



## REDES BOOLEANAS: UMA FERRAMENTA PARA O ESTUDO DA REGULAÇÃO GÊNICA

Herman Fialho Fumiã,  
Luciano Dias de Sousa,  
Flávio Aparecido de Almeida,  
Wiliam Romualdo Vaz de Souza Leal,  
Lucas Bocard Cancela

**RESUMO** :A regulação gênica é um fenômeno que envolve a interação de múltiplos elementos concomitantemente. Uma possível abordagem para o estudo desse fenômeno são os modelos booleanos de redes de regulação. O objetivo deste artigo é apresentação dos conceitos introdutórios relacionados a essa categoria de modelo. Dessa forma, apresentamos conceitos básicos sobre a regulação gênica e sobre a modelagem de redes genéticas por meio de grafos direcionados. Apresentamos também dois tipos de regras utilizadas para estudar a dinâmica da regulação. Finalmente, explicamos os resultados tipicamente produzidos por esses modelos e os interpretamos de acordo com a hipótese fundamental de Kauffman.

**Palavras-chave:** Regulação gênica. Redes booleanas. Kauffman.

### 1 INTRODUÇÃO

Os genes são partes do DNA onde se encontram codificadas informações que podem ser utilizadas para a produção de uma ou mais proteínas. Quando essas informações estão sendo “lidas” e “executadas” dizemos que o gene está sendo expresso. O conjunto de todos os genes sendo expressos em uma dada célula em um determinado momento é denominado de padrão de expressão gênica (ALBERTS et al., 2013).

Tipos celulares distintos possuem os mesmos genes, mas exibem diferentes padrões de expressão gênica. Alterações no ambiente em que uma célula se encontra também podem ocasionar uma mudança no padrão de expressão dos genes. A diminuição dos níveis normais de oxigênio, por exemplo, leva à inibição e à ativação da expressão de vários genes relacionados à adaptação a esse novo ambiente (MUZ *et al.*, 2015).

A flexibilidade no padrão de expressão é possível devido à existência de mecanismos de regulação. Entre a leitura da informação em um gene até a síntese da proteína (ou proteínas) associada ao mesmo, existem vários pontos nos quais a expressão pode



ser regulada. No entanto, como a síntese proteica é energeticamente dispendiosa, o principal ponto de regulação ocorre no início do processo, ou seja, na leitura da informação. Essa etapa de leitura denomina-se transcrição gênica (ALBERTS et al., 2013).

A transcrição inicia-se quando uma enzima denominada RNA polimerase se liga a uma região do DNA denominada promotor que fica próxima ao gene que vai ser transcrito. A RNA polimerase possui pouca afinidade com o promotor, o que significa que essa ligação ocorre a uma taxa muito baixa. É aqui que proteínas denominadas de fatores de transcrição entram em cena; elas são capazes de alterar a taxa com a qual a RNA polimerase se liga ao promotor. Essa alteração pode ser tanto um aumento quanto uma diminuição. No primeiro caso, dizemos que o fator ativa (ou induz) e, no segundo, que ele inibe a transcrição (ALBERTS et al., 2013).

Cada gene é regulado por um conjunto variável tanto em tipos quanto em números de fatores de transcrição. Genes responsáveis por processos celulares essenciais, por exemplo, podem estar sujeitos a uma regulação complexa que requer a ação de inúmeros fatores de transcrição (ALBERTS et al., 2013).

A complexidade da regulação gênica aumenta consideravelmente quando consideramos vários genes. Tomemos um determinado fator de transcrição  $a$ , por exemplo, que regule  $n$  genes. Como os fatores de transcrição são proteínas,  $a$  está presente na célula quando o gene  $A$  que o codifica está expresso. Ora, mas o gene  $A$  também é regulado por fatores de transcrição que, por sua vez, são produzidos por outros genes também regulados por fatores de transcrição. Dessa forma, podemos pensar no genoma como uma grande rede de genes que se conectam através de interações regulatórias mediadas pelos fatores de transcrição. Esse entendimento está na base da modelagem das redes de regulação.

Nesse artigo, apresentamos alguns conceitos fundamentais para a compreensão da modelagem da regulação gênica via modelo de redes booleanas. O modelo de redes



booleanas é uma abordagem simples para o estudo das redes de regulação. Ele exige poucas informações (informações experimentais qualitativas costumam ser suficientes), é de fácil implementação computacional e fácil de ser analisado mesmo para grandes sistemas. Além disso, os resultados obtidos geralmente permitem a obtenção de várias informações úteis sobre a dinâmica do sistema (RIVES; GALITSKI, 2003).

Inicialmente, discutimos a representação das redes genéticas por meio de grafos. Posteriormente, discutimos a aproximação booleana para o estado de ativação de um dado gene e duas regras ou funções de atualização que podem ser utilizadas para modelar a interação regulatória. Finalmente, discutimos o significado biológico dos atratores; estruturas que aparecem como resultado da dinâmica desse tipo de modelo.

## 2 MODELANDO REDES GENÉTICAS

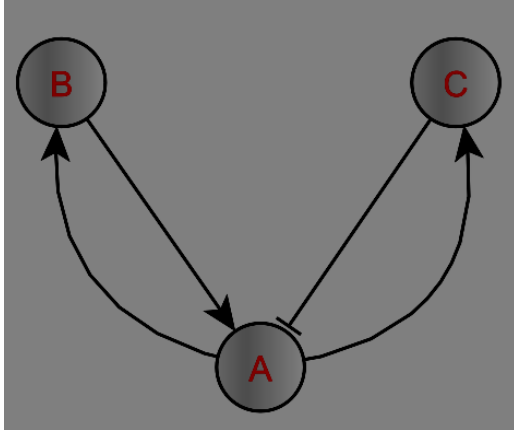
Suponha que tenhamos uma pequena rede de regulação formada pelos genes  $A$ ,  $B$  e  $C$  que codificam os fatores de transcrição  $a$ ,  $b$  e  $c$ , respectivamente. Digamos que o fator  $a$  ative os genes  $B$  e  $C$  e que o gene  $A$  seja ativado pelo fator  $b$  e inibido pelo fator  $c$ . Uma forma de representar essa rede é através de um grafo direcionado, ferramenta muito utilizada na ciência da computação.

Um grafo direcionado consiste em um conjunto  $N$  de nós (ou vértices) e um conjunto  $A$  de arestas ou ligações. Em nosso caso, o conjunto de nós é o conjunto dos genes  $N = \{A, B, C\}$ . As arestas correspondem às interações regulatórias entre os genes. Por exemplo, para o gene  $B$  que sintetiza o fator de transcrição  $b$  que ativa o gene  $A$ , temos a aresta  $(B, A, +)$ . A primeira coordenada indica qual gene é o regulador, a segunda, qual gene é o regulado e, a terceira, se a interação é ativadora ou inibidora. Portanto, o conjunto das arestas no presente caso é  $A = \{(B, A, +), (C, A, -), (A, B, +), (A, C, +)\}$ . O grafo pode ser representado como na Figura 1, na qual os nós são representados por círculos e as arestas por linhas com setas ou linhas



perpendiculares na extremidade próxima aos nós que estão sendo regulados. As setas representam ativação e as linhas perpendiculares, inibição.

**Figura 1: Rede de regulação.** Os círculos representam os nós (genes), as arestas (interações regulatórias) são linhas com setas (+) ou linhas perpendiculares (-) na extremidade próxima aos nós que estão sendo regulados.



### 3 REDES BOOLEANAS

A concentração celular de uma dada proteína  $[P]$  é uma medida da intensidade da expressão do gene que a codifica<sup>1</sup>. Essa concentração está diretamente associada à concentração dos fatores de transcrição que regulam o gene em questão. Como estas duas grandezas estão relacionadas?

<sup>1</sup>Desde que o principal ponto de regulação seja a transcrição.



Existe uma extensa documentação experimental apontando que a intensidade da expressão gênica pode se relacionar de forma sigmoideal com a concentração dos fatores de transcrição (GJUVSLAND; PLAHTÉ; OMHOLT, 2007). Os modelos matemáticos sugerem que em situações em que há cooperatividade entre diferentes fatores de transcrição (BINTU *et al.*, 2005), em que os fatores de transcrição do mesmo tipo formam dímeros, trímeros ou multímeros, ou em que os fatores de transcrição são rapidamente degradados (BOWER; BOLOURI, 2004), essa resposta sigmoideal é observada. Acredita-se que essas situações sejam comuns em organismos eucariotos.

Na Figura 2, vemos que uma função sigmoideal pode ser aproximada por uma função do tipo degrau, o que é denominado de aproximação booleana. Aceitando-a, um gene  $i$  pode ser representado por um elemento binário  $\sigma_i$  que está no estado 1 quando o gene está expresso e no estado 0 quando não está. Conseqüentemente, uma rede de regulação gênica com  $N$  genes torna-se um conjunto  $\{\sigma_1, \sigma_2, \dots, \sigma_N\}$  de  $N$  elementos binários interconectados.

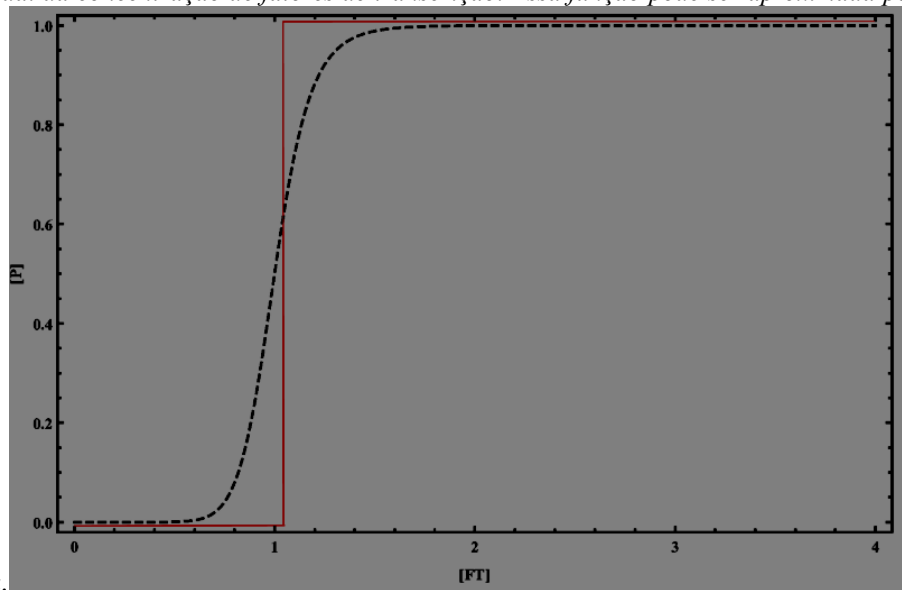
Em uma rede de regulação, cada gene  $i$  possui um número  $k_i$  de outros genes que o regulam, número que depende da topologia da rede em questão. A dinâmica da rede é obtida pela atualização simultânea de todos os genes da rede em um tempo  $t+1$  de acordo com

$$\sigma_i(t + 1) = F_i(\sigma_{i1}(t), \sigma_{i2}(t), \sigma_{i3}(t), \dots, \sigma_{iN}(t))$$

em que  $F_i$  é uma regra específica para o gene  $i$  que combina todos os seus reguladores.



**Figura 2: Aproximação booleana.** A concentração de uma determinada proteína (intensidade de expressão do gene) é uma função sigmoidal da concentração de fatores de transcrição. Essa função pode ser aproximada por



uma função do tipo degrau.

Para um gene  $i$  com  $k_i$  reguladores existem  $2^{2^{k_i}}$  regras lógicas. Portanto, o espaço de regras possíveis é enorme mesmo para pequenos valores de  $k_i$ . Salvo alguns poucos casos, não se pode determinar uma única regra adequada para um gene, mas sim um conjunto delas. Isso ocorre porque as redes biológicas ainda não são muito bem



compreendidas e os dados disponíveis são limitados (HINKELMANN; JARRAH, 2012).

Várias hipóteses adicionais são propostas para limitar o conjunto de regras possíveis. Exemplos de conjuntos de regras bastante utilizados, que derivam da adoção de determinadas hipóteses, são as regras canalizadoras e as regras do tipo limiar.

### 3.1 Regras canalizadoras

Na Biologia, o termo canalização refere-se a estabilidade fenotípica dos organismos às mutações e às alterações ambientais<sup>2</sup>. Em outras palavras, as células dos organismos conseguem se comportar da mesma forma apesar de as perturbações externas e internas. As regras ditas canalizadoras produzem uma dinâmica estável no que diz respeito às alterações nos genes da rede, o que justifica sua denominação.

Seja  $i$  um gene regulado por  $k$  outros genes. Seja  $j$  um gene pertencente ao conjunto de reguladores de  $i$ . Se o gene  $j$  no estado  $a$  determina que o gene  $i$  está no estado  $b$  independentemente do estado dos demais reguladores de  $i$ , temos uma regra canalizadora. Denominamos o gene  $j$  de gene canalizador e dizemos que a regra possui valor canalizador  $a$  e saída canalizada  $b$ .

Seja  $A$  um gene regulado por dois outros genes  $B$  e  $C$ . Existem 16 regras lógicas possíveis (Figura 3), mas quais destas são canalizadoras? As regras podem ser canalizadoras com valor canalizador 0 ou 1 e com relação aos genes  $B$  ou  $C$ . Para exemplificar, tomemos  $C$  como gene canalizador e o valor canalizador como 1. Note que se  $\sigma_C = 1$ , as Regras 13, 15 e 16 determinam que  $\sigma_A = 1$ , e as Regras 1, 2, 3 e 4 determinam que  $\sigma_A = 0$ , independentemente do estado do gene  $B$ . Logo, todas são regras canalizadoras, sendo que as Regras 13, 15 e 16 possuem saída canalizada 1, e as Regras 1, 2, 3 e 4 possuem saída canalizada 0.

---

<sup>2</sup>Essa estabilidade se deve a mecanismos como redundância genética, regulação por *feedback*, e interações bioquímicas cooperativas.




**Figura 3: Regras Lógicas.** Todas as 16 regras possíveis para um gene *A* que é regulado por dois outros genes *B* e *C*.

<i>B</i>	<i>C</i>	Regra 1	Regra 2	Regra 3	Regra 4	Regra 5	Regra 6	Regra 7
0	0	0	1	0	1	0	1	0
1	0	0	0	1	1	0	0	1
0	1	0	0	0	0	1	1	1
1	1	0	0	0	0	0	0	0

Regra 8	Regra 9	Regra 10	Regra 11	Regra 12	Regra 13	Regra 14	Regra 15	Regra 16
1	0	1	0	1	0	1	0	1
1	0	0	1	1	0	0	1	1
1	0	0	0	0	1	0	1	1
0	1	1	1	1	1	1	1	1

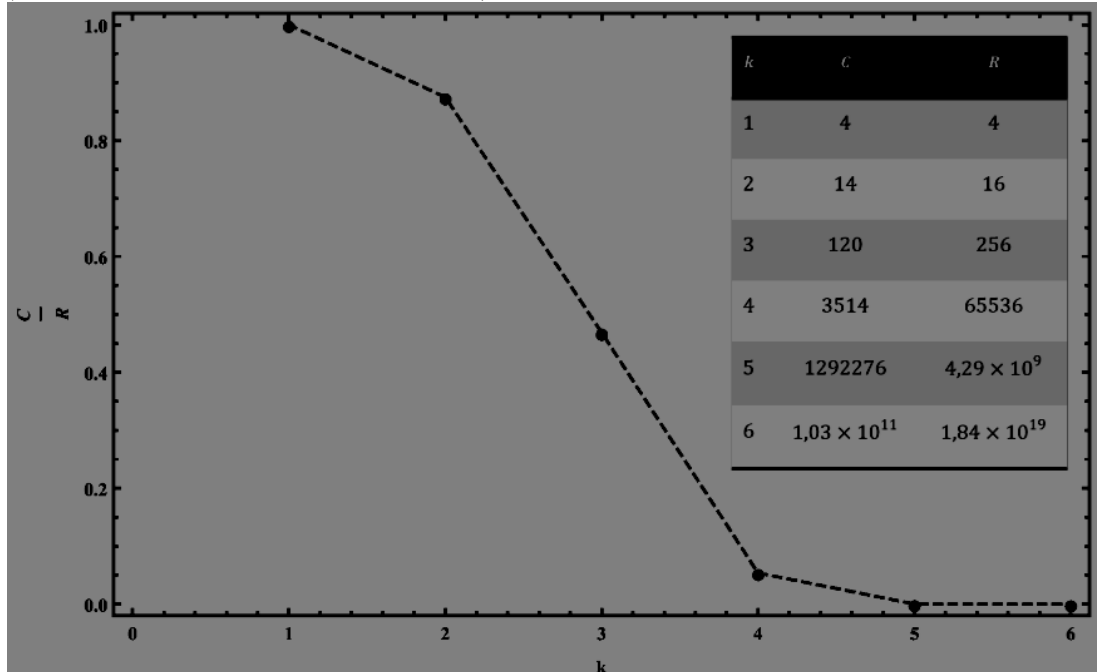
Pode-se mostrar que 14 das 16 regras lógicas são canalizadoras. Embora neste caso a maior parte das regras sejam canalizadoras, a razão entre o número de regras canalizadoras *C* e o número de regras possíveis *R* diminui rapidamente com o número de reguladores *k* do gene (Figura 4).

Em uma rede real os genes frequentemente possuem vários reguladores. Se um gene com *k* reguladores tem a dinâmica controlada por uma regra canalizadora, alterações em qualquer um dos  genes, com exceção do gene canalizador, produzem a mesma saída canalizada. Portanto, quanto maior o número de genes associados a regras canalizadoras, maior a estabilidade da mesma.





**Figura 4: Número de Regras Canalizadoras.** O número de regras canalizadoras  $C$  aumenta com o número  $k$  de reguladores do gene. Contudo, a razão entre o número de regras canalizadoras  $C$  e o número total de regras  $R$  diminui bruscamente. A informação sobre o número de regras canalizadoras até  $k = 10$  pode ser obtida em (JUST; SHMULEVICH; KONVALINA, 2004).



### 3.2 Regras do tipo limiar

A hipótese central das regras do tipo limiar é que a interação entre os distintos fatores de transcrição na regulação de um dado gene é do tipo aditiva. Um gene  $j$  que regula um gene  $i$ , o faz com um determinado peso  $J_{ji}$ . Se o gene  $j$  ativa a expressão do gene  $i$  temos que  $J_{ji} > 0$ . Se ele a inibe, temos que  $J_{ji} < 0$ . O estado final do gene  $i$  é determinado somando as contribuições de todos os seus reguladores. As regras do tipo limiar possuem a seguinte forma funcional:





Portanto, o estado de ativação de um gene  $i$  é determinado ao se comparar o estímulo total dado por todos os genes que o regulam com o limiar de ativação  $\theta_i$  do gene em questão.

Certamente, um dos pontos fortes das regras do tipo limiar é a descrição das interações regulatórias como ativadoras ou inibidoras, o que a aproxima da nomenclatura biológica. Apesar da sua hipótese central restritiva, as regras do tipo limiar foram utilizadas com sucesso para descrever a sequência biológica correta dos eventos no ciclo celular das leveduras *S. cerevisiae*<sup>3</sup> (LI *et al.*, 2004) e *S. pombe*<sup>4</sup> (BORNHOLDT, 2008). Esses resultados sugerem que, pelos menos em determinados casos, as interações regulatórias podem ser consideradas como puramente aditivas (KAUFFMAN, 1993).

## 4 A DINÂMICA DOS MODELOS BOOLEANOS

### 4.1 Os atratores

Seja uma rede com  $N$  nós que está sujeita a uma determinada regra de atualização. Sendo  $\sigma_i$  o estado de ativação (0 ou 1) do nó  $i$ , dizemos que uma configuração é um conjunto  $\langle \sigma_1, \sigma_2, \dots, \sigma_N \rangle$  com  $N$  elementos que indica o estado de todos os nós da rede. Ao conjunto de todas as configurações possíveis ( $2^N$  configurações), denominamos de espaço das configurações.

Se escolhermos uma configuração  $A$  aleatória e aplicarmos a regra de atualização, uma nova configuração  $B$  é produzida<sup>5</sup>. Se tornarmos a aplicar a regra na configuração  $B$ , produziremos uma outra configuração  $C$ . Ao executarmos esse procedimento inúmeras vezes estaremos gerando “caminhos” ou “fluxos” no espaço das configurações. Entretanto, como o espaço de configurações é finito, o fluxo das

---

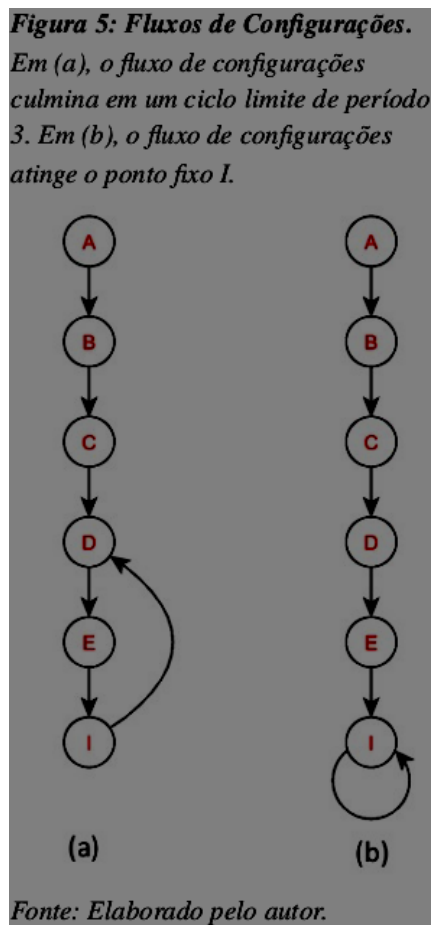
<sup>3</sup>*Saccharomyces cerevisiae* é um fungo unicelular. É utilizado na produção de pão, cerveja e etanol.

<sup>4</sup>*Schizosaccharomyces pombe* também é um fungo unicelular. É utilizado como organismo modelo em biologia molecular e celular.

<sup>5</sup>A não ser no caso de escolhermos justamente um ponto fixo.



configurações também será. Em algum momento, a aplicação da regra de atualização produzirá uma configuração que já foi produzida. Denominemos a configuração antecedente à configuração já produzida de I. Existem apenas duas possibilidades para a configuração sucessora de I; ou ela é uma configuração anterior pertencente a algum ponto do fluxo (por exemplo, a configuração D na Figura 5) ou ela é a própria configuração I.



Observe que no caso da atualização continuar a

No caso da Figura 5(b), a configuração produzida é sempre a I.

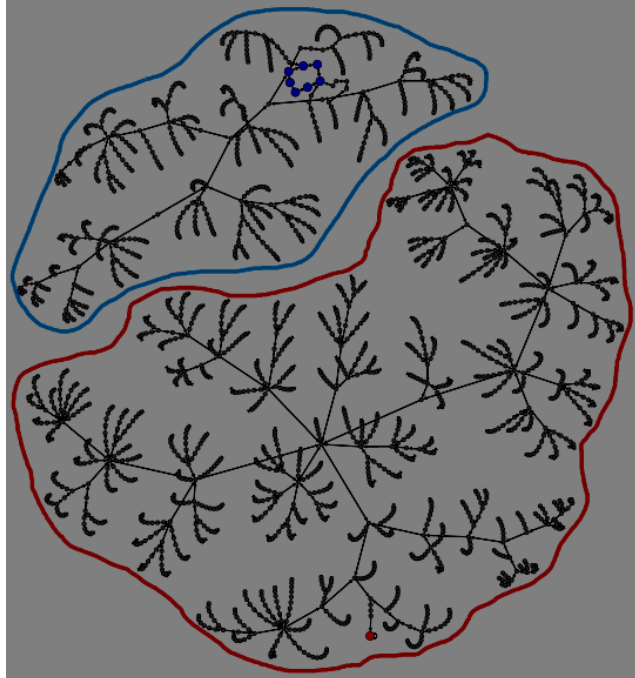
Ao conjunto de configurações para o qual a dinâmica do sistema converge denominamos de atrator. O período de um atrator corresponde ao número de configurações que ele possui. Atratores que possuem período maior que 1, como o da



Figura 5(a), que possui período igual a 3, são denominados de ciclos limites. Atratores de período 1 são denominados pontos fixos (como o da Figura 5(b)).

Distintamente do caso ilustrativo anterior, o espaço de configurações pode ser muito complexo. Em geral, existem vários fluxos de configurações que convergem para um ou mais atratores. A Figura 6 ilustra um caso em que o espaço de configurações possui dois atratores. Um deles é um ciclo limite de período 7 (configurações em azul) e, o outro, um ponto fixo (configuração em vermelho).

*Figura 5: Bacias de atração. No espaço de configurações, conjuntos de configurações são atraídos pelos atratores. Denominamos esses conjuntos de bacias de atração.*



Observe que cada atrator possui um determinado número de configurações as quais ele atrai. Denominamos esses conjuntos de configurações atraídas de bacias de atração. O tamanho da bacia de atração de um atrator está diretamente relacionado à sua importância na dinâmica do sistema. Quanto maior for, maior a probabilidade de que a dinâmica culmine no dito atrator.

#### 4.2 As células “funcionam” nos atratores



Uma configuração é um conjunto ~~de~~ com  $N$  elementos que indica o estado de todos os nós da rede. Em outras palavras, uma configuração representa um determinado padrão de expressão gênica. Portanto, um atrator com  $p$  configurações (período  $p$ ) representa um conjunto limitado de padrões de expressão gênica que se repetem ciclicamente.

Os atratores assemelham-se ao que se observa nas células. As células são sistemas estáveis, caso não haja modificações nas condições ambientais ou mutações genéticas, o seu comportamento não se altera, ou seja, o seu padrão de expressão.

Essas considerações levaram Kauffman a admitir como hipótese que as células operam nos atratores do sistema (KAUFFMAN, 1993). Portanto, um atrator representa um determinado comportamento celular. Dessa forma, espera-se que uma célula em ciclo esteja em um atrator distinto de uma célula que não está se dividindo, por exemplo.

Independentemente da rede de regulação ou da regra adotada, todos os modelos booleanos produzem atratores no espaço das configurações. Contudo, o número, o tamanho da bacia de atração, e o período dos atratores produzidos dependem tanto da rede quanto da regra. Essas quantidades são importantes na análise da adequabilidade do modelo para descrever um problema em questão. É razoável esperar, por exemplo, que o atrator com a maior bacia de atração descreva um comportamento bem conhecido do sistema, já que, segundo a hipótese de Kauffman, é o comportamento celular mais provável.

## 5 CONCLUSÃO

Este artigo pretende ser um compilado dos conceitos introdutórios necessários para aqueles que pretendam estudar a modelagem da regulação gênica via redes booleanas. Acreditamos que isso possa ser útil principalmente para os profissionais das áreas biológicas que, embora conheçam sobremaneira os mecanismos de



regulação, não estão familiarizados com as ferramentas e conceitos matemáticos utilizados nesse tipo de modelo.

O conceito de modelagem das redes de regulação via grafos, ferramenta típica da ciência da computação e da matemática, foi apresentado. A aproximação booleana, hipótese fundamental para o tipo de abordagem que estamos utilizando, também foi apresentada.

Discutimos duas regras bastante utilizadas nesse tipo de modelo: as regras ditas canalizadoras e do tipo limiar. As regras canalizadoras, são inspiradas na estabilidade fenotípica dos organismos às mutações e às alterações ambientais. As regras do tipo limiar são mais simples e foram aplicadas com sucesso em determinados modelos.

Por último, discutimos o conceito de atratores; que são conjuntos de configurações a cuja dinâmica desse tipo de modelo leva. Aliado a isso, discutimos a hipótese de Kauffman que dá, por assim dizer, significado biológico aos resultados exibidos.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

ALBERTS, Bruce et al. **Essential cell biology**. Garland Science, 2013.

BINTU, L. et al. **Transcriptional regulation by the numbers: applications**. Current opinion in genetics & development, v. 15, n. 2, p. 125-135, 2005.

BORNHOLDT, S.. **Boolean network models of cellular regulation: prospects and limitations**. Journal of the Royal Society Interface, v. 5, n. suppl\_1, p. S85-S94, 2008.

BOWER, J. M.; BOLOURI, H. (Ed.). **Computational modeling of genetic and biochemical networks**. MIT press, 2004.

GJUVSLAND, A. B; PLAHTÉ, E.; OMHOLT, S. W. **Threshold-dominated regulation hides genetic variation in gene expression networks**. BMC Syst Biol. 2007;1:57.

HINKELMANN, F.; JARRAH, A. S. **Inferring biologically relevant models: nested canalizing functions**. ISRN Biomathematics, 2012.



JUST, W.; SHMULEVICH, I.; KONVALINA, J. **The number and probability of canalizing functions.** Physica D: Nonlinear Phenomena, v. 197, n. 3-4, p. 211-221, 2004.

LI, F. et al. **The yeast cell-cycle network is robustly designed.** Proceedings of the National Academy of Sciences, 101(14), 4781-4786, 2004.

KAUFFMAN, S. A. **The origins of order: Self-organization and selection in evolution.** OUP USA, 1993.

MUZ, B., de la PUENTE, P., AZAB, F., & AZAB, A. K. (2015). **The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy.** Hypoxia (Auckland, N.Z.), 3, 83–92. doi:10.2147/HP.S93413.

RIVES, A. W.; GALITSKI, T. **Modular organization of cellular networks.** Proceedings of the National Academy of Sciences, v. 100, n. 3, p. 1128-1133, 2003.