntese de tRNA e Robert Holley (1922-1993) descreve a estrutura do tRNA, esclarecendo, assim, uma relação direta entre DNA, RNA e proteína. Isso mostrava que enzimas responsáveis por reações específicas têm estruturas e mecanismos similares em outros seres vivos, evidenciando o caráter universal do código genético. De fato, a descoberta da estrutura de dupla-hélice do DNA e o estabelecimento de uma relação direta entre DNA, RNA e proteína (ou seja, uma relação direta entre o material hereditário que codifica RNA e que, por sua vez, codifica moléculas que executam funções bioquímicas) que exibia a universalidade do código genético.

Com o desenvolvimento da genética, novas questões aparecem. Entre elas, o problema da regulação da expressão gênica, caracterizado pelo fato de que as células de um organismo não produzem proteínas a todo instante e nem em toda e qualquer parte do organismo pressupõe a existência de controle na produção de proteínas. O trabalho

desenvolvido nessa tese está inserido nesse contexto.

Os primeiros modelos para a regulação gênica são propostos após os estudos do ciclo lisogênico de um vírus bacteriófago e da adaptação de bactérias a novas fontes de nutrientes.

O primeiro estudo, realizado no Instituto Pasteur em Paris por André Lwoff (1902-1994) e colegas, descreveu que bactérias *E. coli*, após irradiadas por doses moderadas de luz ultravioleta, paravam de crescer e, cerca de 90 minutos depois, sofriam lise, espalhando no meio de cultura cópias de vírus bacteriófagos (chamados λ), infectando, assim, outras bactérias. Enquanto algumas bactérias sofriam lise pouco depois de serem infectadas e, com isso, liberavam mais cópias de λ , outras bactérias infectadas sobreviviam carregando λ em forma dormente. Essas bactérias cresciam e se dividiam normalmente até que a cultura sofria a incidência de luz ultravioleta, acarretando em lises e liberação de cópias de vírus no meio de cultura. Lwoff e seus colegas François Jacob (1920-) e Jacques Monod (1910-1976) perceberam que os dois estados de virulência — inativo quando a bactéria cresce e se divide e ativo após a cultura ser irradiada — eram um exemplo de um processo biológico fundamental: o ligamento (ativação) e o desligamento (inativação) de genes. De fato, em qualquer instante, cada célula utiliza um subconjunto de seus genes para direcionar a produção de outras moléculas: denominamos como *ligados* os genes que estão sendo expressos e como *desligados* aqueles que não estão sendo expressos e dizemos, então, que a expressão dos genes está sendo regulada(9).

Já o segundo estudo foi realizado com culturas de *E. coli* em um meio composto por várias fontes de açúcar como, por exemplo, uma mistura de glicose e lactose. Monod observou que a cultura apresentava duas fases de crescimento separadas por um determinado intervalo de tempo em que a população não se alterava. Na primeira fase, a população crescia consumindo glicose disponível e, na segunda, consumindo lactose. Monod e Jacob, então, propuseram que a enzima que degrada a lactose (β -galactosidase) é o produto da transcrição de um óperon (chamado óperon *lac*), regulado negativamente por um gene constitutivo que codifica o repressor *lacI*: se não há lactose no meio de cultura, o óperon *lac* encontra-se desligado, com o repressor *lacI* em seu sítio operador; se há lactose no

meio, a atuação do repressor é inativada e o óperon está livre para ser transcrito. Isto é, esse modelo de óperon apresenta um gene que codifica uma proteína repressora que regula a transcrição de um conjunto de genes. De fato, a essência da regulação no modelo de Jacob e Monod reside na hipótese de que a regulação transcricional ocorre na transcrição por meio de proteínas que interagem diretamente com o DNA(10,11).

Historicamente, os estudos sobre regulação gênica também se inserem no contexto da biologia do desenvolvimento de organismos, compreendendo não somente o estudo dos processos e transformações que ocorrem durante a fase e embrião do organismo como também nas mudanças progressivas que um organismo pode sofrer após o nascimento e durante o resto de sua vida. Questões como a diferenciação celular (geração de diversidade celular), morfogênese (criação e distribuição ordenada de estruturas celulares), crescimento, reprodução, evolução e a influência do ambiente sobre o desenvolvimento do organismo são objetos de estudo da biologia do desenvolvimento (12) e, conforme já dito, a regulação gênica se insere naturalmente nesse contexto. O trabalho desenvolvido nessa tese, além de inserido no contexto de regulação gênica, também está inserido nesse contexto uma vez que tratará de um processo transcricional que ocorre durante a embriogênese.

Nesse âmbito, podemos citar os trabalhos de Donald Poulson sobre embriologia em *Drosophila* nas décadas de 1930 e 1940. Poulson chegou a propor um experimento importante que consistia em injetar citoplasma proveniente de um oócito do tipo selvagem em um ovo materno mutante e verificar se a expressão seria recuperada. Tal experimento, que demonstraria o papel fundamental do material genético materno, foi realizado por Alan Gehring e Walter Gehring em 1972 e mostraram que o citoplasma do tipo selvagem recuperaria o fenótipo. Outro nome que merece destaque é Edward Lewis (1918 – 2004). Lewis, que foi aluno de Alfred Sturtevant (que, por sua vez, havia trabalhado com Morgan), instigou a análise genética do desenvolvimento utilizando *Drosophila* como modelo, identificou o complexo *Bithorax* em *Drosophila melanogaster* e elucidou seu funcionamento, caracterizando também uma das primeiras mutações homeóticas (12) e constituindo um paradigma no qual a análise do desenvolvimento de *Drosophila* passou a se basear(13). A veracidade da ideia de famílias de genes estruturalmente relacionados

controlando a especificação do corpo de *Drosophila* apoiou as estratégias em clonagem molecular e caracterização de genes desenvolvimentais em *Drosophila*(13). Em particular, na década de 1980, os trabalhos de Christiane Nüsslein-Volhard e Eric Wieschaus que identificaram um conjunto de genes responsáveis pelo processo de segmentação no embrião de *Drosophila* e também caracterizaram um número grande de casos de embriões com mutações nesses genes(14).

Com relação aos envolvidos no processo de segmentação de *Drosophila*, destacamos trabalhos da segunda metade da década de 1980 e da década de 1990, como os trabalhos de Carroll (1990), Pankratz e Jackle (1990) e de Small e Levine (1991), que observaram que o processo de segmentação em *Drosophila* envolve um refinamento progressivo nos padrões de expressão(16, 17, 130). Mais especificamente, esses padrões foram identificados como resultantes de uma cascata de interações nas quais os genes responsáveis são transcritos de acordo com uma certa hierarquia. Em particular, destacamos também os trabalhos de Harding et al (1986), MacDonald (1986) e de Frasch et al (1987), que elucidam como é formado o padrão de expressão do gene *even-skipped*, caracterizado pela exibição de sete faixas (ou bandas) ao longo do eixo anterior-posterior do embrião(18–20), os trabalhos de Frasch e Levine (1987), Goto et al (1989), Stanojevic (1989, 1991) e Small et al (1991), que apresentaram evidências de que um certo grupo de genes — chamados genes *gap* — estaria envolvido na regulação de *even-skipped* *(22–25). É importante mencionar também

Estudos indicaram que a expressão das faixas de *even-skipped* era regulada por elementos de ação *cis* e, nesse aspecto, os trabalhos de Goto et al (1989), Harding et al (1989) e Small e Levine (1993, 1996) identificaram regiões discretas no DNA como elementos de regulação de faixas individuais da expressão de *even-skipped*(23, 26–28). Nesse sentido, muitas pesquisas foram realizadas e muitas informações a respeito da regulação da faixa

*Hoje é conhecido que os produtos gênicos maternos regulam genes zigóticos da seguinte maneira: os produtos gênicos maternos regulam um grupo de genes, denominados *gap*, cujos produtos gênicos regulam outro grupo de genes zigóticos, denominados *pair-rule*, que, por sua vez, regulam o grupo dos genes *segment-polarity* e que, por fim, regularão os genes homeóticos. Apesar de não termos mencionado explicitamente, os produtos gênicos maternos também podem regular a expressão de genes *pair-rule*, assim como genes *gap* e *pair-rule* também podem regular genes homeóticos. Ademais, também pode ocorrer a regulação entre genes pertencentes a um mesmo grupo(21)

2 da expressão de *even-skipped* foram obtidas de modo que estudos sobre a regulação de *even-skipped* naturalmente foram conduzidas para o estudo da expressão da faixa 2 de *even-skipped*. Particularmente, com relação a esta faixa, os trabalhos de Stephen Small e Michael Levine (1992) e Arnosti et al (1996) identificam os sítios de ligação responsáveis pela regulação da faixa 2 e apontam para as contribuições de cada sítio ao exibirem padrões de expressão de mRNA de *even-skipped* em embriões nos quais sítios específicos sofreram mutagênese(29,30).

Também na década de 1990 podemos destacar os trabalhos de John Reinitz que envolvendo experimentos para estudo da formação de padrões no embrião de *Drosophila* a partir das interações entre genes envolvidos na segmentação bem como desenvolvendo técnicas para aquisição de dados que permitem tratar de forma quantitativa os níveis de expressão a partir das intensidades de fluorescência obtidas via hibridização *in situ* (FISH). Os dados experimentais analisados em nosso trabalho são provenientes de seus trabalhos (31–33).

De modo geral, podemos listar uma série de trabalhos teóricos e experimentais que visam a compreensão dos mecanismos fundamentais para a expressão gênica. Contudo, apesar da biologia molecular ter avançado bastante nesses últimos anos através dos sequenciamentos de DNA, das descobertas de vários genes e interações gênicas, da resolução da estrutura terciária de várias proteínas, ainda sabemos pouco sobre o controle espacial e temporal do processo de síntese proteica ou, mais especificamente, de como se dá a regulação gênica muito embora tenham havido diversas investigações nesse sentido(34–43). Desse modo, são demandados novos conhecimentos, não apenas em Biologia mas também conhecimentos e técnicas provenientes de outras áreas, como a Física e Química, por exemplo, caracterizando a necessidade de interdisciplinaridade no estudo de fenômenos biológicos.

No que se refere às abordagens que visam modelar a regulação transcricional, Delbrück aponta que as flutuações moleculares são significativas nos processos biológicos pelo fato de processos biológicos envolverem muitas interações entre moléculas de várias espécies e pelo fato de ser relativamente pequeno o número de moléculas de cada espécie presente

em uma célula. Dessa forma, uma descrição estocástica é melhor do que uma descrição determinística(44).

Nesse sentido, podemos citar uma série de trabalhos propondo modelagens estocásticas para a regulação gênica.

O primeiro a ser destacado é o trabalho de Gillespie desenvolvendo um algoritmo baseado na formulação estocástica de cinéticas químicas de todas as reações envolvidas. Como as equações obtidas são numerosas e por vezes matematicamente pouco ou nada tratáveis, Gillespie propôs um algoritmo derivado do método de Monte Carlo e que considera as flutuações aleatórias para simular numericamente a evolução temporal de um sistema bioquímico(45,46). Como exemplo da utilização do algoritmo de Gillespie, podemos citar os trabalhos de MacAdams e de Becskei, realizando simulações numéricas para analisar redes gênicas(47–49).

Destacamos também o trabalho de Shea, Ackers e Johnson que, no estudo do bacteriófago λ , implementa pela primeira vez métodos termodinâmicos para o estudo da regulação gênica (50, 51). Este modelo é aplicado posteriormente em sistemas nos quais a regulação envolve simples eventos de ligação ao DNA(52–54).

Uma segunda abordagem que destacamos é utilizada por Collins, Elowitz e Paulsson, que trabalha com equações diferenciais para as concentrações dos reagentes bioquímicos e introduz as flutuações aleatórias através de um termo perturbativo estocástico dependente do tempo via mecanismo de Langevin (55–59).

Outra alternativa de abordagem estocástica é a descrição microscópica que, no lugar de lidar com as inúmeras reações envolvidas, trabalha com o número de partículas transcritas como sua variável estocástica. Na primeira classe de modelos desse tipo, proposta por van Oudenaarden (60,61), o gene apresenta um único estado de ativação e a transcrição, dada por um processo de produção e degradação de mRNA, é acoplada ao processo de tradução. A segunda classe de modelos microscópicos é proposta por Wolynes. Vislumbrando o entendimento de redes genéticas, é proposta a exploração do problema da regulação gênica por meio de uma analogia com sistemas quânticos de muitos corpos. O modelo consiste essencialmente em mapear a dinâmica do chaveamento gênico em termos do modelo de

spin-bóson de um pólaron de dois estados (ou à reação de transferência de um elétron) de modo que redes gênicas possam ser aproximadas por sistemas de spins(62). Do ponto de vista do gene, é introduzido um segundo nível de ativação e termos de acoplamento entre os níveis de modo que o gene pode migrar de uma configuração favorável à transcrição para outra desfavorável à transcrição e vice-versa.

A partir de 2005, os trabalhos do grupo coordenado pelo Professor José Eduardo Hornos, do Instituto de Física de São Carlos da Universidade de São Paulo (IFSC - USP), passaram a explorar o modelo proposto por Wolynes. Com a utilização de funções complexas, foram obtidas soluções exatas para os casos de autorregulação e regulação externa tanto para regime estacionário quanto para o regime dependente do tempo(63–66) e suas simetrias foram identificadas(67,68).

De acordo com a linha de pesquisa do grupo, nesta tese trabalharemos com o modelo proposto por Wolynes com o intuito de analisar os dados experimentais advindos da transcrição da faixa 2 do mRNA de *even-skipped* que ocorre durante a embriogênese de *Drosophila melanogaster*(32). A abordagem utilizada caracteriza-se por um modelo estocástico que considera a dinâmica de chaveamento de um gene binário (dois estados de ativação) regulado externamente e do número de moléculas de mRNA transcritas. A variável estocástica refere-se ao número de moléculas de mRNA transcritas e estabelece que as probabilidades do gene estar em alguma configuração favorável à transcrição (nível ligado) ou desfavorável à transcrição (nível desligado) quando houver um determinado número de transcritos obedecem equações que descrevem processos de nascimento-e-morte (representados pelas taxas de produção e de degradação de mRNA) combinados com termos lineares de acoplamento dos níveis (chaveamento) cujas dependências temporais são descritas por funções de Heaviside. Nosso objetivo é fornecer uma descrição fenomenológica do processo transcricional da faixa 2 do mRNA de *even-skipped* e, para isso, buscaremos obter, além da descrição do número de moléculas de mRNA transcritas no decorrer do tempo e ao longo do eixo anterior-posterior do embrião de *Drosophila melanogaster*, a distribuição de probabilidade do número de transcritos, quantidade que concentra todas as informações do processo e nos permite acompanhar a evolução do modo de produção

de mRNA. Também calculamos o desvio quadrático (flutuação) e o ruído da expressão transcricional, quantidades que também podem ser mensuradas experimentalmente e que têm grande relevância para a descrição e o entendimento do fenômeno. E, por fim, estudamos uma conexão entre os parâmetros do modelo e as principais espécies moleculares envolvidas na transcrição da faixa 2 e propomos uma função em termos dos gradientes de concentração das proteínas regulatórias com o intuito de gerar o mesmo efeito regulatório que o concebido durante o desenvolvimento modelo.

A apresentação desta tese está organizada nos seguintes capítulos. No Capítulo 2, apresentamos o fenômeno de interesse, a transcrição da faixa 2 do mRNA do gene *even-skipped* de *Drosophila*. Para isso, inicialmente fazemos uma breve introdução sobre a transcrição e a tradução, fenômenos envolvidos na síntese proteica, e sobre a regulação (modulação) que tais processos podem sofrer. Posteriormente, direcionamos nossa atenção para a regulação no início da transcrição, mais precisamente na participação de proteínas que, ao interagirem com o DNA, acabam regulando a produção de moléculas de mRNA. Na seção seguinte, apresentamos brevemente a embriogênese de *Drosophila*, contextualizando biologicamente o fenômeno o qual analisaremos nessa tese. Na última seção desse capítulo, apresentamos os dados experimentais a serem analisados. No Capítulo 3, apresentamos a teoria de spins e bósons para a expressão estocástica de um gene, a qual pode modelar um gene binário regulado por um tipo de proteína que pode se ligar e se desligar de um sítio operador. Uma vez concebido o modelo, calculamos, através do uso de funções geratriz, o primeiro e segundo momentos da distribuição de probabilidade para condições iniciais arbitrárias, fornecendo, assim, o número médio de moléculas de mRNA transcritas, a flutuação e o ruído da expressão, e reproduzimos o cálculo da distribuição de probabilidade, realizado primeiramente por Innocentini e Hornos(65). No Capítulo 4, analisamos objetivamente o problema da transcrição da faixa 2 de mRNA do gene *even-skipped* e desenvolvemos um modelo estocástico baseado no modelo apresentado no Capítulo 3. No Capítulo 5, calculamos o desvio quadrático (flutuação), o ruído e a distribuição de probabilidade do processo transcricional e, na tentativa de aprimorar o suporte biológico do modelo, buscamos uma conexão entre os parâmetros do

modelo e objetos biológicos observáveis, como os gradientes de concentração das proteínas regulatórias. Por fim, dedicamos o capítulo 6 a sumarizar os resultados, discutir as conclusões e apresentar perspectivas de estudo que se abrem com o modelo desenvolvido nesse trabalho.

Capítulo 2

A expressão do gene *even-skipped* no desenvolvimento embrionário de *Drosophila melanogaster*

Nesse capítulo, tratamos inicialmente da síntese proteica, mecanismo fundamental para o metabolismo e sobrevivência de um ser vivo, e de sua regulação, presente em todo esse processo. Dentre os mecanismos-chave que compõem a síntese proteica, transcrição e tradução, focamos na transcrição. Mais especificamente, nossa atenção está voltada para a iniciação da transcrição, que compreende as circunstâncias necessárias para que esse processo ocorra, e para as proteínas que, ligando-se ao DNA, regulam tal processo. Posteriormente, traçamos um breve resumo da embriogênese de *Drosophila melanogaster* e da expressão do gene *even-skipped*, determinante para o processo que determina os segmentos do corpo do indivíduo adulto e objeto de estudo do nesse trabalho. Por fim, apresentamos os dados experimentais da expressão de *even-skipped* com os quais trabalharemos.

2.1 Transcrição, Tradução e Regulação Gênica

As proteínas são peças fundamentais para o funcionamento, metabolismo e sobrevivência das células. No entanto, as células não obtêm geralmente as proteínas necessárias para sua manutenção e funcionamento diretamente do ambiente e, por isso, devem ser capazes de sintetizar, a partir dos aminoácidos obtidos através dos alimentos, as proteínas específicas das quais necessitam. E é o DNA, uma estrutura molecular composta por bases nitrogenadas (adenina, timina, citosina e guanina) dispostas em duas fitas complementares* e pareadas que acompanham uma forma helicoidal, como mostra a Figura 2.1, o portador das informações sobre as proteínas que deverão ser sintetizadas na célula. Na realidade, o DNA contém as versões codificadas das proteínas a serem sintetizadas em sequências denominadas *genes*.

Figura 2.1 – O DNA é uma estrutura composta por duas fitas formadas por quatro tipos de bases nitrogenadas, adenina (A), timina (T), guanina (G) e citosina (C). Estas fitas se acoplam seguindo uma estrutura helicoidal de modo que as bases A, T, G e C de uma fita sejam pareadas respectivamente com as bases T, A, C e G da segunda fita. Imagens extraídas de (9).

A síntese de proteínas segue um roteiro bem estabelecido através de dois processos principais, *transcrição* e *tradução*, compondo um cenário mais amplo que estabelece o sentido da transferência de informações entre as três classes de biomoléculas (DNA, RNA e proteína), conforme mostra as Figuras 2.2 e 2.3.

Primeiramente, na transcrição, um complexo enzimático encabeçado pela enzima RNA-polimerase sintetiza uma sequência complementar[†] à sequência do gene, chamada RNA, conforme mostra a Figura 2.4. Depois, na tradução, esta sequência de RNA, for-

*As duas fitas que compõem o DNA permanecem ligadas através de pontes de hidrogênio estabelecidas entre as bases nitrogenadas. A complementaridade está em dizer que adenina (A) se liga a timina (T) enquanto que guanina (G) se liga a citosina (C).

[†]No caso da transcrição, adenina (A) presente numa sequência de DNA é pareada com uracila (U). O pareamento entre guanina e citosina é inalterado.

Figura 2.2 – Relações existentes entre as três classes de biomoléculas (DNA, RNA e proteína) e o sentido das transferências de informações entre elas. *Transferências gerais* (setas laranjas), que ocorrem normalmente na maioria das células: replicação do DNA (DNA → DNA), transcrição (DNA → RNA) e tradução (RNA → proteína). *Transferências especiais* (setas verdes), que ocorrem somente sob condições específicas: transcrição reversa e replicação do RNA. Imagem extraída de (69).

Figura 2.3 – Esquema do processo de síntese de uma proteína em eucariotos. No núcleo ocorre a primeira parte do processo, que é a transcrição da informação contida em uma sequência de DNA em uma sequência complementar, RNA. O RNA transcrito é processado e transportado para o citoplasma, onde ocorrerá o segunda parte, a tradução. (Em procariotos, uma vez que o mRNA está transcrito, ele já está pronto para servir de molde para a síntese de proteína; portanto, os processos indicados na figura — “capping” e transporte para fora do núcleo, entre outros — não ocorrem(70).) Na tradução, o ribossomo sintetiza, a partir da fita de RNA, uma cadeia polipeptídica que posteriormente configurará a proteína. Imagem extraída de (71).

mada por trincas de nucleotídeos, é convertida em aminoácidos, estruturas que formam proteínas.

Como o DNA está presente em todas as células de um organismo, todas as células deste organismo têm, a princípio, condições de sintetizar todas as proteínas que estão

Figura 2.4 – Na transcrição, o complexo enzimático monta uma fita de RNA (azul), uma sequência complementar à sequência lida de DNA (cinza) no sentido de que as bases A, T, G e C da sequência de DNA sejam pareadas respectivamente com U (uracila), A, C e G. Evidentemente, durante a transcrição, a região por onde passa a RNA-polimerase deve ter as ligações entre as bases interrompidas para que a transcrição ocorra. Imagens extraídas de (9).

codificadas no DNA. No entanto, não é isto que ocorre: mesmo possuindo DNA e, com isso, a possibilidade de sintetizar todas as proteínas, as células de um organismo não expressam todas as proteínas ao mesmo tempo[‡].

De fato, muito da vida depende da capacidade das células de expressar seus genes em diferentes combinações em diferentes momentos e em diferentes locais. Se, de um lado, uma simples bactéria expressa somente alguns de seus genes em um dado momento (o que lhe garante sintetizar enzimas necessárias para o metabolismo de certos nutrientes disponíveis enquanto não estiver sintetizando enzimas para metabolizar outros nutrientes), de outro lado, o desenvolvimento de organismos multicelulares exhibe uma modulação tanto espacial quanto temporal da expressão gênica. É claro que, imediatamente, poderíamos pensar que as expressões proteicas controladas espacial e temporalmente devem-se à condição dos tipos celulares já atingido seus destinos celulares. No entanto, podemos ver que a regulação gênica também está inserida no contexto de especificação celular como, por exemplo, gradientes morfogenéticos (proteínas ou mRNAs) que, uma vez distribuídos em diferentes regiões do embrião, são responsáveis por padrões de expressão, o que os significa que atuam no processo de especificação o tipo celular(12), o que, por sua vez, classifica a regulação da expressão genética como um fenômeno de

[‡]Por exemplo, uma célula muscular expressa um conjunto de genes diferentes, pelo menos em parte, daquele expresso por um neurônio — e não necessariamente ao mesmo tempo — apesar de ambos possuírem os mesmos genes.

extrema relevância. Em geral, a origem dessas modulações ocorre no nível da transcrição (mais comumente, na iniciação da transcrição)(70). Toda a modulação da síntese de um transcrito gênico e de seu produto proteico denomina-se *regulação gênica* (ou *regulação genética*).

De modo geral, qualquer passo da expressão de um gene pode ser regulado. Isso significa que, durante todo o caminho percorrido pelas informações que codificam proteínas, podem ocorrer modificações. Estas modificações podem ser alterações químicas e estruturais no DNA e/ou na cromatina, alterações na transcrição (ou mesmo depois desta), no transporte e degradação de RNA e mesmo após a tradução e influenciarão, consequentemente, o produto gênico e a expressão gênica finais.

No entanto, neste trabalho, concentramo-nos no estudo da regulação gênica durante a transcrição ou, mais especificamente, no início da transcrição. Contudo, antes de prosseguir, é conveniente traçarmos um breve esquema do processo transcricional de um gene uma vez que alguns nomes e conceitos mais específicos poderão ser frequentemente invocados ao longo deste trabalho.

Como já dissemos anteriormente, genes são transcritos em RNA pela enzima RNA-polimerase e, em termos gerais, os RNAs formados a partir da transcrição podem ser de três tipos: mRNA (mensageiro), que contém a seqüência que codifica uma proteína e é o RNA com o qual lidaremos durante todo o trabalho; tRNA (transportador), que conduz os aminoácidos até os ribossomos e possibilita a leitura da informação contida no mRNA durante a tradução; e rRNA (ribossômico), que faz parte da estrutura dos ribossomos.

De modo geral, o processo de transcrição de um gene é essencialmente o mesmo em organismos procariotos e eucariotos, como mostra a Fig.2.5. O processo se inicia com a ligação da RNA-polimerase a uma região próxima do início do gene de aproximadamente 60 pares de bases (bp) de extensão, chamada *promotor*. Uma vez que a RNA-polimerase se liga ao promotor, um segmento de DNA tem suas ligações desfeitas de modo que as duas fitas fiquem relaxadas para que uma delas seja usada como molde de RNA e a inserção de bases para compor a fita de RNA segue de modo que é sempre a mesma parte da RNA-polimerase que fica em contato o segmento relaxado de DNA. No final do processo,

Figura 2.5 – Esquema simplificado de um gene submetido ao processo de transcrição. **2.5(a)** O promotor, a RNA-polimerase e o gene. A RNA-polimerase liga-se ao promotor e inicia a transcrição do gene no sentido indicado pela seta. **2.5(b)** Dinâmica do processo transcricional. Imagens extraídas de (9).

a fita de RNA construída é liberada. Esta representação simplificada, mas paradigmática, é a que adotamos no decorrer desse trabalho.

Sem dúvida, esta representação não leva em consideração um conjunto imenso de fatores que tornam o processo de transcrição por si só bastante complexo. Um exemplo está no fato de que o mecanismo transcricional de um organismo eucarioto (no nosso caso, *Drosophila melanogaster*) precisa de vários fatores, chamados de *fatores de transcrição gerais*, que atuam mediando a aproximação, reconhecimento e ligação da RNA-polimerase ao DNA a fim de uma iniciação eficiente (em bactérias, por exemplo, é necessário somente um fator de iniciação, conhecido como fator σ)(70,72). Além disso, apenas os fatores de transcrição gerais não são, no entanto, suficientes para promover uma expressão significativa de modo que fatores adicionais são necessários, como proteínas regulatórias (também denotadas como *fatores de transcrição específicos*, que se ligam ao DNA), o Complexo Mediador(70,73,74) e enzimas modificadoras da cromatina (que, ao deformarem a cromatina, alteram a acessibilidade do DNA à ação de proteínas regulatórias e à ação da RNA-polimerase)(70).

Uma vez adotada a representação mostrada pela Figura 2.5 para o estudo da regulação transcricional, consideraremos, dentre todos os fatores envolvidos na iniciação da transcrição, somente o papel desempenhado pelas proteínas regulatórias.

2.2 Ativadores, Repressores e *Enhancers*

No processo transcricional, há proteínas regulatórias que, estando ligadas ao DNA, podem estimular a transcrição ou podem dificultar (ou mesmo impedir) sua iniciação. As proteínas do primeiro tipo são chamadas de *ativadoras* (ou *regulatórias positivas*) enquanto que as proteínas do segundo tipo são chamadas *repressoras* (ou *regulatórias negativas*). Mais precisamente, estas proteínas se ligam a sequências não-codificantes (denominadas *elementos cis-regulatórios*[§]) denominadas *sítios operadores* (para procariotos) ou *sítios de ligação* (para eucariotos). Como estudaremos o processo transcricional em um organismo eucarioto, denotaremos as sequências regulatórias como sítios de ligação.

Do ponto de vista fenomenológico, uma proteína ativadora, ao se ligar a um sítio de ligação, facilita ou estimula a aproximação da ação da RNA-polimerase ao promotor e, portanto, contribui para a ocorrência do processo transcricional, conforme a Figura 2.6(a). Já uma proteína repressora, segundo a Figura 2.6(b), compromete a transcrição (podendo mesmo anulá-la) ao se ligar em um sítio de ligação.

Figura 2.6 – Esquema de atuação de uma proteína regulatória. **2.6(a)** A ligação de uma proteína ativadora ao sítio de ligação faz com que a transcrição de um gene ocorra (ou seja amplificada, no caso de já haver uma transcrição à uma taxa basal). **2.6(b)** Quando a proteína repressora se liga ao sítio de ligação, a transcrição de um gene, inicialmente ativo, é atenuada ou mesmo anulada. Imagens extraídas de (71).

Do ponto de vista molecular, a ativação transcricional pode se dar através de uma interação direta entre a proteína ativadora (ligada ao sítio de ligação) e a RNA-polimerase, através de uma interação entre a proteína e um complexo mediador que, por sua vez, interage com a RNA-polimerase e também através de deformações locais na cromatina (provocadas pela ocupação do sítio de ligação) deixando o promotor mais exposto à ação da RNA-polimerase. Por outro lado, no caso da proteína repressora ligada ao sítio de

[§]No DNA, além dos genes, há também sequências que não codificam nem RNA e nem proteínas mas que, por exemplo, atuam como sítios de ligação para proteínas regulatórias, por exemplo.

ligação, podem ocorrer, além da interação direta dessa proteína com a RNA-polimerase, a interação da proteína com complexos proteicos que, por sua vez, interagem com a RNA-polimerase (impedindo que ela inicie a transcrição), a interação com complexos que atenuam os efeitos de ativadores já ligadas a sítios vizinhos relativamente próximos — efeito conhecido na literatura como *quenching*(30, 70, 75, 76) — ou, como sugerem alguns estudos, a interação com a própria maquinaria transcricional, mesmo na presença de proteínas ativadoras, comprometendo algum passo da iniciação da transcrição (como a ligação da RNA-polimerase ao DNA, isomerização do complexo DNA-proteína e a formação de ligações fosfodiéster, mecanismos importantes para a concretização do processo transcricional)(75).

Enquanto o controle transcricional em procariotos é geralmente mediado por algumas unidades de sítios de ligação, a região de controle de um gene eucarioto frequentemente envolve centenas ou milhares de sítios de ligação, em que muitos destes ligam-se simultaneamente a fatores perfazendo um conjunto de mecanismos de regulação gênica consideravelmente mais complexo, conforme mostram as Figuras 2.7(a) e 2.7(b). Na Figura 2.7(a), vemos que a flexibilidade da molécula de DNA permite que reguladores ligados a sítios de ligação relativamente distantes de dezenas ou centenas de pares de base podem interagir simultaneamente com moléculas que, por sua vez, interagem com a RNA-polimerase de modo que a transcrição seja iniciada[¶]. A Figura 2.7(b) implicitamente sugere que, embora eucariotos possuam genomas com um número de bases e de genes maiores do que os procariotos, a densidade de genes nesses organismos é menor(77, 78). Além disso, é importante dizer que, em eucariotos, a maior parte do genoma é representada por sequências não-codificantes (no genoma humano, por exemplo, mais de 98%)(79) corresponde a material não-codificante e isso está diretamente relacionado à regulação(70, 78, 79).

Frequentemente, estes sítios de ligação estão agrupados em módulos reguladores denominados *enhancers*, tipicamente da ordem de 100 a 1000 pares de bases de comprimento.

[¶]A distância entre sítios de ligação é medida em termos do número de pares de bases que os separa, imaginando a molécula de DNA como uma estrutura retilínea. Contudo, como a conformação espacial da molécula de DNA não é retilínea, sítios distantes, por exemplo, de centenas de pares de base (no cenário em que a fita de DNA está esticada) podem estar separados, na verdade, por uma distância menor.

Figura 2.7 – 2.7(a) Exemplo de como a regulação em eucariotos pode ser mediada através de várias proteínas regulatórias ligadas a vários sítios de ligação. Nessa representação, o fato de as proteínas estarem ligadas ao DNA e este não apresentar um perfil retilíneo, permite com que as proteínas regulatórias (no caso, ativadoras) interajam com uma molécula mediadora que, por sua vez, interage com a RNA-polimerase, facilitando a transcrição. Além das proteínas regulatórias, essa representação também contempla, mesmo que de forma pictórica, o recrutamento de complexos remodeladores da cromatina, do Complexo Mediador bem como de fatores de transcrição gerais associados à enzima RNA-polimerase. Imagem extraída de (81). **2.7(b)** Ilustração mostrando o aumento de complexidade no que se refere ao número de sequências regulatórias (regiões azuis) e em sua disposição ao longo do DNA (aqui representado na forma de um tubo). Em cada caso, o promotor (região roxa) é mostrado no sítio onde a transcrição é iniciada. Considerando que o sentido da transcrição (ou seja, o sentido do trajeto da RNA-polimerase) se dá no sentido 5' → 3', conforme indica a seta, vemos que enquanto no DNA bacteriano a sequência regulatória se encontra próxima à extremidade 5' do promotor (*upstream*), em um DNA eucariótico de um organismo mais complexo (exemplificado aqui pelo DNA humano), as sequências regulatórias podem se encontrar situadas tanto na direção 5' (*upstream*) quanto na direção 3' (*downstream*). Imagem extraída de (70).

Módulos reguladores aparecem pela primeira vez na literatura no trabalho de Banerji e colaboradores(80). Neste trabalho, eles notaram que a presença de sequências de um DNA viral aumentou em 200 vezes^{||} a transcrição de um gene de β -globina de coelho, após a inserção destes dois elementos no DNA de células HeLa. E mais, notaram também que ocorria a transcrição gênica independentemente da posição e orientação destas sequências em relação ao sítio de iniciação transcritcional. Após este trabalho, outros trabalhos constataram a existência de módulos reguladores em outros organismos^{**}. A literatura então nos diz que os *enhancers* são classicamente definidos como elementos de DNA regulatórios de ação *cis*(ou seja, não codificam RNA nem proteínas regulatórias mas participam da regulação ao interagirem com elementos ativadores e repressores) que estimulam a transcrição, independentemente de suas posições e orientações com respeito

^{||}Daí o nome *enhancer*, que pode ser traduzido do inglês como *acentuador*. Apesar disso, ao longo deste trabalho, seguiremos utilizando o termo *enhancer*.

^{**}Embora a descoberta e a própria descrição original dos *enhancers* fossem vinculadas à idéia de ativação, muitos *enhancers* celulares descobertos, no entanto, apresentam interações com ativadores e repressores. Deste modo, uma descrição completa destes elementos regulatórios deve necessariamente levar em conta ativação e repressão.

ao sítio de iniciação da transcrição.

Desde o início, *enhancers* sempre estiveram associados a ideia de alta aglomeração de sítios de ligação que estão próximos uns dos outros ao longo da fita de DNA. No entanto, vários estudos(32,82–84) indicam um cenário mais geral para a ideia de *enhancer*, como sugere a Fig.2.8.

Figura 2.8 – Diferentes arranjos para sítios de ligação. **a.** Concepção típica de um *enhancer*: um aglomerado denso de sítios relativamente próximos ao longo da fita. **b.** Vários sítios de ligação não necessariamente próximos ao longo da fita que, devido a modificações na cromatina, acabam se aproximando. **c.** Um *enhancer* composto. **d.** Sítios espalhados ao longo da fita, não necessariamente restritos a uma pequena região do DNA. Imagem extraída de (83).

Um ponto interessante dos *enhancers* é como se dá seu funcionamento. Recentes estudos têm fornecido novos conhecimentos sobre o tipo de informação processada que ocorre nos *enhancers*(85). Basicamente, há dois modos que tentam explicar o funcionamento de *enhancers*. O primeiro propõe que o *enhancer* e os reguladores associados a ele atuem coordenada e cooperativamente, de modo que alterações mínimas em sítios de ligação podem acarretar em drásticos efeitos na transcrição. Isso significa que o funcionamento deste complexo (nomeado *enhanceosome*) é mais do que a soma de contribuições de fatores individuais, emergindo de toda uma rede de interações. Já o segundo modo de funcionamento de um *enhancer* sugere sítios de ligação mais flexivelmente dispostos de modo que as proteínas ligadas ao *enhancer* não operariam como uma unidade mas como um conjunto de fatores separadamente agentes ou como pequenos grupos de fatores que

independentemente interagiriam com seus alvos, lendo e interpretando as informações apresentadas pelo *enhancer*, como se este fosse um painel. Deste modo, a flexibilidade dos sítios seria suficiente para conseguir iniciar uma transcrição mesmo com alterações em sítios de ligação.

Em nosso trabalho, investigaremos este cenário apresentado pelas proteínas ativadoras e repressoras mais especificamente no contexto da transcrição do gene *even-skipped* de *Drosophila melanogaster*. Para isto, antes de iniciarmos tecnicamente a investigação deste assunto, trataremos a seguir, mesmo que de forma resumida e simplificada, dos aspectos básicos, mas importantes, da embriogênese da *Drosophila melanogaster*, um dos mais conhecidos organismos modelo e do qual é extraída uma quantidade considerável de informações a respeito de mecanismos genéticos.

2.3 Embriogênese de *Drosophila melanogaster*

A embriogênese da *Drosophila melanogaster* é um processo que leva um dia, compreendendo desde a fecundação do ovo até a incubação do primeiro estágio larval. A geometria do embrião (assim como a do óvulo não fecundado) pode ser aproximada por um elipsóide cujos eixos maior e menor corresponderão, respectivamente, aos eixos anterior-posterior (AP) e dorsal-ventral (DV). Após a fecundação, o núcleo diplóide sofre 13 rápidas divisões mitóticas, sendo que as divisões são quase sincrônicas(70). Para melhor compreensão, divide-se estágio em 13 ciclos de clivagens, onde o n -ésimo ciclo compreende o final da $(n - 1)$ -ésima mitose e o final da n -ésima mitose. Os primeiros 9 ciclos duram, cada um, cerca de 10 minutos. Durante este tempo, uma fração de núcleos dirige-se para a região periférica do embrião (denominada *córtex*). Este movimento inicia-se durante o ciclo 7 e termina no início do ciclo 10. Uma segunda fração de núcleos torna-se gema e uma terceira fração torna-se células polares(86). Esse estágio embrionário, caracterizado pela maioria dos núcleos localizados na região periférica do embrião, é denominado

blastoderme. A configuração caracterizada por núcleos sem membranas celulares recebe o nome de *sincício*. No ciclo 14A, a população de núcleos atinge a marca de cerca de 6000 unidades(70). O ciclo 14A é o último ciclo antes da gastrulação, estágio embrionário em que ocorre um movimento de invaginação em uma região específica do embrião formando uma cavidade (chamada arquênteron) e as três camadas germinativas (ectoderme, mesoderme e endoderme, responsáveis pela formação de tecidos, órgãos no indivíduo adulto). (70,87)

Nosso interesse está focado especificamente no ciclo 14, cuja duração é cerca de 50 minutos(87). Os dados coletados da expressão do gene *even-skipped* e que serão estudados nesse trabalho foram obtidos durante este período. É importante mencionar que o fenômeno da celularização também ocorre durante esse ciclo(70).

Figura 2.9 – Imagem composta por uma série de embriões de *Drosophila* dispostos de forma cronológica de acordo com o número de clivagens sofridas após a fecundação do óvulo. A orientação dos embriões é tal que a parte anterior está apontada para cima. Os núcleos estão representados pelos pontos brancos e a sequência numérica indica a cronologia da série de clivagens. De 1 para 2 ocorre a primeira clivagem, compreendendo portanto o primeiro ciclo; de 2 para 3, ocorre a segunda clivagem e assim por diante. (A partir de 11, é possível observar bem que a concentração de núcleos é maior na periferia do embrião, configurando o estágio de sincício.) Assim, a partir de 14 é iniciado o ciclo 14, a partir do qual ocorrem o fenômeno da celularização e a expressão experimentalmente observada do gene de segmentação *even-skipped* cuja regulação é objeto de nosso estudo. Imagem extraída de (70).

A celularização caracteriza-se pelo processo de formação de membranas celulares em torno dos núcleos e enquanto a celularização não está concluída, os núcleos não estão isolados e proteínas são trocadas entre estes.

Um dos vários processos que ocorrem durante o ciclo 14A é o de definição de domínios ao longo do embrião diretamente relacionados à formação e, conseqüentemente, desenvolvimento dos segmentos observados no corpo do indivíduo adulto. Morfologicamente, o corpo de uma espécie de *Drosophila melanogaster* consiste de 14 segmentos (3 na cabeça, 3 no tórax e 8 no abdômen) que são determinados durante as primeiras horas após a fecundação. O processo de segmentação do corpo do organismo resulta da expressão de um grupo de genes denominados *genes de segmentação*.

Em particular, estamos interessados na expressão de um desses genes, chamado *even-skipped* ou simplesmente *eve*, que codifica uma proteína, Eve, cuja expressão é caracterizada por um padrão de faixas ou bandas — sete, ao todo — ao longo do eixo anterior-posterior do embrião (eixo AP), como ilustra a Fig.2.10^{††}. O padrão de expressão da

Figura 2.10 – Padrão de expressão da proteína Eve ao longo do eixo anterior-posterior (anterior à esquerda e posterior à direita) do embrião de *Drosophila*. Extraído de (70).

proteína Eve observado no embrião é resultante de uma “cascata” de interações entre genes zigóticos e maternos interagentes. Esta cascata é hierarquicamente organizada em níveis de genes zigóticos, como mostra a Figura 2.11. O primeiro nível é constituído pelos genes *gap*, que são regulados pelos produtos gênicos dos genes maternos, cujos mRNAs já se encontram transcritos no citoplasma do óvulo antes da fecundação^{††}. O segundo nível é constituído pelos genes *pair-rule*, cuja expressão é regulada pelos produtos gênicos dos genes *gap* e dos genes maternos. Este é o nível ao qual pertence o gene *eve*.

^{††}A notação adotada para símbolos de genes e proteínas é a seguinte: o nome ou símbolo da proteína deve ser escrito com a primeira letra maiúscula e o do gene deve ser escrito em itálico. Mesmo em casos em que o nome ou símbolo do gene é originalmente escrito com letra maiúscula a regra é mantida, como no caso do gene *Krüppel* (*Kr*) que codifica a proteína Krüppel (Kr). Outros exemplos: gene *eve* e proteína Eve; gene *bicoid* (*bcd*) e proteína Bicoid; gene *hunchback* (*hb*) e proteína Hunchback (Hb).

^{††}Num primeiro momento, a síntese de certas proteínas no zigoto ocorre a partir de mRNAs já presentes no citoplasma antes da fecundação do óvulo. Com o decorrer do processo embrionário e a degradação dos mRNAs maternos, algumas dessas proteínas eventualmente precisam continuar sendo produzidas de modo que passam a ser codificadas a partir do DNA do próprio embrião.

Figura 2.11 – Cascata de interações entre os genes envolvidos no processo de segmentação do corpo de *Drosophila melanogaster*. Genes maternos regulam a transcrição dos genes zigóticos *gap*. Produtos gênicos dos genes maternos e *gap* regulam genes *pair-rule*. A expressão dos genes *pair-rule* regula a expressão dos genes *segment-polarity*. E os produtos gênicos dos genes *gap*, *pair-rule* e *segment-polarity* regulam os genes homeóticos que, uma vez estabelecidos os limites de cada segmento, especificam suas estruturas, sendo, desse modo, responsáveis pela identidade de cada segmento. É importante mencionar que genes de um mesmo nível podem estabelecer vínculos regulatórios. Imagem extraída de (21).

Os produtos gênicos dos genes *pair-rule* regulam os genes que compõem o terceiro nível, chamado de genes *segment-polarity*^{§§}. E os produtos gênicos dos genes *gap*, *pair-rule* e *segment-polarity* regulam os genes homeóticos, que são genes responsáveis pela identidade de cada segmento, especificando suas estruturas uma vez estabelecidos os limites de cada segmento. É importante mencionar que pode haver regulação entre genes pertencentes ao mesmo nível(21, 88, 89). O nome de cada nível de genes zigóticos deve-se ao fenótipo gerado devido a uma mutação sofrida no gene do nível em questão, constatado experimentalmente em (14) e esquematizado na Figura 2.12. Há vários genes pertencentes a cada um desses níveis. No entanto, como estamos interessados somente na expressão do gene *eve*, devemos considerar, como veremos adiante, somente alguns genes de cada nível, aqueles que participam da regulação da transcrição de *eve*.

Em termos dos níveis de expressão de proteínas ao longo do eixo anterior-posterior, os genes *gap* geram um a três vastos domínios de expressão, os genes *pair-rule* aparecem

^{§§}Embora os nomes dos níveis *pair-rule* e *segment-polarity* possam ser traduzidas — “genes regra-dos-pares” e “genes de polaridade do segmento”, respectivamente — nós optamos, pelo fato de haver uma certa familiaridade advinda dos muitos trabalhos nessa área escritos em inglês, por utilizar ao longo desse trabalho as denominações “*pair-rule*” e “*segment-polarity*”.

Figura 2.12 – Comparações entre fenótipos do tipo selvagem e fenótipos gerados decorrentes de mutações em genes *gap*, *pair-rule* e *segment-polarity*. Para cada gene de cada grupo são mostrados (à esquerda) o fenótipo do tipo selvagem e (à direita) o fenótipo resultante de uma mutação no referido gene. Os retângulos rosas destacados nos fenótipos selvagens representam os domínios do eixo AP que estão ausentes nos embriões mutantes. No primeiro grupo apresentado, são exibidas as comparações envolvendo os fenótipos resultantes de mutações nos genes *Krüppel* e *knirps*, ambos genes *gap*. No segundo grupo, as comparações envolvem os genes *even-skipped*, *odd-skipped*, *paired* e *runt*, todos genes *pair-rule*. E no terceiro grupo, as comparações envolvem os genes *gooseberry* e *patched*, genes *segment-polarity*. Imagem extraída de (90).

em sete faixas transversas e os genes de polaridade (*segment-polarity genes*) aparecem em padrões de quatorze faixas por quase a toda a célula(90)^{¶¶}. Este padrão inicial dos segmentos determinará as posições das extensões dos parasegmentos, que é, por definição, a região compreendida entre a metade posterior de um segmento até a metade anterior do segmento imediatamente adjacente. Determinados os parasegmentos, tornam-se morfológicamente definidos os segmentos do embrião e, conseqüentemente, da larva, como esquematiza a Figura 2.13.

Do ponto de vista molecular, a expressão das sete faixas de Eve é regulada por cinco *enhancers*: três regulam individualmente a expressão das faixas 1, 2 e 5 enquanto dois regulam simultaneamente duas faixas cada, um regulando as faixas 3 e 7 e outro regulando as faixas 4 e 6, conforme mostra a Figura 2.14. Entretanto, na literatura, resultados experimentais sugerem que o *enhancer* responsável pela faixa 2 possa também regular parcialmente a expressão da faixa 7(32).

^{¶¶}Embora não seja do escopo desse trabalho, é importante mencionar que os genes *gap* e *pair-rule* possuem vínculos regulatórios com os genes homeóticos dos complexos *Antennapedia* (ANT-C) e *Bithorax* (BX-C), responsáveis pela identidade segmental nas regiões da cabeça e da porção anterior do tórax e pela identidade segmental nas regiões do abdome e da porção posterior do tórax, respectivamente(12,90).

Figura 2.13 – (a) A expressão dos genes *pair-rule*, regulada pelos produtos gênicos dos genes maternos e *gap*, define os domínios que compreenderão os limites dos parasegmentos que, por sua vez, definem morfologicamente os segmentos do embrião. Imagem extraída de (93). (b) No topo, as expressões dos genes *even-skipped* (azul) e *fushi tarazu* (marrom), ambos genes *pair-rule*: caracterizadas por faixas intercaladas ao longo do embrião de *Drosophila*. Imagem extraída de (90). Abaixo, um exemplo de mutação em gene *pair-rule*, nesse caso, no gene *ftz*: em A, o embrião do tipo selvagem cuja morfologia é caracterizada por segmentos; em B, embrião com mutação: a definição dos parasegmentos é comprometida e, conseqüentemente, a definição morfológica de segmentos. Imagem extraída de (12).

Figura 2.14 – Locus do gene *eve*. Contendo cerca de 12 kb de DNA regulatório distribuídas em 5 *enhancers*: três regulam a expressão das faixas 1, 2 e 5 individualmente e dois regulam simultaneamente duas faixas cada, um regulando as faixas 3 e 7 e outro regulando as faixas 4 e 6. A imagem não exibe a região compreendida entre -5.9 e -5.2 kb responsável por autorregulação de *eve*(29). Imagem extraída de (70).

Em nosso trabalho, o interesse recai sobre a expressão da faixa 2 e, para tanto, devemos considerar, dentre todos os genes envolvidos no processo de segmentação, somente aqueles cujas proteínas codificadas atuam como reguladoras da expressão da faixa 2. De acordo com a literatura, a expressão da faixa 2 do mRNA de *eve* é inicialmente atribuída a um conjunto de sítios de ligação denominado *minimal stripe element* (MSE), compreendidos em uma região de 480 bp compreendida entre -1.6 e -1.1 kb em relação ao sítio de iniciação

da transcrição de *eve*. O MSE contém um total de 12 sítios de ligação: 5 para proteína ativadora Bcd, 1 para a ativadora Hb, 3 para a repressora Kr e 3 para a repressora Gt(25, 29, 30). Contudo, trabalhos mostraram que sem a região MSE ainda assim havia uma expressão residual da faixa 2(91). Dessa forma, novos experimentos foram realizados e a partir daí foi proposto que a expressão da faixa 2 do mRNA de *eve* seria regulada por mais sítios, interagindo um total de 34 unidades: além dos 12 sítios já relacionados, são considerados mais 2 para Hb, 3 para Kr, 6 para a ativadora Cad, 6 para a repressora Kni e 5 para a repressora Tll(32). Sendo assim, à medida em que for necessário invocar as proteínas regulatórias da faixa 2 da expressão de *eve*, levaremos em conta todos os estudos relacionados de modo que, em princípio, consideraremos como ativadoras Bicoid (Bcd), Hunchback (Hb) e Caudal (Cad) e como repressoras as proteínas Tailless (Tll), Giant (Gt), Knirps (Kni) e Krüppel (Kr).

É bastante oportuno mencionar que no *locus* do gene *eve* há também, além dos *enhancers*, uma região de 100 bp compreendida entre -5.9 e -5.2 kb em relação ao sítio de iniciação da transcrição que contém sítios de ligação para a proteína Eve, ou seja, o gene *eve* também é regulado pela própria proteína que codifica (autorregulação). Entretanto, de acordo com a literatura especializada, é desconhecida a ocorrência de autorregulação de *eve* durante estágios anteriores à celularização de modo que sua importância está relacionada à manutenção e refinamento da expressão de *eve* durante a gastrulação e a alongação das camadas germinativas. Desse modo, como a expressão de *eve* a ser estudada ocorre durante o ciclo 14 (e, portanto, o embrião ainda não se configura como blastoderme celular), entendemos que efeitos de autorregulação, ainda que ocorram, não deverão influenciar o suficiente de modo que desconsideraremos quaisquer possibilidade de autorregulação durante o intervalo de tempo analisado (23, 25, 26, 28, 29, 91, 92). Também é importante dizer que não encontramos na literatura nenhuma informação que relacione a expressão de *eve* e o processo de celularização de modo que consideraremos estes fenômenos independentes***.

É interessante mencionar que, embora no citoplasma do óvulo, antes do início dos

***Essa consideração poderá ser útil novamente, quando modelarmos a transcrição de *eve* ao longo do embrião no Capítulo 4.

ciclos de clivagens, já haja mRNAs maternos de *bcd*, *hb*, *cad* e *tll*, com o passar do tempo, a síntese das respectivas proteínas passa a ser exclusivamente zigótica de modo que, no ciclo 14A, quase todas proteínas envolvidas na regulação do gene *eve* que são encontradas durante tal regulação estão, de fato, sendo sintetizadas a partir de DNA do próprio embrião (com exceção da proteína Hb pois, ainda que sua expressão deixe gradualmente de ser realizada a partir de mRNA materno para ser realizada a partir de material zigótico, durante a celularização, há ainda uma parcela de Hb de origem materna(89)).

O padrão de expressão do gene *eve* observado, monitorado e com o qual trabalhamos ocorre durante o ciclo 14 e, para efeito de análise quantitativa, todo o processo de expressão é dividido em intervalos de tempo, chamados *classes temporais*, de aproximadamente 6,5 minutos cada(32).

2.4 Dados experimentais da expressão de *even-skipped*

Os dados experimentais a partir dos quais realizamos nossos estudos foram obtidos por Janssens et al.(32) e correspondem às intensidades de fluorescência associadas à presença de moléculas de mRNA transcrito do gene repórter *lacZ* de *Escherichia coli*. Essencialmente, o procedimento experimental para obtenção desses dados consiste na conjunção de dois procedimentos: primeiramente, a técnica do gene repórter e depois a hibridização fluorescente *in situ* (FISH).

A técnica do gene repórter consiste em fundir dois genes: o gene de interesse (no caso, a sequência de *eve*) e o gene cujo transcrito pode ser facilmente detectado (no caso, *lacZ* de *Escherichia coli*). Dessa forma, as duas sequências fundidas são reguladas pelo mesmos sítios de ligação (*enhancer* 2) e são transcritas em uma única molécula de mRNA. Com esse constructo, a detecção visual do produto gênico de *lacZ* indica a presença do mRNA de *eve*(94,95). A partir daí, a detecção da presença de material transcrito é re-

alizada através da hibridização fluorescente *in situ*. Essencialmente, esta técnica utiliza grupos moleculares fluorescentes (*fluoróforos*) como sondas para detecção visual da sequência de DNA de interesse (DNA-alvo), como esquematiza a Figura 2.15. A sonda de

Figura 2.15 – (a) Os principais componentes para a realização da hibridização fluorescente *in situ* (FISH) são a sonda (geralmente, uma sequência de DNA clonado) e o DNA-alvo (sequência de interesse). (b) Antes da hibridização, a sonda de DNA pode ser marcada por meio de várias técnicas possíveis (*nick translation*, *random prime* e PCR). Comumente são utilizadas duas estratégias para marcar o DNA: indireta (à esquerda) e direta (direita). Para a marcação indireta, as sondas são marcadas com nucleotídeos que contêm um hapteno, enquanto que marcação direta utiliza nucleotídeos modificados para se ligar a um fluoróforo. (c) A sonda marcada e o DNA-alvo são desnaturados. (d) As sequências complementares da sonda e do DNA-alvo se ligam. (e) No caso de marcação indireta, é necessário utilizar sistema de detecção enzimática ou imunológica do hapteno não-fluorescente. Se, por um lado, a marcação direta proporciona um processo de análise mais rápido, a marcação indireta tem a vantagem de permitir a amplificação de sinal utilizando várias camadas de anticorpos de modo que o sinal produzido é mais brilhante quando comparado aos sinais de fundo (*background*). Imagem extraída de (96).

DNA é geralmente um fragmento de DNA clonado (vermelho) complementar à sequência do DNA-alvo (azul). A primeira tarefa é obter uma cópia fluorescente da sonda (Figura 2.15(b), coluna do meio) ou uma cópia modificada contendo haptenos^{†††} aos quais moléculas fluorescentes(99–101) podem se ligar (Figura 2.15(b), coluna da esquerda). Com isso, a sonda e o DNA-alvo são desnaturados quimicamente ou através de aquecimento (Figura 2.15c) de modo que as ligações de hidrogênio (linhas pretas) que unem as duas fitas da hélice de DNA sejam desfeitas. Isso é necessário para que novas ligações de hidrogênio sejam formadas entre a sonda e o DNA-alvo durante o próximo passo, a hibridização,

^{†††}Hapteno é uma molécula que não é imunogênica, ou seja, não é capaz de ativar linfócitos quando administrada sozinha. Entretanto, podem provocar resposta imune somente quando ligadas covalentemente a macromoléculas (*carreadoras*) tais como proteínas(97,98).

quando ambas as sequências forem misturadas (Figura 2.15(d)) e as sondas se ligarem às sequências de DNA-alvo. Se a sonda é fluorescente (coluna do meio) é possível detectar a hibridização diretamente. Caso contrário (coluna da esquerda), é necessário introduzir fluoróforos para que se liguem às sequências hibridizadas (Figura 2.15(e)). Os arranjos híbridos formados pelas sondas e pelo DNA-alvo podem ser detectados utilizando microscopia de fluorescência(99). Por fim, a microscopia confocal é utilizada para fornecer imagens prontas para serem processadas(29,32,87,102–112).

O resultado de todos esses procedimentos é apresentado pelas Figuras 2.16 e 2.17. Elas apresentam o monitoramento da expressão de mRNA de *eve* e dos gradientes das proteínas envolvidos na regulação da faixa 2 durante o ciclo 14A, com aproximadamente 50 minutos de duração e dividido em intervalos de tempo denominados *classes temporais*, cada qual com aproximadamente 6.5 minutos(108–110). A classificação temporal baseada na dinâmica do padrão de expressão de *eve* corresponde ao grau de invaginação das membranas (uma vez que a celularização ocorre durante o ciclo 14A), que é o marcador morfológico utilizado para organizar embriões no ciclo 14A(109,113). Tal monitoramento é realizado em uma região cujo comprimento compreende o eixo AP do embrião e cuja largura equivale a 10% do eixo DV do embrião e o processamento é tal que as intensidades de fluorescência apresentadas nos gráficos correspondem às intensidades de fluorescência emitidas por núcleo(106–110).

A Figura 2.16 mostra os perfis exibidos pela intensidade de fluorescência do mRNA de *lacZ* ao longo do embrião de *Drosophila* no decorrer das classes temporais. Com relação à evolução do perfil de mRNA transcrito, podemos notar que a configuração espacial inicial, marcada por um pico de intensidade com considerável largura, começa a adquirir o caráter de faixa, marcado pelo pico da intensidade que é não apenas maior do que antes mas também grande comparado à sua própria largura e aos valores de intensidade observados no restante do embrião. Podemos observar que, de C13 a T2, a expressão, caracterizada por uma grande mancha, exibe duas tendências perceptíveis (e já mencionadas): a diminuição da largura da mancha e o aumento de intensidade em torno de AP = 41.5, consistindo na própria caracterização da faixa. Posteriormente,

Figura 2.16 – Dinâmica do perfil de mRNA de *lacZ* (preto) em diferentes classes temporais: (a)C13, (b)T1, (c)T2, (d)T3, (e)T4, (f)T5, (g)T6, (h)T7 e (i)T8. Para cada instante são exibidos o perfil da expressão de *lacZ* no embrião e o gráfico exibindo a intensidade de fluorescência do respectivo perfil. A correspondente expressão endógena da proteína Eve aparece ao fundo (cinza). No eixo vertical, está representada a escala de intensidade de fluorescência e, no eixo horizontal, está representado o eixo AP. As intensidades exibidas nos gráficos correspondem a valores detectados dentro de uma faixa cujo comprimento compreende todo o eixo AP (se estendendo do extremo anterior até o extremo posterior do embrião) e cuja espessura compreende uma região intermediária do eixo dorsal-ventral (DV), equivalente a 10% do tamanho do eixo DV (mais precisamente, entre 45% e 55%). Grosso modo, o que é feito no processamento de dados é dividir o eixo AP em 100 partes iguais tal que cada ponto do gráfico corresponde à intensidade média de fluorescência emitida por um núcleo dessa região de comprimento equivalente a 1% do eixo AP e de largura equivalente a 10% do eixo DV. Os embriões estão orientados com a parte anterior para a esquerda e com a parte dorsal para cima. A posição no eixo AP está em unidades de porcentagem e aumenta do extremo anterior para o extremo posterior. A expressão observada em torno de 25% a partir de T5 é devido às sequências presentes no vetor de transformação que contém o transposon P e os genes *eve* e *lacZ*. Imagens extraídas de (32).

no intervalo de T3 a T5, a expressão já apresenta um perfil muito mais próximo do formato característico de uma faixa. Durante este intervalo, o processo de estreitamento do pico de expressão continua, porém, a intensidade no pico de expressão não é alterada de forma considerável. Após o perfil da expressão ter assumido a forma de faixa em T5,

Figura 2.17 – Intensidades de fluorescência correspondentes dos gradientes das proteínas regulatórias Bicoid (marrom), Hunchback (amarelo), Giant (azul), Krüppel (verde), Caudal (ciano), Knirps (vermelho) e Tailless (roxo), proteínas envolvidas na regulação da expressão do mRNA de *eve*, comparadas com a intensidade de fluorescência correspondente à expressão de mRNA de *lacZ* (preto) — gene repórter utilizado para monitorar a transcrição de *eve* — em cada classe temporal (C13, T1, T2, T3, T4, T5, T6, T7 e T8). No eixo vertical, está representada a escala de intensidade de fluorescência e, no eixo horizontal, está representado o eixo AP. Assim como nos dados relativos ao mRNA, cada ponto do gráfico corresponde à intensidade média de fluorescência emitida por um núcleo de uma região de comprimento equivalente a 1% do eixo AP e de largura equivalente a 10% do eixo DV. A parte anterior está orientada para a esquerda e a posição no eixo AP está em unidades de porcentagem, aumentando do extremo anterior para o extremo posterior. Dados obtidos de FlyEx Database(33).

a partir T6 o caráter de faixa tende a ser mantido mas a intensidade do pico decresce gradualmente até T8. Ao longo do processo de aparecimento e atenuação da faixa de mRNA em torno de AP = 41.5, ocorre, a partir de T5, o aparecimento de um pequeno pico de expressão em torno de AP = 80.5, cuja intensidade aparentemente aumenta até T7 e depois começa a decrescer. Vale mencionar que a expressão observada em torno de AP = 25 é devido à sequência de transposon *P* presente no vetor de transformação que contém *eve-lacZ*^{†††}. De qualquer modo, não devemos nos preocupar com a expressão nessa região

^{†††} *Transposon P* ou *elemento P* é um elemento de transposição de DNA, isto é, um segmento de DNA capaz de se mover de uma posição no genoma para outra. Esse segmento contém o gene da proteína transposase, enzima que realiza a transposição. O elemento *P* é um transposon característico de *Drosophila* e pode ser utilizado para o estudo de regulação gênica. Para isso, é necessário realizar a fusão das sequências do transposon, do gene de interesse (no caso, *eve*) e do gene repórter

pois analisaremos apenas a expressão na região que compreende de $AP = 35.5\%$ a $AP = 91.5\%$. Lembrando que os dados são coletados ao longo de todo o eixo AP compreendidos em uma região equivalente a 10% do tamanho do eixo DV (mais precisamente, entre 45% e 55%), a análise da expressão entre $AP = 35.5\%$ a $AP = 91.5\%$ é justificada pelo fato de que nessa região os padrões estabelecidos ao longo do eixo AP independem dos padrões estabelecidos ao longo do eixo DV. Morfologicamente, a região entre $AP = 35.5\%$ a $AP = 91.5\%$ corresponde, grosso modo, à região que originará os segmentos torácicos e abdominais(124). Cabe aqui dizer que a formação de padrões de expressão é independente das divisões mitóticas uma vez que estamos trabalhando com dados experimentais obtidos em C13 e que entre C13 e T1 ocorre um ciclo de clivagens(125).

A Figura 2.17 apresenta as evoluções temporais das intensidades de fluorescência associadas à expressão do mRNA de *eve* e aos gradientes de concentração das proteínas regulatórias envolvidas na regulação da faixa 2. Além das proteínas Bicoid (Bcd), Hunchback (Hb), Giant (Gt) e Krüppel (Kr), cujas contribuições regulatórias são experimentalmente conhecidas (29,30), há também a presença de Caudal (Cad), Knirps (Kni) e Tailless (Tll), cujas contribuições na regulação são propostas nos trabalhos de Berman e Janssens.(32,126). No primeiro grupo mencionado, as proteínas Bcd e Hb são conhecidas por serem ativadoras da transcrição de *eve* enquanto que Gt e Kr são repressoras e essas quatro proteínas regulatórias têm disponíveis, ao todo, doze sítios de ligação, cada um específico para uma proteína. No segundo grupo, a proteína Cad é apresentada como uma terceira ativadora e as proteínas Kni e Tll como mais duas repressoras e o número total de sítios de ligação envolvidos aumenta para trinta e quatro. Em virtude de nossos estudos não serem de natureza experimental mas de natureza teórica e, sobretudo, com caráter fenomenológico, iremos considerar, na medida do possível, os resultados advindos dos dois trabalhos. Com respeito à natureza regulatória das proteínas citadas, podemos perceber, através da Figura 2.17, alguns pontos: **(i)** a expressão de mRNA de *eve* é fortemente reprimida entre $AP \approx 45\%$ e $AP \approx 77\%$ e isso se deve às regulações de Kr, Kni

(*lacZ*). Esses constructos, chamados de *vetores de transformação*, são injetados no embrião, inserem-se aleatoriamente no genoma, sendo, a partir daí, expressos durante o desenvolvimento(70,90,114–121). Contudo, o padrão de expressão pode ser influenciado por sequências próximas do sítio de inserção, causando dificuldades na interpretação dos padrões gerados(86,122,123).

e Gt, que formam uma espécie de “bloco” repressor; (ii) a expressão da faixa 2 ocorre em uma região onde há um “vale” de repressão e onde a expressão de Hb atinge seus valores máximos; (iii) a intensidade da faixa 2 diminui a partir de T5, assim que diminui a intensidade da expressão de Hb; (iv) a formação de um segundo pico de expressão em T5, em torno de AP $\approx 80\%$, pode ser entendida como resultante do aumento da expressão de Hb nessa região, iniciado em T3.

Como nossa abordagem do processo de transcrição necessita do número de moléculas de mRNA transcritas por núcleo, nosso primeiro passo é, então, adquirir os valores correspondentes aos dados mostrados pela Figura 2.16. Para isso, as imagens foram digitalizadas através do programa *DigitizeIt* 1.5 e assim tivemos acesso aos valores publicados. Tendo em mãos os valores das intensidades de fluorescência e sabendo que eles estão diretamente relacionados ao número de partículas, é necessário converter as unidades de intensidade de fluorescência em número de moléculas de mRNA por núcleo, com o qual iremos trabalhar.

Muito embora haja uma vasta literatura tratando de expressão gênica em *Drosophila*, de técnicas e procedimentos para aquisição e processamento de imagens bem como resultados experimentais, no entanto, até o presente momento, não encontramos informações precisas do número exato de moléculas de mRNA de *eve* transcritas na região onde ocorre seu pico de expressão (e nem ao longo do embrião) e nem fatores para conversão de unidades de intensidade de fluorescência para concentração de mRNA de *eve*.

Por isso, lançamos mão de informações sobre os genes *eve*, *ftz* e *Ubx* (abreviação para *Ultrabithorax*, gene pertencente ao já mencionado complexo homeótico BX-C e regulado diretamente por vários genes, entre eles, *eve* e *ftz*(127–129)) — cujas expressões ocorrem durante a fase de blastoderme do embrião(90, 130–132) e estão intimamente envolvidos na definição e desenvolvimento de segmentos bem como em vários outros processos(133) — com o intuito de que nos auxiliem na obtenção de uma estimativa para o número de moléculas de mRNA de *eve* transcritas.

1. A proteína Ftz, codificada pelo gene *ftz* (de cuja regulação *eve* também participa), apresenta um padrão de expressão caracterizado por faixas que intercalam com

as faixas de Eve e as intensidades de fluorescência de ambas as expressões são comparáveis entre si(33,110).

2. Os mRNAs codificados pelos genes *ftz* e *Ubx* apresentam abundâncias comparáveis(134).
3. A concentração de moléculas de mRNA de *Ubx* presentes no embrião de *Drosophila* durante o ciclo 11 atinge 50 moléculas/ $250\mu\text{m}^3$ (135).

Considerando estas três informações e supondo que (i) a informação 2 implique, de alguma forma, em concentrações de mRNA de *eve* e de *ftz* comparáveis (de mesma ordem de magnitude) e que (ii) a relação entre intensidade de fluorescência medida (já corrigida de um sinal de fundo) e o número de moléculas de mRNA seja linear, temos que a quantidade de 50 moléculas/ $250\mu\text{m}^3$ pode ser adotada como valor de referência para o número de moléculas de mRNA de *eve* e utilizaremos tal valor para caracterizar o ponto onde a máxima intensidade de fluorescência é registrada (AP = 40.5 em T5, segundo os dados digitalizados). Esperamos que, a exemplo do que ocorre na evolução dos gradientes de diversas proteínas envolvidas no processo de segmentação(33), não haja uma diferença de várias ordens de magnitude entre os valores de mRNA de *Ubx* medidos no ciclo 11 e os valores reais no ciclo 14 de modo que o referido valor possa ainda assim ser considerado como referência. E, considerando novamente a hipótese de uma relação linear entre intensidade de fluorescência e número de moléculas de mRNA, podemos estimar o número de moléculas de mRNA para o restante do embrião.

Como o diâmetro dos núcleos durante o ciclo 14 é de $6,5\mu\text{m}$ (136), o que corresponde a um volume nuclear de $143\mu\text{m}^3$, temos que a concentração de moléculas de mRNA de *eve* corresponde a cerca de 28.75 moléculas por núcleo, o que está dentro da escala de valores típicos para eucariotos(134,137), e, como dissemos anteriormente, esse valor será adotado para estimar o número de moléculas de mRNA de *eve* no ponto de maior intensidade de fluorescência por núcleo registrada (que ocorre em AP = 40.5%). Daí, utilizando a suposição de linearidade entre intensidade de fluorescência e número de moléculas, podemos converter os perfis de intensidade exibidos pela Figura 2.16 em perfis que exibem o número de moléculas de mRNA em cada núcleo ao longo do eixo AP, como mostra a

Figura 2.18.

Figura 2.18 – Evolução temporal do número estimado de moléculas de mRNA ao longo do eixo AP.

De fato, dadas as circunstâncias, a estimativa obtida nos fornece um razoável ponto de partida para nosso trabalho. Entretanto, é imprescindível mencionar que nossa estimativa ignora considerações substanciais. O problema está, especificamente, na suposição de que as intensidades de fluorescência das proteínas Eve e Ftz, uma vez comparáveis, refletiriam concentrações de mRNA de *eve* e de *ftz* também comparáveis, o que tem grandes chances de ser uma consideração equívoca. O fato é que essa suposição ignora completamente aspectos importantes na síntese de ambas as proteínas como, por exemplo, as taxas de produção e de degradação de cada mRNA, a informação de quantas vezes uma mesma fita de mRNA de *eve* ou de *ftz* pode ser traduzida e também aspectos experimentais como quantos moléculas fluorescentes estão associadas a cada molécula de cada proteína e a conversão de unidades de intensidade de fluorescência para número de moléculas presentes. O procedimento que culmina na obtenção de uma estimativa para o número de moléculas de mRNA representa, na verdade, um grande apelo, decorrente da ausência

dos informações experimentais. De qualquer forma, esse valor adotado merece ser revisto e, se possível, uma estimativa baseada em resultados experimentais deve ser realizada.

Por outro lado, é importante mencionar que, caso essa estimativa se revele discrepante frente a valores reais da concentração de mRNA de *eve* (mesmo com a hipótese da linearidade se mostrando razoável), veremos, nos capítulos 3 e 4, que, ainda assim, o modelo utilizado pode ser reescalado apenas alterando um único parâmetro e as condições iniciais, uma vez que todos os termos presentes no primeiro e no segundo momentos (que compõem o valor médio, a flutuação e o ruído do processo) ou são multiplicados pelo referido parâmetro ou são as próprias condições iniciais (e, portanto, podem ser automaticamente reescaladas). Além disso, o referido parâmetro não define nenhuma escala de tempo característica do processo e, sendo assim, qualquer alteração nesse referido parâmetro não comprometerá a dinâmica do sistema.

Capítulo 3

A teoria de spins e bósons para expressão estocástica de um gene

Em uma célula ocorrem inúmeras transformações químicas simultaneamente. Muitas dessas transformações estão associadas à expressão gênica de modo que o conjunto de interações entre genes de uma célula se revela como portador de uma imensa complexidade. Além da diversidade de fenômenos e espécies biomoleculares, a complexidade da expressão gênica também se deve pequeno número de moléculas de cada espécie envolvidas no fenômeno, o que torna as flutuações moleculares consideráveis de maneira que uma abordagem estocástica é a melhor estratégia para o estudo de fenômenos de regulação gênica.

Uma das abordagens estocásticas existentes é o modelo proposto por Sasai e Wolynes. Esse modelo mantém estreita relação com um sistema quântico de muitos corpos ao propor que a ativação e repressão do gene (chaveamento) devido às proteínas regulatórias podem ser descritas como spins imersos em uma atmosfera bosônica tendo suas polarizações constante alteradas pelas interações entre o spin e os bósons. A dinâmica do chaveamento gênico bem como do número de mRNA transcritos (ou de proteínas sintetizadas, no caso de acoplar a transcrição e tradução) é descrita em termos de de ligação de spin enquanto que as flutuações dos produtos gênicos são determinadas utilizando operadores de criação e destruição, analogamente ao caso de osciladores harmônicos quânticos(62). Em outras

palavras, o problema da expressão gênica é descrito analogamente ao problema de spin-bóson que caracteriza a teoria de pólarons e transferência de elétrons. Em princípio, esse formalismo permite descrever uma rede de vários spins interagindo com uma atmosfera de bósons, o que, por sua vez, permite a descrição e o entendimento de uma rede genética.

Posteriormente, esse modelo foi resolvido exatamente por Hornos e colaboradores(63) e investigado em detalhes por Innocentini(64,65) e Ramos(66–68). Dedicamos a primeira parte desse capítulo a apresentá-lo sua apresentação, na qual detalhamos as equações principais e os procedimentos para obtenção das principais quantidades utilizadas para a descrição da expressão gênica. Depois, calcularemos a evolução temporal das probabilidades (do gene estar ativado ou reprimido), do número médio de moléculas de mRNA transcritas e da variância associada ao processo transcricional considerando condições iniciais arbitrárias, enquanto que, nas referências (64) e (65), as respectivas quantidades são obtidas considerando o estado inicial estacionário (exceto as probabilidades, cuja evolução temporal é calculada para condições iniciais quaisquer). E, por fim, calcularemos as distribuições de probabilidade no caso estacionário e no caso dependente do tempo, considerando o sistema partindo do regime estacionário, conforme feito em (65).

3.1 O modelo

Originalmente, nos trabalhos de Wolynes(62), Hornos(63) e Innocentini(64), é considerado um gene com dois estados — um com alta atividade transcricional (ligado) e outro com baixa atividade transcricional (desligado) — que é regulado por proteínas de um mesmo tipo que podem se ligar a apenas um sítio de ligação. Supondo tal proteína ativadora, a ligação desta ao sítio de ligação estimula a transcrição e o gene se encontra ligado; se tal proteína não se liga ao sítio de ligação, a transcrição não ocorre ou ocorre a taxas basais. E, como vimos, se a proteína regulatória for repressora, ocorre o oposto: a sua ligação ao sítio de ligação reprime a transcrição e seu desligamento do sítio de ligação

estimula a transcrição. O estado do sistema é caracterizado pelo nível de ativação (alta ou baixa) e também pelo número de moléculas de mRNA transcritas.

É realizada uma abordagem estatística microscópica, descrevendo o processo transcricional em termos das probabilidades de se encontrar o gene no estado ligado e desligado quando há n moléculas de mRNA transcritas. Essas probabilidades são representadas por $\alpha_n(t)$ e $\beta_n(t)$, respectivamente. Dessa forma, é intuitivo concluir que, se o estado do sistema é caracterizado pelo nível de atividade transcricional e pelo número de moléculas transcritas, então as equações satisfeitas pelas probabilidades $\alpha_n(t)$ e $\beta_n(t)$ devem conter termos relativos a processos que alteram o número de moléculas e termos relativos a processos que alteram o nível transcricional. De fato, as equações satisfeitas por $\alpha_n(t)$ e $\beta_n(t)$, dadas por

$$\frac{d\alpha_n(t)}{dt} = k[\alpha_{n-1}(t) - \alpha_n(t)] + \rho[(n+1)\alpha_{n+1}(t) - n\alpha_n(t)] - h\alpha_n(t) + f\beta_n(t) \quad (3.1)$$

$$\frac{d\beta_n(t)}{dt} = k\chi[\beta_{n-1}(t) - \beta_n(t)] + \rho[(n+1)\beta_{n+1}(t) - n\beta_n(t)] + h\alpha_n(t) - f\beta_n(t), \quad (3.2)$$

afirmam que as taxas de variação temporal de $\alpha_n(t)$ e $\beta_n(t)$ resultam de um processo de nascimento e morte de moléculas de mRNA e da possibilidade de transição entre níveis de ativação. Os termos de nascimento e morte envolvem a taxa de produção (taxa à qual uma fita de mRNA é sintetizada por uma RNA-polimerase), representada pelo parâmetro k , e a taxa de degradação de moléculas de mRNA, representada pelo parâmetro ρ . As possibilidades de transição entre níveis de ativação são representadas pelos parâmetros h e f , taxas às quais o gene realiza as transições ligado→desligado e desligado→ligado respectivamente, representando a eficiência da ligação da proteína regulatória ao sítio de ligação. Se as taxas h e f forem nulas, as equações se desacoplam e as soluções $\alpha_n(t)$ e $\beta_n(t)$ serão distribuições de Poisson com valores médios k/ρ e $k\chi/\rho$, respectivamente. O parâmetro χ , usualmente considerado entre 0 e 1, quantifica a magnitude da expressão do gene em seu estado desligado de modo que, para valores pequenos de χ , o estado desligado corresponde a uma taxa mínima de produção de mRNA e, para $\chi = 0$, a transcrição é totalmente reprimida*.

*A exemplo do que é apresentado em (62), podemos mostrar que a analogia a um sistema quântico torna-

Como já adiantamos no Capítulo 2, por buscarmos uma descrição fenomenológica da transcrição, nossa abordagem desconsidera várias interações entre as espécies moleculares que podem influenciar no produto gênico final. Além disso, ela faz uso de algumas simplificações matemáticas ao supor que os parâmetros envolvidos nos processos — k , ρ , χ , h e f — independem da variável estocástica n ou do tempo t . Supor que os parâmetros k , χ e ρ independem de n e t implica dizer que os processos de produção e degradação de mRNA resultam da conjunção de características intrínsecas da maquinaria transcricional, da sequência gênica, da molécula de mRNA e das ribonucleases, de modo que não seria esperada uma dependência específica de n ou t . A suposição de que h e f independem de t implica dizer que as características regulatórias (da proteína ou do sítio de ligação) não são alteradas durante o processo. Contudo, podemos esperar de antemão que tal independência nem sempre exista durante o processo transcricional, já que a presença e abundância da própria proteína regulatória é resultado de um processo de regulação e que, portanto, nem sempre apresentará um nível constante de abundância. Por sua vez, a suposição de que h e f independem de n apenas reforça que a proteína regulatória não é produto gênico do processo em questão, isto é, não há autorregulação. Essa suposição é sustentada pelo fato de que, em toda a literatura à qual tivemos acesso, não há indícios de que a proteína Eve regule a própria transcrição.

Certamente, dada a diversidade e complexidade das interações gênicas, é possível

se evidente quando, nas Eqs.(3.1) e (3.2), os termos relativos aos processos de produção e degradação do produto gênico se apresentam em termos de operadores de criação e destruição. Para um processo envolvendo somente um tipo de produto gênico decorrente de um único gene, um vetor de estado ψ pode ser escrito como $\psi = \sum_n P(n, t) |n\rangle$, onde $P(n, t)$ corresponde à probabilidade de haver n partículas e $|n\rangle$ corresponde ao estado caracterizado por n partículas. Assim, as Eqs.(3.1) e (3.2) podem ser reescritas como $\partial\psi/\partial t = \Omega\psi$, onde Ω , que atua como um hamiltoniano é dado por

$$\Omega = (\bar{k}I + \delta k\sigma_z)(a^\dagger - 1) + \rho(a - a^\dagger a) + \mu^+(-1 + \sigma_x) + \mu^-(-i\sigma_y + \sigma_z),$$

onde

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad i\sigma_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

e $\bar{k} = (k + k\chi)/2$, $\delta k = (k - k\chi)/2$, $\mu^+ = (h + f)/2$ e $\mu^- = (h - f)/2$. Os operadores a e $a^\dagger$ atuam sobre $|n\rangle$ da seguinte forma:

$$a|n\rangle = n|n-1\rangle, \quad a^\dagger|n\rangle = |n+1\rangle$$

e satisfazem a condição

$$[a, a^\dagger] = 1.$$

encontrar casos que, em termos do modelo apresentado aqui, devem apresentar $h = h(n, t)$ e $f = f(n, t)$. Um exemplo é o caso estudado e resolvido por Hornos e colaboradores(63), no qual um gene é autorregulado negativamente, ou seja, a proteína expressa reprime a transcrição do próprio gene que a codifica. Nesse caso, as equações utilizadas são idênticas às Eqs.(3.1) e (3.2) com exceção do termo referente à transição do nível ligado para desligado, o qual é alterado de h para hn .

Para resolver as equações (3.1) e (3.2), que são equações do tipo diferencial-diferença, devemos introduzir funções complexas — chamadas funções geratrizes — de modo que as equações em questão possam ser reescritas na forma de equações diferenciais e que quantidades importantes para análise estatística podem ser obtidas a partir destas funções. Tais funções geratrizes são séries cujos coeficientes são dados por $\alpha_n(t)$ e $\beta_n(t)$ (139). Dessa forma, temos

$$\alpha(z, t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n, \quad \beta(z, t) = \sum_{n=0}^{\infty} \beta_n(t) z^n, \quad \phi(z, t) = \sum_{n=0}^{\infty} \phi_n(t) z^n, \quad (3.3)$$

onde $\phi_n(t)$, dada por

$$\phi_n(t) = \alpha_n(t) + \beta_n(t),$$

é a probabilidade de que haja n moléculas de mRNA transcritas no instante t independentemente do estado de ativação transcricional do gene. Daí, é imediato ver que $\phi(z, t) = \alpha(z, t) + \beta(z, t)$.

O uso das funções geratrizes nos permite então escrever as equações (3.1) e (3.2) na forma

$$\begin{aligned} \frac{\partial \alpha(z, t)}{\partial t} &= (z - 1) \left[k\alpha(z, t) - \rho \frac{\partial \alpha(z, t)}{\partial z} \right] - h\alpha(z, t) + f\beta(z, t) \\ \frac{\partial \beta(z, t)}{\partial t} &= (z - 1) \left[k\chi\beta(z, t) - \rho \frac{\partial \beta(z, t)}{\partial z} \right] + h\alpha(z, t) - f\beta(z, t). \end{aligned} \quad (3.4)$$

Resolvendo essas equações, obtemos as funções geratrizes $\alpha(z, t)$ e $\beta(z, t)$ e, com elas em

mãos, podemos obter as distribuições de probabilidade $\alpha_n(t)$ e $\beta_n(t)$, dadas por

$$\begin{aligned}\alpha_n(t) &= \left[\frac{1}{n!} \frac{\partial^n \alpha(z, t)}{\partial z^n} \right]_{z=0} \\ \beta_n(t) &= \left[\frac{1}{n!} \frac{\partial^n \beta(z, t)}{\partial z^n} \right]_{z=0},\end{aligned}\tag{3.5}$$

e as probabilidades de que o gene esteja ativo e inativo independentemente do número de mRNAs transcritos, representadas por $P_\alpha(t)$ e $P_\beta(t)$, respectivamente, e dadas por

$$P_\alpha(t) = \alpha(1, t) = \sum_{n=0}^{\infty} \alpha_n(t), \quad P_\beta(t) = \beta(1, t) = \sum_{n=0}^{\infty} \beta_n(t).\tag{3.6}$$

Os momentos das distribuições de probabilidade também podem ser obtidos através de derivadas das funções geratrizes, agora avaliadas em $z = 1$, conforme as seguintes fórmulas:

$$\begin{aligned}\langle n_\alpha^j \rangle(t) &= \left[\left(z \frac{\partial}{\partial z} \right)^j \alpha(z, t) \right]_{z=1} = \sum_{n=0}^{\infty} n^j \alpha_n(t) \\ \langle n_\beta^j \rangle(t) &= \left[\left(z \frac{\partial}{\partial z} \right)^j \beta(z, t) \right]_{z=1} = \sum_{n=0}^{\infty} n^j \beta_n(t).\end{aligned}\tag{3.7}$$

Tendo realizada a formulação do modelo, são necessária uma observação. No Capítulo 2, vimos que a literatura a respeito da expressão do gene *eve* de *Drosophila* nos diz que a formação da listra 2 resulta da ação de diversas proteínas regulatórias e diversos sítios de ligação. Comparada ao cenário mostrado pela literatura especializada, a formulação proposta aqui pelo modelo estocástico pode, em princípio, parecer pouco adequada. De fato, esperamos que um gene regulado por proteínas de vários tipos podendo se ligar a vários sítios acabe apresentando vários níveis de ativação, evidenciando a complexidade do fenômeno biológico. Por outro lado, ao considerarmos um gene com dois níveis e regulado por proteínas de um único tipo e um único sítio de ligação, estamos propondo uma abordagem do tipo “tudo ou nada”, preocupada não com a contribuição de cada configuração mas sim com o efeito regulatório total sobre a transcrição. É esta perspectiva que norteia nossa abordagem. Isso será recapitulado nos Capítulos 4 e 5.

3.2 Probabilidade, valor médio e variância

Para obter a probabilidade de que o gene esteja no estado ativo, independentemente do número de moléculas de mRNA transcritas, basta tomar $z = 1$ na Eq.(3.4) e temos equações diferenciais para as probabilidades $P_\alpha(t)$ e $P_\beta(t)$:

$$\begin{aligned}\frac{dP_\alpha(t)}{dt} &= -hP_\alpha(t) + fP_\beta(t) \\ \frac{dP_\beta(t)}{dt} &= hP_\alpha(t) - fP_\beta(t).\end{aligned}\tag{3.8}$$

É imediato ver que a quantidade $P_\alpha(t) + P_\beta(t)$ é constante. E, pelo fato de $P_\alpha(t)$ e $P_\beta(t)$ serem probabilidades, essa constante deve ser igual à unidade. O sistema (3.8) pode ser escrito na forma matricial

$$\frac{dP}{dt} = \mathbb{H}P,\tag{3.9}$$

onde

$$P = \begin{pmatrix} P_\alpha(t) \\ P_\beta(t) \end{pmatrix}, \quad \mathbb{H} = \begin{pmatrix} -h & f \\ h & -f \end{pmatrix}.$$

Para resolver a Eq. (3.9), teremos de tornar diagonal a matriz $\mathbb{H}$ a fim de que as equações diferenciais para $P_\alpha(t)$ e $P_\beta(t)$ sejam desacopladas e possam ser resolvidas independentemente. A diagonalização de $\mathbb{H}$ consiste em obter uma matriz $\mathbb{S}$ e sua inversa $\mathbb{S}^{-1}$ tais que a matriz $\mathbb{D}$, dada pela transformação de semelhança

$$\mathbb{D} = \mathbb{S}^{-1}\mathbb{H}\mathbb{S},$$

seja diagonal e cujos elementos de sua diagonal principal sejam os autovalores de $\mathbb{H}$. Para isso, a equação de autovalor

$$\mathbb{H}v = \lambda v,$$

deve ser resolvida. Sendo uma matriz 2×2 , $\mathbb{H}$ possui dois autovalores, λ_1 e λ_2 , e dois autovetores, v_1 e v_2 , associados respectivamente a λ_1 e λ_2 . É fácil mostrar que os autovalores

de $\mathbb{H}$ são $\lambda_1 = 0$ e $\lambda_2 = -(h + f)$ e seus respectivos autovetores são

$$v_1 = \begin{pmatrix} f \\ h \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Os autovetores de $\mathbb{H}$ compõem as colunas da matriz $\mathbb{S}$

$$\mathbb{S} = \begin{pmatrix} f & 1 \\ h & -1 \end{pmatrix}, \quad \mathbb{S}^{-1} = \frac{1}{h+f} \begin{pmatrix} 1 & 1 \\ h & -f \end{pmatrix}.$$

Fazendo a substituição $P = \mathbb{S}Y$, onde

$$Y = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix},$$

temos

$$\frac{d(\mathbb{S}Y)}{dt} = \mathbb{H}(\mathbb{S}Y).$$

Lembrando que os elementos de $\mathbb{S}$ são constantes e multiplicando à esquerda por $\mathbb{S}^{-1}$, temos

$$\frac{dY}{dt} = \mathbb{D}Y, \quad \mathbb{D} = \begin{pmatrix} 0 & 0 \\ 0 & -(h+f) \end{pmatrix}.$$

Assim, conseguimos desacoplar o sistema (3.8) de modo que resta apenas resolver as seguintes equações diferenciais

$$\begin{aligned} \frac{d}{dt}y_1(t) &= 0 \\ \frac{d}{dt}y_2(t) &= -(h+f)y_2(t), \end{aligned} \tag{3.10}$$

cujas soluções, considerando o instante inicial em $t = t_0$, são

$$\begin{aligned} y_1(t) &= c_1 \\ y_2(t) &= y_2(t_0)e^{-(h+f)(t-t_0)}, \end{aligned} \tag{3.11}$$

onde c_1 é uma constante.

Utilizando a substituição $P = \mathbb{S}Y$, temos

$$P_\alpha(t) = f y_1(t) + y_2(t) \quad (3.12)$$

$$P_\beta(t) = h y_1(t) - y_2(t).$$

Como $P_\alpha(t) + P_\beta(t)$ deve ser igual à unidade, então, $c_1 = (h + f)^{-1}$. Sendo $P_\alpha(t_0)$ e $P_\beta(t_0)$ as condições iniciais de $P_\alpha(t)$ e $P_\beta(t)$ em $t = t_0$ respectivamente, temos, então

$$P_\alpha(t) = p_\alpha + (p_\alpha^0 - p_\alpha) e^{-(h+f)(t-t_0)} \quad (3.13)$$

$$P_\beta(t) = p_\beta + (p_\beta^0 - p_\beta) e^{-(h+f)(t-t_0)},$$

onde $p_\alpha^0 = P_\alpha(t_0)$ e $p_\beta^0 = P_\beta(t_0)$ são as condições iniciais e p_α e p_β são as probabilidades atingidas no limite assintótico ($t \rightarrow \infty$), respectivamente, por $P_\alpha(t)$ e $P_\beta(t)$, dadas por:

$$p_\alpha = \frac{f}{h+f}, \quad p_\beta = \frac{h}{h+f}. \quad (3.14)$$

Podemos perceber, nas Eqs.(3.13), que as taxas h e f , uma vez definidas, estabelecem automaticamente o estado assintótico para o qual o sistema deve evoluir e que a dinâmica exibida pelas probabilidades $P_\alpha(t)$ e $P_\beta(t)$ é caracterizada por duas escalas de tempo, T e T_{SD} , dadas por

$$T = \frac{1}{\rho}, \quad T_{SD} = \frac{1}{h+f}. \quad (3.15)$$

Como podemos ver, T define uma escala para o tempo de vida de uma fita de mRNA até sofrer degradação e T_{SD} define a escala de tempo para a qual o efeito regulatório, inicialmente ativador, se torna repressor e depois volta a ser ativador, denominada tempo de decaimento do chaveamento ou *switch*. T e T_{SD} configuram propriedades intrínsecas do mRNA e do efeito regulatório externo respectivamente e, por isso, é conveniente trabalhar com o parâmetro adimensional ϵ ,

$$\epsilon = \frac{T}{T_{SD}} = \frac{h+f}{\rho}, \quad (3.16)$$

que contém informações sobre a rapidez das transições entre níveis de ativação relativa à degradação do mRNA transcrito. Veremos adiante que, em regime estacionário, enquanto o número médio de molécula de mRNA independe de ϵ , a variância é consideravelmente

sensível a ϵ .

O valor médio, $\langle n \rangle(t)$, uma das quantidades mais importantes a serem analisadas nesse problema, pode ser obtido aplicando a Eq. (3.7) à soma das Eqs. (3.4), lembrando que $\phi(z, t) = \alpha(z, t) + \beta(z, t)$. Fazendo isso, obtemos a equação diferencial para $\langle n \rangle(t)$

$$\frac{d\langle n \rangle}{dt} = -\rho \langle n \rangle + k(1 - \chi)P_\alpha(t) + k\chi, \quad (3.17)$$

cujas solução é

$$\langle n \rangle(t) = n_0 e^{-\rho(t-t_0)} + k(1 - \chi) e^{-\rho t} \int_{t_0}^t e^{\rho s} P_\alpha(s) ds + k\chi e^{-\rho t} \int_{t_0}^t e^{\rho s} ds, \quad (3.18)$$

onde $n_0 = \langle n \rangle(t_0)$ é a condição inicial.

Utilizando as Eqs.(3.13) e (3.14) e considerando $N = k/\rho$, temos

$$\langle n \rangle(t) = n_S + \Delta n e^{-\rho(t-t_0)} + C \left(e^{-\epsilon \rho(t-t_0)} - e^{-\rho(t-t_0)} \right), \quad (3.19)$$

onde

$$n_S = N(1 - \chi)p_\alpha + N\chi \quad (3.20)$$

$$\Delta n = n_0 - n_S$$

$$\Delta p = p_\alpha^0 - p_\alpha$$

$$C = \frac{N(1 - \chi)\Delta p}{1 - \epsilon}.$$

Para tempos suficientemente grandes, o valor assintótico assumido por $\langle n \rangle(t)$ é n_S , onde o índice S se refere ao caráter estacionário da quantidade envolvida. Além disso, é importante mencionar, conforme mostram as Eqs.(3.14), (3.19) e (3.20), que a especificação das taxas h e f determina naturalmente o estado assintótico.

Lembrando da definição do parâmetro ϵ , dada pela Eq.(3.16), podemos notar que as dinâmicas de $P_\alpha(t)$ e de $\langle n \rangle(t)$ — dadas, respectivamente, pelas Eqs.(3.13) e (3.19) — são caracterizadas por escalas diferentes: a primeira dinâmica é marcada por termos proporcionais a $\exp(-\epsilon \rho t)$ enquanto que a segunda, parcialmente influenciada pela primeira, é marcada por termos proporcionais a $\exp(-\rho t)$ e $\exp(-\epsilon \rho t)$. A Eq. (3.19) mostra que, para chaveamentos lentos ($0 < \epsilon < 1$), a dinâmica de $\langle n \rangle(t)$ é governada pela dinâmica do

chaveamento, caracterizada pelo termo proporcional a $\exp(-\rho\epsilon t)$. Contudo, para chaveamentos mais rápidos ($\epsilon > 1$), a dinâmica de transições entre estados ligado e desligado atinge a configuração de equilíbrio antes da transcrição atingir o regime estacionário. Isso significa que, para ϵ grande, $\langle n \rangle(t)$ tem “memória curta” em relação à dinâmica do chaveamento e, portanto, a dinâmica do número médio de mRNA acaba sendo guiada unicamente pelo processo de degradação.

Além do número médio $\langle n \rangle(t)$ de partículas transcritas no instante de tempo t , é possível calcular também qual a parcela desse número que está associada ao estado ligado, $\langle n_\alpha \rangle(t)$, e qual está associada ao estado desligado, $\langle n_\beta \rangle(t)$, de modo que a soma dessas parcelas corresponda ao número médio total $\langle n \rangle(t)$, como esperado. Para obter $\langle n_\beta \rangle(t)$ e $\langle n_\beta \rangle(t)$, basta aplicar a Eq. (3.7) na Eq. (3.4), ou seja, atuar com o de ligação $z \frac{\partial}{\partial z}$ na Eq. (3.4) e, em seguida, avaliar em $z = 1$. Fazendo isso, obtemos o seguinte sistema de equações

$$\begin{aligned} \frac{d\langle n_\alpha \rangle}{dt} &= -\rho \langle n_\alpha \rangle + kP_\alpha(t) - h \langle n_\alpha \rangle + f \langle n_\beta \rangle, \\ \frac{d\langle n_\beta \rangle}{dt} &= -\rho \langle n_\beta \rangle + k\chi P_\beta(t) + h \langle n_\alpha \rangle - f \langle n_\beta \rangle, \end{aligned} \quad (3.21)$$

Tal sistema pode ser escrito na forma matricial:

$$\frac{dX}{dt} = \mathbb{A}X + \mathbb{B}, \quad (3.22)$$

onde

$$X = \begin{pmatrix} \langle n_\alpha \rangle(t) \\ \langle n_\beta \rangle(t) \end{pmatrix}, \quad \mathbb{A} = \begin{pmatrix} -(\rho + h) & f \\ h & -(\rho + f) \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} kP_\alpha(t) \\ k\chi P_\beta(t) \end{pmatrix}.$$

Para resolver a Eq. (3.22), utilizamos o mesmo procedimento adotado para obter as probabilidades $P_\alpha(t)$ e $P_\beta(t)$: encontrar os autovalores e autovetores de $\mathbb{A}$, escrever uma matriz $\mathbb{S}$ que, através da transformação de semelhança $\mathbb{S}^{-1}\mathbb{A}\mathbb{S}$, fornece uma matriz diagonal cujos elementos de sua diagonal principal serão os próprios autovalores de $\mathbb{A}$. Com isso, o sistema (3.22) é desacoplado e as expressões para $\langle n_\alpha \rangle(t)$ e $\langle n_\beta \rangle(t)$ podem ser obtidas.

Os autovalores de $\mathbb{A}$ são $\lambda_1 = -\rho$ e $\lambda_2 = -\rho(1 + \epsilon)$ e seus respectivos autovetores são

$$v_1 = \begin{pmatrix} f \\ h \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

iguais aos autovetores de $\mathbb{H}$. Naturalmente, para diagonalizar a matriz $\mathbb{A}$, utilizaremos uma matriz igual à matriz $\mathbb{S}$

$$\mathbb{S} = \begin{pmatrix} f & 1 \\ h & -1 \end{pmatrix}, \quad \mathbb{S}^{-1} = \frac{1}{\epsilon\rho} \begin{pmatrix} 1 & 1 \\ h & -f \end{pmatrix},$$

onde usamos $h + f = \epsilon\rho$.

Fazendo a substituição $X = \mathbb{S}Y$, onde

$$Y = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix},$$

temos

$$\frac{d(\mathbb{S}Y)}{dt} = \mathbb{A}(\mathbb{S}Y) + \mathbb{B}.$$

Utilizando novamente o fato de que os elementos de $\mathbb{S}$ são constantes e multiplicando à esquerda por $\mathbb{S}^{-1}$, temos

$$\frac{dY}{dt} = \mathbb{D}Y + \mathbb{S}^{-1}\mathbb{B}, \quad \mathbb{D} = \begin{pmatrix} -\rho & 0 \\ 0 & -\rho(1 + \epsilon) \end{pmatrix}.$$

Dessa forma, conseguimos desacoplar as duas equações:

$$\frac{d}{dt} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} -\rho & 0 \\ 0 & -\rho(1 + \epsilon) \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} + k \begin{pmatrix} \frac{(1-\chi)}{\epsilon\rho} P_\alpha(t) + \frac{\chi}{\epsilon\rho} \\ (1 - p_\alpha)P_\alpha(t) + \chi p_\alpha(P_\alpha(t) - 1) \end{pmatrix}, \quad (3.23)$$

que podemos reescrever da seguinte forma:

$$\begin{aligned} \frac{d}{dt} y_1(t) &= -\rho y_1(t) + \frac{k(1-\chi)}{\epsilon\rho} P_\alpha(t) + \frac{k\chi}{\epsilon\rho} \\ \frac{d}{dt} y_2(t) &= -\rho(1 + \epsilon) y_2(t) + k[1 + p_\alpha(\chi - 1)] P_\alpha(t) - k\chi p_\alpha. \end{aligned} \quad (3.24)$$

As soluções $y_1(t)$ e $y_2(t)$ são dadas por

$$\begin{aligned} y_1(t) &= y_1(t_0)e^{-\rho(t-t_0)} + \frac{k(1-\chi)}{\epsilon\rho} \int_{t_0}^t e^{\rho s} P_\alpha(s) ds + \frac{k\chi}{\epsilon\rho} e^{-\rho t} \int_{t_0}^t e^{\rho s} ds \\ y_2(t) &= y_2(t_0)e^{-\rho(1+\epsilon)(t-t_0)} \\ &+ k[1+p_\alpha(\chi-1)]e^{-\rho(1+\epsilon)t} \int_{t_0}^t e^{\rho(1+\epsilon)s} P_\alpha(s) ds - k\chi p_\alpha \int_{t_0}^t e^{\rho(1+\epsilon)s} ds, \end{aligned} \quad (3.25)$$

onde $y_1(t_0)$ e $y_2(t_0)$ são as condições iniciais.

Obtidas as funções $y_1(t)$ e $y_2(t)$, podemos finalmente, através da substituição $X = \mathbb{S}Y$, obter as expressões para $\langle n_\alpha \rangle(t)$ e $\langle n_\beta \rangle(t)$. Escrevendo $\mathbb{S}$ em termos de ϵ e p_α , temos

$$\begin{aligned} \langle n_\alpha \rangle(t) &= \epsilon\rho p_\alpha y_1(t) + y_2(t) \\ \langle n_\beta \rangle(t) &= \epsilon\rho(1-p_\alpha)y_1(t) - y_2(t). \end{aligned} \quad (3.26)$$

As condições iniciais $y_1(t_0)$ e $y_2(t_0)$ podem ser determinadas com ajuda das Eqs.(3.26) e (3.25). Como a soma $\langle n_\alpha \rangle(t) + \langle n_\beta \rangle(t)$ deve recuperar o valor médio total $\langle n \rangle(t)$, no instante $t = t_0$, temos que

$$\langle n_\alpha \rangle(t_0) + \langle n_\beta \rangle(t_0) = \langle n \rangle(t_0) \Rightarrow n_0 = \epsilon\rho y_1(t_0) \Rightarrow y_1(t_0) = \frac{n_0}{\epsilon\rho}. \quad (3.27)$$

Já a condição $y_2(t_0)$ pode ser escrita a partir do valor inicial de $\langle n_\alpha \rangle(t)$, escrita como $\langle n_\alpha \rangle_0$. De acordo com as Eqs.(3.26), (3.25) e (3.27), temos:

$$\langle n_\alpha \rangle_0 = y_1(t_0) + y_2(t_0) \Rightarrow y_2(t_0) = \langle n_\alpha \rangle_0 - p_\alpha n_0. \quad (3.28)$$

Com as Eqs. (3.25), (3.27) e (3.28), podemos reescrever a Eq.(3.26):

$$\begin{aligned}
\langle n_\alpha \rangle(t) &= p_\alpha n_0 e^{-\rho(t-t_0)} \\
&+ k(1-\chi)p_\alpha e^{-\rho t} \int_{t_0}^t e^{\rho s} P_\alpha(s) ds \\
&+ k\chi p_\alpha e^{-\rho t} \int_{t_0}^t e^{\rho s} ds \\
&+ [\langle n_\alpha \rangle_0 - p_\alpha n_0] e^{-\rho(1+\epsilon)(t-t_0)} \\
&+ k[1+p_\alpha(\chi-1)] e^{-\rho(1+\epsilon)t} \int_{t_0}^t e^{\rho(1+\epsilon)s} P_\alpha(s) ds \\
&- k\chi p_\alpha e^{-\rho(1+\epsilon)t} \int_{t_0}^t e^{\rho(1+\epsilon)s} ds \\
\langle n_\beta \rangle(t) &= (1-p_\alpha)n_0 e^{-\rho(t-t_0)} \\
&+ k(1-\chi)(1-p_\alpha) e^{-\rho t} \int_{t_0}^t e^{\rho s} P_\alpha(s) ds \\
&+ k\chi(1-p_\alpha) e^{-\rho t} \int_{t_0}^t e^{\rho s} ds \\
&- [\langle n_\alpha \rangle_0 - p_\alpha n_0] e^{-\rho(1+\epsilon)(t-t_0)} \\
&- k[1+p_\alpha(\chi-1)] e^{-\rho(1+\epsilon)t} \int_{t_0}^t e^{\rho(1+\epsilon)s} P_\alpha(s) ds \\
&+ k\chi p_\alpha e^{-\rho(1+\epsilon)t} \int_{t_0}^t e^{\rho(1+\epsilon)s} ds.
\end{aligned} \tag{3.29}$$

E com ajuda da Eq.(3.13), podemos obter uma expressão para $\langle n_\alpha \rangle(t)$:

$$\langle n_\alpha \rangle(t) = \langle n_\alpha \rangle_S + E e^{-\rho(t-t_0)} + F e^{-\epsilon\rho(t-t_0)} + G e^{-\rho(1+\epsilon)(t-t_0)}, \tag{3.30}$$

onde

$$\begin{aligned}
\langle n_\alpha \rangle_S &= p_\alpha n_S + \frac{N(1-\chi)p_\alpha p_\beta}{1+\epsilon} = N p_\alpha + \frac{N(\chi-1)\epsilon p_\alpha p_\beta}{1+\epsilon} \\
E &= p_\alpha (n_0 - n_S - C) = p_\alpha (\Delta n - C) \\
F &= p_\alpha C + H \\
G &= \langle n_\alpha \rangle_0 - p_\alpha n_0 - \frac{N(1-\chi)p_\alpha p_\beta}{1+\epsilon} - H = \Delta n_\alpha - p_\alpha \Delta n - H \\
H &= N(p_\beta + \chi p_\alpha) \Delta p \\
\Delta n_\alpha &= \langle n_\alpha \rangle_0 - \langle n_\alpha \rangle_S
\end{aligned} \tag{3.31}$$

e $\langle n_\alpha \rangle_S$ é o valor assintótico assumido por $\langle n_\alpha \rangle(t)$.

As dinâmicas de $\langle n_\alpha \rangle(t)$ e $\langle n_\beta \rangle(t)$ são caracterizadas por três escalas de tempo. Além

dos termos proporcionais a $\exp(-\rho t)$ e $\exp(-\epsilon \rho t)$, os quais já sabemos que caracterizam as dinâmicas de $P_\alpha(t)$ e $\langle n \rangle(t)$, há um terceiro termo, que é proporcional a $\exp(-\rho(1+\epsilon)t)$. Esse termo corresponde aos efeitos simultâneos da degradação e do chaveamento sobre o número de moléculas transcritas quando o gene está ligado (n_α) ou quando está desligado (n_β). Em outras palavras, se o gene está ligado (desligado) e há n moléculas de mRNA transcritas, podem ocorrer ao mesmo tempo uma alteração no número de moléculas (produção ou degradação) e uma transição do gene para o nível desligado (ligado). Podemos ver que essa conjunção de efeitos sempre se estabiliza antes da estabilização do chaveamento e do processo transcricional.

Uma constatação importante é que se, por um lado, a degradação e a produção de uma molécula afetam n_α e n_β (dependendo do nível de ativação do gene) e, conseqüentemente, o número médio total de moléculas, por outro lado, as transições ligado $\rightarrow$ desligado e desligado $\rightarrow$ ligado, apesar de importantes para a caracterização de $\langle n_\alpha \rangle$ e $\langle n_\beta \rangle$, não afetam o número médio de moléculas de mRNA, conforme mostram as Eqs.(3.25), (3.26) e (3.19). Isso porque, ao contabilizar $\langle n_\alpha \rangle$ e $\langle n_\beta \rangle$, $\langle n \rangle$ não faz distinção entre os níveis de ativação e, se não há degradação e produção de mRNA, a quantidade acrescida em n_α é a mesma decrescida em n_β e vice-versa. Num certo sentido, as transições entre estados ligado e desligado são análogas às forças de vínculo (internas) de um sistema mecânico.

Além de $P_\alpha(t)$, $\langle n \rangle(t)$ e $\langle n_\alpha \rangle(t)$, outra quantidade importante para nossos propósitos e que pode ser calculada é $\langle n^2 \rangle(t)$. Para obtê-la, devemos novamente aplicar a Eq. (3.7) à soma das Eqs. (3.4), recordando que $\phi(z, t) = \alpha(z, t) + \beta(z, t)$. Isso fornece a seguinte equação diferencial para $\langle n^2 \rangle(t)$:

$$\begin{aligned} \frac{d\langle n^2 \rangle}{dt} &= -2\rho \langle n^2 \rangle + (2k\chi + \rho) \langle n \rangle + 2k(1 - \chi) \langle n_\alpha \rangle \\ &+ k(1 - \chi)P_\alpha(t) + k\chi, \end{aligned} \quad (3.32)$$

cuja solução é dada por

$$\begin{aligned}\langle n^2 \rangle(t) &= \langle n^2 \rangle_0 e^{-2\rho(t-t_0)} + (2k\chi + \rho)e^{-2\rho t} \int_0^t e^{2\rho s} \langle n \rangle(s) ds \\ &+ 2k(1-\chi)e^{-2\rho t} \int_0^t e^{2\rho s} \langle n_\alpha \rangle(s) ds + k(1-\chi)e^{-2\rho t} \int_0^t e^{2\rho s} P_\alpha(s) ds \\ &+ k\chi e^{-2\rho t} \int_0^t e^{2\rho s} ds.\end{aligned}\quad (3.33)$$

onde $\langle n^2 \rangle_0 = \langle n^2 \rangle(t_0)$ é a condição inicial.

Utilizando as Eqs.(3.13), (3.19), (3.20), (3.30) e (3.31), temos

$$\langle n^2 \rangle(t) = \langle n^2 \rangle_S + S e^{-2\rho(t-t_0)} + T e^{-\rho(t-t_0)} + U e^{-\epsilon\rho(t-t_0)} + V e^{-\rho(1+\epsilon)(t-t_0)}, \quad (3.34)$$

onde

$$\begin{aligned}\langle n^2 \rangle_S &= \frac{(2k\chi + \rho)n_S + 2k(1-\chi)\langle n_\alpha \rangle_S + k(1-\chi)p_\alpha + k\chi}{2\rho} \\ &= \langle n \rangle_S^2 + \langle n \rangle_S + \frac{N^2(1-\chi)^2 p_\alpha p_\beta}{1+\epsilon} \\ S &= \Delta n^2 - T - U - V \\ T &= (2n_S + 1)(\Delta n - C) \\ U &= \frac{(2N\chi + 1)C + 2N(1-\chi)F + N(1-\chi)\Delta p}{2-\epsilon} \\ V &= \frac{2N(1-\chi)G}{1-\epsilon} \\ \Delta n^2 &= \langle n^2 \rangle_0 - \langle n^2 \rangle_S.\end{aligned}\quad (3.35)$$

É imediato ver que, para tempos suficientemente grandes, $\langle n^2 \rangle(t)$ assume o valor $\langle n^2 \rangle_S$.

Assim como as dinâmicas de $\langle n \rangle(t)$ e $\langle n_\alpha \rangle(t)$, a dinâmica de $\langle n^2 \rangle(t)$ também será analisada. A Eq. (3.34) mostra que a dinâmica de $\langle n \rangle(t)$ é governada por quatro escalas de tempo características das quais duas, representadas pelos termos $\exp(-\epsilon\rho t)$ e $\exp(-\rho(1+\epsilon)t)$, são herança do chaveamento.

Com as expressões analíticas para $\langle n \rangle$ e para o $\langle n^2 \rangle$ calculadas, podemos escrever a variância σ^2 , dada por

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2, \quad (3.36)$$

e, conseqüentemente, a flutuação, dada pelo desvio quadrático σ . É imediato ver que se $\epsilon > 1$, a dinâmica da flutuação σ é governada, para tempos suficientemente grandes,

pelo termo $\exp(-\rho t)$, ou seja, pela dinâmica da degradação, de modo que o valor médio e a flutuação acabam evoluindo no mesmo ritmo. Também é imediato ver, a partir da definição de variância e do valor assintótico $\langle n^2 \rangle_S$, que o valor assintótico da variância é

$$\sigma_S^2 = \langle n \rangle_S + \frac{N^2(1-\chi)^2 p_\alpha p_\beta}{1+\epsilon}. \quad (3.37)$$

Daí, é imediato obter o valor assintótico do desvio quadrático ou flutuação

$$\sigma_S = \sqrt{N(1-\chi)p_\alpha + N\chi + \frac{N^2(1-\chi)^2 p_\alpha p_\beta}{1+\epsilon}}. \quad (3.38)$$

Das Eqs.(3.37) e (3.38) é imediato ver que a variância é minimizada para probabilidades estacionárias p_α (ou p_β) que sejam próximas de 0 ou de 1, o que é intuitivamente condizente com o fato de se tratar de um sistema de dois estados: em um sistema governado pela estocasticidade e em regime estacionário, quanto mais “certeza” houver para se produzir ou para não se produzir, menores serão as flutuações em torno do valor médio a ser produzido, isto é, quanto mais próxima de 0 ou 1 for a probabilidade do sistema, menor será a variância do processo; longe dessas probabilidades, o sistema sofre as influências do ambiente de modo que as flutuações em torno do valor esperado só tendem a aumentar.

Embora o valor médio $\langle n \rangle$, a variância σ^2 e a flutuação σ forneçam informações quantitativas sobre relevantes o processo estocástico, é conveniente trabalhar também com grandezas que quantifiquem informações entre estas grandezas como o *ruído* e o *fator de Fano*. O ruído, definido pela razão $\sigma/\langle n \rangle$, mede o quão larga é a distribuição de probabilidade ou, em outras palavras, o quão preciso é o processo estocástico e o fator de Fano, dado pela razão $\sigma^2/\langle n \rangle$, mede o desvio da distribuição de probabilidade de interesse com respeito à distribuição de Poisson de modo que o processo em questão pode ser classificado como *poissoniano* (se a razão for igual à unidade), *subpoissoniano* (se a razão for menor que a unidade) ou *superpoissoniano* (se a razão for maior que a unidade). Nesse trabalho, no entanto, restringimos nossa análise ao valor médio, à flutuação e ao ruído.

3.3 Distribuição de probabilidade

Nessa seção, calcularemos as distribuições de probabilidade do estado ligado e do estado desligado, respectivamente dadas por $\alpha_n(t)$ e $\beta_n(t)$, soluções das Eqs.(3.1) e (3.2), bem como a distribuição de probabilidade total $\phi_n(t)$. Contudo, para fins práticos, calcularemos primeiro tais distribuições no caso estacionário, ou seja, quando as distribuições $\alpha_n(t)$ e $\beta_n(t)$ não evoluem temporalmente e depois calcularemos as distribuições para o caso em que exibem evolução temporal.

3.3.1 Regime estacionário

Como no limite assintótico não há evolução temporal das distribuições de probabilidade $\alpha_n(t)$ e $\beta_n(t)$, elas naturalmente serão representadas respectivamente por α_n e β_n e suas respectivas funções geratrizes serão representadas por $\alpha(z)$ e $\beta(z)$. Isso implicará em assumir $\frac{\partial \alpha(z,t)}{\partial t} = 0$ e $\frac{\partial \beta(z,t)}{\partial t} = 0$ nas Eqs.(3.4), o que gera

$$0 = (z - 1) \left[k\alpha_S - \rho \frac{d\alpha_S}{dz} \right] - h\alpha_S + f\beta_S \quad (3.39)$$

e

$$0 = (z - 1) \left[k\chi\beta_S - \rho \frac{d\beta_S}{dz} \right] + h\alpha_S - f\beta_S, \quad (3.40)$$

onde α_S e β_S representam, respectivamente, $\alpha(z)$ e $\beta(z)$ e o índice S indica que a solução é estacionária. O primeiro passo para obter as funções $\alpha(z)$ e $\beta(z)$ é isolar β_S na Eq.(3.40),

$$\beta_S = \frac{1}{f} \left[\rho(z - 1) \frac{d\alpha_S}{dz} - (zk - k - h)\alpha_S \right], \quad (3.41)$$

e substituí-lo na Eq.(3.40). A equação resultante é

$$\frac{d^2\alpha_S}{dz^2} + p \frac{d\alpha_S}{dz} + q\alpha_S = 0, \quad (3.42)$$

onde

$$p(z) = \frac{(1-z)(1+\chi)k + \rho + h + f}{\rho(z-1)}, \quad (3.43)$$

e

$$q(z) = \frac{k^2\chi(1-z) - k(\rho + f + \chi h)}{\rho^2(z-1)}. \quad (3.44)$$

Embora ambas as funções p e q possuam um polo simples em $z = 1$ e uma irregularidade em $z = \infty$, o que, conseqüentemente, faz com que a solução geral seja divergente em $z = 1$, é possível obter uma solução particular regular. Tal solução, que será expressa em termos de funções hipergeométricas, é fruto da seguinte mudança de variáveis

$$\xi = \frac{k(\chi-1)(z-1)}{\rho} \quad \alpha_S = \exp\left[\frac{k(z-1)}{\rho}\right] \tilde{\alpha}(\xi), \quad (3.45)$$

que nos permite reescrever a Eq.(3.42) na forma canônica da equação de Kummer

$$\xi \frac{d^2 \tilde{\alpha}}{d\xi^2} + (b+1-\xi) \frac{d\tilde{\alpha}}{d\xi} - a\tilde{\alpha} = 0, \quad (3.46)$$

onde

$$a = \frac{h}{\rho} = \epsilon p_\beta, \quad b = \frac{h+f}{\rho} = \epsilon. \quad (3.47)$$

A solução geral da Eq.(3.48) é uma combinação linear das funções de Kummer do tipo M e do tipo U

$$\tilde{\alpha}(\xi) = C_1 M(a, b+1, \xi) + C_2 U(a, b+1, \xi), \quad (3.48)$$

onde C_1 e C_2 são constantes(140). Como em $z = 1$ ($\xi = 0$) a solução do tipo M é regular e a do tipo U é irregular, a solução que procuramos para α_S deve ser escrita em termos da solução do tipo M, ou seja, C_2 deve ser nula. Daí, utilizando a Eq.(3.41), impondo a condição de normalização $\alpha(1, t) + \beta(1, t) = 1$ e com o auxílio da fórmula 13.4.10 da referência (140)(uma relação de recorrência para as funções de Kummer), obtemos as soluções α_S e β_S :

$$\begin{aligned} \alpha_S(z) &= \frac{b-a}{b} \exp\left(\frac{k(z-1)}{\rho}\right) M\left(a, b+1; \frac{k(\chi-1)(z-1)}{\rho}\right) \\ &= p_\alpha \exp[N(z-1)] M(\epsilon p_\beta, \epsilon+1; N(\chi-1)(z-1)), \end{aligned} \quad (3.49)$$

$$\begin{aligned}\beta_S(z) &= \frac{a}{b} \exp\left(\frac{k(z-1)}{\rho}\right) M\left(a+1, b+1; \frac{k(\chi-1)(z-1)}{\rho}\right) \\ &= p_\beta \exp[N(z-1)] M(\epsilon p_\beta + 1, \epsilon + 1; N(\chi-1)(z-1)).\end{aligned}\quad (3.50)$$

Utilizando a fórmula 13.4.3 também da referência (140), podemos escrever uma fórmula compacta para ϕ_S :

$$\phi_S(z) = \alpha_S(z) + \beta_S(z) = \exp[N(z-1)] M(\epsilon p_\beta, \epsilon; N(\chi-1)(z-1)). \quad (3.51)$$

As distribuições de probabilidade para o caso estacionário podem, então, ser finalmente obtidas — com o auxílio da regra de Leibniz e das fórmulas para derivadas das funções de Kummer[†] — realizando sucessivas derivadas das funções geratrizes com respeito à variável z e tomando, em seguida, $z = 0$

$$\alpha_n = p_\alpha N^n \frac{\exp(-N)}{n!} \sum_{s=0}^n \binom{n}{s} (\chi-1)^s \frac{(p_\beta \epsilon)_s}{(\epsilon+1)_s} M(p_\beta \epsilon + s, \epsilon + 1 + s; N(1-\chi)) \quad (3.52)$$

$$\beta_n = p_\beta N^n \frac{\exp(-N)}{n!} \sum_{s=0}^n \binom{n}{s} (\chi-1)^s \frac{(p_\beta \epsilon + 1)_s}{(\epsilon+1)_s} M(p_\beta \epsilon + 1 + s, \epsilon + 1 + s; N(1-\chi)) \quad (3.53)$$

$$\phi_n = N^n \frac{\exp(-N)}{n!} \sum_{s=0}^n \binom{n}{s} (\chi-1)^s \frac{(p_\beta \epsilon)_s}{(\epsilon)_s} M(p_\beta \epsilon + s, \epsilon + s; N(1-\chi)). \quad (3.54)$$

3.3.2 Regime dependente do tempo

Para obtermos as distribuições de probabilidade que resolvem as Eqs.(3.4), primeiramente fazemos a seguinte mudança de variável

$$\tau = e^{-\rho t} \quad (3.55)$$

resultando em

$$\begin{aligned}-\tau \frac{\partial \alpha(z, t)}{\partial \tau} &= (z-1) [N \alpha(z, t) - \frac{\partial \alpha(z, t)}{\partial z}] - p_\beta \epsilon \alpha(z, t) + p_\alpha \epsilon \beta(z, t) \\ -\tau \frac{\partial \beta(z, t)}{\partial \tau} &= (z-1) [N \chi \beta(z, t) - \frac{\partial \beta(z, t)}{\partial z}] + p_\beta \epsilon \alpha(z, t) - p_\alpha \epsilon \beta(z, t),\end{aligned}\quad (3.56)$$

[†]Fórmula 13.4.9 da referência (140).

onde também utilizamos as Eqs. (3.14) e (3.16) e a relação $N = k/\rho$.

Uma segunda mudança de variáveis é realizada

$$\mu = 1 + (z - 1)e^{-\rho t} = 1 + (z - 1)\tau \quad (3.57)$$

$$\nu = z,$$

e as Eqs.(3.56) são reescritas

$$(\nu - 1) \left(N\alpha(\mu, \nu) - \frac{\partial \alpha(\mu, \nu)}{\partial \nu} \right) - p_\beta \epsilon \alpha(\mu, \nu) + p_\alpha \epsilon \beta(\mu, \nu) = 0 \quad (3.58)$$

$$(\nu - 1) \left(N\chi\beta(\mu, \nu) - \frac{\partial \beta(\mu, \nu)}{\partial \nu} \right) + p_\beta \epsilon \beta(\mu, \nu) - p_\alpha \epsilon \alpha(\mu, \nu) = 0. \quad (3.59)$$

O sistema composto pelas Eqs (3.58) e (3.59) possuem a mesma estrutura que o sistema formado pelas Eqs.(3.39) e (3.40), ou seja, as transformações propostas deixaram as equações diferenciais idênticas às equações diferenciais que descrevem o caso assintótico. Isso significa que o mesmo procedimento para resolver as Eqs.(3.39) e (3.40) pode ser adotado. Assim, uma equação diferencial de segunda ordem para $\alpha(\mu, \nu)$ é obtida e escrita na forma canônica da equação de Kummer.

Como foi discutido na seção anterior, a solução de uma equação de Kummer é uma combinação linear de funções de Kummer do tipo M e do tipo U. Como há oito soluções para tal equação[‡] e nem todas são linearmente independentes entre si[§], foi escolhido o par de funções linearmente independentes dado pelas fórmulas 13.1.12 e 13.1.13 da referência (140). Assim as soluções $\alpha(\mu, \nu)$ e $\beta(\mu, \nu)$ são dadas pelas combinações lineares

$$\alpha(\mu, \nu) = F(\mu)\alpha_S(\nu) + G(\mu)A(\nu) \quad (3.60)$$

$$\beta(\mu, \nu) = F(\mu)\beta_S(\nu) + G(\mu)B(\nu),$$

[‡]Fórmulas 13.1.12 - 13.1.19 da referência (140).

[§]Fórmulas 13.1.20 - 13.1.26 da referência (140).

onde

$$\begin{aligned}
\alpha_S(\nu) &= p_\alpha \exp[N(\nu - 1)]M(a, 1 + b; \xi) \\
A(\nu) &= -(1 - b)\xi^{(-b)} \exp[N(\nu - 1)]M(a - b, 1 - b; \xi) \\
\beta_S(\nu) &= p_\beta \exp[N(\nu - 1)]M(1 + a, 1 + b; \xi) \\
B(\nu) &= (1 - b)\xi^{(-b)} \exp[N(\nu - 1)]M(1 + a - b, 1 - b; \xi)
\end{aligned} \tag{3.61}$$

e

$$\xi = N(\chi - 1)(\nu - 1). \tag{3.62}$$

As funções $F(\mu)$ e $G(\mu)$ são constantes de integração calculadas a partir das condições iniciais. Com o auxílio das fórmulas 13.4.4 e 13.4.12, podemos obter uma fórmula para a função $\phi(\mu, \nu)$:

$$\begin{aligned}
\phi(\mu, \nu) &= \alpha(\mu, \nu) + \beta(\mu, \nu) \\
&= \exp(N(\nu - 1)) [F(\mu)M(a, b; \xi) + G(\mu)\xi^{(1+b)}M(1 + a - b, 2 - b; \xi)]
\end{aligned} \tag{3.63}$$

As Eqs.(3.60) podem ser reescritas na forma matricial

$$\begin{bmatrix} \alpha(\mu, \nu) \\ \beta(\mu, \nu) \end{bmatrix} = \begin{bmatrix} \alpha_S(\nu) & A(\nu) \\ \beta_S(\nu) & B(\nu) \end{bmatrix} \begin{bmatrix} F(\mu) \\ G(\mu) \end{bmatrix} \tag{3.64}$$

e, resolvendo tal sistema para $F(\mu)$ e $G(\mu)$, temos

$$\begin{bmatrix} F(\mu) \\ G(\mu) \end{bmatrix} = \frac{1}{\alpha_S(\nu)B(\nu) - A(\nu)\beta_S(\nu)} \begin{bmatrix} B(\nu) & -A(\nu) \\ -\beta_S(\nu) & \alpha_S(\nu) \end{bmatrix} \begin{bmatrix} \alpha(\mu, \nu) \\ \beta(\mu, \nu) \end{bmatrix} \tag{3.65}$$

onde a quantidade $\delta = \alpha_S(\nu)B(\nu) - A(\nu)\beta_S(\nu)$ vale:

$$\delta = (1 - \epsilon)[N(\chi - 1)(\nu - 1)]^{-\epsilon} \exp[N(1 + \chi)(\nu - 1)]. \tag{3.66}$$

Uma vez que $F(\mu)$ e $G(\mu)$ são calculadas a partir das condições iniciais, temos de tomar $t = 0$, o que implica em $\tau = 1$ (de acordo com a Eq.(3.55)) e $\mu = \nu$ (de acordo com

a Eq.(3.57)). Dessa forma, as funções $F(\mu)$ e $G(\mu)$ podem, então, ser escritas

$$\begin{aligned} F(\mu) &= \exp(-N\chi(\mu-1))[M(1+a-b, 1-b; \tilde{\xi})C_1(\mu) \\ &+ M(a-b, 1-b; \tilde{\xi})C_2(\mu)] \end{aligned} \quad (3.67)$$

$$\begin{aligned} G(\mu) &= \frac{\tilde{\xi}^{(b)} \exp(-N\chi(\mu-1))}{b-1} [p_\beta M(1+a, 1+b; \tilde{\xi})C_1(\mu) \\ &- p_\alpha M(a, 1+b; \tilde{\xi})C_2(\mu)]. \end{aligned} \quad (3.68)$$

onde $\tilde{\xi} = N(\chi-1)(\mu-1) = N(\chi-1)(z-1)\tau$ e $C_1(\mu)$ e $C_2(\mu)$ são as condições iniciais.

Com as funções $F(\mu)$ e $G(\mu)$ em mãos e voltando para as variáveis z e τ , podemos reescrever as Eqs.(3.60) e, assim, obter as funções $\alpha(z, \tau)$ e $\beta(z, \tau)$:

$$\begin{aligned} \alpha(z, \tau) &= P(z, \tau)C_1(z, \tau) + Q(z, \tau)C_2(z, \tau) \\ \beta(z, \tau) &= S(z, \tau)C_1(z, \tau) + T(z, \tau)C_2(z, \tau) \end{aligned} \quad (3.69)$$

onde

$$\begin{aligned} P(z, \tau) &= \exp(-N(z-1)(1-\tau\chi)) \\ &\times [p_\alpha M(1+a-b, 1-b; N\tau(\chi-1)(z-1))M(a, 1+b; N(\chi-1)(z-1)) \\ &+ p_\beta \tau^{(b)} M(1+a, 1+b; N\tau(\chi-1)(z-1))M(a-b, 1-b; N(\chi-1)(z-1))] \end{aligned} \quad (3.70)$$

$$\begin{aligned} Q(z, \tau) &= \exp(-N(z-1)(1-\tau\chi)) \\ &\times p_\alpha [M(a-b, 1-b; N\tau(\chi-1)(z-1))M(a, 1+b; N(\chi-1)(z-1)) \\ &- \tau^{(b)} M(a, 1+b; N\tau(\chi-1)(z-1))M(a-b, 1-b; N(\chi-1)(z-1))] \end{aligned} \quad (3.71)$$

$$\begin{aligned} S(z, \tau) &= \exp(-N(z-1)(1-\tau\chi)) \\ &\times p_\beta [M(1+a-b, 1-b; N\tau(\chi-1)(z-1))M(1+a, 1+b; N(\chi-1)(z-1)) \\ &- \tau^{(b)} M(1+a, 1+b; N\tau(\chi-1)(z-1))M(1+a-b, 1-b; N(\chi-1)(z-1))] \end{aligned} \quad (3.72)$$

$$\begin{aligned}
T(z, \tau) &= \exp(-N(z-1)(1-\tau\chi)) \\
&\times [p_\beta M(a-b, 1-b; N\tau(\chi-1)(z-1))M(1+a, 1+b; N(\chi-1)(z-1)) \\
&+ p_\alpha \tau^{(b)} M(a, 1+b; N\tau(\chi-1)(z-1))M(1+a-b, 1-b; N(\chi-1)(z-1))].
\end{aligned} \tag{3.73}$$

Com as soluções $\alpha(z, \tau)$ e $\beta(z, \tau)$ e utilizando a Eq.(3.63), a função geratriz $\phi(z, \tau)$ é escrita

$$\begin{aligned}
\phi(z, \tau) &= \exp[-N(z-1)(1-\tau\chi)] \\
&\times \{C_1(z, \tau)[M(1+a-b, 1-b; N\tau(\chi-1)(z-1))M(a, b; N(\chi-1)(z-1)) \\
&- \frac{p_\beta N(\chi-1)(z-1)\tau^b}{1-b} \\
&\times M(1+a, 1+b; N\tau(\chi-1)(z-1))M(1+a-b, 2-b; N(\chi-1)(z-1))] \\
&+ C_2(z, \tau)[M(a-b, 1-b; N\tau(\chi-1)(z-1))M(a, b; N(\chi-1)(z-1)) \\
&+ \frac{p_\alpha N(\chi-1)(z-1)\tau^b}{1-b} \\
&\times M(a, 1+b; N\tau(\chi-1)(z-1))M(1+a-b, 2-b; N(\chi-1)(z-1))]\}
\end{aligned} \tag{3.74}$$

Se as condições iniciais forem escolhidas de modo que, em $t = 0$, o sistema se encontre em um estado estacionário, então temos

$$\begin{aligned}
C_1(z, \tau) &= \overline{p}_\alpha \exp[N\tau(z-1)]M(\epsilon\overline{p}_\beta, 1+\epsilon; N\tau(\chi-1)(z-1)) \\
C_2(z, \tau) &= \overline{p}_\beta \exp[N\tau(z-1)]M(1+\epsilon\overline{p}_\beta, 1+\epsilon; N\tau(\chi-1)(z-1))
\end{aligned} \tag{3.75}$$

onde $\overline{p}_\alpha$ e $\overline{p}_\beta$ são as probabilidades de se encontrar inicialmente o gene nos estados ativo e inativo. Podemos notar que $C_1(z, \tau)$ e $C_2(z, \tau)$ nada mais são do que as expressões para $\alpha_S(z)$ e $\beta_S(z)$, respectivamente, mas com o fator multiplicativo τ no argumento, garantindo que, em $t = 0$, o sistema está em um estado estacionário caracterizado $\overline{p}_\alpha$ e $\overline{p}_\beta$.

Com as funções geratrizes $\alpha(z, \tau)$ e $\beta(z, \tau)$, dadas pelas Eqs.(3.69), e com o auxílio

das Eqs. (3.5), podemos finalmente obter as distribuições de probabilidade $\alpha_n(t)$ e $\beta_n(t)$:

$$\begin{aligned} \alpha_n(\tau) &= \frac{[N(1 + \tau(1 - \chi))]^n \exp[-N(1 + \tau(1 - \chi))]}{n!} \\ &\times \sum_{r=0}^n \sum_{s=0}^r \sum_{p=0}^s \binom{n}{r} \binom{r}{s} \binom{s}{p} \frac{[N(\chi - 1)]^r \tau^s}{[N(1 + \tau(1 - \chi))]^r} \\ &\times \{p_\alpha \bar{p}_\alpha M_1(r, s) M_2(s, p, \tau) M_3(p, \tau) + \tau^\epsilon p_\beta \bar{p}_\alpha M_4(r, s) M_5(s, p, \tau) M_3(p, \tau) \\ &+ p_\alpha \bar{p}_\beta M_1(r, s) M_4(s, p, \tau) M_6(p, \tau) - \tau^\epsilon p_\alpha \bar{p}_\beta M_4(r, s) M_1(s, p, \tau) M_6(p, \tau)\}, \end{aligned} \quad (3.76)$$

$$\begin{aligned} \beta_n(\tau) &= \frac{[N(1 + \tau(1 - \chi))]^n \exp[-N(1 + \tau(1 - \chi))]}{n!} \\ &\times \sum_{r=0}^n \sum_{s=0}^r \sum_{p=0}^s \binom{n}{r} \binom{r}{s} \binom{s}{p} \frac{[N(\chi - 1)]^r \tau^s}{[N(1 + \tau(1 - \chi))]^r} \\ &\times \{p_\beta \bar{p}_\alpha M_5(r, s) M_2(s, p, \tau) M_3(p, \tau) - \tau^\epsilon p_\beta \bar{p}_\alpha M_2(r, s) M_5(s, p, \tau) M_3(p, \tau) \\ &+ p_\beta \bar{p}_\beta M_5(r, s) M_4(s, p, \tau) M_6(p, \tau) + \tau^\epsilon p_\alpha \bar{p}_\beta M_2(r, s) M_1(s, p, \tau) M_6(p, \tau)\}, \end{aligned} \quad (3.77)$$

onde

$$\begin{aligned} M_1(r, s) &= \frac{(\epsilon p_\beta)_{r-s}}{(1 + \epsilon)_{r-s}} M(\epsilon p_\beta + r - s, 1 + \epsilon + r - s; N(1 - \chi)) \\ M_1(s, p, \tau) &= \frac{(\epsilon p_\beta)_{s-p}}{(1 + \epsilon)_{s-p}} M(\epsilon p_\beta + s - p, 1 + \epsilon + s - p; N\tau(1 - \chi)) \\ M_2(s, p, \tau) &= \frac{(1 + \epsilon p_\beta - \epsilon)_{s-p}}{(1 - \epsilon)_{s-p}} M(1 + \epsilon p_\beta - \epsilon + s - p, 1 - \epsilon + s - p; N\tau(1 - \chi)) \\ M_2(r, s) &= \frac{(1 + \epsilon p_\beta - \epsilon)_{r-s}}{(1 - \epsilon)_{r-s}} M(1 + \epsilon p_\beta - \epsilon + r - s, 1 - \epsilon + r - s; N(1 - \chi)) \\ M_3(p, \tau) &= \frac{(\epsilon \bar{p}_\beta)_p}{(1 + \epsilon)_p} M(\epsilon \bar{p}_\beta + p, 1 + \epsilon + p; N\tau(1 - \chi)) \\ M_4(r, s) &= \frac{(\epsilon p_\beta - \epsilon)_{r-s}}{(1 - \epsilon)_{r-s}} M(\epsilon p_\beta - \epsilon + r - s, 1 - \epsilon + r - s; N(1 - \chi)) \\ M_4(s, p, \tau) &= \frac{(\epsilon p_\beta - \epsilon)_{s-p}}{(1 - \epsilon)_{s-p}} M(\epsilon p_\beta - \epsilon + s - p, 1 - \epsilon + s - p; N\tau(1 - \chi)) \\ M_5(s, p, \tau) &= \frac{(1 + \epsilon p_\beta)_{s-p}}{(1 + \epsilon)_{s-p}} M(1 + \epsilon p_\beta + s - p, 1 + \epsilon + s - p; N\tau(1 - \chi)) \\ M_5(r, s) &= \frac{(1 + \epsilon p_\beta)_{r-s}}{(1 + \epsilon)_{r-s}} M(1 + \epsilon p_\beta + r - s, 1 + \epsilon + r - s; N(1 - \chi)) \\ M_6(p, \tau) &= \frac{(1 + \epsilon \bar{p}_\beta)_p}{(1 + \epsilon)_p} M(1 + \epsilon \bar{p}_\beta + p, 1 + \epsilon + p; N\tau(1 - \chi)). \end{aligned} \quad (3.78)$$

Chegamos ao final desse capítulo tendo em mãos expressões que descrevem um processo

de transcrição modelado estocasticamente pelas Eqs.(3.1) e (3.2). A partir do próximo capítulo trataremos de aplicar toda essa abordagem ao processo de transcrição do gene *eve* de *Drosophila* visando descrever o número médio de moléculas de mRNA, predizer o desvio quadrático e o ruído do processo transcricional e obter as distribuições de probabilidade do sistema.

Capítulo 4

Análise da segmentação embrionária em *Drosophila melanogaster*

No Capítulo 2 tivemos o primeiro contato com o problema da transcrição do gene *eve* de *Drosophila* e acesso — ainda que através de uma estimativa — ao número de moléculas de mRNA transcritas ao longo do embrião entre as classes temporais C13 a T8. No Capítulo 3, apresentamos o modelo estocástico que considera um gene binário (dois níveis de ativação) regulado externamente e calculamos grandezas importantes para a descrição estatística proposta. Nos capítulos que se seguem, incluindo este, trataremos de analisar a transcrição do gene *eve* à luz da estocasticidade proposta.

Especificamente nesse capítulo, nossa tarefa é descrever os dados experimentais através da expressão para o valor médio do número de moléculas de mRNA. Para isso, analisamos inicialmente a dinâmica da expressão de *eve* exclusivamente na região onde se forma a faixa 2 e propomos, baseado na modelagem estocástica trabalhada no capítulo anterior, uma abordagem na qual as taxas h e f sejam dependentes do tempo. Posteriormente, após uma breve análise das condições iniciais e do perfil espacial da expressão de mRNA, adicionamos gradualmente dependências espaciais em $h(t)$ e $f(t)$ e, ao final, obtemos uma expressão otimizada que reproduz toda a dinâmica da expressão de mRNA coletada pelos dados experimentais.

4.1 Dependência temporal nas taxas h e f

O primeiro passo para descrever o processo de formação da faixa 2 de mRNA baseado na abordagem estocástica proposta no Capítulo 3 é rever a própria modelagem, conciliando-a com as particularidades do organismo.

No Capítulo 3, tínhamos o seguinte cenário: quando o único sítio de ligação era ocupado por uma proteína ativadora (repressora), o gene passava a ficar ligado (desligado) e quando esta proteína se desligava do sítio, o gene era desligado (ligado). Como há proteínas de uma só espécie e apenas um sítio, as duas configurações de ocupância do sítio definem os dois estados de ativação. Entretanto, no fenômeno estudado, há proteínas de várias espécies e vários sítios de ligação de modo que há várias configurações, cada uma contribuindo de maneira distinta para a regulação, o que implica, em princípio, na ocorrência de vários níveis distintos de ativação.

Sob a perspectiva de um gene binário, propomos, então, separar todas as configurações em dois grupos: um grupo reunindo as configurações que tendem a favorecer a transcrição de mRNA de modo que suas probabilidades de ocorrência compõem p_α e o grupo das configurações que tendem a não favorecer a transcrição, cujas probabilidades de ocorrência compõem p_β .

Desse modo, as taxas h e f continuam sendo interpretadas como taxas às quais o gene realiza as transições ligado→desligado e desligado→ligado respectivamente mas sem estarem associadas à mudança de uma configuração específica para outra também específica. Embora grosseiro, esse procedimento visa não as contribuições de cada configuração mas sim o efeito regulatório total sobre a transcrição, o que, sob um ponto de vista efetivo, pode ajudar na compreensão dos mecanismos mais básicos da regulação transcricional.

O segundo passo para descrever o fenômeno é compreender a dinâmica do número médio de moléculas de mRNA na região onde se forma a referida faixa. Segundo os dados experimentais, verificamos que os registros de maior intensidade ocorrem em $AP = 40.5\%$ e $AP = 41.5\%$ mas, para efeito de análise, escolhemos analisar somente a dinâmica em

AP = 41.5%. Isso significa implicitamente que estamos assumindo, de certa forma, que a expressão em uma posição ao longo do eixo AP independe da expressão em outras regiões. Em outras palavras, estamos desconsiderando efeitos de transporte e de difusão.

A Figura 4.1 mostra o número estimado de moléculas de mRNA formado em AP = 41.5% durante as classes temporais. Os dados experimentais mostram que a expressão de mRNA em AP = 41.5% apresenta duas fases: a primeira, de C13 a T5, caracterizada por um crescimento sigmoidal e que aparentemente tende para um regime estacionário; a segunda, de T5 a T8, marcada por um decréscimo rápido e que, de acordo com os dados, nos permitem supor que a expressão tende para um regime, potencialmente estacionário, marcado por uma produção nula ou basal de mRNA de *eve*.

Figura 4.1 – Dinâmica do valor médio $\langle n \rangle$ em AP = 41.5.

Na primeira tentativa de reproduzir a dinâmica observada na Figura 4.1, uma função parabólica poderia ser proposta. Uma vez desconsiderados efeitos de difusão e de transporte, isso implicaria no fato dos dados experimentais poderem ser descritos por uma família de parábolas distribuídas ao longo do eixo AP. Sendo descrito por uma família de parábolas, a evolução temporal da expressão ao longo do eixo AP deveria ser caracterizada, grosseiramente, por evoluir a partir do padrão observado em C13, atingir o padrão observado durante o intervalo de T3 a T5 e depois retornar, de algum modo e em algum momento, ao mesmo padrão inicial, o que não ocorre, conforme pode ser observado na Figura 2.4. Mesmo assim, poderíamos supor ainda que essas parábolas estivessem, eventualmente, deslocadas no tempo umas em relação às outras. Todavia, essa hipótese e,

naturalmente, a hipótese da dependência temporal quadrática são descartadas à medida em que observamos a evolução temporal da expressão de mRNA em outras posições ao longo do eixo AP, como mostra a Figura 4.2. Assim, as evoluções temporais mostradas

Figura 4.2 – Dinâmica do valor médio $\langle n \rangle$ em AP = 35.5, AP = 38.5, AP = 43.5, AP = 45.5, AP = 60.5 e AP = 80.5. As evoluções temporais mostram claramente que funções parabólicas não são capazes de descrever tais comportamentos.

nas Figuras 4.1 e 4.2 mostram que a dependência temporal da dinâmica da expressão não pode ser dada por uma parábola. Em especial, as partes finais de cada dinâmica apresentam decaimentos, por aparentemente tenderem a zero ou a valores basais (próximos de zero), sugerem que funções exponenciais talvez sejam capazes de descrever a dinâmica da expressão.

Retomando à Figura 4.1, a primeira forma de analisar os dados apresentados é utilizar a Eq.(3.19). No entanto, tal fórmula não consegue reproduzir a dinâmica observada e isso porque, em processos descritos pela Eq.(3.19), comportamentos semelhantes a regimes estacionários — como o que ocorre entre T3 e T5 — são característicos para tempos suficientemente grandes e não de instantes intermediários do processo.

De modo mais detalhado, o que ocorre é que a expressão, como já mencionamos, tende a expressar cada vez menos transcritos, o que nos permite supor que a expressão se encaminha para um regime estacionário caracterizado por um número de transcritos muito pequeno ou mesmo nulo. De acordo com as Eqs.(3.14), (3.19) e (3.20), se o número assintótico de transcritos for pequeno (ou mesmo nulo), então, a probabilidade assintótica p_α deve necessariamente assumir um valor próximo de zero. Sendo p_α pequeno (próximo de zero), a probabilidade P_α e, conseqüentemente, o número de transcritos $\langle n \rangle$ decrescerão exponencialmente. Nesse sentido, atribuindo aos parâmetros k e ρ valores compatíveis com os valores apresentados na literatura, notamos que o decaimento descrito pela Eq.(3.19) ocorre em uma escala de tempo consideravelmente menor do que o observado nos dados experimentais. Além disso, o aumento do número de transcritos ocorre durante um tempo extremamente curto em relação ao observado nos dados experimentais e não atinge valores similares, mostrando que a Eq.(3.19) é incapaz de descrever sozinha os dados experimentais.

Entretanto, apesar do insucesso, isso não significa que devemos abandonar o modelo estocástico do gene binário mas que devemos rever os dados coletados e a concepção da abordagem com um pouco mais de atenção.

Do ponto de vista matemático, se traçarmos uma linha imaginária unindo os pontos experimentais, podemos perceber que a inclinação dessa linha, entre T3 e T5, tende a se anular e é bruscamente alterada em T5. Em termos do modelo do gene binário, uma brusca mudança na dinâmica dos dados experimentais deve resultar de uma mudança, também brusca, em um ou mais parâmetros do modelo.

No Capítulo 3, argumentamos que dentre os parâmetros utilizados no modelo — k , χ , ρ , h e f — os três primeiros (por resultarem de características intrínsecas da maquinaria transcricional, da sequência gênica e da própria molécula de mRNA) não deveriam apresentar nenhuma dependência com n ou com t ao passo que as taxas h e f , dependendo das interações gênicas, poderiam depender tanto de n quanto de t . Especificamente, no caso da formação da faixa 2 de mRNA de *eve*, não há indícios de autorregulação. Portanto, se quisermos analisar a dinâmica de mRNA de *eve* à luz do modelo estocástico do gene

binário, os únicos parâmetros que podem admitir uma dependência temporal são as taxas h e f .

Embora tenhamos admitido h e f como portadores de alguma dependência em t , não saberíamos, em princípio, dizer quais seriam as formas funcionais h e f . Entretanto, podemos concluir informações qualitativas importantes. As duas fases apresentadas pela expressão de mRNA são governadas por configurações distintas de h e f : na primeira fase, a taxa f — relevante para a ativação transcricional e, portanto, para a formação da faixa — deve ser maior que h e, na segunda fase, a taxa h — relevante para a repressão transcricional e, portanto, para o desaparecimento da faixa — deve ser maior do que f . Mesmo identificada uma relação qualitativa entre h e f ao longo do processo transcricional, isso ainda não é suficiente para obtermos informações quantitativas de quais seriam as formas funcionais envolvidas. Contudo, uma vez que o processo de formação da faixa está sendo relatado através de uma abordagem “tudo ou nada” (de modo que ocorre uma alternância entre as relevâncias de cada taxa ao longo do processo), podemos propor, então, dependências temporais consideravelmente simples para as taxas h e f , como

$$\begin{aligned} f(t) &= f_1\Theta(t_0 - t) + f_2\Theta(t - t_0) \\ h(t) &= h_1\Theta(t_0 - t) + h_2\Theta(t - t_0), \end{aligned} \quad (4.1)$$

onde $f_1 > f_2$, $h_1 < h_2$, t_0 é o instante em que o efeito regulatório externo é bruscamente alterado e $\Theta(t)$ é a função de Heaviside, definida como

$$\Theta(t) = \begin{cases} 0 & , t \leq 0 \\ 1 & , t > 0 \end{cases}. \quad (4.2)$$

As Eqs. (4.1) mostram justamente que, no início do processo transcricional, as taxas h e f assumem certos valores e, a partir de um instante t_0 (o qual já esperamos que esteja entre as classes temporais T5 e T6), assumem outros valores, conforme mostra a Figura 4.3. Com efeito, a Eq.(4.1) e Figura 4.3 mostram que as taxas $h(t)$ e $f(t)$ são propostas de modo que atuam como um “interruptor”: antes de t_0 , o efeito regulatório total sobre o gene é ativador e, após t_0 , o efeito regulatório total é repressor. Em outras palavras, a Eq.(4.1)

Figura 4.3 – Gráficos das taxas $f(t)$ e $h(t)$ propostas pelas Eqs.(4.1). O instante de tempo t_0 representa o momento em que o efeito regulatório é ligado/desligado, analogamente a um “interruptor”.

e Figura 4.3 mostram que, antes de t_0 , é muito mais provável encontrar configurações de ocupância nos sítios que favorecem a transcrição do que após t_0 , isto é, apesar de haver constantemente ocupações e desocupações de sítios de ligação, as circunstâncias para o início da transcrição são muito mais favoráveis antes de t_0 do que depois de t_0 .

De fato, a forma funcional proposta nas Eqs.(4.1) é motivada, sobretudo, pelo fato de que a primeira fase da produção de mRNA termina exibindo um caráter de estabilização, o que sugere um regime estacionário e que, por sua vez, é previsto para tempos suficientemente grandes pelo modelo estocástico com taxas h e f constantes. No Capítulo 3, vimos que a especificação das taxas h e f determina automaticamente o estado assintótico. Com as taxas especificadas pelas Eqs.(4.1) as especificações são análogas. Na primeira fase, o sistema evolui para o estado assintótico especificado por h_1 e f_1 e, a partir de t_0 , o sistema tem de evoluir para outro estado estacionário, especificado por h_2 e f_2 . A tarefa agora é verificar se as taxas h e f propostas segundo as Eqs.(4.1) são capazes de descrever a dinâmica da expressão de mRNA mostrada pela Figura 4.1. Em outras palavras, estamos interessados em saber se a dinâmica da expressão de mRNA é reproduzida após especificarmos os estados inicial e final do sistema. Em termos do modelo estocástico do gene binário, vamos aplicar o modelo estocástico duas vezes consecutivas: na primeira, considerando h_1 e f_1 e, em seguida, h_2 e f_2 .

Todavia, antes de prosseguir, devemos realizar algumas observações acerca das Eqs.(4.1) e suas implicações.

A primeira observação é com relação à descontinuidade apresentada em t_0 . Naturalmente, a descrição proposta aqui do processo transcricional é fenomenológica e a concepção das Eqs.(4.1) está inserida exatamente neste tipo de abordagem. No Capítulo 2, vimos que uma proteína ligada ao DNA ativa ou reprime a transcrição por meio de diversos mecanismos (interação entre a proteína regulatória e a RNA-polimerase, interação entre a proteína regulatória e Complexo Mediador — que, por sua vez, interage com a RNA-polimerase — e modificações locais na estrutura do DNA alterando a exposição do promotor à RNA-polimerase) e, na abordagem que considera um gene binário regulado por várias proteínas e vários sítios de ligação, conforme discutido no Capítulo 3, as Eqs.(4.1) representam de forma mascarada todos estes mecanismos citados e o parâmetro t_0 representa o instante em que o efeito regulatório total deixa de ser essencialmente ativador e passa a ser essencialmente repressor. Novamente, dada a necessidade de descrever um fenômeno e, sobretudo, de compreender sua essência, devemos utilizar as ferramentas mais simples e que requerem o menor número de hipóteses. Nesse sentido, o uso das funções descontínuas (bem como o de outros constructos matematicamente simples), por mais distante que estejam do que conhecemos sobre as interações moleculares, representa, como veremos adiante, uma ferramenta poderosa para a descrição e compreensão do processo transcricional e seus mecanismos fundamentais.

A segunda observação vem do fato que, se mantidas constantes, as taxas h e f definem probabilidades assintóticas dadas pela Eq.(3.14). No caso em questão, em que propomos o processo ocorrendo em duas fases, cada uma com taxas constantes características, temos a determinação de dois conjuntos de probabilidades assintóticas, um para cada fase do processo:

$$p_{\alpha,1} = \frac{f_1}{h_1 + f_1}, \quad p_{\beta,1} = \frac{h_1}{h_1 + f_1}, \quad p_{\alpha,2} = \frac{f_2}{h_2 + f_2}, \quad p_{\beta,2} = \frac{h_2}{h_2 + f_2}, \quad (4.3)$$

onde os índices 1 e 2 se referem respectivamente à primeira fase, de C13 a t_0 , e à segunda fase, a partir de t_0 . É imediato ver que $p_{\alpha,1} + p_{\beta,1} = 1$ e $p_{\alpha,2} + p_{\beta,2} = 1$. Temos então que, inicialmente, a evolução temporal do sistema é guiada por h_1 e f_1 e, após t_0 , o sistema é guiado pelas novas taxas h_2 e f_2 . No Capítulo 3, onde as taxas h e f eram

constantes, a razão entre a quantidade $h + f$ e a taxa de degradação ρ definia uma constante adimensional, denominada ϵ . Do mesmo modo, as taxas h_1 , f_1 , h_2 e f_2 — constantes por construção — definem em cada fase um parâmetro adimensional:

$$\epsilon_1 = \frac{h_1 + f_1}{\rho}, \quad \epsilon_2 = \frac{h_2 + f_2}{\rho}. \quad (4.4)$$

Uma vez que ϵ_1 e ϵ_2 definem a mesma propriedade — rapidez do chaveamento — em intervalos distintos, é plausível esperar que tal propriedade, caso seja intrínseca do gene, não sofra nenhuma alteração durante o tempo, o que nos permite escrever

$$\epsilon_1 = \epsilon_2 = \epsilon \Rightarrow h_1 + f_1 = h_2 + f_2 = \epsilon\rho = \text{constante}. \quad (4.5)$$

Dessa hipótese seguem duas implicações. A primeira implicação é que estamos resumindo todo o conjunto de interações existentes entre uma proteína regulatória e o sítio (ao qual se liga) em apenas uma única interação: se, em uma dada configuração, uma proteína se liga facilmente a um sítio (mudando de nível de ativação a uma taxa alta) então é mais difícil o seu desligamento (retornando ao nível anterior a uma taxa baixa) e vice-versa, indicando que esta interação única não pode ser forte e fraca ao mesmo tempo. Isso significa que, por estas taxas estarem atreladas ao mesmo processo, não é mais necessário trabalhar com as taxas $h(t)$ e $f(t)$ simultaneamente pois, segundo a Eq.(4.5), a especificação de uma taxa determina a outra. Desse modo, optamos por trabalhar com $f(t)$. A segunda implicação é a que se segue. Por construção, temos que $h_1 < h_2$ e $f_1 > f_2$. Isso significa que, antes de t_0 , transições desligado $\rightarrow$ ligado levam menos tempo do que a transições ligado $\rightarrow$ desligado e, após t_0 , são as transições ligado $\rightarrow$ desligado que ocorrem mais rapidamente. Na abordagem trabalhada, isso se deve às condições geradas por alterações na abundância das proteínas (que, conseqüentemente, alteram as configurações de ocupância dos sítios de ligação). Surge, então, a motivação para propor $f(t)$ em termos das concentrações das proteínas. De fato, investigaremos essa possibilidade no Capítulo 5, utilizando funções de Hill, amplamente utilizadas no contexto de interações entre proteínas e ligantes(72,141).

A última observação resulta das duas observações anteriores. Uma vez que nós consideramos a Eq.(4.3) e assumimos válida a Eq.(4.5), podemos reescrever as probabilidades

assintóticas como

$$p_{\alpha,1} = \frac{f_1}{\epsilon\rho}, \quad p_{\alpha,2} = \frac{f_2}{\epsilon\rho}. \quad (4.6)$$

A menos do mesmo fator multiplicativo ($\epsilon\rho$), as taxas f_1 e f_2 correspondem às próprias probabilidades assintóticas características de cada fase do processo. Como já adiantamos, mais adiante lidaremos com otimização de expressões e será conveniente trabalharmos com as probabilidades assintóticas no lugar das taxas simplesmente porque, ao contrário das taxas, que podem assumir, em princípio, quaisquer valores não-negativos, os valores assumidos pelas probabilidades estão limitados entre 0 e 1. Então, a partir desse momento, com o intuito de reduzir a notação, vamos representar as probabilidades $p_{\alpha,1}$ e $p_{\alpha,2}$ como Δ e δ , respectivamente, e as taxas f_1 e f_2 são, portanto, reescritas como

$$f_1 = \epsilon\rho\Delta, \quad f_2 = \epsilon\rho\delta. \quad (4.7)$$

Além disso, será conveniente escrever as probabilidades assintóticas Δ e δ como valores assumidos por uma função $q_\alpha(t)$, definida como

$$q_\alpha(t) = \Delta\Theta(t_0 - t) + \delta\Theta(t - t_0), \quad (4.8)$$

que corresponde, na verdade, à própria taxa $f(t)$, dada pelas Eqs.(4.1), reescalada por $\epsilon\rho$. Novamente, podemos entender $q_\alpha(t)$ como a probabilidade — gerada, num *ensemble*, pelas configurações de ocupâncias dos sítios — de a maquinaria transcricional ser atraída para o promotor e iniciar a transcrição.

Finalizadas as análises acerca das implicações da Eq.(4.1), podemos prosseguir com sua aplicação ao problema transcricional. Para isso, devemos reconsiderar as Eqs.(3.13) e (3.19), já escritas com a nova notação, para compormos as expressões para a probabilidade $P_\alpha(t)$ e para o valor médio $\langle n \rangle(t)$. Como o processo transcricional se dá em duas fases, cada uma com taxas características conforme as Eqs.(4.1), as expressões para a probabilidade $P_\alpha(t)$ e para o valor médio $\langle n \rangle(t)$ devem naturalmente apresentar duas fases:

$$P_\alpha(t) = \begin{cases} \Delta + (p_\alpha^0 - \Delta)e^{-\epsilon\rho t} & , t \leq t_0 \\ \delta + (p_\alpha^{t_0} - \delta)e^{-\epsilon\rho(t-t_0)} & , t > t_0 \end{cases} \quad (4.9)$$

$$\langle n \rangle (t) = \begin{cases} A_1 + B_1 e^{-\rho t} + C_1 e^{-\epsilon \rho t} & , t \leq t_0 \\ A_2 + B_2 e^{-\rho(t-t_0)} + C_2 e^{-\epsilon \rho(t-t_0)} & , t > t_0 \end{cases}, \quad (4.10)$$

onde

$$\begin{aligned} A_1 &= N(1 - \chi)\Delta + N\chi \\ B_1 &= n_0 - N(1 - \chi)\Delta - N\chi - \frac{N(1 - \chi)(p_\alpha^0 - \Delta)}{1 - \epsilon} = n_0 - A_1 - C_1 \\ C_1 &= \frac{N(1 - \chi)(p_\alpha^0 - \Delta)}{1 - \epsilon} \\ A_2 &= N(1 - \chi)\delta + N\chi \\ B_2 &= n_{t_0} - N(1 - \chi)\delta - N\chi - \frac{N(1 - \chi)(p_\alpha^{t_0} - \delta)}{1 - \epsilon} = n_{t_0} - A_2 - C_2 \\ C_2 &= \frac{N(1 - \chi)(p_\alpha^{t_0} - \delta)}{1 - \epsilon} \end{aligned} \quad (4.11)$$

e $p_\alpha^0 = P_\alpha(0)$, $p_\alpha^{t_0} = P_\alpha(t_0)$, $n_0 = \langle n \rangle(0)$ e $n_{t_0} = \langle n \rangle(t_0)$ são as condições iniciais em cada fase sendo que $\langle n \rangle_0$ é dada pelo valor experimental em C13. Além disso, as condições iniciais na segunda fase da dinâmica, $p_\alpha^{t_0}$ e n_{t_0} , são determinadas pela primeira fase já que $P_\alpha(t)$ e $\langle n \rangle(t)$ são funções contínuas em t_0 , apesar da descontinuidade de $f(t)$ em t_0 . Em outras palavras, o que vemos é que a imposição, em t_0 , de um novo estado assintótico transforma as condições do sistema no instante t_0 em condições iniciais para a subsequente evolução. Assim, p_α^0 é a única condição inicial sobre a qual não sabemos a priori nenhuma informação.

Tendo agora expressões para a probabilidade $P_\alpha(t)$ e, sobretudo, para o valor médio $\langle n \rangle(t)$, podemos então otimizar a Eq.(4.10) a fim de descrever os dados experimentais da Figura 4.1. Ao todo, há oito parâmetros livres, para os quais $\langle n \rangle(t)$ deve ser otimizada tal que os nove pontos experimentais sejam reproduzidos: k , χ , ϵ , ρ , Δ , δ , t_0 e p_α^0 . Antes de partirmos para a otimização da Eq.(4.10), é necessário fazer, baseado na literatura, estimativas das ordens de magnitude esperadas para os valores dos parâmetros envolvidos.

Na literatura, encontramos que o tamanho da sequência codificante de *eve* é de 1477 bp(142) e a taxa à qual a enzima RNA-polimerase II transcreve uma sequência é 1400 bp/min = 23,33 bp/seg(143, 144). Isso nos permite estimar em 63,3 segundos o tempo para que uma fita de mRNA de *eve* seja produzida, o que corresponde a uma taxa k de

$1/(63,3 \text{ seg}) = 1.579 \cdot 10^{-2} \text{seg}^{-1}$ e também é compatível com outros valores (145).

Para estimarmos a taxa de degradação ρ do mRNA de *eve*, baseamos no tempo de meio-vida do mRNA de *ftz*. Esse procedimento é adotado devido à escassez de informações sobre mRNA de *eve*. O fato de os genes *eve* e *ftz* serem genes *pair-rule*, expressos durante a mesma fase embrionária e possuem padrões de expressão similares nos fornece suporte para estimar que os tempos de meia-vida do mRNA de *eve* e de *ftz* são comparáveis: na literatura(146–148), o tempo de meia-vida para o mRNA de *ftz* está entre 6 e 7 minutos, o que corresponde a uma taxa de degradação da ordem de 10^{-3}seg^{-1} e, como dissemos, estabelecerá a ordem de magnitude para a taxa de degradação ρ do mRNA de *eve*.

O parâmetro χ , que quantifica a proporção entre dois modos de produção, pode ser inicialmente estimado como sendo de 0 simplesmente porque estamos descrevendo o fenômeno utilizando um gene com dois níveis de ativação bem caracterizados (um nível com alta taxa de produção de mRNA e outro com nenhuma taxa de produção), o que pode ser referido como uma abordagem “tudo-ou-nada”.

As estimativas para os parâmetros Δ e δ , probabilidades assintóticas em cada fase, também podem ser concebidas com base no caráter “tudo-ou-nada” da abordagem: uma vez que estamos analisando o fenômeno somente em $AP = 41.5$ e que a produção de mRNA nessa região é consideravelmente maior do que nas demais regiões ao longo do eixo AP e lembrando também que esse padrão de expressão do gene *eve* é o que ocorre em embriões do tipo selvagem (*wild type*), então, podemos estimar $\Delta = 1$, ou seja, supomos que a tendência do sistema biológico é a de produzir *sempre* mRNA de *eve* durante as primeiras classes temporais; após o instante t_0 (cuja ordem de magnitude pode ser graficamente estimada em torno de $5 \times 390 \text{ seg} = 1950 \text{ seg}$), podemos supor que o sistema biológico cessa a produção de mRNA de *eve* e, assim, temos a estimativa de $\delta = 0$.

Com relação à condição inicial p_α^0 em $AP = 41.5$, também estamos longe de inferir qualquer estimativa simplesmente por desconhecermos maiores informações. No entanto, sendo nossa abordagem baseada em um processo de Markov, a tendência é que, com o evolução temporal, o sistema deixe de sentir o efeito da condição inicial e, por isso, o desconhecimento de uma estimativa para p_α^0 — ou mesmo para n_0 — não representa um

grande problema. E quanto ao parâmetro ϵ , desconhecemos informações que possam nos ajudar a estimar seu valor.

Conseguimos, então, reunir estimativas para k , χ , ρ , t_0 , Δ e δ e, embora não tenhamos estimativas concretas para p_α^0 e ϵ , já temos um direcionamento melhor para encontrar um conjunto de parâmetros e, conseqüentemente, descrever a dinâmica transicional.

A otimização via método dos mínimos quadrados, através do programa *Wolfram Mathematica 7.0* e utilizando como condição inicial o próprio valor experimental ($n_0 = 10.6892$), fornece $k = 0.0701998$, $\epsilon = 1.43301$, $\rho = 0.00242665$, $t_0 = 2053.09$ e $p_\alpha^0 = 0.0165307$, e a expressão Eq.(4.10) otimizada com tais parâmetros é comparada aos dados experimentais (Figura 4.4(a)). Realizando o mesmo procedimento sem, no entanto, impor um valor específico para n_0 , permitindo que n_0 assumia valores em torno do valor experimental inicial, os valores gerados são $k = 0.0703439$, $\epsilon = 1.43182$, $\rho = 0.00243314$, $t_0 = 2054.58$, $p_\alpha^0 = 0.0199561$ e $n_0 = 10.6212$, e a expressão Eq.(4.10) otimizada é comparada aos dados experimentais (Figura 4.4(b)).

Figura 4.4 – Comparação entre dados experimentais (Figura 4.1) e a expressão otimizada de $\langle n \rangle(t)$, dada pela Eq.(4.10). **(a)** A otimização, fixando o valor experimental inicial como condição inicial ($n_0 = 10.6892$), fornece os seguintes valores para os parâmetros envolvidos: $k = 0.0701998$, $\epsilon = 1.43301$, $\rho = 0.00242665$, $t_0 = 2053.09$ e $p_\alpha^0 = 0.0165307$. **(b)** A otimização permitindo que n_0 flutue em torno do valor experimental inicial, fornece os seguintes valores: $k = 0.0703439$, $\epsilon = 1.43182$, $\rho = 0.00243314$, $t_0 = 2054.58$, $p_\alpha^0 = 0.0199561$ e $n_0 = 10.6212$. As Figuras 4.4(a) e 4.4(b) mostram que o modelo consegue descrever os dados experimentais com boa concordância e que pequenas diferenças nas condições iniciais não afetam sua evolução temporal.

As Figuras 4.4(a) e 4.4(b) não apenas mostram que a abordagem proposta consegue descrever os dados experimentais com boa concordância mas também que é possível encontrar um conjunto de parâmetros (k , χ , ρ , t_0 , Δ , δ , p_α^0 , ϵ) coerentes com o modelo e com os valores experimentais reportados na literatura. As Figuras 4.4(a) e 4.4(b) ainda

mostram que pequenas alterações nas escolhas das condições não levam a consideráveis diferenças entre os valores gerados de cada otimização e entre as funções interpoladas de cada otimização, constatando a natureza markoviana do modelo e sua robustez. *

A otimização também nos permite graficar a probabilidade $P_\alpha(t)$ de o gene estar ativo em $AP = 41.5$, como mostra a Figura 4.5 (que utiliza o segundo conjunto de parâmetros otimizados). Comparando as Figuras 4.5 e 4.4, podemos notar que, como $\epsilon = 1.50404$, $P_\alpha(t)$ se estabiliza um pouco mais rápido que $\langle n \rangle(t)$.

Figura 4.5 – Previsão para a probabilidade $P_\alpha(t)$ em $AP = 41.5$, utilizando a segunda otimização (n_0 flutuando em torno do valor experimental). Com o auxílio das Eqs.(4.9) e (4.10), podemos notar que, como $\epsilon = 1.50404 > 1$, $P_\alpha(t)$ começa a estabilizar um pouco antes de $\langle n \rangle(t)$.

De fato, o bom desempenho da Eq.(4.10) se deve à boa concordância, sobretudo, nos instantes intermediários (grosso modo, nos intervalos $C13 \rightarrow t_0$ e $t_0 \rightarrow T8$) e não nos instantes finais do processo. Isso porque, diferentemente dos estados nos instantes intermediários, os estados inicial e final de cada fase já são especificados: pelas condições iniciais (cujo efeito sobre a dinâmica, como vimos, é atenuado com o passar do tempo) e pelas taxas h e f respectivamente. Em outras palavras, a especificação dos estados inicial e final garante apenas a concordância do modelo e dados experimentais nos instantes iniciais e finais respectivamente, ou seja, ela não garante em nenhum momento a concordância em estados intermediários, o que significa dizer que a Figura 4.5 repre-

*É importante mencionar também que, apesar da boa concordância entre os dados experimentais e Eq.(4.10) otimizada, não há temos garantia de que o conjunto de valores obtidos para k , χ , ρ , ϵ , Δ , δ e t_0 seja o único capaz de otimizar Eq.(4.10) ou, no caso de não ser, que corresponda a um mínimo global no espaço de parâmetros. Isso porque não foi realizada nenhuma análise mais detalhada da quantidade a ser otimizada. Contudo, podemos nos apoiar no fato de que valores fornecidos pela otimização são bastante coerentes com as referências consultadas.

senta, de fato, um teste para a dinâmica proposta pelo modelo mesmo porque os dados experimentais não apresentam nenhum indício suficientemente claro de que a dinâmica do processo transcricional possa ser descrita especificamente em linguagem matemática através de duas exponenciais, uma referente à dinâmica do chaveamento e outra referente à degradação. A boa concordância entre o modelo e os dados experimentais nos permite dizer que o processo markoviano proposto pelas Eqs.(3.1) e (3.2) consegue descrever com considerável êxito a dinâmica transcricional e que as Eqs.(4.1) são capazes de captar a informação biológica por trás dos dados experimentais.

4.2 Dependência espacial em f

Devido ao bom desempenho da Eq.(4.10) em descrever os dados experimentais em AP = 41.5 e sabendo que a expressão de mRNA caracteriza-se espacialmente pela formação de uma faixa, tratamos de implementar uma dependência espacial em Δ e δ , o que, em última instância, significa implementar a dependência espacial nas taxas $h(t)$ e $f(t)$ de acordo com Eq.(4.7). Isso significa que lidaremos com $P_\alpha(x, t)$ e $\langle n \rangle(x, t)$, nos quais a dependência espacial resultará única e exclusivamente da proteína regulatória, cujo efeito será representado, na nova abordagem, por $h(x, t)$ e $f(x, t)$. Contudo, a implementação de dependências espaciais em Δ e δ só será possível depois que realizarmos otimizações para cada uma das dinâmicas transcricionais observadas em cada posição do eixo AP: a partir dos valores fornecidos para Δ e δ ao longo do embrião poderemos interpolar funções que dependerão da posição ao longo do eixo AP.

Todavia, antes de prosseguir, é importante mencionar três pontos importantes.

O primeiro refere-se aos parâmetros inicialmente considerados pelo modelo: k , χ , ρ , h e f . Argumentamos que os parâmetros k , χ e ρ são assumidos independentes de n e t por serem resultado da conjunção de características intrínsecas da maquinaria transcricional, da sequência gênica, da molécula de mRNA e das ribonucleases que catalisam

a degradação de mRNA. E é exatamente pelo mesmo motivo que não esperamos que tais parâmetros eventualmente apresentem uma dependência espacial: entendemos que, ao longo do embrião, as sequências gênicas a serem transcritas são idênticas entre si, assim como são idênticas entre si as moléculas de RNA-polimerase que as transcrevem e as ribonucleases que catalisam a degradação de mRNA. Desse modo, caberia somente às taxas f e h a possibilidade de apresentar uma dependência espacial até porque, como já dissemos, h e f representam o efeito de proteínas regulatórias cujas expressões, por resultarem de outros processos transcricionais, certamente apresentam dependências espaciais.

O segundo ponto — que já foi, de certa forma, anunciado — é o fato de que, do ponto de vista prático, implementar uma dependência espacial nas taxas h e f significa implementar uma dependência espacial em $q_\alpha(t)$, ou seja, trabalhar com

$$q_\alpha(x, t) = \Delta(x)\Theta(t_0(x) - t) + \delta(x)\Theta(t - t_0(x)). \quad (4.12)$$

Isso mostra que as taxas de transição ao longo do embrião continuam sendo vistas em termos de funções de Heaviside, cada uma assumindo suas próprias intensidades ($\Delta(x)$ e $\delta(x)$) e seus próprios tempos ($t_0(x)$) nos quais ocorrem as mudanças bruscas.

O terceiro ponto é que, ao pretendermos inserir uma dependência espacial somente nas taxas $h(t)$ e $f(t)$ e continuar trabalhando a Eq. (4.10), estamos implicitamente desprezando quaisquer efeitos de difusão e assumindo que os núcleos são independentes, o que significa que o processo transcricional que ocorre em um núcleo não afeta o processo transcricional em núcleos vizinhos.

Terminadas as análises, podemos começar a implementar a dependência espacial no problema. A meta é obter expressões que descrevam, ao longo do eixo AP e em qualquer instante, a probabilidade de o gene estar ativo e para o número de moléculas de mRNA, ou seja, objetivamos formas generalizadas das Eqs.(4.9) e (4.10)

$$P_\alpha(x, t) = \begin{cases} \Delta(x) + [p_\alpha^0(x) - \Delta(x)]e^{-\epsilon\rho t} & , t \leq t_0(x) \\ \delta(x) + [p_\alpha^{t_0}(x) - \delta(x)]e^{-\epsilon\rho(t-t_0(x))} & , t > t_0(x) \end{cases} \quad (4.13)$$

$$\langle n \rangle(x, t) = \begin{cases} A_1(x) + B_1(x)e^{-\rho t} + C_1(x)e^{-\epsilon \rho t} & , t \leq t_0(x) \\ A_2(x) + B_2(x)e^{-\rho(t-t_0)} + C_2(x)e^{-\epsilon \rho(t-t_0(x))} & , t > t_0(x) \end{cases}, \quad (4.14)$$

onde

$$\begin{aligned} A_1(x) &= N(1 - \chi)\Delta(x) + N\chi \\ B_1(x) &= n_0(x) - A_1(x) - C_1(x) \\ C_1(x) &= \frac{N(1 - \chi)[p_\alpha^0(x) - \Delta(x)]}{1 - \epsilon} \\ A_2(x) &= N(1 - \chi)\delta(x) + N\chi \\ B_2(x) &= n_{t_0}(x) - A_2(x) - C_2(x) \\ C_2(x) &= \frac{N(1 - \chi)[p_\alpha^{t_0}(x) - \delta(x)]}{1 - \epsilon}. \end{aligned} \quad (4.15)$$

Para isso, testaremos em $q_\alpha(x, t)$ dependências espaciais que serão gradualmente refinadas: primeiramente, as dependências espaciais em $\Delta(x)$ e $\delta(x)$ serão propostas de modo que a função $q_\alpha(x, t)$ assume valores não-nulos somente em uma limitada região em torno de $AP = 41.5$ e valores nulos nas demais regiões do embrião, o que tridimensionalmente se assemelha a uma “caixa”; depois, a dependência espacial em $\Delta(x)$ e $\delta(x)$ adquire um pouco mais de detalhamento e, por fim, uma dependência espacial mais sofisticada é proposta e o resultado otimizado de sua inserção na Eq.(4.10) é comparado com os dados experimentais.

Contudo, antes de partirmos para a implementação em $q_\alpha(t)$, vamos analisar dependências espaciais nas condições iniciais n_0 e p_α^0 .

4.2.1 Condições iniciais

Como nosso intuito é realizar otimizações para todas as posições ao longo do eixo AP a fim de descrevermos espacial e temporalmente a transcrição de mRNA de *eve* e como vimos que o modelo não é afetado por pequenas diferenças nas condições iniciais, podemos escrever as condições iniciais $n_0(x)$ e $p_\alpha^0(x)$ ao longo do eixo AP utilizando funções

interpoladoras pois os erros associados (no caso de $n_0(x)$, em relação aos valores experimentais) não comprometerão nossa análise e, além disso, fórmulas fechadas para $n_0(x)$ e $p_\alpha^0(x)$ ajudarão nos passos seguintes da implementação, economizando termos e notação.

A primeira condição inicial que analisaremos é o número inicial de mRNA, n_0 . Como o número de moléculas de mRNA ao longo do eixo AP em C13 pode ser aproximadamente descrito por um decaimento exponencial, podemos, então, escrever como condição inicial a função $n_0(x)$, dada por

$$n_0(x) = a_0 + a_1 \exp[-a_2(x - 35.5)] \quad (4.16)$$

onde $a_0 = 1.36468$, $a_1 = 12.3591$ e $a_2 = 0.0575796$. A Figura 4.6 compara os dados experimentais de mRNA ao longo do eixo AP em C13 e a função $n_0(x)$ que os interpola. A

Figura 4.6 – Condição inicial $n_0(x)$ comparada com o número de moléculas de mRNA em C13.

fórmula $n_0(x)$ para os dados experimentais em C13 representa uma considerável economia de parâmetros pois as dezenas de condições iniciais podem ser aproximadamente descritos por apenas 3 parâmetros: a_0 , a_1 e a_2 .

Como dissemos anteriormente, a condição inicial p_α^0 é a única da qual não temos nenhuma informação e, estendendo esse cenário a todas as 57 regiões ao longo do eixo AP, podemos esperar que 57 valores serão necessários para descrever as condições iniciais ao longo do eixo AP. Sendo assim, seria muito oportuno poder implementar uma dependência espacial, constituindo uma expressão de $p_\alpha^0(x)$ e que tal implementação fizesse uso de um número de poucos parâmetros. De fato, embora desconheçamos quaisquer possíveis valores para p_α^0 ao longo do eixo AP, isso não nos impede de estimá-los. Para isso, faremos uso

naturalmente das informações biológicas às quais temos acesso: o número de moléculas de mRNA em C13 (ou $n_0(x)$) e os gradientes de concentração de proteínas Bcd e Hb, conhecidas como ativadoras para a formação da faixa 2.

A Figura 4.7(a) compara número de moléculas de mRNA (representado pelos dados experimentais e por $n_0(x)$) com os perfis de intensidade de fluorescência relativos à presença de Bcd e Hb (ambos divididos por um fator 5) em C13. Vemos que o perfil do número de moléculas ao longo do eixo AP acompanha os perfis dos gradientes das proteínas ativadoras, principalmente o gradiente de Bcd. Assim, somos imediatamente levados a inferir que o perfil do número de moléculas de mRNA corresponde — a menos de um fator multiplicativo ou de uma constante aditiva — ao próprio perfil de Bcd ou a uma combinação dos perfis de Bcd e Hb (para a qual Bcd contribuiria, em princípio, mais do que Hb). Em termos de regulação, a Figura 4.7(a) pode ser entendida de maneira simples: quanto maior a concentração de proteínas ativadoras, maior é a probabilidade de que as proteínas ativadoras se liguem a seus respectivos sítios de ligação e, portanto, maior será a probabilidade de que o gene esteja ativo e, conseqüentemente, maior será o número de moléculas transcritas. Matematicamente, isso está bastante próximo do que representa o termo n_S na Eq. (3.19), o que significa considerar a hipótese de que, em C13, o sistema se encontra em um estado estacionário ou em um estado suficientemente próximo de um estado estacionário de modo que podemos, então, escrever

$$n_0 = N(1 - \chi)p_\alpha^0 + N\chi \Rightarrow p_\alpha^0(x) = \frac{n_0(x) - N\chi}{N(1 - \chi)}, \quad (4.17)$$

onde $p_\alpha^0(x)$ possui uma dependência espacial adquirida automaticamente de $n_0(x)$.

De fato, essa hipótese é suportada por algumas razões além da semelhança entre os perfis de Bcd e do número de moléculas de mRNA. A primeira delas é o fato de que o gradiente de Bcd mantém-se praticamente inalterado desde o ciclo 11 até os últimos 12-18 minutos antes da gastrulação(110). Os gradientes de Cad (suposta ativadora, segundo (32)), Gt e Tll também não apresentam consideráveis mudanças durante os ciclos 11 e 12. Diferentemente dos perfis dos gradientes das proteínas mencionadas, os perfis de Kr e Kni apresentam consideráveis aumentos nos níveis de expressão entre os ciclos 12 e 13

Figura 4.7 – Condições iniciais do problema. **4.7(a)** Número de moléculas de mRNA e a função otimizada n_0 comparados aos perfis dos gradientes das proteínas ativadoras Bcd e Hb (Para efeito de comparação, os gradientes de intensidade de fluorescência de Bcd e Hb foram divididos por um fator 5). Podemos notar uma considerável semelhança entre os perfis exibidos pelo número de moléculas e pelos gradientes de proteínas regulatórias, principalmente Bcd. **4.7(b)** Supondo que, em C13, o sistema esteja em um estado estacionário, a probabilidade $P_\alpha(x, t = C13) = p_\alpha^0(x)$ pode ser obtida a partir de n_0 , conforme mostra a Eq.(4.17).

pórem, tais aumentos ocorrem nas regiões em torno de $AP = 50\%$ e de $AP = 70\%$ (133), respectivamente, onde a produção de mRNA já é menos intensa do que em $AP = 41.5$, diminuindo ainda mais nos instantes subsequentes e permanecendo estabilizada durante a maior parte do processo, conforme mostra a Figura 2.16. Por outro lado, há também pontos desfavoráveis a essa hipótese de estado inicial estacionário. Um deles é o nosso desconhecimento sobre como evoluiu o perfil de mRNA desde, por exemplo, o ciclo 11 até C13 (se a expressão se mantém relativamente constante ou sofre alterações consideráveis) e sobre isso não temos as mínimas condições de argumentar. Outro ponto é o fato de que o perfil de Hb, assim como os de Kr e Kni, apresenta um aumento considerável no nível de expressão porém, como tal aumento ocorre na região anterior ($0 < AP < 50\%$)(133), isso pode não garantir que o sistema se encontre em C13 em um regime estacionário. Contudo, podemos ainda supor que tal aumento na concentração de Hb não seja suficiente para alterar significativamente a probabilidade $p_\alpha^0(x)$ de modo que a condição estacionária possa ainda ser aceita.

De modo geral, as informações apresentadas nos permitem considerar o estado inicial do sistema em C13 como um estado estacionário e mesmo que tal consideração se revele falha, podemos ainda assim conceber o estado inicial como próximo de um estado estacionário e utilizar a Eq. (4.17) como forma de aproximação. Ainda assim, mesmo que a condição inicial seja completamente diferente da que estamos propondo, nós sabemos que

o modelo não é sensível a alterações nas condições iniciais devido à sua natureza markoviana. Por fim, é interessante ressaltar que essas considerações feitas acerca das condições iniciais reforçam que o modelo oferece uma flexibilidade em relação às condições iniciais e mostram que, do ponto de vista analítico, a Eq.(4.17) representa um grande ferramenta porque não apenas permite escrever $p_\alpha^0(x)$ em uma fórmula compacta mas também evita a inclusão de mais parâmetros.

4.2.2 Dependência espacial: tipo “caixa”

A primeira dependência espacial proposta à $q_\alpha(x, t)$ é dada por

$$q_\alpha(x, t) = \begin{cases} \Delta(x) & , t \leq t_0 \\ \delta(x) & , t > t_0 \end{cases} \quad (4.18)$$

onde $t_0 = 2053.09$ e

$$\Delta(x) = \begin{cases} 0 & , 35.5 \geq x < 38.5 \\ 1 & , 38.5 \geq x \leq 43.5 , \quad \delta(x) = 0. \\ 0 & , 43.5 < x \end{cases} \quad (4.19)$$

A função $q_\alpha(x, t)$ é escolhida dessa forma baseado na desconstrução dos dados experimentais, que mostram que, grosso modo, a produção de mRNA se dá em uma região limitada do eixo AP. Os valores assumidos por $\Delta(x)$, $\delta(x)$ e t_0 são os valores gerados pela otimização na análise em AP = 41.5% (considerando como condição inicial n_0 o valor experimental em C13). O intuito é verificar se a ideia aplicada em AP = 41.5% é capaz de produzir, mesmo de forma grosseira, um padrão espacial de expressão. Do ponto de vista biológico, nosso intuito é verificar se um efeito de ativação de duração finita e espacialmente limitado é capaz de gerar a formação e destruição da faixa. O perfil de $q_\alpha(x, t)$ assemelha-se a uma “caixa”, representando o efeito do agente regulatório externo limitado a uma região do embrião e com uma certa duração, conforme mostra a Figura 4.8. Importante notar nas

Figura 4.8 – (a) Gráfico de $q_\alpha(x,t)$, dado pelas Eqs.(4.18) e (4.23). (b) Gráfico de $q_\alpha(x,t)$ e os perfis assumidos por $q_\alpha(x,t)$ ao longo do eixo AP em cada classe temporal (curvas vermelhas). (c) Gráfico de $P_\alpha(x,t)$, resultante da inserção de $q_\alpha(x,t)$. (d) Gráfico de $P_\alpha(x,t)$ e os perfis assumidos por $P_\alpha(x,t)$ ao longo do eixo AP em cada classe temporal (curvas vermelhas).

Figuras 4.8(c) e 4.8(d) que, em T5, classe temporal em que a faixa de mRNA registra sua máxima intensidade, a probabilidade $P_\alpha(x,t)$ praticamente atinge os valores assintóticos possíveis ao longo do embrião: 1 no domínio que compreende o intervalo entre $AP = 38.5\%$ e $AP = 43.5\%$ e 0 no restante do eixo AP[†]. Posteriormente, a tendência é que, após T8, $P_\alpha(x,t)$ volte a atingir um valor assintótico ao longo de todo o eixo AP e, dessa vez, tal valor será a nulidade. Como a dependência espacial de $\Delta(x)$ é pouco sofisticada, $P_\alpha(x,t)$ evolui de maneira diferente para cada domínio mas para pontos de dentro de um

[†]Em termos da taxa de variação temporal de $P_\alpha(x,t)$, vemos que, no intervalo entre $AP = 38.5\%$ e $AP = 43.5\%$, $|\partial P_\alpha(x,t)/\partial t_{t=T5}|$ é da ordem de 10^{-6} e, no restante do eixo AP, $|\partial P_\alpha(x,t)/\partial t_{t=T5}|$ é da ordem de 10^{-6} , podendo chegar a 10^{-7} .

mesmo domínio a evolução temporal ocorre de maneira idêntica (a menos das diferenças devido às condições iniciais). Sendo assim, é esperado que, dentro de cada domínio, o número de moléculas tenda a ser o mesmo. As Figuras 4.8(c) e 4.8(d) sugerem, então, que núcleos imediatamente vizinhos mas pertencentes a domínios distintos apresentem, em T5, valores assintóticos para $P_\alpha(x, t)$ distantes, caracterizando uma descontinuidade espacial.

Com as condições iniciais e as funções $q_\alpha(x, t)$ e $P_\alpha(x, t)$ especificadas, podemos obter o número médio de moléculas de mRNA. Contudo, resta ainda especificar os parâmetros k , χ , ϵ e ρ que serão utilizados. Como as condições iniciais $n_0(x)$ e $p_\alpha^0(x)$ estão, de certa forma, “fixadas” (pelas Eqs. (4.16) e (4.17) respectivamente), tomamos $k = 0.0701998$, $\chi = 0$, $\epsilon = 1.43301$, $\rho = 0.00242665$, que são os valores utilizados para a descrição do processo em AP = 41.5 considerando o dado experimental em C13 como a condição inicial[‡]. O número de moléculas pode, então, ser obtido através da Eq.(4.14) e podemos então compará-lo aos dados experimentais, como mostra a Figura 4.9.

A Figura 4.9 mostra que a ideia de um efeito de ativação de duração finita e espacialmente limitado permite o aparecimento de uma região de maior produção de moléculas de mRNA. O bom desempenho qualitativo nos mostra que, assim como ocorreu na análise em AP = 41.5%, a dinâmica prevista por $\langle n \rangle(x, t)$ “acompanha” a dinâmica dos dados experimentais, ou seja, a previsão de produção de moléculas na região compreendida entre AP = 38.5% e AP = 43.5% acontece nos mesmos ritmo e intensidade que os apresentados pelos dados experimentais. Isso significa que a hipótese de que os processos bioquímicos associados a k , χ , ϵ e ρ independem do eixo AP mostrou-se válida, constatando, portanto, que a dinâmica da formação de mRNA no eixo AP é governada unicamente pela dinâmica do chaveamento e pela degradação. Com relação ao desempenho quantitativo, é imediato ver que as dependências espaciais propostas para t_0 , $\delta(x)$ e, sobretudo, $\Delta(x)$ em Eq.(4.18) carecem de informações mais detalhadas a fim de descrever com maior exatidão a formação das bordas da faixa. Na Figura 4.9, isso pode ser percebido pelo estreitamento

[‡]Como vimos nas Figuras 4.4(a) e 4.4(b), as equações são robustas com relação a pequenas mudanças nas condições iniciais e nos parâmetros de modo que poderíamos utilizar o outro conjunto de parâmetros — obtido deixando a condição inicial n_0 flutuar em torno do valor experimental em C13 — e o resultado seria bastante similar.

Figura 4.9 – Previsão do número médio de moléculas de mRNA a partir da função $q_\alpha(x, t)$ especificada pelas Eqs.(4.18) e (4.23), que determinam um efeito de ativação de duração finita e limitado espacialmente. Os parâmetros utilizados foram $k = 0.0701998$, $\chi = 0$, $\epsilon = 1.43301$ e $\rho = 0.00242665$. Com essas especificações, é possível que uma região do eixo AP apresente maior atividade transcricional e que a formação da “faixa” acompanhe a dinâmica dos dados experimentais. Entretanto, a função $q_\alpha(x, t)$ carece de informações para descrever com mais exatidão a formação das bordas.

sofrido pela faixa e que não é acompanhado pela função $\langle n \rangle(x, t)$.

É imediato notar nas Figuras 4.8 e 4.9 que as análises são restritas até $AP = 76.5\%$ sendo que há a formação de um pequeno pico de expressão em torno de $AP = 80.5\%$, segundo a Figura 2.16. Restringimos nossa análise até $AP = 76.5\%$ simplesmente porque nossa atenção está voltada para a formação da faixa 2. Desse modo, é natural que, assim que tivermos em mãos informações mais detalhadas sobre a formação da faixa 2, nós as estenderemos para analisar a expressão em torno de $AP = 80.5\%$. Na próxima seção, trataremos de refinar a dependência espacial $q_\alpha(x, t)$ buscando expressões mais sofisticadas mas ainda assim permaneceremos restritos a analisar o processo transcricional somente até $AP = 76.5\%$.

4.2.3 Refinando dependências espaciais no modelo: encontrando uma solução

Na seção anterior, vimos que o valor médio $\langle n \rangle(x, t)$ implementado com a função $q_\alpha(x, t)$ dada pela Eq. (4.18) foi capaz de descrever o aparecimento de um pico de expressão e de reproduzir a dinâmica dos dados experimentais porém, vimos também que tal descrição carece de informação necessária para descrever com maior exatidão a intensidade da faixa e a formação de suas bordas.

Para descrever a formação das bordas, faremos duas alterações: a primeira, já esperada, é reescrever $\Delta(x)$ e a segunda é propor uma dependência espacial em t_0 . Tais alterações são baseadas unicamente no desempenho mostrado na Figura 4.9. Nos dados experimentais, vemos que, o número de moléculas não atinge os mesmos valores dentro do domínio definido pela faixa, ou seja, há uma ativação transcricional diferente em cada posição. Ao mesmo tempo, vemos que, em regiões distantes de $AP = 41.5\%$, o número de moléculas tende a zero mais rapidamente do que no entorno de $AP = 41.5\%$. Isso nada mais é do que a dinâmica de formação das bordas da faixa. Essas diferentes “velocidades” para diferentes posições no eixo AP não devem resultar de diferentes valores para ϵ e ρ (pois, como já argumentamos, esses parâmetros representam características intrínsecas de certas moléculas e processos bioquímicos) mas sim do agente regulatório externo. Portanto, além de uma dependência espacial mais refinada em $\Delta(x)$, é necessário também implementar uma dependência espacial em t_0 .

As dependências propostas para $\Delta(x)$ e $t_0(x)$ são

$$\Delta(x) = \exp[-0.04(x - 41)^2] \quad (4.20)$$

$$t_0(x) = 2053.09 \exp[-0.04(x - 41)^2]. \quad (4.21)$$

A dependência espacial proposta para $\Delta(x)$ é baseada unicamente no perfil da expressão assumido entre no intervalo de T3 a T5, sugerindo a tendência para um regime estacionário, enquanto que a dependência proposta para $t_0(x)$ é baseada no processo de afile-

mento sofrido pela perfil da expressão. Vale mencionar que os valores das amplitudes das gaussianas em $\Delta(x)$ e $t_0(x)$ são oriundos dos resultados obtidos na análise feita em AP = 41.5. Assim, temos uma expressão reformulada de $q_\alpha(x, t)$:

$$q_\alpha(x, t) = \begin{cases} \Delta(x) & , t \leq t_0(x) \\ \delta(x) & , t > t_0(x) \end{cases} \quad (4.22)$$

onde

$$\Delta(x) = \exp[-0.04(x - 41)^2] \quad (4.23)$$

$$\delta(x) = 0$$

$$t_0(x) = 2053.09 \exp[-0.04(x - 41)^2].$$

Tendo em mãos a função $q_\alpha(x, t)$ e tomando os parâmetros $k = 0.0701998$, $\chi = 0$, $\epsilon = 1.43301$ e $\rho = 0.00242665$ e as condições iniciais $p_\alpha^0(x)$ e $n_0(x)$, podemos obter a probabilidade de o gene estar ligado, $P_\alpha(x, t)$, e o número de moléculas de mRNA, $\langle n \rangle(x, t)$, através das Eqs.(4.13) e (4.14) como mostram as Figuras 4.10 e 4.11 respectivamente.

A Figura 4.10(a) mostra que a função $q_\alpha(x, t)$ continua limitada a uma região do eixo AP, porém, sem descontinuidade na periferia da faixa, o que, consequentemente, é herdado pela probabilidade $P_\alpha(x, t)$, como mostrado na Figura 4.10(c), e pelo valor médio $\langle n \rangle(x, t)$, mostrado na Figura 4.11. Nas Figuras 4.10(b) e 4.10(d), estão graficadas as funções $q_\alpha(x, t)$ e $P_\alpha(x, t)$ (superfícies azuis) com seus respectivos perfis assumidos em cada classe temporal (curvas vermelhas) e com $t_0(x)$ (curva azul), indicando o instante a partir do qual o caráter regulatório é bruscamente alterado.

É importante mencionar que, enquanto a Figura 4.8 mostrava, em T5, dois domínios onde $P_\alpha(x, t)$ praticamente atingia valores assintóticos e tais domínios eram separados por descontinuidades, a Figura 4.10(a) mostra também dois domínios separados não mais por descontinuidades mas por uma “zona intermediária”. Tais domínios, onde $P_\alpha(x, T5)$ atinge valores praticamente assintóticos, correspondem aos intervalos $40.5\% < AP < 41.5\%$ e $AP > 48.5\%$. A zona intermediária compreende os intervalos $35.5\% < AP < 40.5\%$ e $41.5\% < AP < 48.5\%$ e nessa região $P_\alpha(x, T5)$ não se encontra tão próximo de

Figura 4.10 – (a) Gráfico de $q_\alpha(x,t)$, dado pelas Eqs.(4.18), (4.20) e (4.21). (b) Gráfico de $q_\alpha(x,t)$, os perfis assumidos por $q_\alpha(x,t)$ ao longo do eixo AP em cada classe temporal (curvas vermelhas) e $t_0(x)$ (curva azul). (c) Gráfico de $P_\alpha(x,t)$, resultante da inserção de $q_\alpha(x,t)$. (d) Gráfico de $P_\alpha(x,t)$, os perfis assumidos por $P_\alpha(x,t)$ ao longo do eixo AP em cada classe temporal (curvas vermelhas) e $t_0(x)$ (curva azul).

seus respectivos valores assintóticos, como mostra a Figura 4.10[§]. Isso pode ser explicado comparando T5 (= 1950 segundos) com $t_0(x)$ (Eq.(4.21)): para $AP > 48.5\%$, o valor de $t_0(x)$ é muito menor que até T5 e $P_\alpha(x,t)$ tem tempo suficiente para evoluir e se aproximar bastante de um valor assintótico; para $40.5\% < AP < 41.5\%$, $t_0(x)$ é maior que T5 de modo que o gene está ligado há um tempo suficiente de modo que, em T5, $P_\alpha(x,t)$ esteja muito próximo de valores assintóticos; há muito tempo e, portanto, o valor de $P_\alpha(x,T5)$ corresponde ao valor assintótico (ou a um valor próximo deste); nos intervalos $35.5\% \leq$

[§]Isso pode ser constatado pela taxa de variação temporal de $P_\alpha(x,t)$ em T5. Nos intervalos $40.5\% \leq AP \leq 41.5\%$ e $AP > 48.5\%$, $|\partial P_\alpha / \partial t|_{t=T5}$ é da ordem de 10^{-6} (podendo chegar a 10^{-7}) e, nos intervalos $35.5\% \leq AP < 40.5\%$ e $41.5\% < AP < 49.5\%$, $|\partial P_\alpha / \partial t|_{t=T5}$ chega a atingir valores da ordem de 10^{-3} .

$AP < 40.5\%$ e $41.5\% < AP < 49.5\%$, $t_0(x)$ é pouco menor que T5, o que implica que, em T5, o sistema está recém-desligado de modo que $P_\alpha(x, t)$, em comparação com o que ocorre nos intervalos, está menos próximo de um valor assintótico.

Figura 4.11 – Gráfico do número médio de moléculas de mRNA em função do tempo e da posição sobre o eixo AP, segundo a Eq.(4.14), considerando a função $q_\alpha(x, t)$ especificada pela Eq.(4.22). Utilizando os mesmos parâmetros k , χ , ϵ e ρ ao longo do eixo AP ($k = 0.0701998$, $\chi = 0$, $\epsilon = 1.43301$ e $\rho = 0.00242665$) e os perfis gaussianos para as dependências espaciais de $\Delta(x)$ e $t_0(x)$, dadas pelas Eqs.(4.20) e (4.21) respectivamente, o modelo descreve com grande concordância os dados experimentais. Comparando com a análise anterior, vemos que uma simples sofisticação nas dependências espaciais de $\Delta(x)$ e $t_0(x)$ foi suficiente para reproduzir com grande concordância os dados experimentais e descrever de maneira clara o perfil da faixa e seu processo de afilamento. O modelo estocástico desenvolvido e introdução de uma função $q_\alpha(x, t)$ dada pelas Eqs.(4.22) e (4.23) mostram ser capazes de descrever o processo transcricional, constituindo, portanto, dentro do caráter de nossa abordagem, uma solução capaz de fornecer uma descrição fenomenológica semi-quantitativa para ser capazes de descrever o fenômeno transcricional. A sofisticação das dependências espaciais de $\Delta(x)$ e $t_0(x)$ se faz necessária apenas para garantir maior exatidão à descrição e não fornecerá nenhuma informação substancial do fenômeno uma vez que a essência do processo já foi identificada.

A Figura 4.11 compara o valor médio $\langle n \rangle(x, t)$, descrito pela Eq.(4.14), considerando $q_\alpha(x, t)$ como descrito pela Eq.(4.22). Em termos da dinâmica, não há diferença uma vez que não alteramos os parâmetros ρ e ϵ . O desempenho aqui é obviamente melhor do que o desempenho mostrado na Figura 4.9, no qual propusemos $q_\alpha(x, t)$ como uma

“caixa”. Naturalmente, tal melhoria se deve às alterações na formulação de $q_\alpha(x, t)$, mais especificamente, às reformulações de $\Delta(x)$ e de $t_0(x)$. Os perfis gaussianos de $\Delta(x)$ e de $t_0(x)$ permitem uma maior concordância com entre os perfis gerados pelo valor médio $\langle n \rangle(x, t)$ e os apresentados pelos dados experimentais: enquanto $\Delta(x)$ garante a forma do perfil, $t_0(x)$ garante seu afilamento, de modo que a formação das bordas e o afilamento da faixa sejam descritos com maior exatidão. Isso corrobora a constatação de que a especificação das taxas h e f determina o estado para o qual o sistema evolui. Vale ressaltar que, mesmo sendo um teste, a dependência espacial em $\delta(x)$ (proposta na Eq.(4.23)) não comprometeu o desempenho do modelo, isto é, não houve necessidade de sofisticar tal dependência, o que representa um ponto forte de nossa abordagem.

Um ponto que merece total consideração é o número de parâmetros. A presente análise mostra que são necessários um total de 14 parâmetros — 4 parâmetros intrínsecos (k , χ , ρ , ϵ), 3 parâmetros para compor $\Delta(x)$, 3 para compor $t_0(x)$, 1 para compor $\delta(x)$ e mais 3 para descrever as condições iniciais $n_0(x)$ e $p_\alpha^0(x)$ — para descrever um conjunto de 378 ($= 42 \times 9$) dados experimentais — uma vez que foi analisado a expressão no intervalo de AP = 35.5 a AP = 76.5 e que a dinâmica em cada uma dessas 42 posições é representada por 9 dados experimentais.

Apesar da grande concordância entre modelo e dados experimentais, há um aspecto conceitual que tem de ser destacado. O formato gaussiano de $\Delta(x)$, responsável por conferir ao valor médio $\langle n \rangle(x, t)$ o perfil de faixa, foi proposto baseado (i) no fato de que a especificação das taxas de chaveamento h e f determina o estado assintótico para o qual o sistema evolui. Nesse sentido, o fato de que o perfil da expressão pouco se altera no intervalo de T3 a T5 (especialmente no região onde é formada a faixa) estaria de acordo com a hipótese de que o sistema se encontraria em um estado estacionário ou próxima de um estado estacionário no referido intervalo. Em AP = 41.5%, o cenário vislumbrado pela hipótese parece estar de acordo com os resultados mostrados pelas Figuras 4.4 e 4.5: a partir desses resultados podemos verificar que as taxas de variação temporal de $P_\alpha(x, t)$ e $\langle n \rangle(x, t)$ em T5 são da ordem de 10^{-6} e 10^{-3} respectivamente, o que sugere uma configuração relativamente próxima do regime estacionário.

Entretanto, embora a hipótese original pareça válida, estendê-la para todo o eixo AP requer uma atenção especial. De fato, vimos anteriormente que, em T5, $|\partial P_\alpha(x, t)/\partial t|$ atingem desde valores da ordem de 10^{-7} a valores da ordem de 10^{-3} . E com a Figura 4.11, verificamos que $|\partial \langle n \rangle(x, t)/\partial t|$ assume ao longo do eixo AP desde valores da ordem de 10^{-4} até valores da ordem de 10^{-2} . Ainda que sejam valores consideravelmente menores que a unidade e, por isso, sejam evidências de um regime assintótico, vemos que as taxas de variação temporal de probabilidade $P_\alpha(x, t)$ em posições diferentes sobre eixo AP chegam a diferir em até 10^4 vezes e, no caso $\langle n \rangle(x, t)$, as taxas de variação chegam a diferir em até 10^2 vezes. Desse modo, imediatamente somos levados a questionar o quão próximos de regimes estacionários os sistemas (gene + proteínas regulatórias + maquinaria transcricional) ao longo do eixo AP se encontram, sobretudo, em $AP = 41.5\%$. Contudo, apenas comparando as taxas não nos permitirá verificar tal proximidade. Essa verificação deve ser realizada através da evolução temporal da distribuição de probabilidade total, ϕ_n , da qual podemos obter uma informação quantitativamente mais clara e objetiva da dinâmica do processo. Dessa forma, vemos que, pelo fato da hipótese de regime estacionário em T5 não estar suficientemente constatada, a dependência espacial de $\Delta(x)$ deve ser repensada: ou concebida a partir de outro embasamento conceitual ou derivada de otimizações. Nós adotamos a segunda alternativa.

De fato, a função $q_\alpha(x, t)$ ou, mais precisamente, $\Delta(x)$ e $t_0(x)$, necessita de reparos. Apesar de as Eqs.(4.20) e (4.21) terem proporcionado uma boa concordância com os dados experimentais, é visível que o perfil apresentado por $\langle n \rangle(x, t)$ ainda não descreve o processo afilamento da expressão de forma precisa. Isso significa que não apenas $\Delta(x)$ requer uma reformulação na dependência espacial mas também $t_0(x)$.

Contudo, ainda que estejamos em busca de uma descrição mais exata do número médio de moléculas de mRNA e que tal tarefa passe a requerer otimizações, é imprescindível ressaltar que todo o problema de descrição dos dados experimentais foi reduzido a um processo de otimização, ou seja, faremos otimizações apenas para melhorar a precisão do nosso modelo e não para obtermos informações sobre a essência biológica do processo. Em outras palavras, o modelo estocástico desenvolvido e introdução de uma função $q_\alpha(x, t)$,

dada pelas Eqs.(4.22) e (4.23), mostram ser capazes de descrever o processo transcricional, o que, dentro do caráter de nossa abordagem, faz com que esse constructo se estabeleça como uma solução capaz de fornecer uma descrição fenomenológica semi-quantitativa do processo transcricional.

De fato, após realizarmos as análises **(i)** em $AP = 41.5\%$, **(ii)** ao longo do eixo AP utilizando uma função $q_\alpha(x, t)$ do tipo “caixa” e **(iii)** ao longo do eixo AP utilizando a função $q_\alpha(x, t)$ como uma gaussiana sofrendo um afilamento, podemos afirmar que a essência do processo transcricional foi captada com a suposição de que a regulação externa age como um “interruptor” que desliga a ativação transcricional a partir de um certo instante, segundo as Eqs.(4.1), e com a suposição de que ϵ é um parâmetro intrínseco do gene, segundo a Eq.(4.5). As duas ideias inseridas no modelo do gene binário nos permitiram descrever o valor médio com grande exatidão em $AP = 41.5\%$. A partir desse momento, com a dinâmica experimental descrita, avaliamos que o foco das análises passou a ser, substancialmente, a forma da faixa mas não a essência de sua dinâmica de formação. A análise (ii) utilizando uma função $q_\alpha(x, t)$ do tipo “caixa” foi importante para verificarmos se o modelo com efeito regulatório do tipo “interruptor” poderia ser estendido para todo o eixo AP e se dependências espaciais adicionais (como efeitos difusivos) seriam necessários para tal consideração. Ao fim dessa análise, constatamos que a dinâmica dos dados experimentais foi reproduzida e que era necessária uma reformulação na dependência espacial da função $q_\alpha(x, t)$ para descrever com exatidão o processo de afilamento da faixa. Por fim, na análise (iii), verificamos que uma simples sofisticação nas dependências espaciais de $q_\alpha(x, t)$ foi suficiente para descrever com razoável exatidão a formação da faixa.

Embora a essência biológica do processo tenha sido praticamente capturada na análise em $AP = 41.5\%$, as análises subsequentes foram de extrema importância não apenas para corroborar as principais hipóteses presentes na primeira análise mas também para dar suporte a suposições como a independência dos núcleos (o que nos permitiu desconsiderar efeitos difusivos e de transporte) e o caráter intrínseco dos parâmetros k , χ , ρ e ϵ . Além disso, vemos que a expressão da faixa 2 de mRNA de *eve*, regulada por sete proteínas regulatórias e 34 sítios de ligação, pode ser descrita por um gene binário cujas taxas de

chaveamento atuam como “interruptores”.

Embora as otimizações, tratadas na próxima seção, visando o aumento de precisão na descrição configurem uma tarefa secundária (frente ao objetivo primário, que era descrever espacial e temporalmente a expressão de mRNA), isso não priva seus resultados de importância: os valores obtidos ajudarão a compor uma descrição mais precisa do processo transcricional e, conseqüentemente, serão utilizados em nossas previsões sobre o ruído da expressão e a distribuição de probabilidade total.

4.3 Otimizando um modelo para f com dependências espacial e temporal

As análises realizadas nas seções anteriores de fato nos mostraram que as Eqs.(4.1) e (4.5) podem ser aplicadas ao longo do eixo AP sem considerarmos dependências espaciais adicionais, como efeitos difusivos, por exemplo. Para tanto, basta refinar as dependências espaciais de $q_\alpha(x, t)$. Na seção anterior, vimos também que o refinamento proposto ($\Delta(x)$ e $t_0(x)$ gaussianos) proporcionou um ganho qualitativo mas ainda assim não foi suficiente para descrever a expressão de mRNA com precisão. Por isso, as dependências espaciais necessitam ser revistas e as otimizações nos permitirão fazê-lo. Com as otimizações, esperamos também refinar os valores dos parâmetros k , χ , ρ e ϵ .

Contudo, antes de prosseguirmos, é importante pontuar que, em virtude do relativo sucesso do modelo nas seções anteriores, estenderemos as análises para o restante do eixo AP, isto é, consideraremos todo o intervalo no qual registraram-se os dados: de AP = 35.5% até AP = 91.5%. Dessa forma, reconsideraremos o pico de expressão que ocorre em torno de AP = 80.5%. Em verdade, toda a expressão que ocorre no intervalo entre AP = 77.5% até AP = 91.5% comporta-se de forma peculiar: inicialmente, o número de moléculas aumenta e posteriormente diminui (assim como nas demais posições ao longo do eixo AP) mas, em um determinado momento, o número de moléculas volta a

aumentar e depois diminui novamente. Para descrevermos o segundo aumento e o segundo decréscimo no número de moléculas de mRNA, introduziremos um segundo “interruptor”, representado por

$$q'_\alpha(x, t) = \begin{cases} \Delta(x) & , t \leq t_0(x) \\ \delta(x) & , t_0(x) < t \leq t_{ref}(x) \\ \Delta'(x) & , t_{ref}(x) < t \leq t'(x) \\ \delta'(x) & , t > t'(x) \end{cases}, \quad (4.24)$$

cuja dependência temporal é melhor visualizada na Figura 4.12. Para região entre AP =

Figura 4.12 – Dependência temporal da função $q'_\alpha(x, t)$ na região entre AP = 77.5% até AP = 91.5%, descrita pela Eq.(4.24). Esse gráfico mostra que o efeito regulatório sobre o sistema, caracterizado em dois momentos: no primeiro momento, o sistema está inicialmente ativado, caracterizado por $\Delta(x)$, e, em $t_0(x)$, o sistema passa a sofrer repressão, caracterizada por $\delta(x)$; no segundo momento, iniciado em t_{ref} , o sistema volta a ser ativado, dessa vez, caracterizada por $\Delta'(x)$ e, após t' , é novamente submetido à repressão, caracterizada por $\delta'(x)$.

77.5% até AP = 91.5%, são propostas duas sequências consecutivas: na primeira sequência, o sistema está inicialmente ativado, caracterizado por $\Delta(x)$, e, em $t_0(x)$, o sistema passa a sofrer repressão, caracterizada por $\delta(x)$; na segunda sequência, iniciada em t_{ref} , o sistema volta a ser ativado, dessa vez, caracterizada por $\Delta'(x)$ e, após t' , é novamente submetido à repressão, caracterizada por $\delta'(x)$. A implementação de $q'_\alpha(x, t)$ acarretará em uma probabilidade (do gene estar ativo) P'_α dada por

$$P'_\alpha(x, t) = \begin{cases} \Delta(x) + [p_\alpha^0(x) - \Delta(x)]e^{-\epsilon\rho t} & , t \leq t_0(x) \\ \delta(x) + [p_\alpha^{t_0}(x) - \delta(x)]e^{-\epsilon\rho(t-t_0(x))} & , t_0(x) < t \leq t_{ref}(x) \\ \Delta'(x) + [p_\alpha^{t_{ref}}(x) - \Delta'(x)]e^{-\epsilon\rho t} & , t_{ref}(x) < t \leq t'(x) \\ \delta'(x) + [p_\alpha^{t'}(x) - \delta'(x)]e^{-\epsilon\rho(t-t_0(x))} & , t > t'(x) \end{cases} \quad (4.25)$$

e em um número de moléculas transcritas $\langle n \rangle'$ dado por

$$\langle n \rangle' (x, t) = \begin{cases} A_1(x) + B_1(x)e^{-\rho t} + C_1(x)e^{-\epsilon \rho t} & , t \leq t_0(x) \\ A_2(x) + B_2(x)e^{-\rho(t-t_0(x))} + C_2(x)e^{-\epsilon \rho(t-t_0(x))} & , t_0(x) < t \leq t_{ref}(x) \\ A_3(x) + B_3(x)e^{-\rho(t-t_{ref}(x))} + C_3(x)e^{-\epsilon \rho(t-t_{ref}(x))} & , t_{ref}(x) < t \leq t'(x) \\ A_4(x) + B_4(x)e^{-\rho(t-t'(x))} + C_4(x)e^{-\epsilon \rho(t-t'(x))} & , t > t'(x) \end{cases}, \quad (4.26)$$

onde

$$A_1(x) = N(1 - \chi)\Delta(x) + N\chi \quad (4.27)$$

$$B_1(x) = n_0(x) - A_1(x) - C_1(x)$$

$$C_1(x) = \frac{N(1 - \chi)[p_\alpha^0(x) - \Delta(x)]}{1 - \epsilon}$$

$$A_2(x) = N(1 - \chi)\delta(x) + N\chi$$

$$B_2(x) = n_{t_0}(x) - A_2(x) - C_2(x)$$

$$C_2(x) = \frac{N(1 - \chi)[p_\alpha^{t_0}(x) - \delta(x)]}{1 - \epsilon}$$

$$A_3(x) = N(1 - \chi)\Delta'(x) + N\chi$$

$$B_3(x) = n_{t_{ref}}(x) - A_3(x) - C_3(x)$$

$$C_3(x) = \frac{N(1 - \chi)[p_\alpha^{t_{ref}}(x) - \Delta'(x)]}{1 - \epsilon}$$

$$A_4(x) = N(1 - \chi)\delta'(x) + N\chi$$

$$B_4(x) = n_{t'}(x) - A_4(x) - C_4(x)$$

$$C_4(x) = \frac{N(1 - \chi)[p_\alpha^{t'}(x) - \delta'(x)]}{1 - \epsilon}.$$

Procederemos da seguinte maneira. Aplicaremos as Eqs.(4.14) e (4.26) às evoluções temporais registradas em cada posição AP, ou seja, repetiremos o mesmo procedimento realizado em AP = 41.5% para todas as outras 56 posições ao longo do eixo AP: de AP = 35.5% até AP = 76.5%, utilizamos a Eq.(4.14); de AP = 77.5% até AP = 91.5%, utilizamos a Eq.(4.26). Cada otimização fornecerá valores para k , χ , ρ , ϵ , δ , Δ , t_0 , n_0 e p_α^0 (ou seja, permitiremos que os valores de n_0 ao longo do eixo AP flutuem em torno dos valores experimentais registrados em C13). Na região compreendida entre AP = 77.5%

até $AP = 91.5\%$, serão fornecidos, além dos valores para os parâmetros mencionados, valores para Δ' , δ' , t' e t_{ref} .

Com relação aos parâmetros k , χ , ρ e ϵ , permitiremos que os valores de tais parâmetros flutuem respectivamente em torno de $k = 0.0736267$, $\chi = 0.0102286$, $\epsilon = 1.51154$ e $\rho = 0.00240715$, que são os valores obtidos na análise em $AP = 41.5\%$ quando permitimos que n_0 flutuasse em torno do valor experimental em C13. Por considerarmos tais parâmetros intrínsecos e, por isso, invariantes ao longo do eixo AP, tomaremos médias sobre os valores obtidos.

Com relação aos valores obtidos para Δ , δ , t_0 , Δ' , δ' , t' e t_{ref} ao longo do eixo, nós utilizaremos funções interpoladoras para descrevê-los de modo que possamos representar matematicamente suas dependências espaciais: $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$.

E, por fim, com relação às condições iniciais, os valores fornecidos para n_0 e p_α^0 ao longo do eixo AP pelas otimizações serão descartados. Para descrever as condições iniciais, utilizaremos as Eqs.(4.16) e (4.17). Contudo, isso só será feito após termos estabelecido valores para k , χ , ρ e ϵ e expressões para $\Delta(x)$, $\delta(x)$ e $t_0(x)$. As razões para isso são o fato do modelo não ser sensível às pequenas diferenças entre as condições iniciais e os valores dados pelas Eqs.(4.16) e (4.17) e, principalmente, a economia de parâmetros (a Eq.(4.16) utiliza apenas 3 parâmetros para descrever os 57 valores registrados em C13 e, para inferir os 57 valores da probabilidade inicial, a Eq.(4.17), por ser automaticamente relacionada à Eq.(4.16), evita a utilização de mais parâmetros). Objetivamos, com isso, ter ao final das otimizações, uma expressão com o menor número de parâmetros possível com a qual consigamos não apenas reroduzir os dados experimentais mas também fazer predições sobre flutuação, ruído e distribuição de probabilidade.

4.3.1 Descrevendo o número de moléculas

Realizadas as otimizações e obtidos os valores para k , χ , ρ e ϵ , tomamos a média de

tais valores e obtemos os seguintes valores:

$$k = 0.07559 \quad (4.28)$$

$$\chi = 0.01269$$

$$\epsilon = 1.57962$$

$$\rho = 0.00261,$$

com os quais trabalharemos a partir de agora.

Com relação aos parâmetros Δ , δ , t_0 , Δ' , δ' , t' e t_{ref} , as otimizações fornecem, para cada um, os valores ao longo do eixo AP e, como esperado, tais valores não são se mantêm constante ao longo do eixo AP. A partir dos valores obtidos, tratamos de descrever cada parâmetro como uma função da posição sobre o eixo AP. As funções que, a partir de agora, irão descrever as dependências espaciais de Δ , δ , t_0 , Δ' , δ' , t' e t_{ref} são:

$$\begin{aligned} \Delta(x) = & 0.015 + 0.446\exp[-0.507(x - 36.127)^2] \\ & + 0.948\exp[-0.074(x - 41.150)^2] \\ & + 0.180\exp[-0.044(x - 76.284)^2] \\ & + 0.143\exp[-0.012(x - 85.728)^2] \end{aligned} \quad (4.29)$$

$$\begin{aligned} \delta(x) = & 0.0023 + 0.0191\exp[-0.1433(x - 80.1794)^2] \\ & + 0.0335\exp[-0.0258(x - 83.999)^2] \end{aligned} \quad (4.30)$$

$$\begin{aligned} t_0(x) = & 265.16 + 474.701\exp[-0.021(x - 38.936)^2] \\ & + 867.995\exp[-0.076(x - 41.056)^2] \\ & + 604.109\exp[-0.081(x - 41.502)^2] \\ & + 989.348\exp[-0.069(x - 50.113)^2] \\ & + 489.367\exp[-0.035(x - 52.0)^2] \\ & + 1248.83\exp[-0.019(x - 61.136)^2] \\ & + 678.227\exp[-0.156(x - 68.096)^2] \end{aligned} \quad (4.31)$$

$$\begin{aligned}\Delta'(x) &= 0.0044 + 0.2187\exp[-0.143514(x - 80.5553)^2] \\ &+ 0.0939\exp[-0.2328(x - 84.9694)^2]\end{aligned}\quad (4.32)$$

$$\delta'(x) = 0.00031 + 0.09637\exp[-1.99429(x - 82.89)^2] \quad (4.33)$$

$$\begin{aligned}t'(x) &= 310.177 + 968.194\exp[-0.911(x - 77.844)^2] \\ &+ 1463.673\exp[-0.081(x - 80.136)^2] \\ &+ 512.866\exp[-0.010(x - 84.0)^2] \\ &+ 668.203\exp[-1.871(x - 87.698)^2] \\ &+ 1244.918\exp[-0.013(x - 88.988)^2] \\ &+ 1847.790\exp[-0.056(x - 93.368)^2]\end{aligned}\quad (4.34)$$

$$t_{ref}(x) = \begin{cases} 2340, & 77.5 \leq x < 79.0 \\ 1560, & 79.0 \leq x < 82.0 \\ 1170, & 82.0 \leq x < 85.0 \\ 1560, & 85.0 \leq x < 87.0 \\ 1950, & 87.0 \leq x < 91.5 \end{cases} \quad (4.35)$$

É importante mencionar que na região entre AP = 77.5% até AP = 91.5%, procedemos de uma maneira ligeiramente diferente do que no restante do eixo AP. Por construção, t_{ref} é o instante de referência partir do qual o efeito ativador volta a ser considerável para depois, em t' , tornar-se desprezível. Em cada posição, é possível perceber a classe temporal a partir da qual o número de moléculas volta a crescer, o que fornece uma boa noção do valor de t_{ref} em cada posição. Contudo, para evitarmos a inclusão de mais parâmetros para a descrição de $t_{ref}(x)$, escolhemos tais classes temporais para representar os instantes de referência t_{ref} . Dessa forma, em cada posição, tendo fixado t_{ref} , realizamos a otimização em duas partes: na primeira, utilizamos os dados até t_{ref} para buscar valores para Δ , δ e t_0 ; na segunda, utilizamos os dados a partir de t_{ref} e para obter valores para Δ' , δ' e t' .

A Figura 4.13 mostra os gráficos das Eqs.(4.29), (4.30), (4.31), (4.32), (4.33), (4.34) e (4.35) e os valores obtidos para os parâmetros Δ , δ , t_0 , Δ' , δ' , t' e t_{ref} ao longo do eixo AP através das otimizações. Dentre todos os valores otimizados, observamos primeiramente os

Figura 4.13 – Gráficos mostram os valores obtidos para os parâmetros Δ , δ , t_0 , Δ' , δ' , t' e t_{ref} ao longo do eixo AP através das otimizações com as funções $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$, dadas pelas Eqs.(4.29), (4.30), (4.31), (4.32), (4.33), (4.34) e (4.35) respectivamente.

resultados obtidos para os parâmetros Δ e t_0 . Podemos notar que seus valores otimizados apresentam ao longo do eixo AP um comportamento que condiz com as dependências espaciais propostas nas Eqs.(4.20) e (4.21), sendo matematicamente descritas por perfis gaussianos. Com relação à forma gaussiana, podemos notar que as Eqs.(4.29) e (4.31) fornecem larguras maiores às gaussianas do que as fornecidas pelas funções-teste descritas pelas Eqs.(4.20) e (4.21). Isso implica que, nas vizinhanças de $AP = 41.5\%$, comparando com descrição realizada na seção anterior, há mais núcleos expressando moléculas em grande número, tornando visualmente a faixa mais “espessa”. Também podemos notar

que os valores fornecidos pelas otimizações para t_0 em regiões intermediárias do eixo AP são comparáveis aos valores fornecidos para regiões dentro do domínio de formação da faixa, o que é consideravelmente diferente do proposto pela Eq.(4.21), na qual os valores de t_0 em regiões intermediárias é nulo. Outro ponto interessante que merece atenção são os valores fornecidos para δ ao longo do eixo AP, mais especificamente, na região entre AP = 35.5% até AP = 77.5%: em uma escala de 0 a 1, tais valores praticamente não mudam ao longo do eixo AP (estando, em sua grande maioria, entre 0 e 0.01), o que concorda com as propostas anteriores, nas quais havíamos considerado $\delta(x)$ constante ao longo de AP.

Com relação à região compreendida entre AP = 77.5% até AP = 91.5%, mais especificamente, em torno de AP = 80.5%, podemos ver que: os valores para Δ' fornecidos pelas otimizações chegam a atingir aproximadamente 0.3, o que entendemos que seja suficiente para a formação do pequeno pico de expressão. Quanto aos valores para δ' , a imensa maioria não passa de 0.01. Na qualidade de fator determinante para estados assintóticos, os valores de δ e δ' mostram que o nível de transcrição ao longo de todo o eixo AP tende a ser drasticamente minimizado, isto é, os genes são efetivamente desligados para tempos suficientemente grandes (após $t_0(x)$ e $t'(x)$).

Concluídas as otimizações e, sobretudo, tendo em mãos as dependências espaciais matematicamente estabelecidas para $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$, podemos escrever a função $q_\alpha(x, t)$, a probabilidade de o gene estar ativo ($P_\alpha(x, t)$) e o valor médio ($\langle n \rangle(x, t)$). Para isso, tomamos os valores dos parâmetros k , χ , ϵ e ρ (dados pela Eq.(4.28)), tomamos as funções $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$ (respectivamente dadas pelas Eqs.(4.29), (4.30), (4.31), (4.32), (4.33), (4.34) e (4.35)), voltamos a descrever as condições iniciais com as Eqs.(4.16) e (4.17) (conforme havíamos adiantado na seção anterior) e introduzimos nas Eqs.(4.12),(4.13),(4.14), (4.15), (4.24), (4.25), (4.26) e (4.27). Os resultados obtidos são mostrados pelas Figuras 4.14 e 4.16.

A Figura 4.14 mostra os gráficos de $q_\alpha(x, t)$ e $P_\alpha(x, t)$ em função do tempo e da posição sobre o eixo AP. Nas Figuras 4.14(b) e 4.14(d), além dos gráficos de $q_\alpha(x, t)$ e $P_\alpha(x, t)$ propriamente ditos, estão mostrados os perfis assumidos por $q_\alpha(x, t)$ e $P_\alpha(x, t)$

Figura 4.14 – (a) Gráfico de $q_\alpha(x,t)$, dado pelas Eqs.(4.12), (4.24) com o auxílio das Eqs.(4.29), (4.30), (4.31), (4.32), (4.33), (4.34) e (4.35). (b) Gráfico de $q_\alpha(x,t)$, os perfis assumidos por $q_\alpha(x,t)$ ao longo do eixo AP em cada classe temporal (curvas vermelhas) e $t_0(x)$ (curva azul). (c) Gráfico de $P_\alpha(x,t)$, resultante da inserção de $q_\alpha(x,t)$ nas Eqs.(4.13) e (4.25). (d) Gráfico de $P_\alpha(x,t)$, os perfis assumidos por $P_\alpha(x,t)$ ao longo do eixo AP em cada classe temporal (curvas vermelhas) e $t_0(x)$ (curva azul).

em cada classe temporal (curvas vermelhas) e o perfil de $t_0(x)$ (curva azul). É possível ver que, devido essencialmente a $\Delta'(x)$, $q_\alpha(x,t)$ e $P_\alpha(x,t)$ atingem valores consideráveis na região entre $AP = 35.5\%$ e $AP = 91.5\%$ (embora pequenos, comparados aos atingidos nas imediações de $AP = 41.5\%$), acarretando na formação de um pequeno pico de expressão em torno de $AP = 80.5\%$.

As Figuras 4.15(a) e 4.15(b) mostram a evolução temporal de $q_\alpha(x,t)$ e $P_\alpha(x,t)$ em algumas posições sobre o eixo AP: $AP = 41.5\%$, $AP = 35.5\%$, $AP = 38.5\%$, $AP = 43.5\%$ e $AP = 46.5\%$, pertencentes ao domínio onde se forma a faixa 2. Na Figura 4.15(a) é possível notar o efeito combinado das dependências espaciais em $\Delta(x)$ e $t_0(x)$: com $\Delta(x)$ e $t_0(x)$ assumindo seus valores máximos em $AP = 41.5\%$, é nessa posição onde observamos o

Figura 4.15 – Evolução temporal de $q_\alpha(x, t)$ e $P_\alpha(x, t)$ em $AP = 35.5\%$, $AP = 38.5\%$, $AP = 41.5\%$, $AP = 43.5\%$, $AP = 45.5\%$, $AP = 60.5\%$ e $AP = 80.5\%$. É possível notar em ambas as figuras o efeito combinado das dependências espaciais em $\Delta(x)$ e $t_0(x)$. **(a)** Com $\Delta(x)$ e $t_0(x)$ assumindo seu valores máximos em $AP = 41.5\%$, é nessa posição onde observamos o estímulo ativador atuando de modo mais intenso e por mais tempo e tanto sua intensidade quanto sua duração diminuem gradativamente à medida que nos afastamos de $AP = 41.5\%$ (curva preta $\rightarrow$ curvas azuis $\rightarrow$ curvas vermelhas), o que caracteriza o processo de afilamento do perfil nos perfis espaciais de $q_\alpha(x, t)$ e $P_\alpha(x, t)$ e, naturalmente, estará presente nos perfis do número médio de moléculas. Vale notar que em regiões intermediárias do eixo AP, como $AP = 60.5\%$ (curva verde), o sistema não é efetivamente ativado. **(b)** Podemos notar que, em cada posição, P_α praticamente já atinge valores assintóticos em T2. Além disso, é imediato notar que em $AP = 41.5\%$ P_α permanece em regime assintótico por mais tempo do que nas outras posições e que quanto mais distante de $AP = 41.5\%$ menor é o tempo t_0 de permanência no regime assintótico.

estímulo ativador atuando de modo mais intenso e por mais tempo e tanto sua intensidade quanto sua duração diminuem gradativamente à medida que nos afastamos de $AP = 41.5\%$ (curva preta $\rightarrow$ curvas azuis $\rightarrow$ curvas vermelhas). Isso representa um processo de afilamento nos perfis espaciais de $q_\alpha(x, t)$ e $P_\alpha(x, t)$ e, naturalmente, estará presente nos perfis do número médio de moléculas. Na Figura 4.15(b), podemos notar que, em cada posição, P_α praticamente já atinge valores assintóticos em T2. Como já antevimos de certa forma, é imediato notar que P_α em $AP = 41.5\%$ permanece em regime assintótico por mais tempo do que nas outras posições e que quanto mais distante de $AP = 41.5\%$ menor é o tempo t_0 de permanência no regime assintótico.

A Figura 4.16 mostra o número médio de moléculas de mRNA ao longo do eixo AP em cada classe temporal descrito pelas Eqs.(4.14) e (4.26) e o compara com os dados experimentais. Em comparação com a Figura 4.11, vemos que com a otimização dos parâmetros Δ , δ , t_0 , Δ' , δ' , t' e t_{ref} ao longo do eixo AP e a subsequente determinação das funções $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$ descreve com maior exatidão os dados experimentais. Dentre todas as parametrizações, podemos destacar novamente

Figura 4.16 – Gráfico do número de moléculas de mRNA em função do tempo e da posição sobre o eixo AP, segundo as Eqs.(4.14) e (4.26), considerando as funções $q_\alpha(x, t)$, especificada por Eq.(4.12), e $q_\alpha(x, t)$, especificada pela Eq.(4.24). Utilizando os mesmos parâmetros k , χ , ϵ e ρ ao longo do eixo AP e as funções $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$, a dinâmica do processo é descrita pelo modelo, assim como ocorreu nas análises anteriores. A formação das bordas da faixa e seu processo de afilamento são devidamente descritos após a otimização dos parâmetros, garantindo o bom desempenho do modelo.

a parametrização de $\Delta(x)$ e $t_0(x)$, que descrevem com boa concordância a definição das bordas da faixa e o afilamento do pico de expressão. Além disso, destacamos também a introdução da Eq.(4.24), responsável por descrever o aparecimento do pico de expressão em torno de $AP = 80.5\%$. Recordando que, no modelo original do gene binário, a especificação das taxas h e f determina para onde o sistema deve evoluir, isto é, seu estado final, é imediato constatar a extensão desse raciocínio ao observarmos as Figuras 4.14 e 4.16.

A Figura 4.17 mostra a evolução do valor médio $\langle n \rangle(x, t)$ em $AP = 41.5\%$, $AP = 35.5\%$, $AP = 38.5\%$, $AP = 43.5\%$ e $AP = 46.5\%$. É imediato ver o processo de afilamento do perfil espacial do número médio de moléculas de mRNA, consequência direta da dependência espacial em $t_0(x)$. Devido a diferença entre as escalas temporais características de $P_\alpha(x, t)$

Figura 4.17 – Evolução temporal do valor médio $\langle n \rangle(x, t)$ em AP = 41.5%, AP = 35.5%, AP = 38.5%, AP = 43.5% e AP = 45.5%. A comparação nos mostra nitidamente o processo de afilamento do perfil do número médio de moléculas de mRNA, consequência direta da dependência espacial em $t_0(x)$. Diferentemente do que ocorre com $P_\alpha(x, t)$, o valor médio $\langle n \rangle(x, t)$ não atinge valores assintóticos em T2. Como a dinâmica de $\langle n \rangle(x, t)$ é caracterizada por uma escala temporal definida por $\epsilon\rho$ e ρ e a dinâmica de $P_\alpha(x, t)$ é caracterizada por uma escala temporal definida por $\epsilon\rho$ e $\epsilon > 1$, é necessário mais tempo para que $\langle n \rangle(x, t)$ atinja valores assintóticos. É importante verificar que o valor médio na posição AP = 41.5% em T5, $\langle n \rangle(41.5, T5) = 28.69$, está relativamente próxima de $n_S = 28.22$, número médio de moléculas de mRNA em regime estacionário com probabilidade assintótica $p_\alpha = \Delta(41.5)$ (ver Eq.(4.14)), representando mais uma evidência de que, em T5, o sistema se encontra em um estado assintótico ou muito próximo de um estado estacionário.

e $\langle n \rangle(x, t)$, o valor médio não atinge valores assintóticos em T2. Na verdade, como a dinâmica de $\langle n \rangle(x, t)$ é caracterizada por $\epsilon\rho$ e ρ e a dinâmica de $P_\alpha(x, t)$ é caracterizada por $\epsilon\rho$ e $\epsilon > 1$, o valor médio $\langle n \rangle(x, t)$ atinge valores assintóticos depois de $P_\alpha(x, t)$ atingir os seus valores assintóticos. Com relação a valores assintóticos, é importante verificar que a flutuação em AP = 41.5%, $\sigma(41.5, T5) = 7.13$, está relativamente próxima de $\sigma_S = 7.07$, o valor da flutuação em regime estacionário com probabilidade assintótica $p_\alpha = \Delta(41.5)$ (ver Eq.(5.22)), isso representa mais uma evidência de que T5 se encontra em um estado assintótico ou muito próximo de um estado estacionário.

É interessante observar o número de parâmetros a que estamos recorrendo e compará-los com o número total de dados que estamos descrevendo. Além dos 4 parâmetros propostos pelo modelo do gene binário (k , χ , ρ e ϵ), contamos com 3 parâmetros em $n_0(x)$, 13 parâmetros em $\Delta(x)$, 7 em $\delta(x)$, 22 em $t_0(x)$, 7 em $\Delta'(x)$, 4 em $\delta'(x)$, 19 em $t'(x)$ e 8 em $t_{ref}(x)$, correspondendo a um total de 87 parâmetros. O número total de dados experimentais registrados ao longo do eixo AP durante as classes temporais é de 513 ($= 57 \times 9$).

Se na análise em AP = 41.5% havia 9 parâmetros — k , χ , ρ , ϵ , Δ , δ , p_α^0 e n_0 — para descrever 9 pontos (muito embora as condições iniciais não representem grandes problemas

devido à natureza markoviana do modelo e embora, naquele momento, não estivessem matematicamente relacionadas), agora temos, grosso modo, 1 parâmetro para cada 6 pontos experimentais, o que é razoável. De fato, a economia de parâmetros advém da suposição de que o estado inicial é estacionário (vinculando matematicamente as condições iniciais p_α^0 e n_0) e da invariância assumida de k, χ, ρ, ϵ , considerados parâmetros intrínsecos do sistema uma vez que os genes ao longo do eixo AP são iguais. (Com efeito, se k, χ, ρ, ϵ não fossem assumidos invariantes, haveria 1 parâmetro desses para cada uma das 57 posições sobre o eixo AP de modo que teríamos de lidar com um total de $4 \times 57 = 228$ parâmetros, ou seja, a invariância de k, χ, ρ, ϵ “economizou” 224 parâmetros).

De fato, as otimizações nos permitiram uma descrição espacial mais precisa da formação da faixa. Contudo, devemos lembrar que o bom desempenho apresentado até agora das Eqs.(4.14) e (4.26) deve-se às Eqs.(4.1), na qual propusemos as taxas h e f variando no tempo segundo funções de Heaviside, à Eq.(4.5), na qual supusemos ϵ representando uma característica intrínseca do gene e, sobretudo, às Eqs.(3.1) e (3.2), que modelam o processo transcricional sob uma óptica markoviana: novamente, vale ressaltar que, como os estados inicial e final do sistema já são especificados (pelas condições iniciais e pelas taxas h e f), o grande mérito do modelo reside na concordância apresentada em instantes intermediários ou, em termos matemáticos, no fato da dinâmica dos dados experimentais poder ser descrita por exponenciais.

Do ponto de vista fenomenológico, o resultado mostrado pela Figura 4.16 mostra que as duas fases da expressão da faixa, formação e desaparecimento, apesar de parecerem funcionalmente distintas, podem ser descritas pela mesma estocasticidade (markoviana). Além disso, este resultado mostra efetivamente que a formação de um padrão de expressão pode ser modulada exclusivamente pela regulação externa, ou seja, a expressão gênica modulada espacial e temporalmente não requer alterações físico-químicas no gene e nem processos de difusão ou de transporte ao longo do tempo e do espaço mas apenas alterações no chaveamento. Mais precisamente, tais alterações estariam relacionadas a ocupação e desocupação dos sítios por proteínas regulatórias.

Embora já viéssemos chamando a atenção para esse fato nas seções anteriores, não

podíamos afirmá-lo com convicção durante a análise em $AP = 41.5\%$ por conta de suposições que necessitavam de verificação a partir do momento em que estendemos o modelo para todo o eixo AP (como a invariância dos parâmetros k , χ , ϵ e ρ ao longo do eixo AP e o não-uso de efeitos difusivos e de transporte, por exemplo). Dessa forma, além de verificar as referidas suposições, as análises anteriores ao longo do eixo AP serviram para corroborar a força do modelo e, confirmando observações anteriores, a tarefa de otimização dos parâmetros foi necessária para aumento da precisão do modelo mas não é decisiva para a compreensão do fenômeno biológico.

Tendo descrito o número de moléculas ao longo do eixo AP durante todas as classes temporais com relativa precisão, podemos a partir de agora utilizar os resultados obtidos nessa seção para fazer previsões sobre o desvio quadrático (flutuação), o ruído e a distribuição de probabilidade total, quantidades extremamente relevantes para a descrição e entendimento do processo transcricional da faixa 2 de mRNA de *eve* à luz da abordagem estocástica.

Capítulo 5

Flutuação, ruído, distribuição de probabilidade e um *ansatz* para a função q_α

No capítulo anterior, investigamos o processo transcricional à luz do modelo estocástico que considera um gene binário regulado por várias proteínas e vários sítios de ligação. Considerando que as taxas de transição entre níveis comportam-se no decorrer do tempo como “interruptores” (alterando bruscamente a essência de caráter regulatório de ativador para repressor) e realizando otimizações, as equações fornecidas pelo modelo foram capazes de descrever o número de moléculas transcritas ao longo do eixo AP durante todas as classes temporais com boa concordância.

Dado que realizamos uma descrição espacial e temporal do número médio de moléculas de mRNA transcritas à luz de uma abordagem estocástica, trataremos de fazer, nesse capítulo, predições sobre o desvio quadrático (flutuação), ruído e também sobre a distribuição de probabilidade, o que é de grande relevância para a compreensão do fenômeno.

Contudo, antes de apresentarmos as predições propriamente ditas, é conveniente apresentarmos de forma sistematizada as fórmulas do modelo que nos fornecem a variância, ruído e distribuição de probabilidade. Isso se faz necessário porque, como vimos no capítulo anterior, há a formação de um pequeno pico de expressão na região posterior do

embrião (de AP = 77.5% a AP = 91.5%) de modo que dependências adicionais em $q_\alpha(x, t)$, função que representa o efeito ativador total sobre o gene, tiveram de ser implementadas.

Antes ainda de enunciar as expressões das funções que comporão as expressões da variância, ruído e da distribuição de probabilidade, é preciso relembrar os valores dos parâmetros k , χ , ρ e ϵ , obtidos das otimizações e presentes em praticamente todas as expressões a serem utilizados. Os valores são:

$$k = 0.07559 \quad (5.1)$$

$$\chi = 0.01269$$

$$\epsilon = 1.57962$$

$$\rho = 0.00261.$$

A função $q_\alpha(x, t)$, que expressa o efeito regulatório externo sobre o gene, é dada pela composição das Eqs.(4.12) e (4.24):

$$q_\alpha(x, t) = \begin{cases} q_\alpha^I(x, t) & , 35.5\% \leq x < 77.5\% \\ q_\alpha^{II}(x, t) & , 77.5\% \leq x \leq 91.5\% \end{cases} \quad (5.2)$$

onde

$$q_\alpha^I(x, t) = \begin{cases} \Delta(x) & , t \leq t_0(x) \\ \delta(x) & , t > t_0(x) \end{cases}, \quad (5.3)$$

$$q_\alpha^{II}(x, t) = \begin{cases} \Delta(x) & , t \leq t_0(x) \\ \delta(x) & , t_0(x) < t \leq t_{ref}(x) \\ \Delta'(x) & , t_{ref}(x) < t \leq t'(x) \\ \delta'(x) & , t > t'(x) \end{cases} \quad (5.4)$$

e as funções $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$ são descritas pelas Eqs.(4.29), (4.30), (4.31), (4.32), (4.33), (4.34) e (4.35) respectivamente.

Com a definição de $q_\alpha(x, t)$, a probabilidade $P_\alpha(x, t)$ pode também ser definida, através

da composição das Eqs.(4.13) e (4.25):

$$P_\alpha(x, t) = \begin{cases} P_\alpha^I(x, t) & , 35.5\% \leq x < 77.5\% \\ P_\alpha^{II}(x, t) & , 77.5\% \leq x \leq 91.5\% \end{cases} \quad (5.5)$$

onde

$$P_\alpha^I(x, t) = \begin{cases} \Delta(x) + [p_\alpha^0(x) - \Delta(x)]e^{-\epsilon\rho t} & , t \leq t_0(x) \\ \delta(x) + [p_\alpha^{t_0}(x) - \delta(x)]e^{-\epsilon\rho(t-t_0(x))} & , t > t_0(x) \end{cases} \quad (5.6)$$

$$P_\alpha^{II}(x, t) = \begin{cases} \Delta(x) + [p_\alpha^0(x) - \Delta(x)]e^{-\epsilon\rho t} & , t \leq t_0(x) \\ \delta(x) + [p_\alpha^{t_0}(x) - \delta(x)]e^{-\epsilon\rho(t-t_0(x))} & , t_0(x) < t \leq t_{ref}(x) \\ \Delta'(x) + [p_\alpha^{t_{ref}}(x) - \Delta'(x)]e^{-\epsilon\rho(t-t_{ref}(x))} & , t_{ref}(x) < t \leq t'(x) \\ \delta'(x) + [p_\alpha^{t'}(x) - \delta'(x)]e^{-\epsilon\rho(t-t'(x))} & , t > t'(x) \end{cases} . \quad (5.7)$$

A continuidade temporal é estabelecida pois $p_\alpha^{t_0}(x) = P_\alpha(x, t_0)$, $p_\alpha^{t_{ref}}(x) = P_\alpha(x, t_{ref})$ e $p_\alpha^{t'}(x) = P_\alpha(x, t')$.

Com as funções $q_\alpha(x, t)$ e $P_\alpha(x, t)$ definidas, o número médio de moléculas de mRNA pode ser calculado e será descrito pela composição das Eqs.(4.14) e (4.26):

$$\langle n \rangle(x, t) = \begin{cases} \langle n \rangle_I(x, t) & , 35.5\% \leq x < 77.5\% \\ \langle n \rangle_{II}(x, t) & , 77.5\% \leq x \leq 91.5\% \end{cases} \quad (5.8)$$

onde

$$\langle n \rangle_I(x, t) = \begin{cases} A_1(x) + B_1(x)e^{-\rho t} + C_1(x)e^{-\epsilon\rho t} & , t \leq t_0(x) \\ A_2(x) + B_2(x)e^{-\rho(t-t_0)} + C_2(x)e^{-\epsilon\rho(t-t_0(x))} & , t > t_0(x) \end{cases} , \quad (5.9)$$

$$\langle n \rangle_{II}(x, t) = \begin{cases} A_1(x) + B_1(x)e^{-\rho t} + C_1(x)e^{-\epsilon\rho t} & , t \leq t_0(x) \\ A_2(x) + B_2(x)e^{-\rho(t-t_0(x))} + C_2(x)e^{-\epsilon\rho(t-t_0(x))} & , t_0(x) < t \leq t_{ref}(x) \\ A_3(x) + B_3(x)e^{-\rho(t-t_{ref}(x))} + C_3(x)e^{-\epsilon\rho(t-t_{ref}(x))} & , t_{ref}(x) < t \leq t'(x) \\ A_4(x) + B_4(x)e^{-\rho(t-t'(x))} + C_4(x)e^{-\epsilon\rho(t-t'(x))} & , t > t'(x) \end{cases} , \quad (5.10)$$

e

$$A_1(x) = N(1 - \chi)\Delta(x) + N\chi \quad B_1(x) = n_0(x) - A_1(x) - C_1(x)$$

$$C_1(x) = \frac{N(1 - \chi)[p_\alpha^0(x) - \Delta(x)]}{1 - \epsilon}$$

$$A_2(x) = N(1 - \chi)\delta(x) + N\chi \quad B_2(x) = n_{t_0}(x) - A_2(x) - C_2(x)$$

$$C_2(x) = \frac{N(1 - \chi)[p_\alpha^{t_0}(x) - \delta(x)]}{1 - \epsilon}$$

$$A_3(x) = N(1 - \chi)\Delta'(x) + N\chi \quad B_3(x) = n_{t_{ref}}(x) - A_3(x) - C_3(x)$$

$$C_3(x) = \frac{N(1 - \chi)[p_\alpha^{t_{ref}}(x) - \Delta'(x)]}{1 - \epsilon}$$

$$A_4(x) = N(1 - \chi)\delta'(x) + N\chi \quad B_4(x) = n_{t'}(x) - A_4(x) - C_4(x)$$

$$C_4(x) = \frac{N(1 - \chi)[p_\alpha^{t'}(x) - \delta'(x)]}{1 - \epsilon}.$$

A continuidade temporal é estabelecida pois $n_{t_0}(x) = n(x, t_0)$, $n_{t_{ref}}(x) = n(x, t_{ref})$ e $n_{t'}(x) = n(x, t')$.

Com as expressões para $q_\alpha(x, t)$, $P_\alpha(x, t)$ e $\langle n \rangle(x, t)$ definidas, resta-nos explicitar as expressões utilizadas para o número de moléculas quando o gene se encontra ativo, $\langle n_\alpha \rangle(x, t)$, e para o segundo momento da distribuição, $\langle n^2 \rangle(x, t)$. A expressão para $\langle n_\alpha \rangle(x, t)$ é dada por

$$\langle n_\alpha \rangle(x, t) = \begin{cases} \langle n_\alpha \rangle_I(x, t) & , 35.5\% \leq x < 77.5\% \\ \langle n_\alpha \rangle_{II}(x, t) & , 77.5\% \leq x \leq 91.5\% \end{cases} \quad (5.11)$$

onde

$$\langle n_\alpha \rangle_I(x, t) = \begin{cases} D_1(x) + E_1(x)e^{-\rho t} + \\ F_1(x)e^{-\epsilon \rho t} + G_1(x)e^{-\rho(1+\epsilon)t} & , \quad t \leq t_0(x) \\ D_2(x) + E_2(x)e^{-\rho(t-t_0(x))} + \\ F_2(x)e^{-\epsilon \rho(t-t_0(x))} + G_2(x)e^{-\rho(1+\epsilon)(t-t_0(x))} & , \quad t > t_0(x) \end{cases}, \quad (5.12)$$

$$\langle n_\alpha \rangle_{II}(x, t) = \begin{cases} D_1(x) + E_1(x)e^{-\rho t} + \\ F_1(x)e^{-\epsilon\rho t} + G_1(x)e^{-\rho(1+\epsilon)t} \quad , \quad t \leq t_0(x) \\ D_2(x) + E_2(x)e^{-\rho(t-t_0(x))} + \\ F_2(x)e^{-\epsilon\rho(t-t_0(x))} + G_2(x)e^{-\rho(1+\epsilon)(t-t_0(x))} \quad , \quad t_0(x) < t \leq t_{ref}(x) \\ D_3(x) + E_3(x)e^{-\rho(t-t_{ref}(x))} + \\ F_3(x)e^{-\epsilon\rho(t-t_{ref}(x))} + G_3(x)e^{-\rho(1+\epsilon)(t-t_{ref}(x))} \quad , \quad t_{ref}(x) < t \leq t'(x) \\ D_4(x) + E_4(x)e^{-\rho(t-t'(x))} + \\ F_4(x)e^{-\epsilon\rho(t-t'(x))} + G_4(x)e^{-\rho(1+\epsilon)(t-t'(x))} \quad , \quad t > t_0(x) \end{cases} \quad (5.13)$$

e

$$\begin{aligned} D_1(x) &= \Delta(x)A_1(x) + \frac{N(1-\chi)\Delta(x)[1-\Delta(x)]}{1+\epsilon} \\ D_2(x) &= \delta(x)A_2(x) + \frac{N(1-\chi)\delta(x)[1-\delta(x)]}{1+\epsilon} \\ D_3(x) &= \Delta'(x)A_3(x) + \frac{N(1-\chi)\Delta'(x)[1-\Delta'(x)]}{1+\epsilon} \\ D_4(x) &= \delta'(x)A_4(x) + \frac{N(1-\chi)\delta'(x)[1-\delta'(x)]}{1+\epsilon} \\ E_1(x) &= \Delta(x)B_1(x) \quad F_1(x) = \Delta(x)C_1(x) + H_1(x) \\ E_2(x) &= \delta(x)B_2(x) \quad F_2(x) = \delta(x)C_2(x) + H_2(x) \\ E_3(x) &= \Delta'(x)B_3(x) \quad F_3(x) = \Delta'(x)C_3(x) + H_3(x) \\ E_4(x) &= \delta'(x)B_4(x) \quad F_4(x) = \delta'(x)C_4(x) + H_4(x) \\ G_1(x) &= \langle n_\alpha \rangle_0(x) - D_1(x) - \Delta(x)[n_0(x) - A_1(x)] - H_1(x) \\ G_2(x) &= \langle n_\alpha \rangle_{t_0}(x) - D_2(x) - \delta(x)[n_{t_0}(x) - A_2(x)] - H_2(x) \\ G_3(x) &= \langle n_\alpha \rangle_{t_{ref}}(x) - D_3(x) - \Delta'(x)[n_{t_{ref}}(x) - A_3(x)] - H_3(x) \\ G_4(x) &= \langle n_\alpha \rangle_{t'}(x) - D_4(x) - \delta'(x)[n_{t'}(x) - A_4(x)] - H_4(x) \end{aligned}$$

$$H_1(x) = N(1 - \Delta(x) + \chi\Delta(x))[p_\alpha^0(x) - \Delta(x)]$$

$$H_2(x) = N(1 - \delta(x) + \chi\delta(x))[p_\alpha^{t_0}(x) - \delta(x)]$$

$$H_3(x) = N(1 - \Delta'(x) + \chi\Delta'(x))[p_\alpha^{t_{ref}}(x) - \Delta'(x)]$$

$$H_4(x) = N(1 - \delta'(x) + \chi\Delta(x))[p_\alpha^{t'}(x) - \delta'(x)]$$

A continuidade temporal é estabelecida pois $\langle n_\alpha \rangle_{t_0}(x) = \langle n_\alpha \rangle(x, t_0)$, $\langle n_\alpha \rangle_{t_{ref}}(x) = \langle n_\alpha \rangle(x, t_{ref})$ e $\langle n_\alpha \rangle_{t'}(x) = \langle n_\alpha \rangle(x, t')$.

E a expressão para $\langle n^2 \rangle(x, t)$ é dada por

$$\langle n^2 \rangle(x, t) = \begin{cases} \langle n^2 \rangle_I(x, t) & , 35.5\% \leq x < 77.5\% \\ \langle n^2 \rangle_{II}(x, t) & , 77.5\% \leq x \leq 91.5\% \end{cases} \quad (5.14)$$

onde

$$\langle n^2 \rangle_I(x, t) = \begin{cases} R_1(x) + S_1(x)e^{-2\rho(t-t_0(x))} + T_1(x)e^{-\rho(t-t_0(x))} + \\ U_1(x)e^{-\epsilon\rho(t-t_0(x))} + V_1(x)e^{-\rho(1+\epsilon)(t-t_0(x))} & , t \leq t_0(x) \\ R_2(x) + S_2(x)e^{-2\rho(t-t_0(x))} + T_2(x)e^{-\rho(t-t_0(x))} + \\ U_2(x)e^{-\epsilon\rho(t-t_0(x))} + V_2(x)e^{-\rho(1+\epsilon)(t-t_0(x))} & , t > t_0(x) \end{cases} \quad (5.15)$$

$$\langle n^2 \rangle_{II}(x, t) = \begin{cases} R_1(x) + S_1(x)e^{-2\rho t} + T_1(x)e^{-\rho t} + \\ U_1(x)e^{-\epsilon\rho t} + V_1(x)e^{-\rho(1+\epsilon)t} & , t \leq t_0(x) \\ R_2(x) + S_2(x)e^{-2\rho(t-t_0(x))} + T_2(x)e^{-\rho(t-t_0(x))} + \\ U_2(x)e^{-\epsilon\rho(t-t_0(x))} + V_2(x)e^{-\rho(1+\epsilon)(t-t_0(x))} & , t_0(x) < t \leq t_{ref}(x) \\ R_3(x) + S_3(x)e^{-2\rho(t-t_{ref}(x))} + T_3(x)e^{-\rho(t-t_{ref}(x))} + \\ U_3(x)e^{-\epsilon\rho(t-t_{ref}(x))} + V_3(x)e^{-\rho(1+\epsilon)(t-t_{ref}(x))} & , t_{ref}(x) < t \leq t'(x) \\ R_4(x) + S_4(x)e^{-2\rho(t-t'(x))} + T_4(x)e^{-\rho(t-t'(x))} + \\ U_4(x)e^{-\epsilon\rho(t-t'(x))} + V_4(x)e^{-\rho(1+\epsilon)(t-t'(x))} & , t > t'(x) \end{cases} \quad (5.16)$$

e

$$R_1(x) = A_1(x)^2 + A_1(x) + \frac{N^2(1-\chi)^2\Delta(x)[1-\Delta(x)]}{1+\epsilon}$$

$$R_2(x) = A_2(x)^2 + A_2(x) + \frac{N^2(1-\chi)^2\delta(x)[1-\delta(x)]}{1+\epsilon}$$

$$R_3(x) = A_3(x)^2 + A_3(x) + \frac{N^2(1-\chi)^2\Delta'(x)[1-\Delta'(x)]}{1+\epsilon}$$

$$R_4(x) = A_4(x)^2 + A_4(x) + \frac{N^2(1-\chi)^2\delta'(x)[1-\delta'(x)]}{1+\epsilon}$$

$$S_1(x) = \langle n^2 \rangle_0(x) - R_1(x) - T_1(x) - U_1(x) - V_1(x)$$

$$S_2(x) = \langle n^2 \rangle_{t_0}(x) - R_2(x) - T_2(x) - U_2(x) - V_2(x)$$

$$S_3(x) = \langle n^2 \rangle_{t_{ref}}(x) - R_3(x) - T_3(x) - U_3(x) - V_3(x)$$

$$S_4(x) = \langle n^2 \rangle_{t'}(x) - R_4(x) - T_4(x) - U_4(x) - V_4(x)$$

$$T_1(x) = (2A_1(x) + 1)B_1(x)$$

$$T_2(x) = (2A_2(x) + 1)B_2(x)$$

$$T_3(x) = (2A_3(x) + 1)B_3(x)$$

$$T_4(x) = (2A_4(x) + 1)B_4(x)$$

$$U_1(x) = \frac{(2N\chi + 1)C_1(x) + 2N(1-\chi)F_1(x) + N(1-\chi)[p_\alpha^0(x) - \Delta(x)]}{2-\epsilon}$$

$$U_2(x) = \frac{(2N\chi + 1)C_2(x) + 2N(1-\chi)F_2(x) + N(1-\chi)[p_\alpha^{t_0}(x) - \delta(x)]}{2-\epsilon}$$

$$U_3(x) = \frac{(2N\chi + 1)C_3(x) + 2N(1-\chi)F_3(x) + N(1-\chi)[p_\alpha^{t_{ref}}(x) - \Delta'(x)]}{2-\epsilon}$$

$$U_4(x) = \frac{(2N\chi + 1)C_4(x) + 2N(1-\chi)F_4(x) + N(1-\chi)[p_\alpha^{t'}(x) - \delta'(x)]}{2-\epsilon}$$

$$V_1(x) = \frac{2N(1-\chi)G_1(x)}{1-\epsilon}$$

$$V_2(x) = \frac{2N(1-\chi)G_2(x)}{1-\epsilon}$$

$$V_3(x) = \frac{2N(1-\chi)G_3(x)}{1-\epsilon}$$

$$V_4(x) = \frac{2N(1 - \chi)G_4(x)}{1 - \epsilon}.$$

A continuidade temporal é estabelecida pois $\langle n^2 \rangle_{t_0}(x) = \langle n^2 \rangle(x, t_0)$, $\langle n^2 \rangle_{t_{ref}}(x) = \langle n^2 \rangle(x, t_{ref})$ e $\langle n^2 \rangle_{t'}(x) = \langle n^2 \rangle(x, t')$.

Com relação às condições iniciais, no Capítulo 4, utilizamos a Eq.(4.16) para escrever $n_0(x)$, o número médio de moléculas de mRNA em C13,

$$n_0(x) = 1.36468 + 12.3591 \exp[-0.0575796(x - 35.5)] \quad (5.17)$$

e, pela Eq.(4.17), escrevemos a condição inicial $p_\alpha^0(x)$ como

$$p_\alpha^0(x) = \frac{n_0(x) - N\chi}{N(1 - \chi)}, \quad (5.18)$$

supondo que, em C13, o sistema se encontre em regime estacionário. Com tal suposição, as condições iniciais $\langle n_\alpha \rangle_0(x)$ e $\langle n^2 \rangle_0(x)$ devem ser dadas por

$$\langle n_\alpha \rangle_0(x) = p_\alpha^0(x)n_0(x) + \frac{N(1 - \chi)p_\alpha^0(x)[1 - p_\alpha^0(x)]}{1 + \epsilon} \quad (5.19)$$

$$\langle n^2 \rangle_0(x) = [n_0(x)]^2 + n_0(x) + \frac{N^2(1 - \chi)^2 p_\alpha^0(x)[1 - p_\alpha^0(x)]}{1 + \epsilon}. \quad (5.20)$$

Com as expressões para $\langle n \rangle$ e para o $\langle n^2 \rangle$, temos condições de avaliar o desvio quadrático (flutuação), dada por

$$\sigma(x, t) = \sqrt{\langle n^2 \rangle(x, t) - \langle n \rangle(x, t)^2}, \quad (5.21)$$

e, conseqüentemente, o ruído, dado pela razão $\sigma / \langle n \rangle$.

5.1 Flutuação e ruído

As predições do nosso modelo para a flutuação $\sigma(x, t)$, dada pela Eq.(5.21), são mostradas pela Figura 5.1. Nela são exibidos os perfis assumidos pela flutuação $\sigma(x, t)$ ao longo do eixo AP durante as classes temporais.

A principal característica apresentada pela flutuação $\sigma(x, t)$ é um processo de min-

Figura 5.1 – Evolução temporal da flutuação, dada pela Eq.(5.21). A principal característica apresentada pela flutuação $\sigma(x, t)$ é um processo de minimização local em torno de $AP = 41.5\%$. Mais precisamente, tal processo ocorre na região entre $AP = 38.5\%$ e $AP = 43.5\%$, iniciando em T2 e terminando em T6.

imização local na região onde a faixa 2 é formada, ou seja, em torno de $AP = 41.5\%$. Mais precisamente, tal processo ocorre na região entre $AP = 38.5\%$ e $AP = 43.5\%$, iniciando em T2 e terminando em T6, e a flutuação é máxima exatamente nessas posições. Do ponto de vista biológico, a minimização local de flutuação em torno de $AP = 41\%$ durante a formação da faixa 2 de mRNA é de extrema relevância uma vez que, como a expressão de *eve* se dá durante a fase embrionária e está inserida em um processo ligado à segmentação do corpo do indivíduo adulto, os processos regulatórios devem apresentar em algum momento uma relativa precisão a fim de garantir, durante a embriogênese, a formação de domínios e padrões em núcleos a partir dos quais serão compostos futuros tecidos.

De modo mais detalhado, a evolução dos perfis se dá da seguinte maneira. Em C13,

o sistema parte da condição de regime estacionário. Em T1, inicia-se a formação de duas regiões exibindo picos de flutuação: a primeira em torno de $AP = 41.5\%$ e a segunda em torno de 80.5% . Em T2, enquanto a intensidade da flutuação nas regiões intermediárias do embrião diminui, em torno de $AP = 41.5\%$ e de 80.5% a intensidade da flutuação praticamente se mantém, deixando mais pronunciados os dois picos. Além disso, no pico em torno de $AP = 41.5\%$ surge o processo de minimização local. Em T3, essa minimização se intensifica, fato que continua ocorrendo até T5. Em T4, a minimização local em torno de $AP = 41.5\%$ continua e, em torno de $AP = 80.5\%$, o pico de flutuação começa a aumentar, fato que dura até T7. Em T5, os picos de flutuação em torno de $AP = 41.5\%$ e de 80.5% se encontram muito bem definidos e o processo de minimização local em torno de $AP = 41.5\%$ atinge seu auge, isto é, nesse momento, o “poço” de variância em torno de $AP = 41.5\%$ atinge sua maior profundidade. É durante essa classe temporal que o perfil da faixa 2 de mRNA atinge maior intensidade e definição. Em especial, é importante verificar que a flutuação em $AP = 41.5\%$, $\sigma(41.5, T5) = 7.13$, está relativamente próxima de $\sigma_S = 7.07$, o valor da flutuação em regime estacionário com probabilidade assintótica $p_\alpha = \Delta(41.5)$ (ver Eq.(5.22)), o que representa mais uma evidência de que, em T5, o sistema se encontra em um estado assintótico ou muito próximo de um estado estacionário. A partir de T6, a faixa 2 começa a desaparecer e o processo de minimização começa a ser desfeito, perdendo intensidade mas não alterando seu perfil. Enquanto isso, em $AP = 80.5\%$, o pico de flutuação segue aumentando. Em T7, a minimização local da flutuação já não ocorre mais de modo que o perfil da flutuação ao longo do eixo AP é caracterizado por dois picos com praticamente a mesma intensidade. Em T8, apesar de menos intensos, as intensidades dos picos se mantém comparáveis.

A Figura 5.2 mostra, sob a perspectiva das evoluções temporais em $AP = 41.5\%$, $AP = 35.5\%$, $AP = 38.5\%$, $AP = 41.5\%$, $AP = 43.5\%$, $AP = 45.5\%$ e $AP = 80.5\%$, como ocorre ao longo do tempo o controle de variância na região onde se forma a faixa 2. Além da já observada redução da variância entre $AP = 38.5\%$ e $AP = 43.5\%$ iniciada em T1, podemos ver que, nas bordas desse “poço” de variância, a flutuação praticamente não se altera, passando a diminuir a partir de T5, classe temporal em que a faixa se encontra mais intensa

Figura 5.2 – Evolução temporal da flutuação $\sigma(x, t)$ em $AP = 35.5\%$, $AP = 38.5\%$, $AP = 41.5\%$, $AP = 43.5\%$, $AP = 45.5\%$, $AP = 60.5\%$ e $AP = 80.5\%$. A comparação nos mostra nitidamente o processo de afilamento do pico de flutuação exibido na Figura 5.1. Comparativamente, vemos que nas posições $AP = 41.5\%$ (curva preta) e $AP = 80.5\%$ (curva roxa) — em torno das quais $\langle n \rangle$ e σ apresentam picos — a flutuação apresenta essencialmente o mesmo comportamento, caracterizado por um aumento de C13 até aproximadamente T1, um decréscimo (até T6 e T4, respectivamente), seguido novamente por um aumento até T7 aproximadamente e um decréscimo. Isso mostra que há também um processo minimização da flutuação em $AP = 80.5\%$ embora pouco perceptível. É importante verificar que a flutuação em $AP = 41.5\%$, $\sigma(41.5, T5) = 7.13$, está relativamente próxima de $\sigma_S = 7.07$, o valor da flutuação em regime estacionário com probabilidade assintótica $p_\alpha = \Delta(41.5)$ (ver Eq.(3.38)). É importante verificar que a flutuação em $AP = 41.5\%$, $\sigma(41.5, T5) = 7.13$, está relativamente próxima de $\sigma_S = 7.07$, o valor da flutuação em regime estacionário com probabilidade assintótica $p_\alpha = \Delta(41.5)$ (ver Eq.(3.38)), o que representa mais uma evidência de que, em T5, o sistema se encontra em um estado assintótico ou muito próximo de um estado estacionário. Nas posições $AP = 38.5\%$ e $AP = 43.5\%$ (curva azuis), entre as quais ocorre a minimização de flutuação, a flutuação se mantém praticamente constante até aproximadamente T5, caindo posteriormente. Já nas posições $AP = 35.5\%$ e $AP = 45.5\%$ (curvas vermelhas), à base da faixa, a flutuação se mantém praticamente constante até T3, diminuindo depois. Em $AP = 60.5\%$ (curva verde), onde a transcrição é altamente reprimida (assim como em demais posições ao longo da região intermediária do eixo AP), vemos que a flutuação essencialmente diminui. A linha rosa, que tangencia as evoluções da flutuação em $AP = 38.5\%$ e em $AP = 43.5\%$, corresponde ao valor máximo, σ_{max} , que a flutuação pode atingir quando o sistema estiver em regime estacionário.

e definida e a partir da qual a faixa começa a desaparecer. Comparativamente, vemos que nas posições em torno das quais se ocorrem picos, $AP = 41.5\%$ (curva preta) e $AP = 80.5\%$ (curva roxa), observamos essencialmente o mesmo comportamento, caracterizado por um aumento da flutuação de C13 até aproximadamente T1, um decréscimo (até T6 e T4, respectivamente), seguido novamente por um aumento até T7 aproximadamente e um decréscimo. Isso mostra que a evolução da flutuação em $AP = 80.5\%$ ocorre tal qual a evolução em $AP = 41.5\%$ mas, no entanto, pouco perceptível. Nas posições $AP = 38.5\%$ e $AP = 43.5\%$ (curva azuis), entre as quais ocorre a minimização de flutuação e onde se localiza a base da faixa, a flutuação se mantém praticamente constante até aproximadamente T5, caindo posteriormente. Na posição $AP = 60.5\%$ (curva verde), onde a transcrição de mRNA é altamente reprimida, a flutuação essencialmente diminui.

Os aumentos apresentados pela flutuação em $AP = 41.5\%$ de C13 a T1 e de T6 a T7 podem ser explicados como consequência direta das alterações do modo de tran-

scrição ocorridas em C13 e, depois, em t_0 . Mais precisamente, tanto em instantes antes de C13 quanto em instantes antes de t_0 o sistema está em regime estacionário (ou próximo de um) e, de repente, tem de evoluir para outro estado estacionário. A imposição de um novo estado assintótico significa a imposição de um novo modo de transcrição, caracterizado por um novo ritmo de produção de mRNA, o que está relacionado exclusivamente a alterações no efeito regulatório externo (uma vez que os parâmetros k , χ e ρ permanecem os mesmos) e que, em última instância, está relacionado às alterações na ocupação e desocupação dos sítios de ligação pelas proteínas regulatórias (chaveamento). Apesar de ϵ também permanecer constante ao longo do processo, as taxas que o compõem não permanecem constantes. Dessa forma, é perfeitamente esperado que alterações na dinâmica do chaveamento tendam a causar um aumento na flutuação (nos primeiros instantes após tais alterações) e tal aumento será maior quanto mais distante o sistema estiver de seu novo estado assintótico, ou seja, quanto maiores forem as diferenças entre as condições do sistema em t_0 e as condições do sistema no novo regime estacionário. Em suma, o aumento da flutuação durante esses intervalos específicos pode ser visto também como a destruição das configurações vigentes enquanto o sistema evolui para um outro conjunto de configurações. Matematicamente, essas alterações na dinâmica aparecem mascaradas pelos termos $\exp[-\epsilon\rho t]$ e $\exp[-\rho(1+\epsilon)t]$ que compõem a Eq.(5.21). Como $\epsilon = 1.57962 > 1$, a dinâmica do chaveamento se estabiliza antes da dinâmica do valor médio e, portanto, após um tempo da ordem de $1/(\epsilon\rho)$, esses efeitos já terão se estabilizado de modo que a flutuação passa a ser governada somente pela dinâmica da degradação $\exp(-\rho t)$, estabilizando-se monotonicamente. Embora tenhamos argumentado utilizando o comportamento da flutuação em $AP = 41.5\%$, esse argumento pode ser estendido para as evoluções de σ nas demais posições sobre ao longo do eixo AP.

Novamente, as dependências espaciais contidas em $q_\alpha(x, t)$ são capazes de explicar porque as mudanças de comportamento mais visíveis na evolução da flutuação ocorrem primeiro em regiões intermediárias do eixo AP (onde não há transcrição e nem ativação suficiente para tal) e depois na região em torno de $AP = 41.5\%$. Contudo, somente tais dependências não são capazes de explicar porque em algumas regiões a flutuação,

comparada com seu valor inicial, tende a diminuir ou aumentar.

Figura 5.3 – As evoluções temporais observadas na Figura 5.2 podem ser particularmente analisadas através do gráfico da flutuação σ em função da probabilidade P_α . **(a)** Temos que se o processo de transcrição é uma sucessão de estados estacionários separados por processos infinitesimais, então os valores assumidos por σ em função dos valores assumidos por P_α devem reproduzir trajetórias sobre o gráfico de $\sigma_S \times p_\alpha$, gerado a partir da Eq.(5.22). Implementando os valores de k , χ , ϵ e ρ dados pela Eq.(5.1), temos que $p_{\max} = 0.54$ é o valor de probabilidade assintótica que corresponde ao maior valor de flutuação, $\sigma_{\max} = 9.71$. **(b)** A curva preta representa os valores assumidos por σ em função dos valores assumidos por P_α durante o processo transicional em AP = 41.5%. Os pontos azuis correspondem aos valores de σ e P_α assumidos nas classes temporais. Temos que inicialmente o sistema se encontra em um estado estacionário a com probabilidade inicial $p_\alpha^0 = 0.323$ e, devido à regulação externa, evolui para um estado assintótico b cuja probabilidade de encontrarmos o gene ligado é dada por $\Delta = 0.937$. Se a função q_α se mantiver constante por tempo suficiente, o sistema atingirá o estado estacionário especificado e veremos que a variação total da flutuação, $\Delta\sigma = \sigma_S(\Delta) - \sigma_S(p_\alpha^0)$, será negativa. Em $t_0(41.5)$, o sistema — com $P_\alpha \approx \Delta$ — tem seu efeito regulatório externo alterado bruscamente de modo que tem de evoluir para outro estado assintótico c cuja probabilidade de encontrarmos o gene ligado é dada por $\delta = 0.002$. Se tal estado for atingido, a variação total da flutuação será dada por $\Delta\sigma = \sigma_S(\delta) - \sigma_S(p_\alpha^0)$ e veremos que também será negativa. Podemos perceber, nos intervalos de C13 a T1 e de T5 a T6, P_α é consideravelmente alterado enquanto que a flutuação pouco se altera. Isso mostra que a regulação (ativação em C13 e repressão a partir de t_0) consegue, num primeiro momento, alterar enormemente a probabilidade do gene estar ativo sem que haja efetivamente uma variação considerável na flutuação. A trajetória gerada nos permite fazer duas observações: (i) o processo transicional não é quasi-estático (ou seja, ele não é uma sucessão de estados estacionários) e (ii) o sistema atinge em T5 o estado estacionário b ou, pelo menos, um estado consideravelmente próximo de b .

De fato, para compreender melhor essa dinâmica, podemos analisar o processo através da relação entre a flutuação σ e a probabilidade P_α . Em particular, podemos fazer uso da Eq.(3.38),

$$\sigma_S = \sqrt{N(1 - \chi)p_\alpha + N\chi + \frac{N^2(1 - \chi)^2 p_\alpha(1 - p_\alpha)}{1 + \epsilon}}, \quad (5.22)$$

que relaciona, para um sistema em regime estacionário, o desvio quadrático σ_S em função da probabilidade assintótica p_α . Como o estado inicial do sistema é considerado estacionário e como a especificação das taxas h e f (aqui nesse caso, representadas por q_α) especifica também o estado final (também estacionário) para o qual o sistema invariavelmente tende a evoluir, podemos, então, prever se a variação total de flutuação será

positiva ou negativa. Para isso, basta calcular a diferença $\Delta\sigma = \sigma_S(p_\alpha^f) - \sigma_S(p_\alpha^i)$, onde p_α^i e p_α^f são as probabilidades de o gene estar ativo nos estados assintóticos inicial e final, respectivamente. Além disso, a Eq.(5.22) nos diz que, se o processo de transcrição for quasi-estático (ou seja, se o processo de transcrição é uma sucessão de estados estacionários, atingidos via processos infinitesimais), então todos valores assumidos por σ se relacionarão com os valores assumidos por P_α segundo a própria Eq.(5.22), o que, graficamente, representará trajetórias sobre a curva $\sigma_S \times p_\alpha$, mostrada na Figura 5.3(a). O valor máximo que σ_S pode atingir, σ_{max} , ocorre para

$$p_{max} = \frac{1}{2} + \frac{1 + \epsilon}{2N(1 - \chi)}$$

e está representado na Figura 5.2 (linha rosa).

Como exemplo, analisemos a evolução temporal da flutuação em $AP = 41.5\%$, como mostra a Figura 5.3(b). O estado inicial, considerado estacionário *a* caracterizado por uma probabilidade de o gene estar ativo $p_\alpha^0(41.5) = 0.32$. Num primeiro momento (para $t \leq t_0(41.5) \approx 2154$ seg), a função q_α determina que o sistema evolua para um estado estacionário *b* caracterizado por uma probabilidade assintótica p_α dada por $\Delta(41.5) = 0.937$. Assim, se a função q_α se mantiver constante por tempo suficiente, o sistema atingirá o estado estacionário especificado (*b*) e, com isso, apresentará uma redução na flutuação dada por $\Delta\sigma = \sigma_S(0.937) - \sigma_S(0.32) = -2.41 < 0$, corroborando o comportamento observado na Figura 5.2. A partir de $t = t_0(41.5)$, o sistema, independentemente se atingiu ou não o estado estacionário especificado, passar a evoluir para outro estado estacionário, *c*, caracterizado pela probabilidade do gene estar ligado dada por $\delta(41.5) = 0.002$. Novamente, se a função q_α se mantiver constante por tempo suficiente, o sistema atingirá o novo estado estacionário (*c*) e a variação total de flutuação será dada por $\Delta\sigma = \sigma_S(0.002) - \sigma_S(0.32) = -8.37 < 0$, condizente com o observado na Figura 5.2. Esse argumento pode ser estendido a outras posições sobre o eixo AP para explicar as evoluções temporais e também a formação do “poço de flutuação”.

Ao coletarmos os valores assumidos por σ em função dos valores assumidos por P_α durante o processo transcricional, podemos gerar uma curva $\sigma \times P_\alpha$, mostrada na Figura

5.3(b) (curva preta). Através dessa curva, podemos ver que o sistema parte do estado estacionário a , segue em direção ao estado estacionário b e, devido à mudança brusca de q_α em t_0 , parte para o estado estacionário c . Os pontos azuis correspondem aos valores de σ e P_α assumidos nas classes temporais. Podemos perceber que, em intervalos específicos do processo como o intervalo de C13 a T1, o intervalo de T4 a T5 (estabilização de P_α) e o intervalo de T5 a T6, a flutuação σ é pouco alterada. Em especial, nos intervalos de C13 a T1 e de T5 a T6, P_α é consideravelmente alterado enquanto que a flutuação pouco se altera. Isso mostra que a regulação (ativação em C13 e repressão a partir de t_0) consegue, num primeiro momento, alterar consideravelmente a probabilidade do gene estar ativo sem que haja efetivamente uma variação considerável na flutuação. Comparada à curva $\sigma_S \times p_\alpha$, a trajetória $\sigma \times P_\alpha$ mostra claramente que os valores assumidos por σ e P_α não são valores compatíveis com a ideia de evolução via processos quasi-estáticos, ou seja, a transcrição não ocorre via uma sucessão de estados estacionários separados por processos infinitesimais. Ao mesmo tempo, a curva $\sigma \times P_\alpha$ mostra que, em T5, o sistema praticamente atinge o estado estacionário b ou, pelo menos, um estado consideravelmente próximo de b , o que condiz com a evolução mostrada na Figura 5.2 e representa mais uma evidência de que, em AP = 41.5%, o sistema atinge o regime estacionário em T5. É interessante observar também que tanto de C13 a t_0 quanto de t_0 a T8, a flutuação evolui de maneira similar em direção aos estados assintóticos estabelecidos (primeiro em C13, depois e em t_0): num primeiro momento, a flutuação atinge valores menores do que os valores estacionários e, comparada à probabilidade P_α , é pouco alterada; num segundo momento, o cenário de certo modo se inverte tal que os valores de flutuação atingidos são maiores que os valores estacionários e, comparada à P_α , sofre alterações consideráveis. Em suma, podemos observar os estados de equilíbrio objetivados pelo processo são atingidos por processos fora do equilíbrio e durante boa parte de sua evolução pode atingir valores de flutuação tanto maiores quanto menores do que os respectivos valores se fosse realizado em regime de equilíbrio.

Naturalmente, toda essa análise pode ser estendida para analisar a flutuação do sistema nas demais posições sobre o eixo AP. Em particular, um fato que merece destaque na

Figura 5.2 é que a flutuação é máxima nas bordas do “poço de flutuação”, $AP = 38.5\%$ e $AP = 43.5\%$, e permanece constante praticamente durante todo o tempo durante o qual é formada a faixa. Sua tendência em se manter praticamente constante se deve, sobretudo, à relativa proximidade entre as probabilidades iniciais assintóticas, $p_\alpha^0(38.5)$ e $p_\alpha^0(43.5)$, e entre os valores assumidos por q_α em cada posição. Contudo, podemos ver que os valores de flutuação atingidos ao longo do eixo AP no decorrer do processo transcricional são limitados superiormente por σ_{max} , o que sugere mecanismos de otimização até o momento não investigados. Acreditamos que isso está relacionado ao fato de termos considerado o estado inicial como estado estacionário. E, nesse sentido, podemos especular que, para um estado inicial arbitrário tal que a flutuação inicial $\sigma_0 > \sigma_{max}$, esperamos que observar, no plano $\sigma \times P_\alpha$, uma trajetória partindo de σ_0 e evoluindo invariavelmente para flutuações estacionárias $\sigma_S \leq \sigma_{max}$ de modo que a flutuação, uma vez atingindo um valor menor ou igual a σ_{max} , permanecerá sempre assumindo valores menores ou iguais a σ_{max} até se estabilizar. Em algum sentido, essa inferência se relaciona com a ideia de atratores. De qualquer forma, o aparecimento de um limite superior σ_{max} para a variância merece ser investigado com mais detalhes.

Como dissemos anteriormente, embora o valor médio e a flutuação forneçam informações relevantes o processo estocástico, é conveniente analisar também como se relacionam estas grandezas. Por isso, nosso próximo ponto de análise é o ruído, dado por $\sigma/\langle n \rangle$. As predições do nosso modelo para o ruído são mostradas pela Figura 5.4, que exhibe os perfis assumidos pelo ruído ao longo do eixo AP durante as classes temporais.

A principal característica apresentada pelo ruído é um processo de minimização local em torno de $AP = 41.5\%$ e em torno de $AP = 80.5\%$, herança da flutuação $\sigma(x, t)$. Contudo, há algumas diferenças interessantes que merecem destaque. A primeira está nos perfis exibidos em C13: enquanto que a flutuação, assim como o valor médio, exhibe um perfil decrescente no sentido anterior-posterior, o ruído exhibe um perfil decrescente no sentido posterior-anterior. A segunda delas é que o processo de minimização em torno de $AP = 80.5\%$ é completamente perceptível. Outra diferença reside nos domínios onde ocorrem a formação dos “poços” de ruído, que são maiores: em torno de $AP = 41.5\%$, o

Figura 5.4 – Evolução temporal do ruído, dado pela razão $\sigma/\langle n \rangle$. A principal característica apresentada pelo ruído é um processo de minimização local em torno de $AP = 41.5\%$ e em torno de $AP = 80.5\%$, herança da flutuação $\sigma(x, t)$. Diferente do processo apresentado pela flutuação, o processo de minimização em torno de $AP = 41.5\%$ ocorre em um domínio maior, chegando a se estender em T6 de $AP = 35.5\%$ até $AP \approx 50.5\%$, e também persiste por mais tempo, iniciando em T1 e persistindo até T8.

“poço de ruído” chega a se estender de $AP = 35.5\%$ até $AP \approx 50.5\%$ (em T6) e, em torno de $AP = 80.5\%$, o domínio chega a compreender de $AP \approx 70.5\%$ até $AP = 91.5\%$ (em T3). Por fim, outra diferença está na duração: enquanto que o processo de minimização da flutuação se inicia quase que imperceptivelmente em T1 e termina em T6, os perfis da minimização local de ruído se iniciam em T1 e persistem até T8.

Do ponto de vista biológico, se a minimização local da flutuação em torno de $AP = 41\%$ durante a formação da faixa 2 de mRNA é considerada de extrema relevância, o mesmo pode ser considerado para o processo de minimização local do ruído em torno de $AP = 41\%$ e as razões para tal são as mesmas. Definido como a razão $\sigma/\langle n \rangle$, o ruído reflete a ideia de variação percentual em relação ao valor esperado, ou, em outras

palavras, da precisão no processo de transcrição e o conceito de precisão em um processo transcricional pode estar associado ao mesmo tempo à ideia função biológica exercida pelo transcrito. Para isso, entendemos que uma molécula de mRNA transcrito terá função biológica enquanto estiver presente na célula (e ser detectada pelo ribossomo e posteriormete traduzida) e assumimos que deixará de ser importante para a célula a partir do momento em que a relação $\sigma / \langle n \rangle$ se tornar favorável à flutuação, ou seja, quando $\sigma / \langle n \rangle > 1$. Assim, a minimização de ruído vai ao encontro do esperado: aumento de precisão na região onde se forma a faixa de mRNA de *eve*, cuja expressão se insere no processo de segmentação do corpo do indivíduo adulto e, como tal, deve especificar com precisão a formação de subsequentes domínios e padrões.

Detalhadamente, a evolução dos perfis de ruído ocorre da seguinte maneira. Em C13, há um gradiente ao longo do eixo AP, que aumenta no sentido anterior-posterior. Em T1, iniciam-se os processos de minimização aos arredores de AP = 41.5% e AP = 80.5%, onde serão formados, respectivamente, a faixa 2 e o pequeno pico de expressão que contribuirá para a formação da faixa 7, e que serão intensificados até T5 e T7 respectivamente. Apesar de ambos serem intensificados, a minimização em torno de AP = 41.5% é muito mais intensa do que a minimização em torno de AP = 80.5% e isso se deve ao fato de que q_α ser muito mais próximo de 1 de AP = 41.5% do que de AP = 80.5%. As minimizações em torno de AP = 41.5% e de AP = 80.5% atingem seus auges em T5 e T7 respectivamente, exatamente quando os respectivos picos de expressão atingem maior intensidade. A partir desses instantes, os picos de expressão começam a diminuir e os processos de minimização reduzem suas intensidades mas os perfis de ruído ao longo do eixo AP continuam exibindo os “poços” de ruído até T8. Na região intermediária do eixo AP, o ruído apresenta um comportamento diferente, aumentando de C13 até T5 e se estabilizando depois. Em termos de precisão na formação da faixa, o processo apresenta $\sigma / \langle n \rangle < 1$ em torno de AP = 41.5% precisamente: de C13 a T3, entre AP = 35.5% e AP = 45.5%; em T4, entre AP = 36.5% e AP = 45.0%; em T5, entre AP = 38.0% e AP = 44.0%; em T6, entre AP = 38.5% e AP = 43.5%; e em T7, entre AP = 39.5% e AP = 43.0%. Daí, é imediato notar que em AP = 41.5%, mesmo após t_0 , o ruído atinge

valores menores que a unidade. Dentro de nosso entendimento, isso sugere que a repressão talvez possua uma importância biológica que não se resumiria somente a desativação de proteínas.

Figura 5.5 – Evolução temporal do ruído $\sigma/\langle n \rangle$ em AP = 35.5%, AP = 38.5%, AP = 41.5%, AP = 43.5%, AP = 45.5%, AP = 60.5% e AP = 80.5%. A comparação nos mostra nitidamente o processo de redução do ruído dentro dos domínios da faixa. Mais que isso: é na região onde se forma a faixa 2 onde o ruído do sistema atinge seu valor mínimo, o que vai ao encontro da ideia de precisão no processo transcricional ainda mais quando tal processo está inserido em um processo embrionário no qual esperamos que haja uma tendência dos sistemas biológicos em aumentar a precisão de processos justamente para formação de domínios e padrões relacionados a futuros tecidos. Nas posições AP = 38.5% e AP = 43.5% (curvas azuis), onde a flutuação permanecia estabilizada durante todo o tempo em que a faixa se formava, o ruído também é reduzido temporalmente mas a um valor ainda acima do valor atingido em AP = 41.5%. Nas posições AP = 35.5% e AP = 45.5% (curvas vermelhas), à base da faixa, o ruído mantém-se inicialmente constante, aumentando após T2-T3 em direção a uma estabilização. Em AP = 60.5% (curva verde), onde a transcrição é altamente reprimida (assim como em demais posições ao longo da região intermediária do eixo AP), vemos que o ruído atinge os maiores valores e isso se deve, sobretudo, ao fato de o número médio de moléculas transcritas ser pequeno nessa região. A linha rosa corresponde ao valor máximo que o ruído pode atingir quando o sistema estiver em regime estacionário.

Analogamente ao que fizemos para analisar a evolução temporal da flutuação em diversas posições, podemos compreender melhor a evolução temporal do ruído considerando o espaço $\sigma/\langle n \rangle \times P_\alpha$. Temos que se o processo de transcrição for uma sucessão de estados estacionários separados por processos infinitesimais, então os valores assumidos pelo ruído $\sigma/\langle n \rangle$ em função dos valores assumidos por P_α deverão reproduzir trajetórias sobre o gráfico de $\sigma_S/n_S \times p_\alpha$, onde simplesmente tomamos a Eq.(5.22) e a dividimos por $\langle n \rangle_S$, conforme mostra a Figura 5.6(a). Enquanto, na Figura 5.3(a), o gráfico de $\sigma \times P_\alpha$ em regime estacionário se assemelhava a uma parábola tal que os valores mínimos de flutuação, comparáveis entre si, correspondiam a p_α 0 ou 1, na Figura 5.6(a), vemos que, no regime estacionário, o valor mínimo para o ruído ocorre preferencialmente para $p_\alpha = 1$. Isso faz sentido porque, para o ruído ser baixo, o número de transcritos deve ser muito maior que σ e, para o número de transcritos ser grande, p_α deve ser 1.

Na Figura 5.6(b), comparamos o gráfico de $\sigma_S/n_S \times p_\alpha$ com os valores assumidos por $\sigma/\langle n \rangle$ em função dos valores assumidos por P_α durante o processo transcricional em AP = 41.5%, representados pela curva preta. Os pontos azuis correspondem aos valores de $\sigma/\langle n \rangle$ e P_α assumidos nas classes temporais. Vemos que parte de um estado estacionário a com probabilidade inicial $p_\alpha^0 = 0.323$ e evolui para um estado assintótico b cuja probabilidade de encontrarmos o gene ligado é dada por $\Delta = 0.937$. Se a função q_α se mantiver constante por tempo suficiente, o sistema atingirá o estado estacionário especificado no qual o ruído é muito próximo do valor mínimo possível. A partir de $t_0(41.5)$, o sistema passa a evoluir para outro estado assintótico c cuja probabilidade de encontrarmos o gene ligado é dada por $\delta = 0.002$. Comparativamente à Figura 5.3, vemos na Figura 5.6 que a substancial variação de P_α nos intervalos de C13 a T1 e de T5 a T6 causa, no primeiro intervalo, uma redução visível do ruído enquanto que, no segundo intervalo, o ruído pouco se altera. Isso mostra que impor subitamente o sistema a um efeitos ativadores ou reguladores intenso e constante (ativação em C13 e repressão a partir de t_0) garante uma variação considerável na probabilidade do gene estar ativo sem, no entanto, alterar o ruído. Isso significa que, num primeiro momento após a ativação ou repressão, as condições que dão possibilidade à transcrição podem ser enormemente alteradas sem que isso gere uma alteração substancial no ruído já existente. A trajetória gerada mostra que, de C13 a t_0 , o ruído apresentado pelo sistema é pouco maior do que o que seria apresentado caso o processo ocorresse via sucessão de estados estacionários e constata que, após t_0 , o ruído apresentado é não apenas consideravelmente menor do que o apresentado pela curva $\sigma_S/n_S \times p_\alpha$ mas também menor que a unidade, o que pode significar que mesmo a desativação da transcrição possua uma importância biológica além da próprio fato da transcrição reprimida. E podemos ver também que a trajetória praticamente atinge o ponto b ou, pelo menos, um estado consideravelmente próximo deste.

De modo geral, interessante observar que o ruído se comporta de modo bastante interessante durante o processo. Podemos ver que o ruído evolui, de C13 a t_0 , assumindo valores maiores que o ruído estacionário mas decrescendo monotonicamente até t_0 . Após

t_0 , o ruído passa a aumentar monotonicamente mas assumindo valores consideravelmente menores do que os valores estacionários e menores que a unidade durante praticamente todo o intervalo de tempo. Novamente, vemos que a transcrição de mRNA de *eve*, do modo como o modelamos, revela um processo relativamente preciso do ponto de vista dinâmico: porque enquanto a faixa é formada o ruído diminui e, após a desativação transcricional, o ruído consideravelmente menor do que os respectivos valores assintóticos e também menor que 1, alcançando esse valor somente no fim do processo, quando o número de transcritos é bastante próximo do nível basal.

Figura 5.6 – As evoluções temporais observadas na Figura 5.5 podem ser particularmente analisadas através do gráfico de $\sigma/\langle n \rangle$ em função de P_α . **(a)** Analogamente à Figura 5.3(a), temos que se o processo de transcrição for uma sucessão de estados estacionários separados por processos infinitesimais, então os valores assumidos pelo ruído $\sigma/\langle n \rangle$ em função dos valores assumidos por P_α devem reproduzir trajetórias sobre o gráfico de $\sigma_S/n_S \times p_\alpha$, gerado novamente a partir da Eq.(5.22) dividindo σ_S por $\langle n \rangle_S$. **(b)** A curva preta representa os valores assumidos por $\sigma/\langle n \rangle$ em função dos valores assumidos por P_α durante o processo transcricional em AP = 41.5%. Os pontos azuis correspondem aos valores de σ e P_α assumidos nas classes temporais. Temos que inicialmente o sistema se encontra em um estado estacionário a com probabilidade inicial $p_\alpha^0 = 0.323$ e evolui para um estado assintótico b cuja probabilidade de encontrarmos o gene ligado é dada por $\Delta = 0.937$. Se a função q_α se mantiver constante por tempo suficiente, o sistema atingirá o estado estacionário especificado no qual o ruído é muito próximo do valor mínimo possível. A partir de $t_0(41.5)$, o sistema — com $P_\alpha \approx \Delta$ — passa a evoluir para outro estado assintótico c cuja probabilidade de encontrarmos o gene ligado é dada por $\delta = 0.002$. Vemos que uma substancial variação de P_α nos intervalos de C13 a T1 e de T5 a T6 causa, no primeiro intervalo, uma redução visível do ruído enquanto que, no segundo intervalo, o ruído pouco se altera. Isso mostra que impor subitamente o sistema a um efeitos ativadores ou reguladores intenso e constante (ativação em C13 e repressão a partir de t_0) garante uma variação considerável na probabilidade do gene estar ativo sem, no entanto, alterar o ruído. A trajetória gerada mostra que, de C13 a t_0 , o ruído apresentado pelo sistema é pouco maior do que o que seria apresentado caso o processo ocorresse via sucessão de estados estacionários e, após t_0 , o ruído apresentado é consideravelmente menor do que o apresentado pela curva $\sigma_S/n_S \times p_\alpha$. Além disso, podemos ver que a trajetória praticamente atinge o ponto b ou, pelo menos, um estado consideravelmente próximo deste.

É importante mencionar que o fato do ruído na região intermediária do eixo AP ser alto em comparação ao observado em torno de AP = 41.5% e AP = 80.5% se deve muito mais ao número relativamente baixo de moléculas transcritas nessa região do que ao próprio

valor da flutuação.

Consideremos $AP = 60.5\%$ em T5. Nessa posição, temos $\langle n \rangle(60.5, T5) = 0.74$ e $\sigma(60.5, T5) = 2.12$. Isso significa que o valor médio de moléculas transcritas deve flutuar em cerca de 2 moléculas. Sob a perspectiva da própria natureza biológica da embriogênese de *Drosophila*, entendemos que não deve haver necessidade biológica de produção nessa região de mRNA regulado pelo *enhancer* 2 e que mesmo que haja a transcrição de algumas poucas moléculas ainda assim o processo de segmentação do embrião bem como a vida do indivíduo adulto não são comprometidos. Então, à luz dos resultados de nosso modelo, podemos inferir que, apesar de o ruído atingir os maiores valores na região intermediária, o organismo é capaz de tolerar a produção de 2 ou 3 moléculas em uma região onde a transcrição é fortemente reprimida. O mesmo argumento pode ser utilizado para analisar onde se estabelece o pico da faixa 2. Em $AP = 41.5\%$, o valor médio, a flutuação e, conseqüentemente, o ruído atingem seus valores extremos em T5: $\langle n \rangle(41.5, T5) = 26.86$, $\sigma(41.5, T5) = 6.77$ e $\sigma / \langle n \rangle = 0.25$. Assim, nosso modelo prevê que, em uma região onde são necessárias cerca de 27 moléculas, o sistema pode tolerar um excesso ou uma falta de até cerca de 7 moléculas. Essa análise pode ser melhor observada na Figura 5.7, onde são comparados sob diversos pontos de vista o valor médio com a flutuação e o ruído.

Na Figura 5.7, os gráficos 5.7(a) e 5.7(b) comparam os valor médio com flutuação e o ruído, respectivamente, durante T5, classe temporal em que a faixa está mais bem pronunciada. Na comparação entre valor médio e ruído, este último foi reescalado por um fator 10. As Figuras 5.7(a) e 5.7(b) chamam a atenção pelo fato de que o “poço de ruído” é mais largo que o “poço de flutuação”. Podemos observar que o “poço de ruído” compreende o intervalo entre $AP = 35.5\%$ e $AP = 48.5\%$ e atinge valores menores que a unidade entre $AP \approx 38\%$ e $AP \approx 44\%$. Lembrando, conforme vimos no Capítulo 2, que cada posição especificada no eixo AP corresponde a um núcleo dentre os 100 núcleos dispostos ao longo do eixo AP e que, portanto, os valores de $\langle n \rangle$ e σ especificadas para cada posição se referem a previsões do modelo para a transcrição em núcleos individuais, então as Figuras 5.7(a) e 5.7(b) indicam que a estocasticidade do sistema e a regulação externa permitem que haja não apenas um aumento de precisão na expressão do mRNA

Figura 5.7 – (a) Comparação entre o valor médio $\langle n \rangle$ e a flutuação σ ao longo do eixo AP em T5, classe temporal em que a faixa está mais bem pronunciada. (b) Comparação entre o valor médio $\langle n \rangle$ e o ruído $\sigma / \langle n \rangle$ ao longo do eixo AP em T5. Aqui, o valor médio está reescalado por um fator 10. Na comparação entre valor médio e ruído, este último foi reescalado por um fator 10. As Figuras 5.7(a) e 5.7(b) chamam a atenção pelo fato de que o “poço de ruído” é mais largo que o “poço de flutuação”. O “poço de ruído” compreende o intervalo entre AP = 35.5% e AP = 48.5% e o ruído atinge valores menores que a unidade entre AP $\approx$ 38% e AP $\approx$ 44%, o que indica que a estocasticidade do sistema e a regulação externa permitem que haja não apenas um aumento de precisão na expressão do mRNA em AP = 41.5% como também em posições compreendidas dentro do domínio de formação da faixa, implicando em uma especificação mais precisa da faixa como um todo, incluindo suas fronteiras. (c) Também em T5, comparação entre o perfil do valor médio e os perfis assumidos pelo mesmo valor médio quando acrescido e quando subtraído da flutuação, traduzindo a ideia de “faixa de tolerância” ao longo do eixo AP dentro da qual o número total de transcritos em cada posição pode variar sem comprometer o processo biológico. (d) Na mesma linha de raciocínio, a Figura 5.7(d) exhibe a evolução temporal de $\langle n \rangle$, $\langle n \rangle - \sigma$ e $\langle n \rangle + \sigma$. Interessante observar que a “espessura” de tal “faixa”, como sendo dada por $2\sigma(x, t)$, diminui de C13 a T5, o que indica um aumento de precisão durante a formação da faixa.

em AP = 41.5% como também em posições compreendidas dentro do domínio de formação da faixa, o que significa uma especificação mais precisa da faixa como um todo, incluindo suas fronteiras. Também em T5, os gráficos 5.7(c) comparam, ao longo do eixo AP, o perfil do valor médio com os perfis assumidos pelo mesmo valor médio quando acrescido

e quando subtraído da flutuação, o que nos fornece a ideia de uma “faixa de tolerância” dentro da qual o número total de transcritos em cada posição pode variar sem, no entanto, comprometer o processo biológico. Na mesma linha de raciocínio, a Figura 5.7(d) exibe a mesma comparação mas analisando essa como se dá a evolução temporal dessa “faixa” na posição $AP = 41.5\%$. É interessante observar que se a “espessura” de tal “faixa” é dada por $2\sigma(x, t)$, então é imedito ver que tal espessura diminui de C13 a T5, o que indica um aumento da supressão de flutuações e, portanto, um aumento de precisão durante a formação da faixa.

Nesse momento em que analisamos o valor médio, a flutuação e o ruído previstos, podemos notar que $\Delta(x)$ e $t_0(x)$ ou, mais precisamente, suas dependências espaciais têm um papel importante na regulação. Ao mesmo tempo que $t_0(x)$ indica por quanto tempo o estímulo transcricional durará em cada posição, $\Delta(x)$ quantifica o estímulo transcricional e, como vimos nas Figuras 5.3, determinará a flutuação no estado estacionário final, que será menor quanto mais próximo de 0 ou 1 os valores de $\Delta(x)$ e $\delta(x)$ estiverem. De modo coordenado pela função $q_\alpha(x, t)$, as dependências espaciais em $\Delta(x)$ e $t_0(x)$, que se assemelham à gaussianas centradas em torno de $AP = 41.5\%$, proporcionam não apenas o crescimento e afilamento em torno de $AP = 41.5\%$ observado na evolução dos perfis do número médio de transcritos (como mostra a Figura 4.16) mas também tempo suficiente para que, em torno de $AP = 41.5\%$, a faixa de mRNA adquira uma configuração estável e o sistema atinja um estado com flutuação e ruído minimizados e estabilizados (como mostra a Figura 5.8).

Se, por um lado, a flutuação σ pouco varia em comparação com o evolução temporal do valor médio $\langle n \rangle$, como mostra os gráficos 5.3(b) e 5.8(a), por outro lado, em termos de precisão do processo transcricional, podemos perceber que o ruído, assim como a flutuação, atinge T5 praticamente estabilizado e, após t_0 , começa a aumentar, como mostra os gráficos 5.6(b) e 5.8(b) e como também já observamos anteriormente. O fato de o ruído aumentar significa que a precisão do processo começa a diminuir. Contudo, como já havíamos observado também, o ruído ainda permanece menor que a unidade, sendo maior que esta somente em T8. Dada nossa expectativa de que processos durante a embriogênese

Figura 5.8 – (a) Evolução temporal do valor médio $\langle n \rangle$ e da flutuação σ em AP = 41.5%. (b) Evolução temporal do valor médio $\langle n \rangle$ (aqui reescalado por um fator 10) ao ruído $\sigma / \langle n \rangle$ em AP = 41.5%. Ambos os gráficos mostram que como o sistema se aproxima de um estado estacionário — taxas de variação temporal do valor médio $\langle n \rangle$ e flutuação σ praticamente nulas em T5 — com flutuação e ruído minimizados.

sejam precisos, esse resultado sugere que a própria repressão (após um período de intensa ativação), por se revelar relativamente precisa, talvez tenha uma importância biológica que não é somente a desativação da transcrição. Em outras palavras, isso pode significar que tão importante quanto transcrever mRNAs em locais e momentos desejados é não transcrevê-los em locais e momentos indesejados (ainda que haja uma certa faixa de tolerância).

Até o momento, tivemos oportunidade de verificar o efeito do regulador na formação e desaparecimento da faixa por meio de $\Delta(x)$, $\delta(x)$ e $t_0(x)$. Desde a concepção do modelo, já sabíamos que a especificação de $\Delta(x)$ e $\delta(x)$ determina o estado assintótico para o qual o sistema evoluirá e, além disso, verificamos nessa seção que quanto mais próximos de 0 ou 1 estiverem os valores assumidos por $\Delta(x)$ e $\delta(x)$ menor será a flutuação do sistema quando este atingir em regime estacionário, o que também já era previsto pelo modelo. Entretanto, em todas as análises realizadas (principalmente as realizadas nessa seção), o parâmetro ϵ foi mantido constante simplesmente porque representa uma característica intrínseca do sistema, indicando o quão rápido é o chaveamento em comparação com a degradação.

Entretanto, com o intuito de compreendermos mais sobre os efeitos da regulação externa sobre o processo transcricional, faremos algumas análises para diferentes valores de ϵ

mas preservando os parâmetros intrínsecos k , χ e ρ e as dependências espaciais $\Delta(x)$, $\delta(x)$, $t_0(x)$, $\Delta'(x)$, $\delta'(x)$, $t'(x)$ e $t_{ref}(x)$. Como dissemos, ϵ está intrinsecamente relacionado ao processo de chaveamento ou, mais especificamente, à dinâmica de ocupação e desocupação dos sítios de ligação pelas proteínas regulatórias. Portanto, introduzir variações em ϵ é uma maneira interessante de verificarmos, sob a perspectiva do modelo, as contribuições do chaveamento para as dinâmicas do valor médio e da flutuação. De fato, já temos, através das Eqs.(5.5), (5.8), (5.11) e (5.14), um certo entendimento do efeito de ϵ sobre a probabilidade do gene estar ativo, o valor médio, a flutuação e o ruído. Entretanto, como já descrevemos os dados experimentais e discutimos nossas previsões para a flutuação e o ruído do processo, aproveitaremos o presente problema para visualizar melhor os efeitos que alterações em ϵ causam à transcrição. Para tanto, consideramos $\epsilon = 2.37$ e $\epsilon = 0.79$ (valores respectivamente 50% maior e 50% menor do que o valor de ϵ utilizado) e obtemos as evoluções temporais de P_α , $\langle n \rangle$, σ e $\sigma/\langle n \rangle$ nas posições AP = 35.5%, AP = 38.5%, AP = 41.5%, AP = 43.5%, AP = 45.5%, AP = 60.5% e AP = 80.5%. Os resultados são mostrados pela Figura 5.9.

Comparando os resultados apresentados na Figura 5.9 com o resultados apresentados nas Figuras 4.15(b), 4.17, 5.2 e 5.5, observamos que, com o aumento de ϵ , P_α , σ e $\sigma/\langle n \rangle$ estabilizam-se mais rápido. Isso decorre do fato do processo de chaveamento continuar se estabilizando mais rápido que o processo de degradação. Com a diminuição de ϵ (mais precisamente, para um valor entre 0 e 1), o processo de chaveamento estabiliza-se depois da degradação levando P_α , σ e $\sigma/\langle n \rangle$ a se estabilizarem mais tarde. Observamos que, embora $\langle n \rangle$ não se revele muito sensível às alterações em ϵ , o mesmo não ocorre com σ e, consequentemente, para $\sigma/\langle n \rangle$: com o aumento de ϵ , os valores atingidos pela flutuação ao longo do embrião diminuem e o inverso ocorre com a diminuição de ϵ , que faz com que a flutuação ao longo do embrião atinja valores maiores. A sensibilidade de σ a alterações em ϵ já era esperada devido não somente à presença de termos $\exp[-\epsilon\rho t]$ e $\exp[-\rho(1+\epsilon)t]$ mas também ao fato de que a grande maioria dos fatores são proporcionais a $1/(2-\epsilon)$ — termo $U(x)$ —, a $1/(1+\epsilon)$ — termos $D(x)$, $G(x)$, $R(x)$, $S(x)$ — e, sobretudo, proporcionais a $1/(1-\epsilon)$ — termos $B(x)$, $C(x)$, $E(x)$, $F(x)$, $S(x)$, $T(x)$, $U(x)$ e $V(x)$ —, como mostram

Figura 5.9 – Evolução temporal de P_α , $\langle n \rangle$, σ e $\sigma/\langle n \rangle$ para diferentes valores de ϵ . À esquerda, os resultados para ϵ 50% maior ($\epsilon = 2.26$). À direita, os resultados ϵ 50% menor ($\epsilon = 0.75$). Comparado às Figuras 4.15(b), 4.17, 5.2 e 5.5, podemos perceber que, com o aumento de ϵ , o processo de chaveamento estabiliza-se ainda mais rápido que o de degradação de modo que P_α , σ e $\sigma/\langle n \rangle$ estabilizam-se mais rápido. Com a diminuição de ϵ (mais precisamente, para um valor entre 0 e 1), o processo de chaveamento estabiliza-se depois da degradação de modo que P_α , σ e $\sigma/\langle n \rangle$ levam mais tempo para estabilizarem. Verificamos que, embora $\langle n \rangle$ não se revele muito sensível às alterações em ϵ , podemos ver claramente que o mesmo não ocorre para σ e, consequentemente, para $\sigma/\langle n \rangle$: com o aumento de ϵ , os valores atingidos pela flutuação ao longo do embrião diminuem e o inverso ocorre com a diminuição de ϵ , que faz com que a flutuação ao longo do embrião atinja valores maiores.

as Eqs.(5.5), (5.8), (5.11) e (5.14).

Do ponto de vista molecular, podemos analisar o efeito de ϵ comparando as escalas

temporais envolvidas. Da Eq.(5.1), temos que a escala temporal envolvida na degradação de um molécula é $T_\rho = 1/\rho = 382.6$ segundos enquanto que para a produção de uma molécula de mRNA a escala temporal é $T_k = 1/k = 13.2$ segundos quando o gene está ativo e $T_{k\chi} = 1/(k\chi) = 1042.4$ segundos quando o gene está reprimido. O valor de ϵ com o qual estamos trabalhando é $\epsilon = 1.57962$, de modo que a escala temporal associada ao chaveamento é $T_\epsilon = 1/(\epsilon\rho) = 242.2$ segundos o que indica que a dinâmica do chaveamento se estabiliza pouco mais rápido do que a dinâmica de degradação. A influência do chaveamento na flutuação pode ser entendida da seguinte maneira. Grosso modo, consideremos que T_ϵ é o tempo que o gene inicialmente ligado (desligado) leva para ser desligado (ligado) e voltar a ser ligado (desligado), T_ρ o tempo para que ocorra a degradação de uma molécula e T_k o tempo para que uma molécula nova de mRNA seja transcrita pela RNA-polimerase. Tendo em vista que T_k é muito menor em comparação às escalas de chaveamento e degradação, consideremos também dois chaveamentos distintos: um que leva um longo tempo para ligar e desligar o gene de modo que $T_\epsilon > T_\rho$ e outro que liga e desliga o gene bem menos tempo de modo que $T_\epsilon < T_\rho$. No primeiro caso, temos que, durante o longo tempo em que o gene permanece desligado, nada é produzido e, durante o longo tempo em que permanece ligado, muitas moléculas são transcritas de modo que, ao longo do tempo haverá intervalos bem definidos apresentando total atividade e outros com total inatividade transcricional. Além disso, como $T_\epsilon > T_\rho$, uma ou mais moléculas devem ser degradadas durante T_ϵ e o valor médio e o desvio quadrático ao longo do tempo nos são fornecidos a partir de um *ensemble* de sistemas como esse. No segundo caso, como o chaveamento leva menos tempo para ligar e desligar o gene, a tendência é que haja menos lacunas ao longo do tempo, isto é, de modo o número de moléculas se apresente de modo mais uniforme. Paralelamente, como $T_\epsilon < T_\rho$, moléculas só são degradadas depois de passado um intervalo de tempo T_ϵ de modo que um *ensemble* de sistemas como esse nos fornece um menor desvio quadrático ao longo do tempo. Isto é, para o mesmo instante de tempo, o segundo *ensemble* fornecerá uma flutuação menor.

As análises realizadas nas Figuras 5.2, 5.3, 5.5, 5.6 e 5.7 e 5.9 reúnem um conjunto de informações que nos permitem avaliar o papel da regulação externa na transcrição sob

a óptica de nosso modelo. Contudo, é conveniente relembrarmos resumidamente como a regulação externa aparece em nosso modelo. Primeiramente, a regulação externa aparece por meio das taxas h e f , que representam as taxas com que o gene executa as transições ligado $\rightarrow$ desligado e desligado $\rightarrow$ ligado respectivamente. Estas taxas, uma vez fixadas, especificam as probabilidades assintóticas do gene estar desligado e ligado ($h/(h+f)$ e $f/(h+f)$, respectivamente) e, da comparação com a taxa de degradação ρ , é definido um parâmetro adimensional ϵ ,

$$\epsilon = \frac{h+f}{\rho},$$

que indica quantas vezes o chaveamento é mais “rápido” que a degradação e que é considerado uma característica intrínseca do sistema. Com a introdução de uma dependência temporal nas taxas h e f seguidas do advento das Eqs. (4.1) e (4.5) e também com a subsequente introdução de dependências espaciais, passamos a escrever as taxas em termos de uma função $q_\alpha(x, t)$ segundo a relação

$$f(x, t) = \frac{q_\alpha(x, t)}{\epsilon \rho},$$

onde tal relação nada mais é do que uma generalização no sentido de que $q_\alpha(x, t)$ representa, no caso de permanecer constante, a probabilidade assintótica do gene estar ligado. Lembrando que utilizamos otimizações para a obtenção de vários resultados, é imediato notar que trocar $h(x, t)$ e $f(x, t)$ por ϵ e $q_\alpha(x, t)$ trouxe benefícios: enquanto que as taxas $h(x, t)$ e $f(x, t)$ podem, a princípio, assumir quaisquer valores não-negativos, a função $q_\alpha(x, t)$ é limitada entre 0 e 1 e o parâmetro ϵ , considerado intrínseco, é invariante no tempo e no espaço. Além disso, trabalhar somente com essas grandezas nos fornece uma ideia melhor do regulador externo: q_α mostra o quão provável é a ocorrência de configurações favoráveis à transcrição e ϵ mostra, grosso modo, o quão rápido é (em comparação com a degradação) realizar a sequência de transições ligado $\rightarrow$ desligado $\rightarrow$ ligado.

Nesse sentido, nossos resultados mostram que enquanto $\Delta(x)$ e $\delta(x)$ tem um papel decisivo na determinação dos valores atingidos pelo número médio de moléculas transcritas e na determinação da flutuação do sistema quando atingir o estado assintótico, o parâmetro ϵ é importante para a estabilização de todas as dinâmicas envolvidas e, sobretudo, para

a magnitude da flutuação e do ruído no decorrer do processo.

Reconsiderando as Figuras 5.3 e 5.6, vimos que o processo transcricional não ocorre segundo transformações infinitesimais entre sucessivos estados estacionários e imaginamos que, se assim fosse e não houvesse regulação externa (“guiando”o sistema para um estado estado estacionário), certamente a transcrição levaria um tempo imensamente maior para ocorrer. Nesse sentido, podemos interpretar o regulador externo como um “catalisador transcricional”, que torna a transcrição mais rápida do que seria caso fosse uma sucessão estados estacionários. Ao mesmo tempo, observamos que o efeito regulatório, assim que imposto ao sistema, consegue alterar consideravelmente a probabilidade do gene estar ativo sem, no entanto, causar alterações percentualmente comparáveis à flutuação. Ademais, na Figura 5.6, vemos que a repressão instaurada a partir de t_0 faz com que a evolução para o novo estado assintótico ocorra apresentando um ruído não somente menor do que o apresentado por uma sucessão de estados estacionários (entre os mesmos estados) mas também menor que a unidade, o que pode significar que mesmo a repressão possua uma importância biológica no processo.

Além disso, ao compararmos as Figuras 4.16, 5.1 e 5.4, nossos resultados indicam que a ocorrência de “poços de flutuação” depende da ocorrência de picos de expressão e de uma probabilidade assintótica p_α próxima de 1 (pois implicará em um maior número de transcritos e em uma menor flutuação) ao passo que, para a formação de “poços de ruído”, basta a ocorrência de picos na expressão e, uma vez formado, sua “profundidade” dependerá da probabilidade P_α , que evolui em direção ao valor assintótico p_α , dado por $f/(h + f)$ (quanto mais próxima de 1 p_α for, maior o número de transcritos, menor a flutuação e, conseqüentemente, menor o ruído). Isso poderia ser, de certo modo, antevisto nas Figuras 5.6, ao notarmos que a curva $\sigma_S/n_S \times p_\alpha$. Assim, em última instância, as zonas de menor ruído se devem à regulação externa. Temos, então, que o sistema regido pela estocasticidade do nosso modelo naturalmente apresentará regiões de menor ruído em regiões onde houver maior número de transcritos. Logo, é natural, ao compararmos os perfis do valor médio e ruído, vemos zonas de menor ruído (os perfis de valor médio e ruído parecem ser, guardadas as devidas proporções, reflexos distorcidos um do outro).

Contudo, vale ressaltar que regiões de menor ruído não implicam necessariamente em um processo com alta precisão assim como que regiões de alto ruído não implicam em total descontrole regulatório: para uma melhor compreensão do cenário, o valor médio e a flutuação devem invariavelmente ser verificados bem como o valor do próprio ruído comparado ao valor da unidade (assim como fizemos em $AP = 41.5\%$ e em $AP = 60.5\%$ anteriormente). De qualquer forma, esta breve análise sugere que o modo de atuação do regulador externo, segundo nosso modelo, acaba por naturalmente reduzir o ruído. Na verdade, tanto essa inferência quanto a ideia de que o regulador seria um “catalisador transcricional”, que naturalmente emergem de nossos resultados, são questões que merecem uma atenção especial e uma investigação detalhada.

Uma outra questão interessante naturalmente surge acerca dos limites do parâmetro ϵ no contexto de controle de flutuação e ruído. Em processos como a transcrição de *eve* (que está relacionado com a formação de domínios e padrões durante a embriogênese), acreditamos que haja necessidade de que tais processos operem sob condições de controle máximo de ruído. De modo geral, embora nosso modelo seja relativamente simples, no sentido de simplificar e desconsiderar vários aspectos do fenômeno transcricional, os resultados obtidos evidenciam estratégias a fim de que a expressão gênica ocorra com controle e redução da variância e do ruído: evolução para um estado estacionário, probabilidade assintótica do gene estar ligado/desligado próxima de 0 ou 1 e ϵ com valor grande.

Dentre esses fatores, acreditamos que a evolução para estados estacionários e a ocorrência de probabilidades assintóticas próximas de 0 ou 1 são naturalmente possíveis no ambiente intracelular, não exigindo do meio intracelular condições de especificidade extrema. Isso automaticamente conduz nosso olhar para o parâmetro ϵ e, ao fazermos isso, nossa atenção novamente recai sobre a regulação externa ou, mais especificamente, sobre a pergunta: como deve ser uma proteína regulatória e sua interação com o DNA de modo que ϵ seja o maior possível e, conseqüentemente, o ruído seja minimizado ao máximo? Embora, em princípio, seja possível matematicamente o parâmetro ϵ assumir grandes valores, é preciso investigar, sob a perspectiva de nosso modelo, as implicações de ϵ assumir valores grandes. Apesar de não ser do escopo de nosso trabalho, podemos especular sobre

isso através de uma breve análise qualitativa.

Para isso, imaginemos o caso mais simples: um gene binário regulado por um único tipo de proteína e um único sítio de ligação de modo que as interações entre a proteína e o sítio possam ser resumidas em apenas uma única interação, conforme indica Eq.(4.5). Como o parâmetro ρ permanece inalterado, temos que se o valor de ϵ aumenta, o valor da soma $h + f$ aumenta e isso implica transições ligado $\rightarrow$ desligado e desligado $\rightarrow$ ligado mais rápidas. Ou seja, em nosso modelo, aumentar ϵ implica necessariamente em fazer a proteína regulatória se ligar e/ou se desligar do sítio de ligação em um tempo menor. De acordo com a definição de ϵ , isso pode se dar de três maneiras: (i) reduzindo o tempo que a proteína regulatória leva para se desligar do sítio, (ii) reduzindo o tempo que a proteína regulatória leva para se ligar ao sítio e (iii) reduzindo ambos os tempos. Na situação (i), se a proteína passa a se desligar do sítio cada vez mais rápido, então a interação entre DNA e proteína é cada vez mais fraca de modo que, num caso limite, a proteína praticamente não se liga e, conseqüentemente, não há efeito regulatório. Na situação (ii), se o tempo que a proteína leva para se ligar ao sítio é cada vez menor, isso significa que a interação entre a proteína e o DNA é cada vez mais forte de modo que, para uma interação suficientemente forte, a proteína se liga quase que instantaneamente e, ao eventualmente se desligar do sítio, a interação é forte o suficiente para refazer a ligação de modo que o efeito final é como se praticamente a proteína nunca se desligasse do sítio de ligação, gerando um regime de contínua ativação ou repressão, o que pode ser indesejado para a célula em determinadas circunstâncias. Na situação (iii), se ambos os tempos são reduzidos a zero, implica em dizer que a interação entre DNA e proteína é simultaneamente forte e fraca, o que é fisicamente contraditório: forte a ponto de a proteína se ligar ao sítio quase que instantaneamente e fraca a ponto de se desligar do mesmo sítio quase que instantaneamente também. Logo, temos que, em um sistema biológico real onde processos requerem considerável precisão, o valor de ϵ acabará sendo limitado superiormente em virtude da (i) própria necessidade de haver regulação, (ii) da necessidade de não haver ativação ou repressão continuamente e (iii) da própria natureza física das interações entre DNA e proteína. Portanto, sob a perspectiva do modelo, em um sistema biológico que

requiera uma transcrição relativamente precisa, o valor de ϵ deve corresponder a um valor *ótimo* de modo que as interações entre proteínas regulatórias e DNA devem ser fortes o suficiente para que as proteínas tenham a possibilidade de se ligar aos sítios de ligação relativamente rápido e, ao mesmo tempo, fracas o suficiente para que as proteínas possam se desligar dos sítios relativamente rápido, evitando ativação ou repressão contínua. Vale mencionar que essa análise sobre ϵ , ainda que escassa de análise matemática, estaria em acordo com o fato de as interações entre DNA e proteína serem fracas, de modo que as ligações das proteínas ao DNA são feitas e desfeitas constantemente(70). Ou seja, essa breve especulação sugere que a ocorrência de interações fracas como forma de mediação entre DNA e proteínas regulatórias poderia ser interpretada como resultado de um processo de otimização (que garante uma regulação eficiente e evita a ativação ou repressão contínua) e, do ponto de vista evolutivo, poderia ser interpretada como a forma de mediação selecionada naturalmente. Embora bastante escassa de análise matemática e de aspectos importantes como a quantidade total de proteínas (diretamente relacionada à composição das taxas h e f), essa especulação — consequência direta da invariância de ϵ e do subsequente atrelamento de h e f — instiga questões importantes acerca da natureza do chaveamento e certamente merece uma investigação muito mais cautelosa e detalhada.

5.2 Distribuição de probabilidade

Obtidos os parâmetros intrínsecos e as funções que compõem a função $q_\alpha(x, t)$, estamos em plenas condições de analisar a distribuição de probabilidade total ϕ_n . Contudo, como estamos focados essencialmente na dinâmica transcricional em $AP = 41.5\%$, onde a faixa de mRNA de *eve* atinge seu ponto de máxima expressão, trataremos de analisar a distribuição de probabilidade total somente nessa posição. Uma vez que

$$\phi_n(t) = \alpha_n(t) + \beta_n(t),$$

basta tomarmos as distribuições $\alpha_n(t)$ e $\beta_n(t)$, dadas pelas Eqs.(3.76) e (3.77), respectivamente.

Entretanto, antes de trabalharmos a evolução temporal de $\phi_n(t)$ propriamente dita, é preciso lembrar um importante detalhe. No Capítulo 3, escolhemos as condições iniciais $C_1(z, \tau)$ e $C_2(z, \tau)$, dadas pelas Eqs.(3.75), tais que o sistema se encontrasse em regime estacionário. Sendo assim, a distribuição de probabilidade $\phi_n(t)$ obtida a partir da função geratriz $\phi(z, \tau)$, dada pela Eq.(3.74), satisfaz esta condição inicial.

Isso significa que, em princípio, podemos utilizar a $\phi_n(t)$ para descrever a evolução temporal somente de C13 até T5. Para descrever a distribuição de probabilidade de T6 a T8, é preciso averiguar primeiro se em t_0 o sistema atinge, de fato, o regime estacionário conforme sugere uma série de resultados mostrados ao longo desse trabalho. Para tanto, comparamos a distribuição de probabilidade dependente do tempo calculada em $t = t_0(x = 41.5)$ com a distribuição de probabilidade para o caso estacionário, dada pela Eq.(3.54). A comparação é exibida na Figura 5.10.

Figura 5.10 – Comparação entre distribuição de probabilidade dependente do tempo $\phi_n(t)$ calculada em $t = t_0(x = 41.5)$ com a distribuição de probabilidade para o caso estacionário, dada pela Eq.(3.54). A distribuição $\phi_n(t)$ é dada por $\phi_n(t) = \alpha_n(t) + \beta_n(t)$, onde $\alpha_n(t)$ e $\beta_n(t)$ são dadas, por sua vez, pelas Eqs.3.76 e 3.77 respectivamente. A comparação mostra que a distribuição de probabilidade dependente do tempo $\phi_n(t)$ calculada em $t = t_0$ coincide com a distribuição de probabilidade para o caso estacionário, ou seja, o sistema em AP = 41.5% atinge o regime estacionário quando $t = t_0$.

Na Figura 5.10, vemos que a distribuição de probabilidade dependente do tempo $\phi_n(t)$ calculada em $t = t_0$ coincide com a distribuição de probabilidade para o caso estacionário, ou seja, o sistema em AP = 41.5% atinge o regime estacionário quando $t = t_0$, o que vai

ao encontro de todas as observações feitas no decorrer desse trabalho (de que o sistema em T5 poderia estar em regime estacionário ou muito próximo deste). Isso significa que podemos utilizar $\phi_n(t)$ para avaliar a distribuição de probabilidade de T6 a T8 tomando como condição inicial o estado do sistema em $t = t_0$, agora conhecidamente estacionário. De fato, ao realizarmos tal procedimento, obtemos a evolução temporal da distribuição de probabilidade total em $AP = 41.5\%$, conforme mostrado na Figura 5.11.

Figura 5.11 – Evolução temporal da distribuição de probabilidade total em $AP = 41.5\%$ durante o processo transcricional.

Detalhadamente, a distribuição de probabilidade evolui da seguinte maneira. Em C13, temos a distribuição de probabilidade inicial, associada a um estado estacionário de probabilidade $p_\alpha^0 = 0.35$, o que lhe confere um caráter bimodal sendo que o maior pico da distribuição encontra-se entre $n = 0$. Em T1, o caráter bimodal continua sendo exibido pela distribuição cujo máximo se encontra para n pouco menor que 20. Em T2, o caráter

bimodal, embora pouco expressivo, ainda é exibido e o pico da distribuição se encontra em torno de $n \approx 22$. De T3 a T5, a bimodalidade apresentada anteriormente torna-se cada vez menos expressiva enquanto que os perfis da distribuição de probabilidade sofrem cada vez menos alterações. De fato, isso evidencia não apenas a evolução para um estado estacionário como também o fato de este estado já estar relativamente “perto” de ser atingido. Entre T5 e T6, mais precisamente em t_0 , a função q_α estabelece um novo estado estacionário a ser atingido e o que vemos em T6 é a distribuição de probabilidade evoluindo para este novo estado. De T6 para T7, a distribuição volta a apresentar bimodalidade e o pico da distribuição passa a se encontrar em n pouco menor que 10. Em T8, cada vez mais próximo do novo estado estacionário, o caráter bimodal novamente volta a desaparecer e o pico de distribuição, que se deslocou, encontra-se em torno de $n = 5$.

O que vemos no decorrer da evolução da distribuição de probabilidade especialmente de C13 a T2 e de t_0 a T7 é o acoplamento de dois processos estocásticos com escalas temporais comparáveis: a degradação e o chaveamento. Isso é refletido não apenas na localização dos máximos da distribuição mas também no aparecimento de um caráter bimodal e no aumento do desvio quadrático.

Como já pudemos notar, a distribuição $\phi_n(t)$ apresenta traços de bimodalidade em alguns momentos, especificamente, de C13 a T2 e depois em T7. Tais traços de bimodalidade são efeitos sentidos pelo sistema causados pela ativação e da desativação repentinas em C13 e em t_0 respectivamente e ocorrem porque as duas escalas temporais características do sistema, T_ρ e T_ϵ , são comparáveis entre si. De fato, como vimos na seção anterior, uma vez que $\epsilon = 1.57962$, temos que $T_\rho = 1/\rho = 382.6$ segundos e $T_\epsilon = 1/(\epsilon\rho) = 242.2$ segundos: a primeira pode ser entendida como o tempo para a degradação de uma molécula de mRNA e a segunda pode ser entendida como o tempo para que o estado do gene execute a sequência de transições ligado $\rightarrow$ desligado $\rightarrow$ ligado (ou desligado $\rightarrow$ ligado $\rightarrow$ desligado). Molecularmente, isso significa, grosso modo, que se num intervalo de tempo T (em que o gene sai do estado inativo, vai para o estado ativo e retorna para o estado inativo) uma molécula de mRNA é transcrita, durante praticamente esse mesmo intervalo de tempo uma molécula de mRNA é degradada. Isso explica por que a distribuição de

probabilidade apresente um caráter bimodal. Contudo, esse cenário ainda pode ser alterado de acordo com o grau de ativação do gene, ou seja, de acordo com sua probabilidade de estar no estado ativo ou inativo: se a probabilidade de o gene estar no estado ativo é alta ($P_\alpha > 0.5$), então, a maquinaria transcricional tende a se ligar com mais frequência ao promotor e, superando o efeitos de repressão e de degradação, aumenta o número médio de moléculas de mRNA e, portanto, faz prevalecer, na distribuição de probabilidade, o pico centrado em torno de N ; por outro lado, se a probabilidade de estar no estado ativo é pequena ($P_\alpha < 0.5$), a maquinaria transcricional tende a se ligar menos vezes ao promotor e, portanto, o número de moléculas mRNA tende a ser resultante da regulação negativa e da degradação. Isso explica porque um dos picos — após tempo T suficientemente grande (isto é, muito maior que T_ρ e T_ϵ) — tende a desaparecer quando o sistema evolui para o regime estacionário com probabilidade assintótica perto de 0 ou 1.

Com relação ao aumento na flutuação ocorrendo de C13 a T2 e de t_0 a T7, quando a bimodalidade é mais pronunciada, vale lembrar que, na seção anterior (mais precisamente, ao analisarmos a Figura 5.2), argumentamos que tal aumento ocorria devido à imposição de um novo estado assintótico, o que significava a imposição de um novo ritmo de ocupação e desocupação dos sítios por parte das proteínas regulatórias, destruindo as configurações do sistema a fim de que outras fossem assumidas. Contudo, esse resultado pode ser melhor analisado sob a perspectiva da evolução da distribuição de probabilidade. Quando os dois picos da distribuição de probabilidade estão comparáveis após C13 e após t_0 , significa que há dois modos de transcrição simultaneamente, um operando com alta taxa de produção e outro com baixa, não refletindo, portanto, uma coesão. Tal coesão passa a ser adquirida à medida que a dinâmica do chaveamento se estabiliza e à medida que o sistema evolui para um regime (estacionário) com probabilidade assintótica próxima de 0 ou 1: se for perto de 0, significa que, num *ensemble*, é mais fácil encontrar genes inativos e o número de genes operando a taxas altas é o menor possível, o que faz com que o número de transcritos não desvie muito do número médio; se for perto de 1, situação oposta, é muito mais fácil encontrar genes ativos do que inativos, o que faz com que o número de transcritos não desvie muito do número médio. Assim, quando a distribuição de probabilidade evolui

para um estado estacionário com probabilidade assintótica próxima de 0 ou de 1, duas coisas acontecem: um pico prevalece e a variância diminui.

5.3 A função q_α e os gradientes de proteínas: um *ansatz*

Uma questão que naturalmente surge neste trabalho é como relacionar os parâmetros do modelo com as proteínas regulatórias. De maneira mais específica, nosso modelo mostra que todo o controle espacial da expressão de mRNA de *eve* pode ser modulado exclusivamente pela regulação externa, representada originalmente pelas taxas h e f . Uma vez que ϵ foi assumido invariante, a atenção se volta para a função q_α (versão reescalada de f) e seu comportamento durante a transcrição da faixa 2. Em especial, a função q_α reúne em si não somente o efeito regulatório externo como também é o objeto que representa a probabilidade — gerada pelo efeito regulatório — do estado assintótico para o qual o sistema evolui.

Nesse tese, trabalhamos com um perfil de $q_\alpha(x, t)$ capaz de recuperar os dados experimentais da transcrição de mRNA com boa concordância. No entanto, por maior que seja a concordância do modelo com o dado experimental, a função $q_\alpha(x, t)$ ainda carece de suporte biológico. Em outras palavras, nós não sabemos o que ou qual objeto biológico seria quantitativamente responsável por $q_\alpha(x, t)$. Como já discutimos anteriormente, podemos conceber a função $q_\alpha(x, t)$ como uma combinação das proteínas regulatórias envolvidas na expressão da faixa 2 de mRNA de *eve*. De acordo com todos os trabalhos que pesquisamos até o momento, são listadas sete proteínas cujas participações são conhecidas ou, pelo menos, evidenciadas direta ou indiretamente (através de experimentos ou cálculos computacionais): as ativadoras Bicoid (Bcd), Hunchback (Hb) e Caudal (Cad) e as repressoras Giant (Gt), Krüppel (Kr), Knirps (Kni) e Tailless (Tll). Ao mesmo tempo, as referências na literatura apontam para até 34 o número total de sítios de ligação envolvidos na

expressão da faixa 2.

Conforme discutimos no Capítulo 3, a ideia é construir uma função em termos das frações ocupacionais dos sítios de ligação que se comporte como a função $q_\alpha(x, t)$ construída a partir das otimizações realizadas no Capítulo 4. Essa construção desta função é compreensível: como há uma relação direta entre a ocupação de um sítio por uma proteína ativadora (repressora) e consequente ativação (repressão) transcricional, então a fração ocupacional de um sítio (ou, em outras palavras, a probabilidade de este sítio estar ocupado) nos fornece diretamente a sua contribuição para a probabilidade de que o gene esteja ligado (desligado), que é a própria $q_\alpha(x, t)$. É importante mencionar que para a concepção dessa fórmula não consideraremos detalhadamente vários aspectos envolvidos na interação entre uma proteína e DNA como, por exemplo, tamanho de cada proteína, especificação dos domínios proteicos que se ligam ao DNA e a interação da proteína com a RNA-polimerase ou outros fatores de transcrição.

Matematicamente, representaremos as frações ocupacionais por funções de Hill, usualmente utilizadas no contexto de interação proteína-ligantes (em que as frações ocupacionais são funções das concentrações dos ligantes que se ligam à proteína) e frequentemente utilizadas para descrição de ocupâncias de sítios reguladores(32, 149, 150). Também discutimos que, dado o caráter “tudo ou nada” do modelo, a construção da função q_α se estabelece no sentido de buscarmos qual configuração contribui decisivamente para a transcrição e não em discriminarmos configurações e respectivas contribuições (o que pode tornar desnecessário utilizar todos os sítios relacionados). Uma vez que uma mesma proteína pode ter afinidades diferentes com diferentes sítios de ligação, nossa atenção também se volta para investigar se a informação transcricional está concentrada em alguns (poucos) sítios. Outro aspecto a ser observado diz respeito à *lógica* dos de ligação: por exemplo, dado um conjunto de sítios de ligação, é preciso averiguar se a ativação (repressão) depende da ocupação (desocupação) simultânea dos sítios de ativadores (repressores), ou se apenas um regulador é suficiente para que a transcrição ocorra ou se são necessárias outras combinações entre ativadores e repressores(151–158). De qualquer forma, para podermos propor alguma função q_α , é imprescindível lançar mão de resultados

experimentais conhecidos presentes na literatura.

Um resultado extremamente significativo foi apresentado por Small e colaboradores(29). Nesse trabalho, são identificados 11 sítios de ligação como componentes do *enhancer* que regula a expressão da faixa 2. São eles: 5 sítios (B1, B2, B3, B4 e B5) aos quais a proteína Bcd pode se ligar, 1 sítio (H3) para a proteína Hb, 3 sítios (G1, G2 e G3) para a proteína Gt e 2 sítios (K4 e K5) para a proteína Kr. Quanto à configuração espacial desses sítios, há quatro casos de superposição envolvendo os pares de sítios B1-K3, H3-G1, B4-G3 e B5-K5. Arnosti e colaboradores analisaram as contribuições de cada sítio(30). Mostraram que a ocupação simultânea dos sítios B1 e H3 (por Bcd e Hb, respectivamente) é indispensável e praticamente suficiente para a formação da faixa 2 de mRNA de *eve*. Além disso, Janssens e colaboradores (32) mostraram que embriões mutantes em *Kni* apresentam expressão na região intermediária do eixo AP e propuseram a repressão de *Tll* para conter a ativação de Hb na região posterior do eixo AP. Ao mesmo tempo, no trabalho de Cox (159) são apresentadas e discutidas regras heurísticas acerca da regulação transcricional em *E. coli* as quais acreditamos que podem ser trazidas para o contexto da regulação em *Drosophila*. Em especial, uma das regras afirma que quanto mais próximos os sítios de ligação estão um do outro, mais próxima do tipo AND será a lógica envolvida*.

Mesmo não sendo um resultado derivado de um estudo em *E. coli*, o resultado de Arnosti e colaboradores está em acordo com o resultado de Cox uma vez que as distâncias envolvidas entre os sítios listados são da ordem de poucas dezenas de pares de base (consideradas pequenas se comparadas às dimensões, em pares de base, da região regulatória como um todo, que tem cerca de 2 kb). Além disso, no trabalho de Janssens e colaboradores, na vizinhanças de B1, H3, G1 e K3 estão N3 e T5, sítios aos quais se ligam *Kni* e *Tll*. Sendo assim, esperamos que a expressão para q_α não só deva depender do produto das frações ocupacionais de B1 e H3 como também inclua, na forma de fatores

*Por exemplo, dados dois sítios independentes a poucas bases de distância um do outro, sendo que um sítio pode ser ocupado por uma proteína ativadora A e o outro pode ser ocupado por uma proteína repressora R, o nível de transcrição copiosa é máximo quando simultaneamente a proteína A se encontra ligada ao seu sítio e a proteína R se encontra livre. Matematicamente, isso será representado pelo produto das frações ocupacionais.

multiplicativos, as frações ocupacionais referentes a sítios N3 e T5. Implementando todas essas informações, propomos o seguinte *ansatz*:

$$q_\alpha^{TESTE} = \gamma \frac{\left(\frac{[Bcd]}{K_1}\right)^{n_1}}{1 + \left(\frac{[Bcd]}{K_1}\right)^{n_1} + \left(\frac{[Kr]}{K_2}\right)^{n_2}} \times \frac{\left(\frac{[Hb]}{K_3}\right)^{n_3}}{1 + \left(\frac{[Hb]}{K_3}\right)^{n_3} + \left(\frac{[Gt]}{K_4}\right)^{n_4}} \quad (5.23)$$

$$\times \frac{1}{1 + \left(\frac{[Kni]}{K_5}\right)^{n_5}} \times \frac{1}{1 + \left(\frac{[Tll]}{K_6}\right)^{n_6}},$$

onde γ é um fator de escala, as constantes K_1, K_2, K_3, K_4, K_5 e K_6 podem ser entendidas como as afinidades de cada sítio, os coeficientes n_1, n_2, n_3, n_4, n_5 e n_6 são os coeficientes de Hill de cada espécie, os termos $[Bcd]$, $[Hb]$, $[Gt]$, $[Kr]$, $[Kni]$ e $[Tll]$ representam respectivamente as concentrações de Bicoid, Hunchback, Giant, Krüppel, Knirps e Tailless que, naturalmente, apresentam dependência espacial e temporal. Especificamente, q_α^{TESTE} propõe que a ativação transcricional resulta da ocupação simultânea dos sítios B1 e H3 por Bicoid e Hunchback respectivamente e da não-ocupação, também simultânea, dos sítios N3 e T5 pelas proteínas Knirps e Tailless respectivamente. Em outras palavras, a função q_α^{TESTE} considera as configurações de ocupação/desocupação dos sítios listados e, através de uma lógica específica (AND), contabiliza o resultado como probabilidade assintótica do gene estar ligado. Em outras palavras, a ocupação dos sítios de ligação determina a probabilidade assintótica para a qual o sistema evoluirá.

Na verdade, trabalhamos com os gradientes de intensidade de fluorescência exibidos pela Figura 2.17 e, assim como fizemos para estimar o número de moléculas de mRNA, assumimos a intensidade de fluorescência sendo diretamente proporcional à concentração da proteínas (de modo que as constantes de proporcionalidade de cada gradiente são incorporadas nas constantes K_1, K_2, K_3, K_4, K_5 e K_6).

Com a função q_α^{TESTE} escrita em termos dos gradientes das concentrações, faremos uma otimização (via método dos mínimos quadrados) de q_α^{TESTE} tendo como referência a família de funções de Heaviside compostas pelos valores de $\Delta, \delta, t_0, \Delta', \delta', t'$ e t_{ref} obtidos das otimizações feitas no Capítulo 4 e que, posteriormente, compuseram a função $q_\alpha(x, t)$ dada pela Eq.(5.2).

A Figura 5.12 mostra o resultado da otimização, realizada pelo programa *Wolfram Mathematica 7.0*, apresentando os perfis assumidos por q_α^{TESTE} em cada classe temporal.

Os parâmetros são: $\gamma = 0.96$, $K_1 = 0.00870$, $K_2 = 3.51128$, $K_3 = 0.00589$, $K_4 = 10.545$, $K_5 = 2.90953$, $K_6 = 0.53737$, $n_1 = 4.83786$, $n_2 = 12.5333$, $n_3 = 2.66421$, $n_4 = 14.1316$, $n_5 = 3.44504$ e $n_6 = 49.7949$.

Figura 5.12 – Evolução temporal dos perfis assumidos ao longo do eixo AP pela função q_α^{TESTE} comparada a família de funções de Heaviside compostas pelos valores de Δ , δ , t_0 , Δ' , δ' , t' e t_{ref} gerados das otimizações realizadas no Capítulo 4 e que, posteriormente, compuseram a função $q_\alpha(x, t)$ dada pela Eq.5.2. Os parâmetros de q_α^{TESTE} são: $\gamma = 0.96$, $K_1 = 0.00870$, $K_2 = 3.51128$, $K_3 = 0.00589$, $K_4 = 10.545$, $K_5 = 2.90953$, $K_6 = 0.53737$, $n_1 = 4.83786$, $n_2 = 12.5333$, $n_3 = 2.66421$, $n_4 = 14.1316$, $n_5 = 3.44504$ e $n_6 = 49.7949$.

É imediato notar a concordância entre a função composta pelos valores otimizados (que compõem $q_\alpha(x, t)$) e a função $q_\alpha^{TESTE}(x, t)$: perfis marcados pelo pico de ativação transcricional em torno de $AP = 41.5\%$ e repressão praticamente total na maior parte do embrião, pelo processo de afilamento desse pico, pela alteração brusca entre T5 e T6 e subsequente desativação transcricional.

A concordância exibida pela função $q_\alpha^{TESTE}(x, t)$ sugere que não somente é possível reunir de forma compacta a informação regulatória como também indica que tal informação

pode estar concentrada em poucos sítios, do mesmo modo como mostram os resultados de Arnosti e colaboradores: no nosso caso, utilizamos apenas 4 sítios enquanto que, na literatura, chega a 34 o número total de sítios envolvidos(29,30,32). Além disso, a função $q_\alpha^{TESTE}(x, t)$ sugere que sejam seis tipos de proteínas, dentre todos os sete considerados na literatura(32), os envolvidos de forma substancial na transcrição da faixa 2. Especificamente, a ocupação simultânea de Bicoid e Hunchback em seus respectivos sítios e a não-ocupação de Knirps e Tailless são suficientes para gerar um efeito regulatório consideravelmente próximo do efeito gerado pela função $q_\alpha(x, t)$.

O desempenho de $q_\alpha^{TESTE}(x, t)$, em acordo com os resultados experimentais de Arnosti e colaboradores, indica que o efeito regulatório dos demais sítios envolvidos na expressão da faixa 2 talvez seja um “ajuste fino” da expressão, isto é, contribuindo para que a produção atinja números de transcritos ligeiramente maiores ou menores, auxiliando na delimitação de domínios da expressão ou permitindo — ainda que sejam sítios com afinidades masi baixas — que mRNA continue sendo produzido mesmo quando as contribuições ativadoras dos sítios principais estiverem inviabilizadas por mecanismos repressores quaisquer, o que confere uma certa flexibilidade ao processo transcricional(30,85). Nesse sentido, é provavelmente no uso de poucos sítios e de uma lógica regulatória extremamente simples onde deve residir a razão para as discrepâncias observadas em C13 e T1 e em AP = 80.5%: o fato de não termos realizado uma análise mais detalhada envolvendo mais sítios ou mesmo lógicas um pouco mais complexas pode ter custado a perda significativa de informação regulatória da parte inicial do processo e também do extremo posterior do embrião[†].

É notável também que uma combinação de funções com dependências espaciais e temporais utilizada para compor q_α^{TESTE} — gradientes de Bcd, Hb, Gt, Kr, Kni e Tll — apresente um comportamento consideravelmente semelhante ao da função q_α que regula a faixa 2, ou seja, um efeito regulatório total tal que somente uma pequena região em torno

[†]Assim como discutimos no Capítulo 4 acerca da otimização, é importante mencionar aqui também que, embora a função otimizada q_α^{TESTE} tenha apresentado um bom desempenho, não há garantia de que o conjunto de valores parametrizados obtidos seja o único capaz de otimizar q_α^{TESTE} e, no caso de não ser, também não sabemos se corresponde a um mínimo global no espaço de parâmetros. Novamente, a razão para isso é o fato de que não realizamos nenhuma análise mais detalhada da quantidade a ser otimizada.

de AP = 41.5% é ativada. Isso mostra que ainda que a função q_α^{TESTE} , do modo como foi concebida, não consiga reproduzir fielmente o comportamento de $q_\alpha(x, t)$, a concordância observada revela, em última análise, uma coerência entre o modelo estocástico (gene binário submetido à regulação externa) e a modelagem de q_α^{TESTE} em termos das concentrações de proteínas, ambas abordagens independentes. Tal coerência não só fortalece ambas as abordagens como abre caminho para novas perspectivas como o desenvolvimento de uma função $q_\alpha(x, t)$ dependente dos gradientes mais sofisticada e para a análise das contribuições individuais de cada sítio ou de cada fator de transcrição para o valor médio, flutuação e ruído, o que automaticamente pode ser relacionado a mutações. De modo geral, estamos diante de um rol de questões em aberto e que merecem ser investigadas com mais detalhes.

Como mencionamos anteriormente, um dos principais resultados do modelo estocástico desenvolvido nessa tese é mostrar que o padrão espacial e temporal da expressão gênica pode resultar unicamente da regulação externa. E escrever q_α^{TESTE} em termos dos gradientes de concentração de proteínas é interessante também porque estes são objetos biológicos observáveis experimentalmente.

Diante disso, caso seja possível escrever uma função q_α^{TESTE} dependente das concentrações de proteínas que seja capaz de reproduzir os dados experimentais tal como $q_\alpha(x, t)$ (dada pela Eq.(5.2)) faz, então teremos um suporte biológico para a abordagem estocástica proposta e nos permitirá afirmar que a função regulatória do modelo, $q_\alpha(x, t)$ (dada pela Eq.(5.2)), interpretada simultaneamente como “guia transcricional” e “interruptor gênico”, possa também ser, de fato, descrita matematicamente em termos da ocupação de sítios de ligação.

Capítulo 6

Conclusões e perspectivas

Nesta tese, desenvolvemos um modelo estocástico fenomenológico para a transcrição da faixa 2 do mRNA de gene *even-skipped* de *Drosophila melanogaster* tendo o número de moléculas de mRNA transcritas como variável estocástica e considerando um gene com dois níveis de ativação regulado externamente.

A partir da análise dos dados experimentais em $AP = 41.5\%$ (onde se formará o pico da faixa 2 de mRNA), observamos que o processo transcricional apresenta duas fases: a primeira caracterizada por um crescimento sigmoidal que se aproxima de um regime assintótico, e a segunda, marcada por um decréscimo rápido, pressupondo uma total desativação total. Diante disso, propusemos, então, que os termos de chaveamento apresentassem dependências temporais dadas por funções de Heaviside de modo que no primeiro momento o gene é ligado a uma taxa alta e, posteriormente, sofre uma mudança súbita de modo que passa a ser ligado a uma taxa baixa. Posteriormente, permitindo dependências espaciais somente aos termos referentes ao chaveamento, estendemos esse constructo ao restante do eixo AP e o resultado obtido descreveu os dados experimentais com boa concordância.

Vimos que dependência temporal proposta aos termos do chaveamento revelou-se fundamental para capturar a essência da regulação externa sem a necessidade de fenômenos difusivos e de transporte. Além disso, esse resultado mostra que é possível descrever o processo transcricional, marcado por várias interações entre várias espécies moleculares, em termos de processos de nascimento e morte somados a um processo de chaveamento

entre níveis de ativação. Em outras palavras, toda a complexidade do fenômeno pode ser reduzida por um gene binário e equações lineares. Particularmente, todos os níveis de ativação provenientes da complexa regulação de diversas proteínas e de até dezenas de sítios puderam ser simplificados por apenas dois níveis de ativação, mostrando que o caráter binário do gene é suficiente para fornecer uma descrição semiquantitativa do fenômeno.

Quanto ao desvio quadrático (flutuação) e o ruído, o modelo prevê a ocorrência de uma minimização da flutuação e do ruído em torno de $AP = 41.5\%$, simultaneamente à formação da faixa. Isso se mostra compatível com nossas expectativas sobre o fenômeno uma vez que é esperado que determinados processos — como os que ocorrem durante a embriogênese, por exemplo — requeiram uma produção com certa precisão. Vimos também que, em $AP = 41.5\%$, durante a formação da faixa, ocorre a redução da flutuação e do ruído (este último atingindo valores menores que a unidade) e, após a desativação transcricional, o ruído, embora apresentando um aumento, continua assumindo valores menores que a unidade. Além disso, o modelo também prevê que a ocorrência de mínimos de ruído está intimamente relacionada à formação de picos de expressão. O modelo também mostra que o processo de transcrição não ocorre através sucessão de estados estacionários separados por transformações infinitesimais. Contudo, no cálculo da distribuição de probabilidade, o modelo prevê que, em $AP = 41.5\%$, o regime estacionário é atingido antes da faixa começar a desaparecer. De modo geral, todas as predições feitas pelo modelo vislumbram um cenário biologicamente favorável para processos transcricionais que requeiram precisão.

Além disso, constatamos o efeito do chaveamento (ϵ) sobre as dinâmicas do valor médio, da flutuação e do ruído e, sobretudo, sua contribuição para a intensidade da flutuação e do ruído do processo. Vimos que a invariância de ϵ e o consequente atrelamento entre as taxas de chaveamento levantam questões importantes acerca das interações entre proteínas e DNA e, dadas as implicações físicas e biológicas discutidas, merecem uma investigação muito cautelosa e detalhada.

Também estudamos uma conexão entre parâmetros do modelo — mais precisamente,

os termos de chaveamento — e as proteínas regulatórias envolvidas na transcrição. Baseado em resultados da literatura, propomos uma função em termos dos gradientes das proteínas regulatórias. Na literatura, a regulação gênica da faixa 2 é atribuída a até sete proteínas (Bcd, Hb, Cad, Gt, Kr, Kni e Tll) e a até 34 sítios operadores. Utilizando seis proteínas (Bcd, Hb, Gt, Kr, Kni e Tll) e apenas quatro sítios operadores, propomos um *ansatz* (q_α^{TESTE}) cujo comportamento regulatório se assemelha muito à função regulatória presente no modelo (q_α , com a qual trabalhamos no desenvolvimento do modelo estocástico, no qual revelou-se peça-chave para que o modelo reproduzisse com grande concordância os dados experimentais). O sucesso de q_α^{TESTE} não somente sugere que a informação transcricional está concentrada na regulação de poucos sítios e como também revela uma coerência entre as duas abordagens, concebidas de maneira independente. Tal coerência não apenas fortalece ambas as abordagens como também abre caminho para o desenvolvimento de uma função dependente das concentrações mais sofisticada e para a análise das contribuições individuais de cada sítio ou de cada proteína para a expressão e assim fazer previsões para o caso de mutações.

Vale mencionar aqui que nosso trabalho possui uma conexão interessante com o trabalho apresentado por Alon(152–154) embora ambos os trabalhos tenham se desenvolvido de maneira independente. No trabalho de Alon, a atenção está voltada para padrões de regulação recorrentes em redes genéticas (*network motifs*) que, através de uma determinada lógica, combinam os produtos gênicos (*input*) para regular a expressão de outros produtos gênicos (*output*). Nesse contexto, a expressão de determinado gene (*input*) é ativada por um sinal biológico S que atua com intensidade constante durante um certo tempo e sofrendo uma anulação súbita. Durante o tempo em que o sinal se mantém constante, a expressão ocorre, tendendo a se estabilizar e, após a anulação súbita do sinal, a expressão decai. De fato, a conexão entre os trabalhos reside justamente na dependência temporal do sinal S e da função $q_\alpha(x, t)$ e no efeito gerado na expressão gênica. Entretanto, S e $q_\alpha(x, t)$ representam objetos diferentes: S representa uma molécula indutora ou algum estímulo enquanto que $q_\alpha(x, t)$ deriva dos termos de chaveamento (h e f), que representam taxas às quais o gene executa transições ligado $\rightarrow$ desligado e desligado $\rightarrow$

ligado.

Um ponto de extrema importância é a realização de experimentos que atestem ou invalidem o modelo estocástico apresentado bem como seus resultados e previsões. Nesse sentido, podemos enumerar vários aspectos do modelo que poderiam ser confrontados com experimentos. Primeiramente, seria interessante reproduzir todo o experimento de monitoramento da transcrição de mRNA a fim de obter dados mais precisos e, com isso, verificar se a evolução temporal pode ser descrita pelas funções utilizadas nesse trabalho ou se há a necessidade de introduzir dependências matematicamente mais sofisticadas. Mais especificamente, seria de grande relevância a realização de experimentos que fornecessem o número de transcritos de mRNA de *eve* ao longo do embrião durante o processo. De grande importância seria também a realização de experimentos que verificassem a ocorrência e o grau de importância de processos de difusão ou de transporte no processo de expressão da faixa 2. Outro ponto que merece destaque é o fato de que quase todos os parâmetros utilizados durante o desenvolvimento desse trabalho resultaram de estimativas baseadas em informações de outros genes. Nesse sentido, seria imprescindível a realização de experimentos que obtenham experimentalmente as taxas de produção (k) e degradação (ρ), o instante (t_0) a partir do qual a intensidade da faixa 2 começa a diminuir. Além disso, é imprescindível para a validação do modelo um monitoramento espacial e temporal da flutuação e do ruído tal qual foi realizado para a detecção indireta do número de partículas transcritas.

De maneira geral, os resultados obtidos nessa tese fornecem ferramentas para o estudo da associação dos processos de transcrição e tradução, visando uma maior compreensão da síntese proteica, e para o estudo sobre o papel de um gene na regulação de outros genes (ou, mais especificamente, como a flutuação e o ruído da expressão de um gene afetam a expressão de um segundo gene regulado pelo primeiro), representando uma estratégia para um maior entendimento de redes genéticas.

REFERÊNCIAS

- 1 1902 - Theodor Boveri (1862-1915) and Walter Sutton (1877-1916) propose that chromosomes bear hereditary factors in accordance with Mendelian laws. Disponível em: <http://www.genomenewsnetwork.org/resources/timeline/1902_Boveri_Sutton.php>. Acesso em: 24 nov. 2012.

- 2 1910 - Thomas Hunt Morgan (1866-1945) establishes the chromosomal theory of heredity. Disponível em: <http://www.genomenewsnetwork.org/resources/timeline/1910_Morgan.php>. Acesso em: 24 nov. 2012.

- 3 T. H. MORGAN – great experiments on little flies. Disponível em: <<http://www.mendelmuseum.muni.cz/en/exhibitions/t-h-morgan/>>. Acesso em: 22 nov. 2012.

- 4 THOMAS Hunt Morgan, genetic recombination, and gene mapping. Disponível em: <<http://www.nature.com/scitable/topicpage/thomas-hunt-morgan-genetic-recombination-and-gene-496>>. Acesso em: 22 nov. 2012.

- 5 RANCE, H. *The Present is the key to the past*: 365 topics in historical geology. QCC Press, 1999. F: The age of mammals - The synthetic theory of evolution. Disponível em: <<http://geowords.com/histbookpdf/f14.pdf>>. Acesso em: 10 fev. 2013.

- 6 REIF, W.; JUNKER, T.; HOBFELD, U. The synthetic theory of evolution: general problems and the German contribution to the synthesis. *Journal of Molecular Biology*, v. 3, n. 3, p. 318 - 356, 1961.

- 7 PERUTZ, M. F. Erwin Schrödinger's *What Is Life?* and molecular biology. In: KILMISTER, C. W. *Schrödinger: centenary celebration of polymath*. Cambridge: Cambridge University Press, 1987, Cap. 19, p. 234-251.

- 8 SCHRÖDINGER, E. *What is life?: the physical aspect of the living cell*. Cambridge University Press, 1944.

-
- 9 PTASHNE, M. *A genetic switch: phage lambda revisited*. New York: Cold Spring Harbor Laboratory Press, 2004.
 - 10 MONOD, J.; JACOB, F. Genetic regulatory mechanism in synthesis of protein. *Journal of Molecular Biology*, v. 3, n. 3, p. 318 - 356, 1961.
 - 11 MULLER-HILL, B. *The lac operon*. A short history of a genetic paradigm. Berlin/New York: Walter de Gruyter, 1996.
 - 12 GILBERT, S. F. *Developmental biology*. 6th ed. Sunderland: Sinauer Associates, 2000.
 - 13 INGHAM, P. W. The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature*, v. 335, p. 25 – 34, 1988. DOI:10.1038/335025a0.
 - 14 NÜSSLEIN-VOLHARD, C.; WIESCHAUS, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature*, v. 287, p. 795 - 801, 1980. DOI 10.1038/287795a0.
 - 15 CARROLL, S. B. Zebra patterns in fly embryos: activation of stripes or repression of interstripes? *Cell*, v. 60, n. 1, p. 9 – 16, 1990.
 - 16 PANKRATZ, M. J. et al. Gradients of *Krüppel* and *knirps* gene products direct pair-rule gene stripe patterning in the posterior region of the *Drosophila* embryo. *Cell*, v. 61, n. 2, p. 309 – 317, 1990.
 - 17 SMALL, S.; LEVINE, M. The initiation of pair-rule stripes in the *Drosophila* blastoderm. *Current Opinion in Genetic Development*, v. 1, n. 2, p. 255 – 260, 1991.
 - 18 HARDING, K. et al. Cross-regulatory interactions among pair-rule genes in *Drosophila*. *Science*, v. 233, n. 4767, p. 953 – 959, 1986.
 - 19 MACDONALD, P. M. et al. Isolation, structure, and expression of *even-skipped*: a second pair-rule gene of *Drosophila* containing a homeo box. *Cell*, v. 47, n. 5, p. 721 – 734, 1986.
 - 20 FRASCH, P. M. et al. Characterization and localization of the *even-skipped* protein of *Drosophila*. *The European Molecular Biology Organization Journal*, v. 6, n. 3, p. 749 – 759, 1987.

-
- 21 SCHROEDER, M. D. et al. Transcriptional control in the segmentation gene network of *Drosophila*. *PLOS Biology*, v. 2, n. 9, p. 1396 - 1410, 2004.
- 22 FRASCH, P. M.; LEVINE, M. Complementary patterns of *even-skipped* and *fushi tarazu* expression involve their differential regulation by a common set of segmentation genes in *Drosophila*. *Genes & Development*, v. 1, n. 9, p. 981 - 995, 1987.
- 23 GOTO, T.; MACDONALD, P.; MANIATIS, T. Early and late periodic patterns of *even skipped* expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell*, v. 57, n. 3, p. 413 - 422, 1989.
- 24 STANOJEVIC, D.; SMALL, S.; LEVINE, M. Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. *Cell*, v. 254, n. 5036, p. 1385 - 1387, 1991.
- 25 SMALL, S. et al. Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes & Development*, v. 5, n. 5, p. 827 - 839, 1991.
- 26 HARDING, K. et al. Autoregulatory and gap gene response elements of the *even-skipped* promoter of *Drosophila*. *The European Molecular Biology Organization Journal*, v. 8, n. 4, p. 1205 - 1212, 1989.
- 27 SMALL, S.; ARNOSTI, D.; LEVINE, M. Spacing ensures autonomous expression of different stripe enhancers in the *even-skipped* promoter. *Development*, v. 119, n. 3, p. 767 - 772, 1993.
- 28 SMALL, S.; BLAIR, A.; LEVINE, M. Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo. *Developmental Biology*, v. 175, n. 2, p. 314 - 324, 1996.
- 29 SMALL, S.; BLAIR, A.; LEVINE, M. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *The European Molecular Biology Organization Journal*, v. 11, n. 11, p. 4047 - 4057, 1992.
- 30 ARNOSTI, D. N. et al. The *eve* stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development*, v. 122, n. 1, p. 205-214, 1996.
- 31 REINITZ, J.; HOU, S.; SHARP, D. H. Transcriptional control in *Drosophila*. *ComplexUs*, v. 1, n. 2, p. 54 - 64, 2003.

-
- 32 JANSSENS, J. et al. Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even skipped* gene. *Nature Genetics*, v. 38, n. 10, p. 1159 - 1165, 2006.
- 33 FLYEX Database - A database of segmentation gene expression in *Drosophila*. Disponível em: <<http://urchin.spbcas.ru/flyex/>>. Acesso em: 20 jul. 2012.
- 34 BERG, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of Theoretical Biology*, v. 71, n. 4, p. 587 - 603, 1978.
- 35 BHALLA, U. S.; IYENGAR, R. Emergent properties of networks of biological signaling pathways. *Science*, v. 283, n. 5400, p. 381 - 387, 1999.
- 36 COOK, D. L.; GERBER, A. N.; TAPSCOTT, S. J. Modeling stochastic gene expression: implications for haploinsufficiency. *Proceedings of the National Academy of Sciences USA*, v. 95, n. 26, p. 15641 - 15646, 1998.
- 37 KENNEL, D.; RIEZMAN, H. Transcription and translation initiation frequencies of the Escherichia coli lac operon. *Journal of Molecular Biology*, v. 144, n. 1, p. 1 - 21, 1977.
- 38 KO, M. S. H. A stochastic model for gene induction. *Journal of Theoretical Biology*, v. 153, n. 2, p. 181 - 194, 1991.
- 39 MAHESHRI, N.; O'SHEA, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annual Review of Biophysics and Biomolecular Structure*, v. 36, p. 413 - 434, 2007. DOI 10.1146/annurev.biophys.36.040306.132705
- 40 PUTTE, P.; GOOSEN, N. DNA inversions in phages and bacteria. *Trends in Genetics*, v. 8, n. 12, p. 457 - 462, 1992.
- 41 RAJ, A.; VAN OUDERNAADEN, A. Single-molecule approaches to stochastic gene expression. *Annual Review of Biophysics and Biomolecular Structure*, v. 38, p. 255 - 270, 2009. DOI: 10.1146/annurev.biophys.37.032807.125928.
- 42 SIEGELE, D. A.; HU, J. C. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences USA*, v. 94, n. 15, p. 8168 - 8172, 1997.
- 43 VON DASSOW, G. et al. The segment polarity network is a robust developmental module. *Nature*, v. 406, n. 6792, p. 188 - 192, 2000.

-
- 44 DELBRÜCK, M. Statistical fluctuations in autocatalytic reactions. *Journal of Chemical Physics*, v. 8, n. 1, p. 120-124, 1940.
- 45 GILLESPIE, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, v. 22, n. 4, p. 403 – 434, 1976.
- 46 GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, v. 81, n. 25, p. 2340 – 2361, 1977.
- 47 McADAMS, H. H.; ARKIN, A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, v. 94, n. 3, p. 814-819, 1997.
- 48 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, 1999.
- 49 BECSKEI, A.; SERRANO, L. Engineering stability in gene networks by autoregulation. *Nature*, v. 405, n. 6786, p. 590 - 593, 2000.
- 50 ACKERS, G. K.; JOHNSON A. D.; SHEA, M. A. Quantitative model for gene regulation by λ phage repressor. *Proceedings of the National Academy of Sciences USA*, v. 79, n. 4, p. 1129 - 1133, 1982.
- 51 SHEA, M. A.; ACKERS, G. K. The OR control system of bacteriophage lambda: a physical-chemical model for gene regulation. *Journal of Molecular Biology*, v. 181, n. 2, p. 211 – 230, 1985.
- 52 ARKIN, A.; ROSS, J.; McADAMS, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, v. 149, n. 4, p. 1633 – 1648, 1998.
- 53 AURELL, E. et al. Stability puzzles in phage lambda. *Physical Review E*, v. 65, n. 5, p. 1 – 9, 2002.
- 54 SAIZ, L. The physics of protein-DNA interaction networks in the control of gene expression. *Journal of Physics-Condensed Matter*, v. 24, n. 19, p. 1 – 15, 2012.
- 55 GARDNER, T. S; CANTOR, C. R.; COLLINS, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, v. 403, n. 6767, p. 339 - 342, 2000.

-
- 56 HASTY, T. S et al. Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences*, v. 97, n. 5, p. 2075 - 2080, 2000.
- 57 PAULSSON, J.; BERG, O. G.; EHRENBERG, M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proceedings of the National Academy of Sciences*, v. 97, n. 13, p. 7148 - 7153, 2000.
- 58 ELOWITZ, M. B.; LEIBLER, S. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, v. 403, n. 6767, p. 335 - 338, 2000.
- 59 BLAKE, W. J. et al. Noise in eukaryotic gene expression. *Proceedings of the National Academy of Sciences*, v. 422, n. 6932, p. 633 - 637, 2003.
- 60 THATTAI, M.; VAN OUDERNAADEN, A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, v. 98, n. 15, p. 8614–8619, 2001.
- 61 OZBUDAK, E. M. et al. Regulation of noise in the expression of a single gene. *Nature Genetics*, v. 31, p. 120-124, 2002. DOI: 10.1038/ng869.
- 62 SASAI, M.; WOLYNES, P. G. Stochastic gene expression as a many-body problem. *Proceedings of the National Academy of Sciences*, v. 100, n. 5, p. 2374–2379, 2003.
- 63 HORNOS, J. E. M. et al. Self-regulating gene: an exact solution. *Physical Review E*, v. 72, n. 5, p. 1 - 5, 2005.
- 64 INNOCENTINI, G. C. P.; HORNOS, J. E. M. Modeling stochastic gene expression under repression. *Journal of Mathematical Biology*, v. 55, n. 3, p. 413 - 431, 2007.
- 65 INNOCENTINI, G. C. P. *Modelamento estocástico para a expressão gênica*. 2008. 53 p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 66 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, p. 062902, 2011.
- 67 RAMOS, A. F. et al. Symmetries in biology: from genetic code to stochastic gene regulation. *IET Systems Biology*, v. 4, n. 5, p. 311 - 329 2007.

-
- 68 RAMOS, A. F.; HORNOS, J. E. M.; Symmetry and stochastic gene regulation. *Physical Review Letters*, v. 99, n. 10, p. 108103, 2007.
- 69 CENTRAL dogma of molecular biology. Disponível em: <http://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology>. Acesso em: 29 nov. 2012.
- 70 WATSON, J. D. et al. *Molecular biology of the gene*. 5th ed. California: Benjamin/Cummings Publishing Company, 2004.
- 71 O'BRIEN, E. L.; ITALLIE, E. V.; BENNETT, M. R. Modeling synthetic gene oscillators. *Mathematical Biosciences*, v. 236, n. 1, p. 1 - 15, 2012.
- 72 VOET, D.; VOET J. et al. *Biochemistry*. 2nd ed. Somerset: John Wiley & Sons, 1995.
- 73 BJÖRKLUND, S.; GUSTAFSSON, C. M. The yeast Mediator complex and its regulation. *Trends in Biochemical Sciences*, v. 30, n. 5, p. 240 - 244, 2005.
- 74 REEVES, W. M.; HAHN, S. Activator-independent functions of the yeast mediator Sin4 Complex in preinitiation complex formation and transcription reinitiation. *Molecular and Cellular Biology*, v. 23, n. 1, p. 349 - 358, 2003.
- 75 JOHNSON, A. D. The price of repression. *Cell*, v. 81, n. 5, p. 655-658, 1995.
- 76 LEVINE, M.; MANLEY, J. Transcriptional repression of eukaryotic promoters. *Cell*, v. 59, n. 3, p. 405 - 408, 1989.
- 77 EUKARYOTIC genome complexity. Disponível em: <<http://www.nature.com/scitable/topicpage/eukaryotic-genome-complexity-437>>. Acesso em: 31 jul. 2012.
- 78 LANE, N. Energetics and genetics across the prokaryote-eukaryote divide. *Biology Direct*, v. 6, n. 1, p. 1 - 31, 2011.
- 79 ELGAR, G.; VAVOURI, T. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends in Genetics*, v. 24, n. 7, p. 344 - 352, 2008.
- 80 BANERJI, J.; RUSCONI, S.; SCHAFFNER, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, v. 27, n. 2, p. 299 - 308, 1981.

-
- 81 ALBERTS, B. et al. *Essential cell biology*. 4th ed. New York: Garland Science, 2004.
- 82 KLINGER, M. et al. Disperse versus compact elements for the regulation of *run* stripes in *Drosophila*. *Developmental Biology*, v. 177, n. 1, p. 73 - 84, 1996.
- 83 HALFON, M.S. (Re)modeling the transcriptional enhancer. *Nature Genetics*, v. 38, n. 10, p. 1102 - 1103, 2006.
- 84 YUH, C.H.; BOLOURI, H.; DAVIDSON, E.H. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science*, v. 279, n. 5358, p. 1896 - 1902, 1998.
- 85 ARNOSTI, D. N.; KULKARNI, M. M. Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, v. 94, n. 5, p. 890 - 898, 2005.
- 86 LAWRENCE, P. A. *The making of a fly: the genetics of animal design*. Oxford: Blackwell Scientific Publications, 1992.
- 87 JAEGER, J. et al. Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics*, v. 167, n. 4, p. 1721 - 1737, 2004.
- 88 LEWIN, B. *Genes*. 5th ed. Upper Saddle River: Pearson Prentice Hall, 2004.
- 89 JAEGER, J. D. The gap gene network. *Cellular and Molecular Life Sciences*, v. 68, n. 2, p. 243 - 274, 2011.
- 90 GRIFFITHS, A.; WESSLER, S.; LEWONTIN, R.; CARROLL, S. *Introduction to genetic analysis*. 9th ed. New York: W.H. Freeman & Company, 2008.
- 91 LUDWIG, M. Z. et al. Functional evolution of a cis-regulatory module. *PLOS Biology*, v. 3, n. 4, p. 588 - 598, 2005.
- 92 JIANG, J.; HOEY, T.; LEVINE, M. Autoregulation of a segmentation gene in *Drosophila*: combinatorial interaction of the *even-skipped* homeo box protein with a distal enhancer element. *Genes & Development*, v. 5, n. 2, p. 265 - 277, 1991.
- 93 MOLECULAR biology for masters. Disponível em: <http://mol-biol4masters.masters.grkraj.org/html/Developmental_Biology1-Drosophila.htm>. Acesso em: 7 jul. 2012.

-
- 94 GENE repórter. Disponível em:
<<http://www.icb.ufmg.br/biq/prodap/2000/gfp/generep.html>>. Acesso em: 9 jul. 2012.
- 95 WOOD, K. V. Marker proteins for gene expression. *Current Opinion in Biotechnology*, v. 6, n. 1, p. 50 – 58, 1995.
- 96 SPEICHER, M. R.; CARTER, N. P. The new cytogenetics: blurring the boundaries with molecular biology. *Nature Reviews Genetics*, v. 6, p. 782 – 792, 2005. DOI:10.1038/nrg1692.
- 97 LANDSTEINER, K. *The specificity of serological reactions*. Cambridge: Harvard Univeristy Press, 1945.
- 98 SHREDER, K. Synthetic haptens as probes of antibody response and immunorecognition. *Methods*, v. 20, p. 372 – 379, 2000. DOI:10.1006/meth.1999.0929.
- 99 FLUORESCENCE *In Situ* Hybridization (FISH). Disponível em:
<<http://www.nature.com/scitable/topicpage/fluorescence-in-situ-hybridization-fish-327>>. Acesso em: 26 jul. 2012.
- 100 FARIAS, C.H.; BOLOURI, H.; DAVIDSON, E.H. GFP: uma ferramenta brilhante para a visualização da vida. *Revista Virtual de Química*, v. 1, n. 1, p. 2 - 8, 2009.
- 101 FLUORESCENCE fundamentals. Disponível em:
<<http://www.invitrogen.com/site/us/en/home/References/Molecular-Probes-The-Handbook/Introduction-to-Fluorescence-Techniques.html>>. Acesso em: 10 jul. 2012.
- 102 HUGHES, S. C.; KRAUSE, H.M. Single and double FISH protocols for *Drosophila*. In: PADDOCK, S. W. *Confocal microscopy: methods and protocols*. New Jersey: Humana Press, 1998. Cap. 5, p. 93 - 101.
- 103 WU, X. et al. Thoracic patterning by the *Drosophila* gap gene *hunchback*. *Developmental Biology*, v. 237, n. 1, p. 79 - 92, 2001.
- 104 NAGASO, H. et al. Simultaneous detection of RNA and protein by in situ hybridization and immunological staining. *Journal Histochemistry & Cytochemistry*, v. 49, n. 9, p. 1177 - 1182, 2001.

- 105 KOSMAN, D.; SMALL, S.; REINITZ, J. Rapid preparation of a panel of polyclonal antibodies to *Drosophila* segmentation proteins. *Development Genes and Evolution*, v. 208, n. 5, p. 290 - 294, 1998.
- 106 JANSSENS, H. et al. A high-throughput method for quantifying gene expression data from early *Drosophila* embryos. *Development Genes and Evolution*, v. 215, n. 7, p. 374 - 381, 2005.
- 107 MYASNIKOVA, E. et al. Removal of background signal from in situ data on the expression of segmentation genes in *Drosophila*. *Development Genes and Evolution*, v. 215, n. 6, p. 320 - 326, 2005.
- 108 MYASNIKOVA, E. et al. Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics*, v. 17, n. 1, p. 3 - 12, 2001.
- 109 SURKOVA, E. et al. Pipeline for acquisition of quantitative data on segmentation gene expression from confocal images. *Fly*, v. 2, n. 2, p. 58 - 66, 2008.
- 110 SURKOVA, E. et al. Characterization of the *Drosophila* segment determination morphome. *Developmental Biology*, v. 313, n. 2, p. 844 - 862, 2008.
- 111 ELSNER, J.; TSONIS, A. *Singular spectrum analysis: a new tool in time series analysis*. New York: Plenum, 1996.
- 112 CAPOVILLA, M.; ELDON, E.D.; PIRROTTA, V. The giant gene of *Drosophila* encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development*, v. 114, n. 1, p. 99 - 112, 1992.
- 113 MERRILL, P. T.; SWEETON, D.; WIESCHAUS, E. Requirements for autosomal gene activity during precellular stages of *Drosophila melanogaster*. *Development*, v. 104, n. 3, p. 495 - 509, 1988.
- 114 RUBIN, G. M.; SPRADLING, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science*, v. 218, n. 4570, p. 348 - 353, 1982.
- 115 SPRADLING, A. C.; RUBIN, G. M. Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science*, v. 218, n. 4570, p. 495 - 509, 1982.

-
- 116 THUMMEL, C. S.; BOULET, A. M.; LIPSCHITZ, H. D. Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection. *Gene*, v. 74, n. 2, p. 445 - 456, 1988.
- 117 HARTENSTEIN, V.; JAN, Y. N. Studying *Drosophila* embryogenesis with P-*lacZ* enhancer trap lines. *Roux's Archives of Developmental Biology*, v. 201, n. 4, p. 194 - 220, 1992.
- 118 O'KANE, C. J.; GEHRING, W. J. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proceedings of the National Academy of Sciences USA*, v. 84, n. 24, p. 9123 - 9127, 1987.
- 119 BIER, E. et al. Searching for pattern and mutation in the *Drosophila* genome with a P-*lacZ* vector. *Genes & Development*, v. 3, n. 9, p. 1273 - 1287, 1989.
- 120 FINNEGAN, D. J. Transposable elements. *Current Opinion in Genetics and Development*, v. 2, n. 6, p. 861 - 867, 1992.
- 121 BELLEN, H. J. et al. P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes & Development*, v. 3, n. 9, p. 1288 - 1300, 1989.
- 122 HIROMI, Y.; KUROIWA, A.; GEHRING, W. J. Control elements of the *Drosophila* segmentation gene *fushi tarazu*. *Cell*, v. 43, n. 3, p. 603 - 613, 1985.
- 123 HOUSDEN, B. E.; MILLEN, K.; BRAY, S. J. *Drosophila* reporter vectors compatible with Φ C31 integrase transgenesis techniques and their use to generate new notch reporter fly lines. *G3 (Bethesda)*, v. 2, n. 1, p. 79 - 82, 2012.
- 124 JAEGER, J. et al. Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, v. 430, n. 6997, p. 368 - 371, 2004.
- 125 GURSKY, V. V. et al. Pattern formation and nuclear divisions are uncoupled in *Drosophila* segmentation: comparison of spatially discrete and continuous models. *Physica D*, v. 197, n. 3, p. 286 - 302, 2004.
- 126 BERMAN, J. et al. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proceedings of the National Academy of Sciences*, v. 99, n. 2, p. 757 - 762, 2002.

-
- 127 LIANG, A. et al. Visualization of individual *Scr* mRNAs during *Drosophila* embryogenesis yields evidence for transcriptional bursting. *Current Biology*, v. 19, n. 23, p. 2037 - 2042, 2009.
- 128 MÜLLER, J.; BIENZ, M. Sharp anterior boundary of homeotic gene expression conferred by the *fushi tarazu* protein. *The European Molecular Biology Organization Journal*, v. 11, n. 10, p. 3653 - 3361, 1992.
- 129 TREMML, G.; BIENZ, M. An essential role of *even-skipped* for homeotic gene expression in the *Drosophila* visceral mesoderm. *The European Molecular Biology Organization Journal*, v. 8, n. 9, p. 2687 – 2693, 1989.
- 130 CARROLL, S. B. et al. Temporal and spatial relationships between segmentation and homeotic gene expression in *Drosophila* embryos: distributions of the *fushi tarazu*, *engrailed*, *Sex combs reduced*, *Antennapedia*, and *Ultrabithorax* proteins. *Genes & Development*, v. 2, n. 3, p. 350 - 360, 1988.
- 131 MARTINEZ-ARIAS, A.; LAWRENCE, P. A. Parasegments and compartments in the *Drosophila* embryo. *Nature*, v. 313, p. 639 - 642, 1985. DOI:10.1038/313639a0
- 132 MAHOWALD, A. P.; HARDY, P. A. Genetics of *Drosophila* embryogenesis. *Annual Reviews of Genetics*, v. 19, p. 149 - 177, 1985. DOI: 10.1146/annurev.ge.19.120185.001053
- 133 FLYBASE A database of *Drosophila* genes & genomes. Disponível em: <<http://flybase.org/>>. Acesso em: 25 jul. 2012.
- 134 ROARK, F. et al. Blastoderm-differential and blastoderm-specific genes of *Drosophila melanogaster*. *Developmental Biology*, v. 109, n. 2, p. 476 – 488, 1985.
a
- 135 PARÉ, A. et al. Visualization of individual *Scr* mRNAs during *Drosophila* embryogenesis yields evidence for transcriptional bursting. *Current Biology*, v. 19, n. 23, p. 2037 - 2042, 2009.
- 136 BERGMAN, S. et al. Re-examining the stability of the Bicoid morphogen gradient. *Cell*, v. 132, n. 1, p. 15 – 17, 2008.
- 137 RAJ, A. et al. Stochastic mRNA synthesis in mammalian cells. *PLOS Biology*, v. 4, n. 10, p. 1707 – 1719, 2006.

-
- 138 INNOCENTINI, G. C. P. *Uma generalização do modelo de spins e bósons para a transcrição de genes sob múltiplo controle.* 2012. 77 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.
- 139 VAN KAMPEN, N. G. *Stochastic processes in physics and chemistry.* Amsterdam: North-Holland, 1992.
- 140 ABRAMOWITZ, A.; STEGUN, I. A. *Handbook of mathematical functions with formulas, graphs and mathematical tables.* Nat. Bur. Standards Appl. Series, 55, U.S. Government Printing Office, Washington, D.C. (paperback edition published by Dover Books, New York), 1964.
- 141 HILL, T. L. *An introduction to statistical thermodynamics.* Massachusetts: Addison-Wesley Publishing Company, 1960.
- 142 FLYBASE - gene dmel/eve. Disponível em:
<<http://flybase.org/reports/FBgn0000606.html>>. Acesso em: 20 set. 2012.
- 143 SHERMOEN, A. W.; O'FARRELL, P. H. O. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell*, v. 67, n. 2, p. 303 – 310, 1991.
- 144 BIONUMBERS - The database of useful biological numbers. Disponível em:
<<http://bionumbers.hms.harvard.edu/bionumber.aspx?id=103003&ver=2>>. Acesso em: 20 set. 2012.
- 145 BOLOURI, H.; DAVIDSON, E. H. Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proceedings of the National Academy of Sciences*, v. 100, n. 16, p. 9371 - 9376, 2003.
- 146 EDGAR, H.; DAVIDSON, E. H. Repression and turnover pattern *fushi tarazu* RNA in the early *Drosophila* embryo. *Cell*, v. 47, n. 5, p. 747 - 754, 1986.
- 147 SURDEJ, P.; RIEDL, A.; JACOBS-LORENA, M. Determinants of *Drosophila fushi tarazu* mRNA instability. *Annual Review of Genetics*, v. 28, p. 263 - 282, 1994. DOI: 10.1146/annurev.ge.28.120194.001403
- 148 RIEDL, A.; JACOBS-LORENA, M. Determinants of *Drosophila fushi tarazu* mRNA instability. *Molecular and Cellular Biology*, v. 16, n. 6, p. 3047 - 3053, 1996.

-
- 149 HE, X. et al. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLOS Computational Biology*, v. 6, n. 9, p. 1 - 15, 2010.
- 150 VEITIA, R. A. A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. *Biological Reviews*, v. 78, n. 1, p. 149 - 170, 2003.
- 151 BUCHLER, N. E.; GERLAND, U; HWA, T. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences*, v. 100, n. 9, p. 5136 - 5141, 2003.
- 152 ALON, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, v. 8, n. 6, p. 450 - 461, 2007.
- 153 SHERR-ORR, S. S. et al. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, v. 31, p. 64 - 68, 2002. DOI:10.1038/ng881.
- 154 MAYO, A. E. et al. Plasticity of the cis-regulatory input function of a gene. *PLOS Biology*, v. 4, n. 4, p. 0555 - 0561, 2006.
- 155 RODRIGO, G.; ELENA, S. F. Structural discrimination of robustness in transcriptional feedforward loops for pattern formation. *PLOS One*, v. 6, n. 2, p. 1 - 7, 2011.
- 156 BINTU, L. et al. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*, v. 15, n. 2, p. 116 - 124, 2005.
- 157 BINTU, L. et al. Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development*, v. 15, n. 2, p. 125 - 135, 2005.
- 158 ISHIHARA, S.; FUJIMOTO, K.; SHIBATA, T. Cross talking of network motifs in gene regulation that generates temporal pulses and spatial stripes. *Genes to Cells*, v. 10, n. 11, p. 1025 - 1038, 2005.
- 159 COX III, R. S. *Transcriptional regulation and combinatorial genetic logic in synthetic bacterial circuits*. 2008. 208 p. Ph. D. Thesis, California Institute of Technology, Pasadena, California, 2008.
- 160 HERR, A. J. et al. RNA polymerase IV directs silencing of endogenous DNA. *Science*, v. 308, n. 5718, p. 118-120, 2005.

-
- 161 RÉVET, B. et al. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Current Biology*, v. 9, n. 3, p. 151 - 154, 1999.
- 162 WIERZBICKI, A. T.; REAM, T. S.; HAAG, J. R.; PIKAARD, C. S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genetics*, v. 41, n. 5, p. 630 - 634, 2009.
- 163 GENE expression. Disponível em:
<<http://www.nature.com/scitable/topicpage/gene-expression-14121669>>. Acesso em: 7 jul. 2012.
- 164 PAIR-rule genes. Disponível em:
<<http://scienceblogs.com/pharyngula/2007/06/23/pairrule-genes/>>. Acesso em: 9 jul. 2012.
- 165 QUANTUM Alexa Fluor 647 MESF. Disponível em:
<<http://www.bangslabs.com/sites/default/files/bangs/docs/pdf/PDS%20889.pdf>>. Acesso em: 10 jul. 2012.
- 166 FCSC the corporation that sets Standards. Disponível em:
<http://www.cyto.purdue.edu/cdroms/cyto4/7_spon/fcsc/mesf.htm>. Acesso em: 10 jul. 2012.
- 167 GAIGALAS, A. et al. The development of fluorescence intensity standards. *Journal of Research of the National Institute of Standards and Technology*, v. 106, n. 2, p. 381 - 389, 2001.
- 168 SCHWARTZ, A. et al. Formalization of the MESF unit of fluorescence intensity. *Cytometry Part B: clinical cytometry*, v. 57B, n. 1, p. 1 - 6, 2004.
- 169 POULSTELNIKOVA, E. et al. A database for management of gene expression data *in situ*. *Bioinformatics*, v. 20, n. 14, p. 2212 - 2221, 2004.
- 170 PISAREV, A. et al. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic Acids Research*, v. 37, p. 560 - 566, 2008. DOI:10.1093/nar/gkn717.
- 171 GREGOR, T. et al. Probing the limits to positional information. *Cell*, v. 130, n. 1, p. 153 - 164, 2007.

172 GREGOR, T. et al. Diffusion and scaling during early embryonic pattern formation. *Proceedings of the National Academy of Sciences*, v. 102, n. 51, p. 153 – 164, 2005.

173 MARKOW, T. A.; BEALL, S.; MATZKIN, L. M. Egg size, embryonic development time and ovoviviparity in *Drosophila* species. *Journal of Evolutionary Biology*, v. 22, n. 2, p. 430 – 434, 2009.

174 WU, M.; SATO, T. N. On the mechanics of cardiac function of *Drosophila*. *PLOS ONE*, v. 3, n. 12, p. 1 – 7, 2008.

175 HE, F. et al. Probing intrinsic properties of a robust morphogen gradient in *Drosophila*. *Developmental Cell*, v. 15, n. 4, p. 558 – 567, 2008.

176 METZLER, R.; WOLYNES, P. G. Number fluctuations and the threshold model of kinetic switches. *Chemical Physics*, v. 284, n. 1 - 2, p. 469 - 479, 2002.