

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE AUTOMAÇÃO E SISTEMAS**

Fabio Luis Baldissera

**UMA ABORDAGEM BASEADA NA TEORIA DE
CONTROLE SUPERVISÓRIO PARA CONTROLE DE
REDES DE REGULAÇÃO GÊNICA**

Florianópolis

2014

Fabio Luis Baldissera

**UMA ABORDAGEM BASEADA NA TEORIA DE
CONTROLE SUPERVISÓRIO PARA CONTROLE DE
REDES DE REGULAÇÃO GÊNICA**

Tese submetida ao Programa de Pós-Graduação em Engenharia de Automação e Sistemas para a obtenção do Grau de Doutor em Engenharia de Automação e Sistemas.

Orientador: Prof. Dr. José Eduardo Ribeiro Cury

Florianópolis

2014

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Baldissera, Fabio Luis

Uma Abordagem Baseada na Teoria de Controle Supervisório
para Controle de Redes de Regulação Gênica / Fabio Luis
Baldissera ; orientador, José Eduardo Ribeiro Cury -
Florianópolis, SC, 2014.

159 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Engenharia de Automação e Sistemas.

Inclui referências

1. Engenharia de Automação e Sistemas. 2. Sistemas a
Eventos Discretos. 3. Teoria de Controle Supervisório. 4.
Biologia Sistemica. 5. Biologia Sintética. I. Cury, José
Eduardo Ribeiro. II. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Engenharia de
Automação e Sistemas. III. Título.

Fabio Luis Baldissera

**UMA ABORDAGEM BASEADA NA TEORIA DE
CONTROLE SUPERVISÓRIO PARA CONTROLE DE
REDES DE REGULAÇÃO GÊNICA**

Esta Tese foi julgada adequada para a obtenção do Título de “Doutor em Engenharia de Automação e Sistemas” e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Automação e Sistemas.

Florianópolis, 15 de agosto 2014.

Prof. Dr. Rômulo Silva de Oliveira
DAS/UFSC
Coordenador do Curso

Banca Examinadora:

Prof. Dr. José Eduardo Ribeiro Cury
Orientador

Prof. Dr. João Carlos dos Santos Basílio
DEE/UFRJ

Prof. Dr. Antônio Eduardo Carrilho da Cunha
IME-RJ

Profa. Dra. Patrícia Nascimento Pena
DELT/UFMG

Prof. Dr. Daniel Santos Mansur
MIP/UFSC

Prof. Dr. Max Hering de Queiroz
DAS/UFSC

Prof. Dr. Daniel Juan Pagano
DAS/UFSC

AGRADECIMENTOS

Agradeço aos meus pais pelo amor e apoio desde os tempos de zigoto, e por nunca questionarem as faturas da Livraria Cultura desde os tempos de graduação.

Agradeço ao meu irmão, que soube compreender (eu acho!) por que eu recusava os convites para café e por que não retornava as ligações nos momentos finais da Tese.

Agradeço ao meu orientador, professor Cury, pelo exemplo, pelo tempo despendido nas reuniões, por me tranquilizar nos picos de ansiedade, pela valiosa orientação, que transformou um trabalho em uma Tese.

Agradeço à minha sogra, que sobreviveu, ao longo de quatro anos, aos sustos periódicos de um terceiro entrando em seu apartamento.

Por fim, sou imensamente grato à minha esposa, que, com carinho, cumplicidade e muito amor, foi a grande companheira e incentivadora de cada etapa deste percurso, mesmo - e ainda mais! - nas minhas ausências regulares.

RESUMO

Nesta Tese, trataremos do problema de controle realimentado de sistemas biológicos celulares, mais especificamente, do controle das chamadas redes de regulação gênica, segundo uma abordagem fundamentada na Teoria de Controle Supervisório (TCS).

A capacidade de reprogramar o comportamento das redes de regulação gênica, elementos determinantes da dinâmica celular, encontra aplicações em áreas como produção otimizada de biocombustíveis, tratamento de doenças como o câncer e síntese de tecidos e órgãos humanos *in vitro*. Basicamente, duas abordagens podem ser adotadas para controlar redes biológicas intracelulares. Uma delas baseia-se no emprego de dispositivos técnicos externos à célula, como equipamentos de microscopia, computadores e sistemas de microfluídica. A segunda estratégia prefere fazer uso exclusivo de moléculas sintéticas que, inseridas no meio intracelular, desempenham as funções de sensores, controladores e atuadores.

Utilizaremos a última das abordagens mencionadas e mostraremos como ideias advindas da TCS podem ser empregadas para definir os critérios de projeto de tais moléculas sintéticas controladoras, capazes de alterar a dinâmica de redes de regulação gênica de maneira pré-determinada. Para tanto, partiremos de um modelo em autômato finito da rede a ser controlada, adicionando a esse autômato os eventos e estados decorrentes das possíveis ações de moléculas sintéticas. Sobre esse novo modelo, resolveremos uma tarefa de controle conhecida na literatura como atração de estados, em que o sistema deve ser levado de um estado inicial a um conjunto de estados-alvo, mantendo-se ali indefinidamente. O supervisor resultante da resolução do problema será convertido, então, em critérios de projeto de moléculas sintéticas.

A Tese traz contribuições à TCS e propõe uma formalização inédita para o problema de controle de redes de regulação gênica por moléculas sintéticas. No campo da TCS, estendemos a resolução do problema de atração de estados para o caso em que os eventos da planta, além de serem particionados em controláveis ou não controláveis, são também divididos em forçáveis ou não forçáveis e, por fim, preemptáveis ou não preemptáveis. Esse rico particionamento, como detalhado na Tese, decorre de hipóteses bem fundamentadas biologicamente acerca da dinâmica dos processos de expressão gênica e da forma de intervenção no comportamento celular por nós adotada. Ainda no âmbito do Con-

trole Supervisório, mostramos como conceitos provenientes da teoria de Controle Hierárquico - como abstração por agregação de estados e consistência hierárquica - naturalmente aparecem na resolução do problema de controle de redes de regulação gênica por moléculas sintéticas. A Tese mostra, também, como projetar controladores que induzem o menor consumo de energia celular possível e que sejam realizados por moléculas sintéticas de fácil construção.

Por sua vez, a formalização do problema de controle de redes gênicas por moléculas sintéticas *como um problema da Teoria de Controle* é nova na literatura, haja vista que os estudos correlatos até então divulgados não adotam métodos da Teoria de Controle para sua resolução, o que dificulta a extensão dessas abordagens para alterar o comportamento de redes de regulação com dinâmicas mais complexas.

Palavras-chave: Teoria de Controle Supervisório. Biologia Sistêmica. Biologia Sintética.

ABSTRACT

In this thesis, we tackle the issue of feedback control of intracellular biological networks, namely, the control of gene regulatory networks within a perspective founded on Supervisory Control Theory (SCT). The ability to reprogram the behavior of gene networks, major players in cell dynamics, might be employed, for instance, to optimize the production of biofuels, to treat and cure diseases, such as cancer, and to synthesize human tissues and organs *in vitro*. Two approaches can be adopted to control the dynamics of intracellular networks. The first one employs external technical hardware to close the loop, e.g. microscopy equipment, microfluidics devices and computers. The second, on the other hand, makes use only of synthetic biomolecules that, when inserted in the cell, perform the major tasks of a control system: sensing, control and actuation.

We adopt the second of the mentioned strategies and show how ideas from SCT can be employed to define design criteria of synthetic biomolecules that are able to alter cell dynamics, so to achieve a set of specifications. In order to do so, we start with a finite state machine model of the native gene network to be controlled and update this model with the new states (and events) that can be achieved (generated) by programmable synthetic biomolecules. This enlarged model is used to solve a control task known as *state attraction*, which can be stated as follows: find a supervisor that brings the system from an initial state to a set of target states, keeping it there indefinitely. This supervisor is then translated into design criteria of synthetic biomolecules.

Our thesis brings theoretical contributions to SCT and proposes a novel approach to formalize the control of gene networks by means of synthetic biomolecules. In the field of SCT, we extend the solution of the state attraction problem for the case where plant events, besides being partitioned into controllable and uncontrollable ones, are also divided into forcible and non forcible, preemptable and non preemptable. This rich partition, as shown throughout the thesis, is a result of concrete biological assumptions regarding the dynamics of native and synthetic genes. Still concerning the field of SCT, we show how concepts from Hierarchical Control Theory – e.g. abstraction by state aggregation and hierarchical consistency – naturally appear in the process of finding a solution for the posed control problem. The thesis shows also how to design controllers that minimize the energy consumption and

that are implemented by synthetic biomolecules that are more easily constructed.

The formalization we propose for the control of gene networks with synthetic genes *as a control problem* is new in the literature, given that the results concerning the intervention in the dynamics of biological networks by means of synthetic molecules do not adopt procedures from Control Theory to solve the problem, thus making it harder to extend such approaches to networks with a more complex dynamics.

Keywords: Supervisory Control Theory. Systems Biology. Synthetic Biology.

LISTA DE FIGURAS

Figura 1	Diagrama simplificado de uma malha de controle	23
Figura 2	Estrutura da Tese	28
Figura 3	Sugestões de leitura da Tese	30
Figura 4	Ativação de proteína por metabólitos	32
Figura 5	Proteínas como enzimas e sensores	33
Figura 6	Representação da transcrição e tradução gênicas	34
Figura 7	Parte da rede de regulação gênica de levedura	37
Figura 8	Representação de um gene	37
Figura 9	Resposta de bactérias a altas temperaturas	39
Figura 10	Rede de regulação gênica mutuamente repressora	43
Figura 11	Relação de modelagem	44
Figura 12	Grafo associado à rede da Figura 10	46
Figura 13	Comportamento temporal de x_1 e x_2 : caso 1	49
Figura 14	Comportamento temporal de x_1 e x_2 : caso 2	50
Figura 15	Espaço de fases para o sistema da Figura 12	51
Figura 16	Metáfora do comportamento de um sistema dinâmico	52
Figura 17	Função de Hill <i>versus</i> função do tipo degrau	55
Figura 18	Diagrama de transição de estados	57
Figura 19	Abstração discreta de planta contínua	57
Figura 20	Nodos externos para controle de redes Booleanas	64
Figura 21	Controle <i>in silico</i> de luminescência	66
Figura 22	Sistema de controle <i>in silico</i> de luminescência	67
Figura 23	Uso de proteínas exógenas para controle celular	70
Figura 24	Rede de Regulação Gênica: Exemplo	76
Figura 25	Dinâmica da rede da Figura 24	76
Figura 26	Exemplo do conceito de controle da Tese	78
Figura 27	Autômato G^{au} para o exemplo	81
Figura 28	Autômato G^{hi} para o exemplo	83
Figura 29	Transições $-u_i$ em G^{au}	86
Figura 30	Transições $+u_i$ em G^{au}	87
Figura 31	Relação entre $\pm v_i^s$ e ocorrência de $-u_i$	97
Figura 32	Controle da expressão gênica	100

Figura 33	Preemptabilidade e dinâmica na expressão gênica.....	102
Figura 34	Implementação de funções lógicas com genes.....	104
Figura 35	Região de regulação de v_1^s e v_3^s	106
Figura 36	Consistência Hierárquica Fraca.....	108
Figura 37	Representação esquemática do Capítulo 5.....	119
Figura 38	Exemplo do Capítulo 5 e S^{hi}	123
Figura 39	Mapa de Karnaugh e simplificação de f_i^s	125
Figura 40	Novo exemplo para minimização de energia.....	128
Figura 41	Atribuição de custos aos macroestados.....	128
Figura 42	Controle de redes modeladas por EDOs.....	132
Figura 43	Rede de regulação a ser modelada por EDOs.....	133
Figura 44	Partição Booleana de espaço de estados.....	135
Figura 45	Algoritmo para construção de abstrações discretas.....	137
Figura 46	Diagrama de transição de estados do sistema de EDOs.....	138
Figura 47	Abstração G^{hi} do sistema de EDOs.....	139
Figura 48	Conversão de f_i^s em EDOs.....	140
Figura 49	Simulação do sistema em malha fechada contínuo.....	143
Figura 50	Dinâmica temporal das variáveis u_1 , u_2 e u_3	144
Figura 51	Simulação do sistema em malha fechada contínuo.....	145
Figura 52	Dinâmica temporal das variáveis u_1 , u_2 e u_3	146
Figura 53	Evolução temporal de u_1 , u_2 e u_3	147

LISTA DE SÍMBOLOS

G^{au}	Autômato de baixo nível	78
G^{hi}	Autômato de alto nível	82
$\Sigma_{cf}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível controláveis e forçáveis em x	89
$\Sigma_{cf}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível controláveis e não forçáveis em x	89
$\Sigma_{cfp}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível não controláveis, forçáveis e preemptáveis em x	89
$\Sigma_{cf\bar{p}}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível não controláveis, forçáveis e não preemptáveis em x	89
$\Sigma_{cfp}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível não controláveis, não forçáveis e preemptáveis em x	89
$\Sigma_{cf\bar{p}}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível não controláveis, não forçáveis e não preemptáveis em x	89
$\Sigma_{cfD}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível controláveis, não forçáveis e que podem ser gerados a partir de qualquer microestado $q \in x$	89
$\Sigma_{cf\bar{D}}^{\text{hi}}(x)$	Conjunto dos eventos de alto nível controláveis, não forçáveis e que não podem ser gerados a partir de qualquer microestado $q \in x$	89
S^{hi}	Supervisor de alto nível	90
$S^{\text{hi}}(x)$	Ação de controle do alto nível no macroestado x	90
$\mathcal{A}_S^{\text{hi}}(x)$	Ações de controle do alto nível admissíveis no macroestado x	90
$\mathcal{H}_\phi^{\text{hi}}(x)$	Imagem do macroestado x para entrada de controle admissível $\phi(x)$	90
P	Predicado $P : X \rightarrow \{0, 1\}$ associado a uma propriedade biológica de interesse	91
$\Phi_W(Z)$	Função controlável aplicada a um conjunto de macroestados $Z \in 2^X$	92
$\Theta_P(Z)$	Função de alcançabilidade estrita aplicada a um conjunto de macroestados $Z \in 2^X$	93
SR^P	Predicado de alcançabilidade estrita	93
$S^{\text{au}}(q)$	Ação de controle do baixo nível no microestado q	107

SUMÁRIO

1 INTRODUÇÃO	19
1.1 MOTIVAÇÃO	19
1.2 ESCOPO DA TESE E CONTEXTO DA PESQUISA	21
1.3 CONTRIBUIÇÕES	26
1.4 ORGANIZAÇÃO DO DOCUMENTO	28
2 FUNDAMENTOS DE BIOLOGIA CELULAR	31
2.1 PROTEÍNAS E FUNÇÕES CELULARES VITAIS	31
2.2 EXPRESSÃO GÊNICA: TRANSCRIÇÃO E TRADUÇÃO ..	32
2.3 REDES DE REGULAÇÃO GÊNICA	36
2.4 SISTEMA DE CONTROLE CELULAR: UM EXEMPLO ...	38
3 MODELAGEM DE REDES DE REGULAÇÃO GÊNICA	41
3.1 CONCEITOS PRELIMINARES	41
3.2 MODELAGEM DE REDES DE REGULAÇÃO GÊNICA ...	46
3.2.1 Modelos Contínuos	47
3.2.2 Modelos Lógicos	53
3.2.3 Modelos Estocásticos	59
4 CONTROLE DE REDES DE REGULAÇÃO GÊNICA	61
4.1 ABORDAGENS TEÓRICAS	61
4.1.1 Controle Ótimo e Redes Booleanas Probabilísticas .	61
4.1.2 <i>Model-checking</i> e Redes Booleanas	63
4.1.3 Controle de Sistemas Diferenciais Afins por Partes .	64
4.2 ABORDAGENS PRÁTICAS	65
4.2.1 Controle <i>In silico</i>	66
4.2.2 Controle com Circuitos Biológicos Sintéticos	67
5 CONTROLE SUPERVISÓRIO APLICADO À INTERVENÇÃO EM REDES DE REGULAÇÃO GÊNICA .	73
5.1 VISÃO GERAL	73
5.1.1 Exemplo a ser Tratado no Capítulo	75
5.2 FORMALIZAÇÃO DO PROBLEMA DE CONTROLE	77
5.2.1 Modelo da Planta: Autômato G^{au}	78
5.2.2 Abstração de G^{au} : Autômato G^{hi}	82
5.2.3 Definição das Entradas de Controle	89
5.2.4 Problema de Alcançabilidade Estrita	91
5.3 CONDIÇÕES PARA EXISTÊNCIA DE UM SUPERVISOR	92
5.3.1 Exemplo: Parte I	94
5.3.2 Escolha de um Supervisor	95

5.4 REALIZAÇÃO DO SUPERVISOR	96
5.4.1 Conversão de $S^{hi}(x)$ em f_i^s	96
5.4.2 Exemplo: Parte II	98
5.4.3 Projeto de Genes Sintéticos	99
5.4.3.1 Hipótese de Preemptabilidade	99
5.4.3.2 Hipótese de Programação da Região de Regulação	101
5.4.4 Exemplo: Parte III	105
5.5 CONSISTÊNCIA HIERÁRQUICA	107
5.5.1 Entradas de Controle $S^{au}(q)$ para $S^{hi}(x) = \{\pi_f\} \cup$ $\Sigma_{\bar{c}p}^{hi}(x)$, com $\pi_f \in \Sigma_f^{hi}(x)$	109
5.5.2 Entradas de Controle $S^{au}(q)$ para $S^{hi}(x) = \lambda \cup$ $\Sigma_{\bar{c}}^{hi}(x)$, com $\lambda \in 2^{\Sigma_{\bar{c}f}^{hi}(x)}$	112
5.5.3 Ausência de Consistência Hierárquica Estrita	118
5.6 RESUMO	118
6 OTIMIZAÇÃO DO SUPERVISOR	121
6.1 SIMPLIFICAÇÃO DA REGIÃO DE REGULAÇÃO DE GE- NES SINTÉTICOS	121
6.2 MINIMIZAÇÃO DA ENERGIA CONSUMIDA	125
7 CONTROLE DE REDES MODELADAS POR EQUA- ÇÕES DIFERENCIAIS	131
7.1 ABSTRAÇÕES A EVENTOS DISCRETOS DE EDO	132
7.2 CONVERSÃO DO SUPERVISOR EM EDO	136
7.3 ANÁLISE QUANTITATIVA DA FORÇABILIDADE	142
8 CONSIDERAÇÕES FINAIS	149
REFERÊNCIAS	153

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

Pouco menos de 60 anos separam a elucidação da estrutura em dupla hélice do DNA por Watson e Crick da síntese completa em 2010, pela equipe de Craig Venter, de um genoma capaz de coordenar todas as funções celulares de uma bactéria.

Ao longo dessas seis últimas décadas, cientistas colecionaram uma quantidade expressiva de resultados importantes para compreensão minuciosa desta peculiar forma de vida que habita a Terra, uma vida codificada em linguagem quaternária (i.e. A, T, C e G, letras identificadoras dos quatro diferentes blocos construtores que compõem a molécula do polímero DNA) e escrita como um complexo programa, ainda não completamente compreendido, cuja execução é tarefa de uma maquinaria celular igualmente abstrusa.

Decifrar os meandros da vida como a conhecemos não é somente um capricho da curiosidade humana. O entendimento pormenorizado do *software* e do *hardware* que carregam os seres vivos é ponto de partida para: a) melhor **controlar a dinâmica** de processos biológicos existentes; e b) **engendrar novas funções** (ou, até mesmo, formas de vida) ainda não implementadas pela natureza, mas úteis aos propósitos do homem contemporâneo.

Tal intervenção na dinâmica de seres vivos, bem como a construção de novas formas de vida, visando fins específicos não é empreendimento novo na história da humanidade. A antiga prática de ingestão de plantas para fins medicinais e o processo de seleção artificial de animais conduzido desde a pré-história (este último fazendo emergir, de uma única espécie de lobo, cães das mais variadas estirpes) deixam evidente a predisposição do homem para alterar o fluxo normal da dinâmica de sistemas biológicos ou conferir-lhes novas matizes. No entanto, foi o recente progresso científico na área de Biologia Molecular, e sua articulação com disciplinas das Engenharias, Ciências da Computação e Matemática, que alçou as possibilidades de manipulação da matéria orgânica a um patamar sem precedentes em nossa história, desvelando novas oportunidades e desafios, mas, em contrapartida, suscitando também muitas controvérsias (vide as contendas sobre pesquisa com células-tronco ou uso de alimentos geneticamente modificados).

O rol de benefícios que a ciência moderna nos permite vislumbrar é grande e inclui, em lista não exaustiva, a reengenharia do metabo-

lismo de microorganismos para produção de compostos de interesse econômico, como biocombustíveis, o controle da resposta do sistema imune para minimizar riscos de rejeição em transplantes de órgãos, a manipulação dos processos de diferenciação e organização espacial celulares para fabricar tecidos e órgãos, o controle de aspectos da fisiologia humana para tratar ou curar doenças (BENENSON, 2012; KARLSSON; WEBER, 2012) e o projeto e síntese de microorganismos inteiramente novos, capazes de degradar poluentes nocivos em locais de difícil acesso ou de reconhecer e invadir tumores humanos, destruindo-os (PURNICK; WEISS, 2009).

Entretanto, a escolha de estratégias eficazes para intervenção em seres vivos de modo a resolver os problemas contemporâneos mais prementes continua a ser tarefa desafiadora. Seres vivos são sistemas organizados em níveis biológicos interdependentes (células, tecidos, órgãos) e, já no primeiro nível da escada biológica, o celular, deparamo-nos com um subsistema composto por diversos e heterogêneos compostos químicos (i.e. DNA, RNA, proteínas, metabólitos, etc.), que interagem entre si de maneira complexa, com dinâmicas variando de microssegundos a horas, e dentro de um ambiente dominado pela aleatoriedade no encontro entre as moléculas.

Modelar tais sistemas, como se intui, é tarefa árdua. E esse é somente o primeiro passo na longa trajetória de idealização e realização de um outro sistema, que será doravante chamado controlador, capaz de, quando colocado em contato com o sistema biológico sob análise, alterar o comportamento deste de maneira pré-especificada. Uma temática com tamanhos desafios e potenciais impactos motiva-nos a estudá-la e a trazer contribuições para seu avanço, o que faremos nesta Tese.

Para isso, o documento excursionará por temas como modelagem matemática de redes intracelulares e síntese de novas moléculas, expondo como contribuição central uma proposta para formalização do processo de síntese e realização destas entidades denominadas controladores (ou supervisores), que, ao interagirem com um dado sistema biológico celular, possuem a propriedade de alterar sua dinâmica.

Nossa Tese almeja, portanto, galgar um pequeníssimo passo para a escrita dos próximos anos dessa história, que começou com a descoberta do DNA e que traz um final tão desconhecido quanto desafiador.

1.2 ESCOPO DA TESE E CONTEXTO DA PESQUISA

Neste trabalho, trataremos do *controle* de sistemas biológicos em nível *celular*, mais especificamente, do controle das chamadas *redes de regulação gênica*. Para adequada compreensão dos objetivos da Tese, e da contribuição que ela encerra, faz-se necessário que discorramos brevemente sobre: a) o sistema a ser controlado, ou *planta*; b) os trabalhos correlatos publicados na literatura técnica; e c) os métodos formais a serem empregados nesta Tese. Um tratamento mais aprofundado desses temas será feito em capítulos posteriores.

Antes, cabe tecer algumas palavras sobre a própria escolha do nível celular para condução da pesquisa. Como argumentam Wolkenhauer e coautores (WOLKENHAUER; MESAROVIC; WELLSTEAD, 2007), a célula é a unidade básica dos níveis mais altos de organização biológica em seres multicelulares, o que as torna, assim, o principal foco de pesquisa da comunidade dedicada à aplicação de métodos formais e sistêmicos às Ciências Biológicas. Por mais que a batida coordenada de um coração ou os pensamentos de um cérebro humano não possam ser explicados *unicamente* pelas propriedades isoladas das células que os compõem, compreender o comportamento daquelas é um passo necessário para desvelar os mecanismos que as levam a interagir, criando entidades e fenômenos mais complexos. Ademais, em vários casos relevantes para a biotecnologia, os organismos envolvidos são unicelulares (ex.: bactérias) e, portanto, a célula confunde-se com o próprio organismo.

A escolha desse nível biológico, por si só, não determina de forma inequívoca as propriedades do sistema sob análise. Embora todas as células conhecidas compartilhem várias características – como o armazenamento da informação hereditária na molécula de DNA, a utilização de proteínas como catalisadores das reações químicas, o emprego do RNA como molécula intermediária entre o DNA e as proteínas (ALBERTS et al., 2008) – a descrição e/ou a própria existência de muitos processos biológicos intracelulares é afetada sobremaneira pela estrutura e organização internas da célula em questão. As células de eucariotos, seres vivos cujos materiais genéticos encontram-se separados em um compartimento denominado núcleo, são, por exemplo, significativamente mais complexas do que aquelas de procariotos, seres que não possuem seção distinta para abrigar o DNA. Por mais que haja diferenças entre as variadas células, empregaremos nesta Tese, majoritariamente, princípios de funcionamento que são universais a todos os sistemas vivos. Quando esta premissa, entretanto, não for válida, faremos o devido

apontamento.

Independentemente de ser a célula um ser vivo autônomo, ou um elemento integrante de um organismo multicelular, esse notável bloco construtor da vida pode ser retratado como um *sistema dinâmico* que reage a estímulos externos e a estados internos, tomando decisões que implicam, por exemplo, diferenciação, divisão e motilidade celulares, além de secreção de substâncias e indução programada de sua morte. É a combinação coordenada dessas possíveis respostas, fruto de sofisticada computação molecular, que faz emergir os complexos comportamentos adaptativos tão característicos da vida em todos os níveis hierárquicos da organização biológica (BENENSON, 2012).

Ao longo do ciclo de vida de uma célula, portanto, ela constantemente: a) recebe informações do ambiente extracelular e mede o estado intracelular; b) integra essas informações recebidas, reorganizando seu padrão de ativação gênica; e c) absorve e degrada moléculas grandes, a fim de obter energia e sintetizar outras moléculas necessárias para concretizar as mais diferentes ações. De maneira simplificada, e não completamente fidedigna, as funções (a), (b) e (c) podem ser associadas às redes biológicas intracelulares conhecidas como *redes de sinalização*, *regulação gênica* e *metabólicas*, respectivamente. Cada uma dessas redes é formada por um conjunto particular de biomoléculas que interagem de maneira específica. É importante ressaltar que, embora as redes sejam divididas para fins de estudo, elas agem de maneira harmonizada na realização das funções celulares vitais.

Como mencionado no primeiro parágrafo desta seção, dedicaremos ao controle das redes de regulação gênica. Essa classe particular das redes intracelulares foi escolhida como porta-voz do comportamento celular, pois é ela quem coordena os processos mais importantes (ALBERTS et al., 2008), como a diferenciação celular, o ciclo que compreende crescimento, replicação do material genético e divisão celular, bem como a adaptação da célula às diferentes condições ambientais (variação de temperatura, disponibilidade de nutrientes, etc.).

O comportamento das redes biológicas intracelulares em geral, e das redes de regulação gênica em particular, é de difícil compreensão e análise se descrito somente por meio de linguagem natural, uma forma imprecisa e incompleta de representação do conhecimento. Tal fato ajudou a disseminar o emprego de modelos matemáticos para formalização dos fenômenos biológicos observados, e posterior análise do comportamento dos sistemas envolvidos. A aplicação de métodos formais a sistemas biológicos não é um empreendimento novo na ciência, como atestam os esforços passados da comunidade de Teoria de Siste-

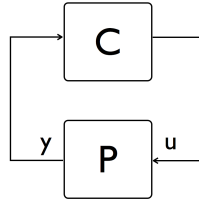


Figura 1 – Diagrama representativo de uma malha de controle tradicional. O controlador recebe medidas y sobre o comportamento do sistema P e, com base nessa informação, intervém em sua dinâmica pela geração de uma entrada u .

mas e Cibernética (WOLKENHAUER, 2001). No entanto, foram os recentes avanços nas tecnologias e no conhecimento de áreas relacionadas à Biologia Molecular que aceleraram a velocidade com a qual modelos matemáticos de sistemas biológicos são construídos, culminando em um expressivo número de publicações na área de pesquisa denominada Biologia Sistêmica (MACHADO et al., 2011; IDEKER; GALITSKI; HOOD, 2001; KITANO, 2002), do inglês *Systems Biology*.

Esses modelos matemáticos são empregados para diferentes propósitos (EPSTEIN, 2008), dentre os quais destacamos o *controle* da dinâmica dos sistemas sob análise, com aplicações já descritas na seção anterior, e tema desta Tese. A área de Biologia Sistêmica é, assim, ponto de partida para as discussões apresentadas neste documento. Interessa-nos aqui o controle *realimentado* de um sistema biológico, dentro de uma perspectiva fundada na Teoria de Controle. Nesse paradigma, dados um sistema e uma especificação a que esse sistema deve satisfazer, objetiva-se projetar um controlador C que intervém no comportamento do sistema (entrada u) em função das saídas y que ele observa, ou mede, como mostrado na Figura 1. À operação conjunta de C e P dá-se o nome de malha fechada. O comportamento de P , desprovido da intervenção do controlador C , é conhecido como comportamento em malha aberta.

A temática de intervenção na dinâmica de redes intracelulares, segundo as concepções da Teoria de Controle, já foi objeto de diversos estudos da comunidade científica (CURY; BALDISSERA, 2013). De maneira esquemática, podemos categorizá-los em pesquisas teóricas, conduzidas estritamente no âmbito de simulações computacionais, e práticas, em que experimentos para controle *in vivo* de redes biológicas foram efetivamente realizados.

Quanto às pesquisas teóricas, Datta e colaboradores, por exemplo, basearam-se em técnicas de programação dinâmica e controle ótimo para alterar o comportamento de redes de regulação gênica representadas por modelos probabilísticos (DATTA et al., 2007). Nesse trabalho, supôs-se que o sinal de controle pode ser implementado por meio da ativação e/ou repressão de genes já existentes na rede, embora não se tenha detalhado tal procedimento ou se justificado tal hipótese. Langmead e Jha, por outro lado, analisaram o problema de controle de redes de regulação gênica sob a perspectiva de *model-checking* (CLARKE, 1999) e supuseram a possibilidade de introdução de novos nodos para implementar o sinal de controle obtido (LANGMEAD; JHA, 2008), também sem oferecer um tratamento mais aprofundado de como fazê-lo e de quais são as restrições biológicas intrínsecas ao método. Em outro trabalho, Chaves e colegas derivaram controladores para redes de regulação gênica modeladas por equações diferenciais lineares por partes (CHAVES; GOUZÉ, 2011), tendo o aspecto da realização do controlador sofrido das mesmas deficiências daqueles artigos já citados.

Dentre os experimentos práticos, duas abordagens podem ser elencadas. Na primeira delas, os autores valeram-se de dispositivos técnicos externos para controlar em tempo real a dinâmica de redes biológicas sintéticas inseridas em uma população de fungos (MILIAS-ARGEITIS; SUMMERS; LYGEROS, 2011; MENOLASCINA; BERNARDO; BERNARDO, 2011). As variáveis de interesse, em ambos os trabalhos citados, foram medidas por um sistema de microscopia externo e enviadas a um computador, no qual algoritmos decidiam, então, qual entrada de controle deveria ser aplicada para atender às especificações. A entrada era, por fim, aplicada ao sistema por meio de dispositivos de microfluídica (MENOLASCINA; BERNARDO; BERNARDO, 2011) ou por um gerador de pulsos de onda eletromagnética (MILIAS-ARGEITIS; SUMMERS; LYGEROS, 2011).

Uma maneira diferente de abordar o problema de controle de redes biológicas intracelulares foi apresentada por Bagh e colaboradores (BAGH; MANDAL; MCMILLEN, 2011), que projetaram *genes sintéticos* e os introduziram em bactérias, com o propósito de controlar a dinâmica da rede de regulação gênica envolvida com a ruptura da parede celular de bactérias infectadas por vírus. Os genes sintéticos inseridos mediam as variáveis biológicas de interesse e respondiam de forma a satisfazer a especificação desejada, qual seja, a de evitar a ruptura das paredes celulares. Nesse trabalho, no entanto, nenhuma disciplina formal das Engenharias respaldou o projeto dos genes sintéticos, já que a dinâmica do sistema sob controle era demasiado simples.

Adotaremos nesta Tese a segunda dessas abordagens, com o objetivo de garantir que o sistema em malha fechada atinja duas classes possíveis de especificações:

- Evitar um determinado estado celular, ou uma região dos estados possivelmente atingida pelo sistema biológico em malha aberta. Os estados a serem evitados podem, por exemplo, estar associados a distúrbios na fisiologia celular, como parece ser o caso de atratores ligados a diferentes tipos de câncer (HUANG; ERNBERG; KAUFFMAN, 2009);
- Conduzir a célula a um conjunto específico de estados, mantendo-a aí indefinidamente. A relevância biológica dessa especificação pode ser entendida por meio de dois exemplos. No primeiro caso, se os diferentes tipos celulares (neurônios, células musculares, células epiteliais, etc.) são interpretados como *atratores* do espaço possível de fenótipos (KAUFFMAN, 1993), quer dizer, como estados estáveis a pequenas perturbações, então levar uma célula de um atrator a outro pode ser interpretado como induzir a diferenciação celular. Promover tais “caminhadas” no espaço de estados celulares poderia resultar em enormes avanços em áreas como Engenharia Tecidual e terapia com células-tronco (WEBER; FUSSENEGGER, 2009). Como segundo exemplo, citamos aquele de conduzir uma célula tumoral para a região do espaço de estados associada à morte celular programada.

Tais especificações para o sistema em malha fechada mostram que muitos problemas biológicos relevantes podem ser resolvidos, mesmo que o sistema em malha aberta não satisfaça aquelas condições mais conservadoras de controlabilidade difundidas na literatura da Teoria de Controle, analisadas, para o caso específico das redes biológicas, no artigo de Liu et al. (LIU; SLOTINE; BARABÁSI, 2011). Em outras palavras, não é necessário que o sistema biológico se deixe controlar “livremente” para que a aplicabilidade das técnicas de controle se potencialize.

Para projetar um controlador que altere o comportamento de uma dada rede de regulação gênica, utilizaremos como arcabouço teórico ideias advindas da Teoria de Controle Supervisório, TCS (CASSANDRAS; LAFORTUNE, 2008). O emprego de tais técnicas para o problema citado é novo na literatura. Brevemente, na TCS, a síntese de controladores é realizada por meio de algoritmos (procedimentos bem definidos) que recebem como entrada um modelo lógico da planta e as especificações para o sistema em malha fechada. Ambos, em geral, são representados por autômatos finitos, objetos matemáticos caracterizados, a

grosso modo, por um conjunto discreto de estados e por uma função de transição de estados mediada pela ocorrência de eventos. Como saída, o algoritmo provê a especificação de um supervisor S , o controlador, que interage com a planta da seguinte forma: S observa a ocorrência de um evento da planta em malha aberta (alguns eventos podem ser não observáveis). S , então, decide quais eventos da planta devem ser restringidos, ou mesmo, forçados. A ação de controle do supervisor S pode mudar a cada nova ocorrência de um evento observável da planta (CASSANDRAS; LAFORTUNE, 2008).

Aqui cabe voltar, brevemente, ao problema de modelagem. Dado que a TCS será o arcabouço de controle utilizado, admitiremos que o comportamento das células pode ser abstraído por meio de autômatos finitos. Note que essa maneira de tratar sistemas biológicos não é, de forma alguma, artificial, pois: a) o conceito de um conjunto discreto de estados aparece naturalmente na Biologia (p.e. os seres vivos multicelulares possuem um número finito de tipos celulares, dentre eles as células neuronais, epiteliais, musculares, etc.); e b) a dinâmica orientada a eventos também é ubíqua no comportamento biológico (p.e. a ativação de uma proteína por um metabólito pode causar mudanças abruptas no estado celular, como a ativação de um gene).

Uma vez encontrado o supervisor que satisfaça as especificações para o sistema em malha fechada, suporemos, nesta Tese, que sua realização se dará por meio de genes sintéticos, ou seja, genes que são projetados e construídos com o único propósito de alterar o comportamento da célula em questão. Como já argumentado em outro artigo (BENENSON, 2012), os circuitos existentes na natureza, capazes de executar computações sofisticadas, são, por si só, uma clara indicação de que as funções de sensoriamento, controle e atuação podem ser realizadas por moléculas. Ademais, moléculas sintéticas interagem com redes biológicas nativas sem demandar quaisquer alterações do material genético celular já existente, como requerido pela abordagem que emprega equipamentos técnicos externos. Logo, em nosso trabalho, será estabelecida uma ponte natural entre a Biologia Sistêmica e a área denominada Biologia Sintética (ANDRIANANTOANDRO et al., 2006; KHALIL; COLLINS, 2010), dentro do contexto de controle de sistemas biológicos.

1.3 CONTRIBUIÇÕES

As principais contribuições desta tese são:

- 1) Formalização do problema de *controle de redes de regulação gênica*

por meio de genes sintéticos como um problema de Controle Supervisório;

- 2) Extensão da Teoria de Controle Supervisório para tratar o problema de alcançabilidade estrita de estados (i.e. levar o sistema de um estado inicial a um conjunto de estados finais, mantendo-o nesse conjunto), quando ao supervisor é permitido: a) habilitar ou desabilitar certos eventos; e b) forçar eventos, preemptando um subconjunto dos possíveis eventos da planta. Outros pesquisadores abordaram o problema de alcançabilidade estrita, sem, no entanto, conferir ao supervisor a capacidade de preemptar outros eventos (BRAVE; HEYMANN, 1989; KUMAR; GARG; MARCUS, 1993). Como veremos, a preempção em nosso trabalho está intrinsecamente vinculada a hipóteses sobre a dinâmica dos genes nativos e sintéticos. Preempção de eventos também já foi tema de pesquisa na literatura de Sistemas a Eventos Discretos (GOLASZEWSKI; RAMADGE, 1987; BRANDIN; WONHAM, 1994). No trabalho de Golaszewski e Ramadge (GOLASZEWSKI; RAMADGE, 1987), por exemplo, eventos *forçados* são aqueles impostos por uma entrada externa, capazes de preemptar *todos* os eventos da planta (hipótese de preempção forte). No trabalho de Brandin e Wonham, por outro lado, eventos *forçáveis* são introduzidos no contexto de modelos temporizados (aqueles em que a passagem do tempo aparece explicitamente no modelo), preemptando somente os eventos associados à passagem do tempo. Em nossa Tese, adotaremos essa noção fraca de preempção, com uma diferença substancial: o conjunto de eventos que podem ser preemptados pode ser qualquer subconjunto dos eventos da planta, e não somente aqueles que captam o fluir do tempo, e tal escolha, como mencionado acima, dar-se-á em função da dinâmica dos processos biológicos associados à expressão dos genes nativos e sintéticos;
- 3) Determinação de condições necessárias e suficientes para existência de supervisores que resolvem o problema formalizado;
- 4) Desenvolvimento de algoritmo capaz de selecionar aqueles supervisores que: a) resolvem o problema de controle formalizado e b) são realizados por genes sintéticos com menor complexidade de construção;
- 5) Desenvolvimento de algoritmo capaz de selecionar aqueles supervisores que: a) resolvem o problema de controle formalizado e b) resultam no menor consumo de energia para a célula;

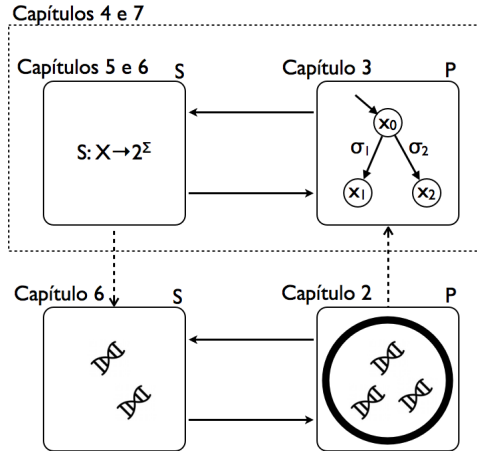


Figura 2 – Estrutura dos capítulos da Tese segundo o diagrama tradicional da Teoria de Controle.

- 6) Avaliação do desempenho, das hipóteses e dos desafios da proposta de controle desenvolvida nesta Tese, quando aplicada ao controle de redes de regulação gênica modeladas por equações diferenciais.

1.4 ORGANIZAÇÃO DO DOCUMENTO

A estrutura linear deste documento pode ser enquadrada logicamente no clássico diagrama de realimentação da Teoria de Controle, representado na Figura 2, em que P e S denotam, respectivamente, o sistema a ser controlado e o supervisor (ou controlador).

Vinculados à planta P da Figura 2, estão os capítulos 2 e 3, que tratam do sistema a ser controlado de uma perspectiva (i) conceitual e (ii) formal, isto é, fundada na apresentação dos diferentes modelos matemáticos existentes na literatura para redes de regulação gênica.

No capítulo 4, mostram-se diferentes estratégias para controlar a dinâmica de sistemas biológicos celulares, articulando, desta forma, todos os elementos da Figura 2. Esse capítulo inclui, também, os conceitos principais para compreensão de como moléculas capazes de medir, controlar e agir podem ser fabricadas; ou seja, como sensores, controladores e atuadores podem ser traduzidos do universo matemático para o nível molecular.

O capítulo 5 traz os principais resultados e contribuições do trabalho, como formalização do problema de controle e apresentação das condições matemáticas necessárias e suficientes para sua resolução.

Se mais de um supervisor S resolve o problema de controle, pode-se levantar a questão de quais soluções são as “melhores”, em que a comparação de uma solução com a outra depende da definição de uma métrica (p.e. consumo de energia). Esse problema de otimidade é tratado no capítulo 6.

O emprego da Teoria de Controle Supervisório pressupõe a existência de um modelo a eventos discretos para o sistema a ser controlado, ou seja, um modelo calcado em dois conceitos básicos: a) a existência de um conjunto discreto de estados; e b) a existência de eventos que levam o sistema de um estado a outro, conferindo-lhe dinâmica. No entanto, as redes de regulação gênica, que correspondem às plantas P neste trabalho, são comumente modeladas por um formalismo matemático diferente, as equações diferenciais ordinárias, em que o conjunto de estados que o sistema pode assumir é contínuo, sendo a mudança de estados determinada pela passagem do tempo e não pela ocorrência de eventos. Portanto, é de interesse investigar como o ferramental desenvolvido nesta Tese se articula com plantas modeladas por equações diferenciais, o que é feito no capítulo 7.

Finalmente, no capítulo 8, faz-se uma discussão global do trabalho realizado, pontuando os resultados, desafios e perspectivas para trabalhos futuros.

Por se tratar de uma tese interdisciplinar, seu conteúdo costura tecidos de variados domínios, em diferentes níveis de abordagem e voltados a públicos específicos. Sugerimos, pois, três maneiras distintas de lê-la (Figura 3): uma voltada aos pesquisadores advindos das Engenharias, Matemática ou Ciências da Computação (grupo 1); outra direcionada àqueles advindos de uma formação fortemente biológica (grupo 2); e, por último, uma apropriada aos membros da comunidade de Biologia Sistêmica, familiarizados com o problema biológico e com os respectivos métodos formais (grupo 3). Essas três maneiras distintas de olhar a mesma Tese são esquematizadas na Figura 3.

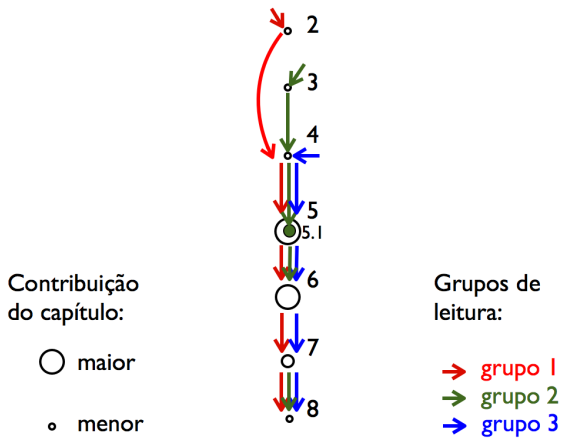


Figura 3 – Sugestões de leitura dos capítulos da Tese. As circunferências maiores indicam os capítulos com mais contribuições originais do documento.

2 FUNDAMENTOS DE BIOLOGIA CELULAR

2.1 PROTEÍNAS E FUNÇÕES CELULARES VITAIS

As células são a unidade básica da vida. Presentes na totalidade das dezenas de milhões de espécies existentes na Terra, as células exibem todas aquelas características que demarcam a fronteira entre a vida e a matéria morta (ALBERTS et al., 2008). Elas autorregulam-se, ajustando seu meio interno em resposta a variações do ambiente, replicam-se e processam quimicamente moléculas maiores para delas extrair energia e blocos construtores de outras macromoléculas. Comunicam-se entre si e, por isso, permitem-se formar estruturas maiores, que podem alcançar as trilhões de células, fazendo emergir um tipo de organização tão complexa que é capaz de refletir sobre si mesma, como acontece com os seres humanos.

De um ponto de vista mais formal, as células são *sistemas dinâmicos* que a) reconhecem estímulos externos e medem estados internos; b) integram essas informações para tomada de decisões coordenadas; e c) ensejam ações correspondentes, como diferenciação e replicação, por exemplo. Para viabilizar a implementação de todas essas funções, as células necessitam de energia livre, que obtêm a partir da degradação de compostos advindos do meio externo. Todas as funções lógicas citadas resultam de interações físico-químicas entre inúmeros compostos, interações estas altamente dependentes da *configuração espacial* de cada um deles. Forma e função, portanto, não estão dissociadas, é a natureza adiantando-se aos preceitos da arquitetura moderna.

Em todas as funções celulares concebíveis, as proteínas exercem um papel de destaque. Segundo um excerto do livro *Biologia Molecular da Célula* (ALBERTS et al., 2008):

As enzimas [tipo de proteínas] constituem as intrincadas superfícies moleculares que promovem suas múltiplas reações químicas em uma célula. As proteínas imersas na membrana plasmática formam canais e bombas que controlam a passagem de pequenas moléculas para dentro e para fora das células. Outras proteínas carregam mensagens de uma célula para a outra, ou agem como integradoras de sinais, direcionando conjuntos de sinais da membrana plasmática para o núcleo celular. Outras, ainda, funcionam como minúsculas máquinas moleculares com partes móveis: as cine-

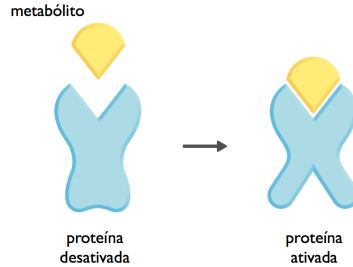


Figura 4 – Ativação de uma proteína após ligar-se a um metabólito.

sinas, por exemplo, impulsionam organelas pelo citoplasma (...). Proteínas especializadas agem como anticorpos, toxinas, hormônios, moléculas anticongelantes, fibras elásticas, cordas ou como fontes de luminescência.

As Figuras 4 e 5 ilustram, respectivamente, uma possível mudança de conformação na estrutura de uma proteína ao ligar-se a um metabólito intracelular, que a ativa, e duas das possíveis funções celulares conduzidas por proteínas especializadas.

Dado que as proteínas são sintetizadas a partir de informações contidas nos genes, em um processo que será detalhado adiante, resta claro que a compreensão da dinâmica celular passa pelo entendimento de *quando* e *como* cada um dos genes é ativado, disparando a síntese desses relevantes polímeros biológicos que são as proteínas.

2.2 EXPRESSÃO GÊNICA: TRANSCRIÇÃO E TRADUÇÃO

O processo de síntese de proteínas, também chamado de *expressão gênica*, é composto por duas etapas principais: a *transcrição*, em que a informação hereditária armazenada no DNA é utilizada como molde para a geração de uma molécula de RNA, e a *tradução*, em que a proteína é sintetizada a partir dessa molécula intermediária de RNA. Algumas vezes, o produto final da expressão gênica é o próprio RNA e, portanto, o processo finda na transcrição. Quando isto não acontece, a molécula intermediária de RNA é denominada RNA mensageiro, mRNA.

Essa sequência de conversão molecular faz-se necessária, pois o

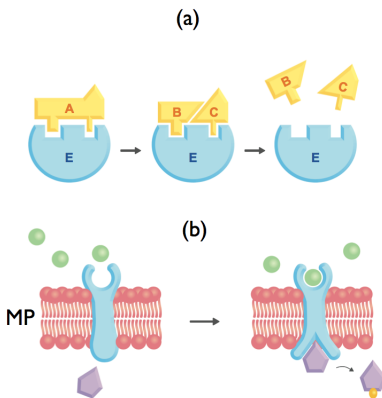


Figura 5 – Dois dos muitos papéis desempenhados pelas proteínas em células: em (a) uma enzima catalisa uma reação química de degradação do composto A; em (b) uma proteína no papel de receptor na membrana plasmática (MP) liga-se a uma molécula extracelular (ex.: hormônios, representados por esferas na figura), resultando em mudanças na estrutura química de moléculas no meio intracelular.

DNA é eficiente como meio de *armazenagem e replicação* da informação hereditária, mas se mostra inapto para executar as mais diversas funções celulares (p.e., catalisar reações químicas). As proteínas, por outro lado, são altamente flexíveis para desempenhar inúmeras tarefas, mas inábeis para se autorreplicar. Utilizando moléculas distintas, a vida extrai o que há de melhor em cada uma delas.

DNA, RNA e proteínas, as três principais moléculas envolvidas na expressão de um gene, são polímeros, ou seja, moléculas formadas pelo encadeamento de compostos químicos menores, os monômeros. O DNA é uma molécula de fita dupla. Cada fita é formada por uma sequência dos monômeros adenina, timina, citosina e guanina, representados pelas letras A, T, C e G, respectivamente. O monômero A de uma fita liga-se, por afinidade química, ao T da outra, enquanto o monômero C pareia-se com o G. Esse pareamento entre as fitas do DNA confere à molécula a sua célebre estrutura espacial rígida no formato helicoidal.

Os blocos construtores do RNA também são quatro, A, U, C e G, com a uracila substituindo a timina que aparece no DNA. Os monômeros A, T, C, G e U são também chamados de *nucleotídeos*. Uma

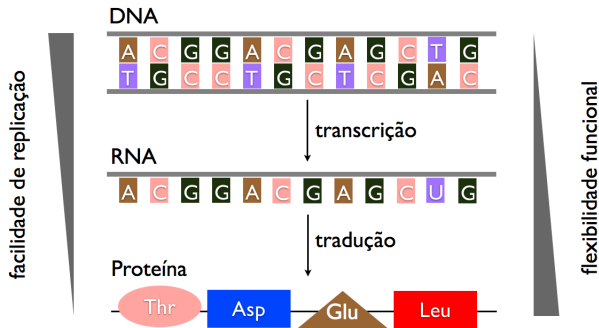


Figura 6 – Representação esquemática dos processos de transcrição e tradução de um gene. Cadeias de aminoácidos (simbolizados aqui pelas iniciais Thr, Asp, Glu, Leu) podem assumir uma gama enorme de conformações espaciais com diferentes propriedades químicas, o que lhes confere grande variedade funcional. A molécula de DNA, por outro lado, apresenta-se em uma configuração rígida, mas extremamente eficiente para replicação e armazenamento da informação hereditária.

outra diferença significativa entre RNA e DNA reside no fato de que o RNA é uma molécula de fita simples. Como consequência, sua estrutural espacial não é rígida como a do DNA, mas flexível, dependente da sequência de nucleotídeos que a compõem.

As proteínas, por sua vez, são constituídas por *aminoácidos*, monômeros radicalmente diferentes daqueles que compõem o DNA e o RNA. São 20 os aminoácidos presentes na natureza. Suas diferentes propriedades químicas conferem às proteínas uma grande flexibilidade em termos de conformação espacial e, portanto, funcional. A Figura 6 mostra esquematicamente as etapas de transcrição e tradução e destaca como a informação contida nos genes é codificada de maneira diferente nas três moléculas, que, como descrito, possuem propriedades distintas.

Ambas as etapas do processo de expressão gênica, transcrição e tradução, demandam a participação de uma complexa maquinaria biomolecular, que “lê” e converte a informação contida no DNA em cadeias proteicas, como se depreende da exposição simplificada contida nos parágrafos que seguem. Uma apresentação mais detalhada dos processos de transcrição e tradução pode ser obtida em referências de *Biologia Celular* (ALBERTS et al., 2008).

A síntese de um RNA a partir da informação contida no gene é catalisada por uma proteína, a enzima RNA polimerase (RNAP). Essa enzima ancora-se em uma região gênica denominada *promotor*, e move-se ao longo do DNA, catalisando a formação de uma cadeia de RNA cuja sequência de nucleotídeos é complementar àquela do trecho de DNA lido. Em outras palavras, sempre que a RNAP confronta-se com um monômero A, T, C ou G no DNA, ela adiciona um monômero U, A, G ou C, respectivamente, à fita do RNA em formação. A RNAP comete, aproximadamente, um erro a cada 10^4 nucleotídeos copiados (por vezes, a RNAP pode corrigir o erro cometido, retirando o nucleotídeo erroneamente adicionado e substituindo-o pelo correto).

A decodificação da informação contida no DNA em uma proteína funcional dá-se na etapa de tradução. É nesse passo que um polímero composto por quatro diferentes nucleotídeos, o RNA, é utilizado para gerar uma molécula escrita em uma linguagem tão diferente como é aquela das proteínas, composta por 20 aminoácidos. Essa tarefa é, em grande parte, conduzida por uma máquina molecular denominada ribossomo, uma associação de mais de 50 proteínas diferentes, e por RNAs especiais chamados RNA transportadores, tRNA. Assim como a RNAP localiza um sítio específico no DNA para se acoplar e iniciar a transcrição, também o ribossomo reconhece uma região específica do RNA antes de dar início à tradução. Uma vez iniciada, a cadeia de RNA é processada pelo ribossomo, que a cada três nucleotídeos associa um aminoácido específico, usando para isso os tRNAs. Ou seja, as “palavras” do RNA, chamadas códons, possuem três letras e são traduzidas, com auxílio de um “dicionário molecular”, para um aminoácido em particular. Existem $4 \times 4 \times 4 = 64$ códons a serem mapeados em um conjunto de 20 aminoácidos. Como tal função não pode ser injetora, diferentes códons codificam, então, para o mesmo aminoácido. Dessa forma, diferentes RNAs podem resultar na mesma proteína. Palavra a palavra, prossegue o ribossomo na tradução, até que o “texto” proteico se finalize.

O processo de leitura do DNA e correspondente síntese proteica não acontece, necessariamente, a todo o momento e com a mesma intensidade. Ele é controlado e modulado pelo contexto químico celular, ou o estado celular, como será detalhado na próxima seção.

2.3 REDES DE REGULAÇÃO GÊNICA

Determinadas proteínas, denominadas *proteínas de regulação gênica* ou simplesmente *fatores de transcrição*, possuem a capacidade de ligar-se a regiões específicas do DNA, interferindo no acoplamento da RNAP à região promotora, ora facilitando-o, ora inibindo-o. Os produtos finais dos genes, as proteínas, podem, portanto, alterar o início da expressão gênica. Logo, o fluxo de informações na célula é bidirecional, do DNA às proteínas, e destas, àquele. A ideia de que genes podem ser ativados ou desativados por proteínas é relativamente nova, tendo sido desenvolvida na metade do século XX a partir de estudos com bactérias (ALBERTS et al., 2008). Estima-se, por exemplo, que aproximadamente 8% de todos os genes em seres humanos (um número em torno de 25.000) codifiquem fatores de transcrição, o que mostra a importância do mecanismo de regulação da ativação gênica.

Um gene pode influenciar a transcrição de um ou mais genes, incluindo a si próprio, por meio do fator de transcrição por ele codificado. Mais de um gene pode impactar na expressão de um mesmo gene, por meio da associação cooperativa das proteínas de regulação gênica (por vezes, centenas de fatores de transcrição controlam a expressão de um único gene). Tais propriedades dos fatores de transcrição implicam a formação das chamadas *redes de regulação gênica*, que, por coordenarem a ativação ou repressão de inúmeros outros genes, exercem papel destacado na dinâmica celular. Nessas redes, os nodos representam genes, e um arco partindo de um gene a outro indica o impacto que aquele possui na transcrição deste, impacto, como já comentado, exercido por meio das proteínas de regulação gênica. A Figura 7 mostra um exemplo de uma rede de regulação em fungos.

Depreende-se do exposto que os genes possuem duas regiões principais, com funções distintas: a *região codante*, que especificará à maquinaria celular qual proteína deve ser produzida, ou seja, qual a sequência de aminoácidos a ser sintetizada (“o que fazer”); e a *região de regulação*, ou região de controle gênico (i.e. promotor mais sítios de ligação de fatores de transcrição ao DNA), que determina as circunstâncias nas quais o gene será transcrito (“quando fazer”). A Figura 8 mostra o padrão gráfico que será empregado neste documento para retratar o gene e suas diferentes regiões.

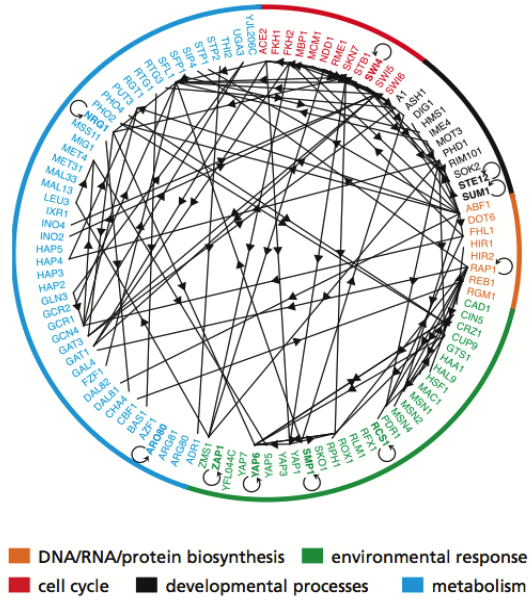


Figura 7 – Parte da rede de regulação gênica da levedura *Saccharomyces cerevisiae*. Os nodos, identificados por códigos alfanuméricos, correspondem aos genes, e as setas representam as relações entre genes. Figura extraída de (ALBERTS et al., 2008).

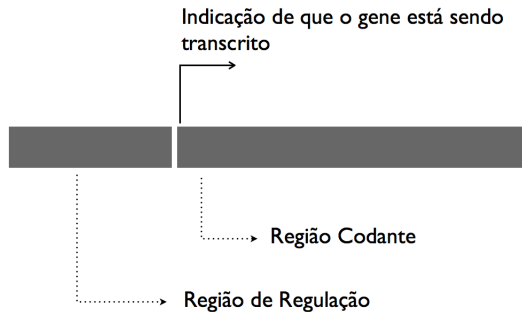


Figura 8 – Representação de um gene e suas regiões codante e de regulação.

2.4 SISTEMA DE CONTROLE CELULAR: UM EXEMPLO

Para finalizar o capítulo, descreveremos um simples exemplo que ilustra como a teia das redes biológicas é tecida de maneira que a célula responda satisfatoriamente às condições do meio externo. O exemplo, uma resposta adaptativa de bactérias a temperaturas externas elevadas, foi retirado do livro *Control Theory and Systems Biology* (IGLESIAS; INGALLS, 2010). Neste sistema realimentado (ver Figura 9), podem-se identificar os seguintes elementos:

1. “quantidade de proteínas desnaturadas”, ou seja, a quantidade de proteínas cujas estruturas tridimensionais foram danificadas, como a variável a ser controlada (a saída y do sistema);
2. “manter baixa a quantidade de proteínas desnaturadas”, como o objetivo de controle, e
3. “quantidade de proteínas remodeladoras da estrutura de outras proteínas”, como a entrada de controle u .

A resposta ao choque térmico em bactérias é uma reação celular a elevadas temperaturas, dentre outras condições adversas. Tais condições frequentemente levam a danos na estrutura de diversas proteínas e, como consequência, à perda de suas funções. Para assegurar a manutenção das funções celulares mesmo em temperaturas elevadas, a célula dispõe das chamadas proteínas de choque térmico (HSP, do inglês *heat shock proteins*). Muitas das HSPs ajudam a remodelar aquelas proteínas desnaturadas; outras HSPs degradam e removem as proteínas sem função. A resposta bacteriana deve manter um fino compromisso entre o efeito restaurador provido pelas HSPs e os custos, material e energético, associados à síntese dessas proteínas. Na bactéria *E. coli*, esse balanço é atingido por meio de uma arquitetura intrincada de laços de controle centrados na proteína σ^{32} , que ao se acoplar com a RNAP, permite o início da transcrição dos genes que codificam as HSPs, iniciando, assim, a síntese delas.

A baixas temperaturas, a tradução do mRNA que codifica para a proteína σ^{32} , *rpoH*, é impossibilitada por questões moleculares estruturais. A elevadas temperaturas, no entanto, essas condições adversas à tradução desaparecem, permitindo a síntese da proteína σ^{32} e, por isso, a produção das HSPs. Esse mecanismo, em linguagem da teoria de controle, implementa, portanto, um sensor de temperatura, associado a uma resposta de controle do tipo *feedforward*, já que a bactéria responde a uma condição externa (o aumento da temperatura), antes

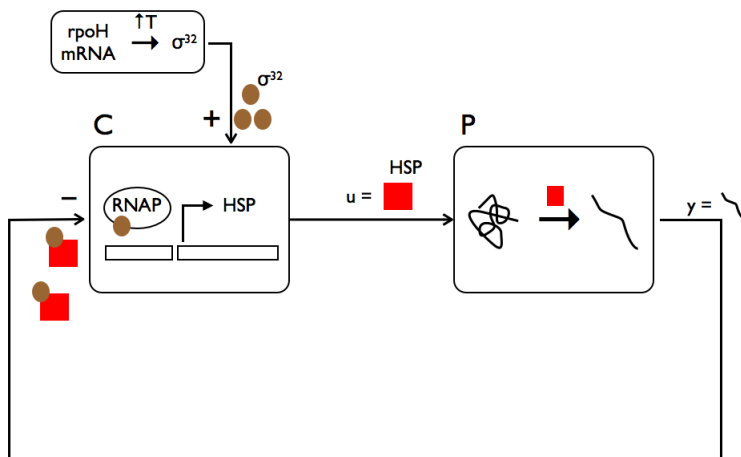


Figura 9 – Representação do sistema de controle implementado por bactérias para reagir ao aumento da temperatura intracelular, evitando desnaturação de proteínas. A linha enovelada em P denota proteínas com estrutura e forma preservadas, ao passo que a linha estendida representa a proteína desnaturada. A variável a ser controlada y é a concentração de proteínas desnaturadas. A regulação dessa variável é obtida pela manipulação de u , a concentração de proteínas HSP.

mesmo de haver desnaturação das proteínas (a variável a ser controlada, ou a saída da planta).

Essa resposta *feedforward* aumenta os níveis de HSPs, que interagem com as proteínas cujas estruturas espaciais foram danificadas, remodelando-as ou degradando-as. As HSPs em excesso, por sua vez, acoplam-se à proteína σ^{32} , desativando-a.

Assim, a realimentação negativa, descrita no parágrafo anterior, e a ação *feedforward* relatada agem orquestradamente para garantir um baixo nível de proteínas desnaturadas, com reduzido custo energético na síntese de biomoléculas.

Note que, conforme apontado por Hana El-Samad (IGLESIAS; INGALLS, 2010), as funções de sensoriamento, controle e atuação da malha de controle biológico descrita não são separadas como acontece na maioria dos sistemas técnicos, embora elas estejam presentes. O exemplo ilustra também que desvendar a função de determinadas vias bioquímicas celulares não é tarefa fácil, já que envolve a participação de inúmeras moléculas que interagem entre si, formando sistemas de controle que, em geral, não revelam, didaticamente, sua organização lógica.

3 MODELAGEM DE REDES DE REGULAÇÃO GÊNICA

3.1 CONCEITOS PRELIMINARES

Discorreremos nesta seção sobre os conceitos de *sistema*, *modelo* e *estado* de um sistema, definições centrais para o entendimento das discussões propostas nesta Tese.

Existem inúmeras acepções para sistema, em um espectro que vai das mais populares e difundidas, como aquelas encontradas em dicionários comuns, às mais técnicas e intrincadas, por exemplo, “sistema é uma família de mapeamentos de um espaço topológico nele mesmo” (MESAROVIC; TAKAHARA, 1989). Consideraremos, para nossos fins, que um sistema é um *conjunto de entidades, concretas ou abstratas, relacionadas entre si e cujo comportamento coletivo se deseja estudar*. A tudo aquilo que está fora do sistema, denominaremos *ambiente*. A definição de sistema, assim, é subjetiva, pois a demarcação da fronteira entre sistema e ambiente depende das motivações de quem o analisa. No contexto biológico, por exemplo, o que é ambiente para um bioquímico que estuda especificamente a via de degradação da glicose, é encampado na concepção de sistema para aquele biólogo interessado no comportamento da célula como um todo. A delimitação entre sistema e ambiente já é, por isso, consequência de uma *abstração* da realidade, um primeiro passo na construção de um *modelo* do objeto sob análise.

Ao se estudar um sistema, deseja-se, em geral, compreender de que forma a manipulação de alguns de seus atributos altera o comportamento de outros atributos que o caracterizam. Por exemplo, interessa ao pesquisador que investiga uma rede de regulação gênica, seu sistema, descrever como varia a taxa de transcrição de um dado gene dessa rede, um atributo do sistema, ao se aumentar ou diminuir a concentração intracelular de uma proteína de regulação gênica específica, outro atributo. Os atributos passíveis de manipulação e aqueles cujos comportamentos pretende-se observar estão associados ao conceito de *entrada* e *saída* do sistema, respectivamente. A observação da saída de um sistema corresponde à observação de um fenômeno por ele gerado. Um sistema pode ser visto, conseqüentemente, como um mapa que relaciona entradas a saídas, ou ainda, como um agente capaz de tomar decisões, produzindo respostas (a saída) a cada estímulo vindo do ambiente (a entrada), de maneira a atingir um determinado objetivo (MESAROVIC; TAKAHARA, 1989).

O estudo de um sistema, entretanto, não se limita somente à observação passiva de seu comportamento quando estimulado. Busca-se na pesquisa entender o “porquê” da saída observada, de maneira que se possa antecipar a resposta do sistema a um estímulo, *sem* a necessidade de experimentação prática, porque impossível, inviável, custosa ou moralmente indesejável (veja-se o embate na área de experimentação animal). O ser humano deseja, assim, prever fenômenos.

A forma adotada pela ciência para lograr êxito nessa tarefa passa pelo estabelecimento de uma relação entre o sistema sob análise e um *sistema formal*, não necessariamente único, caracterizado por *proposições* e *regras de produção* que, aplicadas às proposições, originam novas proposições (ROSEN, 1991). Embora os sistemas formais existam independentemente dos sistemas reais, ambos relacionam-se a partir do estabelecimento de uma *semântica* para o sistema formal, ou seja, uma interpretação das proposições e regras de produção que guarde um vínculo de significação com o sistema real estudado. Assim, intuitivamente, as proposições válidas do sistema formal estão relacionadas aos possíveis fenômenos observados no sistema real, enquanto as regras de produção, ao conjunto de leis que regem o comportamento deste sistema.

Para exemplificarmos as ideias acima, considere uma rede de regulação gênica composta por dois genes que se reprimem mutuamente pela ação dos fatores de transcrição que eles codificam, como mostrado na Figura 10. Considere também um modelo formal composto pelos seguintes elementos:

1. Proposições: $P_1 = (0, 0)$, $P_2 = (0, 1)$, $P_3 = (1, 0)$ e $P_4 = (1, 1)$, em que cada uma delas é interpretada como uma assertiva sobre o estado de cada um dos genes da rede, ativado ou desativado, 1 ou 0, respectivamente. $P_3 = (1, 0)$, por exemplo, denota a situação em que v_1 está ativado (i.e., sendo expresso) e v_2 , desativado;
2. Regra de produção: ‘ $R_1 = \text{Se } (0, 0), \text{ então } (1, 0)$ ’. Essa regra pode ser interpretada como ‘se ambos os genes estão desativados, ou seja, na ausência dos inibidores, v_1 é ativado antes de v_2 ’.

Suponha, ainda, que montemos um experimento em que v_1 e v_2 estejam ambos desativados, situação correspondente à proposição de partida P_1 no sistema formal. Aplicando-se a regra R_1 à proposição P_1 , obtém-se a proposição P_3 . Assim, se o sistema formal capta adequadamente o comportamento do sistema real, poder-se-á inferir que, em algum instante futuro, o gene v_1 será ativado, enquanto v_2 permanecerá

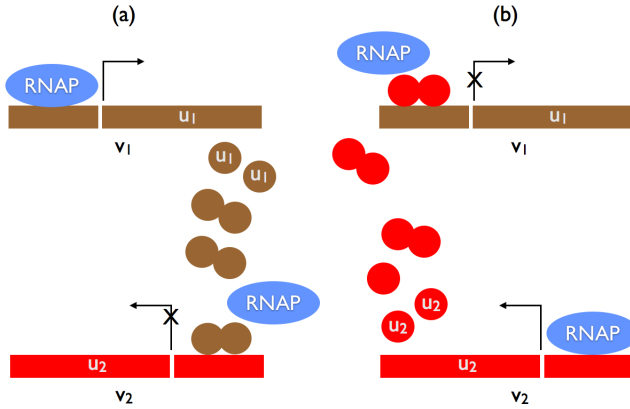


Figura 10 – Configuração de uma rede de regulação gênica mutuamente repressora, composta por dois genes v_1 e v_2 . Os genes v_1 e v_2 são expressos na ausência dos homodímeros formados pelas proteínas u_2 (codificada por v_2) e u_1 (codificada por v_1), respectivamente. Os dímeros acoplam-se à região promotora do gene-alvo, impedindo a ligação do gene com a RNAP. (a) v_1 sendo expresso, inibindo v_2 . (b) v_2 sendo expresso, inibindo v_1 .

desativado. Note que a predição do comportamento da rede de regulação gênica pressupõe: a) uma forma de codificar observações do sistema real em proposições do sistema formal; b) a aplicação da(s) regra(s) de produção do sistema formal à proposição válida; e c) uma maneira de decodificar proposições do sistema formal em observações do sistema real. Essa relação entre o sistema formal e o sistema real está parcialmente representada na Figura 11 (ROSEN, 1991). Idealmente, busca-se um sistema formal que satisfaça as seguintes propriedades:

- se um fenômeno A do sistema real é observado num dado momento, e A é *causa* de um outro fenômeno B , então é verdade que a aplicação da(s) regra(s) de produção do sistema formal à proposição que corresponde à codificação de A resulta em uma proposição que, quando decodificada, é igual a B ; e
- se a aplicação da(a) regra(s) de produção do sistema formal a uma proposição P_i qualquer resulta em uma proposição P_j , então o fenômeno correspondente à decodificação de P_i no sistema real é causa para a ocorrência do fenômeno que resulta da decodificação de P_j .

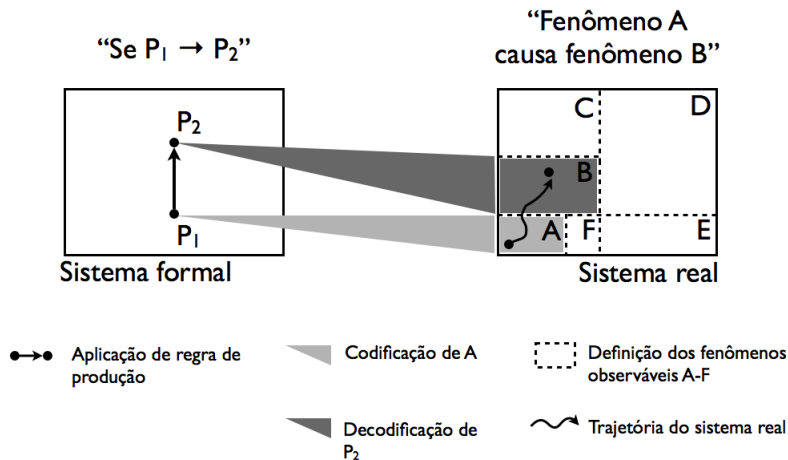


Figura 11 – Sistemas formais podem ser utilizados para prever o comportamento de sistemas reais. Nesse caso, codifica-se um fenômeno observado como uma proposição do sistema formal (na Figura, A é codificado em P_1), aplicam-se as regras de produção do sistema formal à proposição P_1 , decodificando-se as proposições resultantes (na Figura, P_3) em fenômenos do sistema real (saída B).

Se ambas as propriedades são válidas para todas as situações de interesse, tem-se, dentro desse contexto, um *modelo* para o sistema real. Não existe nenhum procedimento conhecido para obtenção de um sistema formal com tal propriedade. A modelagem de um sistema, assim sendo, é uma arte (ROSEN, 1991).

Dados um sistema real e um sistema formal que o modele, o *estado* do sistema real pode ser definido como o *mínimo* conjunto de seus atributos que precisam ser conhecidos (e codificados) no modelo, de modo que a aplicação das regras de produção deste resultem em somente uma *única* proposição possível (e, por conseguinte, uma saída observável).

O conhecimento do estado atual do sistema, assim, elimina a necessidade de registro de seu passado na tarefa de prever o futuro (MESAROVIC; TAKAHARA, 1989) e, além disso, conduz a uma predição *determinística* do comportamento do sistema. Logo, por essa definição, muito influenciada pelos pressupostos da mecânica newtoniana, conhe-

cido o estado atual, só existe um futuro para o sistema em questão, não havendo, pois, espaço para estocasticidade. Tal concepção tornou-se célebre no experimento mental denominado Demônio de Laplace (MITCHELL, 2009), que conclui não existir incerteza para um intelecto capaz de determinar com precisão absoluta o estado atual do universo (que, nesse caso, faz o papel de sistema).

Embora tal determinismo absoluto tenha se eclipsado pelas interpretações modernas da Mecânica Quântica, o conceito de estado continua a desempenhar papel central na ciência moderna (ROSEN, 1991), seja na figura da *função de onda* em sistemas quânticos, seja nas abstrações newtonianas do mundo macroscópico que nos cerca. Há autores, no entanto, que questionam esse paradigma da ciência moderna, calcado na noção de estados, quando aplicado à compreensão dos fenômenos centrais da vida (ROSEN, 1991).

Na maioria das aplicações práticas, no entanto, as informações necessárias para prever unicamente o comportamento de um dado sistema são muitas, o que implicaria dificuldades técnicas, até mesmo impraticabilidade, para medição delas, além de custos computacionais dispendiosos para o cálculo do estado futuro a partir do estado atual (não abordaremos aqui o interessante assunto de como incertezas na determinação do estado inicial de um sistema podem levar a previsões errôneas de seu comportamento futuro, temática presente na Teoria do Caos e bastante explorada em outras referências (STROGATZ, 2001)). Acrescente-se ainda o fato de que nenhum sistema é absolutamente isolado de seu ambiente, permitindo que certas perturbações externas, de difícil quantificação, alterem seu estado. Por isso, escolhe-se, comumente, de todas aquelas informações que deveriam idealmente fazer parte do estado de um sistema, apenas um *subconjunto* delas como representativas de seu estado. Tal solução de compromisso facilita a compreensão do comportamento de sistemas reais às custas da adição de incertezas ou não-determinismos em suas descrições.

Voltemos ao exemplo da rede de regulação gênica da Figura 10. Ao escolhermos unicamente o nível de ativação de cada um dos genes da rede como representativos de seu estado, desprezando, em lista não exaustiva, as concentrações e localizações espaciais de seus fatores de transcrição, bem como as concentrações e localizações das outras moléculas do meio intracelular, naturalmente cometemos erros na previsão do estado futuro da rede (ou seja, no nível de expressão gênica que se observará). Nossa previsão sobre o comportamento do sistema é válida sob certas circunstâncias e dentro de uma margem de erro, maior ou menor em função da magnitude do impacto das variáveis negligenciadas

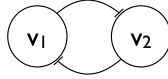


Figura 12 – Grafo associado à rede de regulação gênica da Figura 10.

na expressão dos genes.

Essa digressão mostra que a modelagem matemática de um sistema é resultado de um complexo processo de *abstração* da realidade, fundada em hipóteses cujos graus de validade não são absolutos, mas dependentes dos objetivos e expectativas que norteiam o trabalho de pesquisa.

3.2 MODELAGEM DE REDES DE REGULAÇÃO GÊNICA

A topologia de uma rede de regulação gênica, em outras palavras, sua estrutura, pode ser representada por um objeto matemático denominado *grafo direcionado*, um par (V, E) , no qual: a) $V = \{v_1, \dots, v_N\}$ é o conjunto de nodos da rede que, nesse caso, representam os genes da rede de regulação; e b) E é o conjunto de pares ordenados (v_i, v_j) , denotando que o produto final codificado por v_i , RNA ou fator de transcrição, afeta de alguma maneira a transcrição do gene v_j . Para o simples exemplo da rede de regulação mostrada na Figura 10, tem-se $V = \{v_1, v_2\}$ e $E = \{(v_1, v_2), (v_2, v_1)\}$.

As interações proteína-DNA modeladas pelos arcos do grafo direcionado (V, E) não são qualitativamente iguais. Para incorporar o fato de que proteínas de regulação podem atuar como ativadores ou inibidores transcricionais, utilizaremos dois tipos de arcos para representar um grafo, com o objetivo de diferenciar arcos ativadores dos arcos inibitórios. Aqueles serão representados por setas tradicionais (não mostrados na Figura 12), enquanto estes, por arcos terminados em um segmento de reta (ver Figura 12).

Grafos que captam a topologia da rede, embora importantes para sua compreensão, nada dizem sobre o comportamento *dinâmico* do sistema, ou seja, sobre as possíveis sequências de estados que o sistema pode percorrer. Assim, do ponto de vista dinâmico, a definição da topologia de uma rede traduzida em um grafo é somente um ponto de partida, que será incorporado por modelos matemáticos, juntamente

com outras informações, para construir as possíveis trajetórias que o sistema pode exibir. Existem diversas abordagens matemáticas capazes de conferir dinâmica aos grafos estáticos descritos. Essas abordagens podem ser divididas em três categorias principais: modelos lógicos, contínuos e estocásticos (KARLEBACH; SHAMIR, 2008).

Faremos nas próximas subseções uma breve revisão dos principais representantes dessas classes. É importante notar que cada modelo matemático discutido a seguir faz uso de um sistema formal distinto para explicar os fenômenos observados em redes de regulação gênica. Assim, eles diferem quanto à forma de codificar saídas de uma rede gênica em proposições do sistema formal, bem como quanto à maneira de definir as regras de produção e, finalmente, de decodificar proposições em fenômenos do sistema tratado. *Todos*, no entanto, já foram empregados para compreender algum aspecto do comportamento de redes de regulação gênica reais. Descrições mais detalhadas dos formalismos que serão abordados podem ser encontradas na literatura recente (MACHADO et al., 2011; JONG, 2002; KARLEBACH; SHAMIR, 2008; SCHLITT; BRAZMA, 2007; IGLESIAS; INGALLS, 2010; KELL; KNOWLES, 2010).

3.2.1 Modelos Contínuos

Modelos contínuos descrevem o comportamento de variáveis que assumem valores pertencentes ao conjunto dos números reais, além de incluírem explicitamente o tempo nas equações matemáticas. No contexto das redes de regulação, eles formalizam de que maneira a concentração das moléculas envolvidas na expressão gênica, como mRNA, proteínas e outras moléculas menores, varia ao longo do tempo (KARLEBACH; SHAMIR, 2008; JONG, 2002). Um exemplo de aplicação de tal técnica pode ser encontrado no trabalho de Li e colaboradores, que modelaram o ciclo celular da bactéria *Caulobacter crescentus* (LI; BRAZNIK; TYSON, 2008).

Muitos experimentos biológicos geram dados contínuos (como taxas de reações bioquímicas, variação da massa celular, tamanho de ciclo celular e taxa de transcrição, etc.), que, por essa razão, podem ser diretamente incorporados em modelos contínuos (KARLEBACH; SHAMIR, 2008). Por serem uma representação “mais fina” do comportamento de uma rede biológica, tais modelos são adequados para fins de emulação da dinâmica do sistema em questão. No entanto, a análise de suas propriedades (p.e. a resposta a perguntas do tipo “é possível que a rede de regulação gênica atinja um estado em que este e aquele gene estejam si-

multaneamente ativados?") é tarefa complicada, quando possível, para redes com muitos nodos.

Mostraremos aqui um tipo particular de modelo contínuo, as equações diferenciais ordinárias. Elas constituem-se em representações determinísticas do sistema em questão, ou seja, descrições em que se desprezam o acaso resultante de ruídos e/ou incertezas sobre os parâmetros e os mecanismos de operação da rede biológica.

Considere, como exemplo, a rede de regulação gênica representada pelo grafo da Figura 12. Tomemos como estado desse sistema o par (x_1, x_2) , em que x_1 denota a concentração da proteína u_1 e x_2 , a concentração da proteína u_2 . Para modelar o comportamento dinâmico de x_1 e x_2 , na óptica dos formalismos contínuos, torna-se necessário escrever equações para a *taxa de variação no tempo* de cada uma dessas variáveis, medida em unidades com dimensão de [concentração]/[tempo] e representada matematicamente por $\frac{d(x_i)}{dt}$, com $i \in \{1,2\}$. As equações diferenciais captam matematicamente o balanço entre produção e degradação de cada um dos compostos rastreados pelo modelo. Tem-se, portanto:

$$\frac{d(x_i)}{dt} = \text{taxa de síntese de } x_i - \text{taxa de degradação de } x_i.$$

A primeira parcela da equação anterior, a taxa de síntese de x_i , pode ser modelada matematicamente por expressões derivadas do estudo da cinética de reações bioquímicas, como a dinâmica de Michaelis-Menten ou as curvas de Hill (JONG, 2002). Já a segunda parcela pode ser dada simplesmente por um termo proporcional a x_i , incorporando a ideia de que a diminuição na quantidade de uma dada proteína, pela ação conjunta de degradação, difusão ou diluição devido ao crescimento celular, é tanto maior quanto maior a sua concentração. Assim, uma possível expressão para o sistema composto pelos dois genes mutuamente repressores é:

$$\begin{aligned} \frac{d(x_1)}{dt} &= \frac{k_1^s}{1 + k_1 y_2^{n_y}} - d_1 x_1 \\ \frac{d(x_2)}{dt} &= \frac{k_2^s}{1 + k_2 x_1^{n_x}} - d_2 x_2 \end{aligned}$$

em que as constantes k_i^s , k_i e d_i denotam, respectivamente, a taxa de síntese de uma proteína na ausência de sua inibidora, a magnitude de repressão exercida por uma proteína na transcrição do gene ao qual ela se liga e, finalmente, a constante de degradação de cada uma das

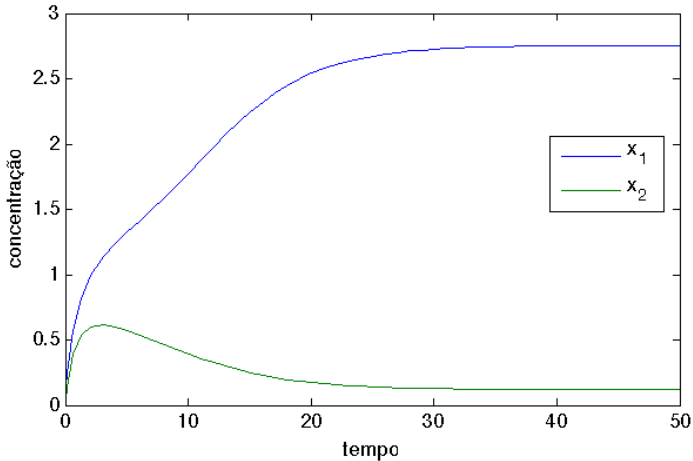


Figura 13 – Comportamento temporal das variáveis x_1 e x_2 quando $x_{0,1} = 0.1$ e $x_{0,2} = 0$.

proteínas. Os valores de n_x e n_y estão associados à multimerização das proteínas de regulação gênica. No exemplo específico tratado, como u_1 e u_2 foram homodímeros (ver Figura 10), $n_x = n_y = 2$.

A resolução das equações diferenciais mostradas acima resulta em curvas para as variáveis x_1 e x_2 que são funções do tempo. As figuras 13 e 14 mostram, para uma seleção arbitrária de parâmetros, o comportamento de cada uma dessas variáveis para duas condições iniciais diferentes, ou seja, para dois pontos de partida distintos em termos de concentrações de x_1 e x_2 , denotadas $x_{0,1}$ e $x_{0,2}$, respectivamente. Observe que o sistema converge para dois “pontos de operação” com características opostas: na Figura 13, a concentração final de x_1 é maior do que aquela de x_2 ; na situação representada pela Figura 14, no entanto, verifica-se o contrário. Comportamento parecido ocorre com a rede de regulação gênica envolvida no ciclo de lise-lisogênese de bactérias infectadas pelo bacteriófago lambda (BAGH; MANDAL; MCMILLEN, 2011). Nesse caso, os dois pontos de operação correspondem aos ciclos (ou estados) lítico e lisogênico (ou de profago), com diferentes concentrações das proteínas cI e cro .

Já a Figura 15 ilustra o chamado *diagrama de fases* do sistema de equações diferenciais modelado anteriormente, em que o tempo é suprimido para dar lugar a um gráfico bidimensional de x_1 versus x_2 .

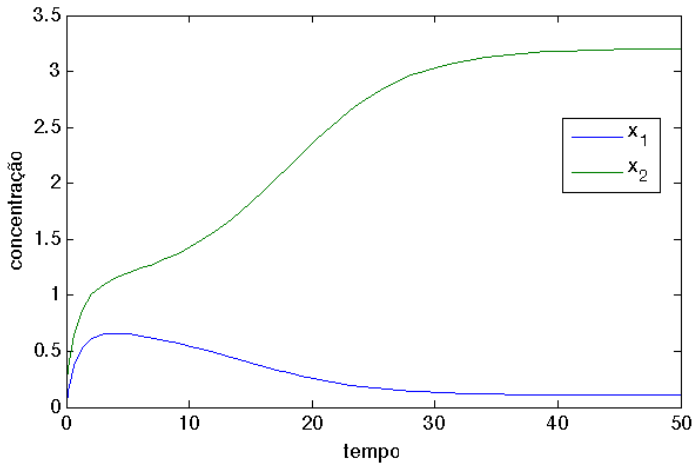


Figura 14 – Comportamento temporal das variáveis x_1 e x_2 quando $x_{0,1} = 0$ e $x_{0,2} = 0.2$.

Cada curva desse gráfico representa uma possível trajetória do sistema, uma sequência de estados por ele percorrida a partir de uma condição inicial. A trajetória representada na Figura 14, por exemplo, corresponde àquela marcada por um asterisco na Figura 15.

Observe que as trajetórias do sistema convergem para um dos dois pontos amplificados na Figura 15, denominados *atratores*, estados em que o sistema repousa de maneira estável. Um sistema dinâmico pode contar com um ou mais atratores, ou mesmo ser desprovido de algum. Quando em um atrator, ao sofrer pequenas perturbações, o sistema descola-se daquele, mas tende a reencontrá-lo depois de um certo lapso temporal. O conceito de atrator encontra ampla aplicabilidade em sistemas biológicos, podendo designar, por exemplo, cada um dos diversos destinos celulares (HUANG; INGBER, 2000; HUANG; ERNBERG; KAUFFMAN, 2009; KAUFFMAN, 1993; DAVIDICH; BORNHOLDT, 2008; MACLEAN; STUDHOLME, 2010), configurações biológicas robustas que dão o tom ordenado à vida como conhecemos.

Na Figura 15, a linha com espessura mais grossa separa o espaço de possíveis estados (x_1, x_2) em duas regiões: aquela que converge para o atrator superior, e a outra que desemboca no atrator inferior. Cada uma dessas regiões corresponde à chamada bacia de atração do atrator correspondente. Quanto mais robusto o estado associado a um atrator,

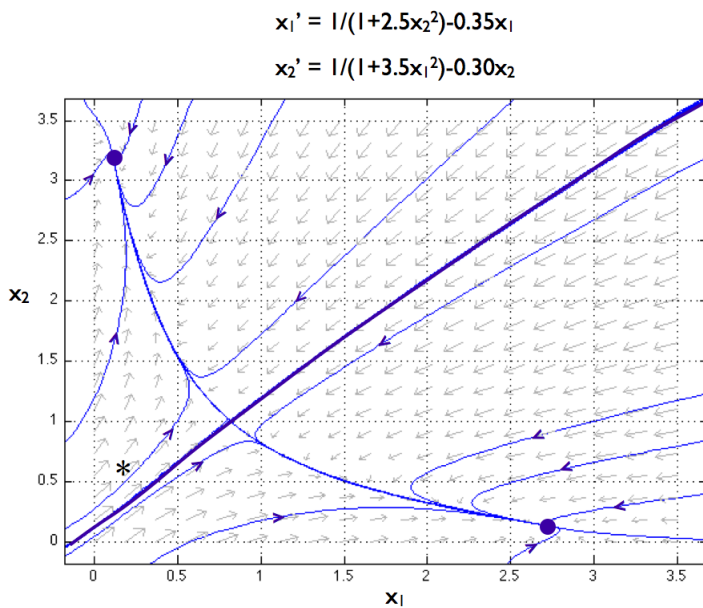


Figura 15 – Espaço de fases associado a um modelo contínuo para o grafo da Figura 12. As variáveis x'_1 e x'_2 denotam a taxa de variação no tempo de x_1 e x_2 , respectivamente.

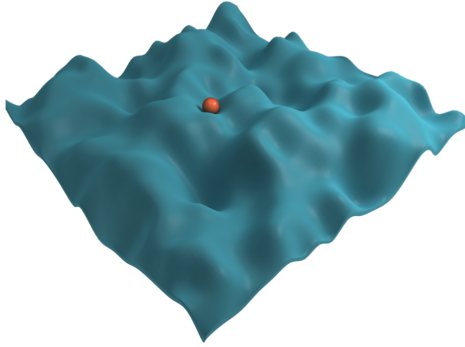


Figura 16 – Metáfora para os conceitos apresentados no texto, como atratores e bacias de atração.

maior a sua bacia de atração. De uma perspectiva matemática, e dependendo do propósito que se tem para o modelo, pode-se definir uma relação de equivalência, na qual os estados contidos em cada uma dessas bacias de atração pertencem a uma mesma classe de equivalência, ou seja, são equivalentes, pois conduzem o sistema a um mesmo estado final.

Uma representação pictórica do espaço de fases de um sistema, de suas trajetórias, dos conceitos de atratores e bacias de atração é o cenário em que uma bolinha salteia em um terreno acidentado, sujeitando-se à ação da gravidade e à irregularidade do terreno, como mostrado na Figura 16. A aplicação desse paralelo mecânico ao comportamento biológico resulta na icônica metáfora da paisagem epigenética de Waddington, bem conhecida da comunidade de biologia do desenvolvimento.

Note que, no modelo matemático discutido acima, adotamos o par (x_1, x_2) , as concentrações de cada um dos fatores de transcrição, como representante do estado do sistema, deixando ausentes da representação matemática as quantidades e localizações espaciais do mRNA, por exemplo. As dinâmicas das etapas de transcrição e tradução, assim, não foram explicitamente modeladas. Como resultado, uma mudança na concentração de x_1 , por exemplo, afeta instantaneamente a taxa de variação no tempo de x_2 . Importante apontar que outros fenômenos foram excluídos da representação matemática, como a difusão das proteínas no ambiente intracelular. Tais etapas intermediárias do gene à proteína podem ser incluídas no modelo, refinando-o, se necessário for

(JONG, 2002). A “complexificação” do modelo, entretanto, dificulta a *análise* de sua dinâmica (i.e. o estudo dos comportamentos possíveis gerados pelo sistema), como já alertado anteriormente.

Por fim, ressaltamos a necessidade de alimentar o modelo em equações diferenciais com parâmetros que não são facilmente mensuráveis, especialmente *in vivo*, como k_1 , k_2 , d_1 e d_2 . Variações nesses parâmetros podem alterar o próprio comportamento qualitativo do sistema, fazendo com que este passe a ter somente um atrator no lugar de dois, por exemplo (uma análise mais aprofundada sobre o impacto de variações paramétricas no comportamento em regime permanente do sistema, em outras palavras, no número de atratores, pode ser encontrada no livro de Strogatz (STROGATZ, 2001)).

3.2.2 Modelos Lógicos

Modelos lógicos incorporam um conjunto *discreto* (muitas vezes, finito) de estados, com a passagem de um estado a outro dando-se pela ocorrência de *eventos* ou pelo disparo de *transições*. Essas duas características os diferenciam dos modelos contínuos, como as equações diferenciais da subseção anterior, em que o espaço de estados é contínuo e a evolução do sistema é impulsionada pela passagem do tempo. As redes booleanas introduzidas por Kauffman e Thomas (KAUFFMAN, 1993; GLASS; KAUFFMAN, 1973; THOMAS, 1973) para modelagem de redes de regulação gênica são representantes da classe de modelos lógicos, que incluem, ainda, outros formalismos não discutidos aqui, como as próprias extensões das redes booleanas (BERNOT; COMET; RICHARD, 2004; JONG, 2002), além das redes de Petri, dos autômatos celulares, da álgebra de processos, dentre outros (FISHER; HENZINGER, 2007; EWALD et al., 2007; CHAUIYA; REMY, 2013).

Podem-se citar algumas motivações para emprego dos modelos lógicos como formalismo para modelagem, análise e controle do comportamento de redes biológicas, em particular das redes de regulação gênica:

- Tais modelos são mais simples de serem analisados do que aqueles baseados em equações diferenciais não-lineares, como as apresentadas na subseção anterior, especialmente para redes de regulação gênica com muitos nodos. Assim, uma vez modelado o sistema, é mais fácil estudar suas propriedades, como número de atratores por exemplo, sem depender de simulações extensivas do modelo matemático em questão. Ferramentas de *model-checking*

(CLARKE, 1999), amplamente difundidas na comunidade de ciências da computação, podem ser empregadas para tal fim;

- Modelos lógicos não dependem de parâmetros cinéticos, como aqueles requeridos para executar equações diferenciais, já que são representações formais *qualitativas* do sistema;
- As funções de Hill, introduzidas na subseção anterior para modelar o impacto de um gene na transcrição de outro, podem ser vistas, em um nível de abstração mais alto, como funções lógicas, em que um gene é ou ativado ou desativado dependendo da concentração da(s) proteína(s) de regulação gênica que afeta(m) sua transcrição (ver exemplo na Figura 17 para o caso de um gene cuja transcrição é afetada por uma proteína de regulação apenas);
- Diversos estudos têm demonstrado que, mesmo sendo simplificações profundas do comportamento real de redes biológicas, tais modelos podem explicar, reproduzir e prever uma grande gama de fenômenos associados às redes de regulação gênica (BORNHOLDT, 2005; DAVIDICH; BORNHOLDT, 2008; ALBERT, 2004; SHMULEVICH; DOUGHERTY; ZHANG, 2002; MACLEAN; STUDHOLME, 2010);
- A transição de estados em sistemas biológicos é, por vezes, dominada por saltos abruptos decorrentes de eventos, como no caso da ativação alostérica de proteínas por determinadas moléculas, (veja-se o caso da ativação do repressor do triptofano (ALBERTS et al., 2008)). Essas situações são mais bem captadas por modelos lógicos;
- A quantidade de cada um dos fatores de transcrição no meio intracelular em um dado momento é relativamente pequena (da ordem de dezenas ou centenas de moléculas), o que não se coaduna com o pressuposto básico empregado para modelar redes de regulação gênica por meio de equações diferenciais ordinárias, o de que o número de moléculas de cada espécie química é suficientemente alto.

Consideremos as redes booleanas como mote para introduzir os modelos lógicos e trazer à tona suas diferenças em relação aos modelos contínuos. Nesse caso específico, o estado de um gene qualquer da rede assume dois valores somente, 0 ou 1, correspondentes às situações de gene não expresso (ou expresso a uma taxa basal) e gene expresso, respectivamente. O estado global da rede pode ser dado, então, pelo

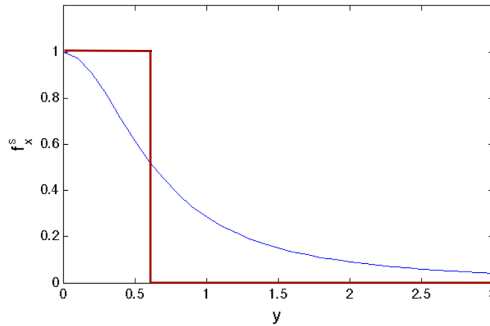


Figura 17 – Sobreposição de uma função contínua a modelar a repressão de uma proteína de regulação gênica sobre a transcrição de um dado gene (em azul) e uma resposta do tipo “tudo-ou-nada” (em vermelho), captando o mesmo fenômeno da repressão de uma proteína sobre a taxa de transcrição de um gene.

estado individual de cada um dos seus genes. Assim, a rede de regulação gênica mutuamente repressora associada ao grafo da Figura 12 poderia encontrar-se em um dos quatro seguintes estados: $(0, 0)$, $(0, 1)$, $(1, 0)$ e $(1, 1)$, em que o primeiro elemento do par corresponde ao estado de v_1 , enquanto o segundo, ao de v_2 . Para o caso geral, escreve-se formalmente que $v_i \in \{0, 1\}$, com v_i denotando aqui a variável que contém o estado do gene nomeado como $v_i \in V$, e $\mathbf{v} = (v_1 \cdots v_N)$ para o vetor que representa o estado global da rede com N genes.

As possíveis transições entre os estados de uma rede booleana podem ser determinadas de duas maneiras distintas. Na primeira, codifica-se o comportamento biológico *diretamente* em termos de *funções de transição* de estados booleanas, como será visto adiante, forma que denominaremos discretização *a priori*. Essas funções são definidas a partir de experimentos laboratoriais de larga-escala (SHMULEVICH; DOUGHERTY; ZHANG, 2002) ou de relatos da literatura que especificam como o estado de um gene varia em função da concentração das proteínas de regulação gênica (DAVIDICH; BORNHOLDT, 2008).

Tome novamente o caso da rede de regulação gênica mutuamente repressora já analisada. Nesse exemplo, poderíamos escrever as seguintes funções de atualização para v_1 e v_2 , respectivamente: $f_1(v_1, v_2) = \bar{v}_2$ e $f_2(v_1, v_2) = \bar{v}_1$. Tais funções são os equivalentes matemáticos do fato

(suponha-se) observado para a rede “um gene ativa-se (mantém-se ativado) sempre que, e somente quando, seu repressor estiver desativado”. A notação \bar{v}_i em f_i indica a negação lógica da variável v_i . Assim, se $v_i = 0$, $\bar{v}_i = 1$; quando $v_i = 1$, $\bar{v}_i = 0$.

Ao se definirem as funções f_1 e f_2 , elas podem ser empregadas para atualizar:

- a) os estados de *todos* os genes de uma só vez (sincronamente, portanto);
- b) o estado de somente um gene por vez (assincronamente); ou
- c) os estados de um subconjunto de V sincronamente e os estados dos genes restantes de maneira assíncrona. Ao leitor interessado em um estudo mais aprofundado sobre o impacto das distintas formas de atualização nas propriedades dinâmicas de uma rede, recomenda-se o trabalho de Saadatpour e coautores (SAADATPOUR; ALBERT; ALBERT, 2010).

Faremos uso exclusivamente da atualização assíncrona nesta Tese, por acreditarmos ser pouco plausível a existência de um *clock* a ordenar a transição de estados dos genes da rede. A Figura 18 ilustra o chamado *diagrama de transição de estados* para a rede mutuamente repressora de nosso exemplo, com as funções de atualização f_1 e f_2 anteriormente definidas e aplicadas para atualização assíncrona do estado da rede. Esse diagrama é o equivalente discreto do espaço de fases da Figura 15. Ele contém todos os estados do sistema e as possíveis transições de um estado a outro. Por exemplo, no estado 00, $v_1 = v_2 = 0$, o que implica $f_1 = f_2 = 1$. Portanto, o sistema pode se deslocar para 01 ou 10 a partir do estado 00.

A segunda forma de obter as transições entre os estados passa por uma etapa de *abstração* de uma dinâmica contínua, ou estocástica, método que chamaremos de discretização *a posteriori*. Nesse caso, *particiona-se* o espaço de estados do sistema contínuo, ou estocástico, utilizando as possíveis trajetórias deste para definir as transições do diagrama de transição de estados do sistema discreto, como esquematizado na Figura 19 para o caso específico do exemplo aqui tratado. No Capítulo 7, voltaremos a esse tópico, com o objetivo de argumentar em favor da conjectura de que as ideias centrais nesta Tese desenvolvidas podem ser também aplicadas para controlar o comportamento de redes de regulação gênica modeladas por equações diferenciais ordinárias.

Naturalmente, as abstrações booleanas são aproximações que não emulam finamente o comportamento de um sistema dinâmico. Não

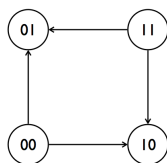


Figura 18 – Diagrama de transição de estados para a rede de regulação gênica associada ao grafo da Figura 12, em que os estados rastreiam os estados individuais de cada gene.

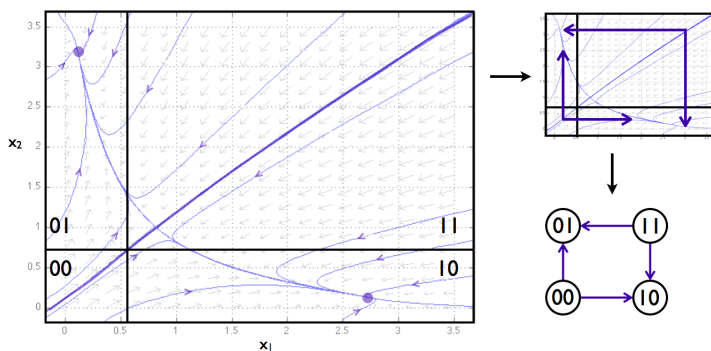


Figura 19 – O espaço de estados do sistema contínuo é particionado (elementos ‘00’, ‘01’, ‘10’ e ‘11’), suas trajetórias são empregadas para determinar que transições entre as partições definidas são possíveis, para então, finalmente, obter uma representação discretizada do sistema contínuo. Note que os elementos ‘01’ e ‘10’, que contêm os atratores do sistema contínuo, são também atratores do sistema discreto (no sentido de que nenhuma transição parte delas).

reside, entretanto, no detalhismo a maior virtude desse tipo de representação, mas sim em sua economicidade, ao tentar captar o essencial de um fenômeno complexo, filtrando detalhes que, para a aplicação específica, são desnecessários.

O diagrama de transição de estados associado a uma rede biológica pode ser utilizado para se obter uma representação em *autômato de estados finitos* dessa rede. Um autômato G é um objeto matemático caracterizado por cinco elementos (CASSANDRAS; LAFORTUNE, 2008):

- 1) Um conjunto finito Q dos *estados* do sistema;
- 2) Um conjunto finito de *eventos* Σ que podem ocorrer no sistema;
- 3) Uma *função de transição* δ , em que $r = \delta(q, e)$ denota que, ao ocorrer o evento $e \in \Sigma$ no estado $q \in Q$, o sistema executa uma transição ao estado $r \in Q$;
- 4) Um *estado inicial* q_0 , dependente da aplicação; e, por fim,
- 5) Um conjunto de *estados marcados* Q_m , que representam os estados associados a uma tarefa completa. No capítulo 5 ficará mais clara a semântica de Q_m no contexto de controle de sistemas biológicos.

Por exemplo, poderíamos definir, para o diagrama de transição de estados da Figura 18 o seguinte autômato $G = (Q, \Sigma, \delta, q_0, Q_m)$:

- 1) $Q = \{00, 01, 10, 11\}$;
- 2) $\Sigma = \{+v_1, +v_2, -v_1, -v_2\}$, em que $+v_i$ e $-v_i$ denotam a ativação e desativação dos genes v_1 e v_2 , respectivamente;
- 3) $\delta(00, +v_1) = 10$, $\delta(00, +v_2) = 01$, $\delta(11, -v_1) = 01$ e $\delta(11, -v_2) = 10$;
- 4) $q_0 = 00$, uma escolha arbitrária do estado de partida do sistema (não definida pelo diagrama de transição de estados); e
- 5) $Q_m = Q$, também uma escolha arbitrária e não determinada pelo diagrama de transição de estados.

Autômatos de estados finitos serão empregados como o formalismo base para o desenvolvimento de nossas ideias no Capítulo 5.

3.2.3 Modelos Estocásticos

A dinâmica estocástica de uma rede biológica manifesta-se de maneira mais pronunciada em redes de regulação gênica, dado que o número típico de moléculas envolvidas nessas redes é relativamente pequeno e, assim, o impacto resultante de flutuações termodinâmicas é mais perceptível (KARLEBACH; SHAMIR, 2008; JONG, 2002).

Logo, com o objetivo de prover uma representação mais detalhada do comportamento individual de uma rede de regulação (em contraste com modelos baseados em equações diferenciais, por exemplo, que privilegiam uma visão do comportamento médio de uma dada população de células), faz-se uso de modelos estocásticos, em que os processos de transcrição e tradução são descritos de maneira mais pormenorizada, oferecendo, assim, oportunidade de incluir flutuações estocásticas nas interações entre os diferentes atores da expressão gênica. Nos modelos estocásticos, o estado do sistema $X(t)$ é dado pelo número de moléculas de cada espécie química no instante de tempo t , e a cada uma das possíveis reações entre essas espécies atribui-se uma *probabilidade* de ocorrência.

Um dos algoritmos mais conhecidos para implementação de modelos estocásticos é o Algoritmo de Simulação Estocástica (SSA, *Stochastic Simulation Algorithm*), desenvolvido por Gillespie. Nesse procedimento, determina-se, dado o estado atual do sistema $X(t)$, *quando* a próxima reação ocorrerá, e *qual* será ela (JONG, 2002). Uma vez determinada essa informação, atualiza-se o estado atual $X(t)$ de forma a refletir a ocorrência da reação obtida, prosseguindo-se assim sucessivamente.

Simulações estocásticas são computacionalmente custosas e diversas abordagens publicadas na literatura ensaiam diferentes maneiras de tornar suas execuções mais eficientes, possibilitando, dessa forma, a aplicação da abordagem a sistemas maiores (KARLEBACH; SHAMIR, 2008; JONG, 2002; SAMAD et al., 2005). Um exemplo do emprego de técnicas estocásticas para estudar a dinâmica de redes de regulação pode ser encontrado no trabalho de Arkin e colaboradores, sobre o comportamento de células de *E. coli* infectadas por bacteriófagos (ARKIN; ROSS; MCADAMS, 1998).

4 CONTROLE DE REDES DE REGULAÇÃO GÊNICA

Neste Capítulo, apresentaremos algumas das principais abordagens existentes na literatura para controle realimentado de redes biológicas intracelulares. Discorreremos somente sobre aqueles casos em que o sistema a ser controlado é uma rede de regulação gênica, ou em que o controlador é realizado por meio de uma rede de regulação gênica sintética.

O controle realimentado de um sistema baseia-se na geração de uma entrada u , dependente da saída y , capaz de alterar o comportamento do sistema segundo critérios pré-estabelecidos.

A Seção 4.1 revisa trabalhos puramente teóricos, em que a validação de u é feita somente por simulações computacionais. A Seção 4.2, por outro lado, aborda as pesquisas de cunho experimental, em que o sinal de controle u projetado é testado *in vivo*.

4.1 ABORDAGENS TEÓRICAS

As pesquisas descritas nesta Seção, conforme já comentado na Introdução desta Tese, carecem de informações detalhadas quanto à *realização* das ações de controle u . Embora tal fato não invalide os resultados, ele dificulta discussões mais aprofundadas sobre abrangência e restrições das ideias apresentadas.

4.1.1 Controle Ótimo e Redes Booleanas Probabilísticas

Começamos pelo trabalho de Datta e colaboradores, que trataram o controle de redes de regulação gênica modeladas como redes Booleanas probabilísticas, RBPs, objetivando conduzi-las de um estado inicial a um estado final em um dado intervalo de tempo e com o menor custo possível (DATTA et al., 2007). Aqui, o estado do sistema congrega o estado individual, e Booleano, de cada um dos N genes da rede. Na concepção dos autores, uma RBP consiste em uma *coleção* de redes Booleanas que compartilham os mesmos nodos. Cada rede possui uma função de atualização de estados *fixa*, que retorna um *único* estado futuro a partir do estado atual. Nesse formalismo, a simulação das trajetórias do sistema começa com a escolha, no estado inicial, de uma dentre aquelas redes Booleanas possíveis (cada rede pode ser escolhida

com uma probabilidade dada *a priori*). Depois, atualiza-se o estado com uma probabilidade p (ou seja, o sistema permanece no mesmo estado com uma probabilidade $1 - p$), segundo a função de atualização da rede Booleana. No estado seguinte, troca-se, com uma probabilidade q , a rede Booleana que rege a dinâmica do sistema, atualiza-se o estado com uma probabilidade p e, assim, prossegue-se por um número de passos finito.

Para estudar analiticamente o comportamento dinâmico de tais redes, definiu-se um vetor de distribuição de probabilidades $w(k)$ de tamanho 2^N , o número máximo de estados para uma rede Booleana com N nodos, em que cada elemento $w_i(k)$ continha a probabilidade de o sistema, no tempo k , encontrar-se no estado indexado por i . A indexação de estados utilizada é trivial: $i = 1$ corresponde ao estado da rede igual a $0 \cdots 0$, $i = 2$ denota o estado $0 \cdots 01$ e assim sucessivamente até o índice $i = 2^N$, que indexa o estado $1 \cdots 1$. A equação que rege o comportamento dinâmico de tais redes pode, então, ser dada por

$$w(k+1) = w(k)A \quad (4.1)$$

em que A é uma matriz de probabilidades com dimensão $2^N \times 2^N$, cujo elemento (i, j) indica a probabilidade de o sistema alcançar no tempo $k+1$ o estado indexado por j , caso $w_i(k) = 1$.

A dinâmica dada pela equação (4.1) não inclui as possibilidades de intervenção, ou seja, ela representa o comportamento do sistema em malha aberta. Os autores assumem, pois, em seguida, a existência de m entradas Booleanas u_1, \dots, u_m que têm a propriedade de *alterar* os elementos da matriz A . Essas entradas podem ser nodos da própria rede nativa, sujeitos a interferências externas, ou nodos adicionados à rede para fins de controle. As 2^m ações de controle possíveis a cada instante de tempo k podem ser condensadas em uma variável inteira $v(k)$ que assume valores pertencentes ao conjunto $\{1, \dots, 2^m\}$, em que $v(k) = 1$ indica $u_1 = \dots = u_m = 0$, $v(k) = 2$ denota $u_1 = \dots = u_{m-1} = 0$ e $u_m = 1$, e assim sucessivamente até chegar-se a $v(k) = 2^m$, representando a situação em que $u_1 = \dots = u_m = 1$. O comportamento do sistema, dotado das entradas de controle definidas, é reescrito, então, como

$$w(k+1) = w(k)A(v(k)). \quad (4.2)$$

A representação contida na equação (4.2) é equivalente à seguinte formulação (DATTA et al., 2007)

$$z(k+1) = f(z(k), v(k), d(k)), \quad (4.3)$$

em que $z(k)$ corresponde a um dos 2^N estados da rede e f é uma função de transição de estados, cujo valor depende também da perturbação $d(k)$, que incorpora os efeitos resultantes da troca de redes com uma probabilidade q e da atualização de estados com uma probabilidade p .

Em seguida, os autores atribuem ao emprego da ação de controle $v(k)$ no estado $z(k)$ um custo $C_k(z(k), v(k))$. Por fim, o problema de controle é formulado como: dado um estado inicial $z(0)$, ache uma lei de controle $\pi = \{v(0), \dots, v(M-1)\}$ que minimiza o funcional

$$J_0(z(0)) = E \left[\sum_{k=0}^{M-1} C_k(z(k), v(k)) + C_M(z(M)) \right],$$

em que $E(\cdot)$ denota o valor estimado de uma variável aleatória. O funcional acima está sujeito à seguinte restrição: a probabilidade de $z(k+1) = j$, dado que $z(k) = i$ e $v(k) = v$, é igual a $a_{ij}(v)$, em que $a_{ij}(v)$ é o elemento na i -ésima linha e j -ésima coluna de $A(v)$.

4.1.2 *Model-checking* e Redes Booleanas

Em uma outra abordagem, Langmead e Jha trataram o problema de conduzir uma rede de regulação gênica Booleana de um estado inicial a um estado final como um problema de *model-checking* (LANGMEAD; JHA, 2008). *Model-checking* compreende, basicamente, técnicas formais que têm por objetivo determinar se um dado modelo \mathcal{M} satisfaz a uma propriedade lógica ϕ (CLARKE, 1999). Tais técnicas são comumente aplicadas para *análise* do comportamento de sistemas técnicos complexos. No contexto biológico, poder-se-iam utilizar ferramentas de *model-checking* para, por exemplo, determinar se é possível que uma rede de regulação gênica atinja um dado estado a partir de um estado inicial qualquer.

Para resolver o problema de controle da dinâmica de uma rede Booleana (V, F) , em que V corresponde ao conjunto de genes nativos e F contém as funções de atualização de cada nodo, os autores supuseram a possibilidade de inserir um conjunto de nodos externos $C = \{c_i\}_i$, também Booleanos, capazes de interagir com os genes nativos de uma maneira pré-determinada. Como exemplo, poderíamos supor, para controlar o comportamento da rede de regulação gênica mutuamente repressora mostrada na Figura 10, a existência de um nodo externo c_1 capaz de reprimir o nodo v_1 e, ao mesmo tempo, induzir a expressão do nodo v_2 , conforme mostrado na Figura 20.

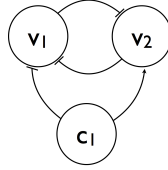


Figura 20 – Exemplo da estratégia apresentada por Langmead e Jha. Supõe-se a possibilidade de inserir genes sintéticos $c_i \in C$ que tenham um efeito pré-determinado sobre a transcrição dos genes nativos $v_j \in V$.

A partir de um modelo formal para essa rede aumentada, contendo genes nativos e sintéticos, os autores empregaram técnicas de *model-checking* para verificar se o modelo satisfazia a uma propriedade lógica ϕ que, traduzida para linguagem natural, tinha a seguinte semântica: “existe uma trajetória a partir do estado inicial tal que, em um tempo não maior que t^* , a rede Booleana atinge o estado final desejado”. Caso a propriedade seja satisfeita, os algoritmos de *model-checking* retornam os valores que os nodos externos c_i assumem ao longo da trajetória que satisfaz ϕ . Esses valores correspondem, então, às entradas de controle. A abordagem foi aplicada a um modelo Booleano de uma rede de regulação gênica que coordena o desenvolvimento embrionário de moscas *Drosophila*.

4.1.3 Controle de Sistemas Diferenciais Afins por Partes

Chavez e Gouzé também abordaram o problema de controle de redes de regulação gênica, no entanto, fizeram-no para redes modeladas por equações diferenciais em que a taxa de síntese das proteínas era dada por uma função do tipo degrau, uma simplificação lógica da dinâmica de Hill, conforme exemplificado na Figura 17 do capítulo anterior (CHAVES; GOUZÉ, 2011).

No referido trabalho, as ideias foram desenvolvidas para uma rede de regulação gênica composta por dois nodos mutuamente repressores. O modelo adotado segue abaixo:

$$\frac{d(x_1)}{dt} = u \cdot k_1 s^-(x_2, \theta_2) - \gamma_1 x_1 \quad (4.4)$$

$$\frac{d(x_2)}{dt} = u \cdot k_2 s^-(x_1, \theta_1) - \gamma_2 x_2 \quad (4.5)$$

As variáveis contínuas x_1 e x_2 representam as concentrações dos fatores de transcrição associados a cada um dos genes. A variável u denota a entrada de controle $u = u(t, X)$, que pode assumir três valores qualitativos pertencentes ao conjunto $\{u_{min}, 1, u_{max}\}$. O controle u depende do tempo t e da partição $X \in \{00, 01, 10, 11\}$ do espaço de estados em que o sistema se encontra. A taxa de síntese de cada um dos fatores de transcrição é dada pela função $s^-(r, \theta)$, definida como:

$$s^-(r, \theta) = \begin{cases} 1, & r < \theta \\ 0, & r > \theta \end{cases}$$

Segundo os autores, a implementação biológica da entrada de controle u , que modula a magnitude da função $s^-(r, \theta)$, baseia-se na possibilidade de projetar genes sintéticos cujas transcrições são reguladas *externamente* por pequenos metabólitos.

O problema de controle foi, finalmente, formalizado como: encontre uma entrada $u(t, X)$ tal que o sistema modelado pelas equações (4.4) e (4.5) convirja globalmente para um dos três pontos de equilíbrio (dois estáveis e um instável) do sistema (assumindo-se que os parâmetros do modelo são tais que o sistema em malha aberta possua três pontos de equilíbrio).

4.2 ABORDAGENS PRÁTICAS

Dentre as abordagens práticas que objetivam implementar uma malha fechada de controle para alterar o comportamento de redes de regulação gênica, identificam-se duas subcategorias, como já mencionado anteriormente: aquela em que equipamentos externos são alocados às tarefas de *sensoreamento*, *computação do sinal de controle* e *atuação*; e uma outra, que emprega somente moléculas sintéticas para desempenhar as referidas funções.

Faz-se ressaltar que, embora existam diversos trabalhos pertencentes à segunda das subcategorias citadas, não identificamos nenhum artigo dentre eles em que a entrada de controle u é obtida por meio de técnicas da Teoria de Controle. Como se depreenderá posteriormente, os problemas tratados por esses artigos são simples do ponto de vista matemático e, portanto, podem ser resolvidos sem o uso de métodos advindos da Teoria de Controle. À medida que o comportamento dinâmico do sistema a ser controlado torna-se mais complexo, no entanto, fica mais difícil a tarefa de obter as entradas u somente por métodos *ad hoc*.

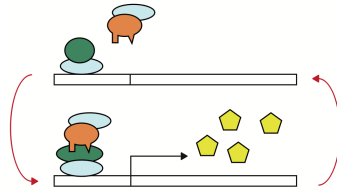


Figura 21 – Módulo composto por moléculas sintéticas inseridas em células de fungo. O módulo induz a produção de uma proteína fluorescente (representada pelo pentágono amarelo), quando excitado por luz na faixa de frequência do vermelho.

4.2.1 Controle *In silico*

O trabalho de Miliás-Argeitis e colaboradores é um dos representantes daquele grupo de pesquisas em que equipamentos técnicos externos são utilizados para implementar as funções típicas de uma malha de controle (MILIAS-ARGEITIS; SUMMERS; LYGEROS, 2011). Nesse trabalho, os pesquisadores inseriram um módulo molecular em uma população de fungos da espécie *S. cerevisiae*. O módulo é constituído por genes e proteínas de regulação gênica sintéticos que, quando exposto a luzes na faixa do vermelho, sintetiza uma proteína fluorescente, conforme mostrado na Figura 21.

Para controlar a fluorescência emitida pelo sistema em questão, a população de fungos dotada do referido módulo, os autores modelaram a rede sintética como um sistema chaveado de equações diferenciais ordinárias de quarta ordem. Os parâmetros livres, em um total de cinco, foram obtidos experimentalmente, após uma série de excitações do sistema em malha aberta com diferentes pulsos de onda eletromagnética na faixa de frequência do vermelho e do vermelho-extremo. Na operação em malha fechada, a fluorescência emitida pelo sistema era medida por equipamentos de microscopia, a uma taxa de amostragem de 30 min, e enviada para um módulo computacional dotado de um filtro de Kalman, capaz de estimar aquelas variáveis de estado não medidas, a partir dos dados de medição. Com base nas informações medidas e estimadas, o controlador, derivado segundo técnicas de controle preditivo baseado em modelos, determinava que pulsos luminosos deveriam ser aplicados ao sistema, de forma a diminuir o erro entre a fluorescência

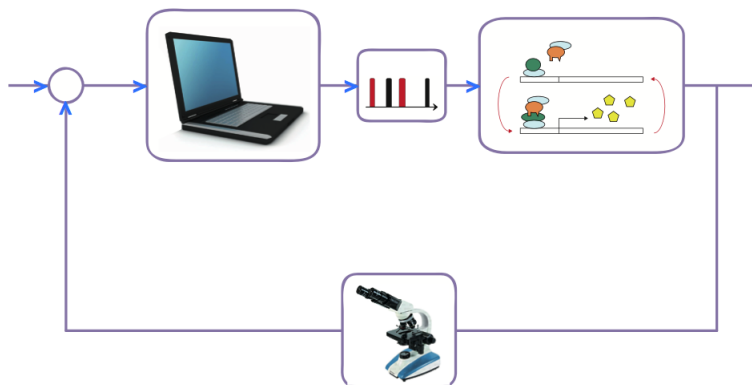


Figura 22 – Visão esquemática do sistema de controle *in silico* utilizado para controlar a luminescência emitida por uma população de fungos portadores de um circuito sintético.

medida e a referência desejada. O sistema de controle completo pode ser visto na Figura 22.

A mesma estratégia, com formalismos matemáticos semelhantes, foi empregada nos últimos anos por outros grupos de pesquisa, visando, igualmente, controlar a expressão de um circuito gênico inserido em população de fungos (UHLENDORF; MIERMONT; DELAVEAU, 2012; MENOLASCINA; BERNARDO; BERNARDO, 2011).

4.2.2 Controle com Circuitos Biológicos Sintéticos

Recentemente, diversos estudos reportaram a construção de moléculas sintéticas capazes de medir estados celulares específicos e disparar ações correspondentes. Na maioria desses estudos, no entanto, os circuitos biomoleculares projetados detectavam *somente um* estado x , produzindo, como resposta, uma entrada de controle u específica. Assim, do ponto de vista da Teoria de Controle, a lei de realimentação de estados $u = g(x)$ foi definida para um único valor de x , restringindo, pois, as possibilidades de comportamento em malha fechada que podem ser obtidas. Entretanto, tais circuitos formam os blocos construtores para implementação de leis de controle mais gerais, em que o domínio de validade da função $g(x)$ é estendido a uma gama maior de estados.

Um exemplo dentre os trabalhos mencionados é aquele de Bagh e colaboradores, em que se demonstrou a construção de genes sintéticos aptos a alterar o destino celular de bactérias infectadas pelo vírus bacteriófago lambda (BAGH; MANDAL; MCMILLEN, 2011). Esse vírus, ao invadir bactérias da espécie *E. coli*, insere seu material genético no genoma hospedeiro e permanece latente, crescendo e se multiplicando juntamente com a bactéria, em um ciclo denominado lisogênese. Na presença de agentes que danificam o DNA bacteriano, como indutores químicos ou luz ultravioleta, a bactéria *chaveia* para um outro estado celular estável, denominado lítico (ou de lise celular), em que o vírus destrói a parede celular bacteriana, liberando aproximadamente 100 novos vírus. O ciclo de lise-lisogênese é coordenado por uma rede de regulação gênica com dois genes mutuamente repressores (denominados *cI* e *cro*, que codificam para proteínas denominadas CI e Cro, respectivamente), cujo espaço de estados pode possuir dois atratores. Nas espécies de bactérias infectadas pelo vírus, cada um desses atratores corresponde a um dos padrões fisiológicos estáveis observáveis: lise (*cI* ativado e *cro* inibido) ou lisogênese (*cI* inibido e *cro* ativado).

Os pesquisadores projetaram diferentes genes sintéticos que, em contato com a rede de regulação gênica formada por *cI* e *cro*, eram capazes de evitar a transição do sistema para o atrator lítico após excitação externa por luz ultravioleta, o comportamento observado no sistema em malha aberta. Os genes sintéticos projetados codificavam a mesma proteína CI já produzida normalmente pela bactéria e foram dotados de uma região de regulação tal que se mantinham reprimidos durante o ciclo lisogênico e ativados quando a célula bacteriana estava na iminência de entrar no ciclo lítico, após a perturbação externa. Em outras palavras, os genes sintéticos detectavam o estado celular associado a baixos níveis da proteína CI, o que poderia levar a bactéria a uma transição para atrator indesejado, e respondiam a essa condição por meio da síntese do fator de transcrição CI, evitando o *chaveamento* gênico.

Um outro trabalho interessante foi publicado por Culler e colaboradores, em que o destino celular foi reprogramado por meio de RNAs sintéticos em vez de genes sintéticos (CULLER; HOFF; SMOLKE, 2010). Nesse estudo, um RNA codificando para uma proteína indutora da morte celular foi projetado de tal maneira que sua ligação com proteínas específicas (representantes de um estado associado a doenças) alterasse sua taxa de tradução, funcionando como um *interruptor molecular*: na presença das proteínas, o RNA era traduzido a uma taxa alta, enquanto, na ausência dessas, a taxa era próxima de zero.

Nesse caso, portanto, a tradução da proteína codificada pelo RNA pode ser vista como o *efeito* associado à detecção de um estado, aquele correspondente à presença das proteínas que se acoplam ao RNA, afetando sua tradução. Note que tal dispositivo implementa no nível molecular uma operação lógica semelhante àquela codificada pelo gene sintético do estudo anterior (do tipo “se <condição satisfeita> então <ação>”), com a diferença de que a “programação” dessa operação, no presente caso, faz-se com um “*hardware*” ligeiramente diferente, baseado em RNA e não em DNA. Assim, no caso do gene sintético, é sua *região de regulação* que reconhece um estado celular, alterando a taxa de *transcrição* do gene como resultado. Já no estudo baseado em RNAs sintéticos, são estruturas denominadas *aptâmeros*, trechos do RNA capazes de se ligar a proteínas, que detectam um estado, acarretando mudanças na *tradução* do RNA. As diferenças no *hardware* implicam características distintas para cada dispositivo. RNAs sintéticos, por exemplo, são mais flexíveis do que DNAs sintéticos, por se ligarem com outras proteínas que não exclusivamente fatores de transcrição, além de o fazerem também com outros metabólitos, aumentando, dessa forma, a gama de estados celulares detectáveis pelo dispositivo de medição-atuação (vide o estudo de Xie e coautores, em que um RNA sintético foi projetado para reconhecer um conjunto de RNAs específicos associados a câncer, disparando, como resposta, a síntese de uma proteína indutora da morte celular (XIE et al., 2011)).

Galloway e colegas construíram recentemente circuitos sintéticos em que tanto a região de regulação da transcrição de genes quanto a região de regulação da tradução de RNAs foram manipuladas, de modo a induzir diferentes destinos celulares em células do fungo *S. cerevisiae* (GALLOWAY; FRANCO; SMOLKE, 2013). Nesse caso, o sinal de controle *u* gerado pelo circuito sintético correspondia à concentração de uma proteína endógena, i.e. naturalmente produzida pela célula a ser controlada, pertencente à via de sinalização MAPK (do inglês *mitogen-activated protein kinase*), uma via bioquímica que desempenha papel importante em processos celulares tais como diferenciação, divisão e morte celular.

Em uma abordagem distinta das anteriores, Nissim e coautores construíram um módulo que, ao identificar um certo estado em células de fungos (correspondente a altas atividades de duas vias metabólicas independentes, uma de síntese do aminoácido metionina e outra de degradação do açúcar galactose), induzia a transcrição de um gene inibidor do crescimento celular (NISSIM; BEATUS; BAR-ZIV, 2007; NISSIM; BAR-ZIV, 2010). Nesse estudo, o funcionamento do circuito baseou-se

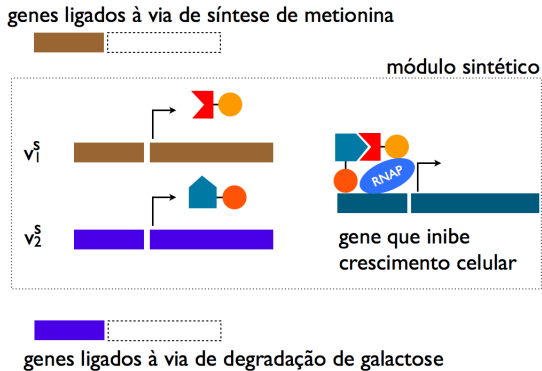


Figura 23 – As regiões de regulação de v_1^s e v_2^s são iguais àquelas dos genes ligados à síntese de metionina e à degradação de galactose, respectivamente. Quando ambas as vias metabólicas estão ativas, as proteínas sintéticas codificadas por v_1^s e v_2^s ligam-se, facilitando o acoplamento da RNA polimerase a um terceiro gene, que induz a morte celular. Figura adaptada do artigo de Nissim e coautores (NISSIM; BEATUS; BAR-ZIV, 2007).

no emprego de dois genes sintéticos v_1^s e v_2^s que, diferentemente das outras abordagens, codificavam para fatores de transcrição *não produzidos naturalmente* pelas células, ou fatores de transcrição exógenos (ver Figura 23). As regiões de regulação desses genes sintéticos foram projetadas de tal forma que eles fossem transcritos sob as mesmas condições em que os genes envolvidos na via de síntese de metionina e de degradação de galactose o eram, como pode ser visto na Figura 23. Dessa forma, na situação em que ambas as vias metabólicas estão ativas, os dois genes sintéticos também serão expressos, fazendo com que seus fatores de transcrição, conjuntamente, induzam a transcrição de um terceiro gene sintético, cuja proteína inibe o crescimento celular.

No estudo publicado por Kemmer e coautores, os genes sintéticos empregados para fins de controle também codificavam para proteínas exógenas, a exemplo do caso anterior, mas com o intuito de manter em níveis adequados a concentração de urato na corrente sanguínea de ratos (KEMMER et al., 2010). Urato é um produto final do metabolismo de compostos denominados purinas e, em níveis elevados, pode levar, por exemplo, à formação de cristais de ácido úrico nos rins, dentre outras patologias. Nesse experimento, a unidade central era composta por

um gene derivado da bactéria *Deinococcus radiodurans*, que se mantém inativo na ausência de ácido úrico e ativo na presença desse metabólito. Esse último trabalho permite vislumbrar o potencial terapêutico advindo do estudo combinado de Biologia Sistêmica, Biologia Sintética e Teoria de Controle.

Como se depreende da leitura desta Seção, a instrução de alto nível “se <condição> então <ação>”, ou $u^* = g(x^*)$ para x^* e u^* representando a condição e a ação, respectivamente, pode ser traduzida para a “linguagem de máquina” molecular, ou “compilada” para tal linguagem, de diferentes maneiras. Esse fato será explorado em nosso trabalho na etapa de realização do supervisor, concebido aqui como uma função matemática que especifica, para os estados alcançáveis de um dado sistema, um conjunto de ações a serem implementadas, de maneira que as especificações de malha fechada sejam satisfeitas.

5 CONTROLE SUPERVISÓRIO APLICADO À INTERVENÇÃO EM REDES DE REGULAÇÃO GÊNICA

5.1 VISÃO GERAL

Neste Capítulo, discorreremos acerca das ideias centrais da Tese, em que se elabora uma proposta, fundamentada na Teoria de Controle Supervisório, para resolver o seguinte problema: dada uma rede de regulação gênica nativa, especifique um conjunto de genes sintéticos que, ao interagirem com a rede nativa, alteram de maneira pré-especificada seu comportamento *dinâmico e estático*.

Nosso procedimento para resolução desse problema recebe como entradas:

- a) um modelo em autômato de estados finitos da rede de regulação gênica a ser controlada;
- b) um subconjunto dos fatores de transcrição dessa rede, aqueles que serão codificados pelos genes sintéticos para alterar o comportamento da rede nativa. Só empregaremos, assim, proteínas *endógenas* para fins de controle, evitando interações não previstas que podem advir da introdução de novas proteínas na célula;
- c) uma especificação de controle que, basicamente, consiste em um conjunto de estados-alvo a serem atingidos pelo sistema e em restrições adicionais sobre os estados permitidos no trajeto que vai do estado inicial até os estados finais.

Como saída, o procedimento retorna:

- a) critérios de projeto para as regiões de regulação de cada um dos genes sintéticos. Esses critérios são expressos como funções de atualização de estados Booleanas, projetadas de tal modo que as especificações de controle em (c) sejam atendidas.

Para obtermos os critérios de projeto dos genes sintéticos – a saída de nosso procedimento – formulamos o problema de controle de redes de regulação gênica como um problema de *Controle Supervisório* (Seção 5.2).

A Teoria de Controle Supervisório (TCS) aborda o problema de controle realimentado de Sistemas a Eventos Discretos (CASSANDRAS;

LAFORTUNE, 2008; RAMADGE; WONHAM, 1987), sistemas dinâmicos cujos *estados*, pertencentes a um conjunto discreto, são alterados pela ocorrência de *eventos*, como a chegada de uma peça para processamento em uma máquina, a falha em um robô, ou, no cenário biológico, a fosforilação de uma proteína ou a ativação de um gene, por exemplo. Originalmente, a TCS foi concebida para alterar a *linguagem gerada* por um dado sistema a ser controlado, também chamado de *planta*. Entende-se por linguagem gerada o conjunto de todas as possíveis sequências de eventos exibidas pela planta. A intervenção no comportamento do sistema sob controle visa garantir que ele atenda a um conjunto de *especificações*, cada uma delas associada a uma linguagem contida na linguagem da planta. As especificações representam os comportamentos desejáveis, admissíveis, para o sistema. Para restringir o comportamento do sistema não controlado, utiliza-se um *supervisor*, uma entidade que observa os *eventos* gerados pela planta e, em função das sequências observadas, desabilita a ocorrência de certos eventos. Nem todos os eventos podem ser desabilitados pelo supervisor, fato que leva à partição dos eventos da planta em eventos *controláveis* e *não controláveis*. Um importante resultado da TCS estabelece a condição que planta e especificação devem satisfazer, para garantir a existência de um supervisor tal que o comportamento do sistema em malha fechada (i.e. planta + supervisor) seja exatamente aquele permitido pelas especificações. Nesse caso, a linguagem dada pela especificação é dita ser *controlável*. Caso a condição mencionada não seja satisfeita, outro resultado da TCS garante a existência da *maior* linguagem controlável contida na linguagem que representa a especificação. Essa nova linguagem corresponde ao comportamento minimamente restritivo capaz de ser obtido por supervisão. Um objeto matemático que permite a implementação computacional dos resultados da TCS, incluindo a síntese de supervisores, são os autômatos finitos.

Em nossa abordagem, o supervisor observa o *estado* em que a rede se encontra e restringe os eventos que podem acontecer ou, ainda, *força* certos eventos, para que as especificações de controle sejam atingidas. A rede evolui, então, segundo as determinações impostas pelo supervisor, que observa o novo estado alcançado por ela e reinicia mais um ciclo de medição e atuação, prosseguindo dessa maneira até que os estados finais sejam atingidos. Na Seção 5.3, determinamos as condições matemáticas que devem ser satisfeitas para que o problema de controle tenha uma solução. Nas Seções 5.2 e 5.3 serão expostas as contribuições à Teoria de Controle Supervisório elencadas na Introdução desta Tese.

A realização do supervisor baseia-se na possibilidade de construir moléculas sintéticas capazes de efetuar a operação lógica “se <estado>, então <ação>”, operação que, como vimos no capítulo passado, já foi construída a partir de diferentes circuitos moleculares, dentre eles, circuitos compostos por genes sintéticos. Na Seção 5.4, mostraremos como tal função pode ser especificamente implementada por genes sintéticos cujas regiões de regulação são programadas para detectar estados, disparando ações específicas, as sínteses das proteínas correspondentes. Ainda na Seção 5.4, apresentaremos resultados da literatura que embasam as nossas duas principais hipóteses biológicas acerca dos genes sintéticos:

- i) As regiões de regulação dos genes sintéticos podem ser programadas para induzir sua transcrição em estados celulares específicos;
- ii) Os genes sintéticos podem ser projetados de forma que a dinâmica dos processos de transcrição e tradução seja mais rápida do que aquela verificada em certos genes nativos (o que se traduz na pre-emptibilidade de eventos, que será discutida no âmbito da Teoria de Controle Supervisório).

5.1.1 Exemplo a ser Tratado no Capítulo

Ao longo do capítulo, aplicaremos as ideias elaboradas a uma rede de regulação gênica teórica. Essa rede consiste em três genes que interagem por meio dos seus fatores de transcrição segundo as funções de atualização de estados descritas na Figura 24.

Consideraremos como estado da rede de regulação gênica na Figura 24 o conjunto dos estados Booleanos individuais de cada um dos seus genes, *ativado* ou *desativado*, estados esses atualizados assincronamente pelas funções f_i . Assim, por exemplo, se o sistema encontra-se em 110, ocorre uma transição para 010, dado que $f_1 = 0$ em 110, pois $v_3 = 0$. Modelaremos a rede como um autômato de estados finitos, seguindo as ideias apresentadas no Capítulo 3. Dadas essas premissas, a Figura 25 mostra o diagrama de transição de estados da rede-exemplo e o autômato correspondente, quando $q_0 = 010$.

Importante ressaltarmos que o modelo da rede de regulação gênica na forma de um autômato de estados finitos pode ser obtido de diferentes maneiras, não estando restrito à forma particular utilizada para obtenção do autômato G da Figura 25(b), como ficará mais claro após a apresentação do Capítulo 7.

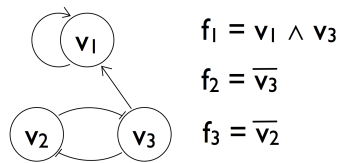


Figura 24 – Rede de regulação gênica que será utilizada como exemplo ao longo deste capítulo. Ela é composta por três genes que interagem por meio de seus fatores de transcrição segundo as funções de atualização de estados f_i . O símbolo ‘ \wedge ’ denota o operador lógico E, enquanto $\overline{v_i}$ é interpretado como a negação lógica da variável v_i . Assim, a expressão $f_3 = \overline{v_2}$ indica que o gene v_3 será desativado quando $v_2 = 1$, o que traduz matematicamente a repressão que v_2 exerce sobre v_3 .

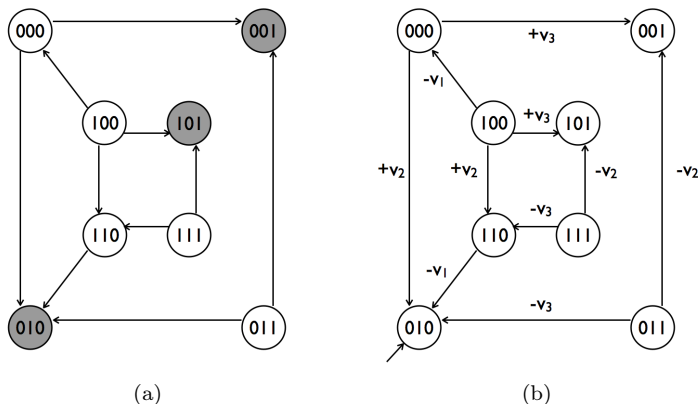


Figura 25 – (a) Diagrama de transição de estados da rede de regulação gênica da Figura 24. Se o sistema encontra-se no estado 110, por exemplo, o gene v_1 será desativado, pois $f_1 = v_1 \wedge v_3 = 1 \wedge 0 = 0$. Assim, o sistema faz uma transição ao estado 010. (b) Representação gráfica do autômato de estados finitos G correspondente ao diagrama em (a), quando 010 é o estado inicial. Estados pintados em cinza correspondem aos atratores da rede. Note que, sem perturbações em 010, o único estado acessível em G é o próprio estado inicial.

5.2 FORMALIZAÇÃO DO PROBLEMA DE CONTROLE

Suponha ser possível inserir na célula genes sintéticos $v_i^s \in V^s$ que satisfazem as seguintes hipóteses:

- a) o gene v_i^s codifica para o mesmo fator de transcrição que o gene nativo $v_i \in V$, o fator de transcrição u_i ;
- b) a região de regulação de v_i^s pode ser construída de tal forma que sua transcrição se inicie em qualquer estado ou combinação de estados. Essa hipótese implica que a ativação ($+v_i^s$) e desativação ($-v_i^s$) do gene v_i^s são eventos *controláveis*, na acepção de que eles podem ser desabilitados;
- c) a eficiência na expressão dos genes sintéticos pode ser ajustada de modo que a dinâmica de seus processos de transcrição e tradução seja mais rápida do que aquela associada a *alguns* genes nativos (BLAZECK et al., 2012; GINGOLD; PILPEL, 2011). Assim, a ativação de um gene sintético v_i^s qualquer pode *preemptar* a ocorrência de eventos relacionados a *alguns* genes nativos. Chamaremos os eventos que podem ser preemptados de *eventos preemptáveis*.

Em nossa abordagem, a modificação no comportamento dinâmico e de regime permanente de uma rede de regulação gênica dar-se-á unicamente pela produção *regulada* de fatores de transcrição endógenos a partir de genes sintéticos com as propriedades listadas no parágrafo anterior. Decorre dessa proposta que a ativação e desativação de genes nativos não pode ser controlada diretamente, motivo pelo qual classificaremos esses eventos como *não controláveis*. Tais eventos podem, ainda, ser subdivididos em preemptáveis ou não preemptáveis, dependendo de quão rápidas são as dinâmicas de seus processos de transcrição e tradução. Preemptabilidade de eventos em nossa abordagem está, pois, ligada à dinâmica temporal *relativa* de processos biológicos.

Como exemplo, suponha que genes sintéticos v_1^s e v_3^s sejam escolhidos para controlar a rede contida na Figura 24. A Figura 26 ilustra esquematicamente a nossa proposta de controle quando aplicada a esse exemplo em particular. Inicialmente, consideraremos uma dinâmica irrestrita para os genes sintéticos (ou seja, eles *podem* ser ativados ou desativados em quaisquer estados) para, após a etapa de síntese do controlador, definir um padrão de ativação que resulte na resolução do problema de controle.

A nossa escolha por genes sintéticos que codificam para os fatores de transcrição u_1 e u_3 para tratar o exemplo foi arbitrária, mas poderia

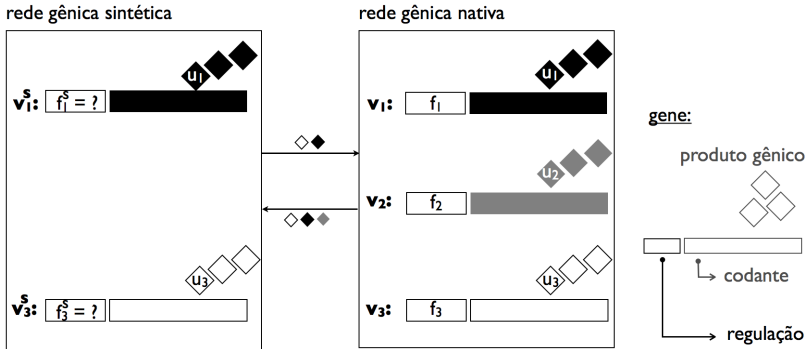


Figura 26 – Exemplo do conceito de controle proposto em nossa Tese. A rede de regulação gênica da Figura 24 é controlada por genes sintéticos que: i) detectam a presença de u_1 , u_2 e u_3 ; e ii) respondem a esses fatores de transcrição por meio da síntese regulada de u_1 e u_3 . As regiões codantes de v_1^s e v_3^s são conhecidas, codificam para u_1 e u_3 , mas o impacto de u_1 , u_2 e u_3 sobre a expressão de v_1^s e v_3^s (representado pelas funções f_1^s e f_3^s) não está determinado ainda. Essa informação só será definida após a resolução do problema de controle.

ter sido o resultado de restrições biológicas concretas. Por exemplo, se a energia necessária para síntese de u_2 é elevada, produções adicionais de tal molécula poderiam causar distúrbios no comportamento celular indesejáveis e difíceis de prever. A escolha dos fatores de transcrição a serem empregados para controle da dinâmica celular pode ainda ser resultado de uma análise de sensibilidade, em que as moléculas com maior impacto na dinâmica da rede de regulação gênica nativa são selecionadas (veja-se, por exemplo, a análise conduzida no trabalho de Galloway e colaboradores (GALLOWAY; FRANCO; SMOLKE, 2013)).

5.2.1 Modelo da Planta: Autômato G^{au}

De modo a considerar as possibilidades de controle obtidas pela inclusão de genes sintéticos na célula, devemos adicionar ao autômato G , que modela o comportamento da rede nativa, os eventos e estados introduzidos pela presença dos genes sintéticos v_i^s , resultando em um novo autômato que chamaremos G^{au} .

Começemos por construir a rede Booleana que descreve o sis-

tema completo, contendo genes nativos e sintéticos. Essa rede é definida pelo par $(V \cup V^s, F^e \cup F^s)$, em que V^s é o conjunto de genes sintéticos v_j^s , $j \in \mathcal{S} \subseteq \{1, \dots, N\}$, usados para fins de controle e $F^e = \{f_i^e\}$, $1 \leq i \leq N$, o conjunto das novas funções de atualização dos genes nativos v_i , que incluem os efeitos regulatórios resultantes da ativação dos genes sintéticos $v_j^s \in V^s$ (lembre-de de que a ativação de v_j^s induz a síntese de u_j). As funções $f_i^e : \{0, 1\}^{N+|\mathcal{S}|} \rightarrow \{0, 1\}$ são dadas por $f_i^e(\mathbf{v} \vee \mathbf{w}(\mathbf{v}^s))$, em que $\mathbf{w} = [w_1 \dots w_N]$ para $w_k = v_k^s$ se $k \in \mathcal{S}$, e $w_k = 0$, caso contrário. A notação $\mathbf{v} \vee \mathbf{w}$ é uma abreviação de $[v_1 \vee w_1 \dots v_N \vee w_N]$. As novas funções $f_i^e \in F^e$ traduzem matematicamente o fato biológico de que o efeito regulatório da proteína produzida a partir de v_j^s é o mesmo daquela sintetizada de v_j . Assim, se um gene v_i é reprimido quando v_j está ativado, por exemplo, ele também será reprimido quando v_j^s estiver ativado, portanto $f_i^e = \overline{(v_j \vee v_j^s)}$.

O conjunto $F^s = \{f_j^s\}_{j \in \mathcal{J}}$ contém as funções de atualização $f_j^s : \{0, 1\}^{|\mathcal{S}|} \rightarrow \{0, 1\}$ que modelam a dinâmica *irrestrita* dos genes sintéticos que podem ser inseridos na célula. Essas funções serão definidas como $f_j^s(v_j^s) = \overline{v_j^s}$, cuja semântica é dada por “toda vez que v_j^s estiver ativado (ou desativado), ele pode ser desativado (ativado)”. Tais f_j^s formalizam a hipótese de que os genes sintéticos podem ser programados para iniciar a transcrição em qualquer estado ou combinação de estados. Como discutiremos adiante, uma vez definido o supervisor que resolve o problema de controle, as funções de atualização dos genes sintéticos assumirão expressões diferentes, que refletem o padrão específico de ativação/desativação requerido pelo supervisor.

É possível obter um diagrama de transição de estados para $(V \cup V^s, F^e \cup F^s)$ da mesma forma como fizemos para a rede (V, F) . Adicionalmente, pode-se associar ao diagrama de transição de estados assim obtido um autômato $G^{\text{au}} = (Q^{\text{au}}, \Sigma^{\text{au}}, \delta^{\text{au}}, q_0^{\text{au}}, Q_m^{\text{au}})$. Esse autômato possui as seguintes propriedades:

- Os estados $q^{\text{au}} \in Q^{\text{au}}$ contêm uma cadeia binária de tamanho $N + |\mathcal{S}|$, com os valores dos estados de $v_i \in V$ e $v_j^s \in V^s$;
- O conjunto Σ^{au} é composto pelos eventos $+v_i$, $-v_i$, $+v_j^s$ e $-v_j^s$, com $1 \leq i \leq N$ e $j \in \mathcal{S}$, associados com a ativação e desativação dos genes nativos e sintéticos, respectivamente;
- $\delta^{\text{au}} : Q^{\text{au}} \times \Sigma^{\text{au}} \rightarrow Q^{\text{au}}$ é a função de transição de estados parcial, trivialmente derivada do diagrama de estados $(V \cup V^s, F^e \cup F^s)$. Como adotamos uma atualização assíncrona da rede completa, as cadeias binárias associadas a dois estados $q, r \in Q^{\text{au}}$ unidos por

uma transição $r = \delta^{\text{au}}(q, \sigma)$ diferem em no máximo um bit. Por exemplo, se $v_i = 0$ em q e $v_i = 1$ em r , então $r = \delta^{\text{au}}(q, +v_i)$;

- q_0^{au} é o estado inicial, em que os valores de v_i são os mesmos adotados em q_0 de G , e os valores de v_i^s são iguais a zero;
- Q_m^{au} é o conjunto de estados associados a uma tarefa completa.

Nem todos os estados q^{au} podem ser alcançados a partir de q_0^{au} . Consideraremos, no restante deste capítulo, que G^{au} é o resultado de $Ac(G^{\text{au}})$, em que $Ac(\cdot)$ denota a operação que retorna a parte acessível de um autômato (CASSANDRAS; LAFORTUNE, 2008).

As hipóteses relativas aos genes sintéticos e nativos implicam uma partição no conjunto de eventos Σ^{au} da forma $\Sigma^{\text{au}} = \Sigma_{cf}^{\text{au}} \cup \Sigma_{cf}^{\text{au}} \cup \Sigma_{cfp}^{\text{au}} \cup \Sigma_{cf\bar{p}}^{\text{au}}$, em que:

- Σ_{cf}^{au} é o conjunto de eventos que podem ser controlados e forçados a preemptar outros eventos em Σ_{cf}^{au} . Esse conjunto é formado exclusivamente pelos eventos $+v_i^s$;
- Σ_{cf}^{au} é o conjunto de eventos que podem ser controlados, mas não forçados a preemptar outros eventos em Σ^{au} . Esse conjunto é formado exclusivamente pelos eventos $-v_i^s$;
- Σ_{cfp}^{au} é o conjunto de eventos relacionados aos genes nativos que podem ser preemptados por eventos em Σ_{cf}^{au} . Todos os eventos ligados aos genes nativos são, pela definição de nossa forma de intervenção, não controláveis e não forçáveis;
- $\Sigma_{cf\bar{p}}^{\text{au}}$ é o conjunto de eventos relacionados aos genes nativos que não podem ser preemptados por nenhum evento em Σ_{cf}^{au} .

A Figura 27 mostra o autômato G^{au} obtido para o nosso exemplo, quando q_0^{au} é tal que $v_1 = v_3 = v_1^s = v_3^s = 0$ e $v_2 = 1$. Na figura, os eventos forçáveis são representados por arcos com um círculo, ao passo que os eventos controláveis são designados por setas cortadas. Determinados estados de G^{au} são caracterizados pela síntese dos mesmos fatores de transcrição. Esses estados são agrupados na Figura 27. Por exemplo, nos estados $\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & \end{smallmatrix}$ e $\begin{smallmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \end{smallmatrix}$, os mesmos fatores de transcrição estão presentes, u_1 e u_3 . No primeiro estado, u_3 é sintetizado a partir de v_3^s , enquanto, no segundo estado, v_3 é o gene responsável pela produção de u_3 . Do ponto de vista fisiológico, no entanto, ambos os estados são equivalentes, dado que é o conjunto de proteínas presentes em um dado momento o fator determinante do comportamento celular.

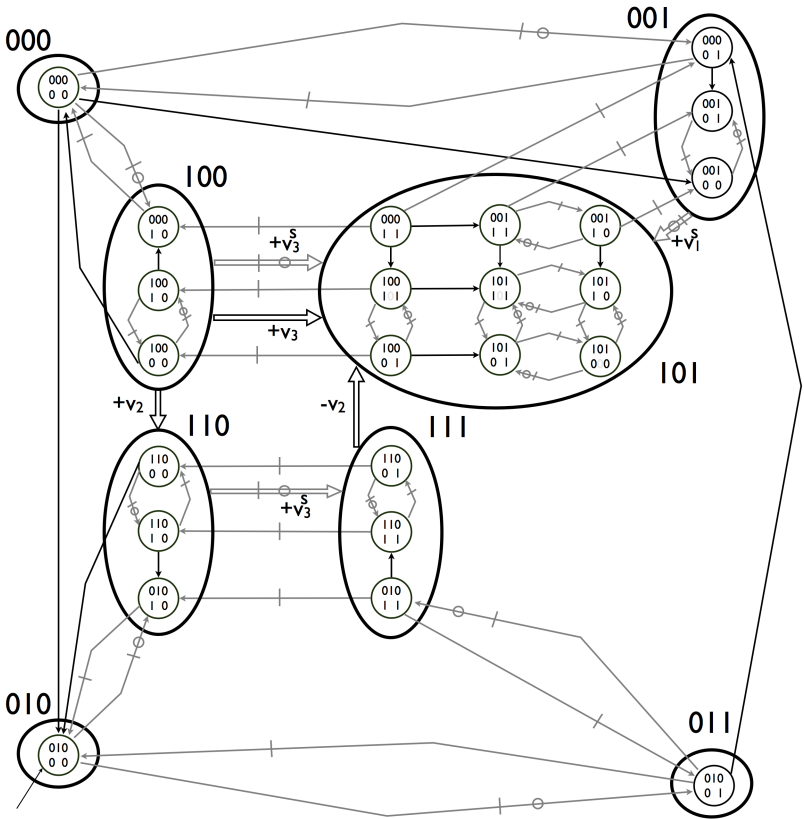


Figura 27 – O autômato G^{au} capta o comportamento da planta contendo os genes nativos v_1, v_2 e v_3 , bem como os genes sintéticos v_1^s e v_3^s . Cada estado de G^{au} é representado por um círculo, em que a primeira linha contém os valores das variáveis de estado v_1, v_2, v_3 , enquanto a segunda linha mostra os valores das variáveis de estado v_1^s e v_3^s . Estados caracterizados pela síntese dos mesmos fatores de transcrição são agrupados (e etiquetados por uma cadeia indicando quais são os fatores de transcrição sintetizados). As setas que representam os eventos $+v_i$ e $-v_i$ (não controláveis) estão pintadas de preto. As setas cinzas são aquelas associadas às transições $+v_i^s$ (forçáveis e controláveis) e $-v_i^s$ (controláveis). Eventos que podem acontecer a partir de *qualquer* estado dentro de um mesmo agrupamento estão desenhados por uma seta com o corpo mais largo.

A agregação de estados fisiologicamente equivalentes sugere a construção de um novo autômato, que chamaremos G^{hi} , cujo número de estados é menor. O procedimento empregado para obtenção de G^{hi} segue os passos mostrados em um trabalho anterior de Hubbard e Caines (HUBBARD; CAINES, 2002), que será detalhado na próxima subseção.

5.2.2 Abstração de G^{au} : Autômato G^{hi}

A introdução de genes sintéticos para fins de controle aumenta significativamente o número de estados do autômato finito que representa o comportamento dinâmico do sistema, como se depreende a partir da comparação entre as figuras 27 e 25(b). O número de estados aumenta de um máximo de 2^N em G , em que N é o número de nós da rede de regulação, a um máximo de $2^{N+|S|}$ em G^{au} , com $|S|$ sendo o tamanho do conjunto V^s , ou seja, o número de genes sintéticos utilizados.

Para evitar um aumento dos custos computacionais durante a etapa de síntese do supervisor, simplificaremos o autômato G^{au} para obter um autômato G^{hi} com menos estados (limitado a um máximo de 2^N). Essa simplificação baseia-se na ideia de que estados de G^{au} que sintetizam os mesmos fatores de transcrição podem ser agregados, dado que esses estados são equivalentes do ponto de vista fisiológico.

A Figura 28 mostra o autômato G^{hi} , derivado do G^{au} contido na Figura 27, para o caso em que todos os eventos não controláveis da rede nativa podem ser preemptados pela ativação dos genes sintéticos v_1^s e v_3^s . Uma análise comparada das Figuras 28 e 27 permite identificar os três principais conceitos subjacentes a essa etapa de simplificação:

- Um estado em G^{hi} corresponde a uma certa combinação de *fatores de transcrição* sintetizados em um dado momento, em vez de uma combinação de estados individuais dos genes v_i e v_i^s . Ressalte-se que um fator de transcrição u_i está presente na célula se, e somente se, v_i ou v_i^s estão ativos;
- Transições entre estados de G^{hi} podem ser interpretadas como: a) um aumento na concentração do fator de transcrição u_i acima de um valor de limiar, evento denotado por $+u_i$; ou b) um decréscimo na concentração do fator de transcrição u_i abaixo de um valor de limiar, evento representado por $-u_i$.
- O conjunto de eventos de G^{hi} é particionado de modo a refletir

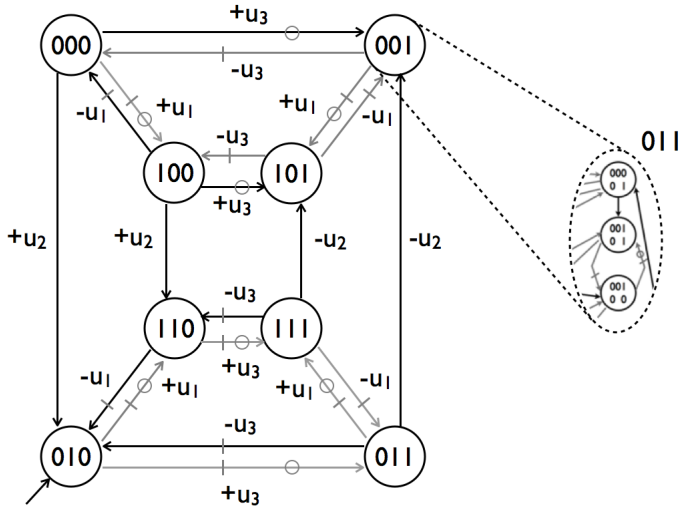


Figura 28 – Autômato G^{hi} obtido a partir de G^{au} . G^{hi} registra quais fatores de transcrição são sintetizados em um dado momento. Os estados de G^{hi} são, assim, uma agregação dos estados de G^{au} , conforme mostrado na Figura 27. G^{hi} pode ser interpretado como uma visão de alto nível do autômato G^{au} . Os elementos da representação gráfica do autômato G^{hi} pintados em cinza correspondem a “fatos novos”, sejam eles eventos ou características qualitativas dos eventos, como controlabilidade e forçabilidade, devidos à introdução de genes sintéticos para fins de controle (compare G^{hi} com o autômato G da Figura 25(b)).

se as transições podem ser controladas, forçadas ou preemptadas. Esse particionamento dependerá da estrutura do autômato de baixo nível G^{au} .

Até o final desta subseção, formalizaremos as ideias delineadas qualitativamente até aqui, de maneira a obter, para o caso geral, o autômato $G^{hi} = (X, \Sigma^{hi}, \delta^{hi}, x_0, X_m)$ de G^{au} .

Inicialmente, considere a função $E : Q^{au} \rightarrow \{0, 1\}^N$, definida da seguinte forma:

$$E_i(q) = \begin{cases} v_i \vee v_i^s, & \text{se } v_i \in V^s \\ v_i, & \text{caso contrário} \end{cases}$$

A função E associa a cada $q \in Q^{au}$ uma cadeia de 0s e 1s com tamanho

N , em que a i -ésima entrada E_i indica se o fator de transcrição u_i é ou não sintetizado naquele estado.

Em seguida, agrupe os estados $q \in Q^{\text{au}}$ de tal forma que $q, r \in Q^{\text{au}}$ pertençam ao mesmo x se, e somente se, $E(q) = E(r)$. Tal agrupamento define uma partição $X = \{x^1, \dots, x^h\}$ sobre Q^{au} , com $x^i \subseteq Q^{\text{au}}$, $\bigcup x^i = Q^{\text{au}}$ e $x^i \cap x^j = \emptyset$, para $i \neq j$, $1 \leq i, j \leq h$. A cada elemento da partição $x \in X$ associa-se uma cadeia binária resultante da aplicação de $E(\cdot)$ aos estados $q \in x$, com x_k denotando o k -ésimo elemento dessa cadeia. Os conjuntos $x \in X$ correspondem aos estados do novo autômato G^{hi} . Note, no entanto, que cada estado x de G^{hi} é também um conjunto de estados $q \in Q^{\text{au}}$ e, por isso, operações definidas sobre conjuntos podem ser aplicadas aos elementos $x \in X$, como $x \cup y$. Ao longo da Tese, para evitar confusões, chamaremos os estados $q \in Q^{\text{au}}$ de *microestados*, e os estados $x \in X$, *macroestados*.

Antes de prosseguirmos com a definição de Σ^{hi} e δ^{hi} , alguns novos conceitos serão apresentados.

Definição 1. *O conjunto $I(x) \subseteq x$ de um macroestado $x \in X$ é o conjunto que contém:*

a) $q = q_0^{\text{au}}$, se $q_0^{\text{au}} \in x$, e

b) *aqueles microestados $q \in x$ para os quais*

$$\exists r \in x^C \text{ tal que } \exists \sigma \in \Sigma^{\text{au}} \text{ com } \delta^{\text{au}}(r, \sigma) = q,$$

em que x^C denota o complemento do conjunto x em Q^{au} .

Ou seja, $I(x)$ é o conjunto de microestados de entrada em x (HUBBARD; CAINES, 2002).

Definição 2. *Um macroestado $x \in X$ é dito ter uma transição para outro macroestado $y \in X$, equivalentemente, a relação $\langle x y \rangle$ é válida, sempre que $\exists q \in I(y)$ e $r \in x$ satisfazendo*

$$\exists \sigma \in \Sigma^{\text{au}} \text{ tal que } \delta^{\text{au}}(r, \sigma) = q.$$

Devido à hipótese de que estados em G^{au} são atualizados assincronamente, a transição $\delta^{\text{au}}(r, \sigma) = q$ somente é definida para microestados q, r cujas cadeias binárias diferem precisamente em um bit apenas. Essa propriedade é herdada por G^{hi} : estados $x, y \in X$ podem ser conectados por um evento se, e somente se, as cadeias binárias a eles associadas diferem em um bit somente.

O conjunto de eventos Σ^{hi} e a função de transição parcial δ^{hi} de G^{hi} são, então, definidas trivialmente da seguinte maneira:

- Se $\langle x, y \rangle$ é válido, com $x_k = 1$, $y_k = 0$, $1 \leq k \leq N$, então $-u_k \in \Sigma^{\text{hi}}$ e $\delta^{\text{hi}}(x, -u_k) = y$;
- Se $\langle x, y \rangle$ é válido, com $x_k = 0$, $y_k = 1$, $1 \leq k \leq N$, então $+u_k \in \Sigma^{\text{hi}}$ e $\delta^{\text{hi}}(x, +u_k) = y$.

Para fins de controle, particionaremos o conjunto Σ^{hi} de acordo com as características qualitativas de seus eventos, como controlabilidade, forçabilidade e preemptabilidade. Para melhor entender esse procedimento, recomendamos ao leitor uma análise cuidadosa das Figuras 29 e 30. Elas mostram, respectivamente, todas as possíveis configurações em que um evento $-u_i$ ou $+u_i$ conectam dois macroestados x e y . Por exemplo, na Figura 29(a), caso em que o gene nativo v_i é ativado pelos fatores de transcrição presentes em x , observa-se que somente a partir de microestados em x que satisfazem $v_i = 0$ e $v_i^s = 1$ é possível atingir microestados em y . Essas transições são sempre controláveis, o que implica $-u_i \in \Sigma_{c\bar{f}}^{\text{hi}}$ para o caso tratado. Nessas figuras, e ao longo do documento, os subscritos de cada elemento da partição de Σ^{hi} denotam se os eventos pertencentes a esse elemento são controláveis ou não (c ou \bar{c}), forçáveis ou não (f ou \bar{f}), preemptáveis ou não (p ou \bar{p}).

Para formalizar a partição de Σ^{hi} , considere as definições a seguir. Elas são escritas de modo a refletir a estrutura particular do autômato G^{au} , bem como o particionamento peculiar dos estados de Q^{au} .

Definição 3. *Uma relação $\langle x y \rangle$ é dita ser forçável, representada por $\langle x y \rangle_f$, se, e somente se,*

$$\forall q \in I(x) : \exists \sigma \in \Sigma_{cf}^{\text{au}} \text{ tal que } \delta^{\text{au}}(q, \sigma) \in y.$$

Assim, a relação $\langle x y \rangle_f$ é válida toda vez que houver um evento forçável em todos os microestados pertencentes a $I(x)$ que leva o sistema a um microestado em y (veja Figura 30, situações (a) e (b)).

Definição 4. *Uma relação $\langle x y \rangle$ é dita ser controlável, representada por $\langle x y \rangle_c$, se, e somente se,*

$$\begin{aligned} \forall q \in I(x) : \exists \sigma \in \Sigma^{\text{au}} \text{ tal que } \delta^{\text{au}}(q, \sigma) \in y \rightarrow & (\sigma \in \Sigma_{cf}^{\text{au}} \cup \Sigma_{c\bar{f}}^{\text{au}}) \vee \\ & [\sigma \in \Sigma_{\bar{c}f\bar{p}}^{\text{au}} \wedge \exists \xi \in \Sigma_{cf}^{\text{au}} \text{ tal que } (r = \delta^{\text{au}}(q, \xi) \in x) \wedge \\ & (\bar{A}w \in (\Sigma_{\bar{c}f\bar{p}}^{\text{au}} \cup \Sigma_{\bar{c}f\bar{p}}^{\text{au}})^* \text{ com } \delta_e^{\text{au}}(r, w) = y \text{ e } \delta_e^{\text{au}}(r, w' < w) = x)]. \end{aligned}$$

A notação δ_e^{au} é empregada para representar a extensão trivial de δ^{au} para cadeias de eventos $w \in (\Sigma^{\text{au}})^*$, enquanto $w' < w$ denota as sub-

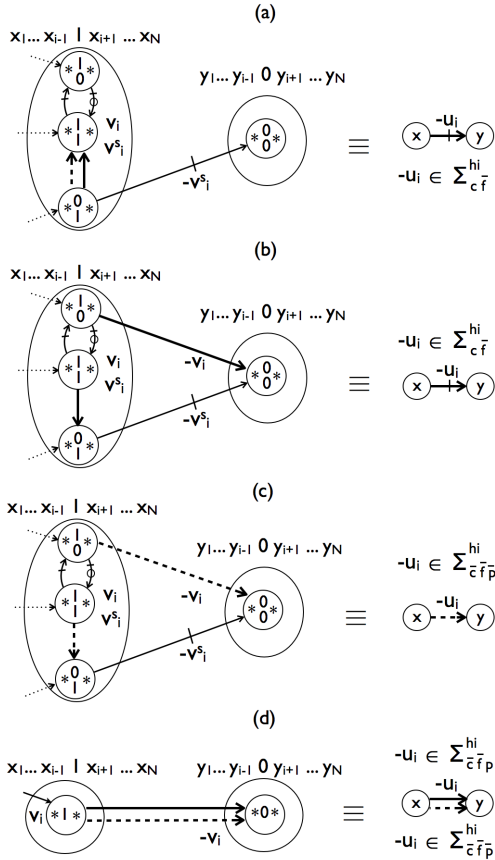


Figura 29 – Situações em G^{au} nas quais o evento de alto nível $-u_i$ conecta o macroestado x a y (cada um etiquetado pela cadeia binária que indica quais fatores de transcrição são sintetizados pela célula naquele macroestado). Os microestados estão agrupados de acordo com uma máscara que considera os valores de v_i e v_i^s somente, dado que essas são as únicas informações relevantes para o estudo de $-u_i$. Note que outros eventos de alto nível podem ocorrer em x e y (eles não são mostrados para reter a simplicidade na representação). Arcos pontilhados denotam microestados em x que *podem* ser microestados de entrada, dependendo do sistema. Os eventos $-v_i$ podem ser preemptíveis (arcos contínuos) ou não preemptíveis (arcos tracejados). (a) v_i é ativado em x ; (b) v_i é desativado em x e $-v_i$ é preemptível; (c) v_i é desativado em x e $-v_i$ é não preemptível; (d) $v_i^s \notin V^s$.

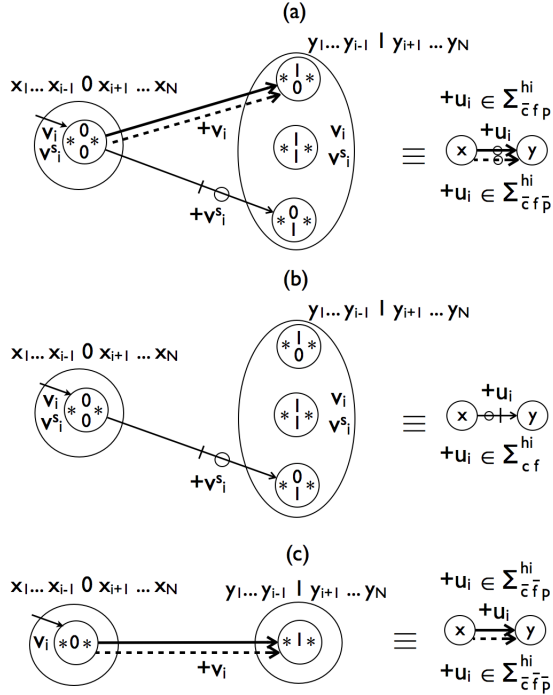


Figura 30 – Situações em G^{au} nas quais o evento de alto nível $+u_i$ conecta o macroestado x a y (cada um etiquetado pela cadeia binária que indica quais fatores de transcrição são sintetizados pela célula naquele macroestado). Os microestados estão agrupados de acordo com uma máscara que considera os valores de v_i e v_i^s somente, dado que essas são as únicas informações relevantes para o estudo de $-u_i$. Note que outros eventos de alto nível podem ocorrer em x e y (eles não são mostrados para reter a simplicidade na representação). Os eventos $+v_i$ podem ser preemptáveis (arcos contínuos) ou não preemptáveis (arcos tracejados). (a) v_i é ativado em x ; (b) v_i é desativado em x ; (c) $v_i^s \notin V^s$.

cadeias (ou prefixos) estritas de w . A relação $\langle x y \rangle_c$ é válida sempre que *todos* os eventos σ que levam o sistema a um microestado em y , a partir de um microestado em $I(x)$, puderem ser evitados, seja: a) desabilitando-os; ou b) preemptando-os, por meio de evento forçável ξ que traz o sistema a um microestado $r \in x$ a partir do qual não existe nenhuma cadeia de eventos não controláveis conduzindo diretamente a y (veja Figura 29(a),(b) e Figura 30(b)).

Definição 5. *Uma relação $\langle x y \rangle$ é dita ser não controlável mas preemptável, representada por $\langle x y \rangle_{\bar{c}p}$, se, e somente se,*

$$\forall q \in x : \exists \sigma \in \Sigma_{\bar{c}fp}^{\text{au}} \text{ tal que } \delta^{\text{au}}(q, \sigma) \in y.$$

A relação $\langle x y \rangle_{\bar{c}p}$ é válida sempre que *todos* os microestados $q \in x$ forem conectados a um microestado em y por meio de um evento σ não controlável, mas preemptável (veja Figuras 29(d) e 30(a),(c)). Nesse caso, a transição a um microestado em y não pode ser desabilitada, mas somente preemptada por um evento que *necessariamente* leva o sistema a um microestado que não pertence a x .

Definição 6. *Uma relação $\langle x y \rangle$ é dita ser não controlável e não preemptável, representada por $\langle x y \rangle_{\bar{c}\bar{p}}$, se, e somente se,*

$$\exists q \in I(x) : \exists \sigma \in \Sigma_{\bar{c}\bar{p}}^{\text{au}} \text{ tal que } \delta^{\text{au}}(q, \sigma) \in y.$$

A relação $\langle x y \rangle_{\bar{c}\bar{p}}$ é válida sempre que existir um microestado $q \in I(x)$ a partir do qual um evento não controlável e não preemptável σ leva o sistema a algum microestado em y (veja Figuras 29(c),(d) e 30(a),(c)).

O conjunto de eventos de alto nível do autômato G^{hi} pode, então, ser particionado, de uma maneira *dependente do estado*, da seguinte forma:

$$\Sigma_{cf}^{\text{hi}}(x) = \{\pi \in \Sigma^{\text{hi}} \mid \exists y \in X \text{ tal que } \delta^{\text{hi}}(x, \pi) = y \wedge \langle x y \rangle_c \text{ é válida} \wedge \langle x y \rangle_f \text{ é válida}\};$$

$$\Sigma_{cf}^{\text{hi}}(x) = \{\pi \in \Sigma^{\text{hi}} \mid \exists y \in X \text{ tal que } \delta^{\text{hi}}(x, \pi) = y \wedge \langle x y \rangle_c \text{ é válida} \wedge \langle x y \rangle_f \text{ não é válida}\};$$

$$\Sigma_{\bar{c}fp}^{\text{hi}}(x) = \{\pi \in \Sigma^{\text{hi}} \mid \exists y \in X \text{ tal que } \delta^{\text{hi}}(x, \pi) = y \wedge \langle x y \rangle_{\bar{c}p} \text{ é válida} \wedge \langle x y \rangle_f \text{ é válida}\};$$

$\Sigma_{\tilde{c}\tilde{f}\tilde{p}}^{\text{hi}}(x) = \{\pi \in \Sigma^{\text{hi}} \mid \exists y \in X \text{ tal que } \delta^{\text{hi}}(x, \pi) = y \wedge \langle x y \rangle_{\tilde{c}\tilde{p}} \text{ é válida} \wedge \langle x y \rangle_f \text{ é válida}\};$

$\Sigma_{\tilde{c}\tilde{f}\tilde{p}}^{\text{hi}}(x) = \{\pi \in \Sigma^{\text{hi}} \mid \exists y \in X \text{ tal que } \delta^{\text{hi}}(x, \pi) = y \wedge \langle x y \rangle_{\tilde{c}\tilde{p}} \text{ é válida} \wedge \langle x y \rangle_f \text{ não é válida}\};$

$\Sigma_{\tilde{c}\tilde{f}\tilde{p}}^{\text{hi}}(x) = \{\pi \in \Sigma^{\text{hi}} \mid \exists y \in X \text{ tal que } \delta^{\text{hi}}(x, \pi) = y \wedge \langle x y \rangle_{\tilde{c}\tilde{p}} \text{ é válida} \wedge \langle x y \rangle_f \text{ não é válida}\}.$

O conjunto $\Sigma_{\tilde{c}\tilde{f}}^{\text{hi}}(x)$ pode ser, ainda, particionado em dois conjuntos, conforme o evento π que vai de x a y possa ocorrer a partir de todo microestado $q \in x$ ou não. Tome $-u_i$ representada na Figura 29(a) como um exemplo. Se o sistema entra no macroestado x por intermédio de microestados em que $v_i^s = 1$ e $v_i = 0$, caso a transição $+v_i$ ocorra, então o sistema se desloca para microestados que não podem alcançar y sem sair de x . Formalmente, dizemos que $\pi \in \Sigma_{\tilde{c}\tilde{f}D}^{\text{hi}}(x)$ se, e somente se, para todo $q \in x$ existe um caminho $w \in (\Sigma^{\text{au}})^*$ tal que $\delta_e^{\text{au}}(q, w) \in y$ e $\delta_e^{\text{au}}(q, w' < w) \in x$ para qualquer subcadeia estrita w' de w . Caso contrário, $\pi \in \Sigma_{\tilde{c}\tilde{f}\tilde{D}}^{\text{hi}}(x)$.

Ao longo do texto, a notação utilizada para indicar a união de diferentes partições em Σ^{hi} será abreviada. Por exemplo, o conjunto de todos os eventos controláveis $\Sigma_{c_f}^{\text{hi}} \cup \Sigma_{\tilde{c}\tilde{f}}^{\text{hi}}$ será escrito simplesmente como $\Sigma_c^{\text{hi}}(x)$.

Finalmente, $x_0 \in X$ é definido como o macroestado que contém $q_0^{\text{au}} \in Q^{\text{au}}$, e a escolha do conjunto marcado X_m depende da tarefa de controle específica a ser resolvida, como veremos adiante. Assim, completamos a definição do autômato de alto nível G^{hi} e da natureza dos eventos a ele associados.

5.2.3 Definição das Entradas de Controle

O problema de controle será formulado como um problema de realimentação de estados, em uma abordagem baseada na Teoria de Controle Supervisório (WONHAM, 2010). Segundo esse corpo teórico, um supervisor para o autômato G^{hi} é definido como uma função $S^{\text{hi}} : X \rightarrow 2^{\Sigma^{\text{hi}}}$, ou seja, o supervisor associa a estados $x \in X$ um subconjunto dos eventos em Σ^{hi} , que denotam quais eventos são permitidos (em nossa abordagem, eles podem ser simplesmente habilitados ou, então, forçados pelo supervisor). Assim, o supervisor desabilita ou preempta

a ocorrência de uma parte dos eventos possíveis em x .

Suporemos que uma entrada de controle válida $S^{\text{hi}}(x)$ em um macroestado x possui uma das duas formas que seguem:

- 1) $S^{\text{hi}}(x) = \lambda \cup \Sigma_{c\bar{p}}^{\text{hi}}(x)$, com $\lambda = \{\pi_f\}$ para $\pi_f \in \Sigma_f^{\text{hi}}(x)$;
- 2) $S^{\text{hi}}(x) = \lambda \cup \Sigma_{c\bar{e}}^{\text{hi}}(x)$, com $\lambda \in 2^{\Sigma_{c\bar{f}}^{\text{hi}}(x)}$.

Tais entradas de controle $S^{\text{hi}}(x)$ serão chamadas de *admissíveis* e o conjunto de todos os controles admissíveis em x será representado por $\mathcal{A}_S^{\text{hi}}(x)$. Por vezes, podemos nos referir a essas entradas de controle como controles do tipo (1) e (2).

A entrada de controle do tipo (1) é empregada para garantir que eventos não controláveis, mas preemptáveis, não ocorram no alto nível. Isso só é possível por meio do uso de eventos forçáveis. Embora tal $S^{\text{hi}}(x)$ contenha possivelmente mais do que um evento forçável (pois além de $\pi_f \in \lambda$, $\Sigma_{c\bar{p}}^{\text{hi}}(x)$ também pode conter eventos forçáveis), destacamos um dentre eles, o evento π_f , para reforçar a ideia de que, na implementação de $S^{\text{hi}}(x)$, optaremos sempre por forçar *somente um* evento de alto nível. A justificativa para essa decisão repousa no fato de que forçar um evento no alto nível implica, biologicamente, ativar genes sintéticos e, por conseguinte, consumir energia celular. Ao forçarmos todos os eventos em $S^{\text{hi}}(x)$ do tipo (1), estaríamos optando por uma realização ineficiente do ponto de vista energético. A escolha aqui feita *não impacta* a resolução do problema de controle, mas tão somente a realização.

A entrada de controle do tipo (2), por outro lado, habilita em x um subconjunto λ dos eventos controláveis (note que λ não pode conter um elemento de $\Sigma_{c\bar{f}}^{\text{hi}}(x)$, dado que isso implicaria forçar um evento, um caso já tratado pelo controle do tipo (1)). A entrada de controle do tipo (2) precisa também habilitar todos os eventos não controláveis em x (mesmo aqueles preemptáveis, já que eles só podem ser preemptados por eventos forçáveis, que não fazem parte dessa entrada).

$S^{\text{hi}}/G^{\text{hi}}$ denota o comportamento em malha fechada do sistema composto por G^{hi} sob ação do supervisor S^{hi} . O estado inicial x_0 de G^{hi} é também o estado inicial de $S^{\text{hi}}/G^{\text{hi}}$. A entrada de controle $S^{\text{hi}}(x) \in \mathcal{A}_S^{\text{hi}}(x)$ aplicada ao estado $x \in S^{\text{hi}}/G^{\text{hi}}$ permite a ocorrência de transições para estados y , satisfazendo $y = \delta^{\text{hi}}(x, \sigma)$ para todos $\sigma \in S^{\text{hi}}(x)$.

Usaremos a notação $\mathcal{H}_\phi^{\text{hi}}(x)$ para representar o conjunto de macroestados y que podem ser alcançados a partir de x (i.e. sua imagem) por meio da entrada de controle admissível $\phi = S^{\text{hi}}(x)$. Embora $\mathcal{H}_\phi^{\text{hi}}(x)$ seja um conjunto de macroestados $y \in X$, $\mathcal{H}_\phi^{\text{hi}}(x)$, por

vezes, será tratado como o conjunto de microestados por ele encampado. Logo, dado $\gamma \in 2^{Q^{\text{au}}}$, sentenças do tipo $\gamma \subseteq \mathcal{H}_\phi^{\text{hi}}(x)$ são válidas. O conjunto $\mathcal{H}_\phi^{\text{hi}}(x)$ é definido da seguinte forma:

- a) se $S^{\text{hi}}(x) \subseteq \Sigma_{c\bar{f}\bar{D}}^{\text{hi}}(x)$, então $\mathcal{H}_\phi^{\text{hi}}(x) = \{x\} \cup \{y \mid y = \delta^{\text{hi}}(x, \pi) \text{ para } \pi \in S^{\text{hi}}(x)\}$. Ou seja, se o supervisor só permite a ocorrência de eventos $\pi \in \Sigma_{c\bar{f}\bar{D}}^{\text{hi}}(x)$, a imagem de x para o controle $\phi(x) = S^{\text{hi}}(x)$ inclui também o macroestado x , pois π pode nunca ocorrer em x ;
- b) se $S^{\text{hi}}(x) \not\subseteq \Sigma_{c\bar{f}\bar{D}}^{\text{hi}}(x)$, então $\mathcal{H}_\phi^{\text{hi}}(x) = \{y \mid y = \delta^{\text{hi}}(x, \pi) \text{ para } \pi \in S^{\text{hi}}(x)\}$.

5.2.4 Problema de Alcançabilidade Estrita

Nesta subseção, formalizaremos o problema de *alcançabilidade estrita*. Qualitativamente, esse problema consiste em levar um sistema qualquer de um estado inicial x_0 a um conjunto final de estados X_m , mantendo-o nesse conjunto por tempo indeterminado. Esses estados-alvo poderiam representar, por exemplo, uma região do espaço de estados celular em que a produção de um dado metabólito em bactérias é otimizado; o “ponto de operação” de uma rede biológica que foi perturbada pela ocorrência de um evento inesperado; um tipo celular especializado a ser obtido a partir de células-tronco, etc. Tal tarefa de controle encontra diversas aplicações em áreas como Engenharia Metabólica, Engenharia Tecidual e pesquisa terapêutica com células-tronco.

Para definir formalmente o problema de alcançabilidade estrita, empregaremos o conceito de *predicado*, uma função $P : X \rightarrow \{0, 1\}$ que usaremos para distinguir macroestados $x \in X$ que satisfazem alguma propriedade biológica de interesse daqueles que não o fazem. Por exemplo, poder-se-ia imaginar um predicado P que assume valor lógico 0 naqueles estados com um alto consumo médio de energia celular. X_P denota o conjunto de estados x para os quais $P(x) = 1$ e $x \models P$ é uma notação sinônima a $x \in X_P$. Adicionalmente, usaremos a abreviação $P = \text{verdadeiro}$ (ou $P = \text{falso}$) para os casos em que $P(x) = 1$ para todos os $x \in X$ ($P(x) = 0$ para todos os $x \in X$).

Problema 1. *Dada uma planta $G^{\text{hi}} = (X, \Sigma^{\text{hi}}, \delta^{\text{hi}}, x_0, X_m)$ e um predicado $P \neq \text{falso}$, com $X_m \subseteq X_P$, encontre um supervisor S^{hi} tal que a sequência $\{Y_i\}_{i=0}^l$, definida por:*

1. $Y_0 = \{x_0\}$; e

2. $Y_i = \bigcup_{x \in Y_{i-1}} \mathcal{H}_\phi^{\text{hi}}(x)$, $1 \leq i \leq l$, para $\phi = S^{\text{hi}}(x)$,

satisfaça:

- i) $Y_i \subseteq X_P$, para $0 \leq i \leq l$;
- ii) $Y_l \subseteq X_m$;
- iii) Existe $Y^* \subseteq X_m$, tal que $Y_l \subseteq Y^*$ e $Y^* = \bigcup_{x \in Y^*} \mathcal{H}_\phi^{\text{hi}}(x)$ para $\phi = S^{\text{hi}}(x)$.

Note que, além de convergir a um subconjunto X_m a partir de x_0 , sem sair de X_P , requer-se que o sistema convirja para macroestados que permaneçam em X_m (esse conjunto invariante é representado por Y^* na formulação acima). O predicado P pode ser adotado para especificar propriedades adicionais a serem satisfeitas pelo sistema de controle em malha fechada na trajetória de x_0 a X_m .

5.3 CONDIÇÕES PARA EXISTÊNCIA DE UM SUPERVISOR

Para determinar as condições que, se satisfeitas, garantem que o problema de alcançabilidade estrita possui solução, algumas definições preliminares são necessárias.

Definição 7. A função controlável $\Phi_W : 2^W \rightarrow 2^W$, para $W \subseteq X$, é definida como

$$\Phi_W(Z) = \{x \in Z \mid \exists \phi \in \mathcal{A}_S^{\text{hi}}(x) \text{ tal que } \mathcal{H}_\phi^{\text{hi}}(x) \subseteq Z\}.$$

Assim, dado um conjunto $Z \in 2^W$, a função $\Phi_W(Z)$ retorna o conjunto de macroestados $x \in Z$ para os quais existe no mínimo uma entrada de controle $\phi(x) \in \mathcal{A}_S^{\text{hi}}(x)$ que faz com que o estado x permaneça em Z no passo seguinte. O conjunto parcialmente ordenado $(2^W, \preceq)$ forma um reticulado completo, com a ordem parcial ' \preceq ' dada pela operação de continência de conjuntos. Adicionalmente, Φ_W é um operador que preserva a ordem induzida por ' \preceq ', pois $\Phi_W(Z_1) \preceq \Phi_W(Z_2)$ sempre que $Z_1 \preceq Z_2$. Logo, o conjunto de pontos fixos de Φ_W tem um ponto fixo maior de acordo com o Teorema de Knaster-Tarski (TARSKI, 1955), ponto fixo que denotaremos $\mu(\Phi_W)$.

Definição 8. A função de alcançabilidade estrita, $\Theta_P : 2^{X_P} \rightarrow 2^{X_P}$, é definida como $\Theta_P(Z) = \{x \in X_P \mid \exists \phi(x) \in \mathcal{A}_S^{\text{hi}}(x) \text{ tal que } \mathcal{H}_\phi^{\text{hi}}(x) \subseteq Z\}$.

Ou seja, a função Θ_P aplicada a $Z \in 2^{X_P}$ retorna o conjunto de macroestados que, além de satisfazerem P , têm imagem contida em Z .

Definição 9. *O predicado de alcançabilidade estrita associado ao autômato G^{hi} e ao predicado P , representado por $SR^P(G^{\text{hi}}) : X \rightarrow \{0, 1\}$, é definido como verdadeiro precisamente nos macroestados x que satisfazem $x \in T^*$, em que T^* é o ponto fixo mínimo de Θ_P que contém o conjunto $X_m^* = \mu(\Phi_W)$, para $W = X_m$.*

Esse predicado contém, assim, todos os macroestados $x \in X_P$ a partir dos quais é possível garantir que existem entradas de controle que os levem a um subconjunto controlável de X_m , com trajetórias que também satisfazem P .

Proposição 1. *O problema de alcançabilidade estrita possui solução se, e somente se,*

1. $X_m^* \neq \emptyset$; e
2. $x_0 \models SR^P(G^{\text{hi}})$.

Demonstração. (se):

Se $X_m^* \neq \emptyset$, então, pela Definição 7, é verdade que, para todo $x \in X_m^*$, existe uma entrada de controle $\phi(x)$ tal que $\mathcal{H}_\phi^{\text{hi}}(x) \subseteq X_m^*$. Escolha, para cada $x \in X_m^*$, uma dessas entradas ϕ e faça $S^{\text{hi}}(x) = \phi(x)$.

O ponto fixo mínimo de Θ_P que contém X_m^* , T^* , pode ser obtido a partir da seguinte iteração:

- $T_0 = X_m^*$;
- $T_{n+1} = \Theta_P(T_n)$, $n \geq 0$;
- $T^* = T_k$, com k tal que $T_{n+1} = T_n$, para $n \geq k$.

Por construção, $T_n \subseteq X_P$, para $n \geq 0$. Se $x_0 \models SR^P(G^{\text{hi}})$, então $x_0 \in T_s$ para algum (possivelmente mais do que um) $0 \leq s \leq k$. Tome o menor de tais s . Então:

- $Y_0 = \{x_0\} \subseteq T_s$;
- $Y_i = \bigcup_{x \in Y_{i-1}} \mathcal{H}_\phi^{\text{hi}}(x)$, $1 \leq i \leq s$, com $\phi(x)$ escolhida de tal forma que $\mathcal{H}_\phi^{\text{hi}}(x) \subseteq T_{j-1}$ para $x \in T_j$. A existência de tais $\phi(x)$ é garantida pela Definição 8.

Por construção, $Y_i \subseteq X_P$, para $0 \leq i \leq s$, e $Y_s \subseteq X_m^*$. Os conjuntos Y_i assim definidos satisfazem as condições (i)-(iii) do problema de alcançabilidade estrita para $l = s$ e $Y^* = X_m^*$.

(somente se):

Assuma agora que existe um supervisor S^{hi} tal que o problema de alcançabilidade estrita possua solução. Assim, pela condição (iii) do enunciado do problema de alcançabilidade e pela definição de X_m^* , chegamos à conclusão de que $X_m^* \neq \emptyset$, pois X_m^* contém Y^* e, por construção, $Y^* \neq \emptyset$.

Usando a definição da função Θ_P e o fato de que ela preserva ordem, obtemos $Y_{l-1} \subseteq \Theta_P(Y_l) \subseteq \Theta_P(X_m^*)$. Iterando essa desigualdade sucessivamente, chegamos a $Y_0 \subseteq \Theta_P(Y_1) \subseteq \Theta_P^l(X_m^*)$. Mas $\Theta_P^l(X_m^*) \subseteq T^*$ e, portanto, os elementos de Y_0 satisfazem o predicado de alcançabilidade estrita associado a G^{hi} e ao predicado P . Assim, $x_0 \models SR^P(G^{\text{hi}})$. □

5.3.1 Exemplo: Parte I

Para o exemplo enunciado no início deste Capítulo, considere $X_m = \{101\}$ e $P = \text{verdadeiro}$. Nesse caso, $X_m^* = X_m = \{101\}$, dado que $\phi(101) = \{\}$ é uma entrada de controle admissível que mantém o sistema no macroestado 101. O cálculo do ponto fixo mínimo T^* de Θ_P que contém X_m^* faz-se de maneira iterativa, resultando na seguinte sequência de conjuntos:

$$T_0 = \{101\}$$

$$T_1 = \{101, 100, 111, 001\}$$

$$T_2 = \{101, 100, 111, 001, 000, 110, 011\}$$

$$T_3 = T^* = \{101, 100, 111, 001, 000, 110, 011, 010\}.$$

O macroestado inicial 010 $\models SR^P(G^{\text{hi}})$ e, por isso, pela Proposição 1, é possível encontrar uma sequência de entradas de controle que conduzem a rede de seu estado inicial 010 ao estado final 101. De fato, podemos verificar por inspeção direta da Figura 25(b) que existem diversas soluções para o problema. Uma delas é dada por:

$S^{\text{hi}}(010) = \{+u_1\}$, $S^{\text{hi}}(110) = \{+u_3\}$, $S^{\text{hi}}(111) = \{-u_2\}$ e, finalmente, $S^{\text{hi}}(101) = \{\}$.

Embora a solução para esse problema seja facilmente determinada sem o emprego de qualquer método formal, o mesmo não seria verdadeiro caso o autômato G^{hi} tivesse centenas ou milhares de estados.

5.3.2 Escolha de um Supervisor

A Proposição 1 fornece as condições necessárias e suficientes para a existência de supervisores que resolvem o problema de controle considerado nesta Tese. Ela não provê, no entanto, um supervisor em particular.

Na demonstração da Proposição 1, uma estratégia específica para construção de supervisores foi adotada. Ela baseia-se na ideia de escolher, para um dado $x \in T_k$, uma entrada de controle $\phi(x)$ tal que $\mathcal{H}_\phi^{\text{hi}}(x) \subseteq T_{k-1}$. Segundo essa abordagem, portanto, a cada passo o sistema fica mais próximo de alcançar o conjunto X_m^* . Essa estratégia minimiza, por conseguinte, o número de iterações necessárias para atingir os estados-alvo a partir de x_0 , mas diferentes métricas poderiam mostrar-se mais úteis em outras aplicações.

Veremos, por exemplo, no próximo capítulo, como obter supervisores que, não somente resolvem o problema de alcançabilidade estrita, mas o fazem minimizando a energia consumida pelo sistema em malha fechada. Esse critério é importante para garantir uma operação adequada do sistema de controle ao longo de várias gerações, pois circuitos celulares supérfluos que drenam muita energia são mais suscetíveis a mutações deletérias.

Poder-se-ia pensar, ainda, em um supervisor tal que sua realização com genes sintéticos se desse com a utilização do menor número possível de genes. Quanto maiores os trechos de DNA inseridos na célula a ser controlada, maior a carga molecular a ser replicada pelo organismo e, assim, maior o custo energético imposto pelo supervisor.

Um outro aspecto que poderia ser considerado é a robustez do sistema de controle em malha fechada. Como incertezas (estruturais e paramétricas) são intrínsecas aos modelos matemáticos de sistemas biológicos, a robustez de um controlador é fator chave para determinar se a solução encontrada apresentará o desempenho desejado em aplicações reais.

5.4 REALIZAÇÃO DO SUPERVISOR

5.4.1 Conversão de $S^{\text{hi}}(x)$ em f_i^s

Os controles admissíveis $S^{\text{hi}}(x) \in \mathcal{A}_S^{\text{hi}}(x)$ dos tipos (1) e (2) baseiam-se em duas intervenções elementares:

- forçar um evento de alto nível $\pi \in \Sigma_f^{\text{hi}}(x)$ em $x \in X$, ou
- desabilitar um evento de alto nível $\sigma \in \Sigma_c^{\text{hi}}$ em $x \in X$.

Somente eventos do tipo $+u_i$ podem ser forçados em um dado macroestado x , e essa ação está *sempre* associada a forçar $+v_i^s$, e, portanto, a ativar v_i^s , em *todos* aqueles microestados de G^{au} contidos em x (veja as Figuras 30(a) e 30(b)). Assim, para implementar tal ação de controle biologicamente, exige-se que o gene sintético tenha uma região de regulação tal que sua transcrição se inicie toda vez que a célula estiver no estado x .

Por outro lado, tanto eventos do tipo $+u_i$ quanto eventos do tipo $-u_i$ podem ser desabilitados, contanto que sejam controláveis. Se o evento a ser desabilitado é do tipo $+u_i \in \Sigma_{cf}^{\text{hi}}(x)$, então $+v_i^s$ deve ser desabilitado nos microestados contidos em x (veja Figura 30(b)). Isso equivale, biologicamente, a desativar v_i^s em todos os microestados pertencentes a x . Lembre-se de que a ocorrência de eventos controláveis e forçáveis no alto nível deve-se exclusivamente à ativação dos genes sintéticos no baixo nível.

Se o evento de alto nível a ser desabilitado em x é do tipo $-u_i \in \Sigma_{c\bar{f}}^{\text{hi}}$, pode-se implementar tal ação no baixo nível ao forçar $+v_i^s$ para preemptar $-v_i$ e/ou desabilitar $-v_i^s$ (Figura 29(b)). Biologicamente, evita-se a diminuição na concentração de um dado fator de transcrição $-u_i$ em x pela ativação do gene sintético v_i^s em x .

A Figura 31 mostra como a ativação e a desativação de genes sintéticos v_i^s nos microestados que compõem um macroestado x implementam ações no baixo nível que desabilitam ou habilitam, por exemplo, eventos controláveis $-u_i$ do alto nível. Análise similar àquela ilustrada pela Figura 31 pode ser conduzida para os casos em que se deseja: a) forçar eventos de alto nível $+u_i$; e b) desabilitar eventos de alto nível do tipo $+u_i$.

Resumidamente, portanto, toda entrada de controle $S^{\text{hi}}(x)$ pode ser implementada por ações no baixo nível que decorrem de uma adequada *ativação* ou *desativação*, em *todos* os microestados pertencentes a x , de cada um dos genes sintéticos $v_j^s \in V^s$ (outras implementações de

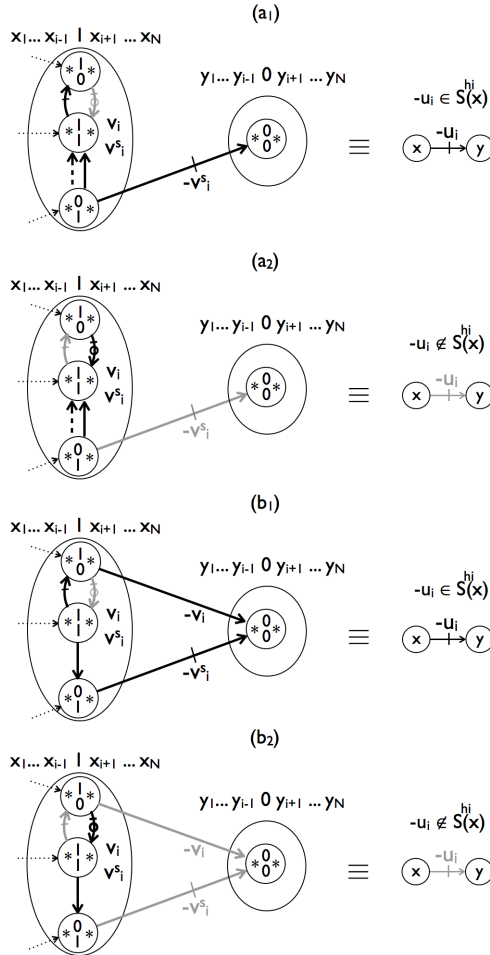


Figura 31 – A ativação e a desativação de genes sintéticos implementam biologicamente as ações de baixo nível que permitem a desativação de eventos do alto nível do tipo $-u_i$. Setas cinzas correspondem a eventos desabilitados ou preemptados, enquanto setas pretas, a eventos habilitados. (a₁) Variante da Figura 29(a) quando $-u_i \in S^{hi}(x)$. O gene sintético v_i^s é ativado pelos fatores de transcrição sintetizados em x . (a₂) Variante da Figura 29(a) quando $-u_i \in S^{hi}(x)$. O gene sintético v_i^s é ativado. (b₁) Variante da Figura 29(b) quando $-u_i \in S^{hi}(x)$. O gene sintético v_i^s é desativado. (b₂) Variante da Figura 29(b) quando $-u_i \in S^{hi}(x)$. O gene sintético v_i^s é desativado.

$S^{\text{hi}}(x)$ no baixo nível seriam possíveis, mas essas alternativas exigiriam a capacidade de distinguir entre dois microestados quaisquer $q, r \in x$, o que, biologicamente, é mais difícil de se implementar, dado que q e r sintetizam os mesmos fatores de transcrição). Logo, a nossa implementação de $S^{\text{hi}}(x)$, de uma perspectiva biológica, requer que os genes sintéticos reconheçam conjuntos de fatores de transcrição, os macroestados, disparando ações correspondentes, quais sejam, a síntese de seus fatores de transcrição.

Com tal implementação biológica, os microestados de G^{au} que são alcançados a partir de x pertencem garantidamente a $\mathcal{H}_\phi^{\text{hi}}(x)$, um tópico que exploraremos de maneira mais aprofundada e formal na seção dedicada à Consistência Hierárquica.

5.4.2 Exemplo: Parte II

Aplicamos, agora, as ideias desenvolvidas na subseção anterior, a fim de determinar o padrão de ativação dos genes sintéticos v_1^s e v_3^s para o exemplo tratado neste capítulo. As ações de controle $S^{\text{hi}}(x)$ a serem aqui utilizadas são aquelas obtidas na Subseção 5.3.1.

Para forçar $+u_1$ em 010, deve-se forçar $+v_1^s$ e, portanto, ativar v_1^s , em todos aqueles microestados contidos em 010. Adicionalmente, desabilitar $-u_1$ em 111 e 101, como requerido por $S^{\text{hi}}(111)$ e $S^{\text{hi}}(101)$, respectivamente, também é realizado biologicamente pela ativação de v_1^s no contexto proteico dado pelos macroestados 111 e 101. O mesmo raciocínio aplica-se para decidir pela ativação de v_3^s nos macroestados 110, 111 e 101.

A definição dos padrões de ativação/desativação de v_1^s e v_3^s como função dos macroestados x é equivalente à definição das funções de atualização f_1^s e f_3^s que, por sua vez, refletem as características das *regiões de regulação* de v_1^s e v_3^s , pois tais padrões especificam exatamente *quando* os genes devem ser ativados ou desativados. Para o exemplo, e considerando as informações do parágrafo anterior, podemos escrever, matematicamente, $f_1^s(u_1, u_2, u_3) = (\bar{u}_1 \wedge u_2 \wedge \bar{u}_3) \vee (u_1 \wedge u_2 \wedge u_3) \vee (u_1 \wedge \bar{u}_2 \wedge u_3)$ e $f_3^s(u_1, u_2, u_3) = (u_1 \wedge u_2 \wedge \bar{u}_3) \vee (u_1 \wedge u_2 \wedge u_3) \vee (u_1 \wedge \bar{u}_2 \wedge u_3)$, em que: u_i é a variável booleana que indica se a concentração de u_i está acima ($u_i = 1$) ou abaixo ($u_i = 0$) de um dado valor de limiar; ‘ \wedge ’ e ‘ \vee ’ denotam os operadores lógicos E (AND) e OU (OR), respectivamente; e \bar{u}_i indica a negação lógica da variável u_i . Assim, f_1^s reflete o fato de que o gene sintético v_1^s deve possuir uma região de regulação tal que sua ativação só se dê nos macroestados 010, 111 ou 101.

No próximo capítulo, mostraremos em detalhes como simplificar as expressões obtidas para as funções f_i^s , de maneira que suas implementações posteriores como regiões de regulação gênica sejam mais factíveis. Essa ideia, que será aqui brevemente aplicada ao exemplo, baseia-se na escolha de entradas de controle $S^{\text{hi}}(y)$ para aqueles macroestados y que *não* pertencem à trajetória em malha fechada do sistema. Tal $S^{\text{hi}}(y)$ é definida somente com o propósito de simplificar as funções f_i^s .

Vejamos um caso concreto. No exemplo aqui tratado, a trajetória em malha fechada visita somente os macroestados 010, 110, 111 e 101. Assim, se fizermos $S^{\text{hi}}(000) = \{+u_1\}$, não alteramos o comportamento em malha fechada do sistema, pois 000 não é um estado alcançável a partir de x_0 . Mas a imposição de tal entrada de controle implica a ativação de v_1^s no macroestado caracterizado pela ausência dos fatores de transcrição u_1 , u_2 e u_3 , resultando em uma nova expressão para f_1^s , que chamaremos h_1^s , igual a $h_1^s(u_1, u_2, u_3) = f_1^s \vee (\bar{u}_1 \wedge \bar{u}_2 \wedge \bar{u}_3) = (\bar{u}_1 \wedge \bar{u}_3) \vee (u_1 \wedge u_3)$. Note que h_1^s contém duas cláusulas com duas variáveis Booleanas cada, enquanto f_1^s possui três variáveis Booleanas em cada cláusula. Explicaremos melhor na próxima subseção por que h_1^s é mais simples de ser implementada do que f_1^s . Assumamos, por ora, que isso é verdadeiro.

A função de atualização f_3^s também pode ser simplificada, por exemplo, fazendo $S^{\text{hi}}(100) = \{+u_3\}$, que implica ativar v_3^s em 100, o que reduz a expressão de f_3^s a $h_3^s(u_1, u_2, u_3) = f_3^s \vee (u_1 \wedge \bar{u}_2 \wedge \bar{u}_3) = u_1$.

5.4.3 Projeto de Genes Sintéticos

5.4.3.1 Hipótese de Preemptabilidade

O processo de expressão gênica é complexo e pode ser regulado em diversos pontos, tanto em bactérias como em células humanas (ALBERTS et al., 2008). A Figura 32 ilustra esse conceito, mostrando os possíveis pontos de regulação da expressão gênica em eucariotos. Até o momento, exploramos nesta Tese somente o controle transcricional, que, embora seja o mais importante para a maioria dos genes, não é o único (ALBERTS et al., 2008). Além disso, mesmo regiões de regulação gênica que induzem o início da transcrição nas mesmas condições celulares, ou seja, na presença dos mesmos fatores de transcrição, fazem-no com dinâmicas e intensidades distintas, resultando em diferentes taxas de síntese e concentração final das moléculas de mRNA (BLAZECK et

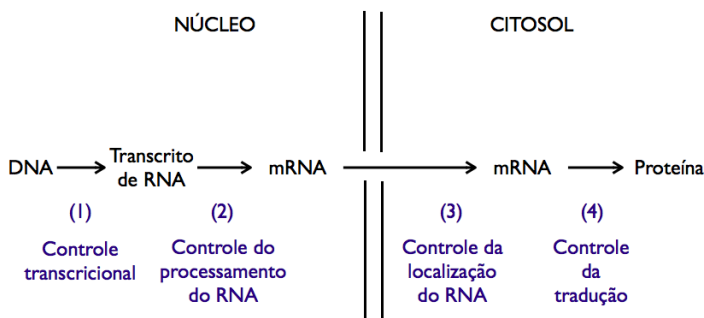


Figura 32 – A expressão gênica pode ser controlada em diversos pontos. Na Tese, concentramo-nos no controle transcricional, indicado na figura pela etapa (1). Em células eucarióticas, a diversidade de passos passíveis de regulação atinge o seu ápice, como ilustrado na figura, adaptada do livro de Alberts et al (ALBERTS et al., 2008). A figura mostra, por exemplo, que a síntese de proteína a partir do mRNA pode ser controlada pela célula na etapa (4). Assim, mesmo que um dado gene esteja ativado, a proteína por ele codificada pode não ser produzida.

al., 2012).

Etapas que ocorrem após finalizada a transcrição também impactam a síntese proteica, dando origem aos chamados controles *pós-transcricionais*. RNAs mensageiros, por exemplo, possuem uma região denominada região não traduzida (UTR, do inglês *UnTranslated Region*), que desempenha um papel importante no controle da tradução. Essas regiões acoplam-se a proteínas, pequenos metabólitos, ou mesmo a outras moléculas de RNA, que, por exemplo, expõem ou bloqueiam as sequências do mRNA que se ligam ao ribossomo, influenciando, assim, o início da tradução (ALBERTS et al., 2008; CULLER; HOFF; SMOLKE, 2010).

A interação de regiões não traduzidas de mRNAs com outras moléculas não constitui o único mecanismo de regulação da tradução. Para compreender um outro mecanismo interessante, lembremos que, como descrito no Capítulo 2, diferentes códons são traduzidos para o mesmo aminoácido. Assim, dois mRNAs com sequências distintas, por exemplo, podem originar a mesma proteína. Embora o produto final

seja o mesmo, a conformação da molécula de mRNA varia em cada um dos casos, o que implica:

- a) diferentes graus de estabilidade das moléculas de mRNA à degradação;
- b) distintas afinidades do mRNA com o ribossomo e, por conseguinte, variada eficiência no início da tradução; e
- c) diferentes taxas de síntese da cadeia proteica (etapa chamada de alongação), seja pela maior ou menor dificuldade do ribossomo em “deslizar” pelo mRNA para “lê-lo”, seja pela necessidade de utilização de tRNAs distintos para adicionar um aminoácido à cadeia de proteína nascente. Em bactérias *E. coli*, por exemplo, códons correspondentes a tRNAs abundantes na célula são traduzidos seis vezes mais rápido do que códons associados a tRNAs que aparecem em concentrações menores (GINGOLD; PILPEL, 2011).

Do exposto, concluímos que dois RNAs mensageiros que codificam para a mesma proteína podem ser regulados de maneiras distintas, apresentar taxas de degradação diversas e, além disso, ser traduzidos em diferentes taxas, dando origem a perfis variados de aumento da concentração proteica na célula (ALBERTS et al., 2008; GINGOLD; PILPEL, 2011).

Todas as variáveis que alteram a dinâmica e a eficiência dos processos de transcrição e tradução constituem possíveis pontos de intervenção da Biologia Sintética no projeto de moléculas que tenham uma *dinâmica mais rápida* ao longo das etapas de síntese proteica, validando, assim, nossa hipótese de preemptabilidade dos eventos associados aos genes sintéticos, como representado esquematicamente na Figura 33. No capítulo dedicado à construção de abstrações discretas a partir de sistemas modelados por equações diferenciais, abordaremos, por meio de um exemplo, o tema da preempção de uma perspectiva mais quantitativa.

5.4.3.2 Hipótese de Programação da Região de Regulação

Trataremos agora da construção de genes sintéticos ativados por combinações lógicas de fatores de transcrição u_i , como $u_1 \wedge u_2$, $u_1 \vee u_2$, $u_1 \wedge \overline{u_2}$, e assim por diante (lembrando que $u_i = 1$ denota que a concentração de u_i é maior do que um dado valor de limiar). Genes sintéticos com tais propriedades implementam as funções “se <estado>,”

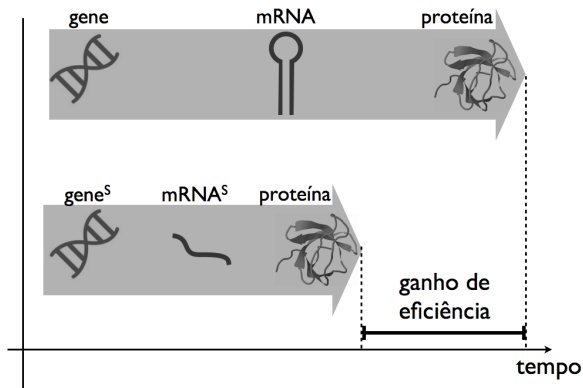


Figura 33 – A ativação dos genes sintéticos pode preempitar eventos da planta, porque é possível projetar genes sintéticos cuja dinâmica da expressão gênica é mais rápida do que aquela relacionada aos genes da planta.

então <ação>”, requeridas para realização biológica do supervisor S^{hi} que resolve o problema de controle.

Um resultado teórico baseado em modelos termodinâmicos para redes de regulação gênica bacterianas mostrou que regiões de regulação podem ser de fato programadas para implementar diferentes funções Booleanas (BUCHLER; GERLAND; HWA, 2003), exatamente o que requer a nossa abordagem. Assim, a computação de funções lógicas por regiões de regulação gênica é não somente uma realidade já implementada nos mais diversos seres vivos ¹, como também pode ser projetada *de novo* em moléculas sintéticas por meio da manipulação de parâmetros biológicos cujos efeitos no início da transcrição gênica são cada vez mais bem compreendidos. Discorreremos mais ao longo dos próximos parágrafos especificamente sobre a ideia publicada por Buchler e colegas (BUCHLER; GERLAND; HWA, 2003), dada a sua importância para compreensão da etapa de realização dos supervisores que requeremos em nossa abordagem.

¹Um exemplo fascinante é dado pelo controle da expressão do gene *Even-skipped* da *Drosophila*, em que mais de 20 mil pares de nucleotídeos são empregados para programar funções lógicas com mais de 20 entradas distintas, dentre elas as concentrações dos fatores de transcrição *Giant*, *Hunchback*, *Bicoid* e *Krüppel* (ALBERTS et al., 2008).

Como vimos no Capítulo 2, a síntese de moléculas de mRNA a partir da informação codificada nos genes é catalisada pela enzima RNA polimerase (RNAP), que se acopla a regiões específicas denominadas *promotores* para dar início à transcrição do gene. Embora haja significativas diferenças entre a transcrição em procariotos e eucariotos, a afinidade da ligação entre a RNAP e a região promotora é, em ambos os casos, um fator crucial para o início da transcrição. É exatamente por facilitar ou dificultar a ligação entre a RNAP e a região promotora que os fatores de transcrição alteram a probabilidade de início da síntese de mRNAs. Por exemplo, quando o sítio de ligação ao DNA de um fator de transcrição sobrepõe-se à região promotora, esse fator de transcrição, sempre que acoplado ao DNA, inibe a ligação da RNAP ao gene, por exclusão direta desta. Por outro lado, se o fator de transcrição, quando ligado ao DNA, facilita o posicionamento da RNAP na região promotora, fornecendo algumas poucas interações favoráveis, a probabilidade de início da transcrição aumenta (ALBERTS et al., 2008). As mesmas proteínas de regulação podem desempenhar o papel de ativadoras ou de repressoras transcricionais, dependendo da localização dos seus sítios de ligação ao DNA (ALBERTS et al., 2008).

O estudo de Buchler e coautores (BUCHLER; GERLAND; HWA, 2003) quantificou o impacto de dois parâmetros na probabilidade de ligação da RNAP à região promotora de um gene, a saber:

- a) a intensidade do acoplamento entre um dado fator de transcrição u_i e o seu sítio de ligação ao DNA, denotado aqui por O_{u_i} ; e
- b) a disposição relativa dos sítios de ligação O_{u_i} ao longo da cadeia de DNA.

Nessa pesquisa, os autores consideraram que a intensidade de acoplamento proteína-DNA pode ser continuamente ajustada, por alteração da sequência de nucleotídeos presente na região de regulação gênica, e que as proteínas possuem uma interação fraca quando colocadas em contato direto. Tal interação fraca entre proteínas é observada tanto em bactérias quanto em eucariotos, sendo, assim, uma hipótese plausível (ALBERTS et al., 2008). Os autores da pesquisa descreveram esses ingredientes como um *kit* de Lego molecular, capaz de ser programado de formas diferentes para implementar diversas funções lógicas (BUCHLER; GERLAND; HWA, 2003). A Figura 34 ilustra alguns exemplos, baseados no artigo citado, de como diferentes configurações de uma dada região de regulação gênica, em termos dos parâmetros (a) e (b) descritos acima, impactam a transcrição gênica.

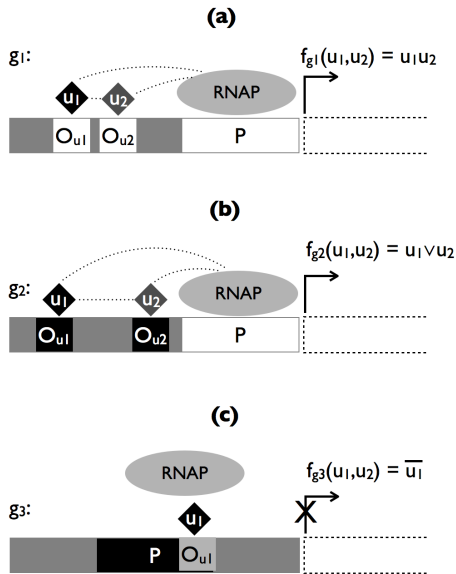


Figura 34 – A localização, bem como a intensidade de interação do sítio de ligação de um fator de transcrição ao DNA afeta a função lógica implementada pela região de regulação. Em (a) e (b) os genes estão sendo transcritos, enquanto em (c) o gene está desativado. Caixas brancas, cinzas e pretas representam sítios de ligação ao DNA com intensidades fraca, moderada e forte, respectivamente. Figura parcialmente reproduzida do artigo de Buchler e coautores (BUCHLER; GERLAND; HWA, 2003).

Por exemplo, sítios de ligação fracos O_{u_1} e O_{u_2} posicionados um próximo ao outro podem ser empregados para construção da função lógica E (que, nesse caso, assume o valor lógico 1 se, e somente se, a concentração de ambos os fatores de transcrição u_1 e u_2 estiver acima de um dado valor de limiar, ou seja, $u_1 = u_2 = 1$). Isso ocorre porque os fatores de transcrição u_1 e u_2 , isoladamente, não conseguem ligar-se ao DNA (dado que a afinidade entre u_i e seus sítios de ligação é fraca). Na presença de ambos, no entanto, a interação cooperativa entre eles facilita uma ligação mais estável com o DNA (Figura 34). Conforme descrito por Alberts et al. (ALBERTS et al., 2008), “interações proteína-proteína que são muito fracas para promover a associação das proteínas em solução podem levar as proteínas a associarem-se ao DNA, com a sequência de DNA atuando como um sítio de cristalização, ou como uma semente para a montagem de um complexo proteico”. Uma vez estavelmente ligados ao DNA, essas proteínas de regulação, conjunta e cooperativamente, facilitam o posicionamento da RNAP na região promotora.

Por outro lado, quando os sítios de ligação ao DNA possuem altas afinidades com seus fatores de transcrição, a interação cooperativa adicional entre as proteínas de regulação não se faz necessária para que haja uma ligação estável dos fatores de transcrição ao DNA, resultando na função lógica OU, $u_1 \vee u_2$ (Figura 34).

Ainda, se um fator de transcrição acopla-se a uma região que se sobrepõe à região promotora, essa proteína de regulação impede a ligação da RNAP ao DNA, inibindo a transcrição, e implementando o NÃO lógico, \bar{u}_1 (Figura 34). A ação repressora por exclusão direta da RNAP, muito comum em bactérias, é somente uma das muitas estratégias empregadas pelos seres vivos para impedir o início da transcrição. Em células eucarióticas, diversos outros mecanismos são utilizados para o mesmo propósito. Alguns deles envolvem interação direta da proteína repressora com a ativadora transcricional (p.e. o repressor pode ligar-se à proteína ativadora, impedindo-a de desempenhar seu papel), enquanto outros intervêm na própria estrutura de empacotamento da molécula de DNA, a cromatina, dificultando a transcrição (ALBERTS et al., 2008).

5.4.4 Exemplo: Parte III

Os princípios delineados até aqui para programação de regiões de regulação de genes sintéticos podem ser aplicados para finalizar o

projeto qualitativo dos genes v_1^s e v_3^s de nosso exemplo, cujas funções de atualização são, respectivamente, $h_1^s(u_1, u_2, u_3) = (u_1 \wedge u_3) \vee (\bar{u}_1 \wedge \bar{u}_3)$ e $h_3^s(u_1, u_2, u_3) = u_1$. Uma representação esquemática das regiões de regulação que implementam tais h_i^s é dada pela Figura 35.

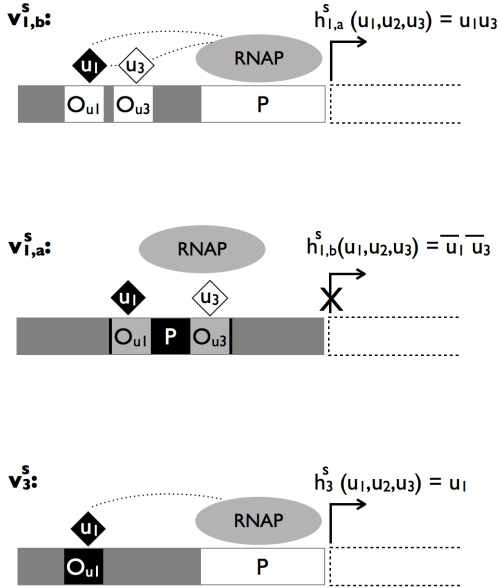


Figura 35 – Descrição qualitativa das regiões de regulação que implementam h_1^s , dividida em duas funções, e h_3^s .

Para obter a função lógica OU requerida por h_1^s , dividimos v_1^s em dois genes, $v_{1,a}^s$ e $v_{1,b}^s$, cada um implementando uma cláusula de h_1^s . A função de atualização $h_{1,a}^s$ pode ser programada em $v_{1,a}^s$ por meio da utilização de dois sítios de ligação O_{u_1} e O_{u_2} com baixa afinidade por u_1 e u_2 , respectivamente, e de um promotor cuja interação com a RNAP é fraca. Nessa situação, uma transcrição eficiente ocorreria somente na presença de ambas as proteínas de regulação gênica u_1 e u_3 , conforme o argumento já apresentado anteriormente. Note que $h_{1,a}^s$ é igual a f_1 , a função de atualização do gene nativo v_1 (o que não implica que os genes v_1^s e v_1 possuam exatamente a mesma sequência de nucleotídeos, já que, por causa da hipótese de preemptabilidade, estas podem diferir).

A função de atualização $h_{1,b}^s$ poderia ser implementada por um promotor com uma forte afinidade pela RNAP e dois sítios de ligação

moderados O_{u_1} e O_{u_3} que se sobrepõem à região promotora. Nessa situação, a transcrição teria início somente na ausência de u_1 e u_3 .

Finalmente, a região de regulação de v_3^s poderia ser obtida pela utilização de um sítio de ligação O_{u_1} com uma forte afinidade por u_1 e pelo emprego de uma região promotora com baixa afinidade pelo promotor de v_3^s (Figura 35).

É importante ressaltarmos que nossa análise qualitativa, e simplificada, para projeto de regiões de regulação fundamenta-se em estudos quantitativos descritos no artigo mencionado (BUCHLER; GERLAND; HWA, 2003). Não é objetivo desta Tese discorrer em detalhes sobre o projeto de genes sintéticos; pretendemos, tão somente, argumentar em favor da coerência e plausibilidade das nossas hipóteses de programação da região de regulação gênica e preempção de eventos da planta.

5.5 CONSISTÊNCIA HIERÁRQUICA

Argumentamos, na Subseção 5.4.1, que as ações de controle no alto nível $S^{\text{hi}}(x)$ podem ser traduzidas para ações no baixo nível (e.g. forçar eventos $+v_i^s$, ou desabilitar eventos $+v_i^s$ e $-v_i^s$) que, por sua vez, formam o ponto de partida para definir o padrão de ativação e desativação dos genes sintéticos v_i^s em cada um dos macroestados x .

Nesta seção, formalizaremos essa relação entre o alto e o baixo nível. Nosso intuito é demonstrar que, a cada entrada de controle do alto nível $S^{\text{hi}}(x)$, pode-se associar um conjunto válido de ações de controle no baixo nível $S^{\text{au}}(q)$, $q \in x$, que implementam $S^{\text{hi}}(x)$. Essa noção está ligada ao conceito de *Consistência Hierárquica Fraca* dentro da Teoria de Controle Hierárquico (ZHONG; WONHAM, 1990).

Começemos pela definição das ações de controle admissíveis $S^{\text{au}}(q)$, $q \in Q^{\text{au}}$:

- 1) $S^{\text{au}}(q) = \lambda \cup \Sigma_{\bar{c}\bar{f}\bar{p}}$, em que $\lambda \in 2^{\Sigma_{ef}}$, e $\lambda \neq \emptyset$;
- 2) $S^{\text{au}}(q) = \lambda \cup \Sigma_{\bar{c}\bar{f}}$, em que $\lambda \in 2^{\Sigma_{cf}}$.

O conjunto de todas as entradas admissíveis em um dado microestado q é representado por $\mathcal{A}_S^{\text{au}}(q)$. A entrada $S^{\text{au}}(q) \in \mathcal{A}_S^{\text{au}}(q)$, aplicada a um microestado $q \in Q^{\text{au}}$ de G^{au} , permite a transição a microestados $r \in Q^{\text{au}}$ que satisfazem $r = \delta^{\text{au}}(q, \sigma)$, para todo $\sigma \in S^{\text{au}}(q)$. Denotaremos o conjunto de tais r por $\mathcal{H}_\psi^{\text{au}}(q)$, com $\psi = S^{\text{au}}(q)$.

Dados os autômatos G^{hi} e G^{au} , dizemos que a entrada de controle $S^{\text{hi}}(x)$ e os controles do baixo nível $S^{\text{au}}(q)$, $q \in x$, satisfazem a

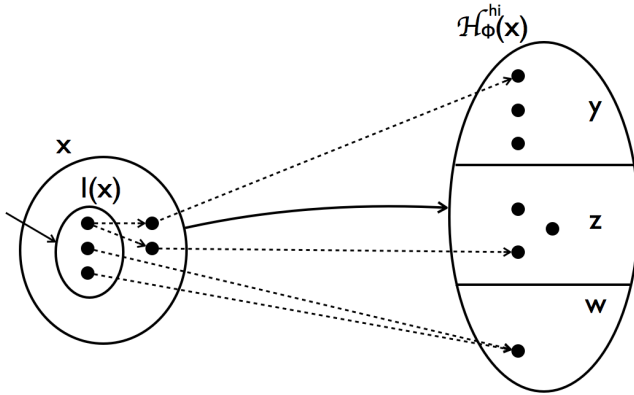


Figura 36 – Representação pictórica das condições de Consistência Hierárquica Fraca dentro do contexto de nossa Tese. Na figura, y , z e w representam macroestados pertencentes à imagem de x , quando a entrada de controle $\phi = S^{\text{hi}}(x)$ é aplicada. Os pontos representam os microestados de G^{au} .

propriedade de Consistência Hierárquica Fraca se, e somente se (ver Figura 36):

- a) as trajetórias induzidas por $S^{\text{au}}(q)$, $q \in x$, a partir de microestados em $I(x)$, conduzem exclusivamente a microestados $s \in \mathcal{H}_\phi^{\text{hi}}(x)$, em que apenas microestados de x são visitados ao longo do percurso até s ;
- b) para todo macroestado $y \in \mathcal{H}_\phi^{\text{hi}}(x)$, as ações de controle $S^{\text{au}}(q)$, $q \in x$, garantem a existência de pelo menos uma trajetória que inicia em algum microestado de $I(x)$ e termina em um microestado $s \in y$, visitando no percurso de $I(x)$ até s somente microestados de x .

A condição de Consistência Hierárquica Fraca descrita no parágrafo anterior é formalizada da seguinte maneira:

Definição 10. *Dados os autômatos G^{hi} e G^{au} , dizemos que a entrada de controle $\phi(x) = S^{\text{hi}}(x)$, para $x \in X$, e os controles do baixo nível $\psi(q) = S^{\text{au}}(q)$, $q \in x$, satisfazem a propriedade de Consistência Hierárquica Fraca se, e somente se, existe uma sequência finita de conjuntos*

de microestados $\{Q_i\}_{i=1}^p$, com $Q_1 = I(x)$, $Q_{i+1} = \bigcup_{q \in Q_i \setminus \mathcal{H}_\phi^{\text{hi}}(x)} \mathcal{H}_\psi^{\text{au}}(q)$, $1 \leq i \leq p-1$, em que:

a) $Q_i \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$, $1 \leq i \leq p-1$, $Q_p \subseteq \mathcal{H}_\phi^{\text{hi}}(x)$;

b) $\forall y \in \mathcal{H}_\phi^{\text{hi}}(x)$, $Q_j \cap y \neq \emptyset$, para alguma $1 \leq j \leq p$.

Note que, na definição acima, os elementos de Q_i pertencentes à $\mathcal{H}_\phi^{\text{hi}}(x)$ não são usados para computar Q_{i+1} .

Cabe-nos, agora, mostrar que os controles de baixo nível $S^{\text{au}}(q)$ qualitativamente descritos na Subseção 5.4.1 satisfazem, juntamente com as entradas de alto nível $S^{\text{hi}}(x)$, as propriedades de Consistência Hierárquica Fraca enunciadas na Definição 10, o que faremos nas próximas duas subseções.

5.5.1 Entradas de Controle $S^{\text{au}}(q)$ para $S^{\text{hi}}(x) = \{\pi_f\} \cup \Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$, com $\pi_f \in \Sigma_f^{\text{hi}}(x)$

Considere uma entrada de controle $S^{\text{hi}}(x)$ do tipo $\phi(x) = S^{\text{hi}}(x) = \{\pi_f\} \cup \Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$, para $\pi_f \in \Sigma_f^{\text{hi}}(x)$. Essa entrada força o evento de alto nível π_f em x e habilita os eventos não preemptáveis $\Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$.

Note que todo evento de alto nível π_f forçável em x , pela definição do conjunto $\Sigma_f^{\text{hi}}(x)$, é gerado por um evento do baixo nível $\sigma_f \in \Sigma_{\bar{c}\bar{f}}^{\text{au}}$. Se $y = \delta^{\text{hi}}(x, \pi_f)$, tal σ_f satisfaz sempre $\delta^{\text{au}}(q, \sigma_f) \in y$ para todo $q \in x$. Nas discussões que seguem, utilizaremos π_f e σ_f para representar precisamente eventos do alto e baixo níveis que guardam essa relação entre si.

Dito isso, dado $\phi(x)$ do tipo (1), considere as seguintes entradas de controle $S^{\text{au}}(q)$ para os microestados $q \in x$:

$$S^{\text{au}}(q) = \{\sigma_f\} \cup (\Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}})^q, \quad (5.1)$$

em que V^q , para $V \in 2^{\Sigma^{\text{au}}}$, denota $V \cap \Gamma^{\text{au}}(q)$, com $\Gamma^{\text{au}}(q)$ contendo os eventos do baixo nível σ para os quais $\delta^{\text{au}}(q, \sigma)$ está definida. Simplicadamente, então, essa entrada de controle $S^{\text{au}}(q)$: a) força o evento σ_f em todos os microestados $q \in x$, gerando, no alto nível, o evento π_f ; e b) habilita todos os eventos não preemptáveis que podem ocorrer em q . Essas ações, como já vimos, são implementadas biologicamente pela ativação de um gene sintético no macroestado x .

Vamos demonstrar formalmente que o controle $S^{\text{au}}(q)$ da Equação 5.1 e a entrada $S^{\text{hi}}(x) = \{\pi_f\} \cup \Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$ satisfazem as propriedades

de consistência enunciadas na Definição 10. Para isso, considere a seguinte Proposição.

Proposição 2. *Sejam $x \in X$ um macroestado de G^{hi} , $\phi(x) = \{\pi_f\} \cup \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{hi}}(x)$ e $\psi(q) = \{\sigma_f\} \cup (\Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}})^q$ as entradas de controle do alto e baixo níveis, respectivamente. A sequência de conjuntos $\{Q_i\}_{i=1}^{|\mathcal{J}|+2}$, definida abaixo, satisfaz os itens (a) e (b) da Definição 10.*

1. $Q_1 = I(x)$;
2. $Q_{i+1} = \bigcup_{q \in Q_i \setminus \mathcal{H}_\phi^{\text{hi}}(x)} \mathcal{H}_\psi^{\text{au}}(q)$, $1 \leq i \leq |\mathcal{J}| + 1$, em que \mathcal{J} denota o conjunto dos j tais que $x_j = 1$.

Para a demonstração da Proposição 2, considere os resultados a seguir.

Lema 1. *Seja $x \in X$ um macroestado qualquer de G^{hi} e $\mathcal{J} \subseteq \{1, \dots, N\}$ o conjunto de índices j tais que $x_j = 1$. A maior cadeia $\gamma \in (\Sigma_{\bar{c}}^{\text{au}})^*$ de eventos unicamente não controláveis que pode ser gerada por G^{au} a partir de $q \in I(x)$, com $\delta_e^{\text{au}}(q, \alpha \leq \gamma) \in x$, possui tamanho limitado a $|\mathcal{J}|$.*

Demonstração. Dado $x \in X$, seja $\gamma \in (\Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}})^*$ uma cadeia não controlável gerada por G^{au} a partir de algum microestado $q \in I(x)$, com $\delta_e^{\text{au}}(q, \alpha \leq \gamma) \in x$. Considere, ainda, os seguintes fatos:

1. Somente eventos $+v_i$ ou $-v_i$, $1 \leq i \leq N$, são não controláveis (os eventos $+v_i^s$ e $-v_i^s$, $i \in \mathcal{S}$, são sempre controláveis).
2. Dado um macroestado $x \in X$, o valor de $f_i^e(\mathbf{v}, \mathbf{v}^s)$ é constante para todo $q \in x$. Assim, ou v_i é ativado ($+v_i$), ou desativado ($-v_i$) em x , nunca ambos;
3. $\forall k \in \{1, \dots, N\} \setminus \mathcal{J}$ e $\forall q \in x$:
 - $-v_k \notin \Gamma^{\text{au}}(q)$, pois $x_k = v_k^s \vee v_k = 0$, o que implica $v_k = 0$ (ou seja, v_k já está desativado em $q \in x$);
 - Se $+v_k \in \Gamma^{\text{au}}(q)$, então $r = \delta^{\text{au}}(q, +v_k) \notin x$, pois $v_k^s \vee v_k = 1$ em r .

De (1) e (3), inferimos que γ só pode ser formada pelos eventos $\sigma \in \{+v_j, -v_j\}_{j \in \mathcal{J}}$. Pelo item (2), deduzimos que, para um mesmo $j \in \mathcal{J}$, a cadeia γ não pode conter $+v_j$ e $-v_j$ simultaneamente. Logo, $|\gamma| \leq |\mathcal{J}|$.

□

Lema 2. *Sejam $x, y \in X$, $x \neq y$, dois macroestados de G^{hi} e $\phi(x) \in \mathcal{A}_S^{\text{hi}}(x)$ um controle admissível em x , ou seja, $\Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x) \subseteq \phi(x)$. Se $\exists q \in x$ tal que $\delta^{\text{au}}(q, \sigma) \in y$, para algum $\sigma \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}}$, então $y \in \mathcal{H}_\phi^{\text{hi}}(x)$.*

Demonstração.

1. Se $\exists q \in x$ tal que $\delta^{\text{au}}(q, \sigma) \in y$ para algum $\sigma \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}}$, então $q \in I(x)$, como mostram as Figuras 29 e 30;
2. Se $\delta^{\text{au}}(q, \sigma) \in y$ para $q \in I(x)$, então, pela definição de $\Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$, $\exists \pi \in \Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$ com $y = \delta^{\text{hi}}(x, \pi)$. Logo $y \in \mathcal{H}_\phi^{\text{hi}}(x)$ para o controle $\phi(x)$ considerado. □

Lema 3. *Para a sequência $\{Q_i\}_{i=1}^{|\mathcal{J}|+2}$ da Proposição 2, se $Q_i \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$, então $Q_{i+1} \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$, $i \geq 1$.*

Demonstração.

1. $Q_{i+1} = \bigcup_{q \in Q_i \setminus \mathcal{H}_\phi^{\text{hi}}(x)} \mathcal{H}_\psi^{\text{au}}(q) \subseteq \bigcup_{q \in x} \mathcal{H}_\psi^{\text{au}}(q)$, pela hipótese sobre Q_i ;
2. $Q_{i+1} \subseteq \mathcal{R} = \{r \mid r = \delta^{\text{au}}(q, \sigma) \text{ para } q \in x \text{ e } \sigma \in \{\sigma_f\} \cup (\Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}})^q\}$;
3. Para $\sigma = \sigma_f$, $r = \delta^{\text{au}}(q, \sigma_f) \in \delta^{\text{hi}}(x, \pi_f) \subseteq \mathcal{H}_\phi^{\text{hi}}(x)$;
4. Para $\sigma \in (\Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}})^q$:
 - ou $r = \delta^{\text{au}}(q, \sigma) \in x$; ou
 - $r = \delta^{\text{au}}(q, \sigma) \notin x$. Neste caso, pelo Lema 2, $r \in y \subseteq \mathcal{H}_\phi^{\text{hi}}(x)$.

Logo, $\mathcal{R} \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$ e, portanto, $Q_{i+1} \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$. □

Passemos agora, finalmente, à demonstração da Proposição 2.

Demonstração.

Parte 1.

1. $Q_1 = I(x) \subseteq x \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$. Logo, pelo Lema 3, $Q_i \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$ para $i \geq 1$. Portanto, os conjuntos Q_i , $1 \leq i \leq |\mathcal{J}|+1$, satisfazem a parte (a) da Definição 10.
2. $Q_{|\mathcal{J}|+2} = \bigcup_{r \in \mathcal{R}} \mathcal{H}_\psi^{\text{au}}(r)$ com $\mathcal{R} = Q_{|\mathcal{J}|+1} \setminus \mathcal{H}_\phi^{\text{hi}}(x) \subseteq x$ (pelo item anterior);

3. O conjunto \mathcal{R} é formado por microestados $r \in x$ tais que $r = \delta_e^{\text{au}}(q, \gamma)$, em que $q \in Q_1 = I(x)$ e $\gamma = \gamma_1 \dots \gamma_{|\mathcal{J}|}$ é uma cadeia de eventos tal que:

- $\delta_e^{\text{au}}(q, \beta \leq \gamma) \in x$;
- $\gamma_i \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}}$, $1 \leq i \leq |\mathcal{J}|$. Isso é verdade, pois se $\gamma_j = +v_i^s$ para algum j , então $\delta_e^{\text{au}}(q, \gamma_1 \dots \gamma_j) \notin x$.

4. Mas, pelo Lema 1, a maior cadeia $\alpha \in (\Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}})^*$ gerada a partir de $I(x)$, e que só visita microestados de x , tem tamanho $|\mathcal{J}|$, ou seja, o mesmo tamanho de γ . Assim, $\forall r \in \mathcal{R}$:

- Se $\sigma \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}} \cap \Gamma^{\text{au}}(r)$, $s = \delta^{\text{au}}(r, \sigma) \notin x$, pois $|\gamma\sigma| = |\mathcal{J}| + 1$. Pelo Lema 2, $s \in \mathcal{H}_\phi^{\text{hi}}(x)$;
- Se $\sigma = \sigma_f$, $s \in \delta^{\text{au}}(r, +v_i^s) \in \mathcal{H}_\phi^{\text{hi}}(x)$;
- Logo, $\mathcal{H}_\psi^{\text{au}}(r) \subseteq \mathcal{H}_\phi^{\text{hi}}(x)$.

Assim, $\{Q_i\}_{i=1}^{|\mathcal{J}|+2}$ satisfaz o item (a) da Definição 10.

Parte 2.

- Seja $y = \delta^{\text{hi}}(x, \pi_f)$. É verdade que para todo $q \in x$, em particular para $q \in I(x)$, $\delta^{\text{au}}(q, \sigma_f) \in y$. Logo, $Q_2 \cap y \neq \emptyset$.
- Seja $\pi \in \Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$ tal que $w = \delta^{\text{hi}}(x, \pi)$. Pela definição de $\Sigma_{\bar{c}\bar{p}}^{\text{hi}}(x)$, é verdade que $\exists q \in I(x)$ satisfazendo $\delta^{\text{au}}(q, \sigma) \in w$ para $\sigma \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}}$. Logo, $Q_2 \cap w \neq \emptyset$. A sequência $\{Q_i\}_{i=1}^{|\mathcal{J}|+2}$ satisfaz o item (b) da Definição 10.

□

5.5.2 Entradas de Controle $S^{\text{au}}(q)$ para $S^{\text{hi}}(x) = \lambda \cup \Sigma_{\bar{c}}^{\text{hi}}(x)$, com $\lambda \in 2^{\Sigma_{\bar{c}\bar{f}}^{\text{hi}}(x)}$

A segunda entrada de controle admissível é do tipo $S^{\text{hi}}(x) = \lambda \cup \Sigma_{\bar{c}}^{\text{hi}}(x)$, com $\lambda \in 2^{\Sigma_{\bar{c}\bar{f}}^{\text{hi}}(x)}$. Formalizaremos, nesta subseção, os controles do baixo nível $S^{\text{au}}(q)$, para $q \in x$, correspondentes a esse $S^{\text{hi}}(x)$.

Lembremos, antes, que todos os eventos $\pi \in \Sigma_{\bar{c}\bar{f}}^{\text{hi}}(x)$ são do tipo $-u_i$, ou seja, denotam uma queda na concentração do fator de transição u_i abaixo de um valor de limiar (a Figura 29 mostra as situações

em que tal evento ocorre). É importante ressaltarmos que os eventos de alto nível $-u_i \in \Sigma_{c\bar{f}}^{\text{hi}}(x)$ podem ser desabilitados pela ativação do gene sintético v_i^s , o que, no baixo nível, corresponde a entradas de controle em $q \in x$ que: a) forçam $+v_i^s$, preemptando, assim, a possível ocorrência de $-v_i$; ou b) desabilitam $-v_i^s$.

Particionemos, então, esses eventos $-u_i \in \Sigma_{c\bar{f}}^{\text{hi}}(x)$ em dois subconjuntos:

- $U_h^-(x) = \{-u_i \in \Sigma_{c\bar{f}}^{\text{hi}}(x) \mid -u_i \in \lambda \subseteq S^{\text{hi}}(x)\}$, o conjunto dos eventos controláveis $-u_i$ habilitados em x ;
- $U_d^-(x) = \{-u_i \in \Sigma_{c\bar{f}}^{\text{hi}}(x) \mid -u_i \notin \lambda \subseteq S^{\text{hi}}(x)\}$, o conjunto dos eventos controláveis $-u_i$ desabilitados em x .

Para facilitar referências futuras, armazenaremos em \mathcal{K} e \mathcal{L} os índices k e l , respectivamente, tais que $-u_k \in U_h^-(x)$ e $-u_l \in U_d^-(x)$. Note que $(\mathcal{K} \cup \mathcal{L}) \subseteq \mathcal{J}$, com \mathcal{J} sendo o conjunto dos índices j para os quais $x_j = 1$.

Por construção, o conjunto $\mathcal{M} = \mathcal{J} \setminus (\mathcal{K} \cup \mathcal{L})$ contém os índices m associados a eventos $-u_m$ que: a) não pertencem a $\Sigma^{\text{hi}}(x)$; ou b) são não controláveis em x e, portanto, pertencem a $S^{\text{hi}}(x)$, mas não a λ . No primeiro caso, forçosamente, $v_m^s \notin V^s$, ou seja, o gene sintético v_m^s não é empregado para fins de controle (caso contrário, existiria um microestado $q \in I(x)$ satisfazendo $v_m^s = 1$ e $v_m = 0$, com $\delta^{\text{au}}(q, -v_m^s) \notin x$ e, portanto, $-u_m \in \Sigma^{\text{hi}}(x)$, segundo ilustra a Figura 29). É verdade também, nesse caso, que $-v_m \notin \Gamma^{\text{au}}(q)$ para nenhum $q \in x$. O caso (b), conforme se depreende da Figura 29, só é possível se $-v_m$ é não preemptável ou se $v_m^s \notin V^s$. Particionaremos o conjunto \mathcal{M} em dois subconjuntos, \mathcal{M}_s e $\mathcal{M}_{\bar{s}}$, de modo que \mathcal{M}_s contenha os índices $m \in \mathcal{M}$ para os quais $v_m^s \in V^s$, e $\mathcal{M}_{\bar{s}}$, aqueles $m \in \mathcal{M}$ tais que $v_m^s \notin V^s$.

Isso posto, visando determinar $S^{\text{au}}(q)$, particionaremos o conjunto de microestados $q \in x$ em dois subconjuntos, definidos a seguir:

- $R_0(x)$, composto por microestados $q \in x$ para os quais $\exists l \in \mathcal{L}$ satisfazendo $v_l^s = 0$;
- $R_1(x)$, composto por todos os outros microestados $q \in x$.

Definamos, então, o controle de baixo nível $S^{\text{au}}(q)$, para $q \in x$, da seguinte maneira:

$$S^{\text{au}}(q) = \begin{cases} (V_{s,l}^+ \cup \Sigma_{c\bar{f}\bar{p}}^{\text{au}})^q, & q \in R_0(x) \\ (V_{s,k}^- \cup V_{s,m}^- \cup \Sigma_{c\bar{f}}^{\text{au}})^q, & q \in R_1(x) \end{cases} \quad (5.2)$$

em que V^q , com $V \in 2^{\Sigma^{\text{au}}}$, é uma notação simplificada para $V \cap \Gamma^{\text{au}}(q)$, $V_{s,l}^+ = \bigcup_{l \in \mathcal{L}} \{+v_l^s\}$, $V_{s,k}^- = \bigcup_{k \in \mathcal{K}} \{-v_k^s\}$ e $V_{s,m}^- = \bigcup_{m \in \mathcal{M}_s} \{-v_m^s\}$. Perceba que os eventos $+v_l^s$ são forçados em $R_0(x)$, enquanto os $-v_l^s$ são desabilitados em $R_1(x)$. Essas ações de controle implementam a ativação do gene sintético v_l^s em todos os microestados de x (ou, equivalentemente, nas condições proteicas que caracterizam x), evitando a ocorrência de $-u_l \notin \lambda$ no alto nível. Para as entradas em $R_0(x)$, é necessário ainda incluir todos os eventos de baixo nível não preemptáveis, para garantir a admissibilidade do controle. Já em $q \in R_1(x)$, como nenhum evento é forçado por $S^{\text{au}}(q)$, torna-se obrigatório habilitar todos os eventos não controláveis possíveis em q , ou seja, $\sigma \in (\Sigma_{\bar{e}\bar{f}})^q$. Por fim, note que os eventos de alto nível $-u_m$, $m \in \mathcal{M}_s$, são não controláveis e, assim, pertencem a $S^{\text{hi}}(x)$. Por isso, a ocorrência dos eventos de baixo nível $-v_m^s$, $m \in \mathcal{M}_s$, não infringe nenhuma restrição imposta pelo supervisor de alto nível $S^{\text{hi}}(x)$, resultando em $-v_m^s \in S^{\text{au}}(q)$ quando $q \in R_1(x)$.

Enunciaremos, agora, a Proposição que versa sobre a consistência do supervisor $S^{\text{au}}(q)$ definido na Equação 5.2.

Proposição 3. *Sejam $x \in X$ um macroestado de G^{hi} , $\phi(x) = \lambda \cup \Sigma_{\bar{e}}^{\text{hi}}(x)$, para $\lambda \in 2^{\Sigma_{\bar{e}\bar{f}}^{\text{hi}}(x)}$, e $\psi(q)$, definida segundo a Equação 5.2, as entradas de controle do alto e baixo níveis, respectivamente. A sequência de conjuntos $\{Q_i\}_{i=1}^n$, descrita abaixo, satisfaz os itens (a) e (b) da Definição 10, para um n finito.*

1. $Q_1 = I(x)$;
2. $Q_{i+1} = \bigcup_{q \in Q_i \setminus \mathcal{H}_{\phi}^{\text{hi}}(x)} \mathcal{H}_{\psi}^{\text{au}}(q)$, $1 \leq i \leq n - 1$.

A fim de demonstrar a Proposição 3, considere os lemas a seguir.

Lema 4. *Para a sequência $\{Q_i\}_{i=1}^n$ da Proposição 3, se $Q_i \subseteq x \cup \mathcal{H}_{\phi}^{\text{hi}}(x)$, então $Q_{i+1} \subseteq x \cup \mathcal{H}_{\phi}^{\text{hi}}(x)$, $i \geq 1$.*

A prova deste Lema segue exatamente o mesmo raciocínio empregado para demonstrar o Lema 3 e, por isso, não será aqui reproduzida.

Lema 5. *Para qualquer $q \in x$, a maior cadeia $\gamma \in (\Sigma^{\text{au}})^*$ gerada por G^{au} , sob o controle das entradas $S^{\text{au}}(q)$, que satisfaz $\delta^{\text{au}}(q, \alpha \leq \gamma) \in x$ tem tamanho finito.*

Demonstração. Seja γ a cadeia de eventos do enunciado do lema. Considere, então, os fatos a seguir:

- 1) As entradas de controle do baixo nível $S^{\text{au}}(q)$ só permitem a ocorrência de eventos σ pertencentes aos conjuntos $V_{s,l}^+$, $V_{s,k}^-$, $V_{s,m}^-$ e $\Sigma_{\bar{c}\bar{f}}^{\text{au}}$;
- 2) Os eventos $\sigma \in \Sigma_{\bar{c}\bar{f}}^{\text{au}}$ só podem ocorrer uma vez em x (lembre-se de que os genes sintéticos v_j são ou ativados, ou desativados para um mesmo x);
- 3) Para qualquer $l \in \mathcal{L}$, o evento $+v_l^s$ pode ocorrer no máximo uma única vez em x . Isso se dá porque, para que haja uma nova ativação do gene v_l^s , é necessária uma desativação de v_l^s . Mas $-v_l^s \notin S^{\text{au}}(q)$ para nenhum $q \in x$;
- 4) Para qualquer $k \in \mathcal{K}$ ($m \in \mathcal{M}_s$), o evento $-v_k^s$ ($-v_m^s$) pode ocorrer no máximo uma única vez em x . Isso se dá porque, para que haja uma nova desativação do gene v_k^s (v_m^s), é necessária uma ativação de v_k^s (v_m^s). Mas $+v_k^s \notin S^{\text{au}}(q)$ ($+v_m^s \notin S^{\text{au}}(q)$) para nenhum $q \in x$;
- 5) Dos itens anteriores, concluímos que nenhuma cadeia γ gerada no interior de x pode conter mais do que uma ocorrência de um dado evento σ habilitado por S^{au} . Como o conjunto de eventos habilitados por $S^{\text{au}}(q)$ é finito, a cadeia γ tem comprimento finito.

□

Lema 6. Para todo $q \in x$, se $q \in \mathcal{H}_{\psi}^{\text{au}}(q)$, para $\psi = S^{\text{au}}(q)$ definida na Equação 5.2, então $q \in \mathcal{H}_{\phi}^{\text{hi}}(x)$.

Demonstração.

- 1) Para todo $\sigma \in \Sigma^{\text{au}}$ e para qualquer $q \in Q^{\text{au}}$, se $\delta^{\text{au}}(q, \sigma)$ está definida, então $\delta^{\text{au}}(q, \sigma) = r \neq q$. Isso acontece, pois os eventos $\sigma \in \Sigma^{\text{au}}$ são *sempre* do tipo $+v_i$, $-v_i$, $+v_i^s$ e $-v_i^s$, com $1 \leq i \leq N$, em que N é o número de genes nativos da rede de regulação gênica. Mas a ocorrência de quaisquer desses eventos em q altera o estado de um dos genes $\{v_i^s, v_i\}$ e, assim, acarreta uma mudança de microestado.
- 2) Do item anterior, concluímos que $q \in \mathcal{H}_{\psi}^{\text{au}}(q)$ somente se $\psi(q) = \emptyset$. Mas:
 - a) se $q \in R_0(x)$, $\psi(q) \neq \emptyset$, pois, por definição, existe em q ao menos um $l \in \mathcal{L}$ tal que $v_l^s = 0$ em q . Nesse caso, $+v_l^s \in \psi(q)$;
 - b) se $q \in R_1(x)$, $\psi(q) = \emptyset$ implica $(V_{s,k}^-)^q = (V_{s,m}^-)^q = (\Sigma_{\bar{c}\bar{f}}^{\text{au}})^q = \emptyset$. Logo:

- $v_k^s = 0$ para todo $k \in \mathcal{K}$ e $-v_k \notin \Gamma^{\text{au}}(q)$. Nesse caso, os eventos de alto nível controláveis $-u_k \in \Sigma_{c\bar{f}\bar{D}}^{\text{hi}}(x)$ e, portanto, $\lambda \subseteq \Sigma_{c\bar{f}\bar{D}}^{\text{hi}}(x)$;
- $v_m^s = 0$ para todo $m \in \mathcal{M}_s$ e $-v_m \notin \Gamma^{\text{au}}(q)$. Mas tal situação nunca pode ocorrer para um evento de alto nível $-u_m \in \Sigma_{\bar{c}}^{\text{hi}}(x)$, conforme mostra a Figura 29(c). Assim, $\mathcal{M}_s = \emptyset$. Como $(\Sigma_{\bar{c}\bar{f}}^{\text{au}})^q = \emptyset$, $\mathcal{M}_{\bar{s}} = \emptyset$, pois se $-u_m \in \Sigma_{\bar{c}}^{\text{hi}}(x)$, $m \in \mathcal{M}_{\bar{s}}$, então, por definição, para todo $q \in x$, $-v_m \in \Gamma^{\text{au}}(q)$, segundo a Figura 29(d). Portanto, não existe $-u_i$ tal que $-u_i \in \Sigma_{\bar{c}}^{\text{hi}}(x)$;
- Não existe $+u_i$ tal que $+u_i \in \Sigma_{\bar{c}}^{\text{hi}}(x)$, pois, caso contrário, $+v_i \in \Sigma_{\bar{c}\bar{f}}^{\text{au}}$ seria possível em q e, assim, $\psi(q)$ seria diferente de vazio.

3) Do item (2), concluímos que $S^{\text{hi}}(x) = \lambda \subseteq \Sigma_{c\bar{f}\bar{D}}^{\text{hi}}(x)$, o que implica $x \in \mathcal{H}_{\phi}^{\text{hi}}(x)$. Como $q \in x$, obtemos $q \in \mathcal{H}_{\phi}^{\text{hi}}(x)$

□

Passemos agora, finalmente, à prova da Proposição 3.

Demonstração.

Parte 1.

- 1) $Q_1 = I(x) \subseteq x \cup \mathcal{H}_{\phi}(x)$. Pelo Lema 4, $Q_i \subseteq x \cup \mathcal{H}_{\phi}(x)$, para $i \geq 1$. Assim, Q_i satisfaz as condições do item (a) da Definição 10;
- 2) Seja p o comprimento da maior cadeia γ produzida em x a partir de um $q \in I(x) = Q_1$. Os microestados $r \in Q_{p+1}$ são tais que uma das duas condições abaixo são válidas:
 - a) r foi obtido a partir de $q \in I(x)$ por uma cadeia α de comprimento $|\alpha| = |\gamma| = p$.
 - b) r foi obtido a partir de $q \in I(x)$ por uma cadeia $\beta < \gamma$ e, nesse caso, r é tal que $r \in \mathcal{H}_{\psi}^{\text{au}}(r)$, o que implica $r \in \mathcal{H}_{\phi}^{\text{hi}}(x)$.
- 3) Por construção, os elementos de Q_{p+2} são formados pela imagem dos microestados r do item anterior, parte (a), quando sob controle da entrada $\psi(q)$. Assim:

- se $S^{\text{au}}(r) \neq \emptyset$, então $s = \delta^{\text{au}}(r, \sigma) \notin x$ para $\sigma \in S^{\text{au}}(r)$, por força do Lema 5. Mas como $s \in Q_{p+2} \subseteq x \cup \mathcal{H}_\phi^{\text{hi}}(x)$, então $s \in \mathcal{H}_\phi^{\text{hi}}(x)$;
- se $S^{\text{au}}(r) = \emptyset$, então $r \in \mathcal{H}_\psi^{\text{au}}(r)$ e, pelo Lema 6, $r \in \mathcal{H}_\phi^{\text{hi}}(x)$.

4) Assim, concluimos que $Q_{p+2} \subseteq \mathcal{H}_\phi^{\text{hi}}(x)$. Portanto, sequência $\{Q_i\}_{i=1}^{p+2}$ satisfaz os requisitos do item (a) da Definição 10.

Parte 2.

Seja $y = \delta^{\text{hi}}(x, \pi)$ para $\pi \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{hi}}(x)$. As Figuras 29 e 30 mostram que $\exists q \in I(x)$ tal que $\delta^{\text{au}}(q, \sigma) \in y$ para algum $\sigma \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}}$. Como tal σ é sempre habilitado por $S^{\text{au}}(q)$, $Q_2 \cap y \neq \emptyset$.

Considere, agora, $y = \delta^{\text{hi}}(x, \pi)$ para $\pi \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{hi}}(x)$. As figuras 29 e 30 também mostram que $\exists q \in I(x)$ tal que $\delta^{\text{au}}(q, \sigma) \in y$ para algum $\sigma \in \Sigma_{\bar{c}\bar{f}\bar{p}}^{\text{au}}$. Tal σ só é habilitado pelo supervisor S^{au} para $q \in R_1(x)$. Portanto, se $Q_1 = I(x)$ contém algum $r \in R_1(x)$, então $Q_2 \cap y \neq \emptyset$. Suponha, agora, que $I(x) \subseteq R_0(x)$. Tome um microestado q qualquer de $I(x)$. É verdade que $v_l^s = 0$, para $l \in \mathcal{L}^* \subseteq \mathcal{L}$. A partir de $q \in I(x)$, é possível alcançar um microestado $r \in R_1(x)$ pela cadeia $\gamma = (+v_{l_1}^s) \dots (+v_{l_i}^s)$, com $l_i \in \mathcal{L}^*$. Assim, $Q_{|\mathcal{L}^*|+2} \cap y \neq \emptyset$.

Por fim, seja $y = \delta^{\text{hi}}(x, \pi)$ para $\pi \in \lambda \subseteq S^{\text{hi}}(x)$, com $\lambda = 2^{\Sigma_{\bar{c}\bar{f}}^{\text{hi}}(x)}$. Lembremos que tais π são do tipo $-u_k$, com $k \in \mathcal{K}$. Além disso, $x_k = v_k^s \vee v_k = 1$ e $y_k = 0$. Suponha que $Q_1 = I(x)$ contenha um microestado $r \in R_1(x)$. O controle $S^{\text{au}}(r)$ permite a ocorrência dos eventos σ em $(V_{s,k}^- \cup V_{s,m}^- \cup \Sigma_{\bar{c}\bar{f}}^{\text{au}})$. Assim:

- se $v_k^s = 1$ e $v_k = 0$, $\delta^{\text{au}}(r, -v_k^s) \in y$. Portanto, $Q_2 \cap y \neq \emptyset$;
- se $v_k^s = 0$ e $v_k = 1$, $\delta^{\text{au}}(r, -v_k) \in y$. Portanto, $Q_2 \cap y \neq \emptyset$;
- se $v_k^s = v_k = 1$, o microestado $s = \delta_e^{\text{au}}(r, \alpha) \in y$, para $\alpha = (-v_k^s)(-v_k)$. Logo, $Q_3 \cap y \neq \emptyset$.

Suponha $Q_1 = I(x) \subseteq R_0(x)$. Tome um microestado q qualquer de $I(x)$. É verdade que $v_l^s = 0$, para $l \in \mathcal{L}^* \subseteq \mathcal{L}$. Assim, ou $Q_{|\mathcal{L}^*|+2} \cap y \neq \emptyset$ ou $Q_{|\mathcal{L}^*|+3} \cap y \neq \emptyset$. A sequência $\{Q_i\}_{i=1}^{p+2}$ satisfaz o item (b) da Definição 10, para p igual ao tamanho da maior cadeia de eventos que pode ser gerada no interior de x sob controle de $S^{\text{au}}(q)$, $q \in x$.

□

5.5.3 Ausência de Consistência Hierárquica Estrita

A ausência de uma solução no alto nível para o problema de atração de estados não implica a não existência de uma solução para o problema no baixo nível. Por solução no baixo nível, entenda-se um supervisor S^{au} que conduza q_0^{au} a microestados pertencentes a X_m^* em um número finito de passos, ao longo de uma trajetória que só visita microestados pertencentes a macroestados que satisfazem o predicado P .

Para exemplificar uma situação em que o problema possui solução no baixo nível, mesmo que não a possua no alto nível, considere duas modificações no exemplo tratado neste capítulo. A primeira é considerar $-v_1 \in \Sigma_{\text{cfp}}^{\text{au}}$; a segunda é fazer P um predicado válido em todos os macroestados de X , exceto em 011.

Essas duas modificações tornam o problema insolúvel no alto nível: nenhuma trajetória pode visitar 011 e, no caminho alternativo a 101, que passa por 110, o conjunto $\{111\}$ não é uma imagem de 110, devido à não preemptibilidade de $-u_1$. Mas perceba que, no autômato de baixo nível G^{au} da Figura 27, se v_1^s é forçado em q_0^{au} , o sistema entra no macroestado 110 por um microestado no qual a volta para 010 é *controlável*. Ou seja, a Figura 27 mostra que, quando o sistema parte de q_0^{au} , ao forçar $+v_1^s$ em $\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{smallmatrix}$, $+v_3^s$ em $\begin{smallmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{smallmatrix}$, e desabilitar $-v_1^s$ e $-v_3^s$ em $\begin{smallmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{smallmatrix}$ e $\begin{smallmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{smallmatrix}$, chega-se a microestados que pertencem a X_m^* .

Essa característica de nossa abordagem implica ausência da chamada Consistência Hierárquica Estrita. Para um sistema de controle hierárquico em que essa propriedade mais forte de consistência é satisfeita, uma solução do problema no alto nível existe se, e somente se, o problema no baixo nível é solúvel. A ausência de tal atributo em nossa metodologia deriva, basicamente, do conservadorismo em nossa definição de eventos do alto nível π não preemptáveis. Conforme se observa na Figura 29(c), um evento π é classificado como não preemptável em x , mesmo quando existem microestados $q \in I(x)$ para os quais a ocorrência de π pode ser desabilitada.

5.6 RESUMO

Este é o capítulo central da Tese, que expõe nossa metodologia para controlar redes de regulação gênica por meio de genes sintéticos (BALDISSERA; CURY, 2012; BALDISSERA; CURY; RAISCH, 2014). A estrutura do capítulo está representada esquematicamente na Figura 37,

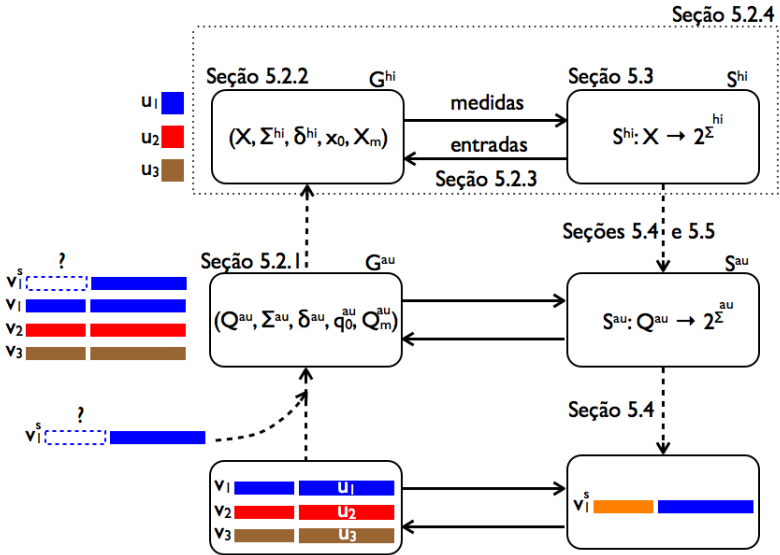


Figura 37 – Representação esquemática dos resultados apresentados no capítulo, segundo uma perspectiva baseada na Teoria de Controle Hierárquico. O autômato G^{au} modela o comportamento dos genes nativos (v_1, v_2 e v_3 na figura) mais os novos eventos possibilitados (e, por conseguinte, os novos estados alcançados) pela utilização de genes sintéticos para fins de controle (na figura, o gene v_1^s). O autômato de alto nível G^{hi} é, então, obtido a partir de G^{au} . Em G^{hi} , os estados guardam a informação referente à presença ou ausência dos fatores de transcrição $u_1 = v_1^s + v_1, u_2 = v_2$ e $u_3 = v_3$. Uma vez obtido o supervisor S^{hi} , fazemos o caminho inverso, descendo aos níveis mais baixos, até chegar à implementação do supervisor. Note que somente após obtidos os supervisores é possível especificar a função de atualização de estados do gene sintético v_1^s, f_1^s , que dará origem à região de regulação de v_1^s .

segundo uma concepção fundada na Teoria de Controle Hierárquico.

Após uma breve descrição qualitativa de nossos propósitos na Seção 5.1, começamos a Seção 5.2 descrevendo a construção de um autômato de baixo nível G^{au} , que modela o comportamento da rede de regulação nativa mais as entradas de controle, i.e. as possíveis ativações

ou desativações dos genes sintéticos $v_i^s \in V^s$. Por admitirmos ser possível programar *quando* cada gene sintético é ativado, os eventos $+v_i^s$ e $-v_i^s$ foram considerados controláveis. Ademais, nossa hipótese de que os genes sintéticos podem ser construídos de modo que suas dinâmicas de expressão sejam mais rápidas do que aquelas associadas a alguns genes sintéticos trouxe ao autômato G^{au} a figura dos eventos forçáveis, capazes de preemptar eventos não controláveis. O autômato G^{au} resultante pode conter até $2^{N+|V^s|}$ estados, em que N é o número de genes nativos v_i e $|V^s|$ é o número de genes sintéticos v_i^s empregados para fins de controle. Isso representa um considerável aumento no número de estados do modelo, quando comparamos G^{au} com G , o autômato que modela a planta em malha aberta e que possui, no máximo, 2^N estados.

No entanto, certos estados de G^{au} são fisiologicamente equivalentes entre si, dado que descrevem estados formados pelos mesmos fatores de transcrição u_i . Com base nessa constatação, agregamos esses estados equivalentes de G^{au} para compor, na Seção 5.2.2, os novos estados x do autômato de alto nível G^{hi} . Os eventos de G^{hi} , bem como sua função de transição, também foram definidos.

Na Seção 5.2.4, formulamos o problema de controle denominado *atração de estados* para a situação em que o conjunto de eventos da planta Σ^{hi} é particionado em eventos controláveis ou não controláveis, forçáveis ou não forçáveis, preemptáveis ou não preemptáveis. Esse problema foi solucionado na Seção 5.3, com o auxílio do conceito de *predicado de alcançabilidade estrita*. Por solução, entenda-se um supervisor S^{hi} que associa a cada macroestado x uma ação de controle, que pode ter duas semânticas distintas: a) um evento é forçado, para evitar a ocorrência de um evento não controlável, mas preemptável; ou b) um conjunto de eventos controláveis é desabilitado.

Na Seção 5.4, mostramos, qualitativamente, como converter esses controles de alto nível em padrões de ativação e desativação de genes sintéticos v_i^s , ou seja, em funções de atualização de estados f_i^s . Ainda na Seção 5.4, descrevemos como essas funções f_i^s podem ser implementadas biologicamente, momento em que embasamos nossa hipótese de programação da região de regulação de genes sintéticos e de preemptabilidade de eventos da planta com resultados da literatura.

Finalmente, na Seção 5.5, formalizamos a conversão das ações de alto nível S^{hi} em diretivas do baixo nível S^{au} , provando que a conversão encontrada é tal que a propriedade de Consistência Hierárquica Fraca é satisfeita.

6 OTIMIZAÇÃO DO SUPERVISOR

Como mostramos no capítulo anterior, o controlador da dinâmica celular em nossa abordagem é implementado por um conjunto de genes sintéticos. Para aplicações reais, é desejável que tais genes sejam de fácil construção e induzam um comportamento em malha fechada com o menor consumo energético possível.

A simplicidade de *design* dos genes sintéticos traduz-se, em nosso caso, em regiões de regulação gênica simples. Trataremos desse tópico na Seção 6.1, objetivando modificar um dado supervisor S^{hi} que atende as especificações de malha fechada, para que ele seja realizado por genes sintéticos cujas funções de atualização f_i^s requeiram o menor número possível de operações lógicas (BALDISSERA; CURY, 2014a).

A Seção 6.2 abordará a temática de minimização da energia total consumida pelo sistema em malha fechada (BALDISSERA; CURY, 2014b). Economia energética é um tema de grande importância em sistemas de controle celulares, dado que circuitos sintéticos não vitais e com alto dispêndio de energia são mais suscetíveis a mutações deletérias.

6.1 SIMPLIFICAÇÃO DA REGIÃO DE REGULAÇÃO DE GENES SINTÉTICOS

As regiões de regulação dos genes sintéticos v_i^s estão diretamente relacionadas às funções de atualização $f_i^s(u_1, \dots, u_N)$, derivadas das ações de controle do alto nível $S^{\text{hi}}(x)$. As funções f_i^s são escritas na chamada *forma normal disjuntiva*, com $f_i^s = s_1 \vee \dots \vee s_m$, em que cada cláusula s_k representa a operação de conjunção lógica entre variáveis u_j ou \bar{u}_j , $1 \leq j \leq N$. Em nossa proposta de realização biológica das funções f_i^s , baseada no estudo de modelos quantitativos da expressão gênica em bactérias (BUCHLER; GERLAND; HWA, 2003):

- 1) a negação lógica da variável u_j (i.e. \bar{u}_j) é implementada por uma região de regulação tal que o sítio de ligação ao DNA O_{u_j} se sobrepõe à região promotora do gene sintético v_i^s ;
- 2) a conjunção lógica entre duas variáveis u_j e u_k (i.e. $u_j \wedge u_k$) é implementada por uma região de regulação em que os sítios de ligação O_{u_j} e O_{u_k} :
 - têm baixa afinidade por u_j e u_k , respectivamente; e

- são posicionados próximos uns aos outros, de modo que a ativação do gene v_i^s só seja possível pela ação sinérgica das proteínas u_j e u_k .
- 3) a disjunção lógica entre as cláusulas s_k é implementada pela “quebra” do gene sintético v_i^s em genes $v_{i,k}^s$, com cada $v_{i,k}^s$ implementando uma cláusula s_k de f_i^s .

Essa implementação traz consigo algumas restrições. Por exemplo, devido ao tamanho finito da região promotora, existe um limite prático para o número de sítios de ligação O_{u_i} que podem se sobrepor a ela. Isso implica que uma mesma cláusula s_k não pode conter um número grande de negações lógicas (o excesso de sítios de ligação coincidentes à região promotora é conhecido por *promoter overcrowding* (BUCHLER; GERLAND; HWA, 2003)). Esse problema poderia ser parcialmente mitigado pela exploração de um fenômeno conhecido como *formação de alças no DNA* (ALBERTS et al., 2008): um mesmo fator de transcrição liga-se a sítios diferentes, forçando dobras na molécula de DNA que impedem ou dificultam o acesso da RNA polimerase ao promotor. Nesse caso, os sítios de ligação O_{u_j} não precisam ser necessariamente aninhados na região promotora. Para facilitar, no entanto, o projeto de redes de regulação, optaremos pela via de *minimização* do número de negações lógicas empregadas por $f_i^s(u_1, \dots, u_N)$.

A implementação da conjunção lógica envolvendo muitos fatores de transcrição u_j também é problemática, já que exige um ajuste muito fino da posição de cada um dos sítios de ligação O_{u_j} , bem como de suas afinidades às proteínas u_j , de maneira que a transcrição só se inicie quando *todos* os fatores de transcrição u_j estiverem presentes. Por isso, é imperativo reduzir o número de variáveis em cada cláusula s_k de f_i^s .

Para reduzir o impacto das restrições enumeradas nos parágrafos anteriores, enunciamos o seguinte problema a ser resolvido:

Problema 2. *Dado um supervisor S^{hi} que atinge as especificações de malha fechada requeridas para a rede de regulação gênica modelada por G^{hi} , ache um novo supervisor S^* tal que:*

- a) $S^*(x) = S^{\text{hi}}(x)$, se $S^{\text{hi}}(x)$ está definida; e
- b) sua implementação como um conjunto de genes sintéticos v_i^s resulte em funções de atualização que, na forma normal disjuntiva, $f_i^s(u_1, \dots, u_N) = s_1 \vee \dots \vee s_m$, contenham o menor número de variáveis em cada cláusula s_k .

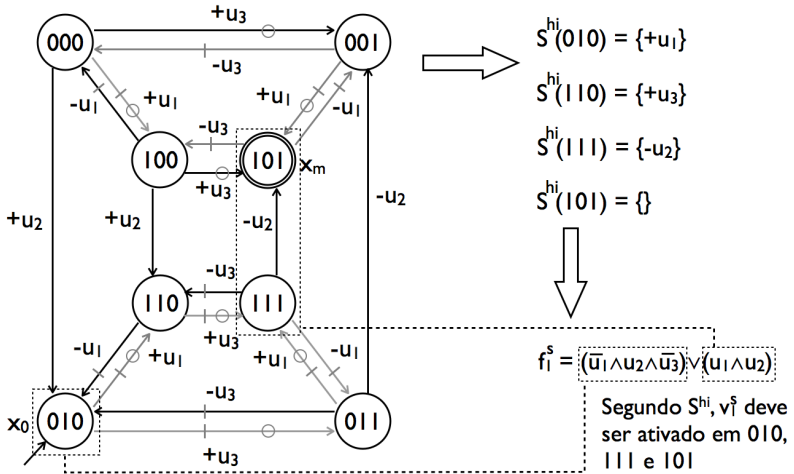


Figura 38 – Exemplo tratado no Capítulo 5, em que se deseja alcançar o macroestado 101 a partir de 010. O supervisor S^{hi} da figura induz a trajetória 010 – 110 – 111 – 101. A função de atualização de v_1^s é dada por f_1^s .

A imposição (a) decorre do fato de que não queremos alterar o comportamento do sistema em malha fechada. A especificação (b), por outro lado, *pode* reduzir o número de conjunções ou de negações lógicas dentro de uma mesma cláusula s_k de f_i^s .

Consideremos o exemplo introduzido no capítulo anterior e reproduzido na Figura 38. Suponha, em contraposição ao supervisor S^{hi} , um supervisor S^* , definido da seguinte forma:

- $S^*(x) = S^{hi}(x)$, para $x \in \{010, 110, 111, 101\}$;
- $S^*(000) = S^*(001) = S^*(011) = \{+u_1\}$.

A ativação adicional de v_1^s nos macroestados 000, 001 e 011, requerida por S^* , simplifica a expressão de $f_1^s(u_1, u_2, u_3) = (\bar{u}_1 \wedge u_2 \wedge \bar{u}_3) \vee (u_1 \wedge u_2 \wedge u_3) \vee (u_1 \wedge \bar{u}_2 \wedge u_3)$, resultando em uma nova função de atualização do estado de v_1^s , dada por $g_1^s(u_1, u_2, u_3) = (\bar{u}_1) \vee (u_1 \wedge u_3)$. Note que a nova expressão utiliza, no máximo, a conjunção de duas variáveis e a negação de somente uma em cada cláusula, um significativo avanço em termos de implementação molecular, quando comparada à função f_1^s .

Além da simplificação do projeto da região de regulação de v_1^s , o novo supervisor S^* aumenta a robustez do comportamento em malha fechada do sistema. Suponha, por exemplo, que haja uma perturbação no macroestado inicial 010, levando o sistema a 011. Nesse caso, a ação de controle $S^*(011) = \{+u_1\}$ recoloca o sistema na trajetória inicialmente planejada, 010-110-111-101. A mesma situação não se verifica quando o supervisor S^{hi} é utilizado, pois como $S^{\text{hi}}(011)$ não está definida, existe a possibilidade de uma transição do macroestado 011 para o atrator local 001, fazendo com que o sistema acomode-se nesse macroestado indefinidamente. A simplificação da região de regulação, portanto, tem como “efeito-colateral”, nesse caso específico, a propriedade de satisfazer as restrições do problema de controle, mesmo face a perturbações não previstas inicialmente no modelo matemático.

Pode-se reduzir a complexidade de uma função Booleana f_i^s por meio da utilização dos mapas de Karnaugh (KARNAUGH, 1953). Considere o exemplo desta seção e suponha que queiramos simplificar a função f_1^s (ver Figura 39). Poderíamos para tal fim empregar um mapa de Karnaugh, construído para as variáveis Booleanas u_1 , u_2 e u_3 . Nesse caso, as entradas do mapa correspondentes a macroestados x para os quais $S^{\text{hi}}(x)$ é definida assumem valor lógico 1, se v_1^s deve ser ativado em x , e 0, caso contrário. Aquelas posições do mapa correspondentes a estados x para os quais $S^{\text{hi}}(x)$ não está definida (ou seja, estados não pertencentes à trajetória em malha fechada do sistema) são preenchidas com o símbolo ‘*’ (interpretado, nas regras do mapa de Karnaugh, como *don't-care-symbol*). Em seguida, podem-se utilizar as regras de manipulação do mapa para simplificar as expressões de f_i^s , substituindo o símbolo ‘*’ por 1, quando necessário. Para o nosso exemplo, a máxima simplificação é obtida quando todos os símbolos ‘*’ da coluna correspondente a $u_1 = 0$ assumem valor lógico 1.

Tal método de simplificação é comumente aplicado a funções Booleanas f_i^s com, no máximo, seis variáveis, limitando, dessa forma, a aplicabilidade do procedimento a redes de regulação gênica de tamanho moderado. Entretanto, outros algoritmos existentes na literatura para simplificação de funções Booleanas podem ser empregados, como o algoritmo de Quine-McCluskey (QUINE, 1952) e outros métodos computacionais nele baseados.

A ideia delineada acima pode ser traduzida para o pseudo-algoritmo a seguir, que se aplica a casos mais gerais.

- 1: A partir de S^{hi} , determine as funções $f_i^s(\mathbf{U})$ na forma disjuntiva normal, para $\mathbf{U} = [u_1 \cdots u_N]$;
- 2: Simplifique as funções Booleanas $f_i^s(\mathbf{U})$;

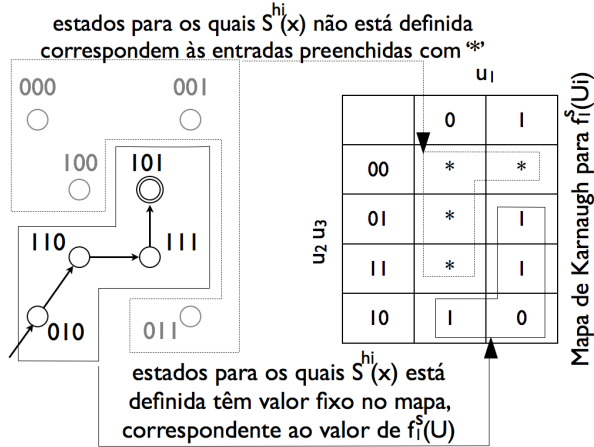


Figura 39 – Representação esquemática do procedimento para simplificar a função de atualização f_1^s por meio de mapas de Karnaugh.

- 3: **for all** $f_i^s(\mathbf{U})$ **do**
- 4: Construa um mapa de Karnaugh com as variáveis Booleanas presentes em f_i^s ;
- 5: **for all** x tal que $S^{\text{hi}}(x)$ é definida **do**
- 6: Preencha com o valor lógico 1 os campos do mapa que satisfazem $f_i^s(\mathbf{U}) = 1$;
- 7: Preencha os outros campos com o valor 0;
- 8: **end for**
- 9: Preencha todas os outros campos com “*” (*don't care symbol*);
- 10: Simplifique f_i^s usando as regras do mapa de Karnaugh com ‘*don't-care-symbols*’;
- 11: **end for**
- 12: Retorne f_i^s ;

6.2 MINIMIZAÇÃO DA ENERGIA CONSUMIDA

A síntese proteica é um empreendimento custoso para as células. Como cada macroestado x de G^{hi} é caracterizado pela síntese de um conjunto distinto de proteínas, cada x demanda uma quantidade de energia específica. Diferentes supervisores induzem caminhos diversos de x_0 a X_m e, assim, impõem *cargas metabólicas* variadas à célula na operação em malha fechada. Células dotadas de circuitos sintéticos

caros do ponto de vista energético, e que não são vitais para o organismo em questão, tendem a eliminá-los por meio de mutação, reduzindo, assim, a funcionalidade do circuito a longo prazo (SLEIGHT et al., 2010).

Nesta seção, consideraremos o problema de minimização da energia consumida no caminho que vai de x_0 a X_m , como uma forma de evitar a desativação do circuito sintético ao longo das gerações celulares. Para isso, algumas definições são necessárias.

Definição 11. *Um caminho W de G^{hi} sob ação do supervisor S^{hi} é definido como uma sequência de estados $W = \{x^k\}_{k=0}^i$, com $x^k \in X$, satisfazendo:*

$$i) \ x^{k+1} = \delta^{\text{hi}}(x^k, \pi_k), \text{ para } 0 \leq k \leq i - 1; \text{ e}$$

$$ii) \ \pi_k \in S^{\text{hi}}(x^k), \ 0 \leq k \leq i - 1.$$

O conjunto de todos os caminhos gerados por G^{hi} sob controle de S^{hi} é representado por $\mathcal{P}(G^{\text{hi}}, S^{\text{hi}})$.

Definição 12. *A função custo de síntese proteica $c_p : U \rightarrow \mathbb{N}$ associa a cada fator de transcrição $u_i \in U$ um número natural relacionado à quantidade de energia consumida no processo de expressão do gene que codifica u_i .*

Essa função depende de características específicas da proteína em questão, como o comprimento da cadeia de aminoácidos, por exemplo.

Definição 13. *A função de custo $c : X \rightarrow \mathbb{N}$ associa a cada macroestado x um número natural que consolida os custos para síntese dos fatores de transcrição presentes em x , ou seja,*

$$c(x) = \sum_{k=1}^{k=N} x_k \times c_p(u_k). \quad (6.1)$$

Os elementos x_k denotam o k -ésimo componente de x , e N é o número de fatores de transcrição presentes na rede de regulação gênica nativa. A função $c(x)$ pode também abarcar outras variáveis, como a síntese de proteínas cuja produção é induzida pelos fatores de transcrição presentes em x (lembre-se de que as redes de regulação gênica são capazes de controlar a expressão de centenas ou até milhares de outros genes). Cabe ressaltar, ainda, que o tempo de permanência do sistema em cada macroestado pode ser variável (algo que não é captado pelo modelo em autômatos finitos que utilizamos). Nesse caso, pode ser desejável quantificar essas diferenças e refleti-las na medida $c(x)$.

Definição 14. *O custo total associado ao caminho $W \in \mathcal{P}(G^{\text{hi}}, S^{\text{hi}})$ é dado por*

$$C(W) = \sum_{k=0}^l c(x^k), \quad (6.2)$$

para $W = x_0 \dots x_l$.

O problema de minimização da energia consumida no trajeto de x_0 a X_m pode ser, finalmente, definido como:

Problema 3. *Dado um problema de atração de estados para a planta G^{hi} , ache um supervisor S^* que: a) atenda às especificações de malha fechada, e b) minimize o máximo custo do caminho $W \in \mathcal{P}(G^{\text{hi}}, S)$ que começa em x_0 e termina em algum estado de X_m^* , ou seja, encontre:*

$$S^* = \arg \min_{S^{\text{hi}}} \left\{ \max_{W \in \mathcal{P}(G^{\text{hi}}, S^{\text{hi}})} C(W) \right\}, \quad (6.3)$$

para $W = x_0, \dots, x_l$, com $x_l \in X_m^*$ (o máximo conjunto controlável de X_m) e $x_j \notin X_m^*$ para $j \neq l$.

Antes de definirmos um algoritmo que resolva o Problema 3 para o caso geral, ilustraremos as ideias por meio de um exemplo. Considere, então, a rede de regulação gênica da Figura 40 e o seu respectivo modelo em autômato de estados finitos G^{hi} quando os genes sintéticos v_2^s e v_3^s são empregados para fins de controle.

A Figura 41 mostra os custos associados a cada estado x de G^{hi} , para uma instância do problema em que $c_p(u_1) = c_p(u_2) = c_p(u_3) = 1$, e aponta também os custos acumulados ao longo de duas trajetórias distintas de x_0 a X_m^* , uma delas visitando o estado de alto custo $x = 111$, em que os três fatores de transcrição são produzidos. É interessante notar como trajetórias caracterizadas pela ocorrência dos mesmos eventos possuem custos totais distintos. Tomemos os dois caminhos indicados na figura, (i) 100–110–010–011 e (ii) 100–101–111–011, por exemplo. Em ambos, dá-se a ocorrência dos eventos $-u_1$, $+u_2$ e $+u_3$, porém, em ordens diferentes, resultando em custos finais distintos (custo de 7 para a trajetória (ii), e 6 para a trajetória (i)).

O nosso algoritmo para resolução do Problema 3 é fortemente baseado naquele construído por Brave e Heymann (BRAVE; HEYMAN, 1993) e, por isso, não demonstraremos que o algoritmo a seguir resolve o problema posto. Introduzimos duas pequenas modificações em relação ao procedimento citado: a) a estrutura de controle que empregamos é mais rica, permitindo preempir eventos não controláveis; e b) os custos são associados a estados e não a ocorrência de eventos.

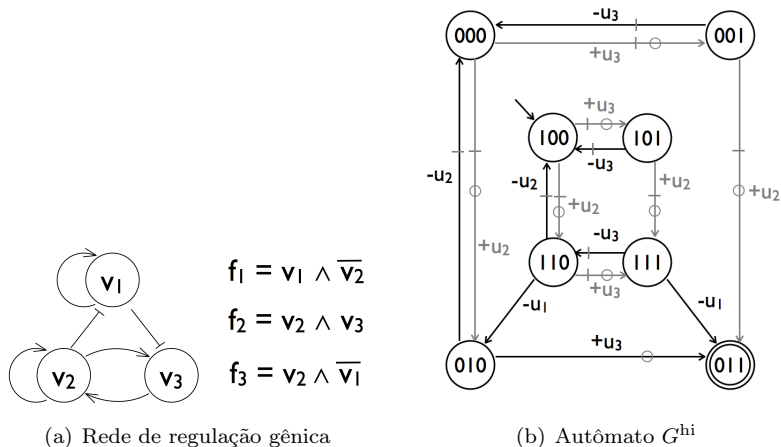


Figura 40 – (a) Rede de regulação gênica utilizada para o problema de minimização da energia consumida no trajeto de $x_0 = 100$ a $X_m = \{011\} = X_m^*$. (b) Autômato G^{hi} construído a partir do modelo aumentado G^{au} (não mostrado na figura), para o caso em que v_2^s e v_3^s são empregados para fins de controle.

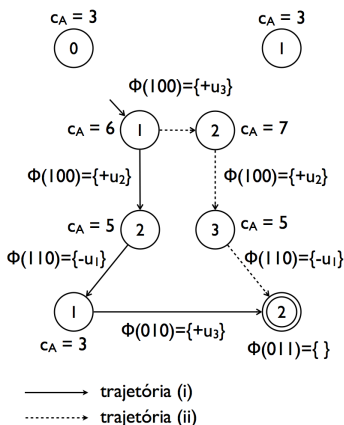


Figura 41 – Associação de custos $c(x)$ a cada macroestado do autômato G^{hi} da Figura 40. As cadeias binárias dos macroestados não estão aqui representadas, mas os macroestados guardam as posições utilizadas na Figura 40. Além dos custos $c(x)$, a figura mostra também os custos acumulados de x até X_m^* , $c_A(x)$.

- 1: Seja $U_0 = X_m^*$;
- 2: Faça $c_A(x) = c(x)$ para $x \in X_m^*$ e $c_A(x) = \infty$, se $x \notin X_m^*$;
- 3: $\forall x \in X_m^*$, faça $S^{\text{hi}}(x) = \phi$, com ϕ satisfazendo $\mathcal{H}_\phi^{\text{hi}}(x) \subseteq X_m^*$;
- 4: **repeat**
- 5: Ache $M = \{x \in (X \setminus U_j) \mid x \models SR^P(G^{\text{hi}}) \wedge \exists \phi \text{ com } \mathcal{H}_\phi^{\text{hi}}(x) \subseteq U_j\}$;
- 6: **for all** $x \in M$ **do**
- 7: $c_A(x) = c(x) + \min_{\phi} \max_{y \in \mathcal{H}_\phi^{\text{hi}}(x)} c_A(y)$;
- 8: $S^{\text{hi}}(x) = \arg \min_{\phi} \max_{y \in \mathcal{H}_\phi^{\text{hi}}(x)} c_A(y)$;
- 9: **end for**
- 10: $V_j = \{x \in M \mid c_A(x) \leq c_A(y) \text{ para qualquer } y \in M\}$;
- 11: $U_{j+1} = U_j \cup V_j$;
- 12: **until** $U_{j+1} = U_j$
- 13: Faça $S^*(x) = S^{\text{hi}}(x)$ para aqueles x que pertencem a algum caminho de malha fechada $W \in \mathcal{P}(G^{\text{hi}}, S^{\text{hi}})$.

Os elementos centrais do algoritmo são o conjunto U_j e a maneira como ele é atualizado. Inicialmente, faz-se $U_0 = X_m^*$ (Linha 1). Em cada iteração j , adicionam-se ao conjunto U_j os macroestados $x \in (X \setminus U_j)$ que chegam a U_j com o menor custo acumulado $c_A(x)$ possível (Linha 11 do algoritmo), macroestados esses que são armazenados em V_j (Linha 10). Por custo acumulado de um macroestado x , entende-se a soma de seu custo $c(x)$ com o maior custo acumulado $c_A(y)$ dentre os macroestados $y \in \mathcal{H}_\phi^{\text{hi}}(x)$, para um ϕ que minimize tal custo acumulado máximo (Linha 7). O custo acumulado dos macroestados em U_0 é o próprio custo do estado (Linha 2).

Assim, o algoritmo vai explorando, de X_m^* em direção a x_0 , os macroestados x de menor custo acumulado até X_m^* , acrescentando-os a U_j . Ao longo da execução do algoritmo os controles S^{hi} que induzem os menores custos são associados a cada macroestado (Linha 8). Ao final, o comportamento de $S^{\text{hi}}/G^{\text{hi}}$ é tal que o maior custo $C(W)$, com $W \in \mathcal{P}(G^{\text{hi}}, S^{\text{hi}})$, é o menor dentre todos os possíveis.

Note que nem todo $x \in U_k$, com U_k sendo o último elemento da iteração do algoritmo, fará parte das trajetórias ótimas de x_0 a X_m^* . Alguns x são ótimos somente no trajeto de x a X_m^* . Ou seja, pode ocorrer que macroestados x cujos custos acumulados $c_A(x)$ sejam baixos só sejam alcançados a partir de x_0 em trajetórias que passam por estados altamente custosos. Por isso que, na Linha 13, atribuem-se ao supervisor $S^*(x)$ somente os controles associados a caminhos gerados por $S^{\text{hi}}/G^{\text{hi}}$, o comportamento em malha fechada a partir de x_0 .

7 CONTROLE DE REDES MODELADAS POR EQUAÇÕES DIFERENCIAIS

Neste capítulo, exploraremos brevemente como as ideias até aqui desenvolvidas aplicam-se ao caso em que a rede de regulação gênica é modelada por um sistema de equações diferenciais ordinárias (EDOs). O objetivo é delinear uma metodologia que permita estender o alcance de nossa abordagem para abarcar também os modelos contínuos disponíveis na literatura.

Tal método consiste em obter um autômato de estados finitos a partir de um modelo em EDOs e, sobre esse novo modelo, derivar um supervisor que satisfaça as especificações de malha fechada. O supervisor é, então, convertido para um conjunto de equações diferenciais correspondentes que, quando acopladas ao modelo contínuo da rede gênica, possa atender aos requisitos de projeto do sistema de controle.

Assim, a metodologia para o controle de uma rede gênica, no caso específico de serem as redes modeladas por EDOs, pode ser bem representada pelo diagrama hierárquico em três níveis da Figura 42. As setas ascendentes do diagrama partem do sistema real a ser controlado e chegam a representações cada vez menos detalhadas, mais abstraídas, do comportamento do sistema. As setas descendentes, por outro lado, fazem o caminho oposto, iniciando o trajeto no simplificado universo das funções $S : X \rightarrow 2^{\Sigma^{\text{hi}}}$, para terminar no mundo detalhado das sequências de nucleotídeos em genes sintéticos. Neste capítulo, trataremos especificamente da interface entre os dois andares mais altos do diagrama na Figura 42, de forma a obter um controle contínuo para o modelo igualmente contínuo da planta. Importante ressaltarmos, no entanto, que o diagrama da figura não implica que a modelagem da rede gênica e a realização biológica do controlador passem, necessariamente, por etapas assentes em formalismos contínuos.

As ideias contidas nas próximas seções descrevem a metodologia esquematizada na Figura 42 e embasam a conjectura de que o formalismo apresentado na Tese não se restringe a tratar plantas cujos comportamentos são modelados *a priori* por formalismos da área de Sistemas a Eventos Discretos.

Argumentaremos aqui em favor dessa conjectura de uma maneira mais empírica, exploratória, expondo as ideias por meio de um exemplo específico, em vez de proposições mais gerais e matematicamente demonstráveis. Ao longo de todo capítulo, apoiaremos-nos sobre resultados e bibliografias já citados anteriormente, especialmente nos

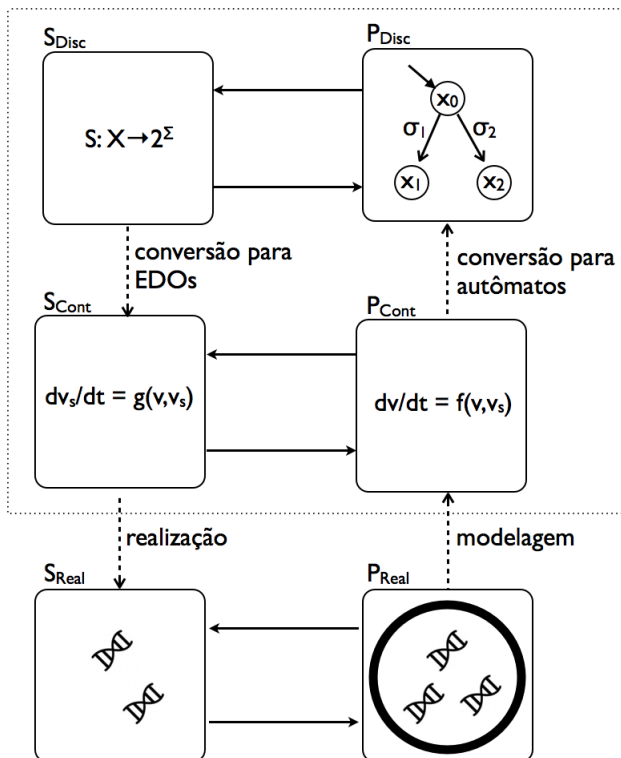


Figura 42 – Representação esquemática de uma metodologia para controle de redes de regulação gênica modeladas por equações diferenciais.

Capítulos 3 e 5.

Este capítulo é, assim, menos formal do que os dois anteriores e os resultados aqui apresentados merecem investigação futura mais aprofundada. Ele é, pois, um capítulo de passagem entre a apresentação das maiores contribuições da Tese e a descrição das perspectivas de trabalhos vindouros que se fará ao final deste documento.

7.1 ABSTRAÇÕES A EVENTOS DISCRETOS DE EDO

Considere a mesma rede de regulação gênica utilizada para ilustrar nossa abordagem em capítulos anteriores, e reproduzida aqui na

Figura 43. Modelemos essa rede por meio de equações diferenciais, seguindo os conceitos apresentados no Capítulo 3, comumente adotados pela comunidade de Biologia Sistêmica para captar a dinâmica de redes de regulação gênica (KARLEBACH; SHAMIR, 2008).

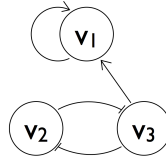


Figura 43 – Rede de regulação gênica que será modelada por um sistema de equações diferenciais.

No sistema de EDOs da equação 7.1, cada variável v_i representa a concentração da proteína produzida *a partir do gene nativo* v_i , um fator de transcrição que temos denominado u_i ao longo da Tese. Como o gene sintético v_i^s também codifica para a proteína u_i , utilizaremos u_i para denotar a soma $v_i + v_i^s$, em que v_i^s é a variável associada com a concentração da proteína u_i produzida *a partir do gene nativo* v_i^s . Assim, u_i , v_i e v_i^s possuem todas um duplo significado: são entidades físicas (p.e. u_i é o fator de transcrição produzido pelos genes v_i e v_i^s) e também grandezas químicas (i.e. concentração proteica). O contexto em que elas aparecem deixará evidente qual a semântica adotada.

$$\begin{aligned}
 \frac{dv_1}{dt} &= k_{1s} \frac{k_{11}v_1^{n_{11}}}{(1 + k_{11}v_1^{n_{11}})} \frac{k_{13}v_3^{n_{13}}}{(1 + k_{13}v_3^{n_{13}})} - \gamma_1v_1 \\
 \frac{dv_2}{dt} &= k_{2s} \frac{1}{1 + k_{23}v_3^{n_{23}}} - \gamma_2v_2 \\
 \frac{dv_3}{dt} &= k_{3s} \frac{1}{1 + k_{32}v_2^{n_{32}}} - \gamma_3v_3
 \end{aligned} \tag{7.1}$$

Considere, ainda, para o restante deste capítulo, que os parâmetros contidos nas equações 7.1 assumem os valores listados na tabela abaixo.

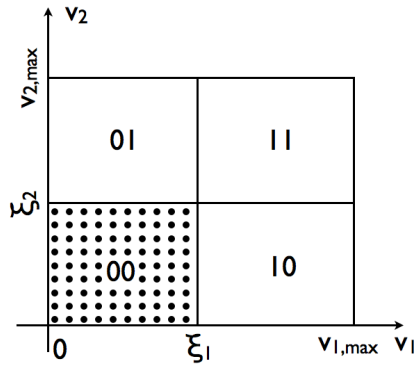
k_{1s}	2	γ_1	1
k_{2s}	2	γ_2	1
k_{3s}	2	γ_3	1
k_{11}	1	n_{11}	2
k_{13}	1	n_{13}	2
k_{23}	2	n_{23}	4
k_{32}	2	n_{32}	4

A fim de obtermos um modelo a eventos discretos associado ao sistema de equações 7.1, é necessário que, primeiramente, particionemos o espaço de estados contínuo formado pelas variáveis v_1 , v_2 e v_3 . Tal particionamento definirá o conjunto de *estados* do modelo discretizado. A Figura 44(a) ilustra uma partição Booleana de um espaço de estados bidimensional $v_1 \times v_2$ qualquer em quatro elementos: 00, 01, 10 e 11. Tais elementos da partição ficam naturalmente determinados, uma vez escolhidos um valor de limiar, ξ_i , bem como um valor máximo $v_{i,\max}$, para cada eixo v_i (supomos, aqui, que todas as variáveis satisfazem $v_i \geq 0$).

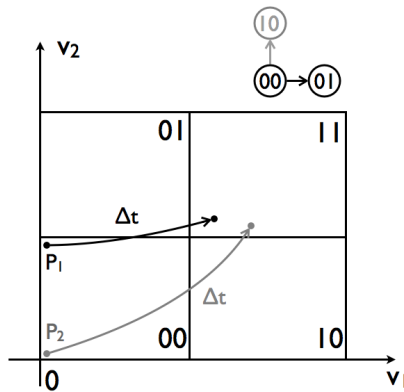
A escolha do parâmetro ξ_i e, portanto, da linha divisória entre “concentração alta de v_i ” e “concentração baixa de v_i ” depende da função matemática que modela a influência da variável v_i sobre a expressão dos genes aos quais o fator de transcrição u_i se acopla. Por exemplo, no caso do sistema de EDOs da equação 7.1, a taxa de síntese de v_3 é reduzida a um terço do seu valor máximo para $v_2 = 1$. É razoável dizer, assim, que valores de v_2 iguais ou maiores que $\xi_2 = 1$ correspondem a concentrações altas de v_2 , dado que o efeito repressor de v_2 sobre v_3 , nesse caso, é pronunciado. Não prolongaremos aqui a discussão interessante acerca da escolha de ξ_i e seu impacto na abstração discreta de uma planta contínua, embora esse tópico seja importante e mereça ser explorado.

Definida a partição do espaço de estados, podem-se emprestar diversas metodologias da comunidade de Sistemas Híbridos, para construir uma abstração a eventos discretos de um modelo em equações diferenciais (RAISCH; O’YOUNG, 1998; CURY; KROGH; NIINOMI, 1998). Uma possível estratégia, que será adotada nas discussões deste capítulo, é definir, para cada elemento da partição, um conjunto de pontos no interior dela, que serão mapeados segundo o modelo contínuo ao longo de um intervalo de tempo fixo Δt , com o objetivo de especificar quais transições entre os estados do modelo discretizado são possíveis.

Analise, por exemplo, a Figura 44(b). Os pontos P_1 e P_2 indicados, ao deixarem o elemento 00, fazem-no passando, *primeiro*, pelos elementos 01 e 10, respectivamente. Isso equivale a dizer que as transi-



(a) Partição do espaço de estados



(b) Mapeamento de pontos de um elemento da partição

Figura 44 – Exemplo de uma partição Booleana de um espaço de estados bidimensional. (a) Cada elemento da partição está associado a um estado do sistema discretizado. A fim de construir o diagrama de transição de estados, seleciona-se um conjunto de pontos para representar cada elemento da partição, como mostrado esquematicamente para 00. (b) Os pontos de um dado elemento da partição são usados como condição inicial do sistema de equações diferenciais, que é, então, simulado por um intervalo de tempo Δt , de maneira a determinar as possíveis transições que comporão o diagrama de transição de estados.

ções $00 \rightarrow 01$ e $00 \rightarrow 10$ são comportamentos factíveis do sistema. Esse processo de mapeamento dos pontos repete-se para todos os pontos do elemento 00 e, posteriormente, para todos os outros pontos pertencentes a 10, 11 e 01.

A Figura 45 traz o algoritmo em Matlab que implementa o procedimento descrito no parágrafo anterior, aplicando-o à planta contínua modelada pelas equações 7.1. Executamos a função `malha_aberta` para `tempo_final = 20` e `in_elemento` pertencente ao conjunto de partições $\{000, 001, 010, 011, 100, 101, 110, 111\}$, obtendo o diagrama de transição de estados da Figura 46(a).

As transições no diagrama da Figura 46(a) são representadas por setas cujas espessuras são proporcionais ao número de pontos que vão de um estado a outro. Por exemplo, as transições de 011 para os estados 001 e 010 acontecem para um número maior de pontos que a transição a 000. Na Figura 46(b), eliminamos as transições com baixa representatividade (nesse caso particular, aquelas que ocorrem para menos de 10% dos pontos do elemento da partição) e desprezamos a informação quantitativa associada à espessura das setas, obtendo um diagrama de transição de estados convencional.

É interessante comparar o diagrama de transição de estados da Figura 46(b) com aquele obtido no capítulo 5, construído diretamente a partir de funções lógicas booleanas. As transições entre os estados em ambos os diagramas são iguais, à exceção da passagem de 101 para 001, que não era observada no caso anterior.

O diagrama de transição de estados da Figura 46(b) pode ser usado como ponto de partida para construção dos autômatos G^{au} e G^{hi} , da mesma forma como fizemos no capítulo 5. Se escolhermos, para fins de controle genes sintéticos que codificam para os fatores de transcrição u_1 e u_3 , ou seja, genes v_1^s e v_3^s , respectivamente, obtemos o autômato G^{hi} graficamente representado na Figura 47 (por questão de simplicidade, não mostramos aqui o autômato intermediário G^{au}). Suporemos que todos os eventos de alto nível não controláveis podem ser preemptados pelos eventos forçáveis.

7.2 CONVERSÃO DO SUPERVISOR EM EDO

Suponha um problema de atração de estados em que $X_m = \{101\}$ e $P = \text{verdadeiro}$. Aplicando-se os conceitos desenvolvidos no capítulo anterior, verifica-se que o macroestado inicial $x_0 = 010$ do autômato G^{hi} na Figura 47 satisfaz o predicado de alcançabilidade estrita. Além

```

function [tempo,matriz_abs,matriz_rel] =...
        malha_aberta(in_elemento, tempo_final)

% tempo_final      : tempo para simulação do sistema em malha aberta
% in_elemento      : indica elemento da partição a ser mapeado
% tempo            : vetor com tempo que vai de 0 a "tempo_final"
% matriz_abs       : [P(000) P(001) ... P(111)], em que P(q)
%                   : é a quantidade de pontos que atingiram "q"
% matriz_rel       : matriz abs/sum(matriz_abs)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% definição dos parâmetros dos elementos das partições

% distância entre os pontos nos eixos v1, v2 e v3, respectivamente
d(1) = .1; d(2) = .1; d(3) = .1;

% valores mínimo, de limiar e máximo de v1, v2 e v3
v1(1) = 0; v1(2) = 1; v1(3) = 4;
v2(1) = 0; v2(2) = 1; v2(3) = 4;
v3(1) = 0; v3(2) = 1; v3(3) = 4;

% função "particao" devolve os pontos pertencentes à "in_elemento"
pontos = particao(in_elemento,d,v1,v2,v3);
[lin_pontos,col_pontos] = size(pontos);

tspan = [0,tempo_final];

% criação do eixo do tempo
tempo = 0:1:tempo_final; size_tempo = length(tempo);

matriz_abs = zeros(1,8); matriz_rel = zeros(1,8);
for i=1:lin_pontos
    % toma o ponto da i-ésima linha como condição inicial
    y0 = pontos(i,:);

    % guarda o resultado da simulação do sistema para y0
    solucao_edo = ode45(@planta,tspan,y0);

    % avalia a solução da edo para os valores contidos no vetor "tempo"
    nivel_expressao_temp = deval(solucao_edo,tempo)';
    for j=1:size_tempo

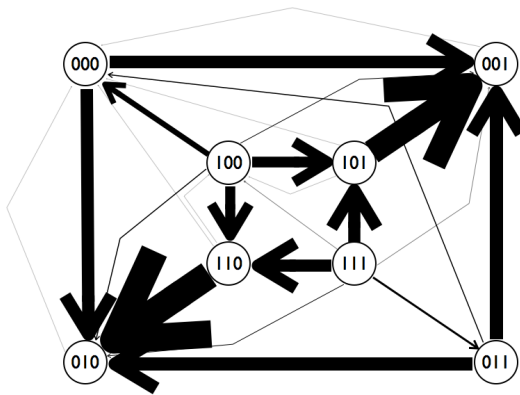
        % função "particiona_pontos" retorna um vetor linha, indicando à
        % qual elemento pertence o ponto "nivel_expressao_temp(j,:)"
        para_onde_ponto_foi =...
            find(particiona_pontos(nivel_expressao_temp(j,:),v1,v2,v3));

        % se o ponto mudou de elemento da partição
        if(para_onde_ponto_foi ~= in_elemento)
            % soma 1 na coluna correspondente da matriz matriz_abs
            matriz_abs(para_onde_ponto_foi) =...
                matriz_abs(para_onde_ponto_foi) + 1;

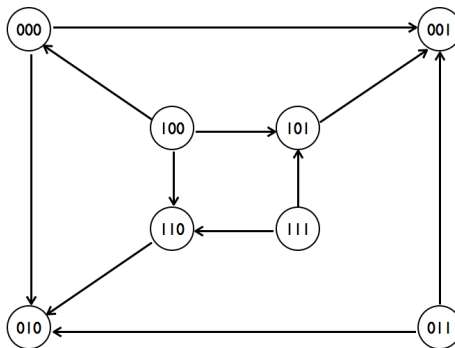
            % encerra execução do laço "for"
            break
        % se o ponto não mudou de elemento ao final da simulação
        elseif(j == size_tempo)
            matriz_abs(in_elemento) = matriz_abs(in_elemento)+1;
        end
    end
end
matriz_rel = matriz_abs/sum(matriz_abs);

```

Figura 45 – Algoritmo implementado em Matlab para construção de uma abstração discreta das equações 7.1, codificadas no arquivo planta.m.



(a) Abstração com todas as transições



(b) Abstração com transições que ocorrem para mais de 10% dos pontos

Figura 46 – (a) Diagrama de transição de estados discretizado do modelo contínuo da rede de regulação gênica. A espessura da seta que liga dois estados q, r é proporcional à quantidade de pontos do elemento da partição q que vão a r , sem passar por outros elementos da partição. (b) Diagrama de transição de estados convencional. As transições de q a r que ocorrem para menos do que 10% dos pontos em q são eliminadas. Não se retém neste autômato a informação quantitativa associada à espessura das transições do diagrama em (a).

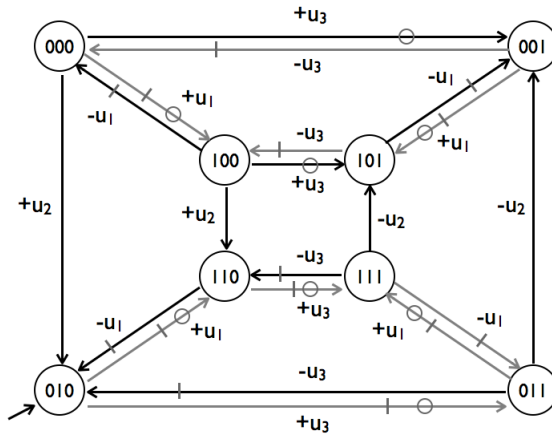


Figura 47 – Autômato de alto nível G^{hi} construído a partir do diagrama de transição de estados da Figura 46(b), quando genes sintéticos v_1^s e v_3^s são utilizados para fins de controle.

disso, $X_m = X_m^* = \{101\}$. Assim, o problema de atração de estados tem solução.

Um possível conjunto de entradas de controle no alto nível que resolve o problema é $S^{\text{hi}}(010) = \{+u_1\}$, $S^{\text{hi}}(110) = \{+u_3\}$, $S^{\text{hi}}(111) = \{-u_2\}$ e, finalmente, $S^{\text{hi}}(101) = \{\}$. De modo a simplificar as regiões de regulação gênica de v_1^s e v_3^s , faremos, ainda, $S^{\text{hi}}(011) = \{+u_1\}$ e $S^{\text{hi}}(100) = \{+u_3\}$. As funções de atualização de estados dos genes sintéticos v_1^s e v_3^s são dadas, então, pelas expressões $f_1^s(u_1, u_2, u_3) = (\bar{u}_1 \wedge u_2) \vee (u_1 \wedge u_3)$ e $f_3^s(u_1, u_2, u_3) = u_1$.

Convertamos agora as funções f_1^s e f_3^s em equações diferenciais ordinárias, seguindo a lógica esquematizada na Figura 48 para a função f_1^s . Os critérios empregados nessa conversão são semelhantes àqueles adotados para modelagem de uma rede de regulação gênica por equações diferenciais. Por exemplo, a cláusula $(\bar{u}_1 \wedge u_2)$ de f_1^s é traduzida para uma parcela de $d(v_1^s)/dt$ cujo valor máximo se obtém para $u_1 = 0$ e $u_2 \rightarrow \infty$. O sistema em malha fechada contínuo é, então, descrito pelas equações 7.2, que modelam a dinâmica das proteínas u_i produzidas a partir dos genes nativos, v_i , e a partir dos genes sintéticos, v_i^s . Como tais fatores de transcrição, por hipótese, desempenham a mesma função regulatória, existe um acoplamento entre a rede nativa e a rede sintética, ou planta e controlador, respectivamente.

$$\begin{aligned}
 h_1^s(u_1, u_2, u_3) &= (\bar{u}_1 \wedge u_2) \vee (u_1 \wedge u_3) \\
 \frac{dv_1^s}{dt} &= k_{4s} \frac{1}{(1 + k_{41}u_1^{n_{41}})} \frac{k_{42}u_2^{n_{42}}}{(1 + k_{42}u_2^{n_{42}})} + \\
 &+ k_{4s} \frac{k_{41}u_1^{n_{41}}}{(1 + k_{41}u_1^{n_{41}})} \frac{u_3^{n_{43}}}{(1 + k_{43}u_3^{n_{43}})} - \gamma_{4s}v_1^s
 \end{aligned}$$

Figura 48 – Obtenção de equações diferenciais que possam implementar, no universo contínuo, um comportamento equivalente àquele observado para genes sintéticos no universo Booleano. Os efeitos regulatórios de um fator de transcrição sobre a dinâmica das outras proteínas são modelados pelas funções de Hill, apresentadas no Capítulo 3. As operações lógicas de conjunção e disjunção são traduzidas para as operações de multiplicação e adição. O fator de degradação proteica $\gamma_{4s}v_1^s$ evita um “efeito memória” no circuito, ou seja, a presença do fator de transcrição u_i em um macroestado x devido tão somente a ativações em macroestados anteriores.

$$\begin{aligned}
\frac{dv_1}{dt} &= k_{1s} \frac{k_{11} u_1^{n_{11}}}{1 + k_{11} u_1^{n_{11}}} \frac{k_{13} u_3^{n_{13}}}{1 + k_{13} u_3^{n_{13}}} - \gamma_1 v_1 \\
\frac{dv_2}{dt} &= k_{2s} \frac{1}{1 + k_{23} u_3^{n_{23}}} - \gamma_2 v_2 \\
\frac{dv_3}{dt} &= k_{3s} \frac{1}{1 + k_{32} u_2^{n_{32}}} - \gamma_3 v_3 \\
\frac{dv_1^s}{dt} &= k_{4s} \frac{1}{(1 + k_{41} u_1^{n_{41}})} \frac{k_{42} u_2^{n_{42}}}{(1 + k_{42} u_2^{n_{42}})} + \\
&\quad + k_{4s} \frac{k_{41} u_1^{n_{41}}}{(1 + k_{41} u_1^{n_{41}})} \frac{u_3^{n_{43}}}{(1 + k_{43} u_3^{n_{43}})} - \gamma_{4s} v_1^s \\
\frac{dv_3^s}{dt} &= k_{5s} \frac{k_{51} u_1^{n_{51}}}{1 + k_{51} u_1^{n_{51}}} - \gamma_{5s} v_3^s \\
u_1 &= v_1 + v_1^s \\
u_3 &= v_3 + v_3^s \\
u_2 &= v_2
\end{aligned} \tag{7.2}$$

Para os novos parâmetros em $d(v_1^s)/dt$ e $d(v_3^s)/dt$, escolhemos os valores da tabela abaixo. As constantes k_{4s} e k_{5s} , que representam as máximas taxas de síntese dos fatores de transcrição u_1 e u_3 a partir dos genes sintéticos v_1^s e v_3^s , são três vezes maiores que seus pares k_{1s} , k_{2s} e k_{3s} , uma forma de garantir a preemptabilidade dos eventos relacionados aos genes sintéticos. Com exceção de k_{4s} e k_{5s} , assumimos para todas as outras constantes valores iguais àqueles escolhidos para os seus pares nos genes nativos. Abstemo-nos aqui de empreender uma investigação mais detalhada sobre a relação entre os valores de cada uma dessas constantes e suas contrapartidas biológicas.

k_{4s}	6	γ_{4s}	1
k_{5s}	6	γ_{5s}	1
k_{41}	1	n_{41}	2
k_{42}	1	n_{42}	4
k_{43}	2	n_{43}	2
k_{51}	1	n_{51}	2

A Figura 49 mostra de forma esquematizada o resultado da simulação do sistema de controle em malha fechada *contínuo* modelado pelas equações 7.2. Nela podemos acompanhar a evolução das variáveis u_1 , u_2 e u_3 , ou seja, a concentração total de cada um dos fatores de

transcrição, para os instantes de tempo indicados. Iniciamos a simulação com 3000 pontos uniformemente distribuídos no macroestado 010. A presença de todos os pontos em 010 é representada na figura pelo círculo cinza que preenche todo o macroestado 010. À medida que o tempo transcorre, pode-se visualizar como a distribuição dos pontos é alterada, em uma trajetória que segue aquela planejada pelo supervisor S^{hi} para o sistema a eventos discretos, culminando com a condução de todos os pontos para o macroestado de destino 101.

A mesma simulação é mostrada sob uma nova perspectiva na Figura 50, que traz o gráfico $\{u_1, u_2, u_3\} \times t$. Note como todos os pontos convergem para uma situação de equilíbrio, interpretada, no nível lógico, como sendo o macroestado 101. As altas constantes k_{4s} e k_{5s} acarretam concentrações finais de u_1 e u_3 quase duas vezes maiores do que aquelas verificadas no sistema em malha aberta. Trataremos desse tema na próxima seção.

7.3 ANÁLISE QUANTITATIVA DA FORÇABILIDADE

Na seção anterior, para garantir a hipótese de forçabilidade, fizemos $k_{4s} = k_{5s} = 6$, ou seja, escolhemos uma taxa de síntese de proteínas a partir dos genes sintéticos v_i^s três vezes maior do que a constante adotada para os genes nativos. Analisaremos agora se essa restrição rígida sobre os genes v_i^s pode ser relaxada, diminuindo as exigências construtivas sobre v_i^s .

Forçar um evento de alto nível em um dado macroestado x nem sempre significa *preemptar* outro $\pi \in \Sigma^{\text{hi}}$. Tome, por exemplo, o controle $S^{\text{hi}}(010) = \{+u_1\}$. Nenhum outro evento que não $+u_1$ parte do macroestado 010. Para induzir a transição $\delta^{\text{hi}}(010, +u_1) = 110$, portanto, o evento de baixo nível $+v_1^s$ não precisa “vencer” nenhum $-v_j$ ou $+v_j$ da rede nativa. Logo, não há necessidade de altas taxas de síntese de u_1 a partir do gene v_1^s em 010.

O controle $S^{\text{hi}}(110) = \{+u_3\}$, por outro lado, demanda a preempção de um evento, especificamente, de $-u_1$. Mas perceba que, nesse caso, o evento de alto nível $-u_1$ em 110 é reflexo da ocorrência de $-v_1^s$ no baixo nível (dado que v_1 não está ativado em 110). Assim, a preempção de $-u_1$ por $+u_3$ requer que a ativação do gene sintético v_3^s em 110 seja mais rápida do que a desativação de v_1^s no mesmo macroestado. Como o gene sintético v_1^s foi construído para se ativar em 010 e se desativar em 110 (veja a região de regulação f_1^s), a taxa de diminuição da concentração da proteína u_1 sintetizada a partir de v_1^s é

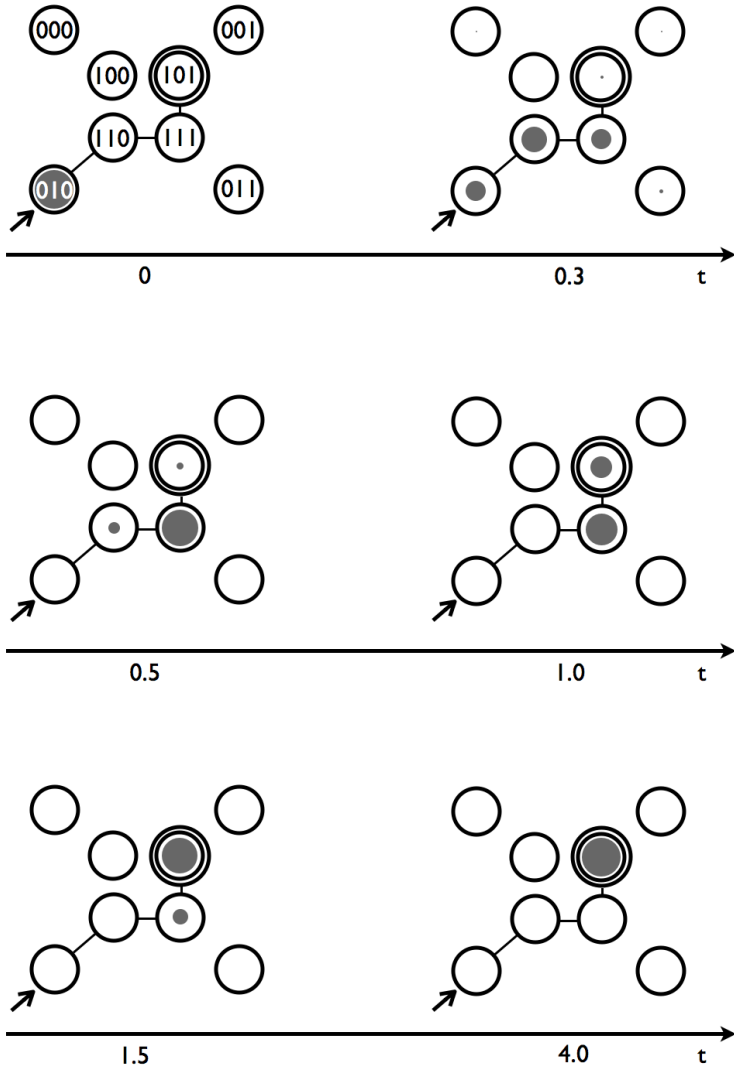


Figura 49 – Representação esquemática da evolução temporal das variáveis *contínuas* u_1 , u_2 e u_3 , segundo a dinâmica dada pelas equações 7.2. A área dos círculos cinzas em cada macroestado é proporcional à quantidade de pontos naquele macroestado no instante de tempo mostrado. Os macroestados conectados por linhas cheias perfazem a trajetória em malha fechada induzida por S^{hi} .

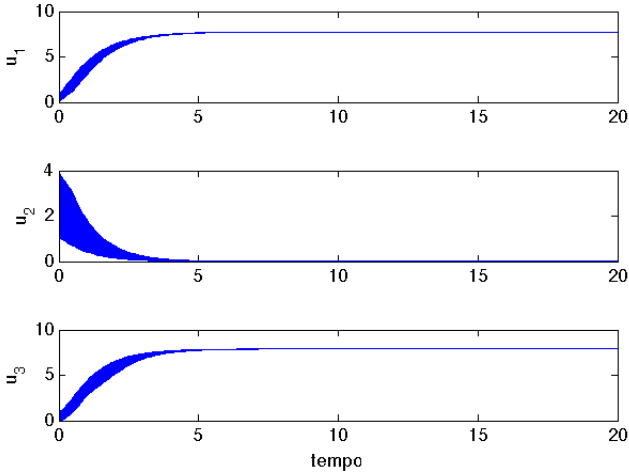


Figura 50 – Dinâmica temporal das variáveis u_1 , u_2 e u_3 , quando todos os pontos pertencentes ao macroestado 010 são tomados como condição inicial do sistema em malha fechada contínuo.

baixa, assim que o sistema entra no macroestado 110. Isso se dá pois o gene v_1^s está na região limítrofe entre os comandos “ative v_1^s ” (válido em 010) e “reprima v_1^s ” (válido em 110). Tal fato sugere que k_{3s} também pode ser reduzida, a julgar pelo comportamento do sistema até o macroestado 110. O mesmo argumento apresentado neste parágrafo, quando aplicado aos outros controles de alto nível $S^{hi}(x)$, embasa uma redução nos valores de k_{4s} e k_{5s} . De fato, como pode ser visto na Figura 51, os requisitos de malha fechada são satisfeitos para o sistema contínuo quando $k_{4s} = k_{5s} = 2$, em outras palavras, quando as taxas de síntese associadas aos genes sintéticos assumem valores iguais àqueles utilizados para os genes nativos.

Adicionalmente, uma redução em k_{4s} e k_{5s} diminui os valores de regime de u_1 e u_3 , como mostrado na Figura 52, o que implica um menor consumo de energia para síntese proteica.

Por fim, a Figura 53 mostra a simulação do sistema em malha fechada contínuo para uma condição inicial particular, qual seja, $v_1 = 0$, $v_2 = 3$, $v_3 = 0$ e $v_1^s = v_3^s = 0$. Essa figura mostra também os sinais de controle v_1^s e v_3^s e ilustra a sequência de transição entre macroestados que caracteriza a evolução temporal de u_1 , u_2 e u_3 .

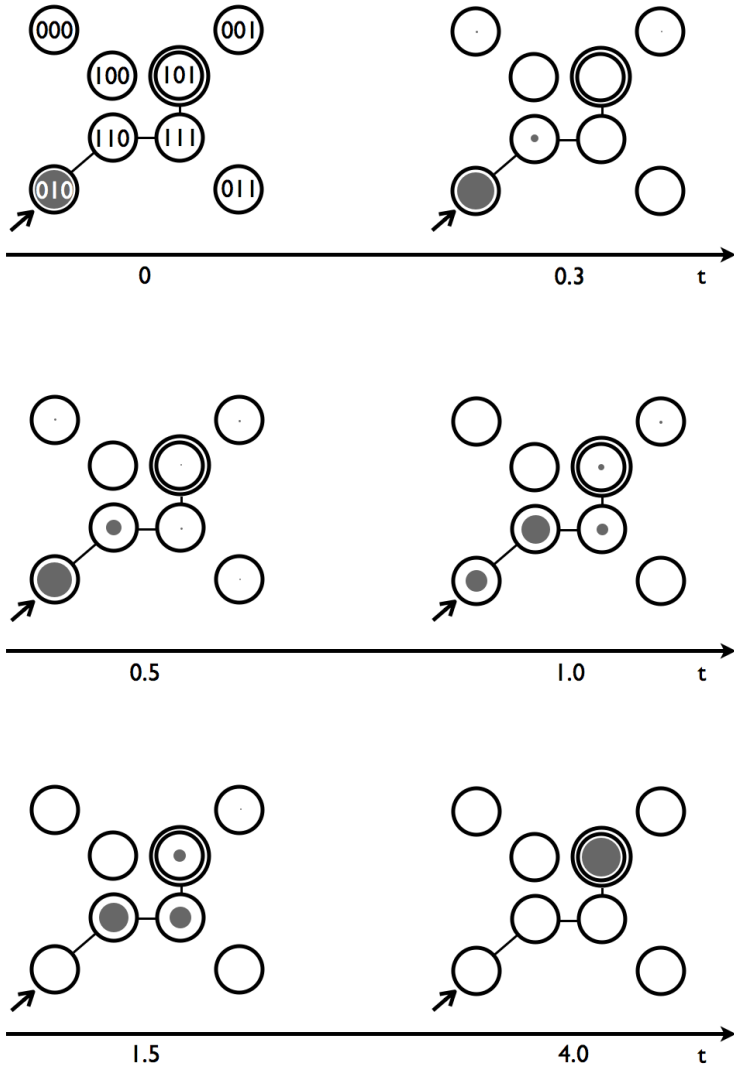


Figura 51 – Representação esquemática da evolução temporal das variáveis *contínuas* u_1 , u_2 e u_3 , segundo a dinâmica dada pelas equações 7.2. A área dos círculos cinzas em cada macroestado é proporcional à quantidade de pontos naquele macroestado no instante de tempo mostrado. Os macroestados conectados por linhas cheias perfazem a trajetória em malha fechada induzida por S^{hi} .

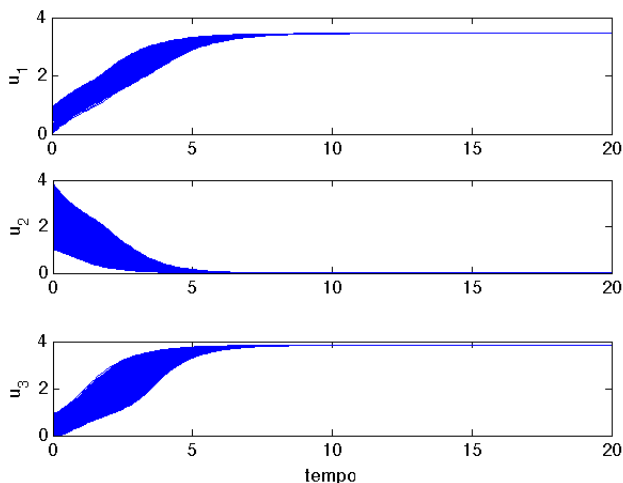


Figura 52 – Dinâmica temporal das variáveis u_1 , u_2 e u_3 , quando todos os pontos pertencentes ao macroestado 010 são tomados como condição inicial do sistema em malha fechada contínuo. A concentração de regime dos fatores de transcrição é menor do que aquela verificada para $k_{4s} = k_{5s} = 6$.

Naturalmente, a discussão qualitativa conduzida nesta seção encontra um paralelo quantitativo, obtido quando da análise dos valores mínimos e máximos, em cada estado, das expressões para $d(v_i^s)/dt$ e $d(v_i)/dt$. Por ora, o nosso objetivo, no entanto, é tão somente alertar para o fato de que satisfazer a hipótese de preemptabilidade no alto nível requer uma escolha de parâmetros no baixo nível que satisfaça um compromisso: nem tão rígida, dificultando a construção de genes sintéticos, e nem tão frouxa, invalidando a hipótese.

Como alertado na introdução deste capítulo, as ideias aqui presentes são, de uma forma geral, quase reflexões sobre um problema instigante a ser explorado. Embora não seja o objeto da Tese trazer respostas efetivas para os tópicos levantados nas últimas seções, as discussões apontam alguns elementos que comporão o espectro de nossas pesquisas futuras.

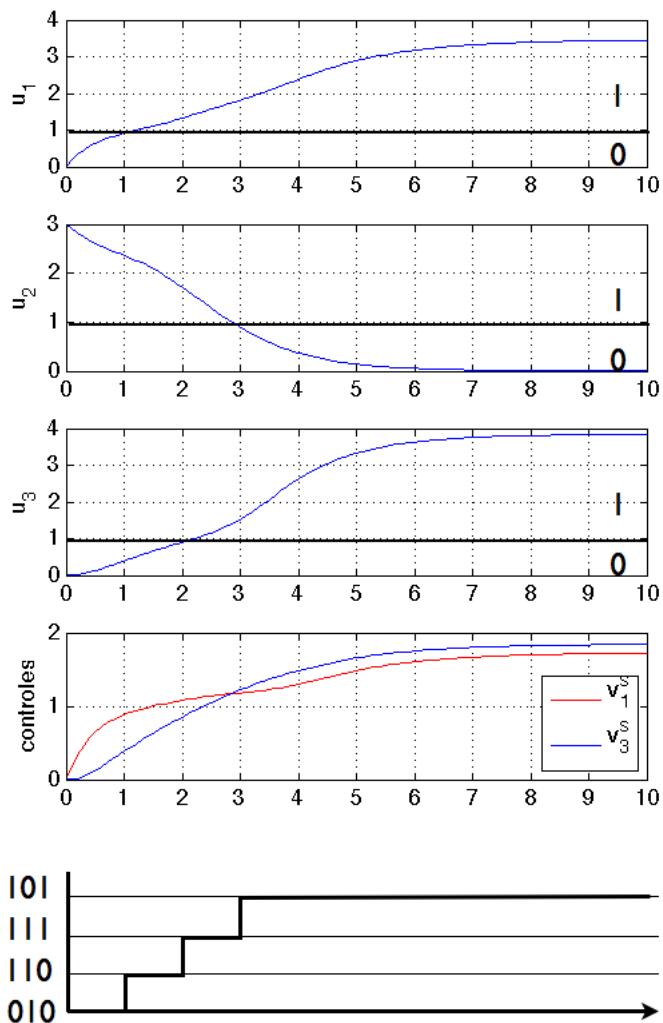


Figura 53 – Evolução temporal de u_1 , u_2 e u_3 para uma condição inicial específica. Os sinais de controle v_1^s e v_3^s são também mostrados. Na base da figura, pode-se acompanhar a trajetória de u_1 , u_2 e u_3 segundo uma perspectiva discretizada. Note como a sequência de macroestados visitados pelo sistema em malha fechada segue a trajetória induzida por S^{hi} sobre G^{hi} .

8 CONSIDERAÇÕES FINAIS

Apresentamos nesta Tese uma metodologia assente na Teoria de Controle Supervisório (TCS), para projetar controladores intracelulares de redes de regulação gênica, implementados biologicamente por genes sintéticos.

A Tese traz duas contribuições principais: a) estende a TCS, a fim de tratar o problema de atração de estados para o caso em que o controlador não somente habilita ou desabilita eventos, mas também preempta certos eventos não controláveis por meio de eventos forçáveis; e b) formaliza e resolve o problema de controle de redes gênicas por moléculas sintéticas, unindo, assim, em uma mesma abordagem, Biologia Sistêmica (para modelagem e análise), Teoria de Controle e, por fim, Biologia Sintética (para realização do controlador).

A extensão teórica da TCS, conduzida ao longo do Capítulo 5, foi fortemente motivada pelo problema de controle biológico tratado. Em outras palavras, a tarefa de intervenção em redes gênicas não foi um mero exemplo escolhido *a posteriori*, por se adequar bem a novidades teóricas. A tarefa biológica constitui-se, antes, no ponto de partida de nossa investigação. Por isso, discorreremos mais detalhadamente sobre a natureza das redes de regulação gênica e seus modelos matemáticos nos capítulos 2 e 3 da Tese.

A rica partição qualitativa do conjunto de eventos a que aludimos acima resulta da estratégia de controle que adotamos, fundada no uso de genes sintéticos. Como não podemos intervir *diretamente* na ativação ou desativação dos genes da rede nativa, os eventos a eles relacionados são não controláveis. A controlabilidade reserva-se, então, somente àqueles eventos vinculados aos genes sintéticos, pois assumimos ser possível especificar *a priori* quando cada um deles será ativado e desativado. Por fim, a hipótese de preemptabilidade de certos eventos da rede nativa fez-se necessária para tornar factível o controle da rede. Fossem todos os eventos da planta não preemptáveis, as possibilidades de resolução do problema de controle seriam muito pequenas. A hipótese de preemptabilidade, como argumentado no Capítulo 5, é, de fato, biologicamente plausível.

É interessante ressaltar, ainda, que o tom hierárquico de nossa abordagem, e.g. definição de planta e supervisor no baixo e alto níveis, é também fruto da estratégia de intervenção por nós escolhida. Como genes nativos e sintéticos codificam para as mesmas proteínas, torna-se oportuno agrupar os estados em que a concentração proteica é a mesma,

independente da fonte que as origina, pois esses estados são equivalentes de uma perspectiva fisiológica. Ao procedermos dessa forma, já colocamos o problema como nos termos de Controle Hierárquico; abstraímos os estados de genes individuais, passando a pensar em termos de concentração proteica. A validade da propriedade de Consistência Hierárquica em nossa abordagem, como mostrado no Capítulo 5, comprova que o passo de abstração do baixo para o alto nível mencionado não prejudica a resolução do problema de controle: o comportamento que exigimos do sistema no nível proteico pode ser induzido pela ação dos genes sintéticos no nível mais baixo.

Do ponto de vista biológico, o corpo conceitual das extensões à TCS que propomos requer que: a) as redes gênicas possam ser abstraídas por modelos a eventos discretos; b) as regiões de regulação dos genes sintéticos possam ser programadas para iniciar a transcrição em estados celulares específicos; c) as sequências de nucleotídeos dos genes sintéticos sejam tais que suas dinâmicas de expressão gênica sejam mais rápidas do que algumas de seus pares nativos. Assim, embora tenhamos desenvolvido a Tese para uma instância particular e simplificada do problema – redes nativas *Booleanas* controladas por genes sintéticos que computam *quaisquer* funções lógicas –, a essência do nosso trabalho pode ser estendida para abarcar os casos mais gerais, e.g. redes gênicas multiníveis (em que os genes assumem mais do que dois estados), regiões de regulação com restrições de programação, etc.

Acerca das hipóteses biológicas levantadas no parágrafo anterior, e sobre as quais se assenta nossa Tese, nenhuma delas é irrealista. Os modelos a eventos discretos pressupõem, basicamente, a existência de um conjunto discreto de estados por entre os quais o sistema transita. Ora, tal característica é uma das mais marcantes em sistemas vivos (e a responsável pela ordem que eles transparecem): os seres multicelulares, por exemplo, são compostos por um conjunto discreto de tipos celulares e não por uma massa de células cujas características fisiológicas enfileiram-se em um espectro contínuo. Citamos, no Capítulo 3, diversos trabalhos publicados que dão respaldo à ideia de que os modelos a eventos discretos captam sim a dinâmica de importantes processos biológicos. Para os casos em que modelos contínuos são mais adequados, no entanto, nossas ideias podem ser empregadas para facilitar a síntese do controlador, como conjecturamos no capítulo 7 da Tese. Em outras palavras, resolver o problema de controle no nível discreto *pode* ser mais simples do que atacar sistemas de equações diferenciais não lineares diretamente.

A programação das regiões de regulação de genes sintéticos, a

segunda das hipóteses que assumimos, não é algo puramente teórico, mas realidade já divulgada em diversos artigos experimentais publicados recentemente na literatura, como mencionamos nos capítulos 4 e 5, na seção dedicada à realização do supervisor. A despeito dos avanços da Biologia Sintética, entretanto, existem diversas restrições para a programação de genes. Esse tópico foi tratado no Capítulo 6, em que mostramos como simplificar a região de regulação de genes sintéticos *após a síntese do supervisor*. A hipótese de preemptabilidade dos eventos ligados aos genes sintéticos também é razoável biologicamente, como mostram os trabalhos científicos que alteram, por exemplo, as taxas de transcrição e tradução de genes, por meio da manipulação das sequências não transcritas, das regiões não traduzidas, ou mesmo dos códons usados para especificar as cadeias proteicas.

O futuro que vislumbramos para o trabalho aqui apresentado passa por dois caminhos: um teórico e o outro prático. No primeiro, pretendemos continuar o trabalho compilado no Capítulo 7, sobre o uso de abstrações discretas para resolver o problema de controle de redes biológicas modeladas por formalismos contínuos. Como derivar “boas” abstrações discretas de plantas contínuas, de que forma ajustar os parâmetros cinéticos das equações diferenciais que implementam o supervisor são duas das principais questões que colocamos como perspectivas de pesquisa. Nesse ponto, os resultados da área de Sistemas Híbridos servirão como referencial teórico de nossas pesquisas.

Ainda no âmbito teórico, temos interesse em investigar com mais detalhes como os conceitos de hierarquia e modularidade, bem estabelecidos nas Engenharias, podem nos ajudar a melhor compreender o comportamento de seres vivos e a intervir em suas dinâmicas. Modularidade é uma característica marcante na Biologia – presente desde a constituição das biomoléculas (e.g. proteínas possuem módulos com competências funcionais distintas), passando pelas topologias modulares de redes intracelulares e chegando até os órgãos especializados em seres multicelulares – e pode ser explorada para diminuir o custo computacional da síntese de controladores. O conceito de hierarquia, por sua vez, constitui-se no elemento central para compreender de que maneira os fenômenos em diferentes níveis da escala biológica se relacionam. Para exemplificar, como pensamentos derivam de estados celulares neuronais, e esses de estados das redes intracelulares. De modo a estender nossas ideias a arranjos biológicos mais complexos do que o nível celular, como tecidos e órgãos, acreditamos ser necessário, portanto, um estudo mais aprofundado da organização hierárquica dos organismos.

Finalmente, desejamos aplicar as ideias aqui desenvolvidas a um problema prático, a fim de testar, validar e rever nossas hipóteses, enriquecer e estender a abordagem teórica apresentada e cumprir o fim a que se devem destinar as teorias na Engenharia: bem compreender os sistemas naturais e utilizar seus princípios para resolver problemas relevantes para a sociedade.

REFERÊNCIAS

- ALBERT, R. Boolean modeling of genetic regulatory networks. *Lecture Notes on Physics*, v. 650, p. 459–481, 2004.
- ALBERTS, B. et al. *Molecular biology of the cell*. New York: Garland Science, 2008.
- ANDRIANANTOANDRO, E. et al. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, v. 2, p. 1–14, 2006.
- ARKIN, A.; ROSS, J.; MCADAMS, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected escherichia coli cells. *Genetics Society of America*, v. 149, p. 1633–1648, 1998.
- BAGH, S.; MANDAL, M.; MCMILLEN, D. An active intracellular device to prevent lethal disease outcomes in virus-infected bacterial cells. *Biotechnology and Bioengineering*, v. 108, p. 645–654, 2011.
- BALDISSERA, F. L.; CURY, J. E. R. Application of supervisory control theory to guide cellular dynamics. In: *Proceedings of WODES 2012*. Guadalajara, Mexico: [s.n.], 2012. p. 384–389.
- BALDISSERA, F. L.; CURY, J. E. R. Dealing with biological constraints in the synthesis of controllers for gene regulatory networks. Aceito para publicação nos Proceedings of the IFAC'14. 2014.
- BALDISSERA, F. L.; CURY, J. E. R. Optimality in the control of gene regulatory networks. Submetido para IEEE 53rd Annual Conference on Decision and Control. 2014.
- BALDISSERA, F. L.; CURY, J. E. R.; RAISCH, J. A supervisory control theory approach for the control of gene regulatory networks. Submetido para IEEE Transactions on Automatic Control. 2014.
- BENENSON, Y. Biomolecular computing systems: principles, progress and potential. *Nature Reviews Genetics*, v. 13, p. 455–468, 2012.
- BERNOT, G.; COMET, J.-P.; RICHARD, A. Application of formal methods to biological regulatory networks. *Journal of Theoretical Biology*, v. 229, p. 339–347, 2004.

- BLAZECK, J. et al. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnology and Bioengineering*, v. 109, p. 1–12, 2012.
- BORNHOLDT, S. Less is more in modeling large genetic networks. *Science*, v. 310, p. 449–451, 2005.
- BRANDIN, B. A.; WONHAM, W. Supervisory control of timed discrete-event systems. *IEEE Transactions on Automatic Control*, v. 39, p. 329–342, 1994.
- BRAVE, Y.; HEYMANN, M. On stabilization of discrete-event processes. In: *Proceeding of the 28th IEEE Conference on Decision and Control*. Tampa, FL: [s.n.], 1989. p. 2737–2742.
- BRAVE, Y.; HEYMANN, M. On optimal attraction in discrete-event processes. *Information Sciences*, v. 67, p. 245 – 276, 1993.
- BUCHLER, N.; GERLAND, U.; HWA terence. On schemes of combinatorial transcription logic. *PNAS*, v. 100, p. 5136–5141, 2003.
- CASSANDRAS, C. G.; LAFORTUNE, S. *Introduction to discrete event systems*. New York: Springer, 2008.
- CHAOUIYA, C.; REMY, E. Logical modelling of regulatory networks, methods and applications. *Bulletin of Mathematical Biology*, v. 75, p. 891–895, 2013.
- CHAVES, M.; GOUZÉ, J.-L. Exact control of genetic networks in a qualitative framework: the bistable switch example. *Automatica*, v. 47, p. 1105–1112, 2011.
- CLARKE, E. M. *Model Checking*. Cambridge, Massachusetts: The MIT Press, 1999.
- CULLER, S.; HOFF, K.; SMOLKE, C. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science*, v. 330, p. 1251–1255, 2010.
- CURY, J. E.; BALDISSERA, F. L. Systems biology, synthetic biology and control theory: A promising golden braid. *Annual Reviews in Control*, v. 37, n. 1, p. 57 – 67, 2013.
- CURY, J. E. R.; KROGH, B.; NIINOMI, T. Synthesis of supervisory controllers for hybrid systems based on approximating automata. *IEEE Transactions on Automatic Control*, v. 43, p. 564–568, 1998.

- DATTA, A. et al. Control approaches for probabilistic gene regulatory networks. *IEEE Signal Processing Magazine*, v. 24, p. 54–63, 2007.
- DAVIDICH, M. I.; BORNHOLDT, S. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS One*, v. 3, p. 1–8, 2008.
- EPSTEIN, J. M. Why model? *Journal of Artificial Societies and Social Simulation*, v. 11, p. 1–5, 2008.
- EWALD, R. et al. Discrete event modelling and simulation in systems biology. *Journal of Simulation*, v. 1, p. 81–96, 2007.
- FISHER, J.; HENZINGER, T. Executable cell biology. *Nature biotechnology*, v. 25, p. 1239–1249, 2007.
- GALLOWAY, K. E.; FRANCO, E.; SMOLKE, C. D. Dynamically reshaping signaling networks to program cell fate via genetic controllers. *Science*, v. 341, p. 12350051–123500519, 2013.
- GINGOLD, H.; PILPEL, Y. Determinants of translation efficiency and accuracy. *Molecular Systems Biology*, v. 7, p. 1–13, 2011.
- GLASS, L.; KAUFFMAN, S. A. The logical analysis of continuous, non-linear biochemical control networks. *Journal of Theoretical Biology*, v. 39, p. 103–129, 1973.
- GOLASZEWSKI, C. H.; RAMADGE, P. Control of discrete event processes with forced events. In: *Proceedings of the 26th Conference on Decision and Control*. [S.l.: s.n.], 1987. p. 247–251.
- HUANG, S.; ERNBERG, I.; KAUFFMAN, S. Cancer attractors: a systems view of tumors from a gene network dynamics. *Seminars in Cell & Developmental Biology*, v. 20, p. 869–876, 2009.
- HUANG, S.; INGBER, D. E. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Experimental cell research*, v. 261, p. 91–103, 2000.
- HUBBARD, P.; CAINES, P. E. Dynamical consistency in hierarchical supervisory control. *IEEE Transactions on Automatic Control*, v. 47, p. 37–52, 2002.
- IDEKER, T.; GALITSKI, T.; HOOD, L. A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics*, v. 2, p. 343–372, 2001.

- IGLESIAS, P. A.; INGALLS, B. P. *Control Theory and Systems Biology*. Cambridge, Massachusetts: The MIT Press, 2010.
- JONG, H. de. Modeling and simulation of genetic regulatory systems. *Journal of Computational Biology*, v. 9, p. 67–103, 2002.
- KARLEBACH, G.; SHAMIR, R. Modeling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, v. 9, p. 770–780, 2008.
- KARLSSON, M.; WEBER, W. Therapeutic synthetic gene networks. *Current Opinion in Biotechnology*, v. 23, p. 1–9, 2012.
- KARNAUGH, M. The map method for synthesis of combinatorial logic circuits. *Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics*, v. 72, p. 593–599, 1953.
- KAUFFMAN, S. A. *The origins of order: self-organization and selection in evolution*. Oxford: Oxford University Press, 1993.
- KELL, D. B.; KNOWLES, J. D. Systems modeling in cellular biology. In: SZALLASI, Z.; STELLING, J.; PERIWAL, V. (Ed.). Cambridge, Massachusetts: The MIT Press, 2010. cap. The role of modeling in systems biology.
- KEMMER, C. et al. Self-sufficient control of irate homeostasis in mice by a synthetic circuit. *Nature Biotechnology*, v. 28, p. 355–360, 2010.
- KHALIL, A. S.; COLLINS, J. J. Synthetic biology: applications come of age. *Nature Reviews, Genetics*, v. 11, p. 367 – 379, 2010.
- KITANO, H. Systems biology: a brief overview. *Science*, v. 295, p. 1662–1664, 2002.
- KUMAR, R.; GARG, V.; MARCUS, S. Language stability and stabilizability of discrete event dynamical systems. *SIAM J. Control Optm.*, v. 35, p. 1294–1320, 1993.
- LANGMEAD, C. J.; JHA, S. K. Symbolic approaches for finding control strategies in boolean networks. *Journal of Bioinformatics and Computational Biology*, v. 7, p. 323–338, 2008.
- LI, S.; BRAZHNIK, P.; TYSON, J. J. A quantitative study of the division cycle of caulobacter crescentus stalked cells. *PLoS Computational Biology*, v. 4, p. 111–129, 2008.

LIU, Y.-Y.; SLOTINE, J.-J.; BARABÁSI, A.-L. Controllability of complex networks. *Nature*, v. 473, p. 167–173, 2011.

MACHADO, D. et al. Modeling formalisms in systems biology. *AMB Express*, v. 1, p. 1–14, 2011.

MACLEAN, D.; STUDHOLME, D. A boolean model of the pseudomonas syringe hrp regulon predicts a tight regulated system. *PLoS One*, v. 5, p. 1–6, 2010.

MENOLASCINA, F.; BERNARDO, M. di; BERNARDO, D. di. Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory network. *Automatica*, v. 47, p. 1265–1270, 2011.

MESAROVIC, M. D.; TAKAHARA, Y. *Abstract Systems Theory*. Heidelberg: Springer-Verlag, 1989.

MILIAS-ARGEITIS, A.; SUMMERS, S.; LYGEROS, J. In silico feedback for in vivo regulation of a gene expression system. *Nature Biotechnology*, v. 29, p. 1114–1116, 2011.

MITCHELL, M. *Complexity: A Guided Tour*. Oxford: Oxford University Press, 2009.

NISSIM, L.; BAR-ZIV, R. A tunable dual-promoter integrator for targeting of cancer cells. *Molecular Systems Biology*, v. 6, p. 1–9, 2010.

NISSIM, L.; BEATUS, T.; BAR-ZIV, R. An autonomous system for identifying and governing a cell's state in yeast. *Physical biology*, v. 4, p. 154–164, 2007.

PURNICK, P.; WEISS, R. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, v. 10, p. 410–422, 2009.

QUINE, W. V. The problem of simplifying truth functions. *The American Mathematical Monthly*, v. 59, p. 521–521, 1952.

RAISCH, J.; O'YOUNG, S. D. Discrete approximation and supervisory control of continuous systems. *IEEE Transactions on Automatic Control*, v. 43, p. 569–573, 1998.

RAMADGE, P. J.; WONHAM, W. M. Supervisory control of a class of discrete-event processes. *SIAM Journal Control and Optimization*, v. 25, p. 475–498, 1987.

ROSEN, R. *Life Itself - A Comprehensive Inquiry into the Nature, Origin and Fabrication of Life*. New York: Columbia University Press, 1991.

SAADATPOUR, A.; ALBERT, I.; ALBERT, R. Attractor analysis of asynchronous boolean models of signal transduction networks. *Journal of Theoretical Biology*, v. 266, p. 641–656, 2010.

SAMAD, H. E. et al. Stochastic modelling of gene regulatory networks. *International Journal of Robust Nonlinear Control*, v. 15, p. 691–711, 2005.

SCHLITT, T.; BRAZMA, A. Current approaches to gene regulatory network modelling. *BMC Bioinformatics*, v. 8, 2007.

SHMULEVICH, I.; DOUGHERTY, E. R.; ZHANG, W. From boolean to probabilistic boolean networks as models of genetic regulatory networks. *Proceedings of the IEEE*, v. 90, p. 1778–1792, 2002.

SLEIGHT, S. C. et al. Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering*, v. 4, p. 12, 2010.

STROGATZ, S. H. *Nonlinear Dynamics and Chaos*. Cambridge, MA: Westview Press, 2001.

TARSKI, A. A lattice-theoretical fixpoint theorem and its applications. *Pacific J. Math.*, v. 5, n. 2, p. 285–309, 1955.

THOMAS, R. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, v. 42, p. 563–585, 1973.

UHLENDORF, J.; MIERMONT, A.; DELAVEAU, T. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences of the United States of America*, v. 109, p. 14271–14276, 2012.

WEBER, W.; FUSSENEGGER, M. Engineering of synthetic mammalian gene networks. *Cell: Chemistry & Biology*, v. 16, p. 287–297, 2009.

WOLKENHAUER, O. Systems biology: the reincarnation of systems theory applied in biology? *Briefings in Bioinformatics*, v. 2, p. 258–270, 2001.

WOLKENHAUER, O.; MESAROVIC, M.; WELLSTEAD, P. A plea for more theory in molecular biology. In: BRINGMANN, P. et al. (Ed.). *Systems Biology*. [S.l.]: Springer Berlin Heidelberg, 2007, (Ernst Schering Foundation Symposium Proceedings).

WONHAM, W. M. *Supervisory Control of Discrete-Event Systems*. 2010. Apostila baseada em notas de aula do autor.

XIE, Z. et al. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science*, v. 333, p. 1307–1311, 2011.

ZHONG, H.; WONHAM, W. M. On the consistency of hierarchical supervision in discrete-event systems. *IEEE Transactions on Automatic Control*, v. 35, p. 1125–1134, 1990.