



ESTUDO COMPARATIVO ENTRE REDES DE REGULAÇÃO GÊNICA COM POTENCIAL PARA OPERAR COMO OSCILADORES CIRCADIANOS

ANA MARIA M. CARPES*, JOSÉ EDUARDO R. CURY*, FABIO L. BALDISSERA*

**Departamento de Automação e Sistemas – DAS/UFSC
Florianópolis, SC, Brasil*

Emails: aninha.mainhardt@gmail.com, jose.cury@ufsc.br, baldissera@gmail.com

Abstract— We propose a comparative study of the dynamics and the robustness to parametric variations among gene regulatory networks that are able to generate and maintain circadian oscillations. Present in most organisms, these oscillations have the role of a clock within the cell, which uses them to anticipate changes in its external environment caused by day and night cycles and to control its behavior accordingly. By means of stochastic models analysis, we conclude that networks comprised of a pair of activating-repressing genes are more robust than those formed by a single self-repressing gene. Moreover, we observe that the formation of compounds among their proteins also favours robustness.

Keywords— Systems Biology, Stochastic Simulation Models, Gene Regulatory Networks, Circadian Oscillations.

Resumo— Propomos aqui um estudo comparativo da dinâmica e da robustez a variações paramétricas entre redes de regulação gênica capazes de gerar e manter oscilações circadianas. Existentes na maior parte dos organismos, essas oscilações têm a função de um relógio interno à célula, que as utiliza para antever as mudanças em seu ambiente externo decorrentes dos ciclos de noite e dia e controlar seu comportamento de acordo. Por meio da análise de modelos estocásticos, concluímos que redes constituídas por um par de genes ativador-repressor são mais robustas perante aquelas formadas por um único gene que se auto reprime. Ademais, vemos que a formação de complexos entre suas proteínas também contribui neste sentido.

Palavras-chave— Biologia Sistemática, Modelos de Simulação Estocástica, Redes de Regulação Gênica, Oscilações Circadianas.

1 Introdução

O estudo da fisiologia celular necessita de uma abordagem sistêmica e interdisciplinar, de modo que seja possível elucidar os mecanismos que regem a vida, prever, ou até mesmo controlar seu comportamento em determinadas circunstâncias (Ideker et al., 2001). O presente trabalho¹ encaixa-se nessa abordagem, uma vez que utilizamos métodos formais de modelagem para analisar a dinâmica de determinadas redes de regulação gênica.

As redes que nos propomos a estudar são aquelas capazes não só de gerar, mas também de manter oscilações circadianas — ou seja, com o período próximo de 24 horas — motivo pelo qual nos referimos a essas redes como osciladores circadianos. Existentes na maior parte dos organismos, essas oscilações assumem um papel importante no controle dos processos celulares (Gonze e Goldbeter, 2006; Alberts et al., 2010). Elas funcionam como um relógio interno à célula, o que lhe permite prever as variações ambientais decorrentes dos ciclos de noite e dia. Assim, apesar de fatores externos como luz e temperatura poderem regular seu período, fase e amplitude, as oscilações são mantidas mesmo sem tais entradas, gerando noites e dias subjetivos à célula. Para terem tais propriedades, duas características são essenciais a essa rede: um laço de realimentação negativa em

seu controle, e um atraso no efeito repressor que provê essa realimentação (Dunlap, 1999; Barkai e Leibler, 2000).

A literatura está repleta de trabalhos nesse contexto — e.g. Gonze e Goldbeter (2006), Dunlap (1999), Barkai e Leibler (2000), Vilar et al. (2002), Blanchini et al. (2014). Porém, não encontramos nenhum que compare a dinâmica e a robustez entre diferentes osciladores. Propomos então um estudo comparativo entre redes de regulação gênica que possuem diferentes estruturas biológicas para implementar as duas características citadas acima, analisando seu potencial para atuar como relógios circadianos e também sua robustez em manter oscilações (circadianas ou não).

2 Redes de Regulação Gênica e o Controle na Célula

A menor unidade viva é a célula, capaz de reproduzir-se e de armazenar as informações hereditárias. Todas as células armazenam essas informações em seu genoma, que são moléculas de DNA de fita dupla. Cada segmento de DNA que transcreve — i.e., que codifica — uma proteína, variante proteica ou RNA representa um *gene*. Ademais, o número dessas moléculas presentes na célula (seus transcritos) indica o que chamamos *nível de expressão* daquele gene. Esse nível pode ser controlado por meio de outros segmentos de DNA, interespaçados entre genes, chamados regiões reguladoras, ou ainda de *promotores*.

¹Trabalho desenvolvido no contexto de um projeto de Mestrado (Carpes, 2015).

As proteínas e RNAs sintetizados pela célula comandam suas funções metabólicas e mecanismos de sinalização, que permitem sua interação com o meio onde se encontra. Ademais, as proteínas podem atuar também como *fatores de transcrição* dos genes, acoplando-se a seus promotores e, desse modo, aumentando ou diminuindo a probabilidade da maquinaria de transcrição se acoplar, i.e., regulando o nível de expressão desses genes. Essa trama de interações forma as redes de regulação gênica. Para essas e mais informações sobre os conceitos básicos da Biologia Celular, consultar Alberts et al. (2010).

Assim, do ponto de vista de um engenheiro de controle, a célula é um complexo sistema, e seu genoma atua como um controlador em malha fechada, recebendo informações do meio intra e extracelular, processando-as e comandando uma resposta através de seus transcritos.

3 Aleatoriedade dos Sistemas Biológicos e o Algoritmo de Simulação Estocástico

A aleatoriedade desempenha um papel fundamental em redes de regulação gênica, e desconsiderá-la na construção de um modelo pode ocasionar uma divergência entre o comportamento previsto e aquele observado (Gillespie, 1977; Elowitz et al., 2002; Samad et al., 2005). De fato, o uso de modelos não deterministas é, muitas vezes, imprescindível para a compreensão de inúmeros fenômenos observados — e.g., como o processo de diferenciação celular (Eldar e Elowitz, 2010), a escolha entre os ciclos lítico e lisogênico em células infectadas por vírus (Arkin et al., 1998) e os diferentes fenótipos observados em populações isogênicas (Elowitz et al., 2002).

São os argumentos acima que nos levam à escolha do algoritmo estocástico de simulação, desenvolvido por Gillespie (1977), para servir como formalismo de base em nossa análise. Esse algoritmo simula um modelo composto por reações químicas R_μ , cada uma associada a uma constante c_μ , tal que a probabilidade dela ocorrer no intervalo infinitesimal de tempo $(t, t + dt)$, num sistema de determinado volume que está no estado \mathbf{x} (definido em termos do número de moléculas de cada tipo nesse volume), é dada pela função $a_\mu = h(\mu, \mathbf{x}) c_\mu dt$ — sendo h a função que indica o número de possíveis combinações entre reagentes de R_μ naquele estado (Gillespie, 1976).

4 Metodologia da Análise

A metodologia da análise consiste, basicamente, nas seguintes etapas²:

- 1 Identificação de redes de regulação gênica que possuem as características essenciais de um

²Para maiores detalhes, consultar Carpes (2015).

oscilador e escolha das que devem ser usadas como alvo de estudo.

- 2 Construção de modelos de simulação estocástica para cada rede visada; de forma concomitante, representação do comportamento observado de um relógio circadiano (de acordo com o que é descrito na literatura, e.g. (Dunlap, 1999)) em alto nível de abstração, por meio de modelo booleano.
- 3 (Re)Definição do conjunto de valores que cada parâmetro c_μ (de cada modelo de simulação) pode assumir, aqui chamado *faixa paramétrica*. Denominamos *configuração* um vetor que fornece os valores de cada um dos parâmetros do modelo.
- 4 Escolha aleatória de 100 configurações para cada modelo estocástico — a que chamamos *casos* para diferenciá-los das demais configurações permitidas pelas faixas paramétricas, mas não selecionadas. Para cada caso, obtenção de uma longa simulação (Banks et al., 1999) equivalente a 264 horas (tempo de *warm-up*, estimado em 24, mais a simulação de 240 horas de regime permanente).
- 5 Extração do comportamento de alto nível de abstração de cada simulação, por meio da definição de valores limítrofes para cada variável do modelo estocástico que é representada no booleano — chamadas aqui *variáveis de interesse* — de modo a obter suas correspondentes booleanas. Para cada caso, aplicamos diferentes valores limítrofes, já que não podemos antever quais são os ideais para eles.
- 6 Cálculo das estimativas da média do período de oscilação $\hat{\mu}_P$ e de sua variância $\hat{\sigma}_P^2$ (para os casos que apresentam oscilações).
- 7 Seleção automática dos melhores casos para cada um dos seguintes critérios: (i) menor $\Delta = |\hat{\mu}_P - 24|$; (ii) menor coeficiente de variação do período $CV = \frac{\hat{\sigma}_P}{\hat{\mu}_P}$; (iii) menor valor para a função $\frac{1}{2}(\Delta^2 + CV^2)$.
- 8 Seleção manual, de acordo com a comparação qualitativa entre o comportamento do modelo booleano e aqueles relativos aos casos selecionados na etapa anterior. Se nenhum dos casos resulta em um comportamento suficientemente próximo daquele esperado por um oscilador, retornamos ao Passo 3.

A simulação de cada modelo para diferentes casos permite-nos analisar seu *potencial como oscilador circadiano*, que é sua capacidade de ser configurada para operar como tal. Ou seja, esse potencial diz respeito à possibilidade das constantes c_μ referentes às reações R_μ serem ajustadas de



Figura 1: Topologia das redes estudadas (\rightarrow indica ativação do elemento de destino pelo de origem, e \dashv repressão no mesmo sentido).

modo que, no conjunto, a rede apresente oscilações de período próximo a 24 horas — o que leva aos critérios de seleção dispostos acima. Atentamos que as variações nesses parâmetros refletem a interação entre os elementos representados nos modelos e os demais presentes na célula, ou seja, podem ser interpretadas como os sinais de entrada do modelo³. Ademais, a simulação desses diferentes casos permite ainda estimar quão robustos os osciladores são a variações paramétricas — pelo menos dentro das faixas estipuladas⁴. A robustez que consideramos aqui é a capacidade de um modelo em manter oscilações, sejam elas circadianas ou não.

5 Modelagem dos Osciladores

5.1 Modelos de Simulação Estocástica

Com base nas características essenciais a um oscilador, elaboramos um conjunto de modelos de redes de regulação que apresentam tais propriedades. Podemos separá-los em dois grupos, de acordo com a topologia de suas redes (Figura 1).

Nas redes do Grupo 1, há apenas um gene, r , que reprime a si mesmo. Essa repressão é transcricional, ou seja, sua proteína R atua como um fator de transcrição. A diferença entre os dois modelos deste grupo reside no mecanismo de atraso, responsável pelo acúmulo de R antes que possa inibir seu gene. O primeiro deles (Figura 2a) é baseado em Gonze e Goldbeter (2006), mas o conceito de proteína ativa (R_A) e inativa (R_I) que utilizamos aqui é mais genérico: ele diz respeito apenas à sua capacidade de atuar como repressor. Assim, a forma ativa pode representar, ou a proteína no núcleo de células eucariotas (como no trabalho citado), ou próxima ao nucleóide de procariontos; pode também referir-se à sua conformação ou ainda à adição/remoção de algum elemento que o permita reprimir seu gene. O segundo modelo desse grupo (Figura 2b) considera que não é R o repressor, mas o complexo formado por duas dessas proteínas, que precisa estar ativo (R_{dA}) para

³Por exemplo, a presença de luz e o aumento de temperatura, importantes na alteração da fase, período e amplitude das oscilações circadianas, possuem tal impacto porque alteram os valores das constantes das reações, i.e., a probabilidade de sua ocorrência.

⁴Claro que devemos estar cientes dos enganos e falhas passíveis de serem cometidos em todo o processo de definição das faixas paramétricas, cálculo de estimativas e seleção de casos.

agir como fator de transcrição.

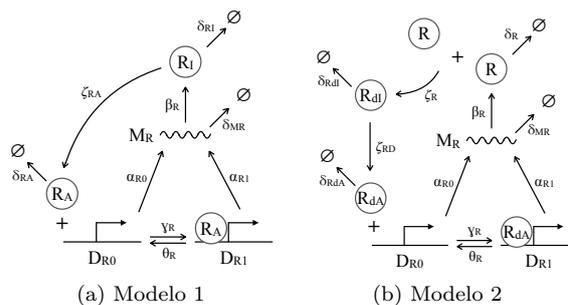


Figura 2: Modelos do Grupo 1. As flechas indicam reações R_μ e as letras gregas associadas indicam os parâmetros c_μ dos modelos. Para suas variáveis, utilizamos letras maiúsculas do alfabeto latino: inicial D indica o promotor de um gene — que pode estar em dois diferentes estados, sem e com fator de transcrição acoplado a ele — e M o RNA transcrito — a partir do qual a proteína é traduzida. O símbolo \emptyset aponta a degradação de um elemento.

As redes que compõem o Grupo 2 são formadas por uma par ativador-repressor, no qual a proteína ativadora, A , ativa o gene da repressora, R , que por sua vez reprime a expressão da primeira — tal que do ponto de vista de ambos os elementos, há um laço de realimentação negativa. Ao contrário do grupo anterior, essa repressão não é transcricional em todos os modelos. No primeiro deles, extraído de Barkai e Leibler (2000) (Figura 3a, adaptada do artigo citado), A é um fator de transcrição que se acopla ao promotor de r e do próprio gene a , ativando-os; R , no entanto, age de modo diferente, formando um complexo com aquela proteína, sequestrando-a e impedindo que faça seu papel como ativadora. No segundo modelo (Figura 3b), ambas as proteínas do par são fatores de transcrição.

Ambas as redes do Grupo 1 e as duas primeiras do Grupo 2 são hipotéticas, mas baseadas em estruturas e mecanismos biologicamente plausíveis (Alberts et al., 2010; Buchler et al., 2003). O terceiro modelo do Grupo 2 (Figura 3c), todavia, foi baseado na descrição de uma rede identificada como cerne das oscilações circadianas em um organismo vivo, o fungo *Neurospora crassa* (Dunlap, 1999; Alberts et al., 2010). Nele, o ativador A é um complexo formado por duas proteínas, cujos genes não são representados porque se considera que A é constitutivamente expresso, i.e., está sempre presente no sistema e em quantidade aproximadamente constante. Ademais, A pode estar em sua forma ativa (A_A) ou inativa (A_I) com relação ao potencial como ativador do gene r . A inativação de A por R dá-se pela transferência de um elemento durante a formação de um complexo entre as duas proteínas⁵.

⁵A correspondência desse modelo com a rede identifi-

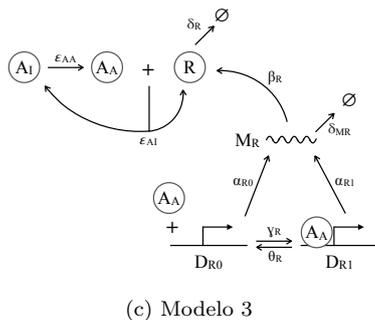
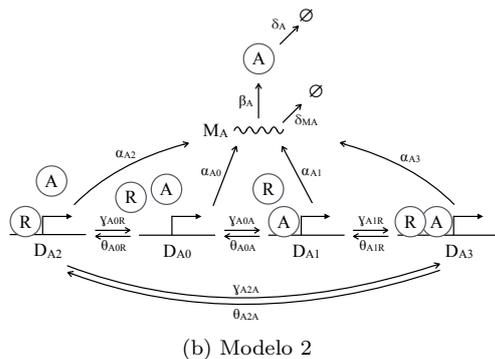
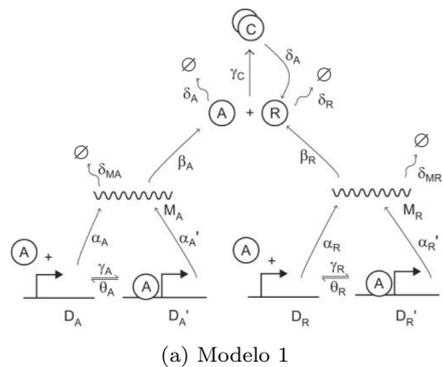


Figura 3: Modelos do Grupo 2 (notação descrita na Figura 2).

5.2 Modelo Booleano do Comportamento Esperado de um Oscilador

Para que o potencial como oscilador circadiano dos modelos apresentados anteriormente possa ser avaliado, precisamos de um modelo de referência que capture os aspectos em comum à dinâmica de qualquer relógio circadiano. Ou seja, precisamos de um modelo com alto nível de abstração de detalhes, o que nos leva ao formalismo denominado Redes Booleanas (Thomas, 1991).

Propomos, então, um modelo booleano para o comportamento de redes identificadas como cerne de oscilações circadianas em organismos, conforme descrição encontrada na literatura (Dunlap, 1999).

cada em *N. crassa* é tal que *A* representa o complexo denominado *white collar complex* (*WCC*), *R* a proteína *FRQ*, e a inativação de *A* ocorre pela sua fosforilação — *WCC* hipofosforilado (*A* ativo) recebe de *FRQ* grupos fosfato, tornando-se hiperfosforilado (*A* inativo), impedindo-o de se ligar ao promotor de *r*.

Tabela 1: Correspondência entre variáveis do modelo booleano e dos modelos de simulação.

Modelo Booleano	Grupo 1		Grupo 2		
	1	2	1	2	3
Reprimido	R_I	R	A		A_A
Repressor	R_A	R_{dA}	R		

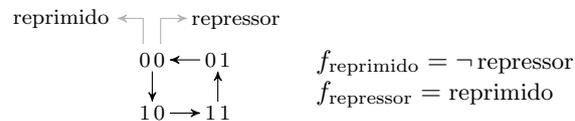


Figura 4: Dinâmica do modelo booleano.

Esse modelo baseia-se apenas nas variáveis mais relevantes das redes tratadas, a que chamamos *variáveis de interesse*. Uma delas está associada ao principal regulador dos demais processos envolvidos do sistema circadiano, i.e., cujo nível de expressão constitui o sinal de controle oscilatório com período próximo de 24 horas⁶. Nos modelos do Grupo 1 ela é a variável repressora, enquanto que nos do Grupo 2 é a ativadora — como indicamos em vermelho na Tabela 1. Outrossim, como é interessante observar ambas as variáveis que atuam diretamente como repressor e reprimido em cada modelo de simulação, o modelo booleano inclui ainda uma segunda variável — referente ao reprimido do Grupo 1, ou ao repressor do Grupo 2.

As funções de atualização das variáveis booleanas que geram o comportamento desejado são mostradas na Figura 4. Perceba que, além de produzirem tal dinâmica, elas ainda condizem com o papel de cada elemento na rede de regulação — i.e., o de repressor e o de reprimido. Ademais, podemos notar que não faz diferença atualizarmos as variáveis de forma síncrona ou assíncrona genérica, ao contrário do que acontece em muitos modelos booleanos (Saadatpour et al., 2010).

6 Potencial como Oscilador de cada Rede de Regulação e sua Robustez

Nesta seção, vamos comparar os resultados obtidos com as simulações dos modelos. Neste contexto, quando nos referimos ao *melhor* caso dentre um conjunto de configurações, queremos dizer aquele para o qual o comportamento do modelo é o mais próximo do esperado, i.e., cujas oscilações podem ser abstraídas conforme o modelo booleano proposto, $\hat{\mu}_P$ é o mais próximo de 24 horas e $\hat{\sigma}_P^2$ (e consequentemente *CV*) possui o menor valor dentre aqueles encontrados.

Para ambos os modelos do Grupo 1, encontramos configurações para as quais eles produzem um

⁶Motivo pelo qual utilizamos essa variável para o cálculo das estimativas relacionadas ao período em cada simulação.

comportamento oscilatório. À medida que vamos do 1 ao 2, todavia, o número de casos selecionados que oscilam aumenta — e assim, melhora o que consideramos por robustez — e a média tomada entre seus respectivos valores $\hat{\mu}_P$ fica mais próxima de 24 horas, bem como a da variância $\hat{\sigma}_P^2$ reduz. Na Figura 5, podemos vislumbrar no melhor caso de cada modelo essa evolução entre eles. Qual seria o motivo disto? Perceba que o período P está diretamente relacionado ao atraso no efeito repressor: quanto maior ele for, maior será o período, e vice-versa. Assim, a análise indica que a reação que forma um complexo entre duas proteínas repressoras tem maior impacto no atraso do que a de *ativação* dessa proteína como agente repressor — o que faz sentido, já que para aquela ocorrer é preciso duas moléculas diferentes do mesmo tipo, enquanto que, para a outra, basta uma.

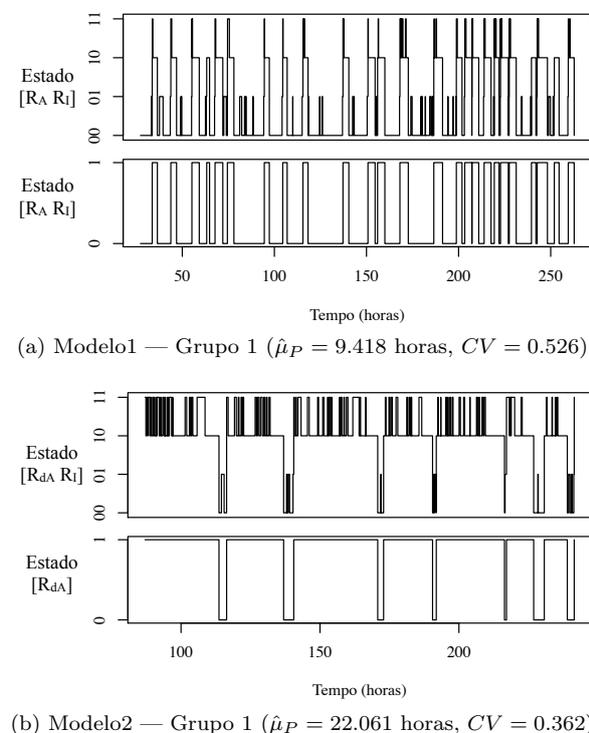


Figura 5: Resultado das simulações (em termos dos estados booleanos das variáveis de interesse) referentes aos melhores casos dos modelos do Grupo 1. Estimativas obtidas para a variável principal (em vermelho na Tabela 1).

Dentre os modelos do Grupo 2, o segundo foi o que resultou nos comportamentos mais distantes daquele capturado pelo modelo booleano, bem como em um menor número de casos oscilatórios. As configurações oscilatórias estão associadas a dinâmicas ruidosas e um baixo número de moléculas repressoras R envolvidas em cada instante de tempo, implicando em $\hat{\sigma}_P^2$ elevado. Dados os resultados do Grupo 1, supomos que se fossem inseridas no modelo as reações de passagem de A e R do citoplasma, onde são traduzidas, para o núcleo,

onde atuam como fatores de transcrição, ele teria um ganho de potencial e de robustez — hipótese esta que precisaria ser testada.

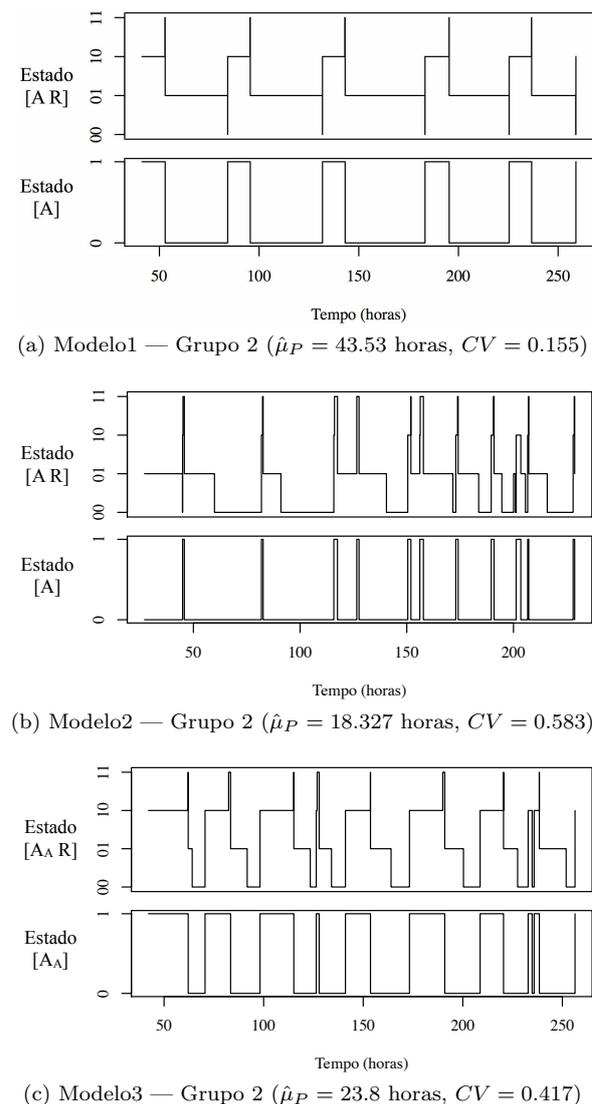


Figura 6: Resultado das simulações (em termos dos estados booleanos das variáveis de interesse) referentes aos melhores casos dos modelos do Grupo 1. Estimativas obtidas para a variável principal (em vermelho na Tabela 1).

Quanto ao primeiro modelo do Grupo 2, apenas com um ciclo de definição das faixas paramétricas já obtemos casos satisfatórios com nossas simulações (exceto pelo período), conforme vemos no exemplo da Figura 6a. O motivo disto é já termos uma configuração para a qual o modelo opera como um oscilador: como mencionamos na Seção 5.1, esse modelo foi extraído de Barkai e Leibler (2000), que fornece valores padrões para seus parâmetros. Assim, no primeiro ciclo as faixas já foram definidas de modo a conter esses valores.

O terceiro e último modelo do Grupo 2 foi baseado na rede de regulação do fungo *Neurospora*. Porquanto ele representa uma rede identificada como cerne dos ciclos circadianos em um

organismo vivo, se não tivéssemos encontrado nenhuma configuração adequada, seria porque omitimos ou deturpamos a representação de algum mecanismo biológico relevante para as oscilações, ou porque não tivemos êxito em uma ou mais etapas no processo de escolha e seleção de casos. Entretanto, os casos que obtemos para este modelo são em sua maioria oscilatórios, e aqueles que selecionamos — e.g., o da Figura 6c — apresentam um comportamento mais próximo do circadiano do que todos os demais modelos.

7 Conclusões

Este trabalho corrobora a ideia de que um oscilador pode ser implementado de maneiras distintas, por meio de diferentes mecanismos biologicamente plausíveis. Todavia, quanto mais complexa a rede, mais custoso é para mantê-la em termos de gasto energético, matéria-prima e ocupação da maquinaria celular; mesmo assim, os relógios circadianos identificados em organismos não são tão simples quanto um oscilador pode ser.

Os resultados aqui obtidos são um indicativo de que osciladores constituídos por um par de genes ativador-repressor são mais robustos e possuem um maior potencial como relógio circadiano perante aqueles formados por um único gene que se auto reprime, o que pode justificar o custo extra empregado pelas células. Ademais, vemos que a formação de complexos entre proteínas — mesmo daqueles que ocorrem apenas durante a transferência de algum elemento — contribui para regular o atraso na realimentação negativa nos modelos; por conseguinte, ela facilita a sintonização dos parâmetros para que as oscilações sejam circadianas, e também provê mais robustez.

Agradecimentos

Agradecemos a CAPES pelo suporte financeiro.

Referências

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. e Walter, P. (2010). *Biologia Molecular da Célula*, 5 ed., Editora Artmed.
- Arkin, A., Ross, J. e McAdams, H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells, *Genetics* **149**(4): 1633–1648.
- Banks, J., Carson, J. e Nelson, B. (1999). *Discrete-Event System Simulation*, 2 ed., Prentice Hall.
- Barkai, N. e Leibler, S. (2000). Biological rhythms: Circadian clocks limited by noise, *Nature* **403**(6767): 267–268.
- Blanchini, F., Samaniego, C., Franco, E. e Giordano, G. (2014). Design of a molecular clock with RNA-mediated regulation, *53rd IEEE Conference on Decision and Control (CDC2014)*, Los Angeles, pp. 4611–4616.
- Buchler, N., Gerland, U. e Hwa, T. (2003). On schemes of combinatorial transcription logic, *PNAS* **100**(9): 5136–5141.
- Carpes, A. M. M. (2015). *Estudo da Dinâmica de Redes de Regulação Gênica que Apresentam Oscilações Circadianas*, Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis, Brasil.
- Dunlap, J. (1999). Molecular bases for circadian clocks, *Cell* **96**(2): 271–290.
- Eldar, A. e Elowitz, M. (2010). Functional roles for noise in genetic circuits, *Nature* **467**(7312): 167–173.
- Elowitz, M., Levine, A., Siggia, E. e Swain, P. (2002). Stochastic gene expression in a single cell, *Science* **297**(5584): 1183–1186.
- Gillespie, D. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *Journal of Computational Physics* **22**(4): 403–434.
- Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions, *Journal of Physical Chemistry* **81**(25): 2340–2361.
- Gonze, D. e Goldbeter, A. (2006). Circadian rhythms and molecular noise, *Chaos* **16**(2): 026110.
- Ideker, T., Galitski, T. e Hood, L. (2001). A new approach to decoding life: systems biology, *Annual review of genomics and human genetics* **2**: 343–372.
- Saadatpour, A., Albert, I. e Albert, R. (2010). Attractor analysis of asynchronous boolean models of signal transduction networks, *Journal of Theoretical Biology* **266**(4): 641–656.
- Samad, H., Khammash, M., Petzold, L. e Gillespie, D. (2005). Stochastic modelling of gene regulatory networks, *Int. J. Robust and Non-linear Control* **15**(15): 691–711.
- Thomas, R. (1991). Regulatory networks seen as asynchronous automata: A logical description, *Journal of Theoretical Biology* **153**(1): 1–23.
- Vilar, J., Kueh, H., Barkai, N. e Leibler, S. (2002). Mechanisms of noise-resistance in genetic oscillators, *PNAS* **99**(9): 5988–5992.