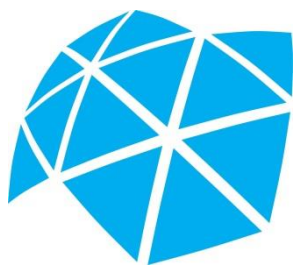


**Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências de Botucatu
Programa de Pós-Graduação em Biotecnologia**

Modelagem do ciclo celular e influência dos lncRNAs em Saccharomyces cerevisiae expostas a altas concentrações de etanol.

Lucas Cardoso Lázari

**Botucatu – SP
2020**



**Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências de Botucatu
Programa de Pós-Graduação em Biotecnologia**

Modelagem do ciclo celular e influência dos lncRNAs em Saccharomyces cerevisiae expostas a altas concentrações de etanol.

**Mestrando: Lucas Cardoso Lázari
Orientador: Dr. Guilherme Targino Valente**

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia do Instituto de Biociências de Botucatu da Universidade Estadual Paulista “Júlio de Mesquita Filho”, para obtenção do título de mestre.

**Botucatu – SP
2020**

L431m Lázari, Lucas Cardoso
Modelagem do ciclo celular e influência dos lncRNAs em *Saccharomyces cerevisiae* expostas a altas concentrações de etanol. / Lucas Cardoso Lázari. -- Botucatu, 2020
117 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Botucatu
Orientador: Guilherme Targino Valente

1. Modelagem. 2. Ciclo celular. 3. lncRNAs. 4. *Saccharomyces cerevisiae*. 5. Bioinformática. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

“There is a fine line between consideration and hesitation. The former is wisdom, the latter is fear. When the time comes to strike, an emperor strikes without hesitation, to entertain doubt is to dance with death’
- *Izaro, Path Of Exile.*

Agradecimentos

Agradeço à minha família por todo o apoio que tenho recebido desde sempre. Aos meus pais, Lígia Cruz Cardoso Lázari e Ovídio Lázari Júnior, por terem cuidado de mim sempre, por não terem deixado faltar nada e por toda motivação dada para eu seguir estudando. À minha irmã, Priscilla Cardoso Lázari, por todo o apoio e ajuda que me deu durante a faculdade e o mestrado, além de ser um exemplo de pessoa e profissional que eu tento seguir sempre. À minha avó, Lygia Prudente Cruz Cardoso, por toda a ajuda que recebi durante minha vida em Botucatu e também pelas visitas constantes. Tive muita sorte de tê-los na minha vida, tudo o que consegui fazer até hoje foi graças à ajuda e presença de vocês, sou muito grato! Vocês são muito importantes para mim, e eu espero poder ajudar mais vocês daqui para frente, da mesma maneira que vocês me ajudaram até hoje.

Agradeço aos meus amigos pela companhia quase diária. Principalmente ao pessoal do SBGL, que deixaram o meu período de mestrado muito mais agradável. Principalmente me forçando a sair de casa – uma tarefa extremamente difícil devido à inércia – para ir aos bares ou fazer qualquer outra coisa. Tenho que admitir que realizaram esse feito com excelência.

Agradeço ao meu orientador, Guilherme Targino Valente, por ter me ajudado a crescer profissionalmente e pessoalmente. Agradeço pelas críticas construtivas e feitas sempre com o intuito de me ajudar a crescer como cientista. Você foi a primeira pessoa durante o período acadêmico que realmente me ajudou a corrigir alguns dos (vários) erros que eu nem sabia que tinha, tanto profissionalmente quanto pessoalmente. A sinceridade das conversas que você tinha comigo foi essencial para tudo isso acontecer. Somente através de um diálogo sincero que é possível diagnosticar os erros e corrigi-los, e sou grato por todas as conversas sinceras que tivemos (embora o meu ego tenha saído ferido na maior parte das vezes, mas ninguém é de ferro, né?). Te admiro como pessoa e como cientista.

Agradeço também a todos os professores da UNESP, principalmente ao professor Rafael Plana Simões, que me ajudou bastante durante o período de mestrado.

Agradeço ao Acaso, pois tudo que tive até hoje foi resultado de muita sorte. Nasci em uma família excelente, tive tudo o que sempre precisei, oportunidades nunca faltaram, tenho amigos muito especiais, trabalhei com pessoas ótimas e não tive nenhum problema causado por puramente azar (apenas os problemas que eu mesmo criei, infelizmente mais do que deveria ter criado, confesso). O Acaso me favoreceu muito ao longo da minha vida, e continuou favorecendo ultimamente. Sinceramente, espero que continue favorecendo no futuro, pois minhas habilidades de planejamento são bem ruins e nem só de bom humor vive o homem.

Agradeço à Universidade Estadual Paulista “Júlio de Mesquita Filho”, pois através dessa universidade pública fui capaz de me graduar e realizar um mestrado. Sou muito grato pela sua existência e espero poder retribuir à sociedade todo o conhecimento obtido.

Agradeço a CAPES pelo suporte financeiro fornecido durante meu período de mestrado. Sem ele eu não seria capaz de me manter durante dois anos aqui em Botucatu, portanto não seria capaz de fazer um mestrado.

Agradeço às intempéries que criei (sem querer, obviamente), pois só a partir do processo de solução delas fui capaz de evoluir como pessoa e identificar que o que mais me prejudicava era uma pequena má interpretação da terceira lei de Newton. De forma geral, agradeço às coisas por serem da maneira que são.

Resumo

A intensa utilização de combustíveis fósseis gera preocupações constantes devido aos impactos de sua combustão ao meio ambiente. Os biocombustíveis são uma alternativa viável aos combustíveis fósseis por apresentarem vantagens como serem menos agressivos ao meio ambiente. O bioetanol é um dos biocombustíveis mais utilizados no mundo e sua produção pode ser feita pela fermentação realizada pela levedura *Saccharomyces cerevisiae*. No entanto, altas concentrações de etanol inibem diversos mecanismos biológicos da levedura, causando a diminuição da produtividade. A partir de resultados prévios, observou-se que o ciclo celular é uma das vias mais afetadas pelo etanol e, além disso, constatou-se a presença de lncRNAs regulando esta via em duas linhagens de *S. cerevisiae*, a BY4742 e SEY6210. Utilizando operadores Booleanos, um modelo lógico discreto foi desenvolvido para o ciclo celular no qual os nós do sistema assumem até quatro valores discretos que representam a quantidade ou o grau de ativação desses nós. O modelo desenvolvido apresentou boa performance preditiva, acertando 87.27% dos 109 fenótipos obtidos da literatura, tornando possível a simulação de novos elementos. Experimentos prévios demonstraram que as leveduras de baixa tolerância ao etanol conseguem retomar o crescimento mais rápido do que as de alta tolerância. Nesse trabalho, simulações feitas com dados de expressão diferencial via RNA-Seq permitiu inferir que isso ocorre porque as linhagens de baixa tolerância sofrem *arrest* na última fase do ciclo celular (fase M), enquanto que as mais tolerantes sofrem *arrest* na primeira fase (fase G1). Os resultados também indicaram que na linhagem SEY6210 (baixa tolerância) a presença de um lncRNA superexpresso permite que esta linhagem complete o ciclo sem sofrer *arrest*. Até o momento, o modelo aqui desenvolvido contempla o maior número de proteínas do ciclo celular.

Palavras-chave: Modelagem lógica, Bioinformática, Ciclo celular, *Saccharomyces cerevisiae*, RNAs longos não codificantes, Tolerância ao etanol

Abstract

The intense use of fossil fuels raised concern about the future due to their negative environmental impact. Bio-fuels are alternatives to the fossil fuels due to be biodegradable and less environmentally harmful. The bio-ethanol is one of the most popular bio-fuel. It can be produced by fermentation using the yeast *Saccharomyces cerevisiae*. However, high ethanol concentration inhibits the yeast decreasing the ethanol yield. Previous data of our groups showed the cell cycle is one of most affected pathways during ethanol stress. Moreover, it was found lncRNAs regulating this pathway in the BY4742 and SEY6210 strains. Using Boolean operators the discrete logical model of the cell cycle was developed. The nodes may get up to four discrete values to represent their abundance of activation degree. This model correctly modeled around 87.27% of correct predictions based on 109 phenotypes from the literature, hence, this model is desirable to predict cell cycle behavior after addition of new elements. According to previous data of our group, the lower tolerant strains recover the normal growth faster than higher tolerant strains after stress relief. The simulations here presented by adding RNA-Seq information into the model, showed a cell cycle arrest at final phase of the cell cycle (M phase) in lower tolerant strains whereas in the higher tolerant ones this arrest occurs at the first phase (G1 phase) during the ethanol treatment. The simulations also indicated that in SEY6210 (low tolerant), the up-regulation of a lncRNA skips the M phase arrest. Finally, this model harbor the highest number of cell cycle's proteins already modeled.

Keywords: long non-coding RNAs, Logical modeling, Cell cycle, Ethanol tolerance, *Saccharomyces cerevisiae*, Bioinformatics

Índice de ilustrações

Figura 1: Sequência de transformações da glicose até o produto final de interesse, o etanol.....	13
Figura 1: Modelo representando as interações entre proteínas do ciclo celular em <i>S. cerevisiae</i>	19
Figura 2: Estrutura do sistema que será modelada pelas equações Booleanas escritas para cada componente da rede.....	31
Figura 3: Grafo de transição de estados de todas as trajetórias do sistema.....	32
Figura 4: Modelo desenvolvido por Li et al. (2004).....	34
Figura 5: Modelo desenvolvido por Fauré et al. (2009).....	36
Figura 6: Modelo de Irons (2009)	38
Figura 7: Estrutura do modelo do ciclo celular desenvolvido por Todd e Helikar (2012).....	40
Figura 8: Estrutura do modelo criado por Alcasabas et al. (2013).	42
Figura 9: Modelo desenvolvido por Rubinstein et al. (2013).....	44
Figura 10: Modelo do ciclo celular desenvolvido nesta dissertação	50
Figura 11: Fluxograma da criação do modelo lógico, adaptado de Wang et al. (2012).....	56
Figura 12: Distribuição do fold-change de todos os transcritos das seis linhagens.....	59
Figura 13: Simulação no modelo de uma levedura wild-type (WT) sem nenhuma perturbação ou checkpoint ativo.....	65
Figura 14: Resultado obtido com a simulação dos dados de transcriptoma das leveduras do grupo HT	71
Figura 15: Resultado obtido com a simulação dos dados de transcriptoma das leveduras do grupo LT as quais, diferente das HTs, sofreram um arresto no final de M.....	71
Figura 16: Resultado obtido para a simulação do lnc_9136 superexpresso no fenótipo LT	73
Figura 17: Resultado obtido pela simulação com a superexpressão constitutiva do lnc_10883 em uma levedura com alto dano ao fuso.	78
Figura 18: Resultados do docking molecular do lnc_9136 com Hsl1p	79
Figura 19: Resultados do docking molecular do lnc_9136 com Gin4p	79
Figura 20: Resultado da predição de estrutura 3D para o lnc_9136.....	81
Figura 20: Resultado obtido pela simulação com a superexpressão constitutiva do lnc_10883 e lnc_9136 como inibidores da ORC	85
Figura 21: Proteínas da sub-rede encontrada na linhagem BMA64-1A.....	96
Figura 21: Simulações do modelo para cada situação.....	98

Índice de tabelas

Tabela 1: Produção de bioetanol de diferentes países no período de 2014-2015	12
Tabela 2: Produção de bioetanol em milhões de galões de 2007 a 2013.....	12
Tabela 3: Principais efeitos do excesso de etanol em leveduras.....	14
Tabela 4: Valores correspondentes à quantidade de genes diferencialmente expressos para cada linhagem	16
Tabela 5: Termos de GO obtidos utilizando os genes diferencialmente expressos (tanto superexpressos quando subexpressos) das linhagens HT e LT.....	16
Tabela 6: Dados de expressão não discretizados utilizados no modelo para diferenciar o fenótipo HT de LT.....	60
Tabela 7: Porcentagem de células que acumulam dano ao DNA.....	63
Tabela 8: Valores de fold-change para cada gene do <i>checkpoint</i> do dano ao DNA para cada linhagem.....	64
Tabela 9: Resultados obtidos da matriz de confusão.....	66
Tabela 10: Algumas mutações testadas no modelo para a validação.	68

Sumário

1. Introdução	11
1.1. Bioetanol	11
1.2. Produção de bioetanol por leveduras	13
1.3. Ciclo celular em <i>Saccharomyces cerevisiae</i>	17
1.4. Modelagem matemática na biologia e modelos do ciclo celular de <i>S. cerevisiae</i>	25
1.5. LncRNAs e suas funções	45
2. Justificativa e objetivos	47
2.1. Objetivos gerais	48
2.2. Objetivos específicos	48
3. Métodos	49
3.1. Definições	49
3.2. Desenvolvimento do modelo lógico	50
3.3. Validação do modelo	57
3.4. Testes dos dados de transcriptoma	59
3.5. <i>Docking</i> molecular	62
3.6. Efeitos do dano ao DNA para cada linhagem	63
3.7. Efeitos dos lncRNAs no complexo ORC	64
4. Resultados e Discussões	64
4.1. Validação do modelo	64
4.2. Simulação dos grupos HT e LT	69
4.3. <i>Docking</i> molecular da interação lnc_9136-Hsl1p/Gin4p	78
4.4. Efeito do dano ao DNA no ciclo celular em cada linhagem	82
4.5. Efeito dos lncRNAs no complex ORC	84
5. Conclusões	85
6. Referências	88
Apêndice 1 – Análise quantitativa	96

1. Introdução

Neste tópico serão abordados os temas relacionados a esta pesquisa com o intuito de demonstrar sua relevância e prover aos leitores informações suficientes para compreender o trabalho. Portanto, será feita uma revisão sobre bioetanol e seu papel no cenário econômico e ambiental atual, seguido por explanações do funcionamento do ciclo celular em *Saccharomyces cerevisiae* e sua relação com a produção de etanol. Dessa forma, será possível compreender como a modelagem pode ajudar no estudo do ciclo celular e, conseqüentemente, na melhora da produção do bioetanol. Uma revisão sobre modelagem matemática de redes também é apresentada para demonstrar suas funções no estudo de sistemas complexos e como ela pode ser útil para este fim. Finalmente, uma revisão sobre RNAs Longos Não Codificantes (lncRNAs) também está presente neste texto, permitindo compreender o papel deles no modelo e como eles podem afetar a função de proteínas.

1.1. Bioetanol

A intensa utilização de combustíveis fósseis demandada pela necessidade de desenvolvimento econômico dos países tem causado uma diminuição progressiva na disponibilidade destes recursos não renováveis gerando problemas econômicos e políticos e tende a piorar em longo prazo. Além disso, é consenso na literatura que estes combustíveis são a principal causa das mudanças climáticas (RASS-HANSEN et al., 2007; SRIVASTAVA et al., 2015). As preocupações acerca do futuro dos combustíveis fósseis e o entendimento dos impactos de sua combustão ao meio ambiente torna necessário o estudo de combustíveis alternativos, procurando aqueles que sejam renováveis e menos agressivos ao meio ambiente. Os biocombustíveis são uma alternativa viável e apresentam vantagens quando comparados aos combustíveis fósseis, tais como: 1- são extraídos de biomassa; 2- são biodegradáveis e, portanto, mais sustentáveis; 3- sua combustão não interfere no ciclo do dióxido de carbono; 4- são menos agressivos ao meio ambiente (GAURAV et al., 2017).

Atualmente, os biocombustíveis mais comuns incluem o bioetanol e biodiesel. A produção destes combustíveis é proveniente das plantações de cana-de-açúcar, beterraba, soja, batata, entre outras (GHADIRYANFAR et al., 2016). O bioetanol é um dos biocombustíveis mais utilizados no mundo, principalmente pela possibilidade de misturá-lo com a gasolina, diminuindo o custo do produto e as emissões de CO₂ (diminuição de 90% das emissões de CO₂ e 60-80% de SO₂ quando misturado com 95% de gasolina) (GUPTA; VERMA, 2015).

No contexto global, há sempre uma tendência de aumento na produção de etanol ao longo dos anos, sendo o Brasil e Estados Unidos líderes na produção de etanol (Tabela 1 e Tabela 2) (NIPHADKAR; BAGADE; AHMED, 2018).

Tabela 1: Produção de bioetanol de diferentes países no período de 2014-2015 (adaptado de Niphadkar et al., 2018).

País	Bioetanol em milhões de galões (valor aproximado)
EUA	15.000
Brasil	65.000
Europa	1.369
China	695
Índia	555
Canadá	522
Tailândia	310
Argentina	160
Outros	726

Tabela 2: Produção de bioetanol em milhões de galões (valor aproximado) de 2007 a 2013 (adaptado de Niphadkar et al., 2018).

País	2007	2008	2009	2010	2011	2012	2013
EUA	6.521	9.309	10.938	13.298	13.948	13.000	13.300
Brasil	5.019	6.472	6.578	6.922	5.573	5.577	6.267
Europa	570	734	1.040	1.209	1.168	1.179	1.371
China	486	502	542	542	555	555	696
Canadá	211	238	291	357	462	449	523
Outros	315	389	914	985	698	752	1272

O Brasil foi recentemente considerado líder global na produção e desenvolvimento tecnológico do bioetanol. Atualmente, é o maior produtor de etanol a partir da cana-de-açúcar. No entanto, como apontado nas tabelas acima, os EUA lidera a produção, sendo que somada à produção brasileira atingem 85% da produção mundial de etanol (SALLES-FILHO et al., 2017). Portanto, visto que o etanol tem uma grande importância econômica e ambiental, é essencial estudar maneiras de aumentar sua produção, melhorando assim a sua competitividade frente aos combustíveis fósseis.

1.2. Produção de bioetanol por leveduras

A produção de bioetanol provém da capacidade das leveduras e outros micro-organismos em converter moléculas de seis carbonos em moléculas de dois carbonos, gerando CO₂ e etanol em vez de lactato, como ocorre nos seres humanos. O processo se inicia na glicólise, a qual a glicose é degradada por uma série de enzimas gerando duas moléculas de piruvato. Os piruvatos são então descarboxilados pela piruvato-descarboxilase formando aldeído, o qual é reduzido a etanol pela álcool desidrogenase (Figura 1) (NELSON; COX, 2014).

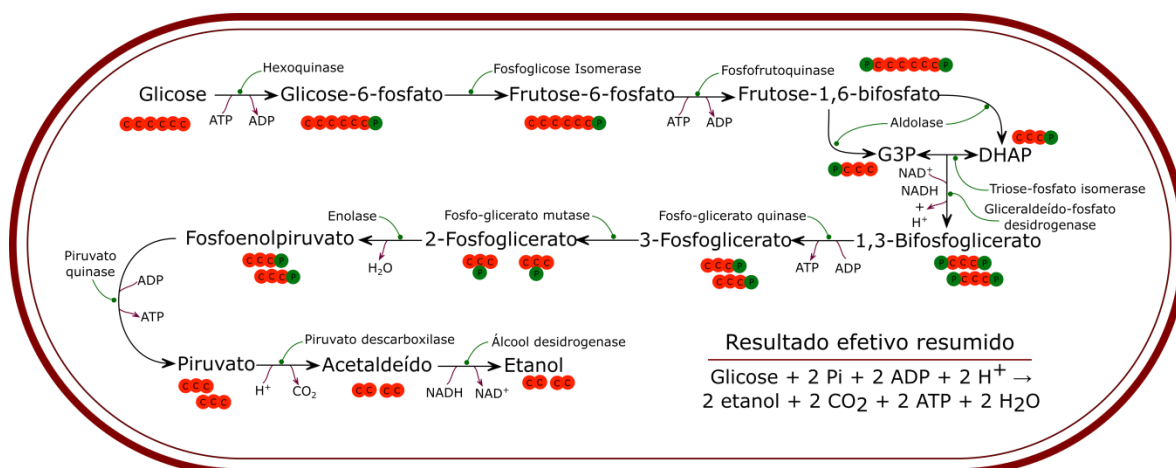


Figura 1: Sequência de transformações da glicose até o produto final de interesse, o etanol. As esferas em laranja e verde representam, respectivamente, o fluxo de carbono e grupos fosfato ao longo do processo.

Dentre as leveduras disponíveis para a produção do bioetanol, a *Saccharomyces cerevisiae* é a mais comumente utilizada. Embora ela seja capaz de produzir etanol com

alto rendimento, a levedura passa a sofrer estresse com o aumento da temperatura e concentração de etanol durante o processo fermentativo, causando perdas da viabilidade celular e diminuição do crescimento (MOHD AZHAR et al., 2017).

O etanol em baixas concentrações já é capaz de inibir o crescimento celular, enquanto que em altas concentrações os efeitos tóxicos do etanol passam a ser mais severos, aumentando a taxa de morte celular do micro-organismo e reduzindo sua vitalidade (atividade e metabolismo da célula) (HU, X. H. et al., 2007). Também, pode causar modificações no metabolismo celular, na biosíntese de macromoléculas, diminuição produção de RNAs e proteínas, aumento na frequência de mutações e alteração na viabilidade e atividade de proteínas intracelulares (Tabela 3) (STANLEY et al., 2010).

Tabela 3: Principais efeitos do excesso de etanol em leveduras (adaptado de Stanley et al., 2010).

Viabilidade e crescimento celular	Inibição do crescimento, da divisão celular e de sua viabilidade; Diminuição do volume celular.
Metabolismo	Diminuição nos níveis de mRNAs e proteínas; Desnaturação e redução de atividade de proteínas; Estímulo para produção de proteínas de resposta ao estresse.
Estrutura celular e função de membrana	Alteração na morfologia do vacúolo; Inibição da endocitose; Perda de gradiente eletrostático e força próton motriz; Inibição de processos de transporte; Inibição da atividade H ⁺ -ATPase; Aumento na fluidez de membrana.

Os mecanismos de resposta ao estresse causado pelo etanol são extremamente complexos e afetam uma variedade enorme de genes e proteínas. Chandler et al. (2004) demonstraram que após uma hora de exposição ao etanol, um total de 374 genes tiveram sua transcrição alterada, sendo que 100 genes apresentaram aumento de transcrição enquanto que 274 tiveram uma diminuição. O número de genes desregulados varia de acordo com o tempo de exposição: após três horas de exposição ao etanol o número de

genes superexpressos caiu para 14, e o número de genes subexpressos caiu para 99 (CHANDLER et al., 2004).

No trabalho publicado por Teixeira et al. (2009), 254 genes foram identificados como responsáveis por atuarem na resistência ao etanol. Analisando a ontologia desses genes candidatos, o grupo identificou que a maioria deles está relacionada com os termos: 1- organização e biogênese celular; 2- estabelecimento da localização celular; 3- transporte intracelular; 4- biogênese e organização de organelas; 5- transporte vacuolar; 6- organização e biogênese de peroxissomos; 7- direcionamento de proteínas para o peroxissomo; 8- biogênese e organização do endossomo; 9- transcrição pelo promotor da RNA polimerase II; 10- biogênese e organização do citoesqueleto (TEIXEIRA et al., 2009).

Ambos os trabalhos apresentados acima utilizaram dados de *microarray* para a identificação de genes importantes para o mecanismo de resistência ao etanol. Conforme o exposto, é possível observar a complexidade desse mecanismo, portanto, estudá-los e encontrar maneiras de desenvolver leveduras cada vez mais resistentes é ainda um desafio para a comunidade científica.

A complexidade fica ainda mais evidente quando tratamos de dados de RNA-Seq, nos quais a quantidade de dados é claramente muito maior. Wolf (2019), obteve o transcriptoma de 6 linhagens de *S. cerevisiae* tanto na condição controle (sem exposição à altos níveis de etanol) quanto tratada. No tratamento citado, foi determinada a quantidade máxima de etanol tolerada para cada linhagem, permitindo separá-las em grupos de alta tolerância (HT) e baixa tolerância (LT) ao etanol (ALMEIDA, 2017). Observou-se que, em média, 50.31% dos genes diferencialmente expressos estão superexpressos e 49.68% estão subexpressos nas HTs, enquanto que nas LTs esse número é de 47.76% e 52.23% respectivamente (Tabela 4). A análise dos GOs dos genes diferencialmente expressos reporta um total de 2.672 termos, os quais foram sumarizados para encontrar os termos que fossem mais relevantes no contexto de resistência ao etanol (Tabela 5) (WOLF, 2019).

Tabela 4: Valores correspondentes à quantidade de genes diferencialmente expressos para cada linhagem (WOLF, 2019).

Linhagem	Grupo	Superexpressos	Subexpressos	Total
BMA64-1A	HT	1319	1240	2559
BY4742	HT	2050	2117	4167
X2180-1A	HT	2144	2126	4270
S288c	LT	2277	2552	4829
BY4741	LT	2116	2256	4372
SEY6210	LT	1640	1795	3435

Tabela 5: Termos de GO obtidos utilizando os genes diferencialmente expressos (tanto superexpressos quando subexpressos) das linhagens HT e LT (WOLF, 2019).

	Alta tolerância	Baixa tolerância
Superexpressos	Energia proveniente da oxidação de compostos orgânicos; Resposta a substâncias tóxicas; Metabolismo de cofatores; Metabolismo de carboidrato celular; Metabolismo de carboidratos; Geração de metabólitos precursores e energia; Metabolismo da coenzima de oxidorredução; Metabolismo composto contendo pirimidina; Reenovelamento de proteínas.	Transporte por transmembrana; Crescimento celular;
Subexpressos	Reparo de rupturas na dupla fita; Mudança na estrutura do DNA; Segregação cromossômica; Regulação biológica; Processos celulares.	Metabolismo de tRNAs; Transporte de RNAs; Biogênese de complexos de ribonucleoproteínas; Processos celulares.

Adicionalmente, foram realizadas análises de enriquecimentos de vias utilizando os dados de expressão diferencial. Nesta etapa foi observado que as vias do ciclo celular, tradução de RNAs e mecanismos de transporte, foram as mais enriquecidas (WOLF, 2019). Além disso, foi observado que linhagens menos tolerantes ao etanol crescem mais rápido e também morrem menos do que as mais tolerantes (ALMEIDA, 2017). Portanto, está claro

que o efeito tóxico do etanol é extremamente complexo, afetando uma variedade enorme de processos celulares tornando difícil o estudo desses mecanismos e o desenvolvimento de leveduras mais tolerantes.

O crescimento celular foi associado com o estresse causado pelo etanol em outros trabalhos. De fato, genes como BUB1 (YGR188C), CDH1 (YGL003C), CLN3 (YAL040C), SWE1 (YJL187C), GRR1 (YJR090C), SWI4 (YER111C) e SWI6 (YRL182W), os quais estão presentes na via que controla o ciclo celular em *S. cerevisiae*, já foram descritos na literatura como genes importantes para o crescimento celular na presença de etanol na concentração de 11% v/v (KUBOTA et al., 2004). Esta informação nos indica que a célula controla a frequência em que se multiplica para contornar o estresse a qual está submetida, logo, estudar a fundo essa via é essencial para desenvolver maneiras de fazer a célula permanecer viável e se multiplicar mesmo na presença do estresse.

O ciclo celular foi escolhido como objeto de estudo deste trabalho dado a descoberta da sua importância no mecanismo de resposta ao etanol e ao fato de haver quatro lncRNAs que interagem com proteínas dessa via nas linhagens SEY6210 e BY4742 (MARQUES, 2019).

1.3. Ciclo celular em *Saccharomyces cerevisiae*

O ciclo celular é o conjunto de fases na qual uma célula deve passar para se multiplicar. Para ocorrer a divisão, a célula cresce continuamente em massa mesmo quando o ciclo celular já foi iniciado. Alguns eventos do ciclo ocorrem apenas uma vez na vida da célula, tais como a replicação do genoma (fase S), a mitose e a divisão celular. Tais eventos acontecem dentro de quatro fases: G1 (*gap1*), S (*synthesis*), G2 (*gap2*) e M (*mitosis*). As fase G1 é responsável por preparar a célula para a síntese de DNA (fase S) e a fase G2 prepara a célula para a mitose (fase M) e divisão celular (ALBERGHINA et al., 2012).

É essencial que os “eventos contínuos do ciclo celular” (aumento de massa) e os “eventos descontínuos do ciclo de divisão do DNA” (replicação do DNA, mitose e divisão celular) aconteçam de forma coordenada a fim de manter a homeostase celular (BARBERIS et al., 2007). O aumento de massa possui um papel importante no ciclo, pois apenas as células que atingem um valor crítico de massa podem iniciar a fase S e, sendo assim, as células que ainda não possuem o tamanho ideal passam por uma fase G1 mais prolongada (RUPE, 2002). Outra característica essencial do ciclo celular é a capacidade de mudar seu caminho de acordo com certas condições; em *S. cerevisiae*, a célula possui três alternativas: 1- seguir o ciclo até a divisão; 2- em ambientes com falta de nutrientes, a célula entra em numa fase estacionária resistente ao calor e a tratamentos químicos; 3- as células haploides podem realizar o *mating* para formar uma célula diploide, a qual pode continuar seu próprio ciclo celular (FORSBURG, 1991).

A progressão do ciclo celular é controlada pelas quinases dependente de ciclinas (CDK). CDKs são treonina e serina quinases e o modo como elas controlam o ciclo celular depende da interação com as ciclinas, sendo estas as suas subunidades de ativação. A disponibilidade das ciclinas no tempo apropriado provoca a progressão do ciclo celular, e essa disponibilidade é controlada pela síntese e degradação destas ciclinas. Em *S. cerevisiae*, a CDK Cdc28p regula a progressão do ciclo celular ligando-se a diferentes ciclinas em diferentes etapas do ciclo celular. Diferentes ciclinas possuem diferentes maneiras de regulação (tanto transcricionalmente quanto pós-transcricionalmente), embora possuam algumas funções que se sobrepõem. Porém, mesmo com funções sobrepostas, as ciclinas são suficientemente diferentes umas das outras, permitindo que o ciclo celular ocorra coordenadamente e nos tempos apropriados (BLOOM; CROSS, 2007).

O ciclo celular é complexo e envolve uma quantidade considerável de genes e proteínas. A progressão do ciclo, a presença de *checkpoints* e a complexa interação entre as ciclinas e seus ativadores e inibidores são diferentes em cada fase. A Figura 1 representa o

ciclo celular em *S. cerevisiae* retirado do KEGG (*Kyoto Encyclopedia of Genes and Genomes*), modelo este utilizado como base para a construção do modelo matemático descrito neste trabalho.

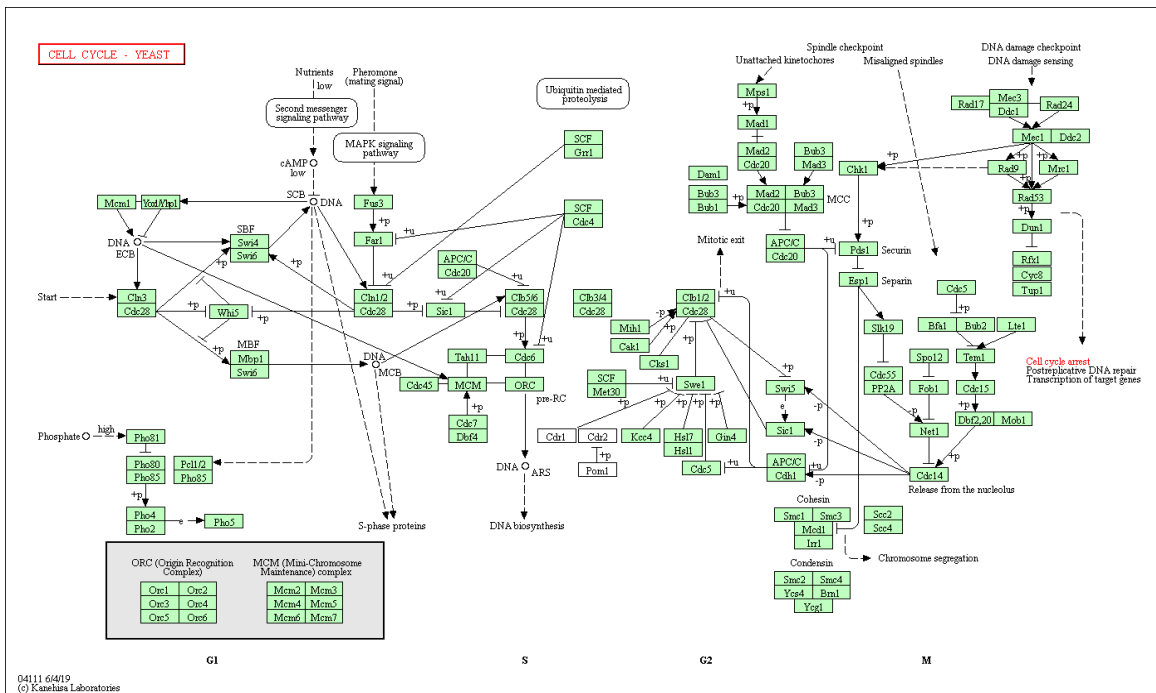


Figura 1: Modelo representando as interações entre proteínas do ciclo celular em *S. cerevisiae*, as setas representam interações de ativação e as retas interações de inibição. Os sinais "+p" e "+u" representam, respectivamente, fosforilação e ubiquitinação (LABORATORIES, 2019).

START e a transição de G1 a S

Como já mencionado, a massa celular é importante para definir se a célula irá se dividir. Além disso, outro requerimento para que as células iniciem o ciclo (*START*) é a necessidade de estarem em constante crescimento e síntese proteica. O *START* necessita de pelo menos uma das ciclinas do tipo Cln (Cln1p, Cln2p ou Cln3p, chamadas de ciclinas do G1) as quais ativam a quinase Cdc28p. As ciclinas do G1 são instáveis implicando que a célula deva continuamente sintetizar essas proteínas para que o *START* ocorra (FUTCHER, 1996). A Cln3p pode ser detectada ao longo do ciclo celular, mas sua expressão é maior no final da fase M e início de G1. A transcrição das outras ciclinas do G1 é controlada pelos fatores de transcrição SFB (Swi4p e Swi6p) e MBF (Mbp1p e Swi6p), sendo que ambos são responsáveis pela expressão de Cln1p e Cln2p produzindo um *feedback* positivo irreversível de ativação do SFB e MBF. Estes dois fatores de transcrição

também ativam Clb5p e Clb6p (duas ciclinas da fase S) (BLOOM; CROSS, 2007). Uma das razões pela qual o Cln3p é a única ciclina expressa no início do ciclo celular é a sua capacidade de fosforilar Whi5p, um repressor que se liga ao promotor de SBF e MBF para impedir a ação destes fatores de transcrição. A fosforilação de Whi5p causa sua inativação e permite a ação de SBF e MBF na transcrição de outros genes (COSTANZO et al., 2004; DE BRUIN et al., 2004). É importante ressaltar que a expressão de Cln3p é feita pela ação do fator de transcrição Mcm1p (PRAMILA et al., 2002).

As ciclinas Cln1p e Cln2p possuem papéis importantes para a polarização da actina cortical, algo essencial para a formação do broto (um crescimento lateral na célula mãe). No entanto, na ausência de Cln1p ou Cln2p, a Clb5p pode suprir suas funções. Portanto, a Clb5p pode causar o *START* do ciclo celular na ausência das três ciclinas do G1 (Cln1p, Cln2p e Cln3p) (CVRCKOVÁ; NASMYTH, 1993; EPSTEIN; CROSS, 1992).

A iniciação da replicação do DNA acontece logo após a ação das Clns. Para ocorrer de maneira apropriada, há necessidade da ação das ciclinas do tipo B (Clb5p ou Clb6p), com suas transcrições necessariamente ocorrendo entre G1 e S. Na ausência destas ciclinas, Clb1p ou Clb2p podem ativar a replicação do DNA, porém com um atraso (BLOOM; CROSS, 2007). A presença das Clbs na transição G1/S não é o suficiente para dar início à replicação, pois o Sic1p é um inibidor estequiométrico que fisicamente bloqueia os complexos Clb-Cdc28; portanto, Sic1p deve ser degradado para que haja replicação do DNA. A fosforilação da Sic1p intermediada pelos complexos Cln-Cdc28 levam a ubiquitinação de Sic1p pelo complexo SCF (Complexo Ubiquitina Ligase) seguida por sua degradação (FELDMAN et al., 1997; VERMA et al., 1997).

Antes da ação das Clbs, é necessária a presença de outras proteínas na origem de replicação. A iniciação da replicação pode ser separada em duas etapas: 1- o complexo de reconhecimento da origem (ORC) de replicação recruta as proteínas de manutenção do minicromossomo (MCM), Cdc6p e Tah11p para formar o complexo pré replicativo (pre-

RC); 2- o pre-RC é ativado pelas quinases da fase S e pelo complexo Dbf4-Cdc7. Estes passos permitem o desempacotamento do DNA e o recrutamento das DNA polimerases (NOUGAREDE et al., 2000; SHEU; STILLMAN, 2010; TANAKA; DIFFLEY, 2002). Para garantir que o genoma da célula fique estável durante o ciclo celular, é necessário garantir que ele seja replicado apenas uma vez por ciclo. As ciclinas do tipo B impedem a dupla-replicação do material genético da célula em um único ciclo, além de serem responsáveis por desencadear o desempacotamento das origens de replicação. Essa regulação acontece de três formas: 1- o complexo Clb-Cdc28 fosforila proteínas do ORC; 2- inibe a atividade da proteína Cdc6p por meio de uma ligação física dependente de fosforilação; e 3- causa exclusão nuclear de proteínas do complexo MCM (MIMURA et al., 2004; NGUYEN; CO; LI, 2001).

A replicação do corpo do pólo do fuso (SPB) ocorre também apenas uma vez por ciclo, sendo regulada pelas ciclinas mitóticas (Clb1p, Clb2p, Clb3p e Clb4p). Já a sua ativação pode ser feita pelas Clns, enquanto que as ciclinas Clb5p e Clb6p podem causar a dupla-replicação na ausência de ciclinas mitóticas (BLOOM; CROSS, 2007).

Além dos mecanismos explicados acima, é necessário também a manutenção da coesão entre as cromátides irmãs até a anáfase, fase essa em que as cromátides devem segregarem-se para as células filhas. Essa coesão é feita pelas coesinas (Smc1p, Smc3p e Mcd1p) (MICHAELIS; CIOSK; NASMYTH, 1997; YANG; REN; ZHANG, 2008).

Transição de S para M

Após a duplicação do DNA e SPB a célula pode iniciar a sua entrada na fase M. As ciclinas do tipo B, mais especificamente as Clb1p, Clb2p, Clb3p e Clb4p, são responsáveis pelos eventos mitóticos tais como a morfogênese do fuso e a inibição da citocinese. Portanto, a degradação de tais ciclinas é necessária no final da anáfase para garantir a divisão celular (BLOOM; CROSS, 2007). A ativação de Cdc28p (responsável por causar a

progressão do ciclo celular) nesta etapa requer a sua junção com as Clbs mitóticas e fosforilação por Cak1 (quinase ativadora de CDKs) (KALDIS; SUTTON; SOLOMON, 1996).

A compreensão de como a ciclinas do tipo B são degradadas é fundamental para compreender o funcionamento da fase M e como a célula completa o ciclo. A proteína Cdc14p desfosforila proteínas-chave levando a inativação (degradação) das ciclinas do tipo B, permitindo a desmontagem do fuso e a citocinese. A Cdc14p fica retida no núcleo durante quase todo o ciclo por meio de uma ligação física com a proteína Net1p, e é transportada para o citosol após a ação das redes MEN (rede de saída da mitose) e FEAR (rede de liberação antecipada da anáfase). A FEAR inicia-se na anáfase com a ativação do complexo APC/C-Cdc20. Essa ativação causa a degradação de Pds1p (securina, inibidora de separase), permitindo a ação da Esp1p (separase que inibe as coesinas) levando a segregação dos cromossomos. A degradação da Pds1p também permite que a Slk19p forme um complexo com Esp1p inibindo a fosfatase de Net1p (PP2A-Cdc55). Essa inibição leva a fosforilação da Net1p pela Clb1-Cdc28 ou Clb2-Cdc28. Consequentemente, a fosforilação da Net1p conduz a saída da Cdc14p do núcleo. A rede MEN tem a função de sustentar FEAR e manter a presença de Cdc14p tanto no núcleo como no citoplasma. Esta rede começa pela proteína de ligação GTP Tem1p, a qual ativa a quinase Cdc15p e que ativa a proteína Dbf2p. A ativação da Dbf2p é responsável pela fosforilação e consequente liberação da Cdc14p do núcleo. Após a ação dessas duas vias, Cdc14p pode desfosforilar e ativar a Swi5p, um fator de transcrição de Sic1p, e desfosforilar Sic1p, impedindo que este seja reconhecido por ubiquitina-ligases (JIMÉNEZ et al., 2005; LEE et al., 2001; ROCK; AMON, 2009). A Cdc14p também desfosforila a Cdh1p para ativar o complexo APC-Cdh1, o qual será responsável pela degradação das ciclinas do tipo B (BLOOM; CROSS, 2007). Portanto, as ciclinas mitóticas acabam sendo inibidas por degradação pelo complexo

APC-Cdh1 e por ligação com o inibidor Sic1p, permitindo que a célula finalize o ciclo com a citocinese.

Checkpoints do ciclo

Os *checkpoints* são mecanismo de vigilância do ciclo celular que garantem a interdependência dos eventos desse ciclo. Os *checkpoints* extrínsecos são aqueles ativados apenas quando perturbações acontecem, enquanto que os intrínsecos são aqueles que ocorrem em condições naturais e visam garantir a ocorrência dos eventos do ciclo celular no tempo correto (BARNUM; O'CONNELL, 2014).

- *Checkpoint do tamanho celular*

Para garantir que as células filhas tenham as quantidades apropriadas de material genético e outros compostos, as células mães devem atingir o tamanho ideal antes de se dividirem (em média dobrar o seu conteúdo), portanto, controlar o tamanho celular é crucial para a distribuição de nutrientes na divisão celular. Um dos possíveis mecanismos de controle do tamanho celular baseia-se no monitoramento da tradução de proteínas, devendo haver certas proteínas que possam exercer um controle sobre o ciclo celular após acumularem uma determinada abundância. Propõe-se que a Cln3p seja uma das proteínas responsáveis por esse fenômeno, corroborando com a sua relação entre massa celular e *START* do ciclo. Outro possível mecanismo sugere que as células monitoram a sua geometria durante a entrada na fase M após atingir um determinado valor de comprimento. No entanto, não está claro como estes mecanismos funcionam e nem se ambos ocorrem concomitantemente (BARNUM; O'CONNELL, 2014).

- Checkpoint de dano ao DNA

Caso ocorra algum dano ao DNA, o *checkpoint* desse dano causa um *arrest* da célula em interfase, permitindo que ela tenha tempo para reparar-se; nesse ponto, é essencial manter as CDKs inativas até finalizar o reparo. Esse processo se inicia pelo reconhecimento das estruturas de DNA específicas ou alterações, uma vez que diferentes tipos de lesão geram diferentes sinais de reparo. Em *S. cerevisiae*, a sinalização de dano ao DNA ocorre por meio das proteínas Mec1p, Chk1p, Rad53p, Dun1p, Cdc5p, Rad9p, Rad24p, Rad17p, Mec3p e Ddc1p. A proteína Rad53p possui papel central neste *checkpoint*, a qual é ativada via fosforilação pela Mec1p, Rad9p e Rad24p. Juntas, a Rad53p e Mec1p fosforilam Dun1p, uma quinase que irá estimular a transcrição de uma série de genes de reparo impedindo que o sistema avance para a fase M. Mec1p também fosforila Chk1p, mantendo assim os altos níveis de Pds1p impedindo então a entrada em anáfase. Além de causar um *arrest* em G2/M, este *checkpoint* é capaz de causar um *arrest* em G1 por meio da fosforilação de Swi6. A fosforilação de Swi6 ocorre pela Rad53p, impedindo a formação do complexo Swi6-Swi4, o qual é responsável pela expressão das ciclinas do G1 (DESANY et al., 2006; LONGHESE et al., 1998; NEECKE; LUCCHINI; LONGHESE, 1999; WEINERT; KISER; HARTWELL, 1994).

- Checkpoint da montagem do fuso

A fim de assegurar a estabilidade do genoma e garantir que os cromossomos sejam segregados corretamente, o *checkpoint* da montagem do fuso atrasa a progressão do ciclo celular para que essas condições se concretizem. Uma montagem correta do fuso necessita que os cromossomos liguem-se aos microtúbulos do fuso, caso contrário, o *checkpoint* se ativa e o ciclo é parado antes de prosseguir para a anáfase; a retomada do ciclo ocorre após a formação correta do fuso. As proteínas-chave deste *checkpoint* são a Mad1p, Mad2p, Mad3p, Bub1p, Bub3p e Mps1p. Primeiramente, Mps1p fosforila Mad2p e permite a

formação do complexo Mad2-Cdc20, o qual se liga ao complexo Mad3-Bub3 formando o complexo do checkpoint mitótico (MCC). Este complexo é responsável por impedir que o complexo de promoção da anáfase (APC/C) recrute as ciclinas do tipo B e a securina (Pds1p) (RUDNER; MURRAY, 1996). Sem a ação de APC/C, não há a entrada para a via do FEAR e, conseqüentemente, não ocorre a liberação de Cdc14p e inibição das ciclinas do tipo B.

Outra maneira de impedir a progressão do ciclo na presença de dano ao fuso ou desalinhamento dos mesmos é por meio do complexo Bfa1-Bub2. A presença desse complexo na presença de problemas no fuso mantém a proteína Tem1p inativa impedindo que a via MEN se ative, resultando na manutenção da Cdc14p no núcleo (HU, Fenghua et al., 2001; WANG, Yanchang; HU; ELLEDGE, 2000).

- Resposta ao fator mating

A presença do fator de *mating* (feromônio que induz a expressão de genes responsáveis pelo mating) gera um *arrest* em G1. Para isso, a quinase Fus3p da via de sinalização MAPK fosforila e ativa Far1p, possibilitando a sua associação com os complexos Cln1-Cdc28, Cln2-Cdc28 e Cln3-Cdc28. Perante essas informações, sem a atividade de pelo menos uma destas Clns a célula não progride além do G1. Esse *arrest* cessa ao final do *matting* (JEOUNG; OEHLLEN; CROSS, 1998).

1.4. Modelagem matemática na biologia e modelos do ciclo celular de *S. cerevisiae*

A modelagem matemática de sistemas biológicos é usada para estudar mecanismos observados experimentalmente ou para realizar predições de novas possibilidades pouco evidentes ou difíceis de serem realizadas experimentalmente. Assim, a modelagem é uma ótima ferramenta para testar hipóteses e entender mecanismos complexos (NOBLE, 2003).

Geralmente a modelagem requer grande quantidade de dados experimentais ou, caso o intuito seja explicar um fenômeno pouco conhecido, requer a hipótese de como seria o funcionamento do sistema a ser modelado. A capacidade de predição de um modelo pode ser relacionada com a quantidade de dados disponíveis para sua montagem. Os modelos de sistemas biológicos mais antigos foram baseados em poucos dados, tornando possível a sua utilização para testar hipóteses oriundas de informações experimentais, mas não são úteis na extrapolação além das informações obtidas pelos experimentos. Com os avanços na biologia molecular, a criação de modelos mais detalhados está se tornando uma realidade (TOMLIN; AXELROD, 2007).

O desenvolvimento de um modelo parte do delineamento da estrutura do sistema a ser modelado. A estrutura é o conjunto de interações entre os componentes do modelo que, no caso do ciclo celular, compõem proteínas, suas interações (ativação ou repressão) e seus efeitos no comportamento do ciclo. Há duas maneiras de criar esta estrutura:

1. A abordagem “*Top-down*” parte da estruturação do modelo como uma representação funcional do sistema como um todo. Nesse caso, o nível mais alto representa fenômenos celulares e processos biológicos. Posteriormente, o sistema é dividido em partes menores acrescido de informações mais detalhadas em menores níveis de funcionamento do sistema, tais como as relações entre proteínas e genes e os mecanismos moleculares.
2. A abordagem “*Bottom-up*” parte das interações mais detalhadas do sistema até chegar à visão geral de seu funcionamento. Porém, esse processo requer uma grande quantidade de informação. Além disso, é necessário saber o que acontece com o sistema nos níveis mais altos, uma vez que modelar as interações nos níveis baixos requer conhecimento de suas ações no contexto geral do sistema.

Independente de qual seja a abordagem adotada, os componentes do modelo serão sempre modelados levando em consideração os dados experimentais ou hipóteses a serem

testadas, pois são essas as informações utilizadas para criar regras e modelar sistemas (NOBLE, 2003).

Com a estrutura delineada, o próximo desafio da modelagem é a implementação das regras matemáticas para cada componente do modelo. A abordagem mais utilizada é descrever os componentes utilizando equações diferenciais ordinárias (ODE), as quais representam aumento ou queda nas concentrações de um determinado elemento ao longo do tempo. Estas equações são construídas levando em consideração uma série de parâmetros, tais como taxas de degradação, taxas de reação e limites de ativação (TOMLIN; AXELROD, 2007). Nessa abordagem, ODEs são construídas para cada variável do modelo para descrever a quantidade livre daquela variável a cada instante de tempo por meio das taxas de variação da concentração ($\frac{d[C]}{dt}$). As equações são diferentes para cada tipo de interação de um sistema biológico.

Supondo que duas proteínas hipotéticas “A” e “B” se associam para formar o complexo proteico “AB”, a reação pode ser então descrita seguindo a lei de ação de massas: $A + B \rightarrow AB$. A variação de “A” num instante de tempo pode ser descrita com a equação 1 (eq. 1):

$$\frac{d[A]}{dt} = \alpha \cdot [A] \cdot [B]. \quad (\text{eq. 1})$$

Na situação acima, a [A] está sendo definida apenas pela sua interação com [B], porém, numa situação real em redes regulatórias, sua concentração seria definida por todas as interações que mudam a quantidade de “A” no sistema. Portanto, na equação acima descrita, são inseridos valores relativos à degradação de “A”, à quantidade de “A” utilizada por outras proteínas e a quantidade de “A” que é expressa a partir de uma quantidade “x” de RNA; em suma, processos que aumentam e diminuem a quantidade de “A” são colocados de forma positiva e negativa, respectivamente.

No caso de reações enzimáticas, a equação de Michaelis-Menten pode ser utilizada para modelar interações envolvendo enzimas, substratos e produtos (INGALLS, 2012). Em

uma reação química hipotética na qual a proteína “A” pode ser fosforilada e desfosforilada, a variação da concentração na forma ativa (fosforilada) pode ser dada pela equação 2 (eq 2):

$$\frac{d[Ap]}{dt} = k1 \frac{[X].[A]}{KM+[A]} - k2 \frac{[Y].[Ap]}{KM+[Ap]} \quad (\text{eq. 2})$$

Sendo $k1$ e $k2$ as taxas para fosforilação e desfosforilação, respectivamente, e A e Ap a proteína desfosforilada e fosforilada, respectivamente.

Reações de inibição também podem ser modeladas. Um exemplo é a inibição alostérica na qual inibidores não se ligam diretamente no sítio de ligação, mas afetam a reação ligando-se a outros sítios (INGALLS, 2012). Essa reação pode ser representada pela equação 3 (eq. 3):

$$v = \frac{Vmax}{1 + \frac{[I]}{KI}} \frac{[X]}{Km+[X]} \quad (\text{eq. 3})$$

Sendo v a taxa de produção do produto, X o substrato, I o inibidor da enzima e $Vmax$, KI e Km são as constantes da reação que variam de acordo com os elementos a serem estudados.

Reações ultrasensíveis são reações que respondem com mais sensibilidade a um sinal. Tais modelos podem ser úteis para explorar funcionamentos de fatores de transcrição. Nestes modelos, pequenas alterações numa enzima menos ativa pode causar um *switch* de estado do substrato, principalmente quando as enzimas possuem V_{max} com pequena diferença. Como exemplo, supomos que a maior parte do substrato “X” encontra-se na forma defosforilada e que ambas as enzimas (fosfatase e quinase) estão com seus substratos saturados no meio. Nesse caso, o aumento da atividade da enzima com menor V_{max} (neste caso a quinase) levará ao aumento na fração de substrato “X” fosforilado; no entanto, se este substrato já estiver saturado, não ocorrerá um aumento na taxa de desfosforilação. Em outra situação, se o V_{max} da quinase ultrapassar o da fosfatase a maioria das proteínas se tornarão fosforiladas (INGALLS, 2012).

A equação de Goldbeter-Koshland possui a capacidade de modelar de forma ultrasensível atividades biológicas que necessitam de sinais e que são controladas por reações químicas reversíveis. Supondo uma proteína hipotética “A”, de forma geral, a equação tem a seguinte forma (eq. 4):

$$\frac{[Ap]}{A_t} = \frac{2u_1J_2}{B + \sqrt{B^2 - 4(u_2 - u_1)u_1J_2}} \quad (\text{eq. 4})$$

Sendo $u_1 = k_1 * S$, $u_2 = k_2$, $J_1 = \frac{Km_1}{A_t}$, $J_2 = \frac{Km_2}{A_t}$, $B = u_2 - u_1 + J_1 * u_2 + J_2 * u_1$. Os termos A_p , S e A_t significam, respectivamente, proteína “A” fosforilada, sinal externo que afeta a quinase, e total de proteínas “A” (fosforiladas + desfosforiladas). Os parâmetros indicados por Km são referentes à equação de Michaelis-Menten.

Estas equações são exemplos de como reações biológicas podem ser traduzidas em expressões matemáticas, mas há muitos outros exemplos de equações que podem ser exploradas para atingir o mesmo objetivo. Dois modelos quantitativos do ciclo celular em *S. cerevisiae* desenvolvidos por Chen et al. (2004) e Kraikivski et al. (2015) são ótimos exemplos de modelos complexos que contém uma grande variedade de expressões matemáticas adequadas para diferentes tipos de reações bioquímicas. Embora muitas expressões utilizadas por eles não sejam exatamente iguais aos exemplos citados, os modelos desenvolvidos são muito precisos provando que há diferentes maneiras de se modelar o mesmo tipo de reação.

Apesar da importância do desenvolvimento das equações, a parametrização dessas equações é a parte mais crucial e difícil da modelagem. Encontrar os valores das constantes de degradação, síntese, inibição, equilíbrio, entre outros parâmetros são um desafio. Tais valores dependem de experimentos de cinética enzimática e interação entre proteínas, os quais são extremamente complexos e demandam muito tempo de pesquisa. Dessa forma, quanto mais desconhecido for o fenômeno biológico estudado, mais escassas serão as informações.

A modelagem quantitativa de sistemas biológicos ainda está em um estágio inicial de desenvolvimento. À medida que as técnicas de biologia molecular avançam, mais dados se tornam disponíveis e, portanto, contribuem para desenvolver modelos mais complexos e mais precisos.

A outra abordagem de implementação matemática é utilizando regras Booleanas para gerar um modelo qualitativo, chamados de modelos de estado discreto. Essa abordagem foi adotada no presente trabalho e será introduzida com mais detalhes.

Modelos discretos

É muito comum o uso da abordagem de redes Booleanas, também chamados de modelos discretos ou lógicos. Segundo Le Novere (2015), uma das grandes vantagens dos modelos lógicos é que os componentes da rede podem representar diversas coisas, desde a presença ou atividade de proteínas e genes, até um evento celular mais abrangente.

As equações dos modelos discretos (chamadas de expressões booleanas) que regem o comportamento dos componentes do modelo seguem uma lógica baseada em conjuntos de regras. Resumidamente, pode-se dividir a criação de um modelo lógico em três principais etapas: 1- montagem da estrutura do modelo; 2- criação de expressões para cada componente; 3- solução do modelo e análise dos atratores.

As redes Booleanas são compostas de nós que possuem valores binários (estados) de 0 ou 1 os quais correspondem a FALSO ou VERDADEIRO. Seu o estado é determinado por outros nós vizinhos por meio de expressões Booleanas. Estas expressões são construídas utilizando os operadores Booleanos tais como AND, NOT, OR (inclusivo) e XOR (OR exclusivo). O valor de cada componente na rede varia de acordo com os valores atuais ou anteriores de seus reguladores, dependendo do tipo de método de atualização utilizado (Figura 2). Esses métodos de atualização podem ser: 1- *Synchronous* (determinístico), no qual todas as variáveis são simultaneamente atualizadas partindo do

estado anterior. Nesse caso, presume-se que todos os eventos biológicos da rede acontecem simultaneamente e com transição sincronizada; 2- *Asynchronous* (estocástico), no qual as variáveis são atualizadas numa escala de tempo (estado) individual. Atualizar um modelo muito grande na forma *Asynchronous* é extremamente difícil do ponto de vista computacional, pois consome muito tempo e poder computacional (WANG; SAADATPOUR; ALBERT, 2012); 3- *Priority classes*, método presente no software GINsim, o qual atualiza o sistema de forma *Synchronous*, porém usando um sistema de classes de prioridade. Nesse caso, o sistema atualiza primeiro os nós pertencente à classe mais importante, seguido pela segunda mais importante, e assim por diante.

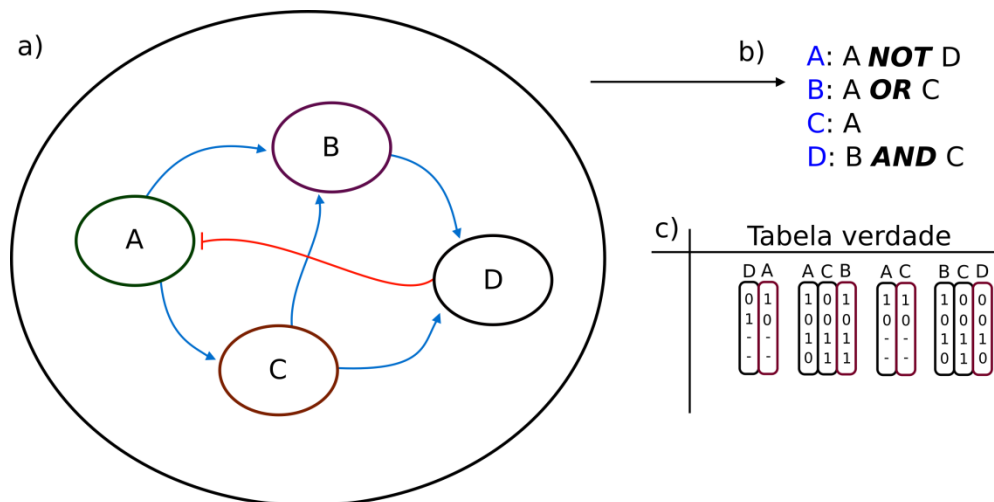


Figura 2: (a) Estrutura do sistema que será modelada pelas (b) equações Booleanas escritas para cada componente da rede. (c) A tabela verdade representa os possíveis outputs (coluna vermelha) a partir de diferentes inputs (coluna preta). O valor 0 representa ausência e 1 presença. As arestas azuis representam uma relação de ativação e as vermelhas representam uma relação de inibição.

A solução de um modelo lógico retorna a variação de cada componente em determinado tempo representado pela transição de estados. O estado compõe valores de todos os componentes da rede e, a partir destes valores, o modelo calcula o próximo estado do sistema. Ao final, o resultado é visualizado em um grafo de transição de estados, o qual pode conter todas as possíveis trajetórias do sistema, ou, a partir de um *input* específico, todas as trajetórias deste *input*. Quanto às informações reveladas pelos *outputs*, o sistema pode estar parado em um estado estável, ou pode permanecer em torno de um número de

estados recorrentes (um ciclo); ambas as situações são chamadas de atratores do sistema (Figura 3) (WANG; SAADATPOUR; ALBERT, 2012).

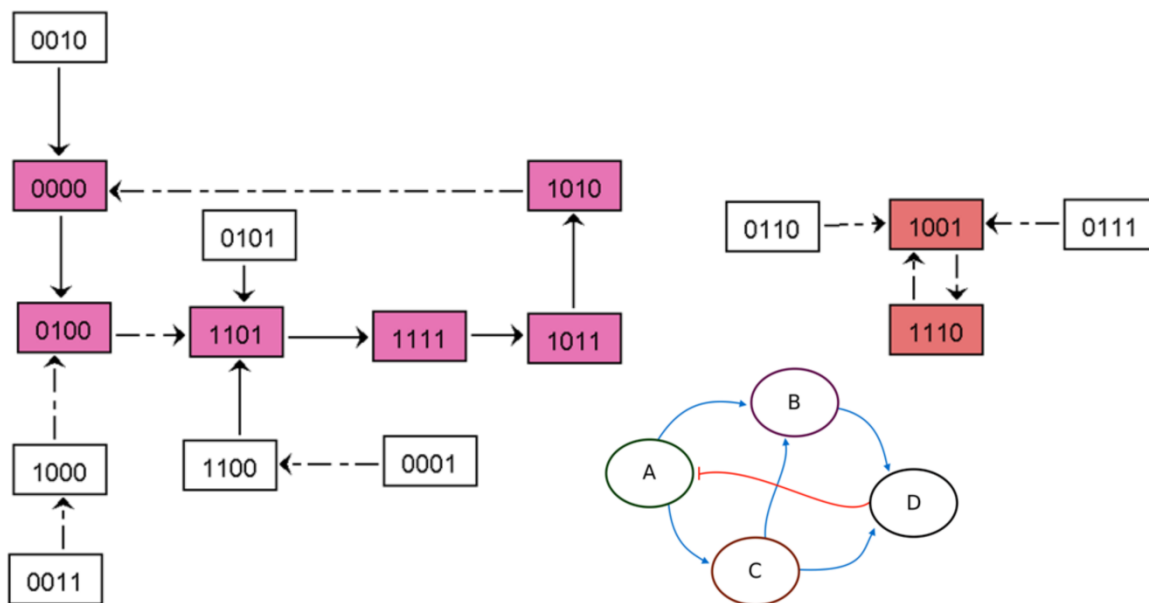


Figura 3: Grafo de transição de estados de todas as trajetórias do sistema. O primeiro número dos retângulos é referente ao componente B do modelo, seguido de A, D e C. Os blocos coloridos representam os possíveis atratores, e neste caso, houve apenas atratores cíclicos. Arestas contínuas significam que apenas uma variável mudou de valor, enquanto que as arestas tracejadas significam que duas ou mais modificações ocorreram. O modelo da Figura 2 foi resolvido utilizando o método de atualização Synchronous no software GINsim.

A atualização dos valores de um nó só acontece quando as equações Booleanas são verdadeiras, ou seja, quando suas exigências são satisfeitas. Utilizando a Figura 2 como exemplo, a equação do nó “D” é “B” AND “C”. Portanto, exigência da equação do nó “D” estará satisfeita somente quando os nós “B” E “C” assumirem o valor = “1”. Como resultado, o valor do nó “D” será atualizado para “1” no próximo estado do sistema.

Analisar os atratores do sistema é o ponto principal da modelagem lógica. Os atratores cíclicos reportam quais são os estados em que o organismo passa mais tempo. Além disso, os atratores estão relacionados a diferentes fenótipos nas redes regulatórias (HOPFENSITZ et al., 2013). Como exemplo, Sahin et al. (2009) utilizou os atratores de seu modelo para procurar novas estratégias para o tratamento do câncer de mama resistente ao trastuzumab, um anticorpo monoclonal utilizado no tratamento desse tipo de câncer,

encontrando uma nova proteína que pode ser utilizada como alvo no tratamento da doença (SAHIN et al., 2009).

As transições de estados levarão a um atrator a partir de um estado inicial. O conjunto de estados iniciais que levam ao mesmo atrator é chamado de bacia de atração e, quanto maior a bacia de atração de um atrator, maior é a sua relevância biológica (HOPFENSITZ et al., 2013). Identificar todos os atratores de um modelo grande é um problema computacionalmente difícil independente do método de atualização utilizado. Uma maneira de contornar este problema é utilizando técnicas de redução do modelo, nas quais variáveis que acabam sempre no mesmo atrator são consideradas irrelevantes para a identificação de todos os atratores do sistema. Nesse caso, essas variáveis são removidas seguidas de uma simplificação das suas equações Booleanas redundantes. No entanto, a utilização de tais métodos podem adicionar algumas oscilações espúrias no novo modelo, mesmo preservando os pontos fixos do sistema (atratores não cíclicos) (WANG; SAADATPOUR; ALBERT, 2012).

Os modelos lógicos de multi-valores abrangem um conjunto de valores além dos Booleanos, assumindo quantos valores discretos forem necessários. Como exemplo, cada componente do sistema (nesse caso, proteínas) pode possuir valor 0 (ausência), 1 (presença), 2 (quantidade abaixo do normal) ou 3 (grande quantidade) (LE NOVERE, 2015). Isso permite uma maior liberdade para o modelo. De fato, essa modelagem vem sendo muito utilizadas para o estudo do ciclo celular em *S. cerevisiae*, como será relatado.

Modelo Booleano de Li

LI et al. (2004) criaram um modelo dinâmico compreendendo todas as fases do ciclo celular de *S. cerevisiae* contendo apenas 11 nós. Os autores separaram os componentes da rede em quatro classes: as ciclínas (Cln1/2, Cln3, Clb1/2, Clb5/6), os inibidores (Sic1, Cdh1, Cdc20, Cdc14), os fatores de transcrição (SBF, MBF, Mcm1, Swi5) e o *checkpoint*

do tamanho celular. Os componentes do modelo podem assumir os valores 0 (ausência) e 1 (presença). Os nós que não possuem um regulador negativo receberam um efeito de auto-degradação, isto é, eles autorregulam-se negativamente (Figura 4).

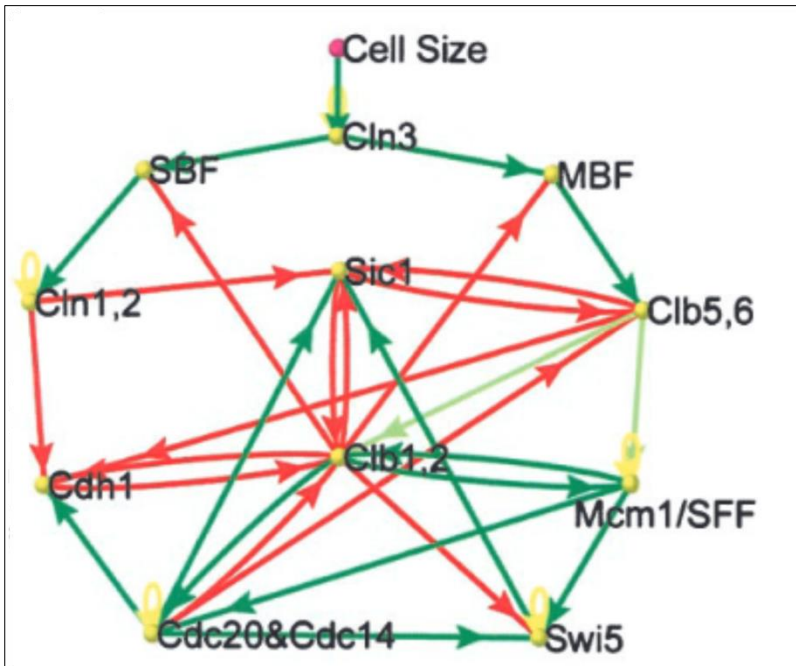


Figura 4: Modelo desenvolvido por Li et al. (2004). Setas em vermelho representam inibição e em verde ativação. As setas em amarelo representam a auto-degradação, portanto, também representam inibição (LI et al., 2004).

Os autores identificaram todos os atratores do sistema partindo de todos os estados iniciais possíveis (2.048). Foi descoberto que, eventualmente, todos os estados iniciais convergiam para um dos sete atratores do sistema, sendo todos então pontos fixos (estados estacionários). A partir desta análise, identificaram que cerca de 86% de todos os estados iniciais convergem para um único atrator, o qual representava a fase G1 do ciclo celular. A explicação dada pelos autores foi que a fase G1 é aquela que garante maior estabilidade para a célula por ficar "estacionada" nesta etapa esperando pela próxima etapa da divisão celular. A fim de testar se há um "caminho" biológico no qual o sistema segue, nesse caso o "caminho" é passar por todas as fases e retornar a primeira (passar por G1, S, G2, M e voltar a G1), iniciou-se o modelo a partir do ponto fixo referente à G1 acrescido de um estímulo do *checkpoint* do tamanho celular. Com isso, foi possível observar que o sistema

passa por todas as fases e retorna ao G1. Além disso, foram testadas perturbações no sistema alterando ou deletando as ligações entre os nós, permitindo constatar que apenas uma pequena fração das perturbações é capaz de eliminar totalmente o ponto fixo (fase G1) (LI et al., 2004).

Os autores concluíram que o ciclo celular em *S. cerevisiae* é robusto, pois grande parte dos estados iniciais foi capaz de permanecer no ponto fixo em G1 e houve resistência às perturbações no sistema. Além disso, concluíram que o sistema é capaz de percorrer o caminho biológico do ciclo celular (LI et al., 2004).

Modelo lógico de Fauré

Neste trabalho (FAURÉ et al., 2009) os autores criaram um modelo lógico de multi-valores do ciclo celular em *S. cerevisiae* contendo um total de 32 nós (Figura 5). Este modelo possui a presença de “nós fenomenológicos”, os quais os autores clamam representar a massa da célula, a citocinese, a formação do broto, a ativação da origem de replicação e a formação do fuso. Os valores de todos os componentes do sistema podem variar numa escala crescente (ausência do componente até altos níveis) atribuídos nos valores 0, 1, 2 ou 3. Além disso, foi incorporado na modelagem as vias FEAR e MEN, acrescentando mais complexidade ao sistema.

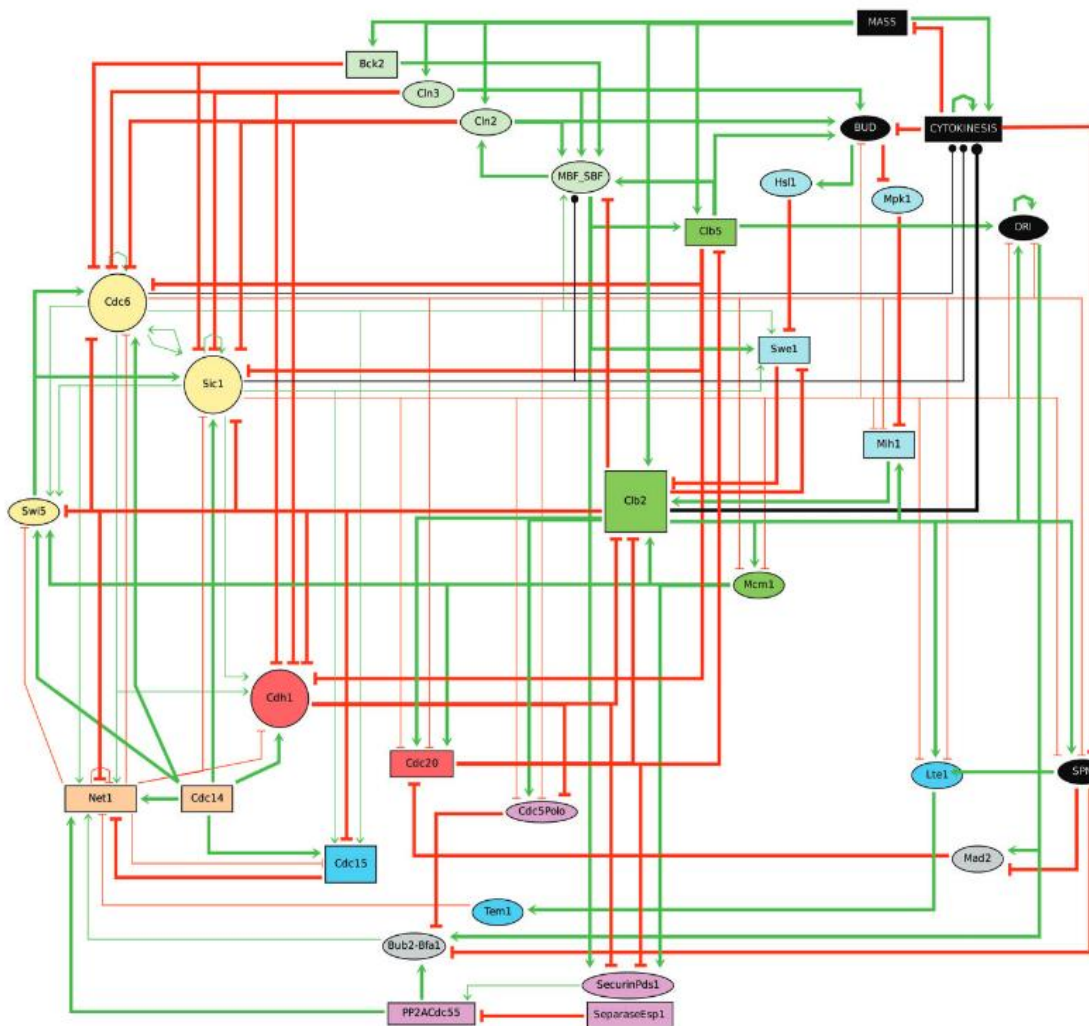


Figura 5: Modelo desenvolvido por Fauré et al. (2009). Arestas verdes significam interações de ativação e as vermelhas interações de inibição. As arestas pretas significam interações desconhecidas, podendo então ser ativação ou inibição de acordo com a equação Booleana do nó regulado.

Para a resolução do modelo, Fauré et al. (2009) escolheram a abordagem *Priority classes*. Os autores criaram quatro classes de prioridades: 1ª prioridade, classe contendo todos os componentes que regulam positivamente o nó “CYTOKINESIS”; 2ª prioridade, classe contendo todos os nós não abordados nas classes de 1ª, 3ª e 4ª prioridade; 3ª prioridade, classe contendo todos os nós que regulam negativamente o “CYTOKINESIS”; 4ª prioridade, classe contendo os componentes que controlam positivamente o nó “MASS”. Dessa forma, indica-se ao modelo que ele deve atualizar o sistema favorecendo a citocinese, seguido do favorecimento da a inibição da citocinese e do estímulo para a massa, resultando no início de um novo ciclo. Os nós restantes como segunda classe de

prioridade permitem que o sistema faça essa progressão entre os nós fenomenológicos (FAURÉ et al., 2009).

Similar ao trabalho de Li et al. (2004), a resolução do modelo indicou que a abordagem lógica da rede regulatória do ciclo celular é suficiente para explicar a maioria das suas propriedades dinâmicas, além de evidenciar sua robustez (FAURÉ et al., 2009). Similar à maioria dos modelos lógicos do ciclo celular já desenvolvidos, o modelo de Fauré et al. (2009) é explanatório, isto é, foi desenvolvido apenas com o intuito de explorar a dinâmica do ciclo celular sem nenhum uso preditivo de novos fenótipos ou de novos componentes na rede.

Interessantemente, este modelo (FAURÉ et al., 2009) foi um dos primeiros a utilizar a análise de comportamento de mutantes como método de validação. Nesta análise, dados disponíveis na literatura sobre fenótipos e comportamentos de leveduras contendo mutações foram comparados com as mesmas mutações aplicadas no modelo, a fim de verificar se ele consegue reproduzir corretamente o fenótipo ou o comportamento observado experimentalmente. Embora o autor tenha realizado mais de 120 testes de mutações com dados experimentais, não há informação quais testes foram corretamente reproduzidos pelo modelo.

Modelo Booleano de Irons

A proposta deste modelo (IRONS, 2009) era melhorar o modelo de Li et al. (2004) adicionando as vias MEN e FEAR, contribuindo para um aumento de complexidade na fase final no ciclo celular. Outro ponto importante deste trabalho foi a utilização de *delays* na atualização de alguns nós específicos para simular velocidade de degradação e ativação, pois, normalmente, isto ocorre em apenas uma transição de estados no modelo Booleano. Com esses *delays*, foi possível especificar quantos estados a mais um determinado nó precisa esperar para que seja ativado ou degradado. Tomando como exemplo a Figura 2,

para a ativação do nó “D” tem-se apenas um estado depois que suas condições de ativação (presença de “B” e “C”) foram alcançadas. Caso fosse adicionado um *delay* de 3 estados, a ativação de “D” só ocorreria depois de quatro estados, ou seja, mesmo se as exigências da equação Booleana fossem cumpridas, o sistema espera o *delay* para atualizar o nó. O autor cria o *delay* adicionando nós falsos que não representam elementos do ciclo celular: eles estão na rede apenas para aumentar o número de estados até a ocorrência de um determinado evento (Figura 6) (IRONS, 2009).

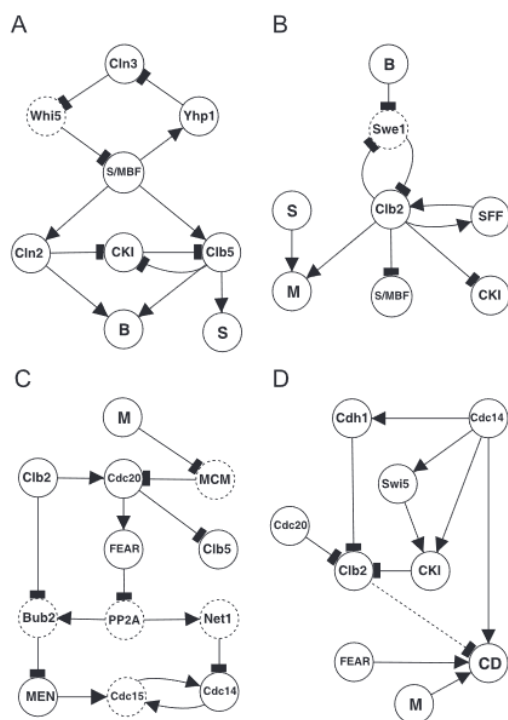


Figura 6: Modelo de Irons (2009). Em A, B, C e D estão representados, respectivamente, a transição de G1 para S, a entrada na fase M, a transição do começo de M até a execução da anáfase, e a saída da mitose (IRONS, 2009).

Para validar o modelo demonstrando que ele é capaz de simular corretamente o ciclo celular, 15 fenótipos presentes na literatura foram testados. Estes fenótipos são os efeitos das mutações em genes do ciclo celular (por exemplo, deleção que acarreta na completa ausência da proteína, ou superexpressão, a qual acarreta no aumento dos níveis de determinada proteína). Como exemplo, sabe-se que a deleção de Swi4 e Swi6 causou um *arrest* em G1, enquanto que a deleção de Esp1 causou um *arrest* em M. Portanto, para o

modelo ser válido, ele deve ser capaz de representar o comportamento correto do maior número possível de mutações presentes na literatura. O modelo de Irons (2009) foi capaz de representar o comportamento dos 15 fenótipos testados (IRONS, 2009), porém, é um número muito baixo quando comparado a quantidade de mutações reportadas na literatura.

O modelo de Irons (2009) também comprovou que o ciclo celular é robusto. Além disso, constatou que o método de atualização *Synchronous* é suficiente para a resolução do sistema, algo também constatado por Fauré et al. (2009).

Como há poucos nós presentes no modelo, foi possível realizar o estudo das sub-redes do sistema. Constatou-se que o sistema pode ser dividido em sete sub-redes, sendo possível criar também uma rede de apenas quatro nós capaz de reproduzir o ciclo celular inteiro (IRONS, 2009). No entanto, não foram discutidos os papéis dessas sub-redes e nem o significado delas existirem.

A introdução dos *delays* foi uma saída para mimetizar o comportamento de um sistema dinâmico, algo que é muito complexo de ser feito com apenas regras booleanas. No entanto, a melhor opção para contornar esse problema seria a utilização de ODEs (NOBLE, 2003).

Modelo estocástico de Todd e Helikar

Os estados irredutíveis são conjuntos ergódicos (conjuntos irredutíveis de um sistema dinâmico) que compõem o grafo de transições de estados de um sistema dinâmico ergódico. Tomando como base a estrutura do ciclo celular desenvolvida por Irons (2009), o modelo de TODD e HELIKAR (2012) parte da ideia de que conjuntos de estados irredutíveis de uma cadeia de Markov em uma rede Booleana probabilística representam fenótipos celulares. Isso porque um sistema dinâmico ergódico é caracterizado por ter o comportamento imutável ao longo do tempo, além disso, independente do *input* inicial, o sistema deve sempre convergir para um atrator específico (Figura 7) (TODD; HELIKAR,

Assim como os outros modelos publicados, Todd e Helikar (2012) demonstraram mais uma vez a robustez do ciclo celular. Os autores reportam que dividir o ciclo celular em conjuntos ergódicos pode modelar fenótipos celulares, pois os perfis de ativação das proteínas do modelo (fases nas quais há aumento ou diminuição nas suas ativações) são similares aos resultados presentes na literatura. A adição de probabilidade ao tamanho celular permitiu que o modelo levemente evitasse a escala discreta dos modelos Booleanos. Nesse caso, o sistema é afetado por incertezas e aleatoriedade, que são variáveis contínuas. Dessa forma, proporciona-se um pouco mais de liberdade ao sistema (TODD; HELIKAR, 2012).

Porém, este estudo permitiu apenas entender o ciclo celular e criar um modelo capaz de representá-lo. Não foram realizados muitos testes de validação do modelo tampouco previsões de fenótipos ou qualquer outro uso prático.

Modelo estocástico de Alcasabas

Tomando como base a estrutura e as equações lógicas desenvolvidas por Fauré *et al.* (2009), Alcasabas *et al.* (2013) adicionaram mais outros 8 nós no sistema e utilizaram a probabilidade para simular diferentes dosagens gênicas. Para simular uma dosagem de 50%, referente a uma deleção heterozigótica, a ativação do nó ocorre com uma probabilidade de 50% e, da mesma forma, para simular dosagens de 0%, 25% e 75%, a probabilidade dos nós voltarem a 0 é de 100%, 75% e 25%, respectivamente. Diferentemente dos outros modelos estocásticos, este modelo não utilizou cadeias de Markov, tornando-o mais simples. O modelo também foi construído para modelar a taxa de crescimento e, para isso, mais de 10.000 simulações foram realizadas para cada mutante ou *wild-type*, sendo computada a quantidade de ciclos completos. Além disso, este foi o primeiro trabalho a construir leveduras mutantes para testar os resultados obtidos da modelagem (ALCASABAS *et al.*, 2013) (Figura 8).

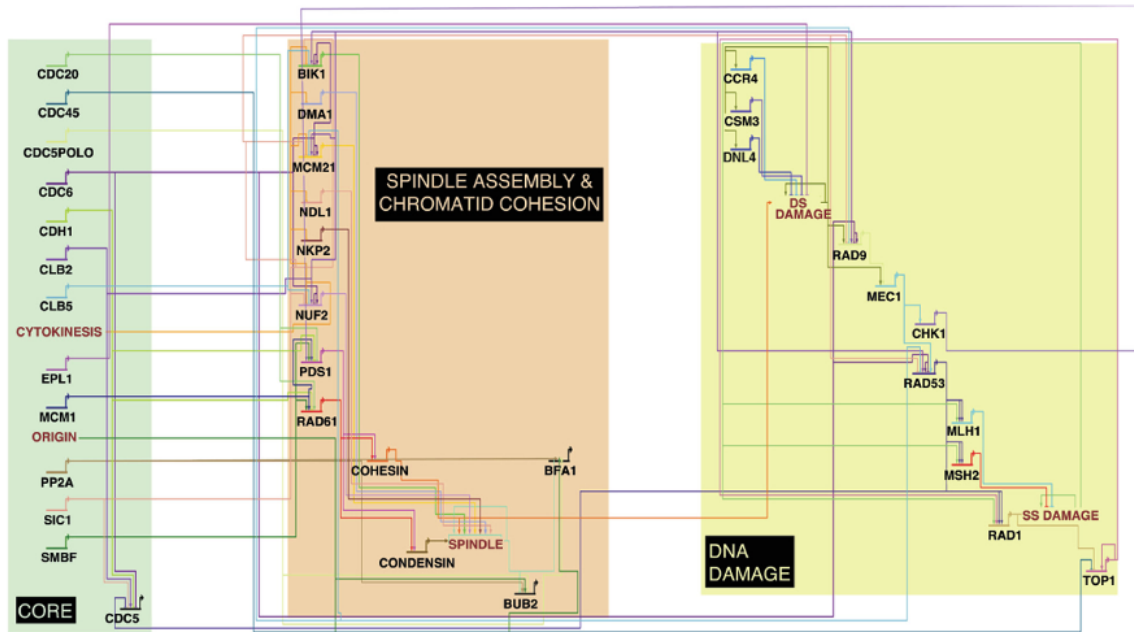


Figura 8: Estrutura do modelo criado por Alcasabas et al. (2013).

Os autores inferiram que os genes SWE1, CLB1, CLB2 e HSL1 exerciam um controle de fluxo de alto nível, ou seja, exerciam grande influência na taxa de crescimento uma vez que sua deleção causava grande impacto na contagem final do número de ciclos. Dos 8 mutantes gerados experimentalmente em uma célula tetraploide (deleção independente dos genes HOG1, MIH1, SLT2, CLB1, CLB2, HSL1, SWE1 e CDC28), o modelo conseguiu prever corretamente 5 fenótipos que, neste caso, eram referentes à variação da taxa de crescimento para cada número de cópias presentes de determinado gene (ALCASABAS et al., 2013).

Com estes resultados, Alcasabas et al. (2013) acrescentaram complexidade ao modelo já desenvolvido por Fauré et al. (2009) e adicionaram módulos de *checkpoint* do fuso e do dano ao DNA, os quais podem ser incorporados nas simulações quando necessário. Ademais, os autores exploraram as previsões errôneas do modelo para tentar inferir funções específicas de tetraplóides e explicar mecanismos pouco conhecidos. A partir dos resultados obtidos, os autores concluem que os modelos lógicos do ciclo celular têm a

capacidade de realizar novas predições e podem ser utilizados para explicar fenômenos pouco conhecidos.

Modelo discreto de Rubinstein

De todos os modelos já apresentados, o modelo de Rubinstein et al. (2013) é o que possui o maior número de nós, totalizando 67. No entanto, isto não significa que houve uma cobertura maior do ciclo celular, pois, neste modelo, um mesmo componente pode possuir mais de um nó relacionado à ele. Como exemplo, a proteína Cln3p possui um nó para seu gene (CLN3), outro para a proteína (Cln3p) e outro para a proteína ligada à quinase Cdc28 (Cln3/Cdk). Esse detalhe é a inovação deste modelo, em virtude de nenhum outro ter ainda estruturado o ciclo celular desta maneira.

O modelo de Rubinstein et al. (2013) também difere na maneira como atualiza os nós, o qual pode assumir um valor discreto de 0 a 9. O método de atualização funciona da seguinte maneira: 1- cada nó sofre um efeito de seu(s) vizinho(s) e, tal como todos modelos lógicos, contem interações de inibição ou ativação; 2- a soma destes efeitos é calculada; 3- caso essa soma seja maior que um limite (*threshold*), o estado do nó regulado irá aumentar na etapa seguinte e diminuirá caso seja menor que o *threshold*. Além disso, os efeitos dos nós vizinhos podem ter intensidades diferentes, por exemplo, um nó ativador pode intensificar o efeito da ativação além de "1" (RUBINSTEIN et al., 2013) (Figura 9).

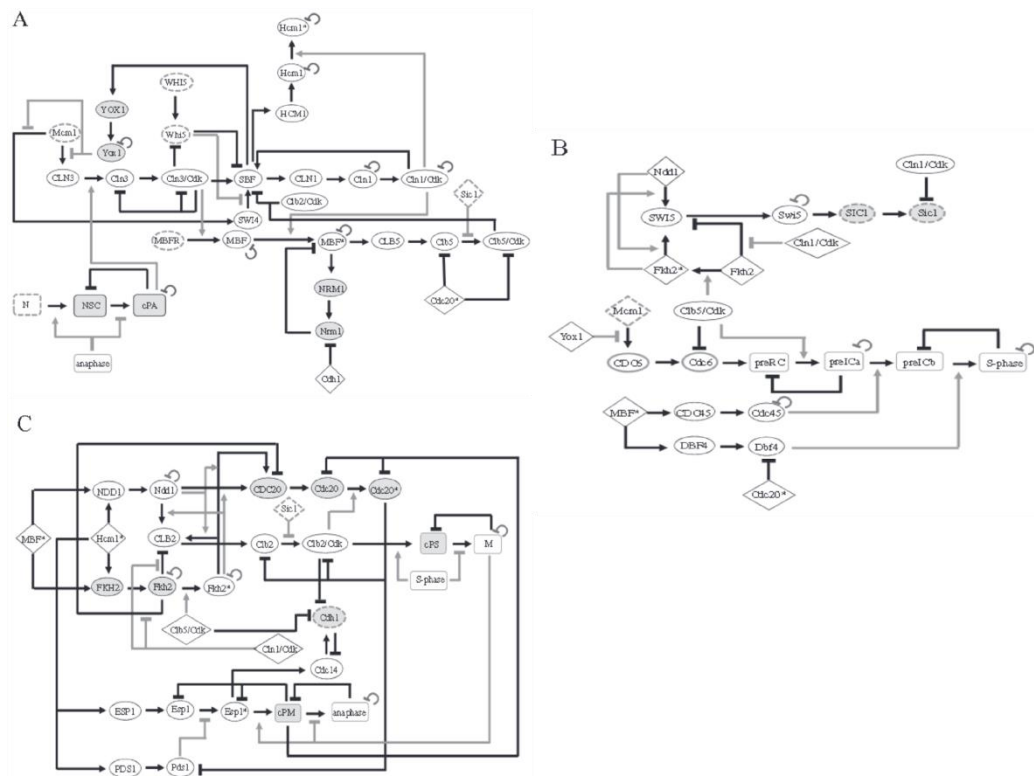


Figura 9: Modelo desenvolvido por Rubinstein et al. (2013). “A” é a parte do modelo referente a G1 e G1/S; “B” representa a fase S; “C” representa as fases entre G2 à anáfase.

Não foram realizados muitos testes para a validação do modelo de Rubinstein et al. (2013). Os autores constatam que a ativação periódica e sequencial de 60 nós do modelo é suficiente para demonstrar o comportamento da rede do ciclo celular. Três mutações *in silico* foram testadas (*knockout* dos genes CDC6, CLB2 e CDC20) e os resultados apresentaram o mesmo comportamento dos dados experimentais publicados. O modelo foi utilizado para examinar três hipóteses quanto a regulação do gene HCM1, possibilitando prever que o complexo Cln1/Cdk era responsável por mediar sua regulação ao invés de Cln3/Cdk ou Clb5/Cdk.

O modelo também foi utilizado para estudar os efeitos do feromônio, o qual serviu para reforçar o que já está presente na literatura, indicando um *arrest* em G1 (JEOUNG; OEHLLEN; CROSS, 1998). Além disso, simulações feitas quanto à depleção de nitrogênio serviram para prever alvos deste sinal, indicando que a disponibilidade da Cln3p e da Cln1p foram afetadas pela depleção, algo que ainda não era claro na literatura. O modelo indicou

também que, possivelmente, a Sic1p não é um alvo essencial da depleção, como era previamente reportado. Portanto, os autores afirmam que o modelo é capaz de realizar predições e testar hipóteses de maneira simples e intuitiva, abrindo portas para predições testáveis experimentalmente (RUBINSTEIN et al., 2013).

1.5. LncRNAs e suas funções

Com os recentes avanços em tecnologias de sequenciamento em *high-throughput*, novas informações foram reveladas a respeito do genoma de eucariotos. Atualmente sabe-se que cerca de 75% do genoma das leveduras *S. cerevisiae* e *Schizosaccharomyces pombe* é transcrito. No entanto, uma grande quantidade destes transcritos não é capaz de codificar uma proteína, sendo então chamados de RNAs não-codificantes. Aqueles que possuem 200 nucleotídeos ou mais são chamados de RNAs longos não-codificantes (lncRNAs) (YAMASHITA; SHICHINO; YAMAMOTO, 2016). De acordo com a versão mais recente do GENCODE (versão 31), o genoma humano contém 48.227 transcritos de lncRNAs, já o genoma de camundongos transcreve 19.112 lncRNAs.

Os lncRNAs são comumente classificados em quatro categorias baseados em sua localização no genoma. No entanto, a classificação é apenas uma forma conveniente de organizá-los, uma vez que não reflete função ou origem evolutiva. Segundo KUNG et al. (2013), as categorias são:

1. *Stand-alone lncRNAs*: engloba lncRNAs que estão presentes em regiões que não sobrepõem genes codificadores de proteínas, comumente referidos como “lincRNAs” (*large intergenic long non-coding RNA*). A maioria deles são transcritos pela RNA polimerase II, são poliadenilados e sofrem *splicing*, possuindo um tamanho médio de 1 kb;
2. *Natural antisense transcripts*: perfaz lncRNA em que a transcrição ocorre de forma mais abundante na fita *antisense*, oposta a fita *sense* de unidades de transcrição

anotadas. Eles tendem a estar presentes próximos dos finais do promotor 5' e do terminador 3' de transcritos *sense*;

3. *Pseudogenes*: os quais são formados a partir de genes que perderam sua capacidade de codificar proteínas devido à mutações.
4. *Long intronic ncRNAs*: enquadram lncRNAs que estão codificados dentro dos introns de genes anotados. Muitos deles apresentam padrões de expressão diferencial e são capazes de responder a estímulos. No entanto, há poucos estudos detalhados quanto a este tipo de lncRNA.

Quanto às suas funções, os lncRNAs participam no desenvolvimento e metabolismo celular por meio de mecanismos como: *imprinting* genético, rearranjo do genoma, modificação da cromatina, regulação do ciclo celular, tradução, *splicing*, transcrição (regulação transcricional e pós-transcricional) e degradação de mRNA (MERCER; DINGER; MATTICK, 2009; ZHU et al., 2013).

Os lncRNAs exercem diversas funções interagindo também com proteínas ou RNAs. Parte deles interage com proteínas próximas ao seu sítio de transcrição regulando a expressão de genes em *cis*; podem servir de *scaffolds*, ligando-se a fatores de transcrição impedindo que se liguem ao DNA; podem também interagir com complexos modificadores de cromatina conduzindo-os aos seus alvos; podem ligar-se a miRNAs servindo de *scaffolds* ou melhorando a atividade de *enhancers*; servem também como base para a formação de novos complexos de macromoléculas (FERRÈ; COLANTONI; HELMER-CITTERICH, 2016).

Existem lncRNAs de *S. cerevisiae* relacionados à regulação da diferenciação sexual por meio do silenciamento do gene *IME1*, mediado pelo lncRNA (*IRT1*), e pela supressão de *IME4*, mediada pela transcrição *antisense* de outro lncRNA (*RME2*) (YAMASHITA; SHICHINO; YAMAMOTO, 2016). Além disso, foram descobertos lncRNAs envolvidos

em mecanismos de resposta a estresse também em leveduras. Como exemplo, Nadal-Ribelles et al. (2014) descobriram que a Hog1p, uma quinase ativada por estresse (SAPK) e essencial para controlar o estresse osmótico através da reprogramação de expressão gênica, é capaz de se associar e induzir um conjunto de lncRNAs em resposta ao estresse osmótico. Um destes lncRNAs (“CDC28 lncRNA”) é expresso na orientação *antisense* do gene de uma proteína quinase CDC28, quinase-chave no ciclo celular. A indução da expressão desse lncRNA aumenta os níveis de Cdc28p permitindo que a célula inicie o ciclo celular mais rapidamente após o estresse (NADAL-RIBELLES et al., 2012; SOLÉ et al., 2015).

Como pode ser observado, os lncRNAs participam de processos celulares importantes, quase sempre exercendo uma função regulatória. As interações entre lncRNAs e proteínas ainda estão sendo exploradas, contendo poucos estudos na literatura sobre o assunto.

2. Justificativa e objetivos

A partir dos modelos reportados, pode-se observar que as capacidades preditivas desses modelos foram pouco exploradas. Com exceção do trabalho de Fauré et al. (2009), pouco esforço foi feito também para validar os modelos utilizando dados experimentais. Embora Rubinstein et al. (2013) tenham criado um modelo relativamente grande, não há relatos na literatura de um modelo completo do ciclo celular de *S. cerevisiae*. Como mencionado anteriormente, Marques (2019) encontrou quatro lncRNAs na via do ciclo celular em linhagens de *S. cerevisiae*. Portanto, é preciso entender como eles agem no sistema a fim de servirem como subsídios para compreender a tolerância ao etanol em *S. cerevisiae*.

Ademais, duas perguntas precisam ser respondidas: 1- o que causa a diferença de crescimento entre HTs e LTs observados por Almeida (2017)?; 2- os lncRNAs possuem algum papel neste sistema? Para respondê-las é necessário ter uma visão sistêmica da via

em virtude de vários genes sofrerem com o efeito do etanol. Além disso, criar hipóteses acerca das funções de lncRNAs, que são elementos pouco conhecidos, ainda é algo extremamente difícil se forem analisados isoladamente. Logo, a modelagem desta via tem como objetivo permitir uma visão sistêmica de forma mais compreensível, algo que seria muito difícil de ser feito laboratorialmente.

2.1. Objetivos gerais

- Criar um modelo que simule o comportamento cíclico do ciclo celular.
- Validar o modelo utilizando dados experimentais reportados na literatura;
- Avaliar o impacto dos lncRNAs e das expressões dos genes em células sob estresse por etanol sobre o ciclo celular.

2.2. Objetivos específicos

- Elaborar um modelo lógico de multi-valores do ciclo celular de *S. cerevisiae* contemplando o maior número possível de nós presentes na estrutura do KEGG (Figura 1);
- Testar dados experimentais de transcriptoma utilizando informações de expressão diferencial para explicar o comportamento de linhagens de leveduras altamente e pouco tolerantes ao etanol;
- Simular a função dos lncRNAs supostamente relacionados ao ciclo celular encontrados nas linhagens SEY6210 e BY4742. Nesse caso, testaram-se diferentes hipóteses para inferir uma possível função para esses RNAs.

3. Métodos

3.1. Definições

Para organizar visualmente os componentes do modelo desenvolvido na dissertação (Figura 10), os componentes foram agrupados da seguinte forma:

1- os círculos pretos representam os fenômenos celulares (nós fenomenológicos) abrangendo a formação do broto, duplicação do DNA, formação do fuso, citocinese e aumento de massa;

2- retângulos vermelhos são fatores de transcrição, sendo que o nó “SBF” é um complexo (Swi4p e Swi6p) tal como o “MBF” (complexo entre Mpb1p e Swi6p);

3- as elipses azuis representam elementos promotores (“ECB”, “MCB” e “SCB”), a origem de replicação (“ARS”) e a expressão de proteínas na fase S e G2 (“S_proteins” e “G2_proteins”, respectivamente);

4- os retângulos roxos representam complexos protéicos entre duas proteínas, ex. “Dbf4_Cdc7”, o qual indica uma união de Dbf4p e Cdc7p, e também as ciclinas, as quais formam um complexo com Cdc28p;

5- os círculos verdes representam complexos protéicos com mais de uma proteína, tal como “Cohesin” (união entre Smc1p, Smc3p, Mcd1p e Irr1p);

6- retângulos amarelos representam proteínas de *checkpoint*;

7- círculos vermelhos são complexos protéicos de mais de uma proteína, porém formados pela ação de outros nós da rede;

8- retângulos brancos são proteínas que não se enquadram nas outras classificações;

9- elipses verdes representam os *checkpoints* do ciclo, os quais não sofrem regulação por serem tratados como *inputs*.

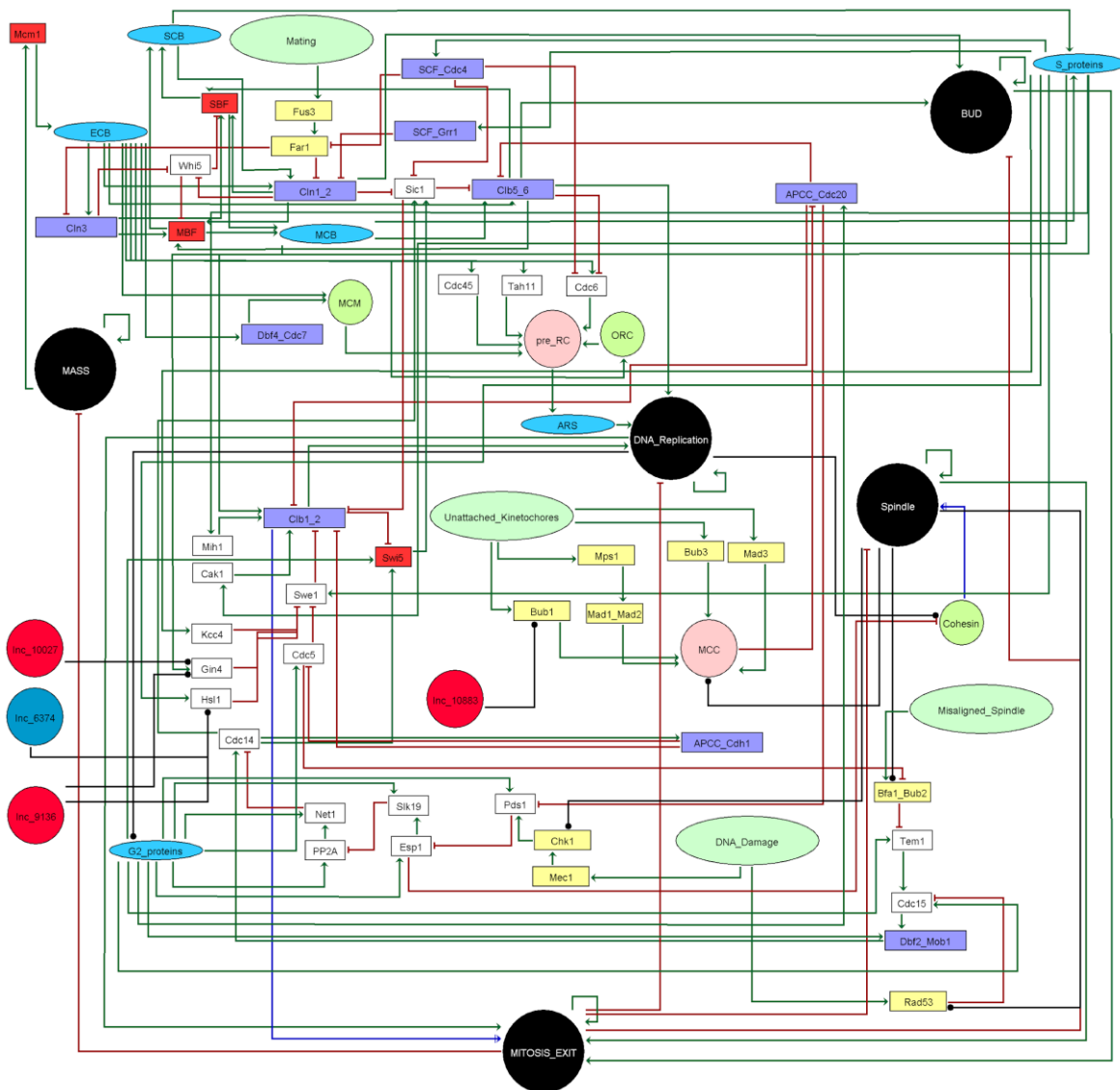


Figura 10: Modelo do ciclo celular desenvolvido nesta dissertação. Arestas verdes indicam interações de ativação, vermelhas, inibição; as arestas azuis são relações duais, indicam que há ativação, mas também inibição; as arestas pretas indicam que as interações ali são desconhecidas. Os nós estão organizados de acordo com o descrito na seção acima (3.1. Definições).

3.2. Desenvolvimento do modelo lógico

Quanto à estrutura do ciclo celular, foi utilizado como base o modelo do KEGG (Figura 1) com o objetivo de contemplar o máximo de nós possíveis. No entanto, a estrutura do KEGG possui algumas limitações: 1- não deixa claro quem são os responsáveis pelos fenômenos tal como a formação do broto, a replicação do DNA e a formação do fuso; 2- não estão presentes algumas interações entre proteínas já descritas na literatura; 3- não há conexões entre proteínas da fase G1/S e G2/M, o que dificulta a modelagem do sistema

como um todo. Portanto, todas as proteínas e suas ligações presentes no KEGG foram incorporadas à estrutura do modelo lógico, que posteriormente foi acrescida de detalhes com dados reportados na literatura.

Embora todo o funcionamento do ciclo celular tenha sido descrito nos tópicos anteriores, é necessário justificar como foi construída a estrutura do modelo a fim de representar o comportamento do ciclo celular. Essa justificativa é feita conectando as informações presentes na introdução com a estrutura do modelo desenvolvido. Portanto, algumas informações serão redundantes, mas necessárias para compreender o porquê de um nó se conectar a outro da maneira ocorre no modelo.

START e transição G1/S

A massa possui uma grande influência no *START*, sendo responsável pela ativação de pelo menos uma Cln. Aqui, estabeleceu-se que quando a massa atingir seu valor máximo, ativará a transcrição dos genes responsáveis pelo G1/S (nesse caso a Mcm1p, tal como reportado por CHANG et al. 2017), culminando na ativação do nó “ECB”. “ECB” levará a ativação dos nós “Cln3” e “Cln1_2”, os quais induzirão a expressão dos fatores de transcrição do SBF e MBF. Nesse momento, “Cln3” deve reprimir “Whi5” para que o “SBF” seja ativado. “SBF” e “MBF” ativarão “SCB” e “MCB”, responsáveis pela transcrição de genes presentes na fase S. “ECB” também ativa vários nós responsáveis pela fase S (“Dbf4_Cdc7”, “MCM”, “Cdc45”, “Tah11”, “Cdc6” e “ORC”) os quais ativarão o nó “ARS” intermediado pelo “pre_RC”. A escolha de ativação destes nós por “ECB” deve-se ao fato deste estar presente na transição G1/S, como reporta o Cyclebase (SANTOS; WERNERSSON; JENSEN, 2015). Posteriormente, “Clb5_6” e ARS ativam o nó “DNA_Replication”. Na ausência de “Cln1_2”, “Clb5_6” pode ativar o nó “BUD”.

G2, entrada na mitose e citocinese

Com a ativação do nó “DNA_Replication” representando que a duplicação do DNA já ocorreu, o sistema segue para a finalização da formação do fuso e entrar na fase M. Isso começa pela ativação do nó “S_proteins” por “MCB” e “SCB”, culminando na ativação dos complexos ubiquitina ligase (SCF), “Cak1” e “Mih1” (GOH; SURANA, 1999). A ativação do nó “G2_proteins” pelo nó “DNA_Replication”, indica a transcrição de genes e a tradução de proteínas do G2. Após a replicação do DNA, a união das cromátides irmãs é representada pelo nó “Cohesin”. Os elementos que compõem a coesina são expressos em G1 (de acordo com Cyclebase), mas no modelo aqui apresentado, o nó “Cohesin” sofre ativação de “DNA_Replication” pois sua ação inicia após esse fenômeno. Por fim, “Cohesin” ativa o nó “Spindle” para indicar o começo da formação do fuso.

A ativação de “G2_proteins” regula positivamente as proteínas das vias MEN (“Tem1”, “Cdc15” e “Dbf2_Mob1”), FEAR (“Pds1”, “Esp1”, “Slk19”, “PP2A” e “Net1”), “Cdc5” e o complexo “APCC_Cdc20”; essas proteínas são expressas em G2/M de acordo com Cyclebase. As vias FEAR e MEN ativam “Cdc14” através da inibição do nó “Net1” e ativação do nó “Dbf2_Mob1”. “Cdc14” ativa o fator de transcrição “Swi5”, “Sic1” e “APCC_Cdh1”. Já o “Sic1” e “APCC_Cdh1” inativam o “Clb1_2”.

A entrada na mitose exige a chegada ao valor “1” do nó “MITOSIS_EXIT”, o qual é atingido quando todos os outros nós fenomenológicos (exceto o da massa, o qual não regula este nó) e “Clb1_2” estão ativos. A inativação do nó “Clb1_2” permite que o nó “MITOSIS_EXIT” atinja o valor “2”, indicando citocinese e, nesse ponto, todos os outros nós fenomenológicos chegam a 0. Um novo ciclo se inicia quando o nó “MASS” iguala a “0”.

Checkpoints

Diferentes dos outros nós, os *checkpoints* foram adicionados como *inputs*, isto é, eles funcionam com valores fixos sendo 0, 1 ou 2; “0” indica ausência do *checkpoint*, “1” presença em níveis normais e “2” é presença em níveis elevados. Como os *checkpoints* são compostos por pelo menos três nós, seria necessário modificar a expressão de cada um individualmente caso o *checkpoint* não fosse tratado como *input*. Da forma aqui apresentada (*inputs*), é necessária apenas a modificação de um deles para garantir o efeito desejado. Isso permite uma maior liberdade ao modelo simplificando a simulação de eventos como danos ao fuso ou ao DNA.

O *checkpoint* do fuso está dividido em duas partes, uma referente ao sinal dado por cinetócoros soltos e outra pelos fusos desalinhados. A primeira parte ocorre pela ativação de uma série de proteínas que culminam na ativação do complexo “MCC”, inibindo “APCC_Cdc20”. A segunda parte ocorre pela ativação de “Bfa1_Bub2”, afetando o nó “Tem1”. O *checkpoint* do dano ao DNA exige a ativação dos nós “Rad53”, “Mec1” e “Chk1”. Os nós referentes ao *checkpoint* do *mating* possuem o valor máximo de “1” e a sua função (inibição das Clns) depende da ativação de “Fus3” e “Far1”.

Para causar um *arrest* na célula, os *checkpoints* devem estar superexpressos. Em condições normais (com os *checkpoints* inativos ou pouco ativos), o modelo progride normalmente. A progressão na presença de baixa atividade (valor “1”) dos *checkpoints* de dano indicam que a célula foi capaz de reparar o dano que sofreu. Esse reparo é indicado por um *feedback* negativo do nó “Spindle” (interagindo com “Bfa1_Bub2” e “MCC”) (Figura 10), o qual foi representado desta maneira devido à sinalização de reparo ultrapassar os limites do ciclo celular.

A estrutura, as equações Booleanas e a resolução do modelo foram desenvolvidas utilizando o *software* GINsim (versão 3.0.0b) (DE JONG, 2002).

Equações Booleanas

Após a montagem da estrutura, foram construídas as expressões lógicas que modelarão cada elemento da rede. As expressões foram escritas levando em consideração o tipo de interação. A lógica foi desenvolvida da seguinte maneira:

1. Em interações de ativação o nó ativador aumenta o valor do nó regulado com base no seu valor atual. Como exemplo, na ativação de “Clb5_6” via “MCB”, ambos nós possuem os mesmos valores.
2. Em interações de inibição por fosforilação, o valor do nó regulado diminui de acordo com o nó regulador. Por exemplo, na regulação de “Swi5” por “Clb1_2”, se “Swi5” possui valor “2”, porém, em seguida, “Clb1_2” atinge o valor “2”, ocorre a desfosforilação de “Swi5”, retornando seu valor para zero.
3. Interações de ativação por fosforilação são inversas às reações de desfosforilação. Tomando o exemplo acima, “Cdc14” fosforila “Swi5”, portanto, ele reverte o efeito de desfosforilação de “Swi5” via “Clb1_2”.
4. Interações de inativação por degradação funcionam pela diminuição direta do valor do nó regulado, como já explicado em 2. Como exemplo, caso “Clb5_6” esteja com valor “2”, ele retornará a zero quando “APCC_Cdc20” atingir “2”.

É importante ressaltar que algumas interações possuem detalhes adicionais, por exemplo, a regulação de “Clb1_2” por “Sic1”. Nesse caso, a “Sic1” sozinha não é suficiente para degradar completamente a “Clb1_2”, necessitando de pelo menos um nó referente ao complexo APC ativo para ocorrer tal inibição. Esses detalhes estão geralmente presentes em nós regulados por muitos outros nós. O Arquivo Suplementar 1 contém todas as explicações para as expressões em que isso acontece.

Como aqui trabalhamos com um modelo de multi-valores, cada elemento pode receber no máximo quatro valores, sendo 0, 1, 2 ou 3, significando respectivamente ausência, baixas quantidades, quantidades normais e grandes quantidades de um

composto/fenômeno. O valor máximo de alguns elementos foi reduzido e outros foram considerados Booleanos (0 ou 1). Para alguns nós, os valores não significam presença ou ausência de uma proteína ou fenômeno biológico, mas sim um grau de ativação. Por exemplo, em algumas proteínas, o valor “2” significa estar ativa o suficiente para exercer sua função ao invés de quantidades normais dessa proteína. Ressalta-se que a escolha do valor máximo do nó foi feita baseada nas informações disponíveis sobre ele, dessa forma, nós pouco conhecidos foram simplificados assumindo no máximo 3 valores, melhorando o desenvolvimento do modelo.

O modelo foi construído para representar o comportamento cíclico do ciclo celular, ou seja, ele precisa ser capaz de iniciar o ciclo, terminar na citocinese e retornar ao início novamente. Em outras palavras, o comportamento de um *wild-type* em condições normais deve ser representado como um atrator cíclico no grafo de transição de estados. A ciclagem do modelo exige a progressão pelos nós fenomenológicos (círculos pretos da Figura 10). Essa progressão inicia-se a partir da formação do broto (“BUD”) e pela duplicação do material genético (“DNA_Replication”); ambos nós podem assumir os valores “0” ou “1” sendo que “1” indica que estes fenômenos terminaram. Em seguida, o modelo passa pela formação do fuso (“Spindle”), o qual pode assumir os valores “0”, “1” ou “2” indicando, respectivamente, que a formação não ocorreu, iniciou-se, ou foi finalizada. O nó “MASS”, o qual representa a massa, pode assumir os valores “0”, “1” ou “2” indicando, respectivamente, um valor de transição entre o fim da mitose e início de um novo ciclo, pouca massa, ou massa adequada. Quando o nó “MITOSIS_EXIT” atinge “2” (mitose completa), significa que houve a citocinese. Com a finalização da mitose, os nós fenomenológicos retornam a “0”, iniciando novamente o ciclo. Esta progressão entre os nós fenomenológicos causa o comportamento cíclico do modelo.

Após a construção das equações, o modelo foi implementado no *software* GINsim utilizando o método *Synchronous* e tendo como *input* inicial o nó “MASS” como valor “1”,

e o restante dos nós receberam valor “0” (explicado no próximo tópico). Quanto aos *checkpoints*, apenas o “Mating” foi definido como “0” (inativo) enquanto que os outros foram assinados como “1” (ativo, mas incapaz de causar *arrest*). Em seguida, o grafo de transição de estados foi separado, os nós que formavam o atrator foram selecionados e o comportamento do ciclo foi analisado. Essa análise foi feita verificando qual é o atrator do sistema e como ocorreu a transição de estados até alcançá-lo. Dessa forma, foi possível compreender se o sistema está ciclando ou se está em algum *arrest*: caso houvesse *arrest*, extraiu-se a fase do ciclo em que isso ocorria.

Nessa etapa, foi avaliado se o modelo realiza sua função normal na forma mais básica, ou seja, sem nenhuma perturbação (simulando uma levedura *wild-type*). Quando o modelo falhou em representar o ciclo celular, revisões foram feitas e mudanças na estrutura (adição ou remoção de interações) ou nas expressões foram empregadas. A partir do momento em que o modelo cumpriu sua função básica, revisões e testes foram realizados para adicionar detalhes ao sistema, permitindo maior poder de predição do comportamento do ciclo celular na presença de perturbações e condições diferentes. Cada vez que o modelo falhou na validação (falha em acertar o fenótipo oriundo de alguma perturbação específica), revisões adicionais foram feitas para ajustar o modelo em todas as suas etapas. Essas etapas foram realizadas de acordo com Wang et al. (2012) (Figura 11).

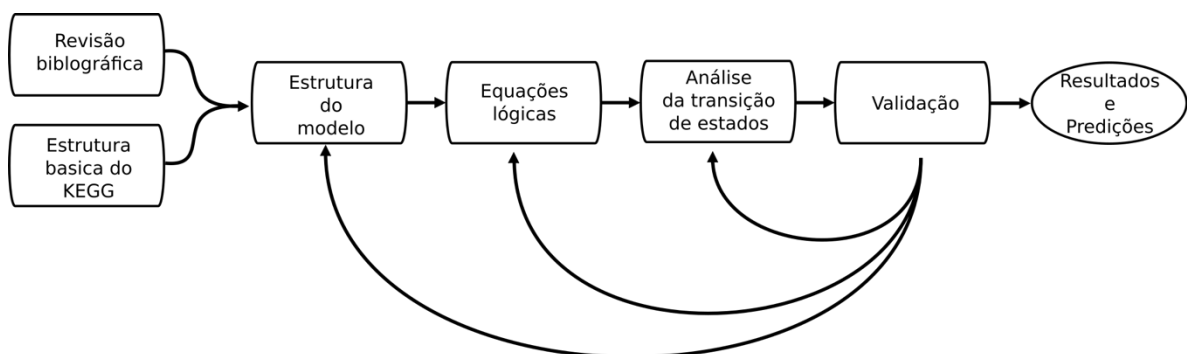


Figura 11: Fluxograma da criação do modelo lógico, adaptado de Wang et al. (2012).

3.3. Validação do modelo

Em virtude do grande tamanho da rede a ser modelada, o método *Synchronous* de atualização foi utilizado tanto para os testes de validação quanto para os testes usando os dados de transcriptoma (posteriormente descritos). Para avaliar se o modelo não está enviesado, vários *inputs* aleatórios foram também testados.

Depois de testar vários *inputs* iniciais, convencionou-se que todos os testes seriam realizados com o sistema partindo do nó MASS com valor “1” e o restante dos nós atribuídos como “0”. Os checkpoints foram tratados como previamente descritos, sendo os mutantes que realmente tinham os *checkpoints* ativos ao máximo as exceções. Esse *input* inicial foi escolhido, pois faz parte da maior bacia de atração do modelo (discutido com detalhes no tópico Discussões). A escolha de usar apenas um *input* na etapa de testes foi baseada na literatura, onde na maioria dos casos as simulações foram feitas partindo de um único *input* inicial.

Para a validação, 109 mutações testadas experimentalmente foram avaliadas pelo modelo buscando checar se o resultado das simulações corresponde aos obtidos experimentalmente: 1- nos testes de *knockout*, o nó relativo ao gene deletado foi eliminado do modelo por meio da atribuição constante do valor “0”; 2- nos testes de superexpressão, a interação entre nó-alvo e seus ativadores era maximizada. Dessa forma, apenas o nó-alvo foi superexpresso enquanto que seus nós-regulados não foram afetados (esse método é mais fidedigno por permitir que o nó sofra inibição de outros elementos da rede); 3- nos testes de sub-expressão, o valor do nó-alvo foi limitado a um alcance de 0 ou 1, mesmo que seu ativador estimule-o ao máximo.

Como o modelo aqui apresentado não distingue as sub-partes dos complexos (por exemplo, “SBF” é uma simplificação do complexo Swi4p e Swi6p), as mutações da literatura que atingiam apenas uma proteína do complexo foram assinadas no modelo como afetando todo o complexo.

Testes que envolviam proteínas termossensíveis foram realizados tal como um *knockout*, pois são proteínas modificadas que, com um aumento de temperatura fora da faixa permitida, perdem sua viabilidade. As proteínas não degradáveis foram simuladas eliminando a interação negativa dos inibidores responsáveis pela degradação, ao mesmo tempo em que permite o nó ser inibido por fosforilação (Arquivo Suplementar 1).

Para simular os efeitos de elevado dano no fuso e danos ao DNA, o *input* para esses nós fenomenológicos foram fixados em seu valor máximo. Para simular os efeitos de *mating*, os inputs do nó "Mating" foram fixados no valor máximo.

Os resultados dos testes das mutações foram aplicados em uma matriz de confusão para avaliar os parâmetros acurácia, especificidade e sensibilidade. Para uma simulação ser pontuada como verdadeiro positivo, ela deveria representar exatamente o que era observado na literatura. Por exemplo, se uma determinada mutação causar *arrest* na fase M, o modelo deve também indicar esse *arrest* nessa fase. Em alguns casos o modelo corretamente previu um *arrest*, mas errou o ponto em que isso acontecia. Embora seja um resultado animador ele não foi computado como acerto.

Algumas mutações presentes na literatura não indicam em qual ponto a levedura cessou o ciclo celular, apenas reportam que o mutante é inviável. No entanto, os resultados das simulações sempre indicam um ponto específico em que a célula parou o ciclo. Como exemplo, a simulação do mutante Cdc15- Δ demonstrou um *arrest* na fase M, enquanto que na literatura indica apenas se tratar de uma mutação inviável. Apesar de "Arrest em M" ser terminologicamente diferente de "Inviável", esses casos foram aqui atribuídos como verdadeiros positivos na matriz de confusão, pois, de fato, o modelo acertou a incompletude do ciclo celular. No modelo quantitativo de Chen et al., (2004), os casos em que isso ocorria também foram considerados como acerto do modelo.

3.4. Testes dos dados de transcriptoma

Após a validação do modelo, os dados de transcriptoma de 6 linhagens de leveduras sob estresse causado por etanol (WOLF, 2019) foram utilizados em simulações a fim de compreender a cinética do ciclo celular frente ao estresse por etanol em dois fenótipos (HT e LT). Como o objetivo é diferenciar o fenótipo HT e LT quanto ao comportamento do ciclo celular, os genes diferencialmente expressos do ciclo (comparando controle vs. tratamento) foram separados. Em seguida, os genes que tinham o mesmo perfil de expressão (sub-expresso ou superexpresso) nas três linhagens de cada grupo (HT ou LT) e aqueles que possuíam o padrão inverso entre os grupos (sub-expresso nas HTs e superexpresso nas LTs ou o oposto) foram selecionados para serem simulados no modelo. Os dados de \log_2 Fold-change foram usados para definir superexpressão ou subexpressão no modelo, porém, por serem contínuos (intervalo de -4 a 4), uma discretização foi realizada da seguinte maneira: 1- genes com valores <-0.5 foram considerados sub-expresso; 2- ≥ -0.5 e ≤ 0.5 foram considerados normais e; 3- >0.5 foram considerados superexpressos. Esses valores foram escolhidos baseados na distribuição dos valores de fold-change (Figura 12). Os dados utilizados nesta etapa estão representados na Tabela 6.

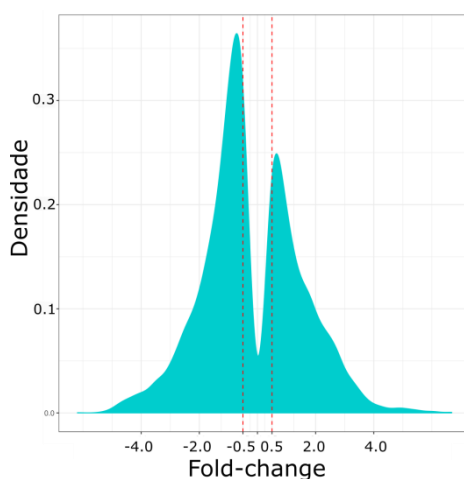


Figura 12: Distribuição do fold-change de todos os transcritos das seis linhagens.

Tabela 6: Dados de expressão não discretizados utilizados no modelo para diferenciar o fenótipo HT de LT. As três primeiras linhagens são HTs, as três últimas são LTs. Os valores são referentes ao Log2 do *fold change* quando comparado o grupo tratamento vs controle.

Genes do ciclo celular	BMA64-1A	BY4742	X2180-1A	SEY6210	BY4741	S288C
YLR079W	-	-	-	-1,08	-1,22	-1,05
YDR328C	0,66	0,52	0,89	-	-	-
YDR113C	-1,26	-0,87	-0,39	-	-	-
YFL029C	0,65	1,39	0,44	2,13	1,58	1,73
YBL023C	-0,36	-1,29	-1,08	-1,19	-2,69	-0,83
YKL101W	-1,31	-1,22	-0,75	-1,09	-3,12	-0,98
YMR199W	-1,51	-3,18	-2,01	-2,51	-0,74	-2,16
YDL003W	-0,78	-2,51	-1,93	-2,24	-2,11	-2,02
YDL155W	-0,73	-1,38	-0,91	-1,01	-0,38	-2,03
YDR507C	-0,95	-1,77	-1,45	-1,68	-1,53	-1,83
YFR028C	-0,59	-2,02	-1,42	-2,41	-1,57	-1,69
YGR188C	-0,83	-1,40	-0,86	-0,94	-2,34	-0,94
YLR210W	-0,97	-2,02	-1,93	-1,47	-3,99	-1,12
YML027W	-1,59	-2,30	-1,93	-2,87	-0,97	-2,42

Os genes resultantes da filtragem possuem um comportamento em comum de acordo com o grupo. Por exemplo, o gene YDR328C está superexpresso em todas as HTs, já YLR079W está sub-expresso em todas LTs. Os genes que possuíam comportamento semelhante entre os dois grupos também foram selecionados, tal como o gene YLR210W, o qual está sub-expresso nos dois grupos. As simulações dos fenótipos LT e HT foram realizadas modificando todos os nós dos genes presentes na Tabela 6. O GINsim permite limitar os valores dos nós em um determinado intervalo. Para simular o fenótipo LT, por exemplo, os genes que estavam superexpressos para as LTs foram limitados para assumir o valor de “3” sempre que ativos, enquanto que aqueles que estavam subexpressos foram limitados no intervalo de “0” ou “1”. O mesmo princípio foi aplicado para as simulações do fenótipo HT. Dessa forma, foi realizada simulações para cada uma das linhagens de acordo

com os perfis de expressão da Tabela 6, e os comportamentos diferentes entre as HTs e entre as LTs foram atribuídos como as características do fenótipo.

As simulações foram realizadas como previamente descrito. Os grafos de transição de estados foram analisados por meio da observação dos atratores do sistema, verificando se o os grafos ciclam ou permanecem em um estado estável. Para isso, apenas os nós do grafo de transição de estados que compunham o atrator foram selecionados e analisados. Definiu-se que a diferença entre os atratores é o que define as diferenças entre os grupos HT e LT, tal como será discutido mais adiante.

No estudo anterior do grupo de pesquisa, os lncRNAs encontrados foram avaliados computacionalmente quanto a sua capacidade de interagir fisicamente com proteínas. Os lncRNAs que interagem com proteínas do ciclo celular foram modelados como sendo reguladores dessas proteínas, pois com base na literatura sabe-se que eles podem exercer essa função. Portanto, os lncRNAs foram simulados como sendo ativadores ou inibidores das proteínas alvo. Para simulá-los como inibidores, a equação de seus alvos foi modificada para receber um efeito negativo (diminuição de valor) sempre que os lncRNAs estiverem ativos. O efeito ativador dos lncRNAs foi definido como um aumento de valor do alvo quando os lncRNAs estiverem ativos; a inibição ou ativação completa dos alvos ocorrem quando o lncRNA é superexpresso. Devido a ausência de informações sobre a ativação estes lncRNAs, dois experimentos foram realizados: 1- atribuindo os lncRNAs como elementos constitutivos, ou seja, presentes durante todo o ciclo; 2- atribuindo os lncRNAs como expressos em fases específicas no ciclo celular.

Como os lncRNAs aqui estudados são de duas linhagens diferentes, para simular a linhagem SEY6210, os lncRNAs referentes a outra linhagem foram retirados da equação, situação semelhante para simular os lncRNAs da BY4742. Portanto, embora eles estejam presentes na estrutura do modelo (Figura 10), eles não são parte de um mesmo sistema. As

equações para todos os componentes da rede estão presentes no Arquivo Suplementar 1 (Tabela S1).

3.5. Docking molecular

Dentre os lncRNAs testados, o transcr_9136 da linhagem SEY6210 foi o que apresentou os resultados mais promissores. Portanto, o *docking* molecular foi realizado para adquirir maiores detalhes de como ocorre a interação entre ele e suas proteínas alvo (Hsl1p e Gin4p).

Primeiramente, a estrutura da proteína Hsl1p (número de acesso: P34244) foi extraída do banco de dados PDB (Protein Data Bank) na forma de um complexo com outras proteínas. Este complexo publicado por Zhang et al. (2016), continha o complexo APC junto com Cdc20p e Hsl1p. No entanto, a Hsl1p presente na estrutura era composta apenas pela porção responsável pela função da proteína. A porção do complexo referente apenas à proteína Hsl1p foi manualmente separada e a energia da estrutura foi minimizada utilizando o GROMACS. Para isso, o arquivo Gromacs foi gerado utilizando o campo de força AMBER99SB-ILDN protein. Em seguida, a molécula foi solvatada, os íons foram gerados e adicionados ao modelo; por fim, a energia da estrutura foi minimizada e estrutura foi utilizada na etapa de *docking*. Para a proteína Gin4p (número de acesso: Q12263), o mesmo procedimento foi realizado; porém, a estrutura publicada por Moravcevic et al. (2010) já continha a porção C-terminal da proteína isolada, sendo apenas necessária a remoção de íons sulfato e gliceróis presentes na estrutura.

Como não há estruturas secundárias e terciárias disponíveis para os lncRNAs aqui estudados, duas etapas de predição foram aplicadas para prever essas estruturas. Na primeira etapa, foi realizada a predição da estrutura 2D utilizando o *web server* RNAfold. Na segunda etapa, a estrutura 2D em formato *dot-bracket* e a sequência de nucleotídeos

foram analisadas usando o *web server* 3dRNA (WANG et al., 2019) a fim de prever uma estrutura 3D.

O *docking* entre os lncRNAs e proteínas-alvo foi realizada no *software* HDOCK (YAN et al., 2017), o qual foi projetado para o estudo de interações proteína-proteína, RNA-proteína ou DNA-proteína. Devido ao tamanho do lncRNA, ele foi escolhido para ser o receptor de moléculas e as proteínas Hsl1p e Gin4p como ligantes. Avaliações da viabilidade dos resultados foram realizadas a fim de suportar aos resultados obtidos. Para isso, a sequência de aminoácidos das proteínas foram aplicadas no *software* DRNApred (YAN and KURGAN, 2017) para avaliar a probabilidade dos resíduos dessas proteínas se ligarem com fitas de DNA ou RNA.

3.6. Efeitos do dano ao DNA para cada linhagem

De acordo com os resultados prévios de citometria de fluxo utilizando o anticorpo H2AX para medir dano ao DNA, há indícios de uma diferença de comportamento entre as linhagens (WOLF, 2019). Esses resultados estão apresentados na Tabela 7.

Tabela 7: Porcentagem de células que acumulam dano ao DNA (WOLF, 2019).

Linhagem	Controle	Tratamento	Fold-change (tratamento/controle)
BMA64-1A	18,03	22,09	1,23
BY4742	7,88	18,15	2,30
X2180-1A	4,14	18,42	4,45
BY4741	12,56	23,72	1,89
S288c	2,88	6,01	2,09
SEY6210	9,93	34,98	3,52

Para entender como essas diferenças afetam o ciclo celular em cada linhagem individualmente e o porquê de haver uma alta porcentagem de células acumulando dano ao DNA, os dados de fold-change (*up* ou *down*) dos genes responsáveis pelo *checkpoint* do dano ao DNA (Tabela 8) foram aplicados no modelo seguindo a mesma lógica explanada

na seção 3.4, limitando os valores dos nós para assumirem “0” ou “1” caso o gene da proteína esteja subexpresso (por exemplo, Rad53p) ou limitando o valor do nó para apenas “3” caso o gene da proteína em questão esteja superexpresso (por exemplo, Mec1p).

Tabela 8: Valores de fold-change para cada gene do *checkpoint* do dano ao DNA para cada linhagem. Os espaços vazios indicam que o gene não está diferencialmente expresso.

	BMA64-1A	BY4742	X2180-1A	S288C	BY4741	SEY6210
MEC1	-	0.61	0.26	-	2.9	-
RAD53	-	-0.62	-0.57	-0.75	-1.8	-
CHK1	-	0.81	0.75	0.4	-	0.56
PDS1	-1,26	-0,87	-0,39	-	-	-

3.7. Efeitos dos lncRNAs no complexo ORC

As linhagens SEY6210 e BY4742 possuem os lncRNA Inc_9136 e Inc_10027, respectivamente, superexpressos e interagindo com o complexo ORC; o complexo ORC é essencial para a duplicação do material genético. Para avaliar como a ação desses lncRNAs influenciam o ciclo celular, ambos foram adicionados ao modelo. As simulações foram conduzidas aplicando os lncRNAs como inibidores ou ativadores de três formas: 1-levando em consideração o grupo em que elas pertenciam (HT ou LT), limitando os valores das proteínas de todo o ciclo; 2-levando em consideração apenas o nível de expressão dos elementos que faziam parte ou interagem com a ORC (nesse caso todos foram atribuídos como subexpressos); 3-considerando uma célula *wild-type*. A simulação foi realizada e os resultados foram obtidos da mesma forma como descrito na seção 3.4.

4. Resultados e Discussões

4.1. Validação do modelo

A situação esperada para a modelagem do ciclo celular é que a simulação de uma condição normal retorne um ciclo de estados no qual o sistema parte de um *input* inicial, passa por todas as fases culminando na citocinese e depois inicie novamente o ciclo. A

Figura 13 representa o resultado obtido para a simulação da levedura *wild-type* (WT) em condições normais, ou seja, sem nenhuma mutação e nem elementos que causem a ativação de um *checkpoint*. O estado inicial do atrator cíclico está indicado como “1” e o estado final como “23”, retornando ao estado “1” novamente. Na simulação WT, é possível observar que o sistema atinge uma estabilidade após o *input* inicial, sendo capaz então de progredir até completar a citocinese e iniciar o ciclo novamente. Esse fenômeno é indicado pela massa, a qual atingiu o valor de “2” no estado “21”, decresceu para “1” no estado “22” e retornou a “0” no estado “23”, reiniciando o ciclo novamente no estado “1”; isso gerou um atrator cíclico. Embora, naturalmente, as células passem a maior parte do tempo em G2, a simulação WT apresentou um maior número de estados em M devido ao fato dos modelos lógicos não serem capazes de trabalhar com a unidade tempo (não simulam nem as reações e nem a quantidade de tempo que cada dura) (LE NOVERE, 2015).

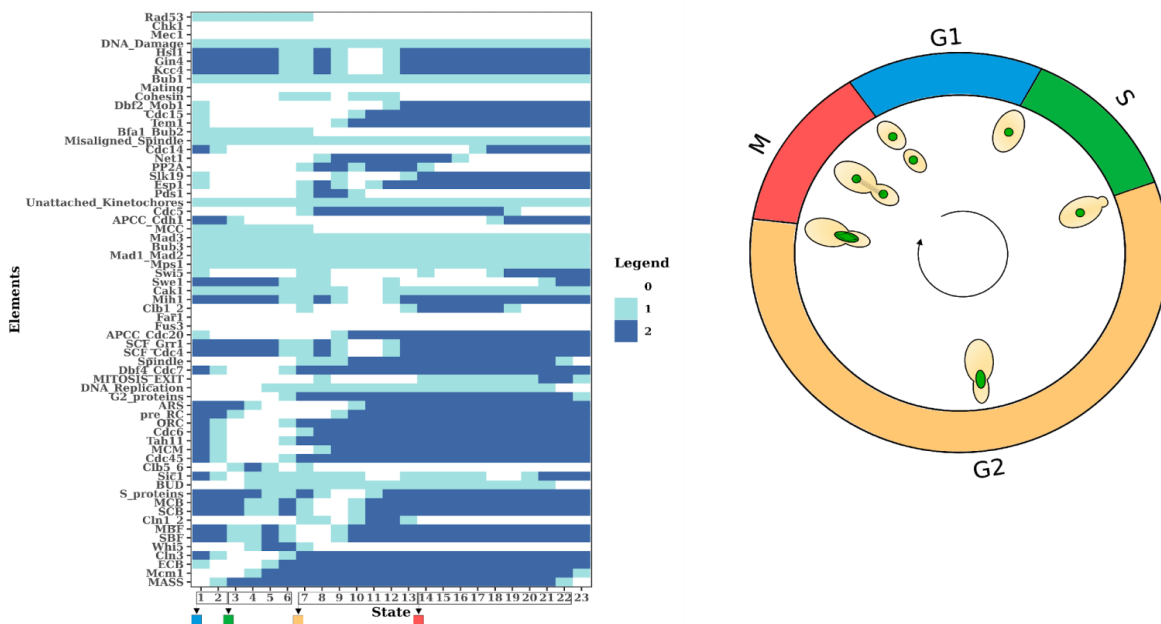


Figura 13: Simulação no modelo de uma levedura *wild-type* (WT) sem nenhuma perturbação ou *checkpoint* ativo. O gráfico reporta as transições de estados para cada elemento da rede, contando apenas os estados que compunham o atrator, sendo o eixo “x” os estados e “y” os componentes do modelo. A coloração reflete o valor de cada componente em cada estado. A figura da direita representa as fases do ciclo celular e, neste caso, indica que o ciclo está ocorrendo normalmente. A coloração dos estados indica em qual ponto do ciclo celular eles estão ocorrendo.

Quando se trata de sistemas biológicos, deve-se levar em conta a sua robustez, isto é, um sistema deve manter-se funcional mesmo perante perturbações. Sabe-se que o ciclo celular é um sistema muito robusto, então é esperado que a maior parte dos atratores dos *inputs* aleatórios sejam atratores cíclicos representando o ciclo celular completo. De fato, dos 100 *inputs* aleatórios testados apenas 18 retornaram um atrator fixo indicando um *arrest* do ciclo celular. Os 82 *inputs* que retornaram um ciclo normal foram simulados simultaneamente, sendo possível observar que todos eles convergem no mesmo atrator cíclico. Portanto, como explicado na introdução, quanto maior a bacia de atração, maior é a relevância biológica dela (HOPFENSITZ et al., 2013). De fato, a relevância biológica do ciclo celular com funcionamento normal é muito maior do que os casos em que ocorrem *arrest*, e por esse motivo, a maioria dos *inputs* convergiram nesse mesmo atrator. Isso comprova que o modelo não está enviesado, pois conseguiu representar algo esperado mesmo partindo de *inputs* randômicos.

Das 109 mutações testadas no modelo (Tabela 10; Tabela S1), o modelo atingiu a acurácia de 87,27%. A Tabela 9 contém os resultados obtidos por meio da matriz de confusão.

Tabela 9: Resultados obtidos da matriz de confusão.

Classe	Especificidade	Sensibilidade
Arrest em G1	1,0000	0,2000
Arrest em S	0,9816	1,0000
Arrest em G2	0,9626	1,0000
Arrest em M	0,9663	0,9524
Inviável	1,0000	0,8667
Viável	0,9167	0,9000

Acurácia: 0,8727 – P-Value: 2,2e-16 – Kappa: 0,8146

Especificidade é a medida da proporção de resultados negativos que são identificados corretamente como sendo negativos, enquanto que sensibilidade é a medida da proporção de resultados positivos que são identificados corretamente como positivos. Dessa forma, é

desejado valores próximos de 1 para ambas métricas, indicando que o número de falsos positivos e falsos negativos são baixos. Dentre os valores presentes na tabela 9, a sensibilidade da classe “*Arrest em G1*” demonstra que o modelo não é capaz de identificar corretamente os mutantes que sofriam *arrest* nesta fase. No entanto, essa baixa sensibilidade pode ter sido causada pelo fato de 7 simulações retornarem como resultado “*Arrest em G1*”, enquanto que o mutante era descrito como inviável pela literatura. Logo, esses 7 elementos da classe “*Arrest em G1*” foram aqui substituídos por “*Inviável*”. Contudo, se contarmos estas 7 simulações como acertos para a classe “*Arrest em G1*”, a sensibilidade sobe para 0,6666, o que continua ainda assim um valor baixo. Esses resultados indicam que a modelagem de perturbações que causam *arrest* em G1 é o ponto que o modelo possui menor precisão (Tabela Suplementar 1).

As três classes que apresentaram os melhores resultados foram “*Arrest em S*”, “*Arrest em G2*” e “*Arrest em M*”, indicando que as perturbações foram muito bem modeladas provavelmente em virtude do grande número de informação disponível acerca de perturbações nestas fases. A baixa especificidade e sensibilidade da classe “*Viável*” e a baixa sensibilidade da classe “*Inviável*” indicam que o modelo precisa de melhorias na identificação desses mutantes.

O coeficiente Kappa é uma métrica que compara a acurácia observada com a acurácia esperada de um modelo randômico, refletindo a concordância do modelo aos dados reais. Quanto maior o valor maior é a diferença entre a acurácia do modelo em comparação à acurácia gerada ao acaso, logo, maior é a confiabilidade do modelo não ser randômico. Valores menores que 0 são caracterizados como *no agreement*, entre 0 e 0,20 como *slight*, de 0,20 à 0,4 como *fair*, de 0,4 a 0,6 como *moderate*, de 0,6 a 0,8 como *substantial* e de 0,8 a 1 como *almost perfect agreement* (SUN, 2011). Como o modelo apresentou um Kappa de 0.81, podemos afirmar que as simulações concordantes com os dados de literatura são realmente fruto da boa modelagem do sistema e se distanciam de um resultado aleatório.

Em suma, pode-se afirmar que o modelo é uma boa representação do comportamento real do ciclo celular, podendo então ser usado para predições de novos fenótipos e estudos mais aprofundados sobre essa via.

Porém, os dois maiores problemas do modelo estão relacionados à dupla-replicação do material genético e a relação entre as vias FEAR e MEN. Para garantir a estabilidade do genoma, o DNA deve ser replicado apenas uma vez por ciclo (MIMURA et al., 2004; NGUYEN; CO; LI, 2001), porém, no modelo desenvolvido, esse mecanismo de inibição da dupla-replicação do DNA não foi modelado. Portanto, qualquer teste de mutações em que a dupla-replicação do DNA cause algum efeito na célula retornará um fenótipo normal. Apesar da interação entre as vias FEAR e MEN estar bem modelada, no modelo aqui apresentado há uma dependência do nó “Dbf2_Mob1” (ativado por “Cdc15”) para a liberação do “Cdc14”. Porém, alguns mutantes conseguem escapar desta dependência de Cdc15p, sendo que grande parte destes mutantes age inibindo Net1p ou sua interação com Cdc14p. Logo, Cdc14p não fica sequestrado no núcleo e consegue realizar sua função sem a necessidade de “Dbf2_Mob1” (SHOU; DESHAIES, 2002). Nas versões mais antigas do modelo, este detalhe foi adicionado, no entanto, pelo fato do modelo lógico não levar em consideração o “tempo” em que Cdc14 leva para deixar o núcleo, a sua atividade apresenta uma queda muito grande no número de estados do ciclo. Este problema acaba afetando o teste de alguns mutantes que tinham seu comportamento simulado corretamente antes da modificação. Logo, para preservar uma maior acurácia do modelo, foi decidido não modelar este detalhe da via do Cdc14.

Tabela 10: Algumas mutações testadas no modelo para a validação. A tabela completa e as referências de cada mutação estão no Arquivo Suplementar 1 (Tabela S1).

Perturbação	Predição do modelo	Comportamento Real
Cdh1- Δ Cdc6- Δ 2-49	Viável	Viável
APCC- Δ GAL_Clb2	<i>Arrest</i> em M	Inviável
Cdc20-ts Bub2- Δ	<i>Arrest</i> em M	<i>Arrest</i> em M
Cdc20-ts GAL-Esp1	<i>Arrest</i> em M	<i>Arrest</i> em M

Cdc20-ts Mad2-Δ	<i>Arrest em M</i>	<i>Arrest em M</i>
Cdc20-Δ	<i>Arrest em M</i>	<i>Arrest em M</i>
Cdc20-Δ (APC-A) Cdh1-Δ	<i>Arrest em M</i>	Letal
Cdc20-Δ (APC-A) Cdh1-Δ Sic1 MC	Viável	Viável
Cdc20-Δ (APC-A) GAL-Clb2	<i>Arrest em M</i>	Inviável
Cdc20-Δ (APC-A) Sic1-Δ	Viável	Viável
Cdc20-Δ Pds1-Δ	Viável	Inviável
Clb5-Δ Cdc20-Δ	<i>Arrest em M</i>	<i>Arrest em M</i>
Clb5-dbΔ	Viável	Viável
Clb5-dbΔ Pds1-Δ	Viável	Viável
Clb5-dbΔ Pds1-Δ Cdc20-Δ	Viável	Viável
Clb5-dbΔ Sic1-Δ	Viável	Viável
Cln1/2-Δ	Viável	Viável
Cln1/2-Δ Clb5/6-Δ	<i>Arrest em G1</i>	<i>Arrest em G1</i>
Cln1/2-Δ Cdh1-Δ GAL-Sic1	Viável	Viável
Cln1/2-Δ Cdh1-Δ	Viável	Viável
Cln1/2-Δ Cln3-Δ	<i>Arrest em G1</i>	Inviável
Cln1/2-Δ Cln3-Δ APC-ts	<i>Arrest em G1</i>	Inviável
Cln1/2-Δ Cln3-Δ Cdh1-Δ	<i>Arrest em G1</i>	Inviável
Cln1/2-Δ Cln3-Δ GAL-Clb5/6	Viável	Viável

4.2. Simulação dos grupos HT e LT

O *arrest* é identificado quando a massa não consegue reduzir seu valor para zero.

Tendo como exemplo a Figura 13, a massa parte de zero no estado 1 e a partir do estado 22 ela começa a diminuir e atinge zero no estado 23, o qual retorna ao estado 1 para repetir o

mesmo processo. Quanto aos testes dos fenótipos das HTs e LTs, observou-se que o ciclo celular em ambos os casos sofrem um *arrest*. Nas figuras 14 e 15 observa-se que o grafo de transição de estados finaliza nos estados 14 e 25, respectivamente, antes da massa retornar ao valor zero para recomeçar o ciclo, indicando os *arrests* dessas simulações.

Quanto aos testes dos fenótipos das HTs, observou-se que o ciclo celular sofre um *arrest* em G1 devido à alta atividade dos complexos SCF, os quais estão superexpressos neste grupo. Porém, sabe-se que a deleção ou expressão condicional (mutante termo-sensível) de uma das proteínas do complexo SCF, a Skp1p, resultou em efeitos deletérios para o ciclo celular, causando *arrests* em G1/G2 e tornando a célula inviável (STEMMANN e LECHNER, 1996; CONNELLY e HIETER, 1996). No entanto, estudos da superexpressão desse complexo ainda precisam ser feitos para avaliar o seu impacto no ciclo celular. Nesse sentido, o modelo contribui indicando que há uma possibilidade de *arrest* em G1 devido à superexpressão desse complexo.

Em contrapartida, para o grupo das LTs, o ciclo celular sofreu um *arrest* em M (Figuras 14 e 15) devido à falta de degradação da proteína Clb1/2p. Há relatos na literatura que a proteólise dessas ciclinas (principalmente mediada pelo complexo APC) é essencial para a saída da mitose (WASH e CROSS, 2002; PFLEGER e KIRSCHNER, 2000). Células com ausência dos complexos APC ou com o mutante Clb2- Δ db (sem os sítios de ligação do APC) mantêm os níveis de Clb1/2p altos durante todo o ciclo celular, tornando a célula incapaz de completar o ciclo celular (CROSS, 2003).

O modelo mostrou que o conjunto de proteínas diferencialmente expressas do grupo LT resultou em algo similar (a não degradação de Clb1/2p), logo a célula dessas leveduras não é capaz de terminar a ciclo da mitose com a citocinese.

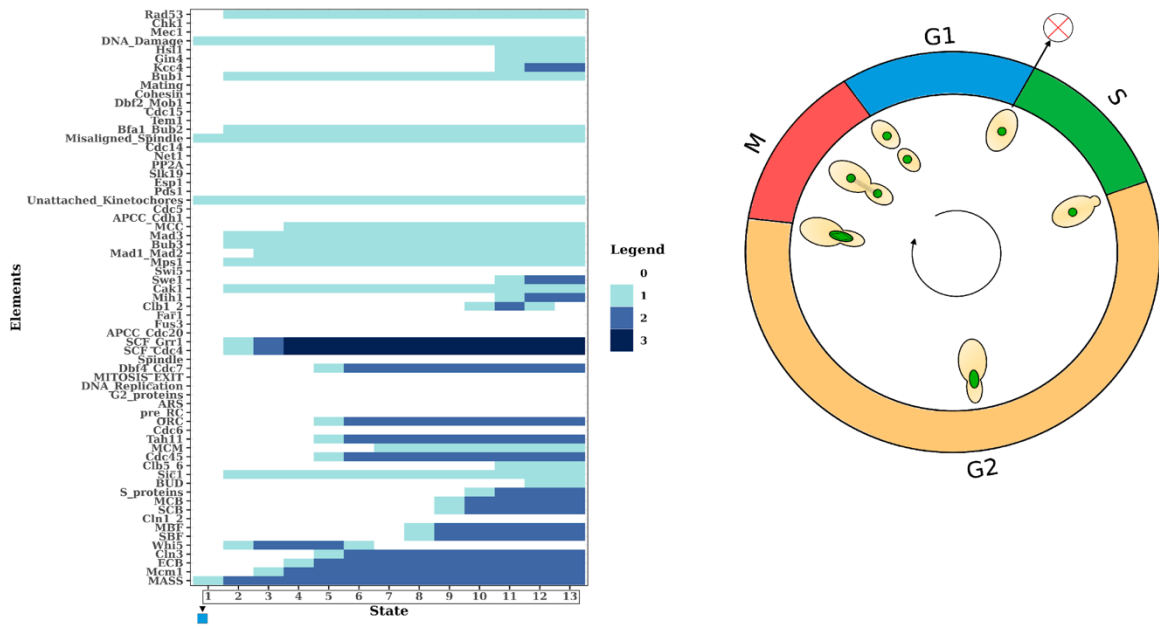


Figura 14: Resultado obtido com a simulação dos dados de transcriptoma das leveduras do grupo HT. A seta preta apontando para o “x” vermelho simboliza um arresto naquela posição.

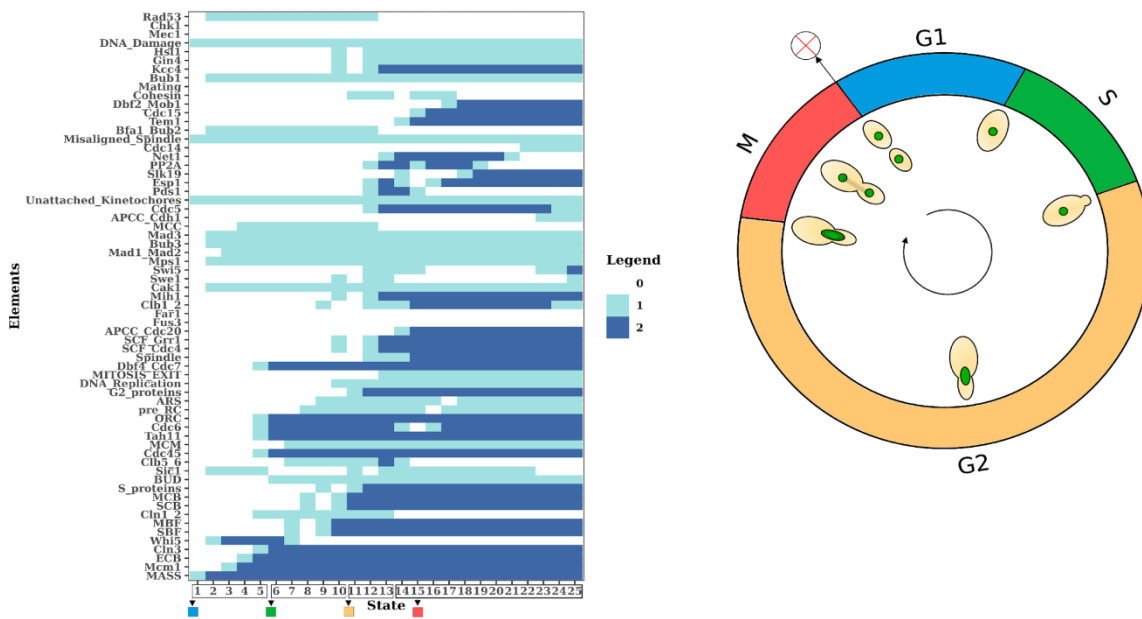


Figura 15: Resultado obtido com a simulação dos dados de transcriptoma das leveduras do grupo LT, diferente das HTs, estas sofreram um arresto no final de M.

De acordo com as informações das curvas de retomada crescimento, um experimento em que as linhagens HTs e LTs, após o severo tratamento com etanol foram cultivadas em um meio sem nenhum estressor por mais 24h (ALMEIDA, 2017), demonstrou-se que as leveduras LTs retomam o crescimento de forma mais acelerada que as HTs. Analisando o

resultado obtido pelas simulações, inferimos que a razão pela qual as LTs não apresentam a mesma deficiência que as HTs se deve ao fato de sofrerem um *arrest* em um estágio mais tardio do ciclo celular. Brewer et al. (1984) demonstraram que o tempo do ciclo celular para células mães é em média 99 minutos, enquanto que para as células filhas é em média 142 minutos. Essa diferença de tempo se dá principalmente na fase G1 do ciclo celular (BREWER; CHLEBOWICZ-SLEDZIEWSKA; FANGMAN, 1984). Em suma, como as LTs já passaram por G1 e estão paradas no final de M, provavelmente já possuem todos os pré-requisitos para sair da mitose, sendo a degradação de Clb1p e Clb2p as únicas exceções. Assim, inferimos que quando realocadas em um meio sem etanol, as LTs geram as células filhas muito mais rapidamente, tendo então um rápido crescimento para um novo ciclo. Isso resulta numa vantagem em tempo para as LTs quando comparadas as HTs, as quais ainda precisam passar por todo o ciclo celular. No entanto, não há estudos que avaliem esta retomada de crescimento de leveduras em diferentes fases do ciclo celular.

LncRNAs podem interagir fisicamente com proteínas-alvo (o tipo testado no modelo) e regulá-las de diversas maneiras. No caso de interações físicas, as duas formas mais comuns de regulação são agindo como “iscas” ou *scaffolds* de proteínas (LI et. al., 2019). Na primeira forma, os lncRNAs se ligam aos seus alvos e impedem que eles exerçam sua função (isso gera efeitos inibidores para o funcionamento da proteína). Por outro lado, na segunda forma de interação os lncRNAs permitem a formação de complexos proteicos em sua superfície podendo gerar efeitos promotores ou inibidores na proteína alvo (WANG e CHANG, 2011). Há relatos na literatura de lncRNAs interagindo com proteínas e as regulando negativamente. Por exemplo, em células humanas, o lncRNA LINC-PINT serve como isca para um complexo chamado *Polycomb Repressive Complex 2* (PRC2), impedindo que ele exerça sua função, a qual, caso esteja desregulada, pode levar a progressão de diversos tipos de cânceres (MARÍN-BEJAR et. al., 2017). Outro exemplo de lncRNA atuando como isca é o caso do lncRNA CAIF, o qual é capaz de se ligar

diretamente com a proteína p53 (fator de transcrição essencial para diversas funções celulares) e impedir a transcrição da proteína miocardina, sendo então capaz de atenuar o infarto do miocárdio (LIU et. al., 2018). Estes exemplos demonstram que os lncRNAs podem ser extremamente relevantes quando agindo como inibidores de seus alvos.

Nas simulações aqui apresentadas, quando os lncRNAs foram atribuídos como inibidores de seus alvos, geraram os resultados mais promissores. Isso indica que os lncRNAs podem ser relevantes também no ciclo celular em *S. cerevisiae*. A interação lnc_9136 de SEY6210 com Hsl1 foi predita por Marques (2019). Esse lncRNA está superexpresso nessa levedura e foi simulado de duas formas pelo modelo: 1- de forma constitutiva com valor fixado em “3”; 2- expressado em seu valor máximo periodicamente em G2. A primeira forma não apresentou uma mudança no fenótipo LT, permanecendo então em *arrest* na fase M. Porém, na segunda simulação, observa-se que o modelo cicla normalmente. Em suma, o lncRNA lnc_9136 da levedura SEY6210 tem a capacidade de desativar os nós “Hsl1” e “Gin4” a partir do valor “2” e, quando superexpresso (valor “3”), ele é capaz de inibir completamente estes nós (Figura 16). As expressões booleanas utilizadas para modelar os lncRNAs estão presentes na Tabela Suplementar 3.

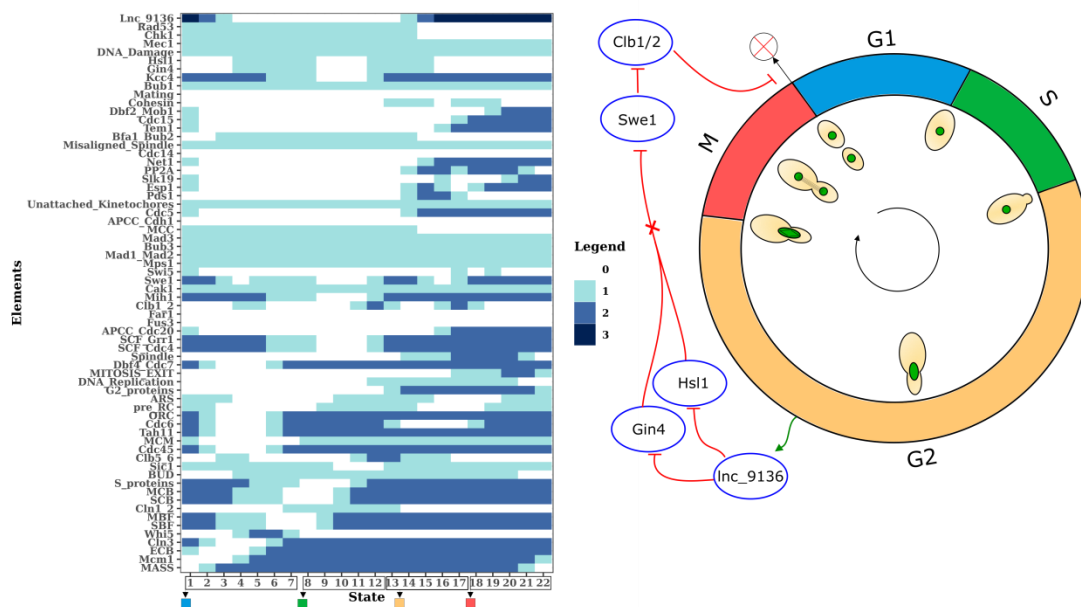


Figura 16: Resultado obtido para a simulação do lnc_9136 superexpresso no fenótipo LT. O esquema da direita indica como o lncRNA age para liberar a célula do arrest. O “X” na aresta entre Hsl1 e Swe1 indica que a interação não está ocorrendo mais devido ao lncRNA.

Como é possível observar na Figura 16, a superexpressão do lncRNA Inc_9136 em G2 foi capaz de permitir a ação de Swe1 sobre Clb1_2, resultando em uma mitose completa por meio da inibição de Clb1p e Clb2p. Sabe-se que a ação conjunta de Hsl1p com Kcc4p e Gin4p é necessária para a organização da septina, conjunto de proteínas que atuam no estágio final da divisão celular (ASANO et. al., 2006). A organização da septina necessita da inibição de Swe1p para que a entrada da mitose (resultado da ação de Clb1/2p) ocorra de forma coordenada com a sua formação. Para isso, as proteínas Hsl1p, Kcc4p e Gin4p são capazes de inibir Swe5p, permitindo que Clb1/2p possa estar presente e iniciar a mitose (BARRAL et. al., 1999). Células mutantes que não possuem uma ou todas essas proteínas (Hsl1p, Kcc4p e Gin4p) são viáveis, apresentando apenas problemas morfológicos causados pela ação prolongada de Swe1 (BARRAL et. al., 1999). As simulações indicam que a inibição de Gin4p e Hsl1p na condição testada (14 proteínas com expressão regulada) resultou na liberação de Swe1p em quantidades suficientes para inibir Clb1/2p e causar a saída da mitose. Corroborando o que fora observado experimentalmente, pois a inibição do lncRNA causou uma perda de função completa da proteína (similar ao *knockout*), retornou um fenótipo viável e aumentou os níveis de Swe1p (como descrito na literatura).

A associação de lncRNAs com efeitos no ciclo-celular é algo que já foi documentado na literatura. Em humanos, por exemplo, o lncRNA GAS5 interage fisicamente com a proteína *Y-box binding protein 1* (YBX1); a inibição desse lncRNA causou um aumento na taxa de *turnover* dessa proteína impedindo o *arrest* em G1 típico de células cancerosas do estômago (LIU et. al., 2015). Outro exemplo é o caso do lncRNA UCA1 em células humanas. Este lncRNA apresentou indícios de ligação ao fator de transcrição FOXM1, pois a subexpressão do lncRNA diminuiu os níveis da proteína. Essa subexpressão teve resultados negativos para a célula causando a inibição da proliferação, invasão, bem como um indução do *arrest* em G1 no ciclo celular (WEI et. al., 2016). Então, observou-se que os lncRNAs podem causar *arrest* celular, algo que foi aqui observado pelas simulações.

Interações entre lncRNAs e ciclinas ou CDKs também já foram documentadas. O lncRNA GADD7 é capaz de regular os níveis da quinase dependente de ciclina Cdk6p através da interação com a proteína *Tar DNA binding protein* (TDP-43); TDP-43p estabiliza o mRNA de Cdk6p, no entanto, a sua ligação com o lncRNA causa a dissociação de TDP-43p do mRNA, permitindo que este seja degradado. A inibição de Cdk6p impede que a transição de G1 para S seja feita (*arrest* em G1) (KITAGAWA et. al., 2013). O lncRNA MALAT1 também é capaz de controlar o ciclo-celular em humanos. Um estudo indicou que a depleção de MALAT1 causa um *arrest* em G1 e aumenta os níveis de alguns fatores de transcrição (como o p53) (TRIPATHI et. al., 2013). Também, a depleção desse lncRNA causa a diminuição dos níveis das ciclinas A2 e de Cdc25A, resultando em *arrest* em G1. No entanto, também há relatos de que ele pode controlar a transição de G2 para M através da interação com um fator de transcrição (B-Myb), responsável pela expressão de diversas ciclinas e CDKs necessárias para essa transição (TRIPATHI et. al., 2013). Novamente, lncRNAs estão associados com *arrest* celular em diferentes fases do ciclo (G1, S ou G2), no entanto, ressalta-se que as relações entre lncRNAs e ciclinas ou CDKs, até o momento, são indiretas, ou seja, a ação dos lncRNAs culmina no controle da disponibilidade de ciclinas e CDKs mas sem uma interação física direta.

Há relatos de lncRNAs em *S. cerevisiae* que atuam em resposta ao estresse osmótico regulando o ciclo celular. O lncRNA CDC28 lncRNA é induzido pela Hog1p na presença de estresse osmótico. Dessa forma, o lncRNA passa a ser expresso na região *antisense* do gene da CDK Cdc28p, uma quinase que interage com todas as ciclinas do ciclo celular (interação necessária para a progressão do ciclo). Com a expressão do lncRNA, os níveis da proteína Cdc28p aumentam significativamente, permitindo que a célula inicie o ciclo celular mais rapidamente após o estresse (NADAL-RIBELLES et al., 2012; SOLÉ et al., 2015). Nesse caso, ressalta-se um efeito positivo da expressão de um lncRNA, tal como observado nas nossas simulações.

Portanto, nossas simulações evidenciam ser possível que lncRNAs possam interferir e controlar positivamente a progressão do ciclo celular em *S. cerevisiae*. Os resultados obtidos nessa etapa apresentam o padrão encontrado nos exemplos citados (lncRNA como agentes de controle da progressão do ciclo celular). Almeida (2017) demonstrou que a linhagem SEY6210 tem a segunda maior retomada de crescimento após estresse, perdendo somente para a linhagem S288C. As simulações elencam o lncRNA lnc_9136 como um regulador-chave dessa ótima retomada da SEY6210, em virtude de tal lncRNA atuar como um mecanismo que sobrepuja o *arrest* do fenótipo LT. O mecanismo molecular que creditamos para esse lncRNA é atuando como isca da Hsl1p e Gin4p, permitindo que Swe1p iniba as ciclinas Clb1/2p (Figura 16). Logo, o lncRNA também pode estar indiretamente relacionado ao controle da progressão do ciclo celular através das ciclinas, corroborando com os exemplos discutidos.

Quanto aos lncRNAs presentes na linhagem BY4742, não houve nenhum caso em que eles contribuíssem para a saída do *arrest* em G1 existente no fenótipo HT. No entanto, o lnc_10883 apresentou um resultado interessante para leveduras que estejam sofrendo com alto dano ao fuso. Estas leveduras normalmente sofrem um *arrest* em M devido ao dano no fuso (STRAIGHT, 1997), porém, uma simulação atribuindo uma superexpressão constitutiva deste lncRNA, não apresentou o *arrest* em M, gerando um ciclo normal (Figura 17). Observamos pela Figura 17 que o lnc_10883 inibe “Bub1”, componente necessário para a formação do MCC e responsável por causar o *arrest* na presença de dano ao fuso (STRAIGHT, 1997). De fato, já é documentado que leveduras com ausência dessa proteína (*knockout*) não são capazes de causar *arrest* em resposta ao tratamento com nocodazole (agente que interfere na polimerização dos microtúbulos), permitindo que completem o ciclo celular mesmo na presença dos cinetócoros desalinhados (GOTO et. al., 2011). O modelo indica que a falta de Bub1p causada pela ação deletéria do lncRNA (similar à um *knockout*) resulta em um efeito similar. Adicionalmente, o modelo permite entender como

isso ocorre, indicando que a falta de Bub1p mantém MCC sempre inativo, garantindo que o “APCC_Cdc20” ative a via FEAR. Com FEAR e “APCC_Cdc20” ativos, Clb1/2p pode ser degradada (CROSS, 2003), levando a completude do ciclo mesmo na presença de danos ao fuso. A interação Inc_10883 com Bub1 foi predita por Marques (2019).

Pouco se sabe sobre os efeitos dos lncRNAs em *checkpoints* do ciclo celular. Há relatos na literatura de que alguns deles podem estar associados com instabilidade genômica (LU et al., 2016; TRACY et. al., 2018). O *checkpoint* do fuso é o responsável por manter a estabilidade do genoma durante o ciclo celular (RUDNER e MURRAY, 1996), logo, é possível que estes lncRNAs estejam relacionados com esse *checkpoint*, porém ainda não foi realizada uma associação direta. Nesse contexto, as simulações contribuem indicando que há um novo possível lncRNA afetando o *checkpoint* do dano ao fuso, além disso, o modelo demonstra como é realizada essa regulação pelo lncRNA.

As simulações realizadas apontam possíveis novos métodos de controle da progressão do ciclo celular em *S. cerevisiae* mediado por lncRNAs. Os resultados possuem padrões observados na literatura e, além disso, a modelagem permitiu visualizar como podem ser os mecanismos utilizados pelos lncRNAs para regulação.

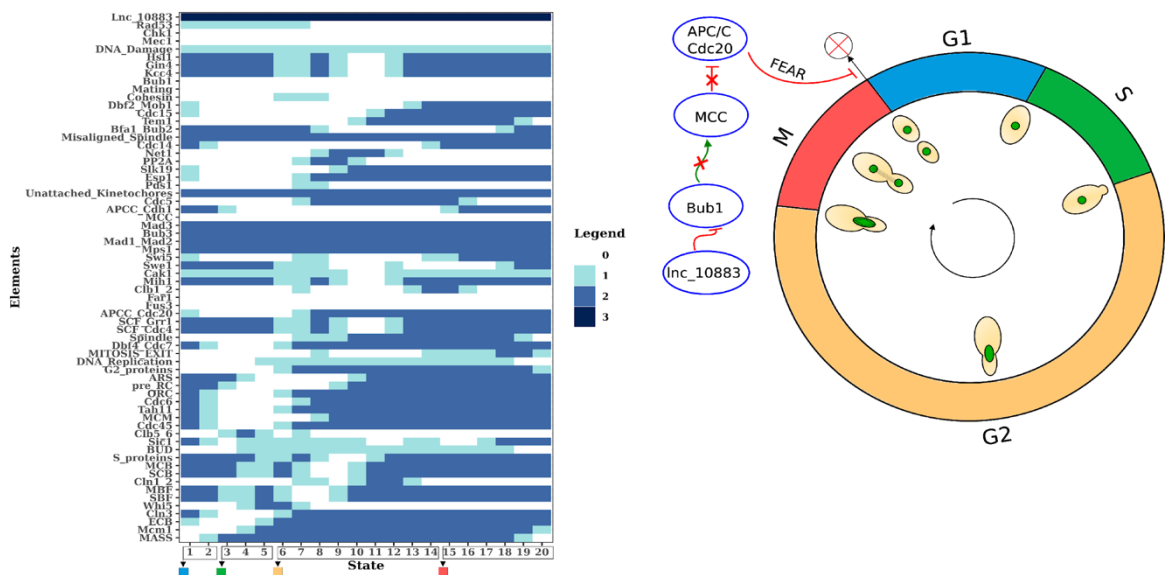


Figura 17: Resultado obtido pela simulação com a superexpressão constitutiva do *lnc_10883* em uma levedura com alto dano ao fuso. O esquema na direita demonstra como a ação do *lncRNA* faz com que a célula não entre em arrest na fase M.

4.3. Docking molecular da interação *lnc_9136*-Hsl1p/Gin4p

O *docking* realizado pelo HDOCK retornou 10 possíveis modelos. A escolha do melhor modelo foi feita levando em consideração, primeiramente, o *score* gerado pelo *software*. No entanto, caso a posição da proteína não apresentasse sentido físico (em alguns casos a proteína se ligava em regiões aparentemente inacessíveis), o segundo melhor *score* era levado em consideração, e assim por diante. Dos 10 resultados obtidos, o terceiro modelo foi selecionado como sendo a melhor opção de *docking* para o *lnc_9136* (SEY6210) com Hsl1p e o primeiro modelo para Gin4p, pois eram aqueles com o maior *score* e com sentido físico. A Figura 18 contém a representação de todos os modelos simultaneamente, do melhor modelo e da proteína Hsl1 após a minimização de energia. A Figura 19 representa os dados para a proteína Gin4p.

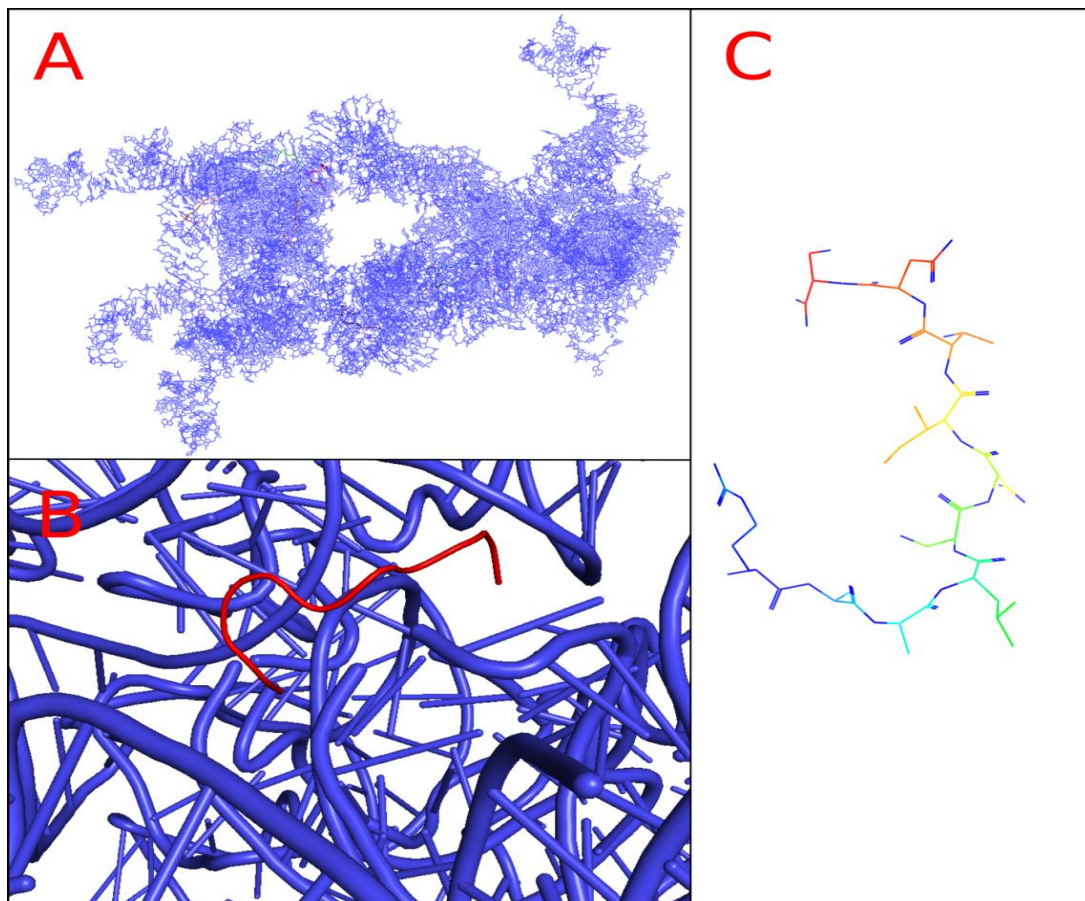


Figura 18: Resultados do docking molecular do lnc_9136 com Hsl1p. Em A, o lncRNA (azul) com todos os modelos de ligação com a proteína (demais cores). Em B, o melhor modelo do docking. Em C, a proteína Hsl1p com a energia minimizada.

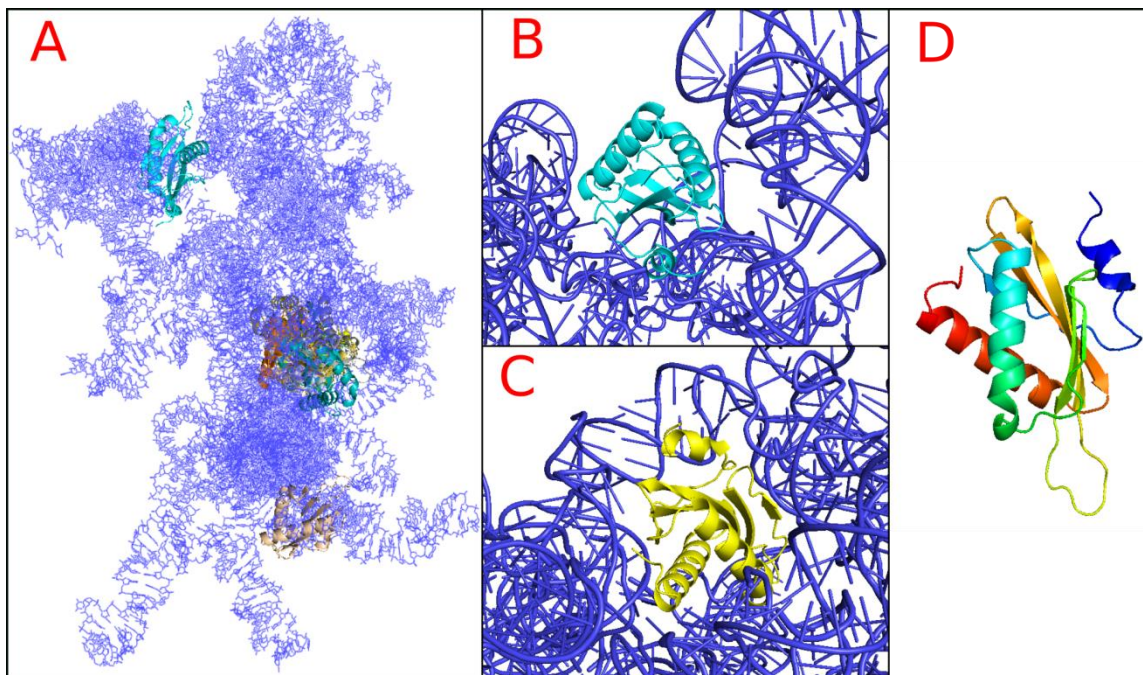


Figura 19: Resultados do docking molecular do lnc_9136 com Gin4p. Em A, o lncRNA (azul) com todos os modelos de ligação com a proteína (demais cores). Em B, o melhor modelo de docking. Em C, uma das proteínas (segundo melhor modelo) que se concentraram em um único ponto do lncRNA. Em D, a estrutura da proteína Gin4p com a energia minimizada.

Embora a metodologia aplicada no *docking* não seja a ideal, não foi possível testar outros tipos de *softwares*. Na literatura há pouco estudo acerca desse tema aplicado à lncRNAs, sendo que a maioria das ferramentas disponíveis são construídas apenas para pequenos RNAs. Ao aplicar essas ferramentas em um lncRNA, praticamente todos retornaram erros, sendo o HDOCK a exceção. Até mesmo uma das ferramentas mais utilizadas para a *docking* de moléculas, o AutoDock Vina, não suportou o tamanho do lncRNA. Além disso, poucas ferramentas estão disponíveis para a predição de estruturas de lncRNAs; embora essas ferramentas aceitem esse *input* não é garantido que a predição possua uma acurácia desejada. A Figura 20-A mostra a estrutura predita do lncRNA lnc_9136, e em B é possível observar um *Gap*, indicando que nessa parte o *software* não foi capaz de prever corretamente a estrutura. Esses *Gaps* aparecem frequentemente ao longo do

lncRNA, provavelmente devido ao fato do *software* utilizado não ser específico para esse tipo de molécula. Dessa forma, como todas as ferramentas utilizadas nesta etapa foram desenvolvidas para pequenos RNAs, afirmações e hipóteses retiradas a partir dos resultados devem ser comedidas.

Quanto à predição dos resíduos que se ligam à RNA ou DNA, a proteína Hsl1p não apresentou nenhum resíduo capaz de se ligar a RNAs. No entanto, 47 resíduos apresentaram probabilidade de se ligarem ao DNA, sendo a maioria desses resíduos concentrados nas extremidades e no centro da proteína. Isso pode explicar a maneira que a proteína se ligou ao lncRNA, envolvendo a dupla fita com as extremidades da proteína com uma ligação mais próxima no centro. É importante ressaltar que a estrutura utilizada da Hsl1p é a porção da proteína referente à sua função como quinase (do aminoácido 81 ao 369), portanto é possível que o lncRNA esteja impedindo sua função de fosforilar Swe1p.

O mesmo foi encontrado para proteína Gin4p, contendo 69 resíduos capazes de se ligarem com DNA concentrados nas folhas betas e na menor alfa-hélice da estrutura. Isso pode ser observado na Figura 19, onde a proteína interage com mais proximidade pelas folhas beta e pela menor alfa-hélice (Figura 19-B e C). No caso da Gin4p, oito modelos ligaram a proteína no centro do lncRNA lnc_9136 indicando que existe uma alta possibilidade de ligarem-se nessa região; um desses modelos está representado na Figura 19-C. Embora tenha um bom *score*, a região da proteína aparenta ser de difícil acesso. A estrutura utilizada para a Gin4p (porção C-terminal) não é a responsável pela sua função de quinase, no entanto, é necessária para sua função como formadora do septo. Logo, a probabilidade de interferência no lncRNA em sua função de fosforilar Swe1p é baixa.

Embora nenhuma das proteínas tenha resultado favorável à ligação com RNAs (apenas com DNA), o fato da estrutura 3D do lncRNA conter várias estruturas em dupla-hélice (Figura 20) indica que é possível que essas proteínas se liguem nessas estruturas do lncRNA, pois possuem alta probabilidade de se ligarem ao DNA (Figura 20-B).

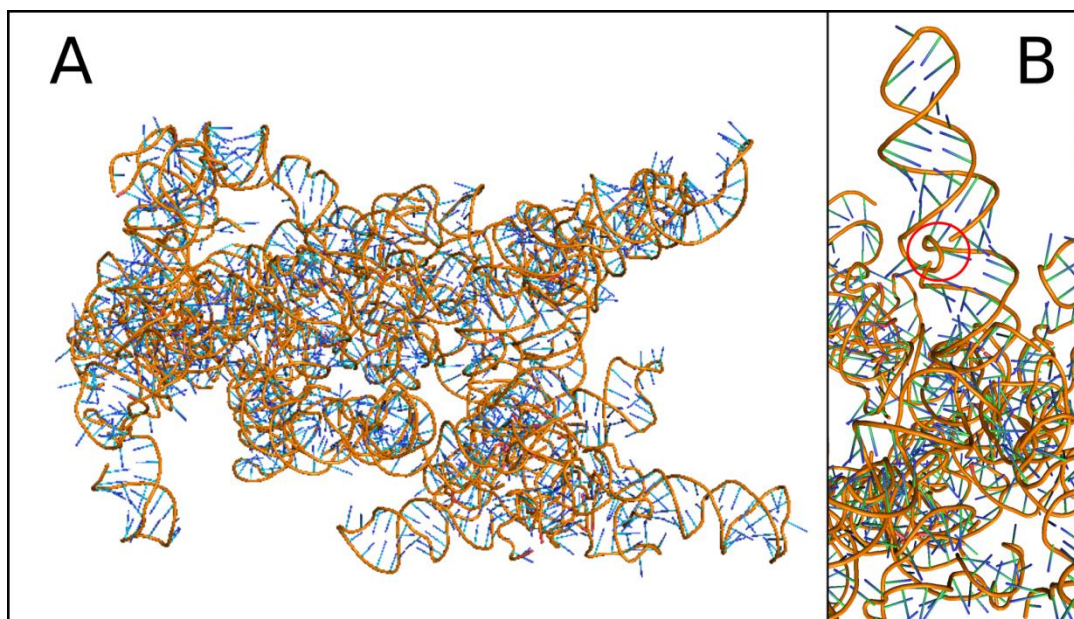


Figura 20: Resultado da predição de estrutura 3D para o lnc_9136. Em A, a estrutura do lncRNA gerada pelo 3dRNA. Em B, o círculo vermelho indica um dos Gaps que podem ser encontrados ao longo da estrutura.

Sabe-se que o tamanho dos lncRNAs conferem a essas moléculas um grande número de sítios de ligação à proteínas, os quais podem ser utilizados como *scaffolds* para a formação de complexos com proteínas (DUSZCYK, 2011), indicando que a sua superfície comporta a interação simultânea com múltiplas proteínas. Quanto aos sítios de reconhecimento de RNAs (RRM) presente em proteínas, sabe-se que esses sítios reconhecem preferencialmente sequências curtas de RNAs de fita simples (ssRNAs). Porém, as ligações podem ocorrer em regiões de dupla-fita do RNA, como os *hairpins* e *bulges* (BLYTHE et. al., 2016). O *docking* realizado corrobora com essas informações. Embora não haja relatos de uma mesma proteína se ligando em mais de uma região de um lncRNA, a análise aqui desenvolvida indica que há essa possibilidade. Também, a análise contribui demonstrando quais regiões das proteínas estudadas se ligam ao lncRNA, principalmente pela interação com os *hairpins* (Figura 18 e 19).

Embora o *docking* não possa confirmar se as interações lncRNA-proteínas realmente ocorrem (a confirmação deve ser proveniente de experimentos laboratoriais), nossas análises demonstram que, embora exista um modelo com melhor *score*, há diversas possibilidades de ligação das proteínas ao lncRNA, todas em lugares completamente

distintas no caso da proteína Hsl1p. O fato da Gin4p apresentar apenas 3 regiões distintas de ligação pode ser devido ao tamanho da proteína (maior que Hsl1p). Além disso, analisando as proteínas, podemos inferir que elas preferencialmente se ligam em regiões onde existam estruturas em dupla-hélice.

4.4. Efeito do dano ao DNA no ciclo celular em cada linhagem

As simulações indicaram que apenas duas linhagens (S288c e X2180-1A) foram capazes de causar um *arrest* no ciclo celular devido ao *checkpoint*. As outras linhagens completaram o ciclo mesmo na presença de dano ao DNA. Além disso, a linhagem BY4742 contém um lncRNA (lnc_10883) interagindo com a proteína Mec1p, o qual, quando modelado como inibidor, permitiu que a linhagem continuasse o ciclo celular mesmo na presença de elevado dano ao DNA.

Quando ocorre a detecção de danos ao material genético, o *checkpoint* é ativado causando a expressão de genes que serão responsáveis por parar a progressão do ciclo celular até que o reparo seja feito. Esse *arrest* pode ocorrer por duas ramificações, partindo da ação de Rad53p ou partindo da ação de Chk1p/Mec1p (normalmente as duas ocorrem simultaneamente) (CHEN e SANCHEZ, 2004). A eliminação de apenas uma ramificação não é o suficiente para evitar o *arrest*, o que acontece apenas quando as duas ramificações estão defeituosas (MELO e TOCZYSKI, 2002).

De acordo com os resultados obtidos pelas simulações, as linhagens S288c e X2180-1A, as quais conseguiram cessar o ciclo celular para realizar o reparo, apresentaram uma dessas ramificações funcionando (a via Chk1); diferentemente, as outras linhagens apresentaram as duas ramificações defeituosas. Esse resultado pode explicar o fato da S288c ser a linhagem que possui a menor porcentagem de células que acumulam dano ao DNA (WOLF, 2019), porém, a mesma afirmação não pode ser feita para X2180-1A. Para as outras linhagens, a alta porcentagem de danos ao DNA (WOLF, 2019) pode ser resultado

da disfunção do *checkpoint* do dano ao DNA, impedindo que a célula interrompa o ciclo celular para realizar o reparo.

Analisando a linhagem BY4742 foi possível identificar duas maneiras pelas quais a linhagem não consegue causar o *arrest* para reparo ao DNA: 1- embora ela tenha dois genes do *checkpoint* superexpressos, a falta de *Pdsp1* nas quantidades normais impediu o *arrest*, pois essa proteína é necessária para a inibição da rede FEAR (*ROCK* e *AMON*, 2009); 2- a presença do lncRNA *lnc_10883* que interage com *Mec1*, quando modelado como inibidor, foi capaz de impedir sozinho a ação da ramificação *Chk1/Mec1*. A modelagem desse lncRNA como ativador de *Mec1* não apresentou resultados significantes nesse caso, no entanto, quando foi simulado em uma levedura em condições normais, o lncRNA causou *arrest* na célula mesmo na ausência de dano ao DNA.

Um estudo já associou lncRNAs com o *checkpoint* do dano ao DNA. Sabe-se que o lncRNA *PANDA* está associado com a capacidade de células humanas a resistirem ao tratamento com agentes intercalantes do DNA (aumento do reparo do DNA danificado) (HUNG et. al., 2011), nesse caso, há uma associação direta entre lncRNA e *checkpoint*, indicando que é possível que lncRNAs afetem esses mecanismos.

De forma geral, podemos inferir que a maior causa do dano elevado ao DNA das leveduras tratadas com etanol se deve à inativação das duas vias responsáveis pela resposta ao dano, corroborando com o encontrado por Melo e Tockzyski (2002). A capacidade da linhagem S288c em lidar com o dano ao DNA se deve à manutenção de pelo menos uma ramificação do *checkpoint*. No entanto, podemos observar na Tabela 7, há um aumento da porcentagem de leveduras que acumulam danos ao DNA após o tratamento provavelmente devido à falta da outra ramificação. Isso acontece pela falta de uma das ramificações pode causar defeitos nesse *checkpoint* (MELO e TOCZYSKI, 2002). Porém, simular essa diferença quantitativa das ramificações é impossível com esse modelo.

4.5. Efeito dos lncRNAs no complex ORC

Na primeira situação testada, levando em consideração o fenótipo (HT ou LT), não houve nenhum caso em que os lncRNAs como ativadores ajudavam as linhagens a contornar os *arrests* característico de cada grupo. Assumindo os lncRNAs como inibidores gerou um efeito mais severo, pois a inibição da ORC não progride a célula para a fase S do ciclo celular, tendo então um *arrest* em G1. Para a linhagem BY4742, como ela já sofria com o *arrest* em G1 característico de seu grupo, o efeito do lncRNA não foi significativo. Porém, para a linhagem SEY6210, a qual possui um fenótipo LT, o lncRNA lnc_9136 impediu que ela progredisse até a fase M, permanecendo no *arrest* característico de seu grupo.

Na segunda situação testada, levando em consideração apenas os elementos que compunham a ORC ou interagiam com ela, o resultado foi o mesmo que o anterior. O intuito desse teste era verificar se os lncRNAs lnc_9136 e lnc_10027 teriam um efeito positivo como ativadores da ORC, visto que o complexo todo estava sub-expresso. No entanto, com ou sem os lncRNAs superexpressos, o sistema conseguiu finalizar o ciclo celular. Ressalta-se aqui uma limitação dos modelos qualitativos: embora a célula consiga finalizar o ciclo com esse complexo sub-expresso, a célula sofre um *delay* na progressão da fase S (TAKEHARA et al., 2008). Logo, se os lncRNAs exercessem uma função de ativação para esse complexo, talvez seria possível observar uma diferença no final caso o modelo fosse quantitativo e levasse em consideração o tempo em que cada reação leva para acontecer.

Na terceira situação testada, considerando uma célula *wild-type*, os lncRNAs como inibidores causaram um *arrest* em G1 em ambas as linhagens. Os efeitos negativos causados pela inibição do complexo ORC corroboram com Giaver et. al.(2002), os quais demonstram que a deleção das proteínas que compõe o complexo causou um fenótipo inviável em leveduras da linhagem S288c (GIAVER et. al., 2002).

Como os resultados dos lncRNAs como inibidores foram idênticos em todos os casos, a

Figura 20 exemplifica todas as modelagens.

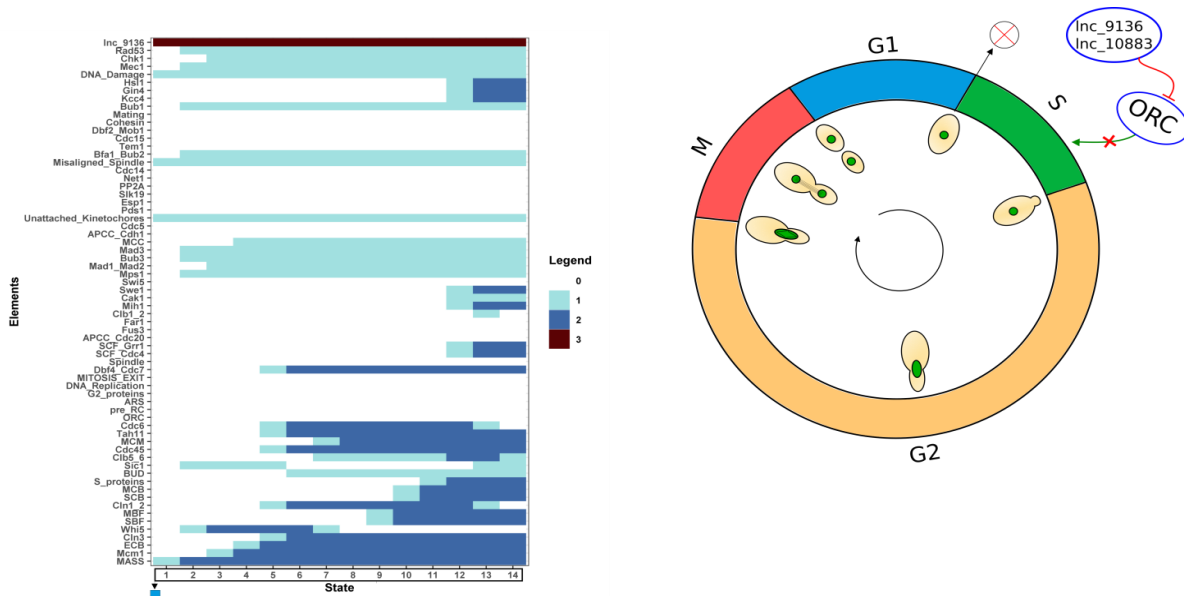


Figura 20: Resultado obtido pela simulação com a superexpressão constitutiva do lnc_10883 e lnc_9136 como inibidores da ORC. Como os resultados foram os mesmos para ambos, apenas o lnc_9136 está representado no heatmap. O esquema à direita demonstra como os lncRNAs impedem o início da fase S.

Esses resultados demonstram que os lncRNAs que interagem com proteínas essenciais para um determinado processo biológico tendem a causar um efeito negativo para o fenômeno do ciclo celular quando modeladas como inibidores.

5. Conclusões

O modelo desenvolvido nesta dissertação possui o maior número de proteínas modeladas para o ciclo celular além de ser o primeiro modelo que contempla lncRNAs nesse processo. É também a primeira utilização dos dados de expressão diferencial em modelagem para prever fenótipos e testar hipóteses. Portanto, o modelo desenvolvido inova ao trazer estes tipos de análise, além de contribuir com um aumento de complexidade ao adicionar mais proteínas na modelagem. Além disso, ele apresentou uma boa precisão e se mostrou livre de vies.

Como o modelo desenvolvido contém as expressões criadas de forma a representar com fidelidade o comportamento real do ciclo celular, ele pode ser utilizado por outros grupos de pesquisa para testar outras hipóteses, desde que se adequem às limitações do modelo.

De forma geral, as simulações dos fenótipos HT e LT sugerem que as LTs retomam o crescimento após o tratamento com etanol de forma mais rápida por estarem mais preparadas para realizar a citocinese em relação às HTs, as quais estão paradas antes mesmo do *START* ocorrer. Foi possível observar também que o alto dano ao DNA após o tratamento com etanol pode estar relacionado com a inibição das duas ramificações do *checkpoint* responsável pelo reparo, e que a manutenção de pelo menos uma dessas ramificações pode ser suficiente para a célula ser capaz de cessar o ciclo e realizar o reparo.

Os lncRNAs demonstram um potencial de serem importantes para o ciclo celular quando tratados como inibidores de seus alvos. De fato, um lncRNA da linhagem SEY6210 é capaz de livrar esta linhagem do *arrest* típico do fenótipo. No entanto, para a outra linhagem, nenhuma simulação envolvendo os lncRNAs foi capaz de contornar o fenótipo característico de HT. Porém, testando isoladamente os efeitos do dano ao DNA, um lncRNA presente na linhagem BY4742 demonstrou ser capaz de contornar no *arrest* característico desse fenótipo permitindo que a célula se divida mesmo com o DNA danificado; isso poderia explicar a alta porcentagem de células que acumulam dano ao DNA nessa linhagem. Embora, nesse caso, os lncRNAs como inibidores tenham agido de forma a favorecer a completude do ciclo celular, a simulação de outros dois lncRNAs que interagem com o complexo ORC demonstrou que a inibição desse complexo pode causar um *arrest* em G1. De forma geral, lncRNAs agindo como inibidores aparentam possuir muita importância no controle do ciclo celular.

As análises de *docking* molecular foram realizadas utilizando *softwares* que não foram desenvolvidos para tratar de lncRNAs, porém, ainda foi possível extrair informações

valiosas. Devido ao tamanho do lncRNA, as proteínas conseguiram se ligar a mais de uma região da molécula. Além disso, a ligação ocorria preferencialmente nas regiões de dupla fita do lncRNA, comportamento que foi encontrado na análise pelo *software* DRNAPred. Esse *software* indicou que ambas as proteínas só conseguem se ligar em duplas fitas e, de fato, os resíduos com maior probabilidade de se ligar a DNA eram os mais próximos do lncRNA no *docking*.

Embora o modelo tenha apresentado bons resultados, vale ressaltar as suas limitações. Por ser do tipo lógico, o modelo não é capaz de prever situações dependentes do tempo, tais como mutações que levam a uma diferença na massa final da levedura devido ao ciclo ter ocorrido mais rápido ou mais devagar que o normal. Também, o fenômeno biológico da reduplicação do DNA é extremamente difícil de ser modelado de maneira qualitativa por depender da variação da concentração de proteínas específicas em tempo real. Outro ponto negativo é a modelagem da intensidade em que algumas proteínas interagem. Por exemplo, a “C1b1/2”, a qual possui mais de um inibidor da proteína e cada um deles possui intensidades de inibição diferentes. Nesse caso, as expressões criadas foram capazes de serem capazes de modelar corretamente o processo, embora sejam extremamente difíceis de serem criadas e confusas de se trabalhar além de serem muito extensas (Arquivo Suplementar 1).

A modelagem de sistemas biológicos se mostrou uma ferramenta importante para o estudo de dados provenientes de técnicas *high-throughput*. Com essa abordagem, é possível criar novas hipóteses e testar as já existentes de uma maneira sistêmica, evitando o gasto desnecessário de tempo e recursos. No entanto, a modelagem de sistemas biológicos, ainda está em um estágio inicial, e possui várias limitações no desenvolvimento de modelos mais complexos. Embora seja possível extrair informações valiosas de um modelo matemático, ele não é o suficiente para provar hipóteses. O futuro da modelagem de sistemas biológicos depende da aliança entre previsões computacionais seguidas de testes em laboratório.

6. Referências

- ALBERGHINA, Lilia et al. Cell growth and cell cycle in *Saccharomyces cerevisiae*: Basic regulatory design and protein-protein interaction network. **Biotechnology Advances**, v. 30, n. 1, p. 52–72, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.biotechadv.2011.07.010>>.
- ALCASABAS, Annette A. et al. Control analysis of the eukaryotic cell cycle using gene copy-number series in yeast tetraploids. **BMC Genomics**, v. 14, n. 1, 2013.
- ALMEIDA, L. F. de. **Análise das linhagens de *Saccharomyces cerevisiae* expostas ao estresse por etanol**. 2017. Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências de Botucatu, 2017. Disponível em: <<http://hdl.handle.net/11449/152537>>.
- ASANO, Satoshi et al. Direct phosphorylation and activation of a Nim1-related kinase Gin4 by Elm1 in budding yeast. **Journal of Biological Chemistry**, v. 281, n. 37, p. 27090–27098, 2006.
- BARBERIS, Matteo et al. Cell size at S phase initiation: An emergent property of the G 1/S network. **PLoS Computational Biology**, v. 3, n. 4, p. 649–666, 2007.
- BARNUM, Kevin J; O’CONNELL, Matthew J. Cell Cycle Regulation by Checkpoints. **Methods Mol Biol**, v. 1170, n. 10, p. 29–40, 2014.
- BARRAL, Yves et al. Nim1-related kinases coordinate cell cycle progression with the organization of the peripheral cytoskeleton in yeast. **Genes & development**, v. 13, n. 2, p. 176–187, 1999.
- BLOOM, Joanna; CROSS, Frederick R. Multiple levels of cyclin specificity in cell-cycle control. **Nature Reviews Molecular Cell Biology**, v. 8, n. 2, p. 149–160, 2007.
- BLYTHE, Amanda J.; FOX, Archa H.; BOND, Charles S. The ins and outs of lncRNA structure: how, why and what comes next?. **Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms**, v. 1859, n. 1, p. 46–58, 2016.
- BREWER, B J; CHLEBOWICZ-SLEDZIEWSKA, E; FANGMAN, W L. Cell cycle phases in the unequal mother/daughter cell cycles of *Saccharomyces cerevisiae*. **Molecular and Cellular Biology**, v. 4, n. 11, p. 2529–2531, 1984.
- CHANDLER, M et al. A genomic approach to defining the ethanol stress response in the yeast *Saccharomyces cerevisiae*. **Annals of Microbiology**, v. 54, n. 4, p. 427–454, 2004.
- CHANG, Ya Lan et al. Yeast Cip1 is activated by environmental stress to inhibit Cdk1-G1 cyclins via Mcm1 and Msn. **Nature Communications**, v. 8, n. 1, p. 1–13, 2017. Disponível em: <<http://dx.doi.org/10.1038/s41467-017-00080-y>>.
- CHEN, Katherine C. et al. Integrative analysis of cell cycle control in budding yeast. **Molecular biology of the cell**, v. 15, n. 8, p. 3841–3862, 2004.

- CHEN, Yinhuai; SANCHEZ, Yolanda. Chk1 in the DNA damage response: conserved roles from yeasts to mammals. *DNA repair*, v. 3, n. 8-9, p. 1025-1032, 2004.
- CONNELLY, Carla; HIETER, Philip. Budding yeast SKP1 encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. *Cell*, v. 86, n. 2, p. 275-285, 1996.
- COSTANZO, Michael et al. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell*, v. 117, n. 7, p. 899–913, 2004.
- CVRCKOVÁ, F.; NASMYTH, K. Yeast G1 cyclins CLN1 and CLN2 and a GAP-like protein have a role in bud formation. *The EMBO Journal*, v. 12, n. 13, p. 5277–5286, 1993.
- CROSS, Frederick R. Two redundant oscillatory mechanisms in the yeast cell cycle. *Developmental cell*, v. 4, n. 5, p. 741-752, 2003.
- DE BRUIN, Robertus A.M. et al. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell*, v. 117, n. 7, p. 887–898, 2004.
- 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. Disponível em: <<http://www.liebertpub.com/doi/10.1089/10665270252833208>>.
- DESANY, B. A. et al. Regulation of RAD53 by the ATM-Like Kinases MEC1 and TEL1 in Yeast Cell Cycle Checkpoint Pathways. *Science*, v. 271, n. 5247, p. 357–360, 2006.
- DUSZCZYK, Malgorzata M. et al. The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. *Rna*, v. 17, n. 11, p. 1973-1982, 2011.
- EPSTEIN, Charles B; CROSS, Frederick R. CLB5: A novel B cyclin from budding yeast with a role in S phase. *Genes and Development*, v. 6, n. 9, p. 1695–1706, 1992.
- FAURÉ, Adrien et al. Modular logical modelling of the budding yeast cell cycle. *Molecular BioSystems*, v. 5, n. 12, p. 1787–1796, 2009.
- FELDMAN, R M Renny et al. A Complex of Cdc4p, Skp1p, and Cdc53p/Cullin Catalyzes Ubiquitination of the Phosphorylated CDK Inhibitor Sic1p* physiological ubiquitination reactions require E3s, and whether all E3s will participate directly in both substrate. *Cell*, v. 91, p. 221–230, 1997. Disponível em: <https://ac.els-cdn.com/S0092867400804043/1-s2.0-S0092867400804043-main.pdf?_tid=594f3dae-5f52-496e-b92c-c04f4f123ad4&acdnat=1528357318_a48d51ddcbc971006770f8c6d0bc9039>.
- FERRÈ, Fabrizio; COLANTONI, Alessio; HELMER-CITTERICH, Manuela. Revealing protein-lncRNA interaction. *Briefings in Bioinformatics*, v. 17, n. 1, p. 106–116, 2016.

- FORSBURG, S. Cell Cycle Regulation In The Yeasts *Saccharomyces Cerevisiae* And *Schizosaccharomyces-Pombe*. **Annual Review of Cell and Developmental Biology**, v. 7, n. 1, p. 227–256, 1991.
- FOSS, Margit et al. Origin recognition complex (ORC) in transcriptional silencing and DNA replication in *S. cerevisiae*. **Science**, v. 262, n. 5141, p. 1838-1844, 1993.
- FUTCHER, Bruce. Cyclins and the wiring of the yeast cell cycle. **Yeast**, v. 12, n. 16, p. 1635–1646, 1996.
- GAURAV, N. et al. Utilization of bioresources for sustainable biofuels: A Review. **Renewable and Sustainable Energy Reviews**, v. 73, n. September 2016, p. 205–214, 2017.
- GHADIRYANFAR, Mohsen et al. A review of macroalgae production, with potential applications in biofuels and bioenergy. **Renewable and Sustainable Energy Reviews**, v. 54, p. 473–481, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2015.10.022>>.
- GIAEVER, Guri et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *nature*, v. 418, n. 6896, p. 387-391, 2002.
- GOH, Phuay-Yee; SURANA, Uttam. Cdc4, a Protein Required for the Onset of S Phase, Serves an Essential Function during G 2 /M Transition in *Saccharomyces cerevisiae* . **Molecular and Cellular Biology**, v. 19, n. 8, p. 5512–5522, 1999.
- GUPTA, Anubhuti; VERMA, Jay Prakash. Sustainable bio-ethanol production from agro-residues: A review. **Renewable and Sustainable Energy Reviews**, v. 41, p. 550–567, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2014.08.032>>.
- HOPFENSITZ, Martin et al. Attractors in Boolean networks: A tutorial. **Computational Statistics**, v. 28, n. 1, p. 19–36, 2013.
- HU, Fenghua et al. Regulation of the Bub2/Bfa1 GAP complex by Cdc5 and cell cycle checkpoints. **Cell**, v. 107, n. 5, p. 655–665, 2001.
- HU, X. H. et al. Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*. **Genetics**, v. 175, n. 3, p. 1479–1487, 2007.
- HUNG, Tiffany et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. **Nature genetics**, v. 43, n. 7, p. 621, 2011.
- INGALLS, Brian P. Mathematical modeling in systems biology: an introduction. **MIT press**, 2013.
- IRONS, D. J. Logical analysis of the budding yeast cell cycle. **Journal of Theoretical Biology**, v. 257, n. 4, p. 543–559, 2009.

- JEOUNG, Doo-Il; OEHLEN, L. J. W. M.; CROSS, Frederick R. Cln3-Associated Kinase Activity in *Saccharomyces cerevisiae* Is Regulated by the Mating Factor Pathway . **Molecular and Cellular Biology**, v. 18, n. 1, p. 433–441, 1998.
- JIMÉNEZ, Javier et al. The role of MEN (mitosis exit network) proteins in the cytokinesis of *Saccharomyces cerevisiae*. **International Microbiology**, v. 8, n. 1, p. 33–42, 2005.
- KALDIS, Philipp; SUTTON, Ann; SOLOMON, Mark J. The Cdk-activating kinase (CAK) from budding yeast. **Cell**, v. 86, n. 4, p. 553–564, 1996.
- KITAGAWA, Masatoshi et al. Cell cycle regulation by long non-coding RNAs. **Cellular and molecular life sciences**, v. 70, n. 24, p. 4785-4794, 2013.
- KUBOTA, Shunsuke et al. Effect of Ethanol on Cell Growth of Budding Yeast: Genes That Are Important for Cell Growth in the Presence of Ethanol. **Bioscience, Biotechnology, and Biochemistry**, v. 68, n. 4, p. 968–972, 2004. Disponível em: <<http://www.tandfonline.com/doi/full/10.1271/bbb.68.968>>.
- KRAIKIVSKI, Pavel et al. From START to FINISH: computational analysis of cell cycle control in budding yeast. **NPJ systems biology and applications**, v. 1, n. 1, p. 1-9, 2015.
- KUNG, Johnny T.Y.; COLOGNORI, David; LEE, Jeannie T. Long noncoding RNAs: Past, present, and future. **Genetics**, v. 193, n. 3, p. 651–669, 2013.
- LABORATORIES, Kanehisa. **Cell Cycle - Yeast - *Saccharomyces cerevisiae* (budding yeast)**. Disponível em: <[https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=sce04111&keyword=cell cycle](https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=sce04111&keyword=cell%20cycle)>. Acesso em: 9 ago. 2019.
- LE NOVERE, Nicolas. Quantitative and logic modelling of molecular and gene networks. **Nature Reviews Genetics**, v. 16, n. 3, p. 146–158, 2015. Disponível em: <<http://dx.doi.org/10.1038/nrg3885>>.
- LEE, Sarah E. et al. Order of function of the budding-yeast mitotic exit-network proteins Tem1, Cdc15, Mob1, Dbf2, and Cdc5. **Current Biology**, v. 11, n. 10, p. 784–788, 2001.
- LI, Fangting et al. The yeast cell cycle is robustly designed. v. 2004, p. 1–6, 2004. Disponível em: <<papers3://publication/uuid/B1097137-BA46-4A56-B6ED-65D450BCD547>>.
- LI, Zhijin et al. The Role of Long Noncoding RNAs in Gene Expression Regulation. In: **Gene Expression Profiling in Cancer**. IntechOpen, 2019.
- LIU, Yongchao et al. lncRNA GAS5 enhances G1 cell cycle arrest via binding to YBX1 to regulate p21 expression in stomach cancer. **Scientific reports**, v. 5, p. 10159, 2015.
- LONGHESE, Maria Pia et al. DNA damage checkpoint in budding yeast. **EMBO Journal**, v. 17, n. 19, p. 5525–5528, 1998.

LU, Lu et al. Epigenetic silencing of miR-218 by the lncRNA CCAT1, acting via BMI1, promotes an altered cell cycle transition in the malignant transformation of HBE cells induced by cigarette smoke extract. **Toxicology and applied pharmacology**, v. 304, p. 30-41, 2016.

MERCER, Tim R.; DINGER, Marcel E.; MATTICK, John S. Long non-coding RNAs: Insights into functions. **Nature Reviews Genetics**, v. 10, n. 3, p. 155–159, 2009.

MARQUES, L. F. **Busca e análise de lncRNAs (long non-coding RNAs) importantes para a tolerância ao etanol em *Saccharomyces cerevisiae***. 2019. Dissertação (Mestrado em Biotecnologia) – Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu.

MARÍN-BÉJAR, Oskar et al. The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. **Genome biology**, v. 18, n. 1, p. 202, 2017.

MELO, Justine; TOCZYSKI, David. A unified view of the DNA-damage checkpoint. *Current opinion in cell biology*, v. 14, n. 2, p. 237-245, 2002.

MICHAELIS, Christine; CIOSK, Rafal; NASMYTH, Kim. Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. **Cell**, v. 91, n. 1, p. 35–45, 1997.

MIMURA, Satoru et al. Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. **Nature**, v. 431, n. 7012, p. 1118–1123, 2004.

MOHD AZHAR, Siti Hajar et al. Yeasts in sustainable bioethanol production: A review. **Biochemistry and Biophysics Reports**, v. 10, n. February, p. 52–61, 2017.

MORAVCEVIC, Katarina et al. Kinase associated-1 domains drive MARK/PAR1 kinases to membrane targets by binding acidic phospholipids. **Cell**, v. 143, n. 6, p. 966-977, 2010.

NADAL-RIBELLES, Mariona et al. Control of Cdc28 CDK1 by a stress-induced lncRNA Mariona. v. 40, n. 6, p. 1301–1315, 2012.

NEECKE, Holger; LUCCHINI, Giovanna; LONGHESE, Maria Pia. Cell cycle progression in the presence of irreparable DNA damage is controlled by a Mec1- and Rad53-dependent checkpoint in budding yeast. **EMBO Journal**, v. 18, n. 16, p. 4485–4497, 1999.

NELSON, David L; COX, Michael M. **Princípios de bioquímica de Lehninger**. 6. ed. [S.l.]: artmed, 2014.

NGUYEN, Van Q.; CO, Carl; LI, Joachim J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. **Nature**, v. 411, n. 6841, p. 1068–1073, 2001.

NIPHADKAR, Shreyas; BAGADE, Praful; AHMED, Shadab. Bioethanol production: insight into past, present and future perspectives. **Biofuels**, v. 9, n. 2, p. 229–238, 2018. Disponível em: <<https://doi.org/10.1080/17597269.2017.1334338>>.

- NOBLE, D. The future: putting Humpty-Dumpty together again. **Biochemical Society Transactions**, v. 31, n. 1, p. 156–158, 2003.
- NOUGAREDE, R. et al. Hierarchy of S-Phase-Promoting Factors: Yeast Dbf4-Cdc7 Kinase Requires Prior S-Phase Cyclin-Dependent Kinase Activation. **Molecular and Cellular Biology**, v. 20, n. 11, p. 3795–3806, 2000.
- PFLEGER, Cathie M.; KIRSCHNER, Marc W. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. **Genes & development**, v. 14, n. 6, p. 655-665, 2000.
- PRAMILA, Tata et al. Conserved homeodomain proteins interact with MADS box protein Mcm1 to restrict ECB-dependent transcription to the M/G1 phase of the cell cycle. **Genes and Development**, v. 16, n. 23, p. 3034–3045, 2002.
- RASS-HANSEN, Jeppe et al. Perspective Bioethanol: fuel or feedstock. **Journal of Chemical Technology and Biotechnology**, v. 82, n. October 2006, p. 329–333, 2007.
- ROCK, Jeremy M.; AMON, Angelika. The FEAR network. **Current Biology**, v. 19, n. 23, p. 1–10, 2009.
- RUBINSTEIN, Amir et al. The effective application of a discrete transition model to explore cell-cycle regulation in yeast. **BMC Research Notes**, v. 6, n. 1, 2013.
- RUDNER, Adam D.; MURRAY, Andrew W. The spindle assembly checkpoint. **Current Opinion in Cell Biology**, v. 8, n. 6, p. 773–780, 1996. Disponível em: <<http://dx.doi.org/10.1016/j.cub.2012.10.006>>.
- RUPE, Ivan. Checking cell size in yeast. v. 18, n. 9, p. 1–7, 2002. Disponível em: <<papers2://publication/uuid/A372A6F4-A3CE-4178-8589-B39EFD800165>>.
- SAHIN, Özgür et al. Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. **BMC Systems Biology**, v. 3, p. 1–20, 2009.
- SALLES-FILHO, Sergio Luiz Monteiro et al. Perspectives for the Brazilian bioethanol sector: The innovation driver. **Energy Policy**, v. 108, n. May, p. 70–77, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.enpol.2017.05.037>>.
- SANTOS, Alberto; WERNERSSON, Rasmus; JENSEN, Lars Juhl. Cyclebase 3.0: A multi-organism database on cell-cycle regulation and phenotypes. **Nucleic Acids Research**, v. 43, n. D1, p. D1140–D1144, 2015.
- SHEU, Yi Jun; STILLMAN, Bruce. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. **Nature**, v. 463, n. 7277, p. 113–117, 2010. Disponível em: <<http://dx.doi.org/10.1038/nature08647>>.

- SHOU, Wenying; DESHAIES, Raymond J. Multiple telophase arrest bypassed (tab) mutants alleviate the essential requirement for Cdc15 in exit from mitosis in *S. cerevisiae*. **BMC Genetics**, v. 3, p. 1–12, 2002.
- SOLÉ, Carme et al. A novel role for lncRNAs in cell cycle control during stress adaptation. **Current Genetics**, v. 61, n. 3, p. 299–308, 2015. Disponível em: <<http://dx.doi.org/10.1007/s00294-014-0453-y>>.
- SRIVASTAVA, Neha et al. A review on fuel ethanol production from lignocellulosic biomass. **International Journal of Green Energy**, v. 12, n. 9, p. 949–960, 2015. Disponível em: <<http://dx.doi.org/10.1080/15435075.2014.890104>>.
- STANLEY, D. et al. The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. **Journal of Applied Microbiology**, v. 109, n. 1, p. 13–24, 2010.
- STEMMANN, Olaf; LECHNER, J. The *Saccharomyces cerevisiae* kinetochore contains a cyclin-CDK complexing homologue, as identified by in vitro reconstitution. **The EMBO journal**, v. 15, n. 14, p. 3611-3620, 1996.
- STRAIGHT, Aaron F. Cell cycle: checkpoint proteins and kinetochores. **Current biology**, v. 7, n. 10, p. R613-R616, 1997.
- SUN, Shuyan. Meta-analysis of Cohen's kappa. **Health Services and Outcomes Research Methodology**, v. 11, n. 3-4, p. 145-163, 2011.
- TANAKA, Seiji; DIFFLEY, John F X. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. **Nature Cell Biology**, v. 4, n. 3, p. 198–207, 2002.
- TAKEHARA, Masaya et al. Analysis of mutant origin recognition complex with reduced ATPase activity in vivo and in vitro. **Biochemical Journal**, v. 413, n. 3, p. 535-543, 2008.
- TEIXEIRA, Miguel C. et al. Genome-wide identification of *Saccharomyces cerevisiae* genes required for maximal tolerance to ethanol. **Applied and Environmental Microbiology**, v. 75, n. 18, p. 5761–5772, 2009.
- TODD, Robert G.; HELIKAR, Tomáš. Ergodic Sets as Cell Phenotype of Budding Yeast Cell Cycle. **PLoS ONE**, v. 7, n. 10, 2012.
- TOMLIN, Claire J.; AXELROD, Jeffrey D. Biology by numbers: Mathematical modelling in developmental biology. **Nature Reviews Genetics**, v. 8, n. 5, p. 331–340, 2007.
- TRACY, Kirsten M. et al. Mitotically-associated lncRNA (MANCR) affects genomic stability and cell division in aggressive breast cancer. **Molecular Cancer Research**, v. 16, n. 4, p. 587-598, 2018.

TRIPATHI, Vidisha et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. **PLoS genetics**, v. 9, n. 3, 2013.

VERMA, R et al. Phosphorylation of Sic1p by G1 Cdk Required for Its Degradation and Entry into S Phase. **Science**, v. 278, n. 5337, p. 455–460, 1997.

WANG, Rui Sheng; SAADATPOUR, Assieh; ALBERT, Réka. Boolean modeling in systems biology: An overview of methodology and applications. **Physical Biology**, v. 9, n. 5, 2012.

WANG, Yanchang; HU, Fenghua; ELLEDGE, Stephen J. The Bfa1/Bub2 GAP complex comprises a universal checkpoint required to prevent mitotic exit. **Current Biology**, v. 10, n. 21, p. 1379–1382, 2000.

WANG, Kevin C.; CHANG, Howard Y. Molecular mechanisms of long noncoding RNAs. **Molecular cell**, v. 43, n. 6, p. 904-914, 2011.

WÄSCH, Ralph; CROSS, Frederick R. APC-dependent proteolysis of the mitotic cyclin Clb2 is essential for mitotic exit. **Nature**, v. 418, n. 6897, p. 556-562, 2002.

WEINERT, Ted A.; KISER, Gretchen L.; HARTWELL, L. H. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. **Genes and Development**, v. 8, n. 6, p. 652–665, 1994.

WEI, Yanping et al. LncRNA UCA1-miR-507-FOXM1 axis is involved in cell proliferation, invasion and G0/G1 cell cycle arrest in melanoma. **Medical Oncology**, v. 33, n. 8, p. 88, 2016.

WOLF, I. R. **Identificação de assinaturas sistêmicas associadas à tolerância ao etanol em linhagens de *Saccharomyces cerevisiae***. 2019. Dissertação (Doutorado em Biotecnologia) – Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu.

YAMASHITA, Akira; SHICHINO, Yuichi; YAMAMOTO, Masayuki. The long non-coding RNA world in yeasts. **Biochimica et Biophysica Acta - Gene Regulatory Mechanisms**, v. 1859, n. 1, p. 147–154, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.bbagr.2015.08.003>>.

YANG, Hui; REN, Qun; ZHANG, Zhaojie. Cleavage of Mcd1 by Caspase-like Protease Esp1 Promotes Apoptosis in Budding Yeast. **Molecular biology of the cell**, v. 19, n. 1, p. 308–317, 2008.

ZHU, Juan Juan et al. Function of lncRNAs and approaches to lncRNA-protein interactions. **Science China Life Sciences**, v. 56, n. 10, p. 876–885, 2013.

Apêndice 1 – Análise quantitativa

Em paralelo com o desenvolvimento do modelo descrito no projeto, foi realizada uma modelagem quantitativa de uma pequena rede de interação entre proteínas, a fim de elaborar hipóteses acerca de um resultado encontrado pelo grupo de pesquisa.

Métodos

A partir dos dados obtidos no projeto do grupo de pesquisa, uma sub-rede da BMA64-1A foi descoberta (Figura 21), envolvendo uma série de proteínas que podem estar relacionadas com a enzima álcool desidrogenase (Adh2p), a qual catalisa a conversão de etanol para acetaldeído. Na rede, há também um lncRNA (transcr_20548) que interage fisicamente com um fator de transcrição de Adh2p (Cat8p). Os dados indicaram que há uma redução drástica no Fold-change de Adh2p, e as causas dessa diminuição ainda não estão elucidadas. Portanto, com o intuito de estudar mais detalhadamente essa rede e avaliar se o lncRNA possui algum papel na inibição de Adh2p, um modelo matemático quantitativo foi desenvolvido para esse sistema.

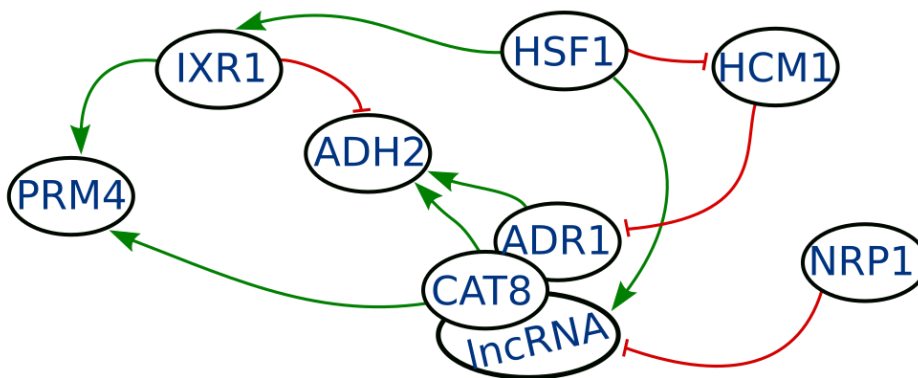


Figura 21: Proteínas da sub-rede encontrada na linhagem BMA64-1A. Arestas verdes representam interações de ativação e as vermelhas representam inibição. Os nós sobrepostos indicam associações de proteínas, nesse caso, Cat8 se associa com o lncRNA e Adr1 se associa com Cat8.

Como o número de proteínas é pequeno, o método de modelagem escolhido foi o quantitativo (utilizando ODEs). Para a criação das equações e as escolhas dos parâmetros

algumas suposições foram feitas: 1- como explicado na introdução, algumas proteínas precisam ser fosforiladas ou desfosforiladas para serem ativas, porém, como os elementos da rede modelada não são bem descritos na literatura, não foi possível adicionar esse nível de complexidade ao modelo; portanto, todas as proteínas foram consideradas como ativas e as interações entre elas se resumem na capacidade de aumentarem ou diminuírem a síntese de seus alvos; 2- a falta de informação acerca das proteínas da rede torna impossível escolher os valores dos parâmetros (degradação, síntese, inibição e associação) com perfeição, portanto, para evitar o viés, todos os parâmetros receberam o mesmo valor (dependendo do tipo). Dessa forma, embora o modelo não represente com exatidão o comportamento real do sistema, ainda é possível analisar a influência de cada elemento baseado apenas em como as interações entre as proteínas estão dispostas no sistema. Porém, para aumentar ao máximo a fidelidade do modelo com o comportamento real, uma revisão bibliográfica foi feita para cada proteína da rede.

De acordo com Kraikviski et al. (2015), taxas de degradação e síntese ocorrem na ordem de aproximadamente 0.1 min^{-1} . Já taxas de associação física entre proteínas ocorrem na ordem de aproximadamente 50 min^{-1} . Ademais, existem dois tipos de taxas de síntese: 1- síntese basal, de menor intensidade, geralmente menores que as taxas de degradação e; 2- sínteses provenientes de fatores de transcrição, de alta intensidade, geralmente maiores que as taxas de degradação (CHEN et al., 2004).

As equações foram construídas levando em consideração todas as informações supracitadas e o modelo foi resolvido no software COPASI. Os valores iniciais das variáveis do modelo foram 0 para o lncRNA, Ixr1, Hsf1 e Prm4, as outras variáveis iniciaram com o valor 1. Essa escolha foi feita com base nos dados de *time-course* da pesquisa do grupo e, se o transcrito iniciasse com alto fold-change e terminasse com um

valor mais baixo, a proteína relacionada a ele recebia o valor inicial 1; caso contrario, recebia 0 (dados não mostrados). Todas as equações e parâmetros estão disponíveis na Tabela Suplementar 4 e 5.

Para avaliar a influência do lncRNA na disponibilidade de Adh2, três simulações foram realizadas. A primeira foi do modelo sem nenhuma perturbação, a segunda foi uma simulação com a proteína Ixr1 (inibidor de Adh2) ausente no sistema (*knockdown*) e a última foi a simulação a ausência do lncRNA. O lncRNA foi considerado apenas como inibidor da sua proteína alvo.

Resultados

As simulações representam a variação de unidades de proteína durante um determinado período de tempo. Embora o modelo seja simples, há informações o suficiente para estudar os efeitos do lncRNA na produção de Adh2p. Os resultados das simulações estão representados na Figura 21.

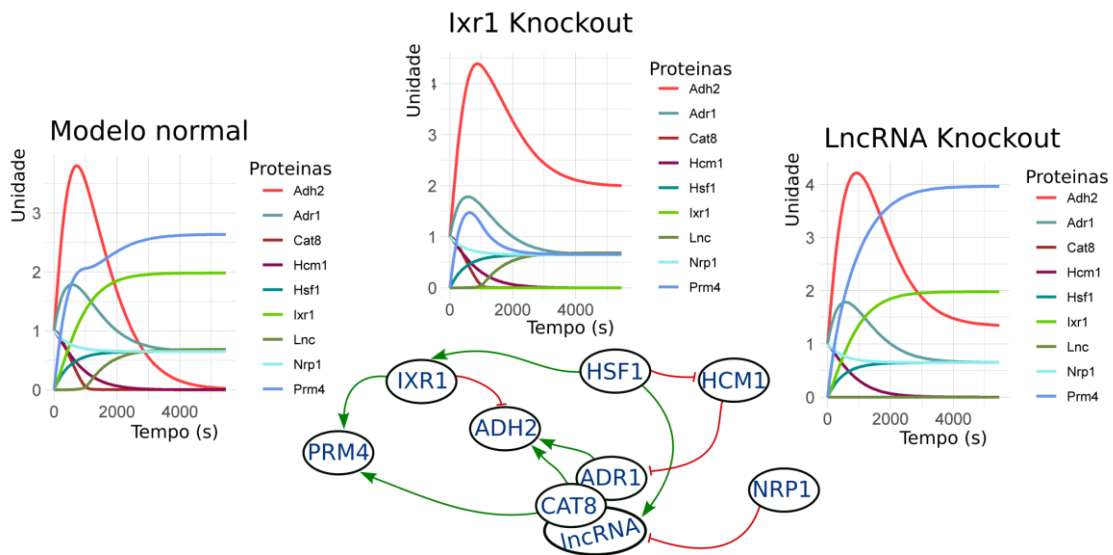


Figura 21: Simulações do modelo para cada situação. A rede foi adicionada na figura para facilitar a visualização do comportamento de cada proteína do modelo.

Os resultados indicam que há um aumento na disponibilidade de Adh2p tanto na ausência do Ixr1p quanto do lncRNA. O maior aumento foi observado na deleção de Ixr1p, no entanto, vale ressaltar que ele é muito mais abundante que o lncRNA (mais de duas vezes). Sendo assim, como a deleção do lncRNA causou um aumento na disponibilidade de Adh2p de apenas 0,5 unidades a menos que Ixr1p, pode-se inferir que ele é um inibidor mais forte de Adh2p.

Essa força de inibição observada é devido à taxa de associação empregada para a formação do complexo Cat8-lncRNA. Como Cat8p é um fator de transcrição para Adh2 (logo, aumenta significativamente sua síntese), a associação com o lncRNA reduz drasticamente a disponibilidade de Cat8p no sistema, o que causa a diminuição da síntese de Adh2p. Portanto, existe a possibilidade do lncRNA ser a principal causa da inibição da síntese de Adh2p. No entanto, podemos inferir também que o sinergismo entre o lncRNA e Ixr1p é a melhor maneira de atingir a redução máxima na disponibilidade da enzima.

Não foi possível modelar o lncRNA como ativador de Cat8p devido à falta de informação sobre essa proteína, pois não há como saber por quais mecanismos o lncRNA o ativaria. Enquanto que a modelagem como inibidor é muito mais simples, pois se sabe que lncRNAs se associam com proteínas e modelar essas associações é muito mais simples do que outros tipos de reações.

ARQUIVO SUPLEMENTAR DA DISSERTAÇÃO DE MESTRADO

Modelagem do ciclo celular e influência dos lncRNAs em Saccharomyces cerevisiae expostas a altas concentrações de etanol.

Lucas Cardoso Lázari

Todos os mutantes testados nesta dissertação (Tabela S1) foram retirados de três trabalhos desenvolvidos por Chen et al. (2000), Chen et al. (2014) e Kraikiviski et al. (2015). São três modelos quantitativos do ciclo celular, que não foram utilizados para a execução da dissertação, apenas a lista de mutantes foi utilizada. As mutações que não foram retiradas deste trabalho estão referenciadas à parte.

Testes para a validação do modelo lógico		
Perturbação	Predição do modelo	Comportamento Real
Cdh1-Δ Cdc6-Δ2-49	Viável	Viável
APCC-Δ GAL_Clb2	Arrest em M	Inviável
Cdc20-ts Bub2-Δ	Arrest em M	Arrest na Metáfase
Cdc20-ts GAL-Esp1	Arrest em M	Arrest na Telófase
Cdc20-ts Mad2-Δ	Arrest em M	Arrest na Metáfase
Cdc20-Δ	Arrest em M	Arrest na Metáfase
Cdc20-Δ (APC-A) Cdh1-Δ	Arrest em M	Letal
Cdc20-Δ (APC-A) Cdh1-Δ Sic1 MC	Viável	Viável
Cdc20-Δ (APC-A) GAL-Clb2	Arrest em M	Inviável
Cdc20-Δ (APC-A) Sic1-Δ	Viável	Viável
Cdc20-Δ Pds1-Δ	Viável	Inviável
Cdh1-Δ Swi5-Δ	Arrest em M	Arrest na Telófase
Bub2-Δ	Viável	Viável
Bub2-Δ - nocodazole	Viável	Viável
Bub2-Δ Mad2-Δ - nocodazole	Viável	Viável
Bfa1/Bub2-Δ Pds1-Δ - nocodazole	Viável	Viável
Cdc14-ts Cdh1-Δ	Arrest em M	Arrest na Telófase
Cdc14-ts GAL-Sic1	Arrest em M	Arrest na Telófase
Cdc14-Δ	Arrest em M	Arrest na Telófase
Cdc14-ts GAL_Cln2	Arrest em S	Inviável

Cdc14-ts Sic1-Δ	Arrest em M	Arrest na Telófase
Cdc15-ts Tem1 MC	Arrest em M	Arrest na Telófase
Cdc15-Δ	Arrest em M	Inviável
Cdc20-ts Net1-ts	Viável	Inviável
Cdc20-Δ Pds1-Δ Clb5-Δ	Viável	Viável
Cdc20-ts Bub2-Δ	Arrest em G2	Arrest na Metáfase
Cdc20-ts Mad2-Δ	Arrest em G2	Arrest na Metáfase
Cdc45-Δ (ZOU et al. 1997)	Arrest em S	Inviável
Cdc5-Δ	Arrest em M	Inviável
Cdc6-Δ	Arrest em S	Arrest em S
Cdc6-Δ ARS-O (HOGAN; KOSHLAND, 1992)	Viável	Viável
Cdc6-Δ2-49	Viável	Viável
Cdc6-Δ2-49 Cdh1-Δ Sic1- Δ	Arrest em M	Arrest na Telófase
Cdc6-Δ2-49 Cdh1-Δ Sic1- Δ GAL-Cdc20	Arrest em S depois de um ciclo	Viável
GAL-Cdh1	Arrest em G2	Arrest em G2
Cdh1-Δ	Viável	Viável
Clb2-Δ	Arrest em G2	Arrest em G2
Clb2-dbΔ	Arrest em M	Arrest na Telófase
Clb2-dbΔ Clb5-Δ	Arrest em M	Arrest na Telófase
Clb2-Δ Pds1-Δ	Arrest em G2	Inviável
Clb5-Δ	Viável	Viável
Clb5-Δ Cdc20-Δ	Arrest em M	Arrest na Metáfase
Clb5-dbΔ	Viável	Viável
Clb5-dbΔ Pds1-Δ	Viável	Viável
Clb5-dbΔ Pds1-Δ Cdc20-Δ	Viável	Viável
Clb5-dbΔ Sic1-Δ	Viável	Inviável
Cln1/2-Δ	Viável	Viável
Cln1/2-Δ Clb5/6-Δ	Arrest em G1	Arrest em G1
Cln1/2-Δ Cdh1-Δ GAL- Sic1	Viável	Viável
Cln1/2-Δ Cdh1-Δ	Viável	Viável
Cln1/2-Δ Cln3-Δ	Arrest em G1	Inviável
Cln1/2-Δ Cln3-Δ APC-ts	Arrest em G1	Inviável
Cln1/2-Δ Cln3-Δ Cdh1-Δ	Arrest em G1	Inviável
Cln1/2-Δ Cln3-Δ GAL- Clb5/6	Viável	Viável

Cln1/2-Δ Cln3-Δ GAL-Cln2*	Arrest em G1	Viável
Cln1/2-Δ Cln3-Δ Mc-Clb5	Viável	Viável
Cln1/2-Δ Cln3-Δ Sic1-Δ	Viável	Viável
Cln1/2-Δ Cln3-Δ Whi5-Δ	Arrest em G1	Inviável
Cln1/2-Δ GAL_Sic1	Arrest em S	Inviável
Cln1/2-Δ GAL-Cln2 GAL-Sic1 Cdh1-Δ	Viável	Viável
Cln1/2-Δ GAL-Sic1 GAL-Cln1/2	Viável	Viável
Cln1/2-Δ Sic1-Δ	Viável	Viável
Cln3-Δ	Viável	Viável
Cln3-Δ Cln1/2-Δ GAL-Cln3 Cdc14-ts	Arrest em S	Inviável
Cln3-Δ Whi5-Δ	Viável	Viável
Cohesin-Δ (Mcd1p) (GUACCI <i>et al.</i> 1997)	Arrest in G2	Inviável
Dbf4-Δ (KITADA <i>et al.</i> , 1992)	Arrest em S	Inviável
Esp1-Δ	Arrest in G2	Inviável
GAL-Clb2 Swi5-Δ	Arrest em M	Arrest na Telófase
GAL-Clb5	Viável	Viável
GAL-Sic1	Viável	Viável
GAL-Sic1-db-Δ	Arrest em G1	Arrest em G1
GAL-Cdc14	Arrest em G2	Inviável
GAL-Clb2 Cdh1-Δ	Arrest em M	Arrest na Telófase
GAL-Clb2 Sic1-Δ	Arrest em M	Arrest na Telófase
GAL-Clb5 Cdh1-Δ	Viável	Viável
GAL-Clb5 Sic1-Δ	Arrest em G1	Inviável
GAL-Cln3	Viável	Viável
GAL-Tem1	Viável	Viável
Mad2-Δ	Viável	Viável
Mad2-Δ - nocodazole	Viável	Viável
Mad2-Δ GAL_Tem1 - nocodazole	Viável	Viável
Mad2-Δ Pds1-Δ - nocodazole	Viável	Viável
Mec1-Δ	Viável	Viável
Mob1-Δ (LUCA; WINEY, 1998)	Arrest em M	Inviável

GAL-Mps1 (HARDWICK et al., 1996)	Viável	Viável
Mps1-Δ (SCHUTZ ;WINEY, 1998)	Viável	Viável
GAL-Net1	Arrest em M	Arrest na Telófase
Net1-Δ	Viável	Viável
Net1-Δ - nocodazole	Viável	Viável
ORC-Δ (GIBSON et al., 2006)	Arrest in G1	Inviável
Pds1-Δ	Viável	Viável
Pds1-Δ - nocodazole	Viável	Arrest na Telófase
PP2A/Cdc55-Δ (HEALY et al., 1991)	Viável	Viável
Rad53-Δ (NEECKE et al., 1999)	Viável	Viável
SBF-MBF-Δ	Inviável	Arrest em G1
SBF-Δ	Inviável	Inviável
SBF-Δ Sic1-Δ	Inviável	Inviável
Sic1-Δ	Viável	Viável
Sic1-Δ Cdc6-Δ2-49	Viável	Viável
Sic1-Δ Cdh1-Δ	Arrest em M	Inviável
Sic1-Δ Cdh1-Δ GAL-Cdc20	Arrest em S depois de um ciclo	Viável
Slk19-Δ (RICHMOND et al., 2013)	Viável	Viável
GAL-Swe1	Arrest em G2	Arrest em G2
Swi5-Δ	Viável	Viável
TAB6-1 Clb1	Viável	Viável
TAB6-1 Clb5-Δ	Viável	Arrest em G1
Tah11-Δ (JACOBSON et al., 2001)	Arrest em S	Inviável
Whi5-Δ	Viável	Viável
α-factor	Arrest em G1	Arrest em G1

Tabela 1S: Todas as mutações testadas para validar o modelo.

Pelo software GINsim, as equações foram escritas para cada nó do sistema, abaixo (Tabela S2) estão todas as equações desenvolvidas para o modelo. Para um determinado nó assumir um valor, deve haver uma equação que indique as condições para este valor ser alcançado. Os símbolos “!”, “&” e “|” são os operadores Booleanos NOT, AND e OR (inclusivo) respectivamente. As linhas que contém “*” recebem um maior detalhamento abaixo.

Nó	Equação
MASS	0 -> MITOSIS_EXIT:2
	1 -> !MASS & !MITOSIS_EXIT
	2 -> MASS:1 MASS:2
Mcm1	0 -> !MASS
	2 -> MASS:2
ECB	0 -> !Mcm1
	1 -> Mcm1:1
	2 -> Mcm1:2 Mcm1:3
Cln3	0 -> !ECB Far1
	1 -> ECB:1
	2 -> ECB:2
	3 -> ECB:3
MBF* e SBF*	0 -> (Whi5:1 & !Cln3 & !Cln1_2 & !Clb5_6) (Whi5:3 & Cln3:1 & !Cln1_2 & !Clb5_6) (Whi5:3 & Cln3:1 & Cln1_2:1 & Clb5_6:1) (Whi5:3 & !Cln3 & Cln1_2:1 & Clb5_6:1) (Whi5:2 & !Cln3 & !Cln1_2 & !Clb5_6) (Whi5:3 & !Cln3 & !Cln1_2 & !Clb5_6) (Whi5:3 & Clb5_6:1 & !Cln1_2 & !Cln3) (Whi5:3 & Clb5_6:1 & Cln3:1 & !Cln1_2) (Whi5:3 & Clb5_6:1 & Cln1_2:1 & !Cln3) (Whi5:3 & !Cln3 & Cln1_2:1 & !Clb5_6)
	1 -> (Cln1_2:1 & !Whi5 & Cln3:1) (Cln1_2:2 & Whi5:2 & Cln3:1) (Cln1_2:1 & !Whi5 & !Cln3) (Cln1_2:2 & Whi5:2 & !Cln3) (Cln1_2:1 & Whi5:2 & Cln3:2) (!Cln1_2 & !Whi5 & Cln3:1) (!Cln1_2 & Whi5:2 & Cln3:2) (Cln1_2:2 & Whi5:3 & Cln3:1) (Cln1_2:2 & Whi5:3 & Cln3:2) (Cln1_2:2 & Whi5:3 & Cln3:2) (Cln1_2:1 & Whi5:3 & Cln3:2) (Cln1_2:2 & Whi5:3 & Cln3:2) (Cln1_2:2 & Whi5:3 & !Cln3) (!Cln1_2 & Whi5:3 & Cln3:2) (Clb5_6:1 & !Cln1_2 & !Cln3)
	2 -> (Cln1_2:2 & !Whi5 & !Cln3) (Cln1_2:2 & !Whi5 & Cln3:1) (Cln1_2:2 & !Whi5 & Cln3:2) (!Cln1_2 & !Whi5 & Cln3:2) (Cln1_2:1 & !Whi5 & Cln3:2) (Cln1_2:3 & Whi5:2 & !Cln3) (Cln1_2:3 & Whi5:2 & Cln3:1) (Cln1_2:3 & Whi5:2 & Cln3:2) (Cln1_2:3 & Whi5:2 & Cln3:3) (Cln1_2:1 & Whi5:2 & Cln3:3) (Cln1_2:2 & Whi5:2 & Cln3:3) (!Cln1_2 & Whi5:2 & Cln3:3) (Clb5_6:2 & !Cln1_2 & !Cln3)
	3 -> (Cln3:3 & !Whi5 & !Cln1_2) (Cln3:3 & !Whi5 & Cln1_2:1) (Cln3:3 & !Whi5 & Cln1_2:2) (Cln3:3 & !Whi5 & Cln1_2:3) (!Cln3 & !Whi5 & Cln1_2:3) (Cln3:1 & !Whi5 & Cln1_2:3) (Cln3:2 & !Whi5 & Cln1_2:3) (Clb5_6:3 & !Cln1_2 & !Cln3)
Whi5	0 -> Cln3:2 Cln1_2:2
	2 -> (!Cln3 & !Cln1_2) (Cln1_2:1 & Cln3:1)

Cln1_2	<p>0 -> Far1:1 (SCF_Grr1 & !SCB & !ECB) (SCF_Grr1:3 & SCB & ECB) (SCF_Grr1:2 & SCB:2 & ECB:2) (SCF_Grr1:2 & SCB:2 & ECB:1) (SCF_Grr1:2 & SCB:1 & ECB:2) (SCF_Grr1:1 & SCB:1 & ECB:1) SCF_Grr1:3</p> <p>1 -> (ECB:1 & SCB:1 & !SCF_Grr1) (ECB:1 & !SCB & !SCF_Grr1) (SCB:1 & !ECB & !SCF_Grr1) (ECB:2 & SCB:2 & SCF_Grr1:1) (ECB:2 & SCB:1 & SCF_Grr1:1) (SCB:2 & ECB:1 & SCF_Grr1:1) (ECB:2 & !SCB & SCF_Grr1:1) (SCB:2 & !ECB & SCF_Grr1:1) (ECB:3 & SCB:1 & SCF_Grr1:2) (ECB:3 & SCB:2 & SCF_Grr1:2) (ECB:3 & SCB:3 & SCF_Grr1:2) (ECB:1 & SCB:3 & SCF_Grr1:2) (ECB:2 & SCB:3 & SCF_Grr1:2)</p> <p>2 -> (ECB:2 & SCB:2 & !SCF_Grr1) (ECB:2 & !SCB & !SCF_Grr1) (SCB:2 & !ECB & !SCF_Grr1) (ECB:3 & SCB:3 & SCF_Grr1:1) (ECB:3 & SCB:1 & SCF_Grr1:1) (ECB:3 & SCB:2 & SCF_Grr1:1) (ECB:1 & SCB:3 & SCF_Grr1:1) (ECB:2 & SCB:3 & SCF_Grr1:1) (ECB:3 & !SCB & SCF_Grr1:1) (!ECB & SCB:3 & SCF_Grr1:1) (ECB:1 & SCB:2 & !SCF_Grr1) (ECB:2 & SCB:1 & !SCF_Grr1)</p> <p>3 -> (ECB:3 & SCB:3 & !SCF_Grr1) (ECB:3 & !SCB & !SCF_Grr1) (SCB:3 & !ECB & !SCF_Grr1) (ECB:3 & SCB:1 & !SCF_Grr1) (ECB:3 & SCB:2 & !SCF_Grr1) (SCB:3 & ECB:1 & !SCF_Grr1) (SCB:3 & ECB:2 & !SCF_Grr1)</p>
SCB	<p>1 -> SBF:1 (MBF:1 & SBF:1)</p> <p>2 -> SBF:2 SBF:3 (MBF:3 & SBF:3)</p>
Fus3	<p>0 -> !Mating</p> <p>1 -> Mating</p>
Far1	<p>0 -> !Fus3 SCF_Cdc4</p> <p>1 -> Fus3 & !SCF_Cdc4</p>
MCB	<p>1 -> (MBF:1 & SBF:1) SBF:1</p> <p>2 -> SBF:2 SBF:3 (MBF:3 & SBF:3)</p>
S_proteins	<p>1 -> SCB:1 & MCB:1</p> <p>2 -> (SCB:2 & MCB:2) (SCB:3 & MCB:3)</p>
SCF_Cdc4	<p>0 -> !S_proteins</p> <p>1 -> S_proteins:1</p> <p>2 -> S_proteins:2</p> <p>3 -> S_proteins:3</p>
SCF_Grr1	<p>0 -> !S_proteins</p> <p>1 -> S_proteins:1</p> <p>2 -> S_proteins:2</p> <p>3 -> S_proteins:3</p>
BUD	<p>0 -> MITOSIS_EXIT:2</p> <p>1 -> Cln1_2 (Clb5_6 & !Cln1_2) BUD:1</p>

Sic1	<p>1 -> (!Swi5 & !Cdc14 & !Cln1_2 & !SCF_Cdc4) (Swi5:1 & !Cln1_2 & !Cdc14) (Swi5:1 & Cln1_2 & Cdc14:3) (Swi5:1 & (Cln1_2:1 Cln1_2:2) & Cdc14:2) (Swi5:1 & Cln1_2:1 & !Cdc14:1) (SCF_Cdc4:1 & Swi5:1 & Cln1_2 & !Cdc14) (SCF_Cdc4:2 & Swi5:2 & Cln1_2 & !Cdc14) (SCF_Cdc4:3 & Swi5 & Cln1_2 & !Cdc14) (SCF_Cdc4 & (!Cln1_2 Cln1_2) & !Swi5 & !Cdc14)</p> <p>2 -> (Swi5:2 & !Cln1_2 & !Cdc14 & !SCF_Cdc4) (Swi5:2 & Cln1_2 & Cdc14:3 & SCF_Cdc4) (Swi5:2 & Cln1_2:1 & Cdc14:1 & (!SCF_Cdc4 SCF_Cdc4:1)) (Swi5:2 & Cln1_2:2 & Cdc14:2 & (!SCF_Cdc4 SCF_Cdc4:1 SCF_Cdc4:2)) (Swi5:3 & Cln1_2:1 & !Cdc14 & !SCF_Cdc4) (Swi5:3 & Cln1_2:2 & Cdc14:1 & (!SCF_Cdc4 SCF_Cdc4:1)) (Swi5:3 & Cln1_2:3 & Cdc14:2 & (!SCF_Cdc4 SCF_Cdc4:1 SCF_Cdc4:2)) (Swi5:2 & Cdc14 & !Cln1_2 & !SCF_Cdc4) (Swi5:2 & Cdc14:1 & !Cln1_2 & (SCF_Cdc4:1 !SCF_Cdc4)) (Swi5:2 & Cdc14:2 & !Cln1_2 & (!SCF_Cdc4 SCF_Cdc4:1 SCF_Cdc4:2)) (Swi5:2 & Cdc14:3 & !Cln1_2 & SCF_Cdc4) (Swi5:2 & Cdc14:2 & (Cln1_2:1 Cln1_2:2 !Cln1_2))</p> <p>3 -> (Swi5:3 & Cdc14:3 & (Cln1_2 !Cln1_2) & (SCF_Cdc4 !SCF_Cdc4)) (Swi5:3 & Cdc14 & !Cln1_2 & (!SCF_Cdc4 SCF_Cdc4)) (Swi5:3 & !Cdc14 & !Cln1_2 & !SCF_Cdc4) (Swi5:3 & Cdc14:2 & (Cln1_2:1 Cln1_2:2 !Cln1_2) & (!SCF_Cdc4 SCF_Cdc4:1))</p>
Clb5_6	<p>0 -> Sic1:3 APCC_Cdc20:3 (Sic1:3 & APCC_Cdc20:3) (Sic1:2 & APCC_Cdc20:2) (Sic1 & APCC_Cdc20 & !MCB) (MCB:1 & APCC_Cdc20:1 & Sic1) (MCB:1 & APCC_Cdc20 & Sic1:1) (MCB:1 & APCC_Cdc20:2 & Sic1) (MCB:1 & APCC_Cdc20 & Sic1:2) (MCB:1 & APCC_Cdc20 & Sic1:3) (MCB:1 & APCC_Cdc20:3 & Sic1)</p> <p>1 -> (MCB:1 & !Sic1 & !APCC_Cdc20) (MCB:2 & Sic1:1 & !APCC_Cdc20) (MCB:2 & !Sic1 & APCC_Cdc20:1) (MCB:3 & Sic1:1 & APCC_Cdc20:1) (MCB:3 & Sic1:2 & !APCC_Cdc20) (MCB:3 & !Sic1 & APCC_Cdc20:2) (ECB:1 & !Sic1 & !APCC_Cdc20 & !MCB)</p> <p>2 -> (MCB:2 & !Sic1 & !APCC_Cdc20) (MCB:3 & Sic1:1 & !APCC_Cdc20) (MCB:3 & !Sic1 & APCC_Cdc20:1)</p> <p>3 -> MCB:3 & !Sic1 & !APCC_Cdc20</p>
APCC_Cdc20	<p>0 -> MCC:1</p> <p>1 -> G2_proteins:1 & !MCC</p> <p>2 -> G2_proteins:2 & !MCC</p> <p>3 -> G2_proteins:3 & !MCC</p>
Cdc45, Tah11,	0 -> !ECB

Dbf4_Cdc7 e ORC	1 -> ECB:1
	2 -> ECB:2
	3 -> ECB:3
MCM	1 -> ECB:1 & Dbf4_Cdc7:1
	2 -> ECB:2 & Dbf4_Cdc7:2
	3 -> ECB:3 & Dbf4_Cdc7:3
Cdc6	0 -> (SCF_Cdc4:3 & ECB) (Clb5_6:3 & ECB) (SCF_Cdc4:2 & Clb5_6:2)
pre_RC	1 -> (Cdc45:1 & Tah11:1 & Cdc6:1 & MCM:1 & ORC:1) (Cdc45:1 & Tah11 & Cdc6 & ORC & MCM) (Cdc45 & Tah11:1 & Cdc6 & ORC & MCM) (Cdc45 & Tah11 & Cdc6:1 & ORC & MCM) (Cdc45 & Tah11 & Cdc6 & ORC:1 & MCM) (Cdc45 & Tah11 & Cdc6 & ORC & MCM:1)
	2 -> (Cdc45:2 & Tah11:2 & Cdc6:2 & ORC:2 & MCM:2) (Cdc45:2 & (Tah11:3 Tah11:2) & (Cdc6:3 Cdc6:2) & (ORC:3 ORC:2) & (MCM:3 MCM:2)) ((Cdc45:3 Cdc45:2) & Tah11:2 & (Cdc6:3 Cdc6:2) & (ORC:3 ORC:2) & (MCM:3 MCM:2)) ((Cdc45:3 Cdc45:2) & (Tah11:3 Tah11:2) & Cdc6:2 & (ORC:3 ORC:2) & (MCM:3 MCM:2)) ((Cdc45:3 Cdc45:2) & (Tah11:3 Tah11:2) & (Cdc6:3 Cdc6:2) & ORC:2 & (MCM:3 MCM:2)) ((Cdc45:3 Cdc45:2) & (Tah11:3 Tah11:2) & (Cdc6:3 Cdc6:2) & (ORC:3 ORC:2) & MCM:2)
	3 -> Cdc45:3 & Tah11:3 & Cdc6:3 & ORC:3 & MCM:3
ARS	1 -> pre_RC:1
	2 -> pre_RC:2
	3 -> pre_RC:3
DNA_Replication	0 -> MITOSIS_EXIT:2
	1 -> ((Clb1_2:2 Clb1_2:3) & ARS) ((Clb5_6:2 Clb5_6:3) & ARS) DNA_Replication:1 (Clb1_2:1 & Clb5_6:1 & ARS)
Spindle	0 -> MITOSIS_EXIT:2
	1 -> Cohesin:1 & (!MITOSIS_EXIT MITOSIS_EXIT:1) & (!Spindle Spindle:1)
	2 -> (Spindle:1 & !Cohesin & (!MITOSIS_EXIT MITOSIS_EXIT:1)) (Spindle:2 & (MITOSIS_EXIT:1 !MITOSIS_EXIT) & (Cohesin !Cohesin))
Clb1_2*	0 -> !Cak1 !MCB Swe1:3 APCC_Cdc20:3 APCC_Cdh1:3 Sic1:3 (MCB:2 & Cak1 & APCC_Cdh1:2) (MCB:2 & Cak1 & ((Sic1:2 & APCC_Cdc20:1) (Sic1:2 & APCC_Cdc20:2))) (MCB:2 & Cak1 & Swe1:2) (MCB:1 & Cak1 & APCC_Cdh1:1) (MCB:1 & Cak1 & Swe1)

	<p>1 -> (MCB:1 & Cak1 & !APCC_Cdh1 & (!Sic1 (Sic1:1 & !APCC_Cdc20) (Sic1:2 & !APCC_Cdc20)) & !Swe1 & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:2 & Cak1 & APCC_Cdh1:1 & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:2 & Cak1 & Swe1:1 & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:3 & Cak1 & APCC_Cdh1:2 & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:3 & Cak1 & ((Sic1:2 & APCC_Cdc20:1) (Sic1:2 & APCC_Cdc20:2)) & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:3 & Cak1 & Swe1:2 & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:1 & Cak1 & !APCC_Cdh1 & ((Sic1:1 & APCC_Cdc20:1) (Sic1:1 & APCC_Cdc20:2)) & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2) & !Swe1)</p>
	<p>2 -> (MCB:2 & Cak1 & !APCC_Cdh1 & (!Sic1 (Sic1:1 & !APCC_Cdc20) (Sic1:2 & !APCC_Cdc20)) & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2) & !Swe1) (MCB:3 & Cak1 & APCC_Cdh1:1 & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:3 & Cak1 & Swe1:1 & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2)) (MCB:2 & Cak1 & !APCC_Cdh1 & ((Sic1:1 & APCC_Cdc20:1) (Sic1:1 & APCC_Cdc20:2)) & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2) & !Swe1)</p>
	<p>3 -> (MCB:3 & Cak1 & !APCC_Cdh1 & (!Sic1 (Sic1:1 & !APCC_Cdc20) (Sic1:2 & !APCC_Cdc20)) & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2) & !Swe1) (MCB:3 & Cak1 & !APCC_Cdh1 & ((Sic1:1 & APCC_Cdc20:1) (Sic1:1 & APCC_Cdc20:2)) & (!APCC_Cdc20 APCC_Cdc20:1 APCC_Cdc20:2) & !Swe1)</p>
Mih1*	<p>0 -> !S_proteins</p> <p>1 -> S_proteins:1</p> <p>2 -> S_proteins:2</p> <p>3 -> S_proteins:3</p>
Cak1*	<p>1 -> S_proteins</p>
Swi5	<p>0 -> (G2_proteins:1 & Cdc14:1 & Clb1_2:2) (G2_proteins:1 & Cdc14:1 & Clb1_2:3) (G2_proteins:1 & Cdc14:2 & Clb1_2:3) (G2_proteins:2 & Cdc14:1 & Clb1_2:3) (G2_proteins:2 & !Cdc14 & Clb1_2:2) (G2_proteins:1 & !Cdc14 & Clb1_2:3) (G2_proteins:3 & !Cdc14 & Clb1_2:3) (G2_proteins:1 & !Cdc14 & Clb1_2:2) (G2_proteins:1 & !Cdc14 & Clb1_2:3)</p>

	<p>1 -> (G2_proteins:1 & !Cdc14 & !Clb1_2) (G2_proteins:1 & Cdc14:1 & Clb1_2:1) (G2_proteins:1 & Cdc14:2 & Clb1_2:2) (G2_proteins:1 & Cdc14:3 & Clb1_2) (G2_proteins:2 & Cdc14:1 & Clb1_2:2) (G2_proteins:2 & Cdc14:2 & Clb1_2:3) (G2_proteins:3 & Cdc14:1 & Clb1_2:3) (G2_proteins:2 & !Cdc14 & Clb1_2:1) (G2_proteins:1 & Cdc14 & !Clb1_2) (G2_proteins:1 & Cdc14:2 & Clb1_2:1)</p>
	<p>2 -> (G2_proteins:2 & Cdc14 & !Clb1_2) (G2_proteins:2 & Cdc14:1 & Clb1_2:1) (G2_proteins:2 & Cdc14:2 & Clb1_2:1) (G2_proteins:2 & Cdc14:2 & Clb1_2:2) (G2_proteins:2 & Cdc14:3 & Clb1_2) (G2_proteins:3 & Cdc14:1 & Clb1_2:2) (G2_proteins:3 & Cdc14:2 & Clb1_2:3)</p>
	<p>3 -> (G2_proteins:3 & !Cdc14 & !Clb1_2) (G2_proteins:3 & Cdc14:3 & Clb1_2) (G2_proteins:3 & Cdc14:1 & Clb1_2:1) (G2_proteins:3 & Cdc14:2 & Clb1_2:2) (G2_proteins:3 & Cdc14:2 & Clb1_2:1)</p>
Swel *	<p>0 -> (S_proteins & Hsl1:3 & Cdc5:3 & (Gin4 !Gin4) & (!Kcc4 Kcc4)) (S_proteins:1 & Hsl1:1 & Cdc5 & (Gin4 !Gin4) & (!Kcc4 Kcc4)) (S_proteins:1 & Hsl1 & Cdc5:1 & (Gin4 !Gin4) & (!Kcc4 Kcc4)) (S_proteins:2 & Hsl1:2 & (Cdc5:2 Cdc5:3) & (Gin4 !Gin4) & (!Kcc4 Kcc4)) (S_proteins:2 & (Hsl1:2 Hsl1:3) & Cdc5:2 & (Gin4 !Gin4) & (!Kcc4 Kcc4))</p>
	<p>1 -> (S_proteins:1 & !Hsl1 & !Cdc5 & !Kcc4 & !Gin4) (S_proteins:1 & ((Hsl1:2 & !Cdc5) (Hsl1:1 & !Cdc5) (!Hsl1 & Cdc5:1) (!Hsl1 & Cdc5:2)) & (Kcc4:1 !Kcc4 Kcc4:2) & (Gin4:1 !Gin4 Gin4:2)) (S_proteins:2 & Hsl1:1 & Cdc5:1 & (Kcc4:1 Kcc4:2 !Kcc4) & (Cdc5:1 Cdc5:2 !Cdc5)) (S_proteins:2 & Kcc4:3 & Gin4:3 & (Hsl1:1 !Hsl1) & (!Cdc5 Cdc5:1))</p>
	<p>2 -> (S_proteins:2 & !Hsl1 & !Cdc5 & !Kcc4 & !Gin4) (S_proteins:2 & ((Hsl1:2 & !Cdc5) (Hsl1:1 & !Cdc5) (!Hsl1 & Cdc5:1) (!Hsl1 & Cdc5:2)) & (Kcc4:1 !Kcc4 Kcc4:2) & (Gin4:1 !Gin4 Gin4:2)) (S_proteins:3 & ((Hsl1:1 & Cdc5:1) (Hsl1:2 & Cdc5:1) (Hsl1:1 & Cdc5:2)) & (Kcc4:1 Kcc4:2 !Kcc4) & (Gin4:1 Gin4:2 !Gin4)) (S_proteins:3 & Kcc4:2 & Gin4:3 & (Hsl1:1 !Hsl1) & (Cdc5:1 !Cdc5)) (S_proteins:3 & Kcc4:3 & Gin4:2 & (Hsl1:1 !Hsl1) & (Cdc5:1 !Cdc5)) (S_proteins:3 & Hsl1:2 & Cdc5:2 & (Kcc4:1 Kcc4:2 !Kcc4) & (Cdc5:1 Cdc5:2 !Cdc5))</p>
	<p>3 -> (S_proteins:3 & !Hsl1 & !Cdc5 & !Kcc4 & !Gin4) (S_proteins:3 & ((Hsl1:2 & !Cdc5) (Hsl1:1 & !Cdc5) (!Hsl1 & Cdc5:1) (!Hsl1 & Cdc5:2)) & (Kcc4:1 !Kcc4 Kcc4:2) & (Gin4:1 !Gin4 Gin4:2))</p>
Bub1, Mps1, Bub3	0 -> !Unattached_Kinetochores

e Mad3	1 -> Unattached_Kinetochores:1 2 -> Unattached_Kinetochores:2
Mad1_Mad2	0 -> !Mps1 1 -> Mps1:1 2 -> Mps1:2
MCC	0 -> (Spindle:1 & Bub1:1 & Mad1_Mad2:1 & Bub3:1 & Mad3:1) (Spindle:2 & Bub1:1 & Mad1_Mad2:1 & Bub3:1 & Mad3:1) 1 -> (Mad1_Mad2:1 & Mad3:1 & Bub3:1 & Bub1:1 & !Spindle) (Mad1_Mad2:2 & Mad3:2 & Bub3:2 & Bub1:2)
Cohesin	0 -> Esp1:2 Esp1:3 !DNA_Replication 1 -> DNA_Replication & (!Esp1 Esp1:1)
Kcc4, Gin4 e Hsl1	0 -> !S_proteins 1 -> S_proteins:1 2 -> S_proteins:2 3 -> S_proteins:3
Cdc5	1 -> (G2_proteins:1 & !APCC_Cdh1) (G2_proteins:2 & APCC_Cdh1:1) (G2_proteins:3 & APCC_Cdh1:2) 2 -> (G2_proteins:2 & !APCC_Cdh1) (G2_proteins:3 & APCC_Cdh1:1) 3 -> G2_proteins:3 & !APCC_Cdh1
Cdc14*	0 -> (Net1:3 & Dbf2_Mob1:1) (Net1:3 & Dbf2_Mob1:2) (!Dbf2_Mob1 & Net1:1) (Net1 & !Dbf2_Mob1) (Net1:1 & Dbf2_Mob1:1) (Net1:2 & Dbf2_Mob1:2) (Net1:3 & Dbf2_Mob1:3) 1 -> (Dbf2_Mob1:1 & !Net1) (Dbf2_Mob1:2 & Net1:1) (Dbf2_Mob1:3 & Net1:2) 2 -> (Dbf2_Mob1:2 & !Net1) (Dbf2_Mob1:3 & Net1:1) 3 -> Dbf2_Mob1:3 & !Net1
G2_proteins	2 -> DNA_Replication
Net1	0 -> !G2_proteins !PP2A 1 -> PP2A & G2_proteins:1 2 -> PP2A & G2_proteins:2 3 -> PP2A & G2_proteins:3
PP2A	0 -> (G2_proteins:1 & Slk19) (G2_proteins:2 & Slk19:2) (G2_proteins:2 & Slk19:3) (G2_proteins & Slk19:3) 1 -> (G2_proteins:1 & !Slk19) (G2_proteins:2 & Slk19:1) (G2_proteins:3 & Slk19:2) 2 -> (G2_proteins:2 & !Slk19) (G2_proteins:3 & Slk19:1) 3 -> G2_proteins:3 & !Slk19
Slk19	1 -> Esp1:1 & G2_proteins:1 2 -> Esp1:2 & G2_proteins:2

	3 -> Esp1:3 & G2_proteins:3
Pds1	0 -> (G2_proteins:1 & APCC_Cdc20 & !Chk1) (G2_proteins:2 & APCC_Cdc20:2 & !Chk1) (G2_proteins & APCC_Cdc20:3 & !Chk1) 1 -> (G2_proteins:1 & !APCC_Cdc20) (G2_proteins:2 & APCC_Cdc20:1 & !Chk1) (G2_proteins:3 & APCC_Cdc20:2 & !Chk1) 2 -> (G2_proteins:2 & !APCC_Cdc20) (G2_proteins:3 & APCC_Cdc20:1) (G2_proteins:2 & !APCC_Cdc20 & Chk1) (G2_proteins:3 & APCC_Cdc20 & Chk1) 3 -> (G2_proteins:3 & !APCC_Cdc20 & !Chk1) (G2_proteins:3 & !APCC_Cdc20 & Chk1)
Chk1	0 -> Spindle & Mec1:1 1 -> Mec1:1 & !Spindle 2 -> Mec1:2
APCC_Cdh1	0 -> !Cdc14 1 -> Cdc14:1 2 -> Cdc14:2 3 -> Cdc14:3
Bfa1_Bub2	0 -> (Cdc5:2 & Misaligned_Spindle) (Cdc5:3 & Misaligned_Spindle) (Spindle:1 & Misaligned_Spindle:1) (Spindle:2 & Misaligned_Spindle:1) 1 -> Misaligned_Spindle:1 & !Spindle & !Cdc5 2 -> Misaligned_Spindle:2 & !Cdc5
Mec1	1 -> DNA_Damage:1 2 -> DNA_Damage:2
Tem1	0 -> Bfa1_Bub2:1 Bfa1_Bub2:2 1 -> G2_proteins:1 & !Bfa1_Bub2 2 -> G2_proteins:2 & !Bfa1_Bub2 3 -> G2_proteins:3 & !Bfa1_Bub2
Cdc15	0 -> Rad53:1 1 -> Tem1 & G2_proteins:1 2 -> Tem1 & G2_proteins:2 3 -> Tem1 & G2_proteins:3
Dbf2_Mob1	1 -> Cdc15:1 & G2_proteins:1 2 -> Cdc15:2 & G2_proteins:2 3 -> Cdc15:3 & G2_proteins:3
Rad53	0 -> Spindle & (!DNA_Damage DNA_Damage:1) 1 -> DNA_Damage:1 & !Spindle 2 -> DNA_Damage:2
MITOSIS_EXIT	0 -> !Spindle !BUD !DNA_Replication 1 -> (Spindle:2 Spindle:1) & BUD:1 & DNA_Replication:1 & (Clb1_2:1 Clb1_2:2 Clb1_2:3) 2 -> Spindle:2 & BUD:1 & DNA_Replication:1 &

Tabela S2: Equações para cada valor de cada nó do sistema. Na primeira coluna estão os nós do sistema, na segunda, estão todas as equações de determinado nó, e cada valor que ele atinge pela equação. Os números após um nó na segunda coluna, seguindo “:”, indica o valor do nó regulador, caso não tenha nenhum, indica que pode ser qualquer valor possível.

Em relação às equações de MBF e SBF, Whi5 não é capaz de inibir completamente SBF ou MBF, pois, mesmo quando Whi5 é superexpresso, não causa um *arrest* ou inibe a ação destas outras proteínas, apenas causa um aumento no tamanho celular (NASH *et al.*, 2001). Como explicado na introdução, Clb5p é capaz de causar o *START* do ciclo, este efeito foi modelado apenas na ausência de Cln1/2 ou Cln3, por este motivo que, na equação, Clb5/6 só tem efeito na ausência destes dois.

Tratando de Clb1/2, quanto à sua inativação, sabe-se que a degradação pelo complexo APC é necessária para a saída da mitose, logo, apenas Sic1 não é capaz de realizar esse feito. Para resolver esta questão, na equação, é necessário que pelo menos um complexo APC esteja ativo (APCC_Cdc20 ou APCC_Cdh1) (WASH; CROSS, 2002). Além disso, inicialmente Mih1 possuía um papel na rede, no entanto, como descrito por Davidich e Bornholdt (2008), este homólogo em *S. cerevisiae* é insignificante para o ciclo, portanto, não possui presença nas equações de Clb1/2, mas foi deixado no modelo para, caso novos estudos apareçam, modifica-lo se necessário. Como não haviam muitas informações sobre Cak1 e sua relação com Clb1/2, e como sabe-se que o fenótipo de *S. cerevisiae* sem ele é nulo (GIAVER *et al.*, 2002), foi colocado no sistema como variável Booleana essencial para a atividade de Clb1/2.

Para a degradação completa de Swe1 foi designado que há a necessidade da presença de Hsl1 e Cdc5 (LIU; WANG, 2006). Kcc4 e Gin4, por falta de informações adicionais, foram designados apenas como elementos capazes de diminuir Swe1, porém, incapazes de degradá-los completamente.

Em Cdc14, como explicado em Resultados e Discussões, há uma dependência em Dbf2_Mob1, ativado por Cdc15; no entanto, sabe-se que há maneiras de se “burlar” esta necessidade de Cdc15 para a liberação de Cdc14 (SHOU; DESHAIES, 2002). Portanto, esta equação não está de pleno acordo com a literatura, necessitando de um estudo mais aprofundado e novas modificações estruturais.

No software GINsim existe um tipo especial de representação para expressões lógicas, para cada valor do nó existe uma expressão que torna aquele valor verdadeiro, por exemplo, para Gin4 obter o valor “1” é preciso que o nó “S_proteins” tenha o valor “1” e que o nó

“lnc_10027” esteja ausente, portanto, a expressão lógica nesse caso seria $1 \rightarrow S_proteins:1 \& !lnc_10027$.

Tabela S3: Equações (à direita) para as interações dos nós (à esquerda) com os lncRNAs de cada linhagem. Os símbolos “&”, “!” e “|” representam, respectivamente, os operadores booleanos AND, NOT e OR. Os números antes das equações são os valores que o nó assume quando a função Booleana da equação é verdadeira.

Gin4 (lnc_10027 - BY4742)	$0 \rightarrow !S_proteins \mid lnc_10027:3 \mid (S_proteins:1 \& lnc_10027) \mid (S_proteins:2 \& lnc_10027:2)$
	$1 \rightarrow S_proteins:1 \& !lnc_10027$
	$2 \rightarrow (S_proteins:2 \& !lnc_10027) \mid (S_proteins:2 \& !lnc_10027) \mid (S_proteins:2 \& lnc_10027:1) \mid (S_proteins:3 \& lnc_10027:2)$
	$3 \rightarrow S_proteins:3 \& !lnc_10027$
Gin4 (lnc_9136 - SEY6210)	$0 \rightarrow !S_proteins \mid lnc_9136:3 \mid (S_proteins:1 \& lnc_9136) \mid (S_proteins:2 \& lnc_9136:2)$
	$1 \rightarrow S_proteins:1 \& !lnc_9136$
	$2 \rightarrow (S_proteins:2 \& !lnc_9136) \mid (S_proteins:2 \& !lnc_9136) \mid (S_proteins:2 \& lnc_9136:1) \mid (S_proteins:3 \& lnc_9136:2)$
	$3 \rightarrow S_proteins:3 \& !lnc_9136$
Hsl1 (lnc_6374 - BY4742)	$0 \rightarrow !S_proteins \mid lnc_6374:3 \mid (S_proteins:1 \& lnc_6374) \mid (S_proteins:2 \& lnc_6374:2)$
	$1 \rightarrow S_proteins:1 \& !lnc_6374$
	$2 \rightarrow (S_proteins:2 \& !lnc_6374) \mid (S_proteins:2 \& !lnc_6374) \mid (S_proteins:2 \& lnc_6374:1) \mid (S_proteins:3 \& lnc_6374:2)$

	$3 \rightarrow S_proteins:3 \ \& \ !Inc_6374$
Hsl1 (Inc_9136 – SEY6210)	$0 \rightarrow !S_proteins \ \ Inc_9136:3 \ \ (S_proteins:1 \ \& \ Inc_9136) \ \ (S_proteins:2 \ \& \ Inc_9136:2)$
	$1 \rightarrow S_proteins:1 \ \& \ !Inc_9136$
	$2 \rightarrow (S_proteins:2 \ \& \ !Inc_9136) \ \ (S_proteins:2 \ \& \ !Inc_9136) \ \ (S_proteins:2 \ \& \ Inc_9136:1) \ \ (S_proteins:3 \ \& \ Inc_9136:2)$
	$3 \rightarrow S_proteins:3 \ \& \ !Inc_9136$
Bub1 (Inc_10883 – BY4742)	$0 \rightarrow Inc_10883:3 \ \ (Unattached_Kinetochores:1 \ \& \ Inc_10883:2)$
	$1 \rightarrow (Unattached_Kinetochores:1 \ \& \ !Inc_10883) \ \ (Unattached_Kinetochores:2 \ \& \ (Inc_10883:1 \ \ Inc_10883:2))$
	$2 \rightarrow Unattached_Kinetochores:2$

Tabela S4: ODEs utilizadas para a modelagem quantitativa da rede envolvendo Adh2p.

Hcm1	$\frac{d[Hcm1]}{dt} = k_{shcm1} - kd_{hcm1} * [Hcm1] - ki_{hsf1} * [Hsf1]$
Cat8	$\frac{d[Cat8]}{dt} = k_{scat8} - kd_{cat8} * [Cat8] - k_{lnc} * [LncRNA] * [Cat8]$
Ixr1	$\frac{d[Ixr1]}{dt} = k_{sixer1} + k_{sshsf1} * [Hsf1] - kd_{ixr1} * [Ixr1]$
Adr1	$\frac{d[Adr1]}{dt} = k_{sadr1} * [Adr1] + k_{sshcm1} * [Hcm1] - kd_{adr1} * [Adr1]$
Adh2	$\frac{d[Adh2]}{dt} = k_{sadh2} + k_{ssc8} * [Cat8] + k_{ssadr1} * [Adr1] - ki_{ixr1} * [Ixr1] - kd_{adh2} * [Adh2]$

Lnc	$\frac{d[Lnc]}{dt} = k_{S_{lnc}} + k_{SS_{hsf1}} * [Hsf1] - k_{a_{lnc}} * [Cat8] * [Adr1] - k_{i_{nrp1}} * [Nrp1] - k_{d_{lnc}} * [Lnc]$
Hsf1	$\frac{d[Hsf1]}{dt} = k_{S_{hsf1}} - k_{d_{hsf1}} * [Hsf1]$
Nrp1	$\frac{d[Nrp1]}{dt} = k_{S_{nrp1}} - k_{d_{nrp1}} * [Nrp1]$
Prm4	$\frac{d[Prm4]}{dt} = k_{S_{prm4}} + k_{SS_{cat8}} * [Cat8] + k_{SS_{ixr1}} * [Ixr1] + k_{d_{prm4}} * [Prm4]$

Tabela S5: Parâmetros utilizados nas ODEs. “ks” é a síntese basal, “kss” é a síntese mediada por fatores de transcrição, “ki” e “kd” são as taxas de inibição e degradação respectivamente, e “ka” é a taxa de associação para formação de complexos proteicos.

	Hcm1	Cat8	Adr1	Adh2	Ixr1	Nrp1	Lnc	Prm4	Hsf1
ks	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
kss	0.0041	0.0041	0.0041	-	0.0041	-	-	-	0.0041
ka	-	-	-	-	-	-	0.66	-	-
ki	-	-	-	-	0.002	0.002	-	-	0.002
kd	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

Referências

NASH, R. S.; VOLPE, T.; FUTCHER, B. Isolation and characterization of WHI3, a size-control gene of *Saccharomyces cerevisiae*. **Genetics**, v. 157, n. 4, p. 1469-1480, 2001.

WÄSCH, Ralph; CROSS, Frederick R. APC-dependent proteolysis of the mitotic cyclin Clb2 is essential for mitotic exit. **Nature**, v. 418, n. 6897, p. 556, 2002.

GIAEVER, Guri et al. Functional profiling of the *Saccharomyces cerevisiae* genome. **nature**, v. 418, n. 6896, p. 387, 2002.

LIU, Hong; WANG, Yanchang. The function and regulation of budding yeast Swe1 in response to interrupted DNA synthesis. **Molecular biology of the cell**, v. 17, n. 6, p. 2746-2756, 2006.

SHOU, Wenying; DESHAIES, Raymond J. Multiple Arrest na Telófase bypassed (tab) mutants alleviate the essential requirement for Cdc15 in exit from mitosis in *S. cerevisiae*. **BMC genetics**, v. 3, n. 1, p. 4, 2002.

KRAIKIVSKI, Pavel et al. From START to FINISH: computational analysis of cell cycle control in budding yeast. **NPJ systems biology and applications**, v. 1, p. 15016, 2015.

CHEN, Katherine C. et al. Integrative analysis of cell cycle control in budding yeast. **Molecular biology of the cell**, v. 15, n. 8, p. 3841-3862, 2004.

CHEN, Katherine C. et al. Kinetic analysis of a molecular model of the budding yeast cell cycle. **Molecular biology of the cell**, v. 11, n. 1, p. 369-391, 2000.

ZOU, Lee; MITCHELL, Jay; STILLMAN, Bruce. CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. **Molecular and cellular biology**, v. 17, n. 2, p. 553-563, 1997.

HOGAN, Eileen; KOSHLAND, Douglas. Addition of extra origins of replication to a minichromosome suppresses its mitotic loss in *cdc6* and *cdc14* mutants of *Saccharomyces cerevisiae*. **Proceedings of the National Academy of Sciences**, v. 89, n. 7, p. 3098-3102, 1992.

KITADA, Kunio et al. Temperature-sensitive *cdc7* mutations of *Saccharomyces cerevisiae* are suppressed by the *DBF4* gene, which is required for the G1/S cell cycle transition. **Genetics**, v. 131, n. 1, p. 21-29, 1992.

GUACCI, Vincent; KOSHLAND, Douglas; STRUNNIKOV, Alexander. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of *MCD1* in *S. cerevisiae*. **Cell**, v. 91, n. 1, p. 47-57, 1997.

HARDWICK, Kevin G. et al. Activation of the budding yeast spindle assembly checkpoint without mitotic spindle disruption. **Science**, v. 273, n. 5277, p. 953-956, 1996.

LUCA, Francis C.; WINEY, Mark. *MOB1*, an essential yeast gene required for completion of mitosis and maintenance of ploidy. **Molecular biology of the cell**, v. 9, n. 1, p. 29-46, 1998.

GIBSON, Daniel G.; BELL, Stephen P.; APARICIO, Oscar M. Cell cycle execution point analysis of ORC function and characterization of the checkpoint response to ORC inactivation in *Saccharomyces cerevisiae*. **Genes to Cells**, v. 11, n. 6, p. 557-573, 2006.

NEECKE, Holger; LUCCHINI, Giovanna; LONGHESE, Maria Pia. Cell cycle progression in the presence of irreparable DNA damage is controlled by a Mec1- and Rad53-dependent checkpoint in budding yeast. **The EMBO Journal**, v. 18, n. 16, p. 4485-4497, 1999.

RICHMOND, Daniel et al. *Slk19* clusters kinetochores and facilitates chromosome bipolar attachment. **Molecular biology of the cell**, v. 24, n. 5, p. 566-577, 2013.

HEALY, Annette M. et al. *CDC55*, a *Saccharomyces cerevisiae* gene involved in cellular morphogenesis: identification, characterization, and homology to the B subunit of mammalian type 2A protein phosphatase. **Molecular and cellular biology**, v. 11, n. 11, p. 5767-5780, 1991.

JACOBSON, Matthew D. et al. Mutations in *SID2*, a novel gene in *Saccharomyces cerevisiae*, cause synthetic lethality with *sic1* deletion and may cause a defect during S phase. **Genetics**, v. 159, n. 1, p. 17-33, 2001.