UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

Shantanu Gupta

MODELAGEM COMPUTACIONAL DE REDES GENÉTICAS REGULATÓRIAS

Santa Maria, RS
2016

Shantanu Gupta

MODELAGEM COMPUTACIONAL DE REDES GENÉTICAS REGULATÓRIAS

Dissertação de Mestrado apresentado ao Programa de Pós-Graduação em Física, Área de Concentração em Sistemas Complexos, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção Obtenção do grau de **Mestre em Física.**

ORIENTADOR: Prof. José Carlos Merino Mombach

Santa Maria, RS
2016

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

Gupta, Shantanu
MODELAGEM COMPUTACIONAL DE REDES GENÉTICAS REGULATÓRIAS / Shantanu Gupta.- 2016.
56 p.; 30 cm

Orientador: José Carlos Merino Mombach
Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Ciências Naturais e Exatas, Programa de Pós-Graduação em Física, RS, 2016

1. Rede Regulatória 2. Redes Regulatória de Genes 3.Expressão Gênica 4. Astrócito 5. Sinergia de Droga I. Mombach, José Carlos Merino II. Título.

Endereço: Rua Appel, nº 431, apto 203

End. Eletr.: shantanubrasil@gmail.com

Shantanu Gupta

MODELAGEM COMPUTACIONAL DE REDES GENÉTICAS REGULATÓRIAS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física, Área de Concentração em Sistemas Complexos, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Mestre em Física.**

Aprovado em 30 de setembro de 2016:

José Carlos Merino Mombach, Dr. (UFSM)
(Presidente/Orientador)

Solange Binotto Fagan, Dr. (Unifra)

Otávio Costa Acevedo, Dr. (UFSM)

Santa Maria, RS
2016

RESUMO

MODELAGEM COMPUTACIONAL DE REDES GENÉTICAS REGULATÓRIAS

AUTOR: Shantanu Gupta
ORIENTADOR: José Carlos Merino Mombach

Em biologia, redes regulatórias são conjuntos de macromoléculas, principalmente proteínas e RNAs que interagem para executar uma tarefa. As proteínas de ligação de DNA, também chamadas de fatores de transcrição, são as principais executoras nas redes regulatórias, visto que modulam o primeiro passo na expressão gênica. Uma rede genética regulatória (RRG) é um conjunto de genes ou proteínas que interagem uns com os outros para controlar uma função celular específica. Redes regulatórias são importantes no desenvolvimento, diferenciação e para responder aos sinais ambientais. Elas são os botões de liga/desliga de uma célula operando no nível do gene e/ou proteína. Seus métodos de modelagem podem ser geralmente classificados em contínuos e discretos. Neste trabalho, dedicamos atenção aos modelos discretos em senescência celular para astrócitos [35], a modelagem de sinergias de drogas para controle do câncer gástrico [38] e também escrevemos um artigo sobre Modelos Discretos e Contínuos, vantagens e desvantagens desses modelos e listagem dos softwares disponíveis para uso nesse tipo de abordagem.

Palavras-chave: Rede Regulatória, Redes Regulatória de Genes, Expressão Gênica, Astrócito, Sinergia de Droga.

ABSTRACT

COMPUTATIONAL MODELLING OF GENE REGULATORY NETWORKS

AUTHOR: Shantanu Gupta
ADVISOR: José Carlos Merino Mombach

In biology, regulatory networks are sets of macromolecules, mostly proteins and RNAs that interact to execute task. The main players in regulatory networks are DNA-binding proteins, also called transcription factors as they modulate the first step in gene expression. A gene regulatory network (GRN) is a set of genes or proteins that interact with each other to control a specific cell function. Gene regulatory networks are important in development, differentiation and to respond to environmental cues. Gene regulatory networks (GRNs) are the on-off switches of a cell operating at the gene and/or protein level. The modeling methods can be broadly categorized into continuous and discrete. In this work , we dedicate attention to discrete models on cell senescence models for Astrocyte [35], the modelling of drug synergies to control gastric cancer [38], and we also wrote a paper about Discrete and Continuous Model, advantage or disadvantage of these models and a list of available softwares for using these kind of approaches.

Keywords: Regulatory Networks, Gene Regulatory Network, Gene Expression, Astrocyte, Drug Synergies.

List of abbreviations

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
APO	Apoptosis
GE	Gene Expression
GRN	Gene Regulatory Network
MiRNAs	MicroRNAs
M	Mitosis
G1	Gap1
G2	Gap2
S	Synthesis phase
G0	Gap 0
ATP	Adenosine triphosphate.
ADP	Adenosine diphosphate.
RRG	Rede genética regulatória

SUMÁRIO

CAPÍTULO1: INTRODUÇÃO

Os grupos de genes, proteínas reguladoras e as suas interações são muitas vezes referidos como redes genéticas regulatórias (RRG). que interagem uns com os outros indiretamente (através do produto da expressão de RNA e proteína) e com outras moléculas na célula,. RRGs fornecem uma compreensão sistemática dos mecanismos moleculares dos processos biológicos subjacentes [1-7]. Os nós dessas redes são genes e as arestas entre os nós representam interações de genes através das quais o produto de um gene afeta o produto de outro, através de interações protéicas. Essas interações podem ser indutivas (representadas por setas), com o aumento na expressão de um levando ao aumento no outro, ou inibitória (representada por martelo), com o aumento em um que conduz a diminuição no outro. Uma série de arestas indica uma cadeia de tais dependências, com ciclos que correspondem a laços de feedback. Por exemplo, vamos considerar um sistema simples discreto de dois elementos (Figura 1). Suponha-se que o produto x ativa o gene Y e o produto y reprime o gene X. Em outras palavras:

X = 1 se y = 0 (X " ativado " se y ausente)
Y = 1 se e somente se x = 1 (Y “ativado" se x presente).

Figura 1. Exemplo de laços de feedback simples. O circuito é um laço negativo (ver abaixo).

Redes regulatórias de genes descrevem o controle no nível da expressão gênica e podem ser inferidas a partir de microRNAs (miRNAs). Padrões regulatórios, perfis de expressão e interações entre alvos regulatórios [8]. A regulação gênica tornou-se importante como informação de interações regulatórias moleculares, tornando-se cada vez mais disponíveis. A importância das redes regulatórias de genes é evidente para todas as espécies

e sistemas biológicos [9], visto que elas desempenham um papel importante na manutenção das funções biológicas do organismo vivo [10]. A inferência de reguladores é o fator central na interpretação das condições regulatórias reais em RRGs [11]. Isso proporciona um modelo mais claro sobre a relação entre os genes e proteínas alvos e os genes reguladores. Previsões precisas do comportamento de redes regulatórias também irão acelerar os projetos biotecnológicos e tais previsões são mais rápidas e baratas do que experimentos de laboratório.

Vários modelos computacionais desenvolvidos para a análise de redes regulatórias podem ser divididos em quatro classes (Figura 2). Um modelo completo de redes regulatórias incorpora o conhecimento experimental sobre os componentes e as suas interações, bem como o estado inicial desses componentes e levam ao conhecido estado final ou comportamento dinâmico da rede. Vemos na fig. 2 que os modelos podem ser classificados em dois tipos grosseiramente: discretos e contínuos, estes últimos baseados em equações lineares ou diferenciais (ver apêndice).

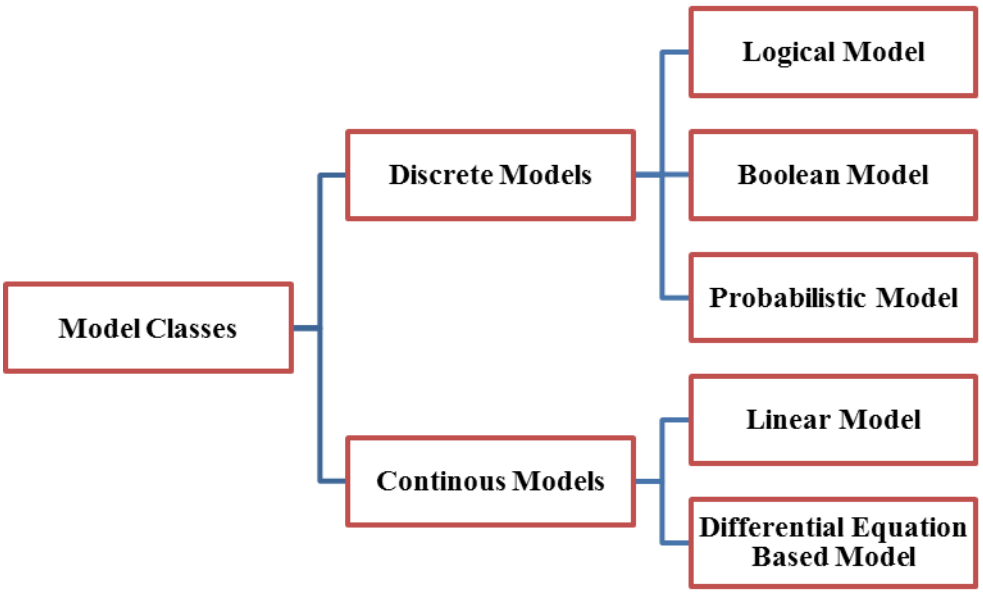


Figura 2. Classificação dos Modelos dividida em duas classes. (A) Modelos discretos, (B) Modelos contínuos.

1.1 Regulação vs. Fenótipo

É geralmente reconhecido que um fenótipo é uma propriedade emergente de interações genótipo-ambiente. Especificamente, um fenótipo resulta de padrões de atividade moleculares e celulares das interações genótipo-ambiente. Isso implica que cada fenótipo observável está associado a redes de genes específicas do fenótipo, pois sem alteração das interações moleculares um fenótipo não pode mudar. Isso significa que todas as alterações no nível do genótipo que irão resultar numa alteração do fenótipo também conduzirão inevitavelmente a uma mudança na estrutura da rede de genes como um mediador entre os dois níveis.

Além disso, uma rede regulatória de genes pode representar potencialmente diversos tipos de interações físicas e bioquímicas entre os genes e produtos de genes [12]. Pode-se esperar que redes regulatórias sejam altamente fenótipo específicas [13,14]. Estabelecer essas relações será, portanto, uma tarefa complexa, mas também oferece uma oportunidade para catalogar quantitativamente fenótipos. Um exemplo para a análise de redes específicas de tecido pode ser encontrado em Guan et al. [15], onde 107 redes específicas de tecido foram estudadas. Atualmente, o número de RRGs é difícil de estimar, mas com base nesses resultados preliminares, pode-se supor que há mais de 200 RRGs diferentes apenas para os seres humanos, uma vez que isso corresponde ao número de diferentes tipos de células. No entanto, células patológicas, tais como tumores, também têm suas próprias redes características [16] implicando provavelmente em milhares de diferentes redes de genes em seres humanos.

CAPÍTULO 2: CONCEITOS BIOLÓGICOS BÁSICOS

2.1 A célula

A célula foi descoberta por Robert Hooke, em 1665, que nomeou a unidade biológica por sua semelhança com as células habitadas por monges cristãos em um mosteiro. [1-2] A teoria celular, desenvolvida pela primeira vez em 1839 por Matthias Schleiden e Theodor Schwann, diz que todos os organismos são constituídos por uma ou mais células, que as células são a unidade fundamental da estrutura e função em todos os organismos vivos, que todas as células são provenientes de células pré-existentes, e que todas as células contêm a informação hereditária necessária para a regulação das funções celulares e para transmitir informações para a próxima geração de células. [3] As células surgiram na Terra, pelo menos 3,5 bilhões de anos atrás. [4-6]

A célula é a unidade básica estrutural, funcional e biológicas de todos os organismos vivos conhecidos. Uma célula é a menor unidade de vida que pode replicar independentemente, e as células são muitas vezes chamadas de "blocos de construção da vida" [1-2].

As células são de dois tipos: eucariontes, que contêm um núcleo, e procariontes, que não contêm (Figura 3).

Células procariontes: células procariontes foram a primeira forma de vida na Terra, caracterizadas por possuírem processos biológicos vitais incluindo a sinalização celular e serem autossustentáveis. Elas são mais simples e menores do que as células eucariontes, e carecem de organelas ligadas à membrana, tais como o núcleo. Os procariontes incluem dois dos domínios de vida: bactérias e archaea. O DNA de uma célula procarionte é constituído por um único cromossomo que está em contato direto com o citoplasma. A região nuclear no citoplasma é chamada de nucleóide. A maioria dos procariontes são os menores de todos os organismos, variando de 0,5 a 2,0 μm de diâmetro. [7]

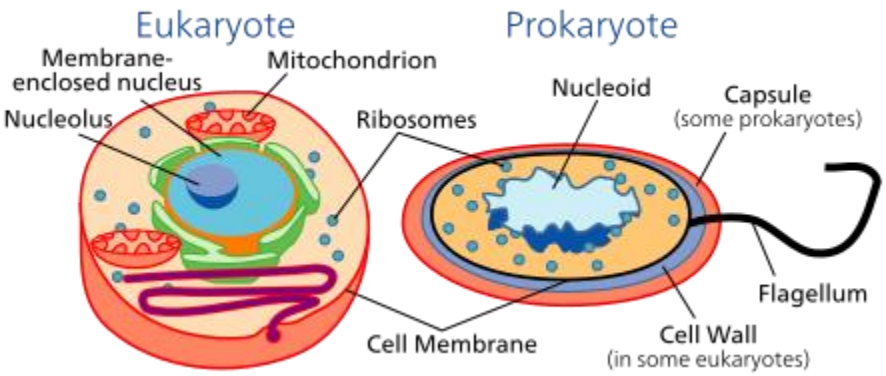


Figura 3. Célula eucarionte e célula procarionte. Figura extraída de https://commons.wikimedia.org/wiki/File:Celltypes.svg

Células eucariontes: Plantas, animais, fungos, moldes de limo, protozoários e algas são todos eucariontes. Estas células são cerca de quinze vezes mais largas do que uma procarionte típica e podem ser tanto quanto mil vezes maior em volume. A principal característica distintiva de eucariontes em comparação com procariontes é a compartimentalização: a presença de organelas ligada à membrana (compartimentos) em que as atividades metabólicas específicas ocorrem. A mais importante delas é o núcleo celular, uma organela que abriga o DNA da célula.

2.2 O DNA

O ácido desoxirribonucléico (DNA) é uma molécula que codifica o mapa genético de um organismo. Em outras palavras, o DNA contém toda a informação necessária para construir e manter um organismo. O DNA foi descoberto em 1868 quando o médico suiço de vinte e quatro anos de idade, Friedrich Miescher, isolou um composto do núcleo de células brancas do sangue. Este composto não era uma proteína, nem um lípido ou um hidrato de carbono, e foi, por conseguinte, um novo tipo de molécula biológica. Miescher nomeou sua descoberta de "nuclein", pois ele tinha a isolado a partir de núcleos de células. Hoje em dia, esta molécula é chamada DNA. Quase todas as células dentro de um único organismo têm exatamente o mesmo DNA. O DNA é uma molécula linear composta (Figura 4) de quatro tipos de moléculas químicas menores chamadas bases de nucleotídeos: adenina (A), citosina (C), guanina (G) e timina (T). A ordem destas bases é chamada de sequência do

DNA. Segmentos de DNA que carregam a informação genética são denominados genes, e eles são herdados dos pais durante a reprodução.

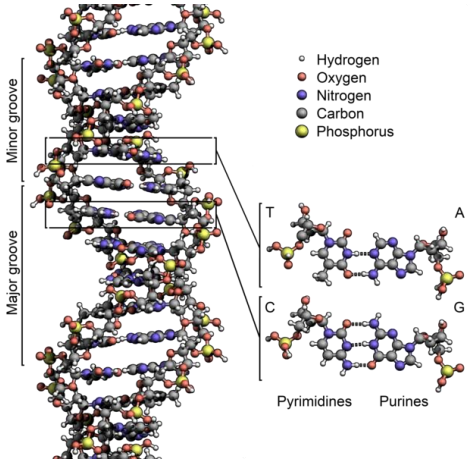


Figura 4. A estrutura de dupla hélice do DNA. Os átomos na estrutura estão em código de cor por elemento e a estrutura detalhada de dois pares de bases é mostrada na parte inferior direita. Figura extraída de https://en.wikipedia.org/wiki/File:DNA_Structure%2BKey%2BLabelled.pn_NoBB.png

Em 1953, Francis Crick e James Watson descreveram a forma molecular do DNA como uma "dupla hélice". O DNA de cadeia dupla é composto por duas cadeias lineares que rodam em frente umas às outras, conhecidas como filamentos antiparalelos; estas vertentes torcem juntas para formar uma dupla hélice. A estrutura do DNA pode também ser descrita como uma escada. As colunas químicas da escada são constituídas por moléculas de açúcar e fosfato, que estão ligados por ligações químicas. Os degraus da escada são pares de unidades entre A e T ou entre C e G. Estes pares são chamados de pares de bases e conectam as duas colunas de açúcar-fosfato por meio de interações chamadas ligações de hidrogênio.

No núcleo, o DNA forma um complexo com as proteínas. Este complexo é chamado de

cromatina e é formado quando o DNA se envolve em torno de proteínas nucleares e, em seguida, envolve-se várias vezes em torno de si mesmo para condensar o DNA em um volume menor. Além disso, os cromossomos do DNA são frequentemente reconhecidos e descritos como estruturas em forma de X. O DNA assume esta forma depois da replicação do DNA durante o processo de divisão celular, quando os cromossomos replicados estão altamente condensados e aparecem em forma de X.

2.3 Gene

Um gene é a unidade física e funcional fundamental da hereditariedade. Os genes, que são compostos de DNA, atuam como instruções para fazer moléculas chamadas proteínas que são as moléculas que realizam a maioria das funções biológicas celulares. Em humanos, os genes variam em tamanho desde poucas centenas de bases de DNA para mais de dois milhões de bases.

A maioria das características biológicas está sob a influência de poligenes (muitos genes diferentes), bem como as interações entre gene e ambiente. Alguns traços genéticos são imediatamente visíveis, como a cor dos olhos ou o número de membros, e alguns não são, como sanguíneo, riscos para doenças específicas, ou os milhares de processos bioquímicos básicos que compõem a vida.

Os genes podem adquirir mutações na sua sequência, conduzindo a diferentes variantes conhecidas como alelos. Esses alelos codificam versões ligeiramente diferentes de uma proteína, que causam diferentes características fenotípicas. O uso coloquial do termo "possuindo um gene" (por exemplo, "bons genes", "gene da cor de cabelo") refere-se tipicamente a ter um alelo diferente do gene. Genes evoluem devido à seleção natural ou a sobrevivência do mais apto dos alelos.

O conceito de um gene continua a ser refinado à medida que novos fenômenos são descobertos. [8] Por exemplo, as regiões reguladoras de um gene podem ser muito longe das suas regiões de codificação, e regiões codificadoras podem ser divididas em vários éxons (ver adiante). Alguns vírus armazenam seu genoma no RNA em vez do DNA e alguns produtos de genes são RNAs funcionais não-codificantes, isto é, não codificam uma

proteína. Portanto, uma ampla definição funcional e moderna de um gene é qualquer locus discreto da sequência hereditária e genômica que afeta as características de um organismo ao ser expressa como um produto funcional ou pela regulação da expressão do gene.

2.4 Proteínas

As proteínas são grandes biomoléculas, ou macromoléculas, que consistem em uma ou mais cadeias longas de resíduos de aminoácidos. Proteínas executam uma vasta gama de funções dentro de organismos vivos, incluindo catalisação de reações metabólicas, replicação do DNA, resposta a estímulos e o transporte de moléculas a partir de um local para outro. As proteínas diferem umas das outras principalmente nas suas sequências de aminoácidos, que são ditadas pela sequência de nucleotídeos dos genes, e que geralmente resultam na proteína se dobrando numa estrutura tridimensional específica que determina a sua atividade. Uma cadeia linear de resíduos de aminoácidos é chamada de polipeptídio. Uma proteína contém pelo menos um polipeptídio de comprimento. Polipeptídios curtos, contendo menos de 20-30 resíduos, raramente são considerados como proteínas e são comumente chamados de peptídeos.

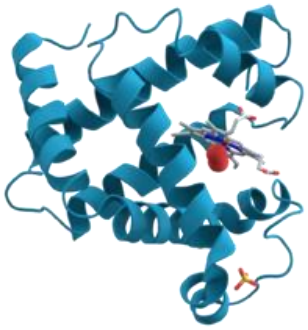


Figura 5. A representação da estrutura 3D da proteína mioglobina mostrando hélices α (turquesa). Esta figura extraído de
https://commons.wikimedia.org/wiki/File:Myoglobin.png

2.5 RNA :

O ácido ribonucléico (RNA) é uma molécula polimérica implicada em vários papéis biológicos de codificação, descodificação, regulação e expressão de genes. RNA e DNA são ácidos nucléicos e, juntamente com as proteínas e os hidratos de carbono, constituem as três principais macromoléculas essenciais para todas as formas de vida conhecidas. Assim como o DNA, o RNA é constituído por uma cadeia de nucleotídeos, mas ao contrário do DNA, é mais frequentemente encontrado na natureza como uma única cadeia dobrada sobre si mesma, em vez de uma dupla cadeia emparelhada. Organismos celulares utilizam RNA mensageiro (RNAm) para transmitir informação genética (utilizando as letras G , U, A e C para denotar a bases nitrogenadas guanina , uracila , adenina e citosina) que dirige a síntese de proteínas específicas. Muitos vírus codificam a informação genética usando um genoma de RNA.

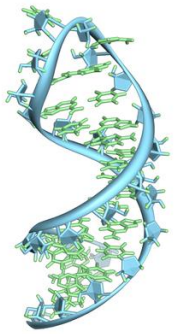


Figura 6. A estrutura de um pré RNAm. Destacam-se as nucleobases (verde) e a coluna de ribose-fosfato (azul). Este valor extraído https://commons.wikimedia.org/wiki/File:Pre-mRNA-1ysv-tubes.png

2.6 O Ciclo Celular

O ciclo celular representa a série de eventos que acontecem numa célula levando à sua divisão e duplicação. Em células sem um núcleo (procariontes), o ciclo celular ocorre através de um processo denominado fissão binária. Nas células com um núcleo

(eucariontes), o ciclo celular pode ser dividido em dois períodos breves: intérfase – durante a qual a célula cresce, acumulando nutrientes necessários para a mitose e duplicando o seu DNA – e a mitose (fase M), durante a qual a célula se divide em duas células distintas, muitas vezes chamadas de "células-filhas". O ciclo de divisão celular é um processo fundamental pelo qual um ovo unicelular fertilizado se desenvolve em um organismo maduro, bem como o processo pelo qual o cabelo, pele, células do sangue e alguns órgãos internos são renovados. [9]

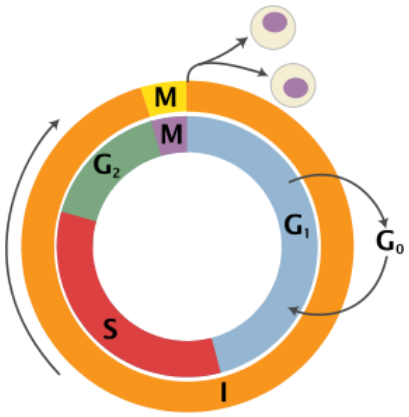


Figura 7. Representação esquemática do ciclo celular. Anel externo: I = Intérfase, M = mitose; anel interno: M = mitose, G1 = Gap 1, G2 = Gap 2, S = Síntese; fora do anel: G0 = Gap 0 / Repouso. Esta figura extraído de https://commons.wikimedia.org/wiki/File:Cell_Cycle_2-2.svg

Fases da divisão celular: O ciclo celular consiste em quatro fases distintas: a fase G1 (Gap1), fase S (síntese), fase G2 (Gap2) (conhecidos coletivamente como intérfase) e fase M (mitose).

Intérfase:

Antes de uma célula entrar na divisão celular, ela precisa de nutrientes. Todas as preparações são feitas durante a intérfase. A intérfase é uma série de mudanças que ocorrem numa célula recém-formada e no seu núcleo, antes de se tornar capaz de divisão novamente. É também chamado de fase preparatória ou intermitose. Anteriormente foi chamada de fase de repouso porque não há nenhuma atividade aparente relacionada com a divisão celular. Tipicamente a intérfase dura pelo menos 90% do tempo total necessário para o ciclo celular.

Fase G0:

A fase G0 é um período no ciclo celular em que existem células num estado quiescente. A fase G0 é visto como uma fase G1 estendida, em que a célula não está dividindo e nem se preparando para dividir, ou um estágio quiescente distinto que ocorre fora do ciclo celular. Por isso, G0 é por vezes referida como um estado "pós-mitótico".

Fase G1:

A primeira fase de interfase é fase G1. O final da fase anterior da mitose até o início da replicação do DNA é chamado de G1 (G indica “Gap”). É também chamada de fase de crescimento. Durante esta fase as atividades biossintéticas das células, que tinham sido consideravelmente retardadas durante a fase M, retomam a uma taxa elevada. Esta fase caracteriza-se por síntese de várias enzimas que são necessárias na fase S, principalmente aquelas necessárias para a replicação do DNA. A duração de G1 é altamente variável, mesmo entre diferentes células da mesma espécie. [9]

Fase S:

A iniciação da replicação do DNA é a indicação da fase S; quando estiver concluída, todos os cromossomos foram replicados e cada cromossomo tem duas cromátides (irmã). Assim, durante esta fase, a quantidade de DNA na célula foi efetivamente dobrada, embora o

número de cromossomos da célula permaneça o mesmo. As taxas de transcrição de RNA e a síntese de proteínas são muito baixas durante esta fase. Uma exceção a isso é a produção da proteína histona, que ocorre principalmente durante a fase S.

Fase G2:

Depois da fase S ou replicação, a célula entra na fase G2, que dura até que a mesma entre na mitose. Mais uma vez, a biossíntese significativa ocorre durante esta fase, envolvendo principalmente a produção de micro túbulos, que são necessários durante o processo de mitose. A inibição da síntese de proteínas durante a fase G2 impede a célula de sofrer mitose.

Mitose:

A mitose é o processo pelo qual uma célula eucarionte separa os cromossomos no seu núcleo em dois conjuntos idênticos em dois núcleos. Em geral, é imediatamente seguida pela citocinese, que divide o núcleo, citoplasma, membrana celular e organelas em duas células que contenham partes aproximadamente iguais desses componentes celulares. A mitose e a citocinese em conjunto definem a fase mitótica (M) do ciclo celular - a divisão da célula mãe em duas células filhas geneticamente idêntica.

2.7 Regulação transcricional de genes:

Em biologia molecular e genética, a regulação transcricional é o meio pelo qual uma célula regula a conversão de DNA para o RNA. Um único gene pode ser regulado numa gama de formas, desde alterar o número de cópias de RNA que são transcritos, até o controle temporal de quando o gene é transcrito. Esse controle permite que a célula ou o organismo responda a uma variedade de sinais intra e extracelulares. A transcrição é o primeiro passo da expressão genética em que um segmento específico de DNA é copiado em RNA (RNAm) pela enzima RNA polimerase. Tanto o DNA e RNA são ácidos

nucleicos, que utilizam pares de bases de nucleotídeos como uma linguagem complementar. Durante a transcrição, uma sequência de DNA é lida por uma polimerase de RNA, o que produz uma cadeia de RNA antiparalela complementar chamada de transcrito primário.

2.7.1.Éxons:

Os éxons são seções de codificação de um transcrito de RNA ou DNA que são traduzidos em proteína. Os éxons podem ser separados por intervenções de seções de DNA que não codificam proteínas, conhecidos como íntrons. Após a transcrição, novos fios imaturos de RNA mensageiro, chamados de pré-mRNA, podem conter tanto íntrons e éxons. Essas moléculas de pré-mRNA passam por um processo de modificação no núcleo chamado 'splicing', durante o qual os íntrons não codificantes são tirados fora e apenas os éxons de codificação permanecem. Esse 'splicing' produz uma molécula de RNA mensageiro madura que é então traduzida em uma proteína.

2.7.2 Íntrons:

Um íntron é qualquer sequência de nucleotídeos dentro de um gene que é removida pelo 'splicing' do RNA durante a maturação do produto final de RNA. O termo íntron refere-se a sequência de DNA dentro de um gene e a sequência correspondente em transcritos de RNA. Sequências essas que são unidas no RNA final depois de splicing de RNA são exons. Os íntrons são encontrados nos genes da maioria dos organismos e vários vírus, e podem estar localizados em uma grande variedade de genes, incluindo os que geram proteínas, RNA ribossômico (RNAr) e RNA transportador (tRNA). Quando as proteínas são geradas a partir de genes contendo íntrons, o RNA 'splicing' ocorre como parte da via de processamento de RNA que segue a transcrição e precede a tradução.

2.7.3 Splicing do RNA:

Em biologia molecular, splicing é a edição do nascente RNA pré-mensageiro (pré- RNAm) transcrito . Depois do splicing, os íntrons são removidos e os éxons são unidos (ligados) em

conjunto. Para genes nucleares codificados, o splicing ocorre dentro do núcleo, quer durante ou imediatamente após a transcrição. Para os genes eucariontes que contém íntrons, o splicing é geralmente necessário para criar uma molécula de RNAm que pode ser traduzida em proteína .

2.8 Interações proteicas:

As proteínas são as trabalhadoras que facilitam a maioria dos processos biológicos em uma célula, incluindo expressão gênica, crescimento celular, proliferação, absorção de nutrientes, morfologia, motilidade, comunicação intercelular e apoptose. Mas as células respondem a uma variedade de estímulos, e, portanto, a expressão proteica é um processo dinâmico, e as proteínas que são utilizadas para completar tarefas específicas talvez não estejam sempre expressas ou ativadas. Além disso, as células não são iguais, e muitas proteínas são expressas de uma forma dependente do tipo de célula. Essas características básicas das proteínas sugerem uma complexidade que pode ser difícil de investigar, especialmente quando se tenta compreender a função da proteína no contexto biológico adequado. Na interação proteína-proteína (PPIs), é importante considerar que as proteínas podem interagir de uma maneira "transitória" (para produzir algum efeito específico em um curto período de tempo) ou interagir com outras proteínas de uma forma "estável" para construir complexos multiproteicos que são máquinas moleculares dentro dos sistemas vivos.

Aspectos importantes necessários para compreender a função de uma proteína incluem:

- Sequência e estrutura proteica - usada para descobrir os motivos que predizem a função da proteína
- História evolutiva e sequências conservadas - identificam resíduos reguladores fundamentais
- Perfil de expressão – revela a especificidade de cada tipo de célula e como a expressão é regulada
- Modificações pós-tradução - fosforilação, desfosforilação e ubiquitinação sugerem

localização, ativação ou inativação da função.

- Interações com outras proteínas – a função pode ser extrapolada sabendo a função dos parceiros de ligação.

2.9 Fosforilação:

A fosforilação reversível de proteínas, principalmente em resíduos de serina, treonina ou tirosina, é uma das mais importantes e bem estudadas modificações pós-tradução. A fosforilação desempenha um papel crítico na regulação de muitos processos celulares, incluindo o ciclo celular, crescimento, apoptose e vias de transdução de sinais. A fosforilação é a adição de um grupo fosfato

2.10 Desfosforilação:

A desfosforilação em células vivas é um processo bioquímico comum, no qual um grupo fosfato é removido de um composto orgânico por meio de hidrólise. Um exemplo desta situação é quando um ATP perde um grupo fosfato, alterando, assim, em ADP.

2.11 Ubiquitinação:

A Ubiquitinação é um processo enzimático que envolve a ligação de uma proteína de ubiquitina a uma proteína do substrato. Isso tem sido por vezes referido como o "beijo da morte" molecular para uma proteína, pois o substrato torna-se usualmente inativado e é marcado para degradação pelo proteassoma através da fixação da molécula ubiquitina.

CHAPTER 3: MODELING GENE REGULATORY NETWORKS

3.1 Gene regulatory networks:

A Gene or Protein that activates (or inhibits) the functionality of another Gene or Protein is said to interact with the target protein by an activating (or an inhibiting) mechanism. A graphical representation where the nodes of the graph represents the functional form of gene or proteins and directed edges represent the activation (or inhibition) mechanism is commonly referred to as a gene regulatory network (GRN). Where activating edges are represented by arrow-headed lines and inhibiting edges are represented by circle-headed lines.

A genetic regulatory network (GRN) is a collection of DNA segments in a cell which interact with each other indirectly (through their RNA and protein expression products) and with other molecules in the cell, thereby governing the rates at which genes in the network are transcribed into mRNA. GRNs provide a systematic understanding of molecular mechanisms underlying biological processes [10-16]. The groups of genes, regulatory proteins and their interactions are often referred to as regulatory networks. The nodes of this network are genes and the edges between nodes represent gene interactions through which the product of one gene affects those of another. These interactions can be inductive (represented by arrowheads), with an increase in the expression of one leading to an increase in the other, or inhibitory (represented by hammerheads), with an increase in one leading to a decrease in the other. A series of edges indicates a chain of such dependencies, with cycles corresponding to feedback loops. For example, let us consider a simple discrete two element system **(**Figure 8**).** Suppose that product $\underline{x}$ activates gene $\underline{Y}$ and product $\underline{y}$ represses gene $\underline{X}$. In other words.

$X = 1$ iff $y = 0$ ($\underline{X}$ `` on´´ iff y absent)

$Y = 1$ iff $x = 1$ ($\underline{Y}$ ``on´´ iff x present).

This can be described by the graph of interactions in Figure 6.

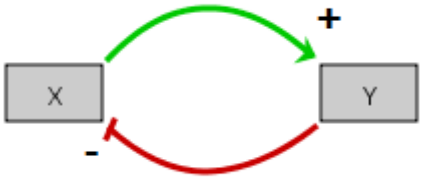


Figure 8. Example of simple feedback loops. Circuit is a negative loop.

Gene regulatory networks describe control at the gene expression level and could be inferred from microRNAs (miRNAs). Gene regulation has become important as information of molecular regulatory interactions, becoming increasingly available. The importance of gene regulatory networks is evident for all biological species and systems [17-18] as they play important role in maintaining the biological functions of living organism [19]. The inference of regulators is the core factor in interpreting the actual regulatory conditions in GRNs [20]. This provides a clearer blueprint on the relationship between target genes and regulator genes. Accurate predictions of the behavior of regulatory networks will also accelerate biotechnological projects and such predictions are quicker and cheaper than lab experiments.

Various computational models developed for regulatory network analysis can be roughly divided into four classes**.** A complete gene regulatory network

model incorporates experimental knowledge about the components and their interactions as well as the initial state of these components, and lead to the known final state or dynamical behavior of the network. In this article we review briefly the various modeling techniques for reconstructing gene regulatory network.

3.2 Modeling techniques:

The GRN modeling techniques can be broadly categorized into continuous and discrete approaches. Here we are interested in using discrete modeling (Boolean or logical modeling).

3.3 Discrete models:

In the discrete modeling of GRNs, a node can take only discrete expression values, unlike the continuous range of values that are possible in models. In the most restrictive case, a node exists in only two expression states to represent active and inactive gene (or protein) represented by Boolean 1 and 0 respectively. In Boolean models, an inhibiting edge can change the expression of the node from 1-to-0 and an activating edge can change the state from 0-to-1 (only when no inhibition is present). A snapshot of the activity level of all the nodes in the GRN at a given time instance is called the state of the network.

The discrete model is based upon a distinction between time-points and time intervals. Typically the time axis is divided into a number of adjacent time-segments (which usually are of fixed length). Both the number of time intervals and time points that are specified are finite. Actually the number of time points = Number of time-intervals +1.

Discrete Systems are dynamic systems that evolve in discrete steps, due to the abrupt occurrence of internal or external events. The system evolution is typically modelled as resulting from firing state transforming rules, which are triggered when certain (internal and/or external) conditions become true. Such systems encompass sequential (also called transformational) algorithms and their implementations as computer programs, but also systems of distributed (asynchronous concurrent) processes.

3.3.1 Logical modeling

The most basic and simplest modelling methodology is discrete and logic-based, and was introduced by Kauffman and Thomas [21, 22]. The reconstruction of the regulatory network that controls the mammalian cell cycle [28], the development of sea urchin embryos [23, 24] is another example of the profound insights that qualitative examination of regulatory network models can provide. Logical model represent the local state of each entity in the system (for example, genes, proteins and small molecules) at any time as a discrete level and the temporal development of the system is often assumed to occur in a synchronous, discrete time steps. Entity levels are updated at each time step according to regulation functions. Discrete modeling allows researchers to rely on purely qualitative knowledge. Such models can be analyzed using a broad range of well-established mathematical and statistical methods.

Logical models are versatile: a variable can represent almost anything, such as a gene activity, the presence of a protein or the state of the cell. They are flexible: the state of a given cellular components can be represented by one or more variables, with different sets of values.

3.3.2 Boolean Network

Boolean regulatory networks were first presented by Kauffman [21] and are a simplified version of logical models. In a Boolean network, A variable can only assume two values; the value are usually represented by 1 and 0 or true and false. The logic operators are ***and, or*** and ***not.*** The variables are divided into two classes for building a model: inactive and active [25]. For example, a gene can be described as expressed or not expressed at any time. A Boolean function is a function in which Boolean variables are linked by logic operators. This means that, in a Boolean network, the state of gene expression level is either ***on*** or ***off***. They are the simplest network models that resemble some of the biological and systemic properties of real gene network [26-27].

A Boolean network can be defined by a directed graph G(X, E), where the nodes, x_i ∈ X, are Boolean variables. To each node, x_i, is associated a boolean function, $b_i(x_i1$, $x_i2\ldots, x_i l)$, $i = 1,2,\ldots. n$, $l \leq n$, $x_{ij} \in$ X, where the arguments. Together, at any given time, the states (values) of all nodes represent the state of network, given by the vector $S(t) = \{x_1(t), x_2(t),\ldots x_n(t)\}$.The state of all nodes are updated at the same time (i.e., synchronous) according to their respective Boolean functions.

$$x_i(t+1) = b_i \{x_i1(t),x_i2(t),\ldots x_i l(t)\}.$$

All states transitions together correspond to a state transition of the network from S (t) to the new network state, S (t+1).The update can also be asynchronous, i.e., one variable at a time.

3.4 Simulation Tools: GINSim

GINsim supports the definition, the simulation and the analysis of regulatory graphs, based on the (multi-valued) logical formalism. Developed in Java, GINsim is platform-independent and only requires a recent Java Virtual Machine. GINsim is freely available for academic users. The latest official version is available from the GINsim web site (http://gin.univ-mrs.fr/GINsim).

3.4.1. Dynamical simulations:

The dynamical behavior of a logical regulatory graph is described as another type of graph called state transition graph. In this graph, each node represents a state of the model, defined by vectors of the component levels. The arcs represent transitions between states. One core function of GINsim is the automatic construction of this graph. However, to use this function judiciously, it is important to understand the principles underlying this construction, which are outlined hereafter. Afterwards, several options to analyse the dynamics of logical regulatory graphs are presented. Two main strategies are commonly

used. In the synchronous updating, all concerned components change their levels simultaneously in a unique transition towards the next state. By contrast, the asynchronous updating generates a successor state for each component updating call.

Building a state transition graph, one can choose to generate the full state transition graph, considering all possible initial conditions, or yet a (sub) graph from specific initial state(s). In such state transition graphs, it is then possible to determine the stable states (defined as nodes with no outgoing arcs), or more complex attractors (defined as terminal maximal strongly connected components, denoting a stable oscillatory behavior). Moreover, it may be interesting to search a path between two states, in particular to verify the reachability of a given stable state from an initial state.

3.4.2 Circuit Analysis

A regulatory circuit is defined by a sequence of interactions forming a simple closed directed path. The sign of a circuit is defined as the product of the signs of its interactions. Consequently, a circuit is positive if it has zero or an even number of inhibitions, it is negative otherwise. R. Thomas [22] proposed that positive circuits are necessary to generate multistationarity, whereas negative circuits are necessary to generate stable oscillations. External regulators might prevent the functioning of a circuit imbedded in a more complex network. Presents a method to determine the functionality context of a circuit in terms of constraints on the levels of its external regulator. A circuit functionality context can be interpreted as the part of the state space where the circuit is functional, i.e. generates the expected dynamical property.

Regulatory circuits are responsible for the emergence of dynamical properties such as multistationarity or stable oscillations. GINsim implements specific algorithms to:

1. Identify all the circuits of a regulatory graph (possibly considering constraints such as length limitation, component exclusion, etc.).

2. Determine the functionality contexts of these circuits.

Notes:-

GINsim allows the user to export logical regulatory graphs as well as state transition graphs towards various formats, facilitating the use of other software's. Logical regulatory graphs may encompass multi-arcs, composed of different interactions having different effects. To each interaction, an interval specifying the range of the source levels for which the interaction occurs must be specified. Intervals assigned to interactions with the same source and target must be disjoint. Moreover, a sign should be specified (positive, negative, or dual) for each regulatory effect. Transitions between states of the state transition graph amount to the update of one (in the asynchronous case) or several (in the synchronous case) components. In any case, the update (increase or decrease) of a component is unitary (current value +1 or 1). Obviously, this remark applies only for multi-valued components (for which the maximal level is greater than 1).

CHAPTER 4. APPLICATION OF GENE REGULATORY NETWORKS

4.1 Introduction:

Computational modeling of genomic regulation has become an important focus of Systems biology or Bioinformatics and genomic signal processing for the past several years [39]. It holds the promise to uncover both the structure and dynamical properties of the complex gene, protein or metabolic networks responsible for the cell functioning in various contexts and regimes. This, in turn, will lead to the development of optimal intervention strategies for prevention and control of disease. At the same time, constructing such computational models faces several challenges. High complexity is one of the major impediments for the practical applications of the models. Thus, reducing the size or complexity of a model becomes a critical issue in problems such as model selection, construction of tractable subnetwork models, and control of its dynamical behavior. This work focus on some Boolean logical models as examples that how it's useful for research in medicine or drug design.

4.2 Networks as biomarkers:

In recent studies, it has been argued that (sub-) networks could also be used as biomarkers, e.g., for diagnostic, predictive or prognostic purposes [29-31]. This is particularly plausible for a complex disorder like cancer, because the hallmarks of cancer are represented by pathways rather than individual genes [32] and the crucial aspect of pathways is that their constituting genes are actively interacting with each other. For this reason, network-based biomarkers can be seen as statistical measures that consider the interaction structure between individual genes explicitly. In contrast, biomarkers based on individual genes neglect these completely [32].

4.3 Network in medicine and also in drug designing:

For establishing a network medicine useful for clinicians, it will be necessary to integrate different types of gene networks with each other, because each network type carries information about particular molecular information's. For example, whereas the

transcriptional regulatory network contains only information about the controlling regulations of gene expression, protein interaction networks represent information about protein-protein complexes [33-34]. Taken together, an integration of various important molecular interaction types results in a comprehensive overview of regulatory programs and organizational architectures.

Our interest here is logical modeling than we are presenting some examples…

4.4 Application to Astrocyte Senescence.

4.4.1. What is Astrocyte?

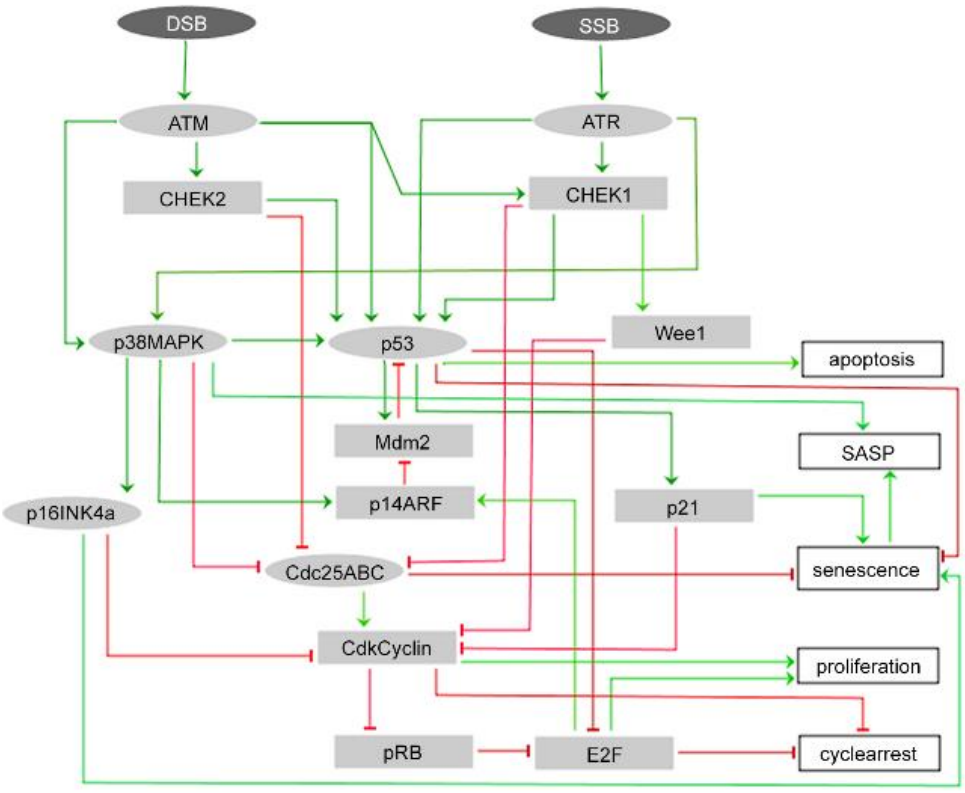
Astrocytes, also known as astroglia, are characteristic star-shaped glial cells in the brain and spinal cord. They perform many functions, including biochemical support of endothelial cells that form the blood–brain barrier.

4.4.2. A Model for p38MAPK-Induced Astrocyte Senescence [35]

According this research article, Senescence is a antitumor program that triggered by telomere shortening and oxidative and oncogene activation. Model support that senescent cells accumulate in ageing mammal tissues and altered phenotype called SASP (Senescence associated secretory phenotype). SASP involve in several aging disease including Alzheimer´s disease. SASP contributes to ‘inflamm-aging’ (the development of a systemic proinflammatory status with normal aging) which means an increase of blood plasma levels of inflammatory cytokines like interleukin 6 (IL-6). Alzheimer disease (AD) is an example of inflammation disease. In AD case, astrocyte senescence is claimed to be an important contributor to the development of the pathology [36]. Astrocytes are sensitive to oxidative stress, which increases with aging and causes DNA damage. So here Author´s propose that p38MAPK-Induction and explain Astrocyte Senescence and SASP. They propose an extended logical model of process integrating checkpoints G1/S and G2/M [37].

Methods-

In most cases the variables are Boolean (0 or 1), but multi-valued variables can represent different influences of a node affecting its targets. Edges represent activators or inhibitory effects and variables denote activity levels with two or more states (multi-valued).



Regulatory network for astrocyte fate decision. Rectangular and elliptic nodes represent Boolean and multi-valued nodes, respectively. The input nodes in dark color at the top of the network denote single (SSB) and double-strand (DSB) DNA breaks, respectively. The output nodes in white color represent the possible cell fate decisions and the internal nodes are the regulators of the outputs.

Figure 9. Interactions in the logical model for astrocyte senescence [35].

As shown in Fig. 9, DNA double-strand breaks (DSB) activate the ATM, either DNA single-strand breaks (SSB) or DSB activate Rad3-related (ATR). Phosphorylation goes downstream of ATM and ATR lead to activation of p53. The kinase checkpoint kinase 2 (CHEK2) is activated by ATM while the kinase checkpoint kinase 1 (CHEK1) is activated

by ATR. CHEK2 and CHEK1 start the arrest upregulating Wee1 G2 checkpoint kinase (Wee1) and inactivating CDC25A/B/C required for both checkpoints to activate protein complexes involving cyclins and cyclin-dependent kinases (CDKs) that determine cell cycle progress, These complexes are cyclin-dependent kinase 4, 6 and cyclin D (Cdk4/6-Cyclin-D) complex, cyclin-dependent kinase 2 and cyclin E (Cdk2/Cyclin-E) complex for checkpoint G1/S, and cyclin-dependent kinase 1 and cyclin B (Cdk1/Cyclin B) complex (which is inhibited by Wee1) for checkpoint G2/M. In addition, phosphorylated p53 mediates the maintenance of arrest through the activation of cyclin-dependent kinase inhibitor 1A (p21), which also inhibits Cdk4/6-Cyclin-D. In the case of checkpoint G1/S, the inhibition of these complexes prevents the phosphorylation of retinoblastoma 1 protein (pRB) and the release of E2F transcription factors that induce the expression of genes required for the cell to enter the S phase. In the case of reparable damage, the complexes are reactivated driving the cell to the next phase of the cycle. E3 ubiquitin protein (Mdm2), p14ARF and p53 form a regulatory circuit. Mdm2 inhibited p53 and Mdm2 is sequestered by p14ARF controlling p53 degradation. The choice between cycle arrest and apoptosis occurs through a threshold mechanism dependent on the activation level of p53 that, when exceeded, triggers apoptosis. Owing to this, in this model, apoptosis is activated only when p53 reaches its highest level which is a strong simplification. p14ARF and cyclin-dependent kinase inhibitor 2A ($p16^{INK4a}$) contribute to cell cycle regulation and senescence, deletion of the locus (CDKN2A) that produces these two proteins enhances astrocyte proliferation. DNA damage can induce a checkpoint arrest through p38MAPK upon joint mechanisms like, upregulation of $p16^{INK4a}$ and p14ARF, inhibition of the protein family Cdc25A/B/C and activation of p53 which, additionally, can lead to apoptosis. Absence of DNA damage implies the activation of nodes CdkCyclin (representing the protein complexes that promote cell cycle for both checkpoints G1/S and G2/M) and ‘proliferation’. The ‘cycle-arrest’ node represents an arrest for repair.

Model Results: Wild Type Case

This model presents deterministic behavior since each combination of the levels of the input nodes DSB and SSB (nine in total) leads to a unique stable state characterized by the

activation or deactivation of the nodes representing fates. The synopsis of the results for the wild type case is that no DNA damage (input: DSB = SSB = 0) leads to proliferation. The highest level of irreparable damage DSB = SSB = 2 increase apoptosis, irreparable DSBs lead to senescence (and SASP) and reparable DSBs result in transient cycle arrest indifferent to the value of SSBs. Lastly, 'apoptosis' is activated by p53 = 2 when SSB = DSB = 2.

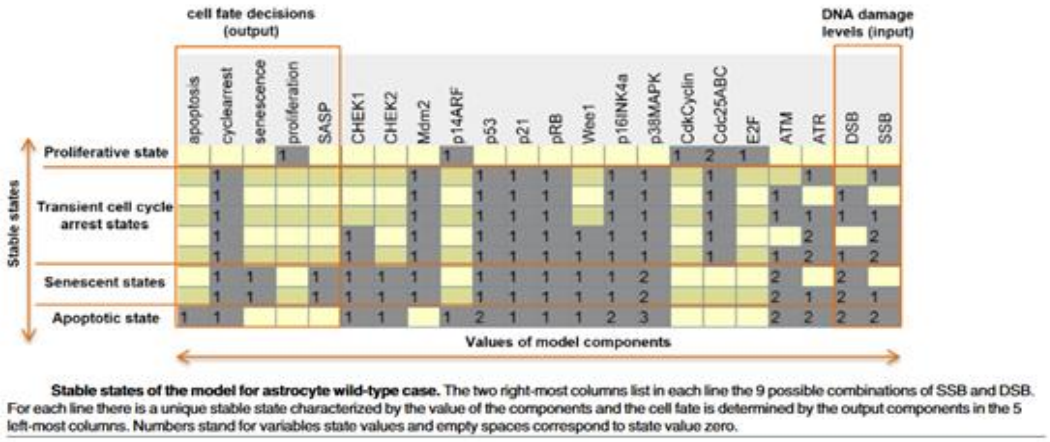


Figure 10. Stable states of the logical model for astrocyte senescence [35].

Model Results: Comparison with experiments

The model can be perturbed through in silico mutations corresponding to loss of function or gain of function experiments. These model perturbations affect the stable states with respect to the wild type case and changes can be related to the predominant growth trend observed in cultures of astrocyte cells undergone LoF (Loss of function) or GoF (Gain of function) experiments involving one or more genes. Comparisons of LoF and GoF experiments with model perturbations can be classified in two cases according to DNA damage: (i) absence or (ii) presence of damage that activates checkpoints. DNA damage is assumed to happen in stress situations as seems to be the case of astrocytes in aged brains or in AD

Comparison with experiments.

Loss of Function (LoF)		
Gene	**Model outcome (No DNA damage / With DNA damage)**	**Experimental outcome**
p38MAPK	No damage: proliferation / With damage: loss of senescence, apoptosis & SASP	[9]
p16INK4a & p14ARF	No damage: proliferation / With damage: loss of senescence & apoptosis	[29]
p16INK4a	No damage: proliferation / With damage: loss of senescence	[29]
p53	No damage: proliferation / With damage: loss of senescence & apoptosis	[39]
ATM	No damage: proliferation / With damage: loss of senescence & apoptosis	[39,40]
ATR & p53	No damage: proliferation / With damage: loss of senescence & apoptosis	[43]
p21	No damage: proliferation / With damage: loss of senescence	?
Cdc25ABC	No damage: cycle arrest / With damage: senescence enhanced	? / Fibroblasts: [46]
E2F	No damage: cycle arrest / With damage: similar to the wild type	? / Fibroblasts: [47]
pRB	No damage: proliferation / With damage: apoptosis for DSB = SSB = 2 and undetermined for other cases	? / Fibroblasts: [48,49]
Gain of Function (GoF)		
Gene	**Model outcome**	**Experimental outcome**
p38MAPK	No damage: [1–2] cycle arrest; [3] apoptosis / With damage: [1–2] senescence enhanced & loss of apoptosis; [3] apoptosis	? / Fibroblasts: [30]
p16INK4a	No damage: cycle arrest / With damage: [1] similar to the wild type; [2] senescence enhanced	? / [33]
p53	No damage: [1] cycle arrest; [2] apoptosis / With damage: [1] senescence enhanced; [2] apoptosis	?
ATM	No damage: [1] cycle arrest; [2] enhanced senescence / With damage: [1] loss of senescence & apoptosis; [2] senescence enhanced	?
p21	No damage: cycle arrest / With damage: similar to the wild type	? / Fibroblasts: [44,45]
Cdc25ABC	No damage: proliferation / With damage: [1–2] loss of senescence	? / Fibroblasts: [46]
E2F	No damage: proliferation / With damage: apoptosis	?
pRB	No damage: cycle arrest / With damage: similar to the wild type	?
LoF + GoF		
p16INK4a & p14ARF LoF + CdkCyclin GoF	No damage: proliferation / With damage: proliferation	[42]

Comparison of results of perturbations of the model with experiments. Cases for which no experimental data were found are indicated by question marks.

Table 1. Analysis of mutations the logical model for astrocyte senescence [35].

So, finally, p38MAPK plays a central role in the explanation of senescence and SASP induction due to DNA damage. One prediction, the confirmation that p38MAPK GoF in astrocytes induces senescence, would give support to the model.

4.5 Application to Drug Synergies in Gastric Cancer Cells [38].

According to this research article [38] developed a computational model, this model based on specific cancer cell biomarker obtained from unperturbed cancer cells. Computational models are increasingly used to predict drug effect, such models would ideally be constructed drug perturbation data and predicative models can be based on molecular data from unperturbed cancer cells.

Construction of a logical model:

The network has 75 signaling and regulatory components (protein, protein complex and gene) and 149 directed interactions. Two nodes (Outputs) named Pro-survival and Anti-survival. This outputs is represent cell fate phenotypes.

In this logical model nodes was represented by a Boolean variable (0 or 1) and few nodes were multi-leveled. Two outputs nodes Pro-survival and Anti-survival taking four values (0,1,2,3) and their immediate upstream nodes, Caspase 3/7 and CCND1, each taking three values (0, 1, 2). These multilevel variable nodes are only used for nodes governing the outputs of the model.

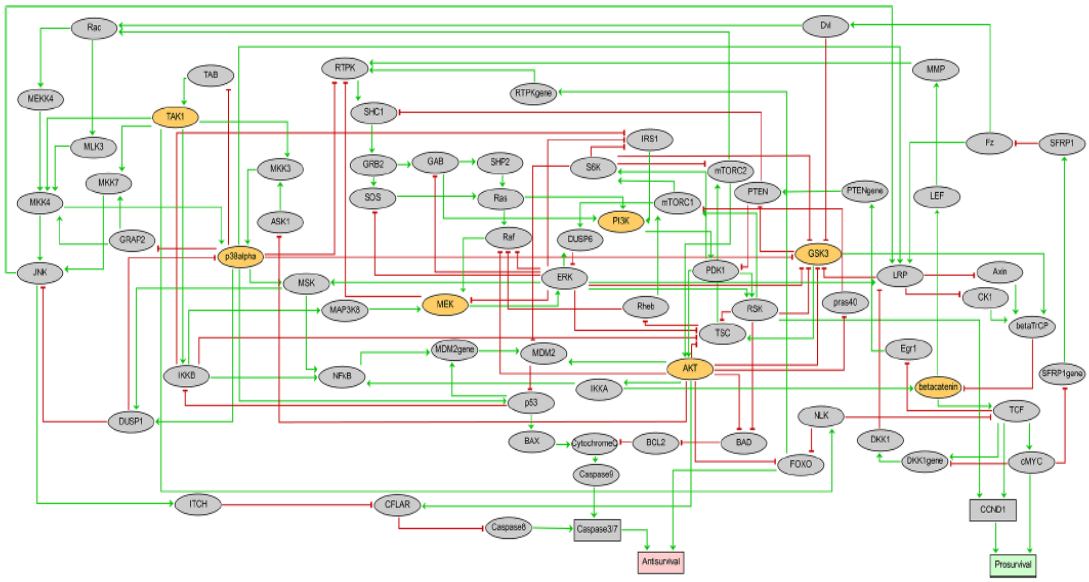


Figure 11. Protein interactions in gastric cancer cells [38].

The network has no external input but encompasses two outputs *Antisurvival* and *Prosurvival* (phenotypic readouts, colored in red for Antisurvival and green for Prosurvival). Activating regulations are denoted by green arrows, while red T arrows denote inhibition. Signaling component nodes (proteins, protein complexes or genes) associated with Boolean variables (taking the values 0, 1) are represented by ellipses, while rectangles depict nodes encoded with multilevel variables. Yellow nodes represent drug

targets.

Reduced Logical Model:

The computation of potential complex attractors is challenging because of the combination of states for large logical models. Model Reduction is best way to solve this problem so here authors used model reductions method to obtain compressed model preserving the selected drug targets and compacted the state transition graph in a Hierarchical graph. So they focused on the systematic inhibition of seven models nodes. These seven nodes (labelled with thick border in large model). They chose this seven nodes for reduce model because these seven nodes are potent and specific chemical inhibiters were available for targeting.

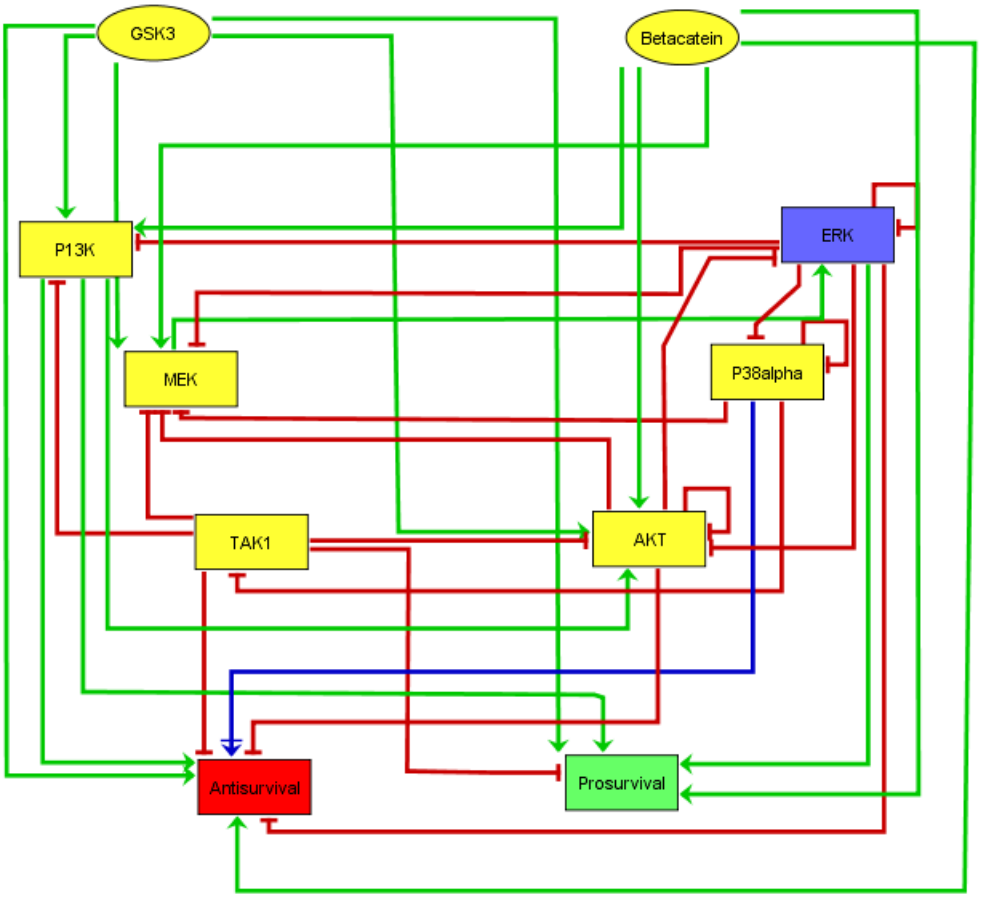


Figure 12. Reduced model Protein interactions in gastric cancer cells [38].

The reduced model encompasses all seven drug targets (yellow) and the two outputs (red for Antisurvival and green for Prosurvival). In addition the ERK node (blue) had to be preserved to maintain dynamical consistency with the large model. Activating regulations are denoted by green arrows, while red T arrows denote inhibition. The blue arc with both arrow and T head (p38alpha to Antisurvival) indicates a dual regulation, i.e. activating and inhibiting. In some contexts p38alpha inhibition will increase Antisurvival, while in others p38alpha inhibition will decrease Antisurvival. Important thing is here that in reduce model, GSK3 and Betacatenin both are Inputs of reduce model.

Nodes of the network by updated logical function of their target nodes. The stable states are conserved by this reduction. Furthermore, each complex attractor of the original model is matched by at least one complex attractor in the reduced model. However, as model reduction generally results in a simplified STG, complex attractors may be split during the reduction process, while attractor reachability might also be affected. Such distortion of the dynamics was assessed by checking the behavior of the original model or by using alternative reductions. For all reduced instances of the model.

Results:

Hierarchical state transition graphs:

The resulting dynamics is represented in terms of a state transition graph (STG). The nodes of the STG denote the states of the system, i.e. discrete vectors encompassing the activity values of all components (Boolean variables, except for the four multi-valued components), while the arcs connect successive states, denoting “state transitions”. Enabled transitions were defined based on an asynchronous updating policy: whenever multiple components are called for a change, all single value changes are considered, leading to the representation of all possible asynchronous trajectories in a single STG. The asymptotic behavior of the system corresponds to the attractors of the dynamics (terminal strongly connected components in graph theoretical terms). Since this AGS model is finite, its dynamics contains at least one attractor. Two types of attractors may occur: stable states (single state attractors) and complex (cyclic) attractors (sets of states from which the system

cannot escape). To identify and characterize the complex attractors for such a large network so Authors [38] did compaction of state transition graphs into hierarchical transitions graphs.

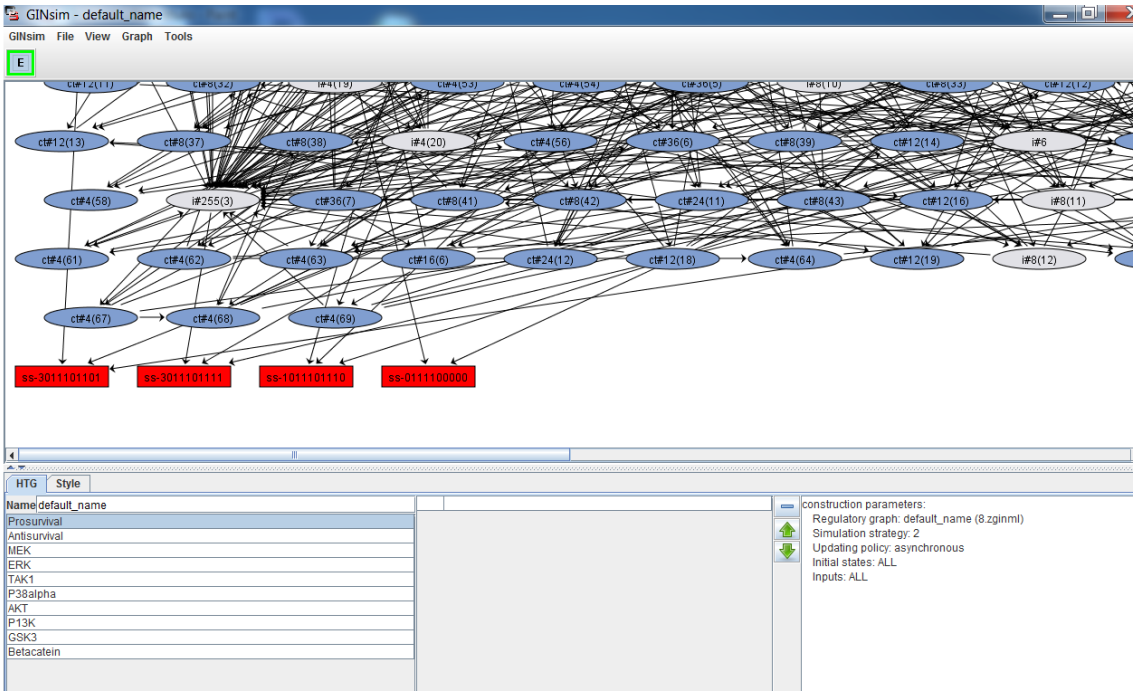


Figure 13. Stable states for the logical model of gastric cancer cells reproduced using GINsim.

The analysis of state transition graphs becomes intractable as their size increases. To ease its interpretation and its manipulation, an STG can be compressed into a hierarchical transition graph (HTG), which preserves its main structural properties: as the STG is constructed, its nodes are gathered into groups of states sharing the same set of successors. The resulting HTG displays all reachable attractors, and their basins of attraction and getting 4 stable states. Which means *MEK-AKT* or *MEK-P13K* and *TAK1-AKT* or *TAK1-P13K* lead to *Prosurvival.* This 4 stable states are predicted synergies and confirmed in drugs used in adenocarcinoma of gastric câncer, showing that discrete models can grasp important features of real systems.

CAPÍTULO 5: CONCLUSÃO

Neste trabalho estudamos métodos de modelagem para redes regulatórias de genes. Os métodos de modelagem para essas redes podem ser classificados em contínuos e discretos. Para os modelos complexos que envolvem muitas proteínas ou genes (10 ou mais), a melhor opção inicial é usar os métodos discretos, pois são mais simples de aplicar, embora qualitativos. No entanto para sermos rigorosos devemos utilizar os métodos baseados em equações diferenciais. Devido aos nossos interesses, neste trabalho dedicamos atenção aos seguintes modelos discretos: senescência celular em astrócitos [35] e sinergia de drogas para controle do câncer gástrico [38]. Além disso, escrevemos um artigo que apresenta um resumo sobre as técnicas de modelagem discreta e contínua, vantagens e desvantagens desses dessas abordagens com uma listagem dos softwares disponíveis para uso nesse tipo de abordagem.

References:

1 KARP, Gerald. **Cell and Molecular Biology: Concepts and Experiments 4th Edition with Study**. John Wiley & Sons, 2004.

2. JENKINS, Philip. **Synthetic panics: The symbolic politics of designer drugs**. NYU Press, 1999.

3. MATON, Anthea et al. **Cells: Building blocks of life**. Pearson Prentice Hall, 1997.

4 SCHOPF, J. William et al. Evidence of Archean life: stromatolites and microfossils. **Precambrian Research**, v. 158, n. 3, p. 141-155, 2007.

5. SCHOPF, J. William. Fossil evidence of Archaean life. **Philosophical Transactions of the Royal Society of London B: Biological Sciences**, v. 361, n. 1470, p. 869-885, 2006.

6. ASIMOV, Isaac. Part I Metaphors of Memory. **Memory in the Twenty-First Century**, v. 30, n. 4, p. 365, 2002.

7. KUBIATKO, Milan; HALÁKOVÁ, Zuzana. Slovak high school students' attitudes to ICT using in biology lesson. **Computers in Human Behavior**, v. 25, n. 3, p. 743-748, 2009.

8. BLACK, Jacquelyn G. **Microbiology: principles and explorations**. John Wiley & Sons, 2008.

9 GERICKE, Niklas Markus; HAGBERG, Mariana. Definition of historical models of gene function and their relation to students' understanding of genetics. **Science & Education**, v. 16, n. 7-8, p. 849-881, 2007.

10. KARLEBACH, Guy; SHAMIR, Ron. Modelling and analysis of gene regulatory networks. **Nature Reviews Molecular Cell Biology**, v. 9, n. 10, p. 770-780, 2008.

11 LEVINE, Michael; DAVIDSON, Eric H. Gene regulatory networks for development. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 14, p. 4936-4942, 2005.

12. HASTY, Jeff et al. Computational studies of gene regulatory networks: in numero molecular biology. **Nature Reviews Genetics**, v. 2, n. 4, p. 268-279, 2001.

13. SWAIN, Martin T.; MANDEL, Johannes J.; DUBITZKY, Werner. Comparative study of three commonly used continuous deterministic methods for modeling gene regulation networks. **BMC bioinformatics**, v. 11, n. 1, p. 1, 2010.

14. WESSELS, Lodewyk FA et al. A comparison of genetic network models. In:**Pacific Symposium on Biocomputing**. 2001. p. 508-519.

15 CHO, K.-H. et al. Reverse engineering of gene regulatory networks. **IET Systems Biology**, v. 1, n. 3, p. 149-163, 2007.

16. DE JONG, Hidde. Modeling and simulation of genetic regulatory systems: a literature review. **Journal of computational biology**, v. 9, n. 1, p. 67-103, 2002.

17. POTKIN, Steven G. et al. Identifying gene regulatory networks in schizophrenia. **Neuroimage**, v. 53, n. 3, p. 839-847, 2010.

18. WILCZYNSKI, Bartek; FURLONG, Eileen EM. Challenges for modeling global gene regulatory networks during development: insights from Drosophila.**Developmental biology**, v. 340, n. 2, p. 161-169, 2010.

19. HORI, Yutaka; KIM, Tae-Hyoung; HARA, Shinji. Existence criteria of periodic oscillations in cyclic gene regulatory networks. **Automatica**, v. 47, n. 6, p. 1203-1209, 2011.

20. LIU, Guixia et al. Combination of neuro-fuzzy network models with biological knowledge for reconstructing gene regulatory networks. **Journal of Bionic Engineering**, v. 8, n. 1, p. 98-106, 2011.

21. GLASS, Leon; KAUFFMAN, Stuart A. The logical analysis of continuous, non-linear biochemical control networks. **Journal of theoretical Biology**, v. 39, n. 1, p. 103-129, 1973.

22. THOMAS, René. Boolean formalization of genetic control circuits. **Journal of theoretical biology**, v. 42, n. 3, p. 563-585, 1973.

23. DAVIDSON, Eric H. et al. A genomic regulatory network for development.**science**, v. 295, n. 5560, p. 1669-1678, 2002.

24. SMITH, Joel; THEODORIS, Christina; DAVIDSON, Eric H. A gene regulatory network subcircuit drives a dynamic pattern of gene expression. **Science**, v. 318, n. 5851, p. 794-797, 2007.

25. DAVIDICH, Maria I.; BORNHOLDT, Stefan. Boolean network model predicts cell cycle sequence of fission yeast. **PloS one**, v. 3, n. 2, p. e1672, 2008.

26. FILKOV, Vladimir. Handbook of computational molecular biology. 2005.

27. FAURÉ, Adrien et al. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. **Bioinformatics**, v. 22, n. 14, p. e124-e131, 2006.

28 FAURÉ, Adrien et al. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. **Bioinformatics**, v. 22, n. 14, p. e124-e131, 2006.

29. CHUANG, Han-Yu et al. Network-based classification of breast cancer metastasis. **Molecular systems biology**, v. 3, n. 1, p. 140, 2007.

30. BEN-HAMO, Rotem; EFRONI, Sol. Gene expression and network-based analysis reveals a novel role for hsa-miR-9 and drug control over the p38 network in glioblastoma multiforme progression. **Genome medicine**, v. 3, n. 11, p. 1, 2011.

31. CHEN, Li et al. Identifying cancer biomarkers by network-constrained support vector machines. **BMC systems biology**, v. 5, n. 1, p. 1, 2011.

32. HANAHAN, Douglas; WEINBERG, Robert A. Hallmarks of cancer: the next generation. **cell**, v. 144, n. 5, p. 646-674, 2011.33.Varol, Chen, et al. "Intestinal lamina propria dendritic cell

subsets have different origin and functions." Immunity 31.3 (2009): 502-512.

34. NEWMAN, Mark; BARABASI, Albert-Laszlo; WATTS, Duncan J. **The structure and dynamics of networks**. Princeton University Press, 2011.

35. MOMBACH, José CM; VENDRUSCULO, Bruno; BUGS, Cristhian A. A Model for p38MAPK-Induced Astrocyte Senescence. **PloS one**, v. 10, n. 5, p. e0125217, 2015.

36. SALMINEN, Antero et al. Astrocytes in the aging brain express characteristics of senescence-associated secretory phenotype. **European Journal of Neuroscience**, v. 34, n. 1, p. 3-11, 2011.

37. NALDI, Aurélien et al. Logical modelling of regulatory networks with GINsim 2.3. **Biosystems**, v. 97, n. 2, p. 134-139, 2009.

38. FLOBAK, Åsmund et al. Discovery of drug synergies in gastric cancer cells predicted by logical modeling. **PLoS Comput Biol**, v. 11, n. 8, p. e1004426, 2015

39. IVANOV, Ivan. Boolean models of genomic regulatory networks: reduction mappings, inference, and external control. **Current genomics**, v. 10, n. 6, p. 375-387, 2009.

6. APPENDIX: An overview on gene regulatory network modelling methods

Article to be submitted for publication that presents a description about Gene regulatory networks and Its applications.

Available online at www.sciencedirect.com

Advances in Biological Regulations

Journal homepage : www.elsevier.com/advances-in-biological-regulation 00 (0000) 0–00

An overview on gene regulatory network modelling methods

Shantanu Gupta[a] , José C. M. Mombach[a]

[a]Universidade Federal de Santa Maria,Santa Maria,RS,Brazil

Abstract

Many biological research areas require gene regulatory networks (GRNs) to provide clear insight and understanding of the cellular process in living cells. This is because interactions between genes and their products play an important role in many molecular processes. Due to its importance, several computational methods have been proposed to gene networks from gene expression data. Computational methods for development of network models and for the analysis of their functionality have proved to be valuable tools in system biology or in bioinformatics applications. In this review, two inference approaches are discussed: Discrete Models and Continuous Models.

Keywords: Gene regulatory network; Computational methods; Gene expression data; Discrete Models; Logical models; Continuous models; Ordinary differential equation based models

1. Introduction

A genetic regulatory network (GRN) is a collection of DNA segments in a cell which interact with each other indirectly (through their RNA and protein expression products) and with other molecules in the cell, thereby governing the rates at which genes in the network are transcribed into mRNA. GRNs provide a

Asystematic understanding of molecular mechanisms underlying biological processes [1-7]. The groups of genes, regulatory proteins and their interactions are often referred to as regulatory networks. The nodes of this network are genes and the edges between nodes represent gene interactions through which the product of one gene affect those of another. These interactions can be inductive (represented by arrowheads), with an increase in the expression of one leading to an increase in the other, or inhibitory (represented by hammerheads), with an increase in one leading to a decrease in the other. A series of edges indicates a chain of such dependencies, with cycles corresponding to feedback loops. For example, let us consider a simple two element system (Figure 1). Suppose that product x activates gene Y and product y represses gene X. In other words.

X = 1 iff y = 0 (X `` on´´ iff y absent)
Y = 1 iff x = 1 (Y ``on´´ iff x present).

This can be described by the graph of interactions in Figure 1.

Figure 1). Example of simple feedback loops. Circuit is negative loop.

Gene regulatory networks describe control at the gene expression level and could be inferred from microRNAs (miRNAs). Regulatory motifs, expression profiles and interactions between regulatory targets [8]. Gene regulation has become important as information of molecular regulatory interactions, becoming increasingly available. The importance of gene regulatory networks is evident for all biological species and systems [9]as they play important role in maintaining the biological functions of living organism[10]. The inference of regulators is the core factor in interpreting the actual regulatory conditions in GRNs[11]. This provides a clearer blueprint on the relationship between target genes and regulator genes. Accurate predictions of the behavior of regulatory networks will also accelerate biotechnological projects and such predictions are quicker and cheaper than lab experiments.

Various computational models developed for regulatory network analysis can be roughly divided into four classes (Figure 2). A complete gene regulatory network model incorporates experimental knowledge about the components and their interactions as well as the initial state of these components, and lead to the known final state or dynamical behavior of the network. In this article we review briefly the various modeling techniques for reconstructing gene regulatory network.

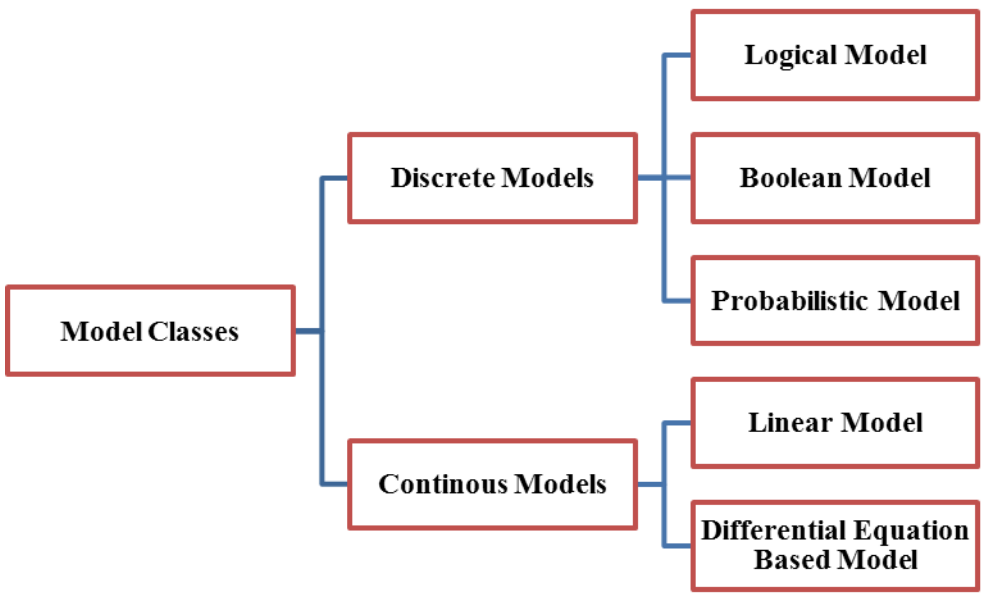


Figure 2. Classification of Models divided into two classes. (A) Discrete Models, (B) Continuous Models.

2. Regulation Vs Phenotype

It is generally acknowledged that a phenotype is an emergent property of genotype-environment interactions. Specifically, a phenotype results from molecular and cellular activity patterns from genotype-environment

interactions. This implies that each observable phenotype is associated with phenotype-specific gene networks, because without changing molecular interactions a phenotype cannot change; this concept is illustrated in gene networks can be seen as a bottleneck between the genotype and the phenotype with respect to their coupling. That means every change on the genotype level that will result in a change of the phenotype will also inevitably lead to a change in the gene network structure as mediator between both levels.

Moreover, a gene regulatory network can potentially represent many types of physical and biochemical interactions among genes and gene products [12] it can be expected that gene regulatory networks to be highly phenotype specific [13,14]. Establishing such relationships will therefore be a complex task, but also provides an opportunity to catalog phenotypes quantitatively. An example for the analysis of tissue-specific networks can be found in Guan et al. [15] where 107 tissue specific networks have been studied. Currently, the number of GRNs is difficult to estimate but based on these preliminary results one can hypothesize that there are more than 200 different GRNs for Humans alone, since this corresponds about the number of different cell types. However, also pathological cells, such as tumors, have their own characteristic networks [16] implying in probably thousands of different gene networks in Humans.

3. Discrete Models

The discrete model is based upon a distinction between time-points and time intervals. Typically the time axis is divided into a number of adjacent time-segments (which usually are of fixed length). Both the number of time intervals and time points that are specified are finite. Actually the number of time points = Number of time-intervals +1.

Discrete Systems are dynamic systems that evolve in discrete steps, due to the abrupt occurrence of internal or external events. The system evolution is typically modelled as resulting from firing state transforming rules, which are triggered when certain (internal and/or external) conditions become true. Such systems encompass sequential (also called transformational) algorithms and their implementations as computer programs, but also systems of distributed (asynchronous concurrent) processes.

3.1. Logical Models

The most basic and simplest modelling methodology is discrete and logic-based, and was introduced by Kauffman and Thomas [17, 18]. The reconstruction of the regulatory network that controls the development of sea urchin embryos [19, 20] is a seminal example of the profound insights that qualitative examination of regulatory network models can provide. Logical model represent the local state of each entity in the system (for example, genes, proteins and small molecules) at any time as a discrete level and the temporal development of the system is often assumed to occur in a synchronous, discrete time steps. Entity levels are updated at each time step according to regulation functions. Discrete modeling allows researchers to rely on purely qualitative knowledge. Such models can be analyzed using a broad range of well-established mathematical and statistical methods.

Logical models are versatile: a variable can be represent

almost anything, such as a gene activity, the presence of a protein or the state of the cell. They are flexible: the state of a given cellular components can be represented by one or more variables, with different sets of values.

3.2. Boolean Network

Boolean regulatory networks were first presented by Kauffman [17] and are a simplified version of logical models. In a Boolean network, A variables can only assume two values; the value are usually represented by 1 and 0 or true and false. The logic operators are *and*, *or* and *not.* The variables are divided into two classes for building a model: inactive and active [21]. For example, a gene can be described as expressed or not expressed at any time. A Boolean function is a function in which Boolean variables are linked by logic operators. This mean that, in a Boolean network, the state of gene expression level is either *on* or *off*. They are the simplest network models that resemble some of the biological and systemic properties of real gene network [22, 23].

A Boolean network can be defined by a directed graph G(X, E), where the nodes, $x_i \in$ X, are Boolean variable. To each node, x_i, is associated a boolean function, $b_i(x_i1, x_i2\ldots, x_il)$, i=1,2,…. n, $l \leq n$, $x_{ij} \in$ X, where the arguments. Together, at any given time, the states (values) of all nodes represent the state of network, given by the vector S(t) = {x_1(t), x_2(t),...x_n(t)}.The state of all nodes are updated at the same time (i.e., synchronous) according to their respective Boolean functions.

$$x_i(t+1) = b_i\{x_i1(t), x_i2(t), \ldots x_il(t)\} \quad (1)$$

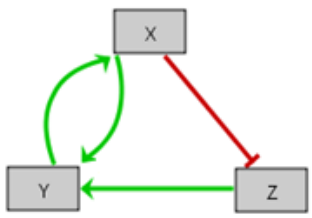
All states transitions together correspond to a state transition of the network from S (t) to the new network state, S (t+1). A sample network is shown in figure 3. The update can also be asynchronous, i.e., one variable at a time.

LIMITATION: These models are ultimately limited by their definition: they are Boolean and asynchronous. In reality, of course, the levels of gene expression do not have only two states but can assume virtually continuous values. Thus discretization of the original data becomes a critical step in the inference and often reducing the values of two states may not suffice.

3.3. Probabilistic Boolean Network

Often, due to insufficient experimental evidence or incomplete understanding of a system, several candidate regulatory functions may be possible for an entity.

(A)



(B)

Logic Rules

Y = X & Z, X = Y, Z =!

(& = AND; ! = NOT)

(D)

INPUT (t)			OUTPUT (t+1)		
X	Y	Z	X	Y	Z
0	0	0	0	0	1
0	0	1	0	0	1
0	1	0	1	0	1
0	1	1	1	0	1
1	0	0	0	0	0
1	0	1	0	1	0
1	1	0	1	0	0
1	1	1	1	1	0

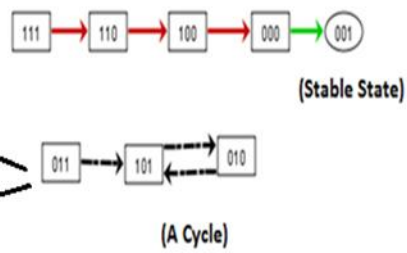


Figure 3. An example Boolean network consisting of 3 genes X, Y, Z and there are different ways for representing the network. (A) Is a gene network modeled as a Boolean network. (B) Boolean rules for state transitions. (C) A complete table of all possible states before and after transition and (D) As a graph representing the state transitions for a stable state and a cycle.

This raises the need to express uncertainty in the regulatory logic. *Shmulevich et al*., [24, 25] addressed this idea by modifying the Boolean network model such that an entity can have several regulation functions, each of which is given a probability based on its compatibility with prior data. At each time step, every entity is subjected to a regulation function that is randomly selected according to the defined probabilities. Hence the model is stochastic and an initial global state can lead to many trajectories of different probabilities. Probabilistic boolean network yield a subclass of Markovian genetic networks in which the Markov chain state space is composed of gene vectors and each Markov chain state space corresponding to a probabilistic boolean network is context specific. It is composed of the current gene vector occupied as well as the current constituent Boolean network. For example, a PBN was used to 15 gene sub network model that was inferred from human glioma expression data [24, 25]. This analysis demonstrates that the stationary distributions of entities may indicate possible regulatory relationships among them: entities that have the same

states in a significant proportion of the global states are likely to be related. As the number of global states in the gene sub network was prohibitively large, one study estimated the stationary distribution by sampling the global states.

LIMITATION: Difficult to apply for large scale of network and High computational complexity and also we cannot cope with instantaneous interactions between variables.

4. Continuous Models

Biological experiments usually produce continuous, rather than discrete valued, measurements. Examples include reaction rates, cell mass [32,33], cell cycle length and gene expression intensities. Logical models require discretization of the real valued data, which reduces the accuracy. Continuous models, using real valued parameters over a continuous timescale, allow a straightforward comparison of the global state and experimental data and can theoretically be more accurate. In practice, however, quantitative measurements are almost Always partial (that is, they cover only a fraction of the system's entities). Therefore, some of the parameters of continuous models are usually based on estimations or inference.

4.1 Linear Model

The defining property of linear models is that each regulator contributes to the input of the regulation function independently of the other regulators, in an additive manner [22]. In other words, the change in the level of each entity depends on a weighted linear sum of the levels of its regulators. This assumption allows a high level of abstraction and efficient inference of network structure and regulation functions.

A biological system can be considered to be a state machine, where the change in internal state of the system depends on the current internal state plus any external inputs. The mRNA levels from an important part of the internal state of a cell (ideally, we also want to measure protein levels, metabolites, etc). As a first approximation, we fit the expression data with a purely linear model, where the change in expression level of each mRNA species is derived as a weighted sum of the expression. Levels of all other genes. Of course, a linear model can never be much more than a caricature of the real system, but perhaps we can still draw some interesting conclusions from it. The basic linear model is of the form:

$$X_i(t + \Delta t) = \sum_j W_{ij} X_j (t) \qquad (2)$$

Where $X_i(t + \Delta t)$ is the expression level of gene i at time (t + Δt), and W_{ij} indicates the strength of the interface between *j* and i.

LIMITATION: Linear additive regulation models revealed certain linear relation in regulatory systems but failed to capture nonlinear dynamics aspects of genes regulation. When higher sensitivity to detail is desired, more complex non-linear models are preferable.

4.2 Ordinary Differential Equation Based Model

Ordinary differential equation (ODE) models use continuous variables, while many other methods use discrete variable models. This is a popular tool to model dynamic system gene regulation [34].

In order to analyses network dynamics, locate limit cycles or investigate bifurcation behavior, ODE are the best approach for non-linear systems. A number of options are available in the extensive literature on different ODE models, such as parameter values, the number of equations and kinetic law in functional form. The differential equation models are more accurate compared to discrete variables. This is because they represent underlying physical phenomena by virtue of their use of continuous variables.

$$\frac{dx_i}{dt} = fi(x_1, x_2, \dots \dots x_n, p, u) \qquad (3)$$

In which x is the expression level of gene i at time t (the independent variable), *N* represent the number of genes, *u* is the external perturbation to the system and P indicates the parameter set of the system. In ODE models, which use continuous time variables with constraints, there is no negative value are allowed in concentration, such as for protein and mRNA molecules. The degradation of mRNA or proteins is assumed to not be regulated.

The rate of change in concentration of a particular transcript is given by an influence function of other RNA concentrations. The non-linear differential equations describe the mutual activating and repressing influences of genes in a GRN at a high-level of abstraction. In particular, it is assumed that the rate of gene expression depends exclusively on the concentration of gene products arising from the nodes (genes) of the GRN. This means that the influence of other molecules (e.g., transcription factor) and cellular processes (translation) is not taken into account directly. Even with these limitations, dynamic GRN models of this kind can be useful in deciphering basic aspects of gene regulatory interactions.

Reverse-engineering algorithms based ODEs relate changes in gene transcript concentration to each other and to an external perturbation. An experimental treatment can alter the transcription rate of genes in a cell i.e. external perturbation. This approach produces signed directed graphs which can be applied to both steady-state and time-series expression profiles. ODE can be used to predict the behaviour of a network under different conditions by using the parameter θ_j for all i known.

$$x_i(t) = f_i(x_1, \dots\dots\dots, x_N, u, \theta_i) \qquad (4)$$

Where θ_i indicates the interactions among genes, $i = 1....N$, $x_i(t)$ describes the concentration of transcript *i* measured at time *t. u* is the external perturbation of the gene node. Reverse-engineering techniques are capable of correctly inferring regulatory interactions between genes [35].

Polynikis et al. [36] aimed to compare different ODE modelling approaches in gene regulatory networks. The model obtained from different modeling approaches let to conflicting conclusions on the existence as well as the stability of equilibria and stable oscillatory behavior. Continuous time variables represented the concentration of protein, mRNA and molecules in

ordinary differential equations. A constraint existed whereby negative concentration was not allowed. Typical translation and transcription process are described as:

$$\textit{Transcription:} \quad \frac{dr_i}{dt} = F\,(f_i^R(p_1), (f_i^R(p_2)..... f_i^R(p_n)) - \Upsilon_i r_i \qquad (5)$$

$$\textit{Translation:} \quad \frac{dp_i}{dt} = f_i^p(r_i) - |\,\delta_i p_i \qquad (6)$$

i: Any given gene;
r_i *:Rate of change of the concentration of the transcribed mRNA;*
p_i*: Rate of change of the concentration of translated protein.*

A quasi- steady state assumption on mRNA dynamics provides a simplified model after a complete non-linear ODE model from which description of protein and mRNA are derived. The occurrence of a hopf bifurcation leading to persistent oscillatory behaviour is shown in the complete non-linear model.

Comparison of different models discussed in this paper is given in Table 1 and freely available software (Tools) used to build models is given in Table 2.

LIMITATIONS: Unless they are restricted to simple function forms, differential equation models involve a large number of parameters (d^2) parameters where d is the number of genes modeled. Moreover, differential equation models require time-series data to learn the parameters.

Table 1. Advantages and disadvantages of the different algorithms for gene network construction.

TECHNIQUE	ADVANTAGES	DISADVANTAGES
Logical and Boolean Networks	A simplistic logical formalism can represent realistic complex biological phenomena such as cellular state dynamics that exhibit switch-like behavior, stability and hysteresis.	Boolean: Two states are not sufficient for the levels of real gene expressions. The updates of the network states in this model are synchronous, whereas biological networks are typically asynchronous.
Probabilistic Boolean Networks	It is stochastic overcome the deterministic rigidity of Boolean networks. They are able to cope with uncertainty both in data and in the model selection.	Difficult to apply for large scale of network and high computational complexity and also we cannot cope with instantaneous interactions between variables.
Linear Models	Linear models do not require extensive knowledge about regulatory mechanisms. It can be used to obtain qualitative insights about regulatory networks.	Failed to capture nonlinear dynamics aspects of gene regulations.
Differential Equation Based Models	Simple homogeneous structure: this allows the settings of parameter discovering softwear to be easily customized for these structures.	Involve a large numberv of parameters (d^2), Where d is the number of genes modeled.

Table 2: Freely available software used to build and analyse models

Name	Features	License	Refs
Gene regulatory network inference			
ARACNE	Information theoretic	Non-commercial license	70
Banjo	Bayesian inferences	Non-commercial license	73
CatNet	Bayesian inferences	General public license	–
Inferelator	ODEs	No license	77
NAIL	Multiple	Apache license	–
NIR	ODEs	Non-commercial license	78
TIGRESS	Regression	General public license	67
Quantitative kinetic modelling			
BIOCHAM	ODEs	General public license	141
CellDesigner	ODEs; stochastic	Free	142
COPASI	ODEs; stochastic	Artistic license	86
DBSolve	ODE	Free	–
E-Cell Project	ODEs; stochastic	General public license	125
iBioSim	ODEs; stochastic	MIT License	143
SBMLsimulator	ODEs	Lesser general public license	62
XPP-Aut	ODEs	General public license	144
Qualitative modelling			
BoolNET	Logic models	Artistic license	145
CellNetOptimizer	Logic models	General public license	121
GINsim	Logic models	General public license	140
Genetic Network Analyzer	Piecewise linear equations	Free for non-profit academic research	105

CONCLUSION:

In this review, several commonly used computational approaches for constructing gene regulatory networks are reviewed, as well as their recent developments and application. **Table 1** shows the advantages and disadvantages for each inference approaches discussed in this review and also we provide **Table 2** is freely available software used to build and analyse the models. As the significance of gene regulatory networks in biological field increase over the years, many computational approaches and models have been developed to construct gene regulatory networks from gene expression data.

Most of the approaches or models share the same principles, which include emphasising the reverse engineering paradigm [44]. Recent studies also showed that gene regulatory networks can be used to highlight genes and proteins that are involved with cell differentiation [45].

As biological research fields began their transition into a new perspective of holistic research, biological data is increasing exponentially at unprecedented speed. Along with the advance of technology, gene regulatory network construction serves as the catalyst to propel future promising achievements in system biology or in bioinformatics.

References

1. Karlebach, Guy, and Ron Shamir. "Modelling and analysis of gene regulatory networks." Nature Reviews Molecular Cell Biology 9.10 (2008): 770-780.
2. Levine, Michael, and Eric H. Davidson. "Gene regulatory networks for development." Proceedings of the National Academy of Sciences of the United States of America 102.14 (2005): 4936-4942.
3. Hasty, Jeff, et al. "Computational studies of gene regulatory networks: in numero molecular biology." Nature Reviews Genetics 2.4 (2001): 268-279.
4. Swain, Martin T., Johannes J. Mandel, and Werner Dubitzky. "Comparative study of three commonly used continuous deterministic methods for modeling gene regulation networks." BMC bioinformatics 11.1 (2010): 1.
5.Wessels, Lodewyk FA, Eugene P. van Someren, and Marcel JT Reinders. "A comparison of genetic network models." pacific Symposium on Biocomputing. Vol. 6. No. 4. 2001.
6 Cho, K-H., et al. "Reverse engineering of gene regulatory networks." IET Systems Biology 1.3 (2007): 149-163.
7. De Jong, Hidde. "Modeling and simulation of genetic regulatory systems: a literature review." Journal of computational biology 9.1 (2002): 67-103.
8.Potkin, Steven G., et al. "Identifying gene regulatory networks in schizophrenia." Neuroimage 53.3 (2010): 839-847.
9.Wilczynski, Bartek, and Eileen EM Furlong. "Challenges for modeling global gene regulatory networks during development: insights from Drosophila." Developmental biology 340.2 (2010): 161-169.
10. Hori, Yutaka, Tae-Hyoung Kim, and Shinji Hara. "Existence criteria of periodic oscillations in cyclic gene regulatory networks." Automatica 47.6 (2011): 1203-1209.
11. Liu, Guixia, et al. "Combination of neuro-fuzzy network models with biological knowledge for reconstructing gene regulatory networks." Journal of Bionic Engineering 8.1 (2011): 98-106.
12. de Matos Simoes, Ricardo, Matthias Dehmer, and Frank Emmert-Streib. "Interfacing cellular networks of S. cerevisiae and E. coli: Connecting dynamic and genetic information." BMC genomics 14.1 (2013): 1.
13. Schadt, Eric E. "Molecular networks as sensors and drivers of common human diseases." Nature 461.7261 (2009): 218-223.
14. Emmert‐Streib, Frank, and Galina V. Glazko. "Network biology: a direct approach to study biological function." Wiley Interdisciplinary Reviews: Systems Biology and Medicine 3.4 (2011): 379-391.
15. Guan, Yuanfang, et al. "Tissue-specific functional networks for prioritizing phenotype and disease genes." PLoS Comput Biol 8.9 (2012): e1002694.
16. Emmert-Streib, Frank, et al. "The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks." Frontiers in genetics 5 (2014): 15.
17. Glass, Leon, and Stuart A. Kauffman. "The logical analysis of continuous, non-linear biochemical control networks." Journal of theoretical Biology 39.1 (1973): 103-129.

18. Thomas, René. "Boolean formalization of genetic control circuits." Journal of theoretical biology 42.3 (1973): 563-585.
19. Davidson, Eric H., et al. "A genomic regulatory network for development." science 295.5560 (2002): 1669-1678.
20. Smith, Joel, Christina Theodoris, and Eric H. Davidson. "A gene regulatory network subcircuit drives a dynamic pattern of gene expression." Science 318.5851 (2007): 794-797.
21. Davidich, Maria I., and Stefan Bornholdt. "Boolean network model predicts cell cycle sequence of fission yeast." PloS one 3.2 (2008): e1672.
22. Filkov, Vladimir. "Handbook of computational molecular biology." (2005).
23. Fauré, Adrien, et al. "Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle." Bioinformatics 22.14 (2006): e124-e131.
24. Shmulevich, Ilya, et al. "Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks." Bioinformatics 18.2 (2002): 261-274.
25. Shmulevich, Ilya, et al. "Steady‐state analysis of genetic regulatory networks modelled by probabilistic Boolean networks." Comparative and functional genomics 4.6 (2003): 601-608.
26. Mason, Oliver, and Mark Verwoerd. "Graph theory and networks in biology." IET systems biology 1.2 (2007): 89-119.
27. Kingsmore, Stephen F. "Multiplexed protein measurement: technologies and applications of protein and antibody arrays." Nature reviews Drug discovery 5.4 (2006): 310-321.
28. Cao, Jiguo, Xin Qi, and Hongyu Zhao. "Modeling gene regulation networks using ordinary differential equations." Next Generation Microarray Bioinformatics: Methods and Protocols (2012): 185-197.
29. Bansal, Mukesh, et al. "How to infer gene networks from expression profiles." Molecular systems biology 3.1 (2007): 78.
31. Ross, Ian L., Catherine M. Browne, and David A. Hume. "Transcription of individual genes in eukaryotic cells occurs randomly and infrequently." Immunology & Cell Biology 72.2 (1994).
32. Bailone, Adriana, Alain Levine, and Raymond Devoret. "Inactivation of prophage λ repressor in vivo." Journal of molecular biology 131.3 (1979): 553-572.
33. Csete, Marie E., and John C. Doyle. "Reverse engineering of biological complexity." science 295.5560 (2002): 1664-1669.
34. Bodaker, Matan, et al. "Genes related to differentiation are correlated with the gene regulatory network structure." Bioinformatics (2013): btt685.
35. Schultz, Daniel, et al. "Molecular level stochastic model for competence cycles in Bacillus subtilis." Proceedings of the National Academy of Sciences 104.45 (2007): 17582-17587.
36. Polynikis, Athanasios, S. J. Hogan, and Mario di Bernardo. "Comparing different ODE modelling approaches for gene regulatory networks." Journal of theoretical biology 261.4 (2009): 511-530.
37. Kim, Haseong, Jae K. Lee, and Taesung Park. "Boolean networks using the chi-square test for inferring large-scale gene regulatory networks." BMC bioinformatics 8.1 (2007): 1.
38. Dorigo, Marco. "Learning by probabilistic Boolean networks." Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on. Vol. 2. IEEE, 1994.
39. Lähdesmäki, Harri, et al. "Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks." Signal Processing 86.4 (2006): 814-834.
40. Pearl, J. "Morgan Kaufmann series in representation and reasoning. Probabilistic reasoning in intelligent systems: Networks of plausible inference." (1988).
41. Friedman, Nir, et al. "Using Bayesian networks to analyze expression data." Journal of computational biology 7.3-4 (2000): 601-620.
42 Du, Pan, et al. "Modeling gene expression networks using fuzzy logic." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 35.6 (2005): 1351-1359.
43. Maraziotis, Ioannis A., Andrei Dragomir, and Dimitris Thanos. "Gene regulatory networks modelling using a dynamic evolutionary hybrid." BMC bioinformatics 11.1 (2010): 1.
44. Csete, Marie E., and John C. Doyle. "Reverse engineering of biological complexity." science 295.5560 (2002): 1664-1669.
45. Bodaker, Matan, et al. "Genes related to differentiation are correlated with the gene regulatory network structure." Bioinformatics (2013): btt685.
46. Schadt, Eric E. "Molecular networks as sensors and drivers of common human diseases." Nature 461.7261 (2009): 218-223.