

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-graduação em Bioquímica e Imunologia

Amanda Cristina Pereira Antônio

**CARACTERIZAÇÃO ESTRUTURAL E FUNCIONAL DA PROTEÍNA RQC1P DE
SACCHAROMYCES CEREVISIAE E AVALIAÇÃO DO SEU PAPEL NAS VIAS DE
RESPOSTA A ESTRESSES CELULARES**

Belo Horizonte

2025

Amanda Cristina Pereira Antônio

**CARACTERIZAÇÃO ESTRUTURAL E FUNCIONAL DA PROTEÍNA RQC1P DE
SACCHAROMYCES CEREVISIAE E AVALIAÇÃO DO SEU PAPEL NAS VIAS DE
RESPOSTA A ESTRESSES CELULARES**

Dissertação apresentada ao Programa de Pós-graduação em Bioquímica e Imunologia da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestra em Bioquímica e Imunologia.

Orientador: Prof. Dr. Erich Birelli Tahara

Coorientador: Prof. Dr. Lucas Bleicher

Belo Horizonte

2025

043

Antônio, Amanda Cristina Pereira.

Caracterização estrutural e funcional da proteína Rqc1p de *Saccharomyces cerevisiae* e avaliação do seu papel nas vias de resposta a estresses celulares [manuscrito] / Amanda Cristina Pereira Antônio. – 2025.

91 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Erich Birelli Tahara. Coorientador: Prof. Dr. Lucas Bleicher.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Bioquímica e Imunologia.

1. Bioquímica e imunologia. 2. Estresse Proteotóxico. 3. *Saccharomyces cerevisiae*. 4. Ribossomos. 5. Biossíntese de Proteínas. I. Tahara, Erich Birelli. II. Bleicher, Lucas. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 577.1



Universidade Federal de Minas Gerais
 Curso de Pós-Graduação em Bioquímica e Imunologia ICB/UFMG
 Av. Antônio Carlos, 6627 – Pampulha
 31270-901 – Belo Horizonte – MG
 e-mail: pg-biq@icb.ufmg.br (31)3409-2615



ATA DA DEFESA DA DISSERTAÇÃO DE MESTRADO DE AMANDA CRISTINA PEREIRA ANTÔNIO. Aos vinte e oito dias do mês de março de 2025 às 13:30 horas, reuniu-se no Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, a Comissão Examinadora da dissertação de Mestrado, indicada *ad referendum* do Colegiado do Curso, para julgar, em exame final, o trabalho intitulado "Caracterização estrutural e funcional da proteína Rqc1p de *Saccharomyces cerevisiae* e avaliação do seu papel nas vias de resposta a estresses celulares", requisito final para a obtenção do grau de Mestre em Bioquímica e Imunologia, área de concentração: Bioquímica. Abrindo a sessão, o Presidente da Comissão, Prof. Erich Birelli Tahara, da Universidade Federal de Minas Gerais, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações: Dra. Katharina de Oliveira Barros (Universidade Federal de Minas Gerais), aprovada; Dr. Dawidson Assis Gomes (Universidade Federal de Minas Gerais), aprovada; Dr. Lucas Bleicher - Coorientador (Universidade Federal de Minas Gerais), aprovada; Dr. Erich Birelli Tahara - Orientador (Universidade Federal de Minas Gerais), aprovada. Pelas indicações a candidata foi considerada:

APROVADA
 REPROVADA

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente da Comissão encerrou a reunião e lavrou a presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 28 de março de 2025.


 Dra. Katharina de Oliveira Barros (UFMG)


 Dr. Dawidson Assis Gomes (UFMG)


 Dr. Lucas Bleicher - Coorientador (UFMG)


 Dr. Erich Birelli Tahara - Orientador (UFMG)

AGRADECIMENTOS

Ao meu orientador, Prof. Erich Birelli Tahara, o qual faltam palavras para descrever a imensa gratidão por toda atenção, dedicação e apoio, ao longo de 6 anos. Agradeço por me ensinar grande parte de tudo que sei hoje e que fazer ciência é sempre ir em busca do conhecimento.

Ao Prof. Lucas Bleicher, que atravessou meus caminhos acadêmicos em 2018 e indiretamente me apresentou a Bioquímica, o que mudou todo o rumo da minha carreira e que anos depois veio a contribuir e muito com este trabalho.

Aos meus amigos do grupo de Bioquímica e Biologia Molecular de Leveduras, carinhosamente apelidados de ETs, Erik Marques e Aysllan Coelho, por toda parceria, deixando os dias no laboratório mais leves e divertidos.

À Prof. Mariana Quezado, por toda contribuição intelectual e pelos dias mais animados no laboratório. Aos demais membros e ex membros do Laboratório de Biofísica de Macromoléculas: Ana Carolina Batista, Ana Tedeschi, Bruno Santos, Camila Yamada, Flávio Júnior, Gabriel Portwood, Juliana Rodrigues, John Amaya, Pedro Campos, Pedro Gontijo, Rafael Lemos, Victor Fabri e, em especial, Julia Rodrigues, com os quais pude aprender ainda mais sobre o mundo das proteínas.

Aos professores do Laboratório de Genética Bioquímica, prof. Glória Regina Franco e Carlos Renato Machado, pela colaboração e disponibilização do espaço e dos equipamentos sempre que necessário. Aos atuais e ex alunos do LGB: Adalberto Miranda, Amanda Carolina, André Rodrigues, Bárbara Marinho, Bruno Resende, Dáfne Oliveira, Daniela de Laet, Gabriel Mendes, Gabrielle Melo, Heron Hilário, Jéssica Duarte, Júlia Raspante, Lorrane Diniz, Veronica Santana e Wesley Ferreira, que contribuíram para minha formação profissional e pessoal.

Ao Prof. Mário Henrique de Barros e aos membros do Laboratório de Biogênese Mitocondrial do ICB-USP, em especial, Marcelo Mesa, Leticia Veloso e Ana Paula Barbosa, pela amizade e auxílio nos experimentos de clonagem gênica.

Aos demais laboratórios do ICB-UFMG, técnicos e servidores, em especial, Alexandre, Eneida, Elimar, Renata, Jamil e Patrícia, que contribuíram de alguma forma para que este trabalho pudesse ser realizado.

Ao Programa de Pós-Graduação em Bioquímica e Imunologia, por todo conhecimento adquirido ao longo dos anos que agregou para a minha formação profissional, e às agências de fomento, pelo financiamento deste projeto.

À minha mãe Nair Aparecida e à minha avó Raimunda Cristiana, que são as minhas maiores inspirações e que batalharam para que eu pudesse ter a oportunidade de estudar e seguir nesta carreira.

À minha irmã, Daniela Luciana, pela paciência, companheirismo e apoio incondicional diário desde os 3 anos de idade e que sem ela, nada disso seria possível.

Ao meu tio, Gilmar Santana, que não pôde me ver formar, mas torceu pelo meu sucesso e se fez presente desde o meu primeiro dia de vida.

Ao Freddy, por todo amor, dedicação e apoio, estando comigo desde o primeiro dia da minha jornada acadêmica o qual compartilhamos todos os nossos objetivos, sonhos e frustrações e que se tornou meu parceiro de vida e sem o qual eu não teria chegado até aqui.

Aos meus amigos da vida, Marcus Santos, Kahena Barros, Gustavo Almeida, Juliano Guerra, Fernanda Frias e Gustavo Santos, por todo companheirismo e paciência ao longo dos anos e que tornaram meus dias fora da universidade mais felizes. À Thais Bassani, que sobreviveu em bases comigo e foi companhia essencial ao longo destes dois últimos anos para compartilhar os dias bons e superar os dias difíceis dentro e fora do trabalho.

Por fim, gostaria de agradecer à UFMG, por ter sido o meu lar durante quase 8 anos e ter mudado definitivamente a minha vida. Agradeço pelo conhecimento imensurável adquirido durante a graduação em Ciências Biológicas e durante o mestrado em Bioquímica, pelas conexões geradas ao longo dos anos e pelos amigos que vou carregar durante a vida.

RESUMO

Condições de estresse celular promovem remodelamento do proteoma para permitir que as células se adaptem às novas condições e, neste contexto, ocorrem processos de re-enovelamento e degradação de proteínas defeituosas. Desta forma, mecanismos de vigilância da qualidade de proteínas que identificam a síntese de proteínas aberrantes e previnem seu acúmulo são fundamentais para a manutenção da proteostase celular. Um destes processos de vigilância é realizado pela via do complexo de controle de qualidade de proteína associado ao ribossomo (RQC) que tem como alvo os polipeptídeos defeituosos recém-sintetizados presos ao ribossomo, marcando-os para degradação via sistema ubiquitina-proteassomo. Na ausência de proteínas do complexo RQC, é observado estresse proteotóxico em leveduras e neurodegeneração em camundongos. A proteína Rqc1p é uma das proteínas integrantes do RQC, cuja função é pouco compreendida. Neste estudo, caracterizamos estrutural e funcionalmente Rqc1p de *Saccharomyces cerevisiae*, investigando seu papel na resposta ao estresse traducional e térmico. Nossas análises *in silico* indicaram que Rqc1p é conservada entre os eucariotos e que a região polibásica próxima ao N-terminal e o domínio TCF25 desempenham papéis essenciais em sua função biológica no núcleo e no citoplasma. Além disso, predições de localização subcelular sugerem que a região polibásica medeia a translocação nuclear de Rqc1p e pode estar associada à sua função como fator de transcrição. Análises fenotípicas mostraram que a deleção de *RQC1* resultou em maior sensibilidade ao choque térmico e reduziu a velocidade específica máxima de crescimento celular nesta condição. Notavelmente, a complementação de células desprovidas de *RQC1* com vetor de expressão contendo este gene foi capaz de restaurar o fenótipo selvagem. Nossos achados expandem o entendimento sobre o papel de Rqc1p e sugerem novos caminhos para investigação da sua função na homeostase proteica durante a resposta ao estresse celular.

Palavras-chave: estresse; levedura; ribossomo; tradução.

ABSTRACT

Cellular stress conditions promote proteome remodeling to allow cells to adapt to ever changing environmental conditions. In this context, protein turnover and degradation of defective proteins play a crucial role in removing substandard proteins from cells. Thus, protein surveillance and quality control mechanisms that detect aberrant protein synthesis and prevent their accumulation are essential for maintaining cellular proteostasis. One such surveillance mechanism is the ribosome-associated protein quality control (RQC), which targets newly synthesized defective polypeptides stalled on the ribosome, marking them for degradation via the ubiquitin-proteasome system. In the absence of RQC complex proteins, proteotoxic stress is observed in yeast, while neurodegeneration occurs in mice. Rqc1p is a subunit of the RQC complex, but its function remains poorly understood. In this study, we performed structural and functional characterization of the *Saccharomyces cerevisiae* Rqc1 protein and investigated its role in the response to both translational and heat shock stresses. *In silico* analyses revealed that Rqc1p is conserved among eukaryotes and that the polybasic region in the N-terminal and the TCF25 domain play essential roles in its biological function in both the nucleus and the cytoplasm. Moreover, subcellular localization predictions suggest that the polybasic region mediates the nuclear translocation of Rqc1p and may be associated with its function as a transcription factor. Phenotypic analyses showed that *RQC1* deletion led to increased sensitivity to heat shock stress and reduced the maximum specific growth rate in this condition. Notably, complementation of *RQC1*-deficient cells with an expression vector containing this gene was able to restore the wild-type phenotype. Our findings expand the understanding of Rqc1p's biological role and suggest new ways for investigating its function in protein homeostasis during cellular stress response.

Keywords: stress; yeast; ribosome; translation.

LISTA DE ILUSTRAÇÕES

Fig. 1. Resposta ao estresse e manutenção da proteostase celular.....	14
Fig. 2. Síntese de proteínas eucarióticas.....	15
Fig. 3. Causas da interrupção da tradução durante o alongamento.....	16
Fig. 4. Via do controle de qualidade associada ao ribossomo eucariótico.....	18
Fig. 5. Modelo molecular do complexo RQC associado à subunidade 60S do ribossomo eucariótico.....	39
Fig. 7. Curva de crescimento celular de <i>S. cerevisiae</i> WT e <i>ltn1Δ</i> em diferentes concentrações de higromicina B.....	40
Fig. 8. Velocidade específica máxima de crescimento celular de <i>S. cerevisiae</i> WT e <i>ltn1Δ</i> em higromicina B.....	41
Fig. 9. Tolerância ao choque térmico de <i>S. cerevisiae</i> WT, <i>rqc2Δ</i> , <i>ltn1Δ</i> e <i>rqc1Δ</i>	42
Fig. 10. Tolerância ao choque térmico de <i>S. cerevisiae</i> WT, <i>rqc1Δ</i> , <i>ltn1Δ</i> e <i>rqc1Δltn1Δ</i>	42
Fig. 11. Curva de crescimento e velocidade específica máxima de crescimento celular de <i>S. cerevisiae</i> WT, <i>rqc1Δ</i> , <i>ltn1Δ</i> e <i>rqc1Δltn1Δ</i> em choque térmico.....	44
Fig. 12. Alinhamento múltiplo de sequências de proteínas homólogas de Rqc1p com enfoque nas regiões de maior conservação.....	46
Fig. 13. Previsão da localização subcelular de Rqc1p de <i>S. cerevisiae</i> e sinais de classificação previstos pela ferramenta DeepLoc-2.1.....	48
Fig. 14. Previsão de NLSs pela ferramenta cNLS Mapper.....	49
Fig. 15. Predição da estrutura secundária e de características estruturais de Rqc1p de <i>S. cerevisiae</i> utilizando a ferramenta NetSurfP-3.0.....	51
Fig. 16. Modelos gerados pelo ColabFold e gráficos de ERRAT e Ramachandran...53	
Fig. 17. Resíduos altamente conservados na família TCF25.....	54
Fig. 18. Decomposição em comunidades da rede de correlação gerada para a família TCF25.....	54
Fig. 19. Representação dos resíduos da comunidade 1 decomposta da rede de correlação gerada para a família TCF25.....	55
Fig. 20. Construção dos plasmídeos.....	56
Fig. 21. Mapa das construções e confirmação da clonagem por digestão diagnóstica.....	57
Fig. 22. Tolerância ao choque térmico de <i>S. cerevisiae</i> WT, <i>rqc1Δ</i> , <i>rqc1Δ+Rqc1p</i> , <i>rqc1Δ+Rqc1p1-327</i> e <i>rqc1Δ+Rqc1p328-723</i>	58
Fig. 23. Análise de componentes principais (PCA).....	60
Fig. 24. Genes diferencialmente expressos em choque térmico.....	61
Fig. 25. Rede de interação de Rqc1p obtida pelo STRING v.12.....	62
Fig. 26. Expressão dos genes associados a resposta ao choque térmico e ao controle da síntese de proteínas.....	63

LISTA DE TABELAS

Tabela 1. Descrição dos genótipos de <i>S. cerevisiae</i> utilizados neste estudo.....	23
Tabela 2. <i>Primers</i> utilizados para as clonagens.....	26
Tabela 3. <i>Primers</i> utilizados para o sequenciamento.....	26
Tabela 4. Programas utilizados para amplificação dos fragmentos utilizados para clonagem.....	28
Tabela 5. Descrição das digestões diagnósticas realizadas para cada construção...31	
Tabela 6. Descrição geral das sequências das proteínas ortólogas de organismos modelo selecionadas do Uniprot.....	33
Tabela 7. Descrição das métricas utilizadas para validação dos modelos gerados por abordagens computacionais.....	51
Tabela 8. Descrição dos dados de sequenciamento obtidos no SRA.....	59

LISTA DE ABREVIATURAS E SIGLAS

bHLH	básicas hélice-volta-hélice (<i>basic helix-loop-helix</i>)
CAT-tail	cauda de Alanina-Treonina na extremidade C-terminal
DEGs	genes diferencialmente expressos (<i>differentially expressed genes</i>)
DRCN	decomposição de redes de coevolução de resíduos (<i>decomposition of residue coevolution networks</i>)
eEFs	fatores de liberação de alongamento eucariótico (<i>eukaryotic elongation factors</i>)
eIFs	fatores de iniciação da tradução eucariótica (<i>eukaryotic translation initiation factors</i>)
eRFs	fatores de liberação de cadeia polipeptídica eucariótica (<i>eukaryotic peptide chain release factors</i>)
HSR	resposta ao choque térmico (<i>heat shock response</i>)
Hsf1	fator de choque térmico 1 (<i>heat shock factor 1</i>)
HSP	proteínas do choque térmico (<i>heat shock proteins</i>)
IDR	regiões intrinsecamente desordenadas (<i>intrinsically disordered regions</i>)
K48	lisina 48
mRNA	RNA mensageiro
MSA	alinhamento múltiplo de sequência (<i>multiple sequence alignment</i>)
NES	sinal de exportação nuclear (<i>nuclear export signal</i>)
NFAT	fator nuclear de células T ativadas (<i>nuclear factor of activated T cells</i>)
NLS	sinal de localização nuclear (<i>nuclear localization signal</i>)
ORF	janela de leitura aberta (<i>open reading frame</i>)
PCA	análise de componentes principais
pLDDT	<i>predicted local distance difference test</i>
tRNA	RNA transportador
RQC	complexo de controle de qualidade de proteína associado ao ribossomo (<i>Ribosome-associated Quality Control</i>)
RSA	acessibilidade relativa à superfície (<i>relative surface accessibility</i>)
SRA	<i>Sequence read archive</i>
SRF	fator de resposta ao soro (<i>serum response factor</i>)
$\mu^{\text{máx}}/\text{h}$	velocidade específica máxima de crescimento celular

SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 Estresses e a manutenção da proteostase celular.....	13
1.2 Tradução.....	14
1.3 Controle de qualidade em síntese de proteínas eucarióticas.....	16
1.4 A diversidade funcional da proteína Rqc1.....	19
2 OBJETIVOS.....	22
2.1 Objetivo geral.....	22
2.2 Objetivos específicos.....	22
3 MATERIAL E MÉTODOS.....	23
3.1 Caracterização funcional da proteína Rqc1.....	23
3.1.1 Linhagens de <i>Escherichia coli</i> e <i>Saccharomyces cerevisiae</i>	23
3.1.2 Meios de cultura, cultivo e armazenamento.....	23
3.1.3 Determinação da tolerância celular ao estresse traducional e ao choque térmico.....	24
3.1.4. Determinação da curva de crescimento celular.....	25
3.1.5 Determinação da velocidade específica máxima de crescimento.....	25
3.1.6 Desenho dos <i>primers</i>	25
3.1.7 Extração do DNA genômico de <i>S. cerevisiae</i>	26
3.1.8 Amplificação dos fragmentos por reação da polimerase em cadeia (PCR).....	27
3.1.9 Purificação dos produtos de PCR e digestão.....	28
3.1.10 Digestão enzimática e ligação dos fragmentos.....	29
3.1.11 Transformação de <i>E. coli</i> competente.....	29
3.1.12 Identificação de clones positivos e sequenciamento das construções.....	30
3.1.13 Transformação de <i>S. cerevisiae</i> WT e <i>rqc1</i> Δ.....	31
3.1.14 Análises estatísticas e geração de gráficos.....	32
3.2 Caracterização estrutural in silico da proteína Rqc1.....	32
3.2.1 Alinhamento de sequências.....	32
3.2.2 Predição de localização subcelular e sinais de classificação.....	33
3.2.3 Estrutura secundária.....	34
3.2.4 Modelagem molecular e validação estrutural.....	34
3.2.5 Decomposição de redes de coevolução de resíduos.....	35
3.2.6 Análise de similaridade estrutural.....	36
3.2.7 Visualização das estruturas tridimensionais.....	36
3.3 Metanálise dos transcriptomas de <i>S. cerevisiae</i> em choque térmico.....	36
3.3.1 Aquisição dos dados de sequenciamento de RNA e análise da qualidade das <i>reads</i>	36
3.3.2 Mapeamento no transcriptoma de referência e contagem das <i>reads</i>	37
3.3.3 Análises da expressão diferencial de genes.....	37
3.3.4 Recursos computacionais.....	38

4 RESULTADOS.....	39
4.1 A deleção de <i>LTN1</i> e <i>RQC1</i> reduz a viabilidade celular de <i>S. cerevisiae</i> na presença de higromicina B.....	39
4.2 O impacto da higromicina B na velocidade específica máxima de crescimento é acentuado na ausência de <i>LTN1</i>	39
4.3 A deleção de <i>RQC1</i> compromete a viabilidade celular de <i>S. cerevisiae</i> sob choque térmico enquanto a deleção concomitante de <i>LTN1</i> restaura o fenótipo selvagem.....	41
4.4 A deleção do gene <i>RQC1</i> reduz a velocidade específica máxima de crescimento celular em choque térmico.....	43
4.5 Rqc1p apresenta uma região polibásica N-terminal e um domínio C-terminal conservado em eucariotos.....	45
4.6 A predição de NLSs e a presença de sítios de fosforilação sugerem que Rqc1p pode ser translocada para o núcleo.....	46
4.7 A predição da estrutura secundária de Rqc1p indica a presença de regiões desordenadas no N-terminal e formação de múltiplas α -hélices no domínio TCF25.....	49
4.8 Modelos gerados <i>in silico</i> e validação estrutural.....	51
4.9 Identificação de resíduos de Rqc1p que coevoluíram.....	53
4.10 Construção dos plasmídeos YCplac111+Rqc1p, YCplac111+Rqc1p1-327 e YCplac111+Rqc1p328-723.....	55
4.11 A reintrodução de <i>RQC1</i> em <i>S. cerevisiae rqc1</i> Δ complementa o fenótipo selvagem em choque térmico.....	57
4.12 Bibliotecas selecionadas, mapeamento, contagem de genes e análise de componentes principais.....	58
4.13 Análise da expressão diferencial de genes.....	60
5 DISCUSSÃO.....	64
6 CONCLUSÕES.....	71
7 REFERÊNCIAS BIBLIOGRÁFICAS.....	72
8 APÊNDICES.....	78

1 INTRODUÇÃO

1.1 Estresses e a manutenção da proteostase celular

Condições de estresse impactam a homeostase do proteoma celular, uma vez que podem alterar o enovelamento correto das proteínas levando à perda de sua função. Além disso, proteínas defeituosas não-funcionais tendem a se acumular e serem propensas a agregação, gerando altos níveis de toxicidade celular. Desta forma, vias de resposta a estresses e controle de qualidade de proteínas são essenciais para a manutenção da integridade das funções celulares (RICHTER; HASLBECK; BUCHNER, 2010; SANTIAGO; GONÇALVES; MORANO, 2020).

Um dos mecanismos de adaptação a estresses celulares mais estudados é a resposta ao choque térmico (HSR, *heat shock response*), que é conservada entre os eucariotos. A HSR atua principalmente sob a regulação do fator de transcrição fator de choque térmico 1 (Hsf1, *heat shock factor 1*), que modula genes em resposta ao estresse proteotóxico e oxidativo, assim como em condições de restrição calórica e choque térmico, através da expressão de chaperonas e outros elementos que mantêm a proteostase (HAHN; THIELE, 2004; MORANO; GRANT; MOYE-ROWLEY, 2012). Muitos dos avanços na elucidação dos mecanismos moleculares e regulatórios da HSR vieram de estudos realizados utilizando a levedura *Saccharomyces cerevisiae* como organismo modelo (NUSSBAUM et al., 2014; VERGHESE et al., 2012). Em leveduras, o gene *HSF1* é essencial para a viabilidade celular e, em humanos, apresenta conservação funcional, visto que uma das isoformas de *HSF1* humano complementa a deleção do gene em leveduras (LIU, 1997).

Em resumo, condições de estresse promovem o remodelamento do proteoma para que as células possam se adaptar às condições ambientais que podem apresentar constante mudança. Nesse cenário, há re-enovelamento de proteínas mal dobradas, degradação de proteínas aberrantes e síntese de novas proteínas que irão integrar o programa de resposta a estas condições (Fig. 1) (MÜHLHOFER et al., 2019).

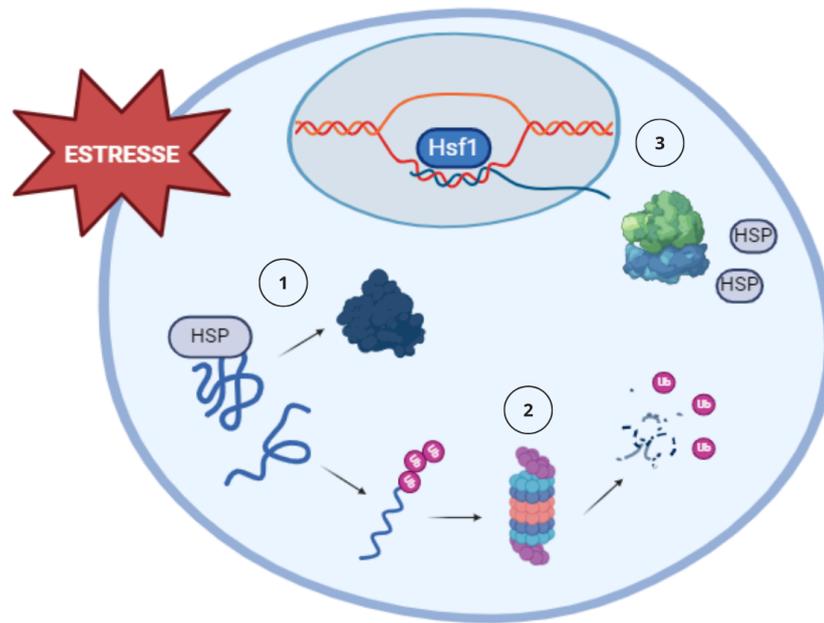


Fig. 1. Resposta ao estresse e manutenção da proteostase celular. A adaptação celular após a exposição a diferentes estresses inclui: (1) re-enovelamento de proteínas mal dobradas através da ação de proteínas do choque térmico (HSPs, *heat shock proteins*), (2) degradação de proteínas aberrantes poliubiquitinadas pelo proteassomo e (3) ativação da transcrição de genes por Hsf1 e síntese de novas proteínas que modulam a resposta ao estresse.

1.2 Tradução

A tradução é o processo de síntese de proteínas celulares realizada pelos ribossomos. Os ribossomos são complexos ribonucleoproteicos que, a partir da leitura da informação contida em uma molécula de RNA mensageiro (mRNA), sintetizam proteínas por meio da ação de RNAs transportadores (tRNA). Para exercer sua função de síntese proteica, as subunidades 40S e 60S dos ribossomos eucarióticos se unem e formam três sítios de ligação para os tRNA, formando o ribossomo 80S. O sítio A é o local de ligação do aminoacil-tRNA que irá ser acoplado na cadeia polipeptídica em formação, enquanto o sítio P é o local onde ocorre a ligação peptídica. O último sítio, denominado de E, é ocupado pelos tRNAs desacilados, os quais, posteriormente, irão se dissociar do ribossomo (ZAHA; FERREIRA; PASSAGGLIA, 2014).

O processo da síntese de proteínas eucarióticas pode ser dividido em quatro etapas principais sendo elas (i) a iniciação; (ii) o alongamento; (iii) a terminação; e (iv) a reciclagem dos ribossomos (Fig. 2). Em eucariotos, a iniciação da tradução é orquestrada pela presença de muitos fatores de iniciação da tradução eucariótica (eIFs, *eukaryotic translation initiation factors*) que coordenam a montagem adequada do ribossomo 80S no códon de iniciação (AUG) com um metionil-tRNA

ligado ao sítio P (BRITO QUERIDO; DÍAZ-LÓPEZ; RAMAKRISHNAN, 2024). A partir disso, durante a etapa de alongamento, por meio da ação dos fatores de alongamento eucariótico (eEFs, *eukaryotic elongation factors*), os ribossomos movimentam-se ao longo do mRNA realizando a leitura dos códons, sintetizando o polipeptídeo de acordo com o códon lido. Ao longo desta etapa, os tRNAs transitam entre os sítios A, P e E de modo que ocorre a decodificação e entrada do aminoacil-tRNA com o anticódon que corresponde ao códon do mRNA no sítio A, a formação da ligação peptídica e a translocação do complexo mRNA-tRNA. O ciclo de alongamento é repetido até que a proteína completa tenha sido sintetizada e que, portanto, o ribossomo encontre um códon de terminação (na maior parte dos casos, UAA, UAG e UGA) no final do quadro de leitura aberto (ORF, *open reading frame*). O códon de terminação é reconhecido por fatores de liberação de cadeia polipeptídica eucariótica (eRFs, *eukaryotic peptide chain release factors*), como o eRF1, que juntamente com a GTPase eRF3, promovem a liberação das proteínas recém sintetizadas. Por fim, na última etapa, o complexo ribossomal é reciclado após a terminação pela ATPase Rli1 (membro 1 da subfamília E do cassete de ligação ao ATP, ABCE1, em mamíferos), que separa as subunidades em 40S e 60S para iniciar o processo novamente (SCHULLER; GREEN, 2018; SHAO et al., 2016).

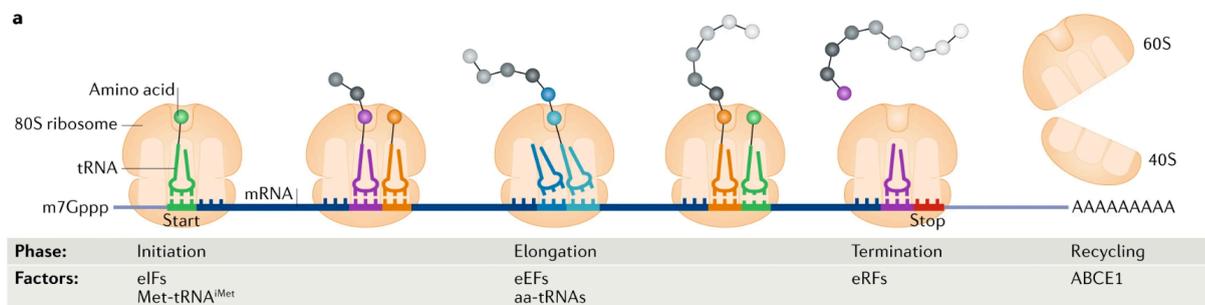


Fig. 2. Síntese de proteínas eucarióticas. A tradução é dividida em quatro etapas principais e se inicia quando as subunidades 60S e 40S, o mRNA, os eIFs e o metionil-tRNA se posicionam no códon de iniciação (AUG). Em seguida, por meio da ação dos eEFs e dos aa-tRNAs, ocorre a etapa de alongamento da cadeia peptídica até que o ribossomo encontre um códon de parada. A terminação ocorre quando os eRFs liberam o peptídeo e as subunidades ficam livres para serem recicladas pelo Rli1 (membro 1 da subfamília E do cassete de ligação ao ATP, ABCE1, em mamíferos). Fonte: SCHULLER; GREEN, 2018

No entanto, durante a etapa de alongamento da tradução, podem ocorrer problemas que causam a paralisação ribossômica, como, por exemplo, (i) a decodificação de códons raros ou subótimos; (ii) flutuações nos níveis ou modificações de tRNAs; (iii) a formação de estruturas secundárias no mRNA; (iv) a

ocorrência de danos químicos ao mRNA ou nos ribossomos; e (v) a clivagem acidental ou poliadenilação prematura que geram mRNAs sem códon de terminação (JOAZEIRO, 2017; CHANDRASEKARAN et al., 2019) (Fig. 3). Estes eventos geram uma cadeia polipeptídica nascente que fica presa ao ribossomo já que, por paralisarem a elongação ribossomal, a tradução não ocorre até a entrada de um códon de terminação no sítio A, acarretando no não-recrutamento de fatores que liberariam o polipeptídeo e que promoveriam a reciclagem dos ribossomos (HELLEN, 2018).

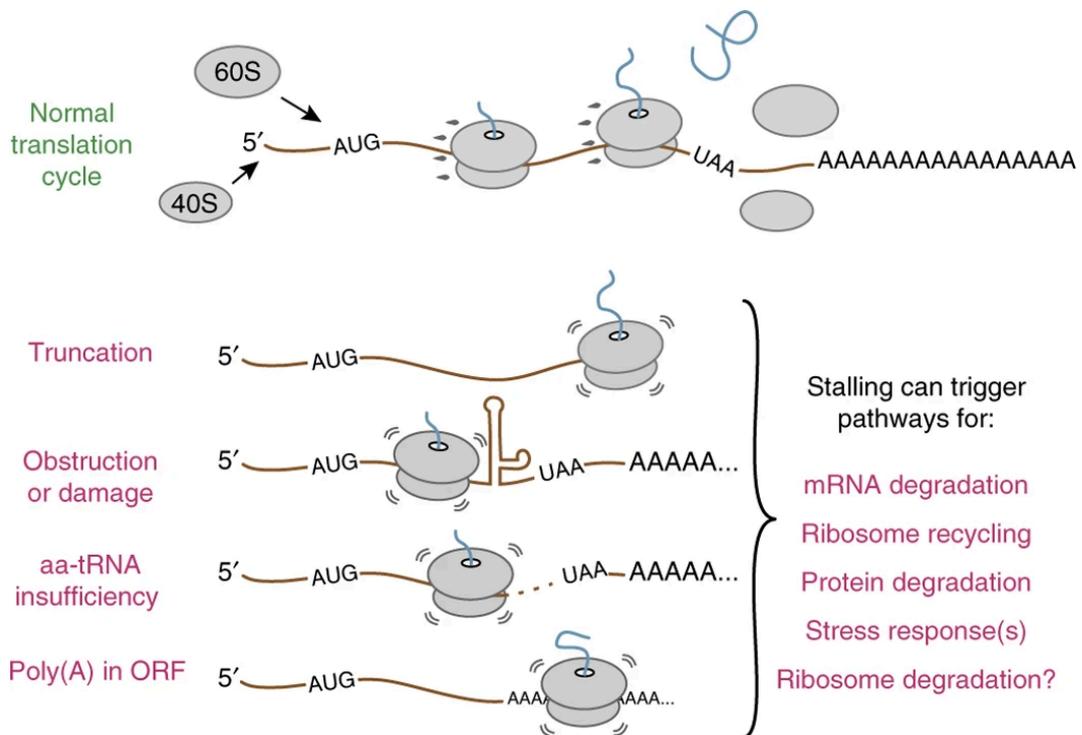


Fig. 3. Causas da interrupção da tradução durante o alongamento. A tradução aberrante pode ocorrer devido a diferentes causas como mRNA clivados sem códon de terminação, obstrução gerada por mRNA que contém formação de estruturas secundárias, tradução de códons raros e da cauda poli-A. Fonte: BRANDMAN; HEGDE, 2016

1.3 Controle de qualidade em síntese de proteínas eucarióticas

Uma vez que as proteínas desempenham papéis essenciais para manutenção da homeostase celular, o sequestro dos ribossomos e a tradução aberrante podem levar ao comprometimento da função celular (GRAILLE; SÉRAPHIN, 2012). Neste cenário, as células possuem mecanismos de vigilância e controle de qualidade de proteínas que identificam a produção de proteínas aberrantes e previnem seu acúmulo (INADA; BECKMANN, 2024). Um desses processos de vigilância é realizado pela via do complexo de controle de qualidade de proteína associado ao ribossomo (RQC, *Ribosome-associated Quality Control*) e tem como alvo os

polipeptídeos defeituosos recém-sintetizados presos ao ribossomo, marcando-os para degradação via sistema ubiquitina-proteassomo (BRANDMAN et al., 2012; DEFENOUILLERE et al., 2013).

O processo geral do controle de qualidade co-traducional eucariótico é desencadeado a partir da paralisação e colisão dos ribossomos e possui as funções de (i) degradação do mRNA aberrante; (ii) reciclagem das subunidades do ribossomo; e (iii) degradação do polipeptídeo defeituoso recém sintetizado (Fig. 4) (BRANDMAN; HEGDE, 2016). Em um primeiro momento, os ribossomos parados e colididos são detectados e poliubiquitinados pela ubiquitina E3 ligase Hel2p (ZNF598 em mamíferos), que juntamente com outros fatores, induzem a dissociação das subunidades 40S e 60S (MATSUO et al., 2020). A subunidade 60S obstruída contendo a cadeia nascente ligada ao tRNA no sítio P, é reconhecida pela subunidade Rqc2p (NEMF em mamíferos) do complexo RQC, que se liga e recruta a ubiquitina E3 ligase do tipo RING, Ltn1p/Rkr1p (Listerin em mamíferos) (BRANDMAN et al., 2012). A proteína Ltn1p apresenta papel central do RQC, já que ela promove a poliubiquitinação dos polipeptídeos defeituosos recém-sintetizados, marcando-os para degradação proteassomal (BENGTSON; JOAZEIRO, 2010) – a marcação de poliubiquitinação pela lisina 48 (K48) é provavelmente mediada pela subunidade Rqc1p (TCF25 em mamíferos). Em seguida, ocorre o recrutamento da ATPase AAA Cdc48p (VCP/p97 em mamíferos) e seus cofatores Ufd1p e Npl4p, que realizam a extração e liberação da cadeia nascente após a clivagem específica do tRNA por Vms1p (ANKZF1 em mamíferos) (KUROHA et al., 2018; VERMA et al., 2013, 2018).

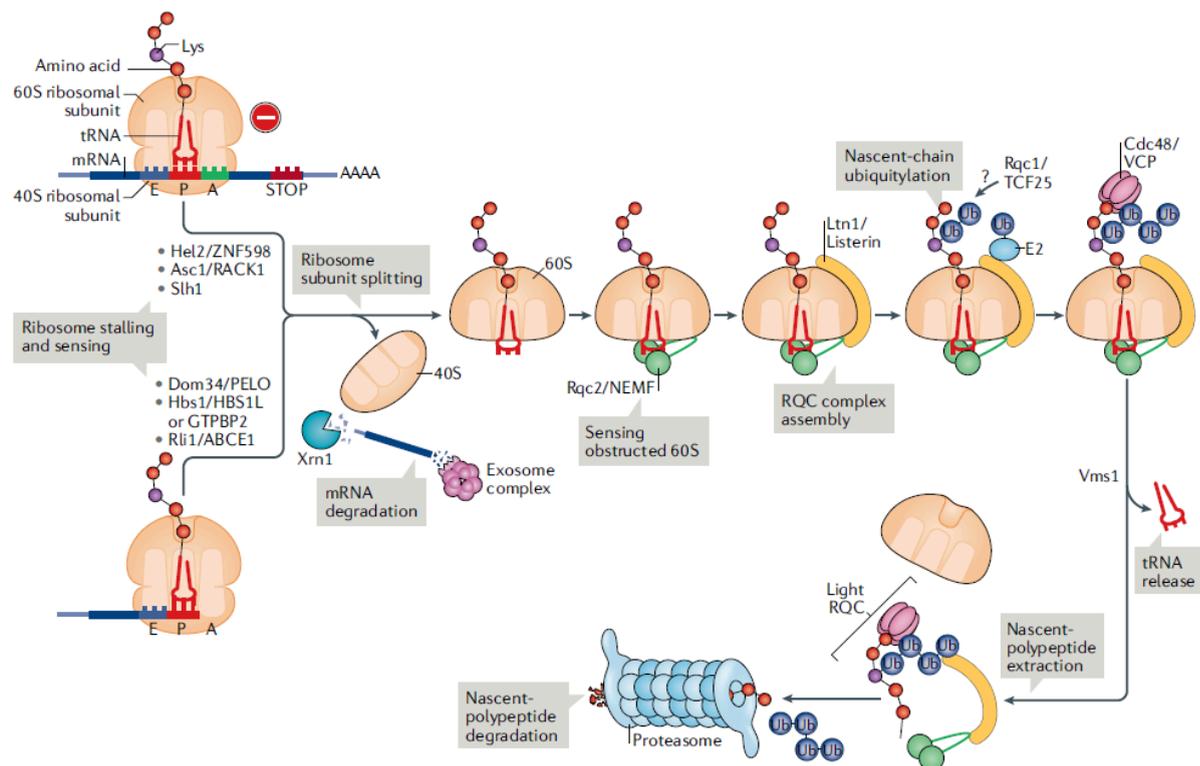


Fig. 4. Via do controle de qualidade associada ao ribossomo eucariótico. Os ribossomos paralisados na extremidade 3' do mRNA são detectados por Hbs1p (HBS1L em mamíferos) e Dom34p (PELO em mamíferos). Já os ribossomos paralisados em sequências internas podem ser reconhecidos por Hel2p (ZNF598 em mamíferos), Asc1p (RACK1) e Slh1p. Dom34p recruta Rli1p (membro 1 da subfamília E do cassete de ligação ao ATP, ABCE1, em mamíferos) que separa as subunidades do ribossomo e o mRNA é liberado e degradado pela exoribonuclease Xrn1p e o complexo exossomo. O peptidil-tRNA permanece ligado à subunidade 60S que é reconhecida por Rqc2p (NEMF em mamíferos) que recruta a estabiliza a ligação de Ltn1p (Listerin em mamíferos). Além disso, Rqc2p adiciona a *CAT-tail* para expor os resíduos de lisina que estão enterrados no túnel de saída. Rqc1p pode estar envolvida na especificidade de poliubiquitinação pela K48 que é reconhecida por Cdc48p (VCP/p97 em mamíferos) e seus cofatores. A cadeia polipeptídica é extraída por Cdc48p e é então degradada pelo proteossomo. Fonte: JOAZEIRO, 2019.

Desde a identificação do complexo central do RQC em *Saccharomyces cerevisiae*, que é composto pelas proteínas Ltn1p, Rqc1p e Rqc2p, esforços vêm sendo realizados para caracterizar e investigar os papéis moleculares de suas proteínas integrantes (JOAZEIRO, 2019). Já se sabe que Rqc2p, além de facilitar a ligação de Ltn1p ao ribossomo, promove a adição de uma cauda de alanina-treonina à extremidade C-terminal (*CAT-tail*) dos polipeptídeos defeituosos, processo que auxilia a exposição dos resíduos de lisina enterrados no túnel de saída do ribossomo para que estes possam ser ubiquitinados pela Ltn1p (SHEN et al., 2015). Além disso, foi observado que a falha em degradar os polipeptídeos que sofreram adição de *CAT-tail* pode resultar em agregação e levar a interrupção da proteostase celular em leveduras (YONASHIRO et al., 2016; SITRON; BRANDMAN, 2019; CHOE et al., 2016). Curiosamente, a mutação de *LTN1* levou à neurodegeneração em

camundongos, o que evidencia a relevância fisiopatológica dos mecanismos que regulam a via do complexo RQC (CHU et al., 2009).

Ademais, estudos estruturais de bancada com foco na elucidação da arquitetura do complexo RQC quando associado ao ribossomo caracterizaram as localizações e as regiões de interação entre Ltn1p e Rqc2p (Fig. 5). Foi descrito que o domínio N-terminal de Ltn1p se liga a Rqc2p e a subunidade 60S e que este é espaçado da região C-terminal por uma região contendo repetições do tipo repetições do tipo HEAT. A região C-terminal, composta pelos domínios RWD e RING, liga-se próxima ao túnel de saída do polipeptídeo, permitindo a poliubiquitinação das cadeias polipeptídicas nascentes (LYUMKIS et al., 2014; SHAO et al., 2015; TESINA et al., 2023). Apesar de o complexo RQC ter sido purificado por afinidade à Rqc1p nestes estudos estruturais, nenhuma densidade foi observada para esta proteína (TESINA et al., 2023).

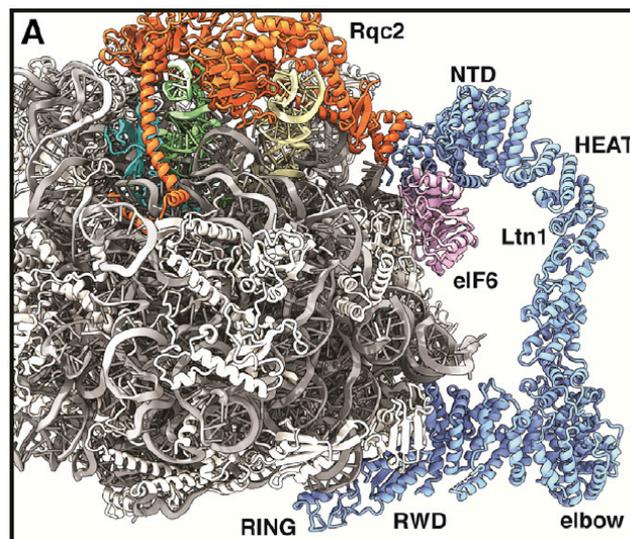


Fig. 5. Modelo molecular do complexo RQC associado à subunidade 60S do ribossomo eucariótico. O modelo destaca os domínios de Ltn1p (azul) e a interação com Rqc2p (laranja) através do domínio N-terminal. Fonte: TESINA, et al. 2019.

1.4 A diversidade funcional da proteína Rqc1

Embora a Rqc1p tenha sido identificada como parte do complexo RQC no início da década de 2010, sua função exata nesta via ainda não está completamente elucidada. Estudos mostraram que a deleção do gene *RQC1* em leveduras não afetou a ligação de Ltn1p e nem comprometeu a extensão da poliubiquitinação de polipeptídeos defeituosos recém-sintetizados. No entanto, a observação do acúmulo de repórteres que contêm a adição de *CAT-tail* indicou que Rqc1p, provavelmente, desempenha um papel nas etapas anteriores ou durante a extração da cadeia

nascente da subunidade 60S (DEFENOUILLÈRE et al., 2016; YONASHIRO et al., 2016). Também foi observado que as cadeias de poliubiquitina que são ligadas via K48 são seletivamente ligadas por Cdc48p e seus cofatores Ufd1p e Npl4p (TSUCHIYA et al., 2017). A partir disso, análises bioquímicas indicaram que cadeias de ubiquitina ligadas via K48 foram preferencialmente formadas sobre outros tipos de cadeia na presença de Rqc1p, sugerindo seu envolvimento na regulação deste processo (KUROHA et al., 2018). No entanto, as tentativas de estabelecer a estrutura e posição de Rqc1p no complexo RQC associado ao ribossomo não foram bem sucedidas e estudos estruturais são essenciais para a compreensão da sua atividade biológica.

Curiosamente, antes da sua associação ao complexo RQC, o gene *TCF25* (inicialmente denominado *NULP1*) foi originalmente identificado em mamíferos, como possível regulador do desenvolvimento inicial de diversos órgãos, com expressão marcante em tecidos embrionários de camundongos, incluindo tecidos neurais, rim, coração e fígado, conforme descrito por Olsson e colaboradores (OLSSON et al., 2002). De fato, neste mesmo trabalho, os autores indicam que a proteína TCF25 fora, no passado, classificada como pertencente à família de proteínas básicas hélice-volta-hélice (bHLH, *basic helix-loop-helix*), as quais contêm fatores de transcrição envolvidos na diferenciação e morfogênese durante o desenvolvimento de diversos órgãos. Posteriormente, outro estudo observou que a superexpressão de *TCF25* de humanos inibiu a atividade transcricional do fator de resposta ao soro (SRF, *serum response factor*), sugerindo seu papel como repressor na via de sinalização do SRF (CAI et al., 2006). Mais recentemente, *TCF25* foi também associada à regulação da expressão do fator nuclear de células T ativadas (NFAT, *nuclear factor of activated T cells*) em cardiomiócitos no contexto de hipertrofia cardíaca patológica (ZHANG et al., 2020).

Desta forma, nota-se que Rqc1p e suas proteínas homólogas desempenham funções celulares bastante diversas, possivelmente reguladas por condições celulares específicas. Enquanto no citoplasma de *S. cerevisiae* a Rqc1p atua no controle da síntese de proteínas associado à via do complexo RQC, no núcleo, a proteína pode desempenhar papéis na regulação transcricional, segundo evidências coletadas de outros organismos. Entretanto, ainda permanece pouco compreendido como essas funções são reguladas e em quais contextos celulares a proteína exerce determinada função. Diante deste cenário, a compreensão dos mecanismos que

determinam seus papéis, especialmente em vias de resposta ao estresse, pode fornecer novos entendimentos sobre sua regulação e até possivelmente sobreposição nos processos de vigilância da tradução e transcrição.

2 OBJETIVOS

2.1 Objetivo geral

Caracterizar a estrutura e função da proteína Rqc1 de *S. cerevisiae* e avaliar o seu papel nas vias de resposta aos estresses traducional e térmico.

2.2 Objetivos específicos

- Verificar a influência do estresse traducional e do choque térmico na viabilidade celular de *S. cerevisiae* desprovida dos genes *RQC1*, *RQC2* e *LTN1*, que estão envolvidos no controle de qualidade em síntese de proteínas;
- Caracterizar a proteína Rqc1 por meio de análises estruturais *in silico* e identificar resíduos conservados e que coevoluíram em seus ortólogos;
- Investigar o impacto do choque térmico na viabilidade celular de *S. cerevisiae rqc1Δ* transformada com plasmídeos que expressam tanto a versão completa da proteína como suas variantes truncadas;
- Analisar a expressão de genes relacionados ao complexo RQC em condição de choque térmico, utilizando dados de sequenciamento de RNA de *S. cerevisiae* WT depositados em bancos de dados públicos.

3 MATERIAL E MÉTODOS

3.1 Caracterização funcional da proteína Rqc1

3.1.1 Linhagens de *Escherichia coli* e *Saccharomyces cerevisiae*

A linhagem de *E. coli* utilizada para a obtenção do DNA plasmidial foi a RR1 (HANAHAN, 1983). A linhagem parental e mutantes de *S. cerevisiae* utilizadas neste estudo foi a BY4741 e seus genótipos estão descritos na Tabela 1. As linhagens mutantes de levedura que possuem a deleção de um gene foram doadas pelo Prof. Dr. Luis Eduardo Soares Netto, do Instituto de Biociências da Universidade de São Paulo, e a linhagem mutante que possui a deleção de dois genes foi previamente gerada pelo nosso próprio grupo (VALDEZ, 2018).

Tabela 1. Descrição dos genótipos de *S. cerevisiae* utilizados neste estudo.

Linhagem	Genótipo	Referência
WT	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	BRACHMANN et al., 1998
<i>ltn1Δ</i>	<i>MATa: his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; ltn1::KanMX6</i>	BRAUN et al., 2007
<i>rqc1Δ</i>	<i>MATa: his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; rqc1::KanMX6</i>	BRANDMAN et al., 2012
<i>rqc1Δltn1Δ</i>	<i>MATa: his3Δ1; leu2Δ0; met15Δ0; ura3Δ0 rqc1::KanMX6; ltn1Δ::His3MX6</i>	VALDEZ, 2018
<i>rqc1Δ</i> promotor+ terminador	<i>MATa: his3Δ1; leu2Δ0; met15Δ0; ura3Δ0 rqc1::KanMX6</i> YCplac111-Rqc1p	Este estudo
<i>rqc1Δ</i> Rqc1p	<i>MATa: his3Δ1; leu2Δ0; met15Δ0; ura3Δ0 rqc1::KanMX6</i> YCplac111-Rqc1p	Este estudo
<i>rqc1Δ</i> Rqc1p ¹⁻³²⁷	<i>MATa: his3Δ1; leu2Δ0; met15Δ0; ura3Δ0 rqc1::KanMX6</i> YCplac111- <i>rqc1Δ</i> Rqc1p ¹⁻³²⁷	Este estudo
<i>rqc1Δ</i> Rqc1p ³²⁸⁻⁷²³	<i>MATa: his3Δ1; leu2Δ0; met15Δ0; ura3Δ0 rqc1::KanMX6</i> YCplac111-Rqc1p ³²⁸⁻⁷²³	Este estudo

3.1.2 Meios de cultura, cultivo e armazenamento

Os meios de cultura utilizados para *E. coli* foram (I) LB líquido (triptona 1%, NaCl 1% e extrato de levedura 0,5%); (II) LB sólido (LB líquido e ágar bacteriológico 1,5%) e (III) LB sólido suplementado com ampicilina (100 µg/mL), todos esterilizados em vapor úmido a 121°C, por 20 minutos. O cultivo das células foi realizado em erlenmeyers de 250 mL fechados com tampas de estopa envoltas em gaze contendo

50 mL de meio LB líquido a 37°C, em incubadora com agitação orbital a aproximadamente 150 rpm.

Os meios de cultura utilizados para *S. cerevisiae* foram (i) YPD líquido (peptona 2%, extrato de levedura 1% e glicose 2%); (ii) YPD sólido (YPD líquido e ágar bacteriológico 1,5%); (iii) SD líquido desprovido de leucina (base nitrogenada para leveduras com sulfato de amônio 0.67%, glicose 2%, sulfato de adenina 10 mg/L, uracila 20 mg/L, triptofano 50 mg/L, histidina 20 mg/L, arginina 50 mg/L, tirosina 50 mg/L, lisina 50 mg/L, fenilalanina 50 mg/L, glutamato 1 g/L, aspartato 1 g/L, valina 140 mg/L, treonina 100 mg/L, serina 4 g/L, metionina 20 mg/L, isoleucina 50 mg/L e asparagina 80 mg/L); e (iv) SD sólido desprovido de leucina (SD líquido suplementado, exceto leucina, e ágar bacteriológico 1,5%), todos esterilizados em vapor úmido a 121°C, por 20 minutos. O cultivo das células foi realizado em erlenmeyers fechados com tampas de estopa envoltas em gaze contendo volumes variáveis de cultura entre 10 a 30 mL, a 28°C ou 38°C, em incubadora operando sob constante agitação orbital a aproximadamente 120 rpm. As pré-culturas eram realizadas entre 16 a 24 horas e o número de células inoculadas por mL de meio fresco variava entre 4×10^6 ($Abs_{600} = 0,005$), 8×10^6 ($Abs_{600} = 0,01$), $1,6 \times 10^7$ ($Abs_{600} = 0,02$), 8×10^7 ($Abs_{600} = 0,1$) e $1,6 \times 10^8$ ($Abs_{600} = 0,2$), a depender do experimento. Os estoques das linhagens foram mantidos a -80°C em meio YPD ou SD líquido desprovido de leucina contendo 25% de glicerol, e a 4°C, em placas de Petri contendo YPD ou SD sólido.

3.1.3 Determinação da tolerância celular ao estresse traducional e ao choque térmico

A determinação da tolerância celular de *S. cerevisiae* ao estresse traducional e térmico foi verificada através da capacidade de formação de *spots* em meio sólido (*Spot-Test*) (KWOLEK-MIREK; ZADRAG-TECZA, 2014). Para isso, as células foram cultivadas a 28°C em meio rico (YPD) ou meio sintético desprovido de leucina (SD-Leu) líquido por até 1 dia. Um volume da cultura de células foi coletado, centrifugado a 1000 x g, a 25°C por 1 minuto e ressuspensão em 1 mL de água ultrapura estéril. A etapa de lavagem foi repetida e as células foram ressuspensas em um volume final de 200 ou 500 µL de água ultrapura estéril. Em seguida, a absorbância (Abs_{600}) foi ajustada inicialmente a 0,2 e foram feitas diluições seriadas até 0,00002 (com fator de diluição de 10). Um volume de 3 ou 5 µL de todas as

diluições foi sequencialmente adicionado em placas contendo YPD sólido suplementado com higromicina B (75, 100 e 125 µg/mL) e SD-Leu sólido, a depender do experimento. As placas foram incubadas a 28°C (controle) e 38°C (choque térmico) no qual foram fotografadas após aproximadamente 48 horas de crescimento.

3.1.4. Determinação da curva de crescimento celular

As curvas de crescimento celular de *S. cerevisiae* cultivadas em condição de estresse traducional e térmico foram realizadas através do monitoramento e obtenção dos valores de Abs_{600} das culturas líquidas. Para os ensaios de crescimento celular em meio YPD contendo higromicina B, a cultura inicial foi realizada com Abs_{600} de 0,1 e as medidas foram feitas em um intervalo de 2 h de crescimento, por até 10 h. Já para os ensaios de crescimento celular em meio YPD sob choque térmico, a Abs_{600} inicial foi de 0,0025 e as alíquotas foram retiradas a cada 1 h no intervalo entre 12 a 24 h de crescimento.

3.1.5 Determinação da velocidade específica máxima de crescimento

Para determinação da velocidade específica máxima de crescimento ($\mu^{m\acute{a}x}/h$) de *S. cerevisiae*, foi plotado inicialmente um gráfico do logaritmo natural da concentração celular em Abs_{600} pelo tempo em horas. O $\mu^{m\acute{a}x}/h$ corresponde ao coeficiente angular da regressão linear obtida através do trecho de pontos da curva de crescimento em logaritmo natural (DORAN, 2004).

3.1.6 Desenho dos primers

A sequência do gene *RQC1* utilizada para o desenho dos primers foi obtida do SGD (*Saccharomyces Genome Database*). Todos os primers foram desenhados utilizando o programa SnapGene Viewer (<https://www.snapgene.com/>) e analisados através da ferramenta Oligocalc (KIBBE, 2007), sendo a síntese dos oligonucleotídeos realizada pela empresa Exxtend (Paulínia, SP). As sequências e a finalidade de cada primer estão descritas na Tabela 2 e 3.

Tabela 2. Primers utilizados para as clonagens. Os sítios de restrição reconhecidos pelas enzimas de restrição estão destacados em negrito e a sequência que codifica para Flag está sublinhada.

Identificador	Sequência nucleotídica (5' para 3')	Finalidade
pRQC1_FW_(HindIII)	ATGCA AAGCTT CTTTTGAAC TTCAAGGACTTCGCTGC	Clonagem promotor
pRQC1_RV_(PstI)	GTAC CTGCAGG TTTCGAGTACTTTACAATATATTTAGATGATTCAACGACC	
RQC1_FW_ALL_(PstI)	ATCG CTGCAGATGAGCTCTAGAGCATTAAAGGAGATTA C	Clonagem Rqc1p e Rqc1p ¹⁻³²⁷
RQC1_RV_ALL_(KpnI)	GTAC GGTACCTTATTTATCGTCATCGTCTTTATAATCAC CCTCATTTCATTGACTCTTCATTTCATG	Clonagem Rqc1p e Rqc1p ³²⁸⁻⁷²³
RQC1_NT_RV_(KpnI)	TAC GGTACCTTATTTATCGTCATCGTCTTTATAATCAT AAAAAGAAATGAATTTTGGCACTGACTG	Clonagem Rqc1p ¹⁻³²⁷
RQC1_CT_FW_(PstI)	ATCG CTGCAGATGAAGTTCGAACCTTTAAATTCCGAC CTGAGC	Clonagem Rqc1p ³²⁸⁻⁷²³
tRQC1_FW_(KpnI)	ATCG GGTACCTAAGAAATGTTGCAGCTCATTTCCTAGT ACGC	Clonagem terminador
tRQC1_RV_(EcoRI)	CGAT GAATTC CGGACTTGGAAAGGACAAACAATGC	

Tabela 3. Primers utilizados para o sequenciamento.

Finalidade	Sequência nucleotídica (5' para 3')
Sequenciamento YCplac111+Rqc1p	CTCGTAAGTTGTCGCGAAACGC
	CGAGGCTGCAAAGACAAGGG
	CCAAAACGATGTCAGTCAGTGGC
	GAATAAATGAGATGAGTTCCGCCCG
	CGCCAGCTGGCGTAATAGCGAAG
Sequenciamento YCplac111+Rqc1p ¹⁻³²⁷	CTCGTAAGTTGTCGCGAAACGC
	CGAGGCTGCAAAGACAAGGG
	CGCCAGCTGGCGTAATAGCGAAG
Sequenciamento YCplac111+Rqc1p ³²⁸⁻⁷²³	CTCGTAAGTTGTCGCGAAACGC
	GAATAAATGAGATGAGTTCCGCCCG

3.1.7 Extração do DNA genômico de *S. cerevisiae*

A extração de DNA genômico de *S. cerevisiae* foi adaptada segundo descrito por Harju e colaboradores (HARJU; FEDOSYUK; PETERSON, 2004). Um volume de 1mL de cultura de *S. cerevisiae* S288C crescida *overnight* em meio YPD foi

centrifugado a 1000 x *g*, a 25°C por 1 min e o sobrenadante foi descartado. As células foram ressuspensas em tampão de lise (2% de triton X-100, 1% de SDS, 100 mM NaCl, 10 mM Tris-HCl [pH 8.0] e 1 mM EDTA [pH 8.0]) e incubadas por 4 min a -80°C sendo transferidas para o banho seco a 95°C por 2 min. O processo foi repetido e o conteúdo foi homogeneizado em vórtex por 30 s. Posteriormente, foram adicionados 200 µL de clorofórmio ao microtubo que foi homogeneizado em vórtex por 2 min e centrifugado a 12000 x *g*, a 25°C por 5 min. A fase aquosa foi transferida para um microtubo contendo 400 µL de etanol 100% gelado que foi homogeneizado por inversão, incubado a 25°C por 5 min e centrifugado a 12000 x *g*, a 25°C por 5 min. O sobrenadante foi removido e o *pellet* de DNA foi seco por 15 min a 45°C e depois ressuspense em 100 µL de tampão TE (10 mM Tris [pH 8.0], 1 mM EDTA [pH 8.0]). A quantificação do DNA genômico obtido foi realizada utilizando o espectrofotômetro NanoDrop ND-1000 (Thermo Scientific).

3.1.8 Amplificação dos fragmentos por reação da polimerase em cadeia (PCR)

Para a amplificação dos fragmentos por PCR, o DNA genômico extraído conforme descrito no Item 3.2.7 foi utilizado como molde. As reações para amplificação dos insertos correspondentes à sequência do promotor e do gene truncado (somente a porção N-terminal, ou C-terminal) foram preparadas para um volume final de 50 µL contendo: tampão de PCR 10X (1X), dNTP 10 mM (0,2 mM), MgCl₂ 50 mM (1,5 mM), primers FW e RV 10 µM cada (0,2 µM cada), DNA genômico S288C (1 µg), Platinum™ Taq DNA Polimerase - Invitrogen (1 U) e água Milli-Q estéril. Já para amplificação do gene completo e do terminador, as reações preparadas para o volume final de 50 µL foram: tampão Phusion GC 5X (1X), dNTP 10 mM (0,2 mM), primers FW e RV 10 µM cada (0,5 µM cada), DNA genômico S288C (1 µg), DMSO (3%), Phusion High-Fidelity DNA Polymerase - New England Biolabs (1 U) e água Milli-Q estéril. As reações foram realizadas utilizando o termociclador MJ Mini (Bio-Rad) e os programas gerais utilizados conforme os fabricantes estão descritos na Tabela 4.

Tabela 4. Programas utilizados para amplificação dos fragmentos utilizados para clonagem.

Ciclos	Etapa	Platinum (Invitrogen) - 10966	Phusion HF (NEB) - M0530S
1	Desnaturação inicial	94°C 2 min	98°C 1 min
30	Desnaturação	94°C 30 seg	98°C 10 seg
	Hibridização	55°C 30 seg	55-58°C 30 seg
	Extensão	72°C 1 min/kb	72°C 30 seg/kb
1	Extensão final	72°C 5 min	72°C 5 min
∞	<i>Hold</i>	8°C	8°C

3.1.9 Purificação dos produtos de PCR e digestão

Para purificação, os produtos de PCR e digestão foram submetidos a eletroforeses de DNA em gel de agarose 1% com tampão TAE (Tris-Cl 40 mM [pH = 8,3], ácido acético 20 mM, e EDTA 1 mM) corado com brometo de etídio (0,5 µg/mL). Para isto, as amostras juntamente com tampão de carregamento, foram adicionadas aos poços do gel de agarose, que foi submetido a eletroforese a 100 V. Após o tempo necessário, o DNA foi visualizado no transiluminador T26M (BioAgency) operando a 302 nm, e as bandas de interesse foram cortadas com bisturi. As bandas foram inseridas em membranas de celulose contendo tampão TAE 0,1 X e estas foram vedadas em suas extremidades com grampos. O sistema montado foi colocado em uma cuba de eletroforese submergido com tampão TAE 0,1 X e a eletroeluição do DNA foi realizada a 300 V por 30 min. Após a eletroeluição, o tampão contido na membrana de celulose – que contém o DNA de interesse – foi coletado e transferido para microtubos de centrífuga. A partir disso, foi adicionado um volume de fenol aos microtubos, que foram então agitados fortemente e centrifugados a 14000 rpm, a 25°C por 3 min. A fase aquosa foi coletada e foi adicionado um volume de clorofórmio aos microtubos, com subsequente agitação e centrifugação tal como anteriormente. A fase aquosa foi transferida a novos microtubos e 1/20 volume de NaCl 5 M e 2,5 volumes de etanol 100% foram adicionados aos sistemas, que foram homogeneizados por inversão. Os microtubos foram incubados no gelo seco por cerca de 10 min e centrifugados a 14000 rpm, a 25°C por 5 min. O sobrenadante foi descartado e o *pellet* que contém o DNA precipitado foi lavado duas vezes com 1 mL de etanol 80%. Após a lavagem, o *pellet* foi seco a 37°C e ressuspenso em uma quantidade variável de água Milli-Q estéril.

3.1.10 Digestão enzimática e ligação dos fragmentos

De forma geral, as digestões enzimáticas dos fragmentos e do plasmídeo YCplac111 foram realizadas em um volume final de 100 μL utilizando aproximadamente 10 U de enzima de restrição para concentrações variadas de DNA purificado. As digestões dos fragmentos correspondentes às sequências do promotor (HindIII e PstI [New England Biolabs]), do gene *RQC1* completo e das porções truncadas (PstI e KpnI [New England Biolabs]) foram realizadas por até 90 min a 37°C. Já a digestão do fragmento do terminador (KpnI e EcoRI [FastDigest/Thermo Fisher]) foi feita por 15 min a 37°C devido à observação de atividade estrela em tempos maiores. Para a digestão do plasmídeo YCplac111 foi adicionado 10 U de fosfatase alcalina (Roche) após 10 min de incubação. Após as digestões os produtos foram purificados, conforme descrito no Item 3.1.9, e a ligação foi realizada através da incubação de 1 U de T4 DNA ligase a temperatura ambiente por 1 h em um volume final de 10 μL .

3.1.11 Transformação de *E. coli* competente

Para a transformação, o protocolo de competência das bactérias foi conduzido utilizando um volume de 50 mL de cultura de *E. coli* RR1 crescida por até 4 h em meio LB. As células então foram centrifugadas a 6000 x *g*, a 4°C por 5 min, tendo sido o sobrenadante descartado e as células ressuspensas em 10 mL de tampão I (CaCl₂ 10 mM, Tris-Cl 10 mM [pH = 7,5]) gelado. O processo de ressuspensão foi repetido, agora utilizando tampão II (CaCl₂ 30 mM, Tris-Cl 10 mM [pH = 7,5]), e as células foram incubadas por 30 min no gelo. Após este tempo, as células foram ressuspensas em 5 mL de tampão II e foram utilizadas para transformação. Um volume de 200 μL de células competentes foi adicionado a 10 μL de ligação e os tubos foram incubados no gelo por 10 min. Posteriormente, o choque térmico foi realizado a 42°C por 2 min no banho molhado e após alguns segundos foi adicionado 1 mL de meio LB. Os tubos foram incubados a 37°C por 30 min sendo homogeneizados por inversão a cada 10 min. Após incubação, os tubos foram centrifugados a 1000 x *g*, a 25°C por 1 min, e o sobrenadante parcialmente retirado até restar, aproximadamente, 200 μL de suspensão celular que foi plaqueada em LB suplementado com ampicilina (100 $\mu\text{g}/\text{mL}$). As placas foram incubadas *overnight* a 37°C até o aparecimento de colônias transformantes que foram estriadas em

pequenos tapetes celulares em novas placas contendo antibiótico para posterior testagem.

3.1.12 Identificação de clones positivos e sequenciamento das construções

As células crescidas em pequenos tapetes conforme descrito no Item 3.1.11, foram coletadas com palito de madeira e adicionadas à microtubos com 200 µL de solução I (50 mM glicose, 10 mM EDTA e 25 mM Tris-Cl [pH 8;0]) contendo 5 mg/mL de lisozima e 2 µg/mL de RNase A. Foram também adicionados aos microtubos 200 µL da solução II (0,2 M NaOH 1% SDS) e 150 µL da solução III (7,5 M NH₄OAc e ácido acético glacial 2%). Após homogeneização por inversão após a adição das soluções, os microtubos foram centrifugados a 14000 rpm, a 25°C por 3 min e os sobrenadantes foram transferidos para novos tubos contendo 300 µL da solução IV (isopropanol 100% e 0,2% de triton X-100). As amostras foram homogeneizadas por inversão, centrifugadas a 14000 rpm, a 25°C por 5 min e o sobrenadante foi descartado. O *pellet* que contém o DNA precipitado foi lavado duas vezes com 1 mL de etanol 80%, seco a 37°C e ressuspenso em 5 µL de água Milli-Q estéril. O DNA plasmidial obtido de cada colônia foi digerido e o produto da digestão foi analisado através de eletroforese em gel de agarose 1%. Após a observação e identificação das bandas esperadas, uma colônia positiva de cada construção foi selecionada e estriada cobrindo a área total de uma placa contendo meio LB suplementado com ampicilina. O tapete celular resultante do crescimento das células na placa mencionada acima foi coletado com o auxílio de uma espátula e adicionado a tubo com 2 mL de solução I (0,25 M sacarose, 0,125 M EDTA e 25 mM Tris [pH 8;0]) contendo 5 mg/mL de lisozima e 10 mg/mL de RNase A. A homogeneização foi realizada utilizando vórtex por alguns segundos e as células foram incubadas por 30 min no gelo. Após incubação, foi adicionado 1 mL de solução II (10% triton X-100, 0,5 M EDTA, 1 M Tris [pH 8]) e as células foram centrifugadas a 65000 rpm, a 4°C por 10 min. O sobrenadante foi coletado e o DNA plasmidial foi extraído utilizando o protocolo de extração por fenol-clorofórmio com modificações (Item 3.1.9). Foi adicionado um volume de fenol, sendo os tubos, em seguida, agitados fortemente e centrifugados a 4000 rpm, a 4°C por 5 min. A fase aquosa foi coletada e foi adicionado um volume de clorofórmio aos tubos, os quais foram agitados fortemente e centrifugados tal como anteriormente. A fase aquosa foi coletada novamente e 1/20 volume de NaCl 5 M e 2,5 volumes de etanol 100% foram adicionados aos

tubos, os quais foram homogeneizados por inversão e centrifugados por 10 min. O sobrenadante foi descartado e foi adicionado 1,5 mL de NH₄OAc e 4,5 mL de etanol 100%. Os tubos foram homogeneizados por inversão e centrifugados por 10 min. A etapa anterior foi repetida, mas com a adição de 300 µL de NH₄OAc e 1 mL de etanol 100% aos tubos que foram centrifugados por 5 min. O sobrenadante foi descartado e o *pellet* que contém o DNA plasmidial precipitado foi lavado duas vezes com 1 mL de etanol 80% a 14000 rpm, a 25°C por 3 min. Após a lavagem, o *pellet* foi seco a 37°C e ressuspenso em 100 µL de água Milli-Q estéril.

A partir do DNA plasmidial obtido em grande quantidade de cada colônia selecionada, foram realizadas digestões diagnósticas para cada construção. As digestões foram realizadas semelhante ao descrito no Item 3.1.10 e estão descritas na Tabela 5. Os produtos da digestão foram analisados em gel de agarose 1%.

Tabela 5. Descrição das digestões diagnósticas realizadas para cada construção.

Construção	Enzima de restrição (tamanho das bandas esperadas em pb)
YCplac111+terminador	BstXI (3676 pb, 2389 pb e 526 pb), BgIII (4596 pb e 1998 pb) e BstII+NsiI (3742 pb e 2849 pb)
YCplac111+promotor+terminador	EcoRI (6602 pb e 983 pb), Scal (3862 pb, 2725 pb e 995 pb), AflII (3832 pb, 3540 pb e 213 pb) e PstI e KpnI (7558 pb e 27 pb)
YCplac111+Rqc1p	AflII (4742 pb, 3540 pb, 1265 pb e 213 pb), BstBI (4986 pb, 4283 pb e 491 pb), ClaI (5623 pb e 4137 pb)
YCplac111+Rqc1p ¹⁻³²⁷	XbaI +NdeI (7701 pb e 871 pb), PstI+BamHI (7736 pb e 831 pb), PstI+BstEII (4870 pb e 3702 pb) e PstI+KpnI (7558 pb e 1014 pb)
YCplac111+Rqc1p ³²⁸⁻⁷²³	HindIII (7225 pb e 1557 pb), Scal (3862 pb, 2725 pb, 1636 pb e 556 pb), ClaI (4645 pb e 4137 pb) e PstI+KpnI (7558 pb e 1224 pb)

Após a confirmação das bandas esperadas, o DNA plasmidial de cada colônia positiva foi extraído utilizando o kit PureYield™ Plasmid Miniprep System (Promega). O DNA plasmidial, juntamente com os *primers* correspondentes para cada construção (Tabela 3), foi então enviado para o Centro de Pesquisa sobre o Genoma Humano e Células-Tronco (USP), onde foi realizado o sequenciamento de DNA pelo método de Sanger.

3.1.13 Transformação de *S. cerevisiae* WT e *rqc1Δ*

A transformação de *S. cerevisiae* foi realizada segundo descrito por Amberg e colaboradores com algumas modificações (AMBER; BURKE; STRATHERN, 2006).

Um volume de células de *S. cerevisiae* WT e *rqc1Δ* cultivadas *overnight* em meio YPD foi coletado e centrifugado a 1000 x *g*, a 25°C por 1 min. As células foram lavadas em 1 mL de acetato de lítio 0,1 M, centrifugadas como anteriormente e ressuspensas em uma solução de transformação (480 µL de PEG 3350 50%, 60 µL de acetato de lítio 1 M e 20 µL de DNA esperma de salmão [10 mg/mL] desnaturado a 100°C por 10 minutos). Foi então adicionado, aos microtubos, aproximadamente 100 ng de plasmídeo, que foram homogeneizados gentilmente e incubados a 42°C por 45 min em banho-maria. Os tubos foram centrifugados, os sobrenadantes foram descartados e as células foram ressuspensas em 1 mL de meio YPD e incubadas a 28°C por 2 h. Após a incubação, as células foram centrifugadas e o *pellet* celular foi lavado em 1 mL de água Milli-q estéril. Uma centrifugação adicional foi realizada, sendo o sobrenadante retirado deixando somente 50 µL finais. Essa suspensão celular foi plaqueada em meio SD desprovido de leucina, a marca de seleção do plasmídeo YCplac111, e as placas foram incubadas a 28°C pelo tempo necessário para o aparecimento de colônias transformantes.

3.1.14 Análises estatísticas e geração de gráficos

As análises estatísticas e os gráficos foram feitos utilizando o programa GraphPad Prism 8 (GraphPad Software).

3.2 Caracterização estrutural *in silico* da proteína Rqc1

3.2.1 Alinhamento de sequências

As sequências das proteínas da família TCF25 utilizadas para as análises *in silico* foram obtidas no banco de dados Uniprot (THE UNIPROT CONSORTIUM et al., 2025). Para a busca, foram utilizadas palavras-chave como “tcf25”, “rqc1” e “nulp1” juntamente com o nome da espécie dos organismos modelos eucarióticos que representam a diversidade biológica. É importante ressaltar que em situações nas quais havia mais de uma entrada para a mesma espécie, foi escolhida a sequência que possuía revisão pelo *Swiss-Prot*, seguida de melhor escore de anotação e nível de existência da proteína. As sequências selecionadas estão listadas na Tabela 6.

Tabela 6. Descrição geral das sequências das proteínas ortólogas de organismos modelo selecionadas do Uniprot.

Código Uniprot	Espécie	Status de revisão	Escore de anotação	Existência da proteína	Tamanho (aminoácidos)
Q05468	<i>Saccharomyces cerevisiae</i>	Swiss-Prot	4/5	Evidência a nível de proteína	723
Q9BQ70	<i>Homo sapiens</i>	Swiss-Prot	5/5	Evidência a nível de proteína	676
Q8R3L2	<i>Mus musculus</i>	Swiss-Prot	5/5	Evidência a nível de proteína	676
A0A8V0Y0J9	<i>Gallus gallus</i>	TrEMBL	1/5	Evidência a nível de proteína	661
Q68EJ4	<i>Danio rerio</i>	TrEMBL	1/5	Evidência a nível de proteína	650
A3KMT4	<i>Xenopus laevis</i>	TrEMBL	1/5	Evidência a nível de transcrito	675
Q9Y109	<i>Drosophila melanogaster</i>	TrEMBL	2/5	Evidência a nível de transcrito	702
P90919	<i>Caenorhabditis elegans</i>	TrEMBL	1/5	Evidência a nível de proteína	647
Q8T2A4	<i>Dictyostelium discoideum</i>	Swiss-Prot	3/5	Inferida por homologia	740
O80734	<i>Arabidopsis thaliana</i>	TrEMBL	1/5	Evidência a nível de proteína	627

O alinhamento múltiplo das sequências foi feito utilizando a ferramenta Muscle disponibilizada pelo EMBL-EBI (*European Bioinformatics Institute*) (MADEIRA et al., 2024). Os alinhamentos foram tratados e visualizados no Jalview v. 2.11.4.0 (WATERHOUSE et al., 2009).

3.2.2 Predição de localização subcelular e sinais de classificação

Para a predição da localização subcelular e presença de sinais de classificação foi utilizado o servidor *web* DeepLoc 2.1, que utiliza uma abordagem de aprendizado profundo para classificação de sequências (ØDUM et al., 2024). Os dados de entrada foram um arquivo no formato FASTA com as sequências das proteínas ortólogas de organismos modelo. Os resultados gerados incluíam tabelas de pontuação de previsão para dez localizações subcelulares e nove sinais de classificação, dentre eles sinais de localização nuclear (NLS, *nuclear localization*

signal) e sinais de exportação nuclear (NES, *nuclear export signal*). As possíveis localizações das proteínas preditas foram destacadas em verde quando ultrapassaram o limiar de probabilidade sendo 0,4761 para localização citoplasmática e 0,5014 para nuclear. Além disso, foi gerado um gráfico que indica as regiões da sequência que são importantes para as previsões listadas.

Para uma análise mais aprofundada das possíveis NLSs identificadas pelo DeepLoc 2.1, foi utilizada a ferramenta cNLS Mapper, que prevê NLSs específicas para a via de importação mediada por importina $\alpha\beta$ (KOSUGI et al., 2009). O cNLS Mapper calcula escores somando as contribuições positivas ou negativas de cada resíduo de aminoácido com base nos valores definidos para perfis de NLS já estudados em leveduras. As sequências das proteínas estudadas foram inseridas individualmente e os parâmetros selecionados tiveram escore de corte de 3,0 e busca por NLSs em toda a extensão da sequência. As NLSs identificadas e seus respectivos escores foram analisados da seguinte forma: escores entre 8 e 10 indicam proteínas com localização nuclear, entre 7 e 8 indicam localização parcial no núcleo, entre 3 e 5 correspondem a proteínas presentes tanto no núcleo quanto no citoplasma.

3.2.3 Estrutura secundária

A análise da estrutura secundária de Rqc1p foi realizada utilizando a ferramenta NetSurfP-3.0 (HØIE et al., 2022). Além de prever a estrutura secundária de proteínas a partir da sua sequência, a ferramenta também gera informações sobre a acessibilidade relativa à superfície (RSA, *relative surface accessibility*) com valores de RSA >0.25 indicando resíduos expostos e RSA <0.25 indicando resíduos enterrados.

3.2.4 Modelagem molecular e validação estrutural

Para modelagem molecular utilizamos duas abordagens distintas: (i) ColabFold v1.5.5, uma ferramenta integrada ao *Google Collaboratory* que combina a busca acelerada por sequências homólogas via MMseqs2 com a predição estrutural baseada no AlphaFold2, que emprega metodologias baseadas em inteligência artificial, aprendizado profundo e redes neurais (JUMPER et al., 2021; MIRDITA et al., 2022); e (ii) I-TASSER, que prediz estruturas combinando o método de *threading*, montagem por fragmentos e refinamento para otimização da estrutura, podendo

recorrer à modelagem *ab initio* quando não há modelos disponíveis (ROY; KUCUKURAL; ZHANG, 2010).

Para isso, preparamos o arquivo de entrada contendo a sequência completa de Rqc1p (723 aminoácidos) e também suas versões truncadas Rqc1p¹⁻³²⁷, que contém a região polibásica, e Rqc1p³²⁸⁻⁷²³, que contém o domínio TCF25. Os três melhores modelos gerados pelo ColabFold foram selecionados de acordo com o *rank* dos valores de pLDDT (*predicted local distance difference test*), que indica a confiabilidade através de uma escala de 0 a 100, sendo que regiões acima de 90 são consideradas de alta confiança e abaixo de 50, de baixa confiança. Já os três melhores modelos gerados pelo I-TASSER foram selecionados com base no valor de *C-score*, que é um indicador de confiança do modelo, variando entre -5 e 2, no qual valores mais altos indicam maior confiança e maior similaridade com estruturas conhecidas.

A validação de todas as estruturas selecionadas foi feita através das ferramentas disponíveis na plataforma *online* SAVES v.6.1 (<https://saves.mbi.ucla.edu/>) sendo elas: (i) ERRAT, que avalia a estabilidade da conformação dos resíduos por meio das interações átomo-átomo não ligadas (COLOVOS; YEATES, 1993); (ii) VERIFY3D, que analisa a compatibilidade entre a estrutura tridimensional e sua estrutura primária, verificando os ambientes locais das cadeias laterais (EISENBERG; LÜTHY; BOWIE, 1997) e (iii) PROCHECK, que gera gráficos de Ramachandran para identificação de resíduos de aminoácidos em regiões desfavoráveis ou não permitidas de acordo com seus ângulos diedros *phi* e *psi* (LASKOWSKI et al., 1993).

3.2.5 Decomposição de redes de coevolução de resíduos

O alinhamento múltiplo de sequências (MSA, *multiple sequence alignment*) da família do domínio TCF25 (código de entrada PF04910) foi obtido no Pfam utilizando o tipo de alinhamento “Uniprot” que continha 3.278 sequências. Para as análises de conservação e decomposição de redes de coevolução de resíduos (DRCN, *decomposition of residue coevolution networks*), foi utilizado o *software* PFstats (FONSECA-JÚNIOR et al., 2018). De forma geral, através deste método, é construída uma rede de correlação obtida através dos sinais de correlação ou anticorrelação entre pares de resíduos de uma família de proteínas em uma determinada posição. Posteriormente, esta rede é decomposta em comunidades de

resíduos coevolúidos que podem ser identificados como relevantes para a estrutura e função de uma família de proteínas (BLEICHER; LEMKE; GARRATT, 2011).

Após a obtenção do MSA, as 3.278 sequências foram filtradas no PFstats utilizando a abordagem por perfil HMM (*hidden markov model*) e os fragmentos e sequências redundantes foram removidos utilizando os filtros de 80% de cobertura mínima, e 80% de identidade máxima, respectivamente. Após a filtragem, permaneceram 740 sequências e os parâmetros utilizados para as análises de conservação e para montagem e decomposição da rede de coevolução foram 0,80 para o escore mínimo de conservação, 10 para o escore mínimo de correlação, 15% de sub-alinhamento mínimo e 0,2 de variação de frequência.

3.2.6 Análise de similaridade estrutural

A análise de similaridade estrutural utilizando como dado de entrada o modelo gerado pelo ColabFold contendo o domínio TCF25 (Rqc1p³²⁸⁻⁷²³) foi realizada utilizando o servidor *online* FoldSeek (VAN KEMPEN et al., 2024). Foram consideradas nas análises somente os 10 primeiros *hits* com estruturas resolvidas experimentalmente que foram depositadas no PDB100.

3.2.7 Visualização das estruturas tridimensionais

As estruturas tridimensionais geradas foram visualizadas e analisadas utilizando o programa UCSF ChimeraX (MENG et al., 2023).

3.3 Metanálise dos transcriptomas de *S. cerevisiae* em choque térmico

3.3.1 Aquisição dos dados de sequenciamento de RNA e análise da qualidade das *reads*

Os dados de sequenciamento de RNA (RNA-seq) de *S. cerevisiae* cultivada em condição de choque térmico foram retirados do maior repositório de dados brutos públicos de sequenciamento, o SRA (*Sequence Read Archive*), mantido pelo NCBI (*National Center for Biotechnology Information*). Para a busca, foram utilizados os termos “*Saccharomyces cerevisiae*” e “*heat shock*”, o que resultou em 28 estudos após a filtragem por somente transcriptomas até setembro de 2024. Para garantir maior homogeneidade e robustez nas análises, foi realizada uma curadoria manual a partir dos 28 estudos para selecionar somente os que possuíam tipo de leitura

paired-end e com pelo menos 3 replicatas biológicas para cada condição. Ao final obtivemos 7 estudos e o *download* de todas as bibliotecas foi realizado utilizando a ferramenta *fasterq-dump* v.3.0.1, do pacote de SRA Toolkit disponível no NCBI.

Para a análise da qualidade das bibliotecas de RNA-seq foi utilizada a ferramenta *FastQC* v.0.11.9 (BABRAHAM BIOINFORMATICS) e a compilação e visualização dos resultados foi realizada utilizando o programa *MultiQC* v.1.13.dev0 (EWELS et al., 2016). Ao final, 4 dos 7 estudos pré-selecionados apresentaram homogeneidade entre a profundidade das bibliotecas e valor de qualidade das *reads* igual ou superior a *phred* 30, que indica a probabilidade de ocorrer erro em 1 nucleotídeo a cada 1000 nucleotídeos sequenciados.

3.3.2 Mapeamento no transcriptoma de referência e contagem das *reads*

Os mapeamentos das *reads* das bibliotecas de RNA-seq foram realizados contra o genoma de referência de *S. cerevisiae* da linhagem S288C (código de entrada no NCBI: GCF_000146045.2) utilizando o programa *STAR* v.2.7.10b (DOBIN et al., 2013). A conversão e ordenação dos arquivos com a extensão de BAM para SAM foi realizada com a ferramenta *Samtools* v.1.16.1 (DANECEK et al., 2021). Já a contagem do número de *reads* mapeadas por gene foi realizada utilizando o *FeatureCounts* v.2.0.3 (LIAO; SMYTH; SHI, 2014), utilizando as tags *-p*, *-t* CDS, *-g* *gene_id*; identificando as *reads* como *paired-end*, as *features* como CDS e retornando os *gene_id* como *meta-features*.

Para a visualização dos padrões de variação entre as condições experimentais de cada estudo individualmente, foi realizada a Análise de Componentes Principais (PCA). O *output* do *FeatureCounts* foi importado para o programa R e utilizado para a geração de uma matriz e foi executada a função *prcomp*, do pacote *stats*, para a realização da análise de PCA.

3.3.3 Análises da expressão diferencial de genes

As análises da expressão diferencial de genes foram realizadas no programa R utilizando o pacote *DESeq2* v.1.40.2 (LOVE; HUBER; ANDERS, 2014). Para considerar um gene como diferencialmente expresso foram adotados como critérios o valor de corte de *log2FoldChange* $\geq |1|$, indicando uma alteração de pelo menos duas vezes na expressão em comparação ao controle, e um valor de *p* ajustado (*padj*) $< 0,05$.

3.3.4 Recursos computacionais

O processamento dos dados e as análises dos transcriptomas foram realizadas nos *clusters* computacionais Gloriosos, ICB-UFMG.

4 RESULTADOS

4.1 A deleção de *LTN1* e *RQC1* reduz a viabilidade celular de *S. cerevisiae* na presença de higromicina B

Com o objetivo de entender o papel das proteínas Rqc1p e Ltn1p na resposta ao estresse traducional, realizamos um ensaio de tolerância celular à higromicina B, no qual verificamos, através de *Spot-Test*, o crescimento de *S. cerevisiae* em placas contendo meio YPD suplementadas com diferentes concentrações deste antibiótico. A higromicina B interage com os sítios A e P do ribossomo, impedindo a translocação, o que gera a parada da tradução (BOROVINSKAYA et al., 2008), sendo células de *S. cerevisiae* desprovidas de *LTN1* conhecidamente sensíveis à este antibiótico (BENGTSON; JOAZEIRO, 2010).

Podemos notar que, de fato, as células desprovidas de Ltn1p possuem menor tolerância ao estresse traducional induzido por higromicina B, sendo esse fenótipo dose-dependente. Além disso, observamos que a ausência de Rqc1p também promove uma maior sensibilidade de *S. cerevisiae* à presença desse antibiótico em relação à célula WT, porém em menor extensão do que aquela verificada para os mutantes *ltn1Δ* (Fig. 6).

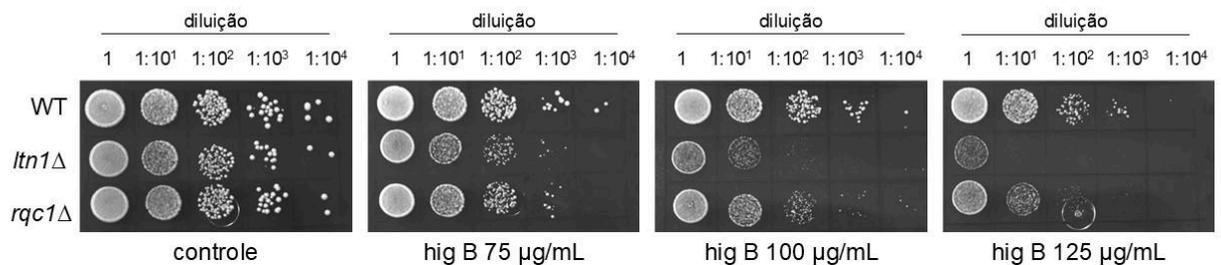


Fig. 6. Tolerância ao estresse traducional induzido por higromicina B de *S. cerevisiae* WT, *rqc1Δ* e *ltn1Δ*. A determinação da tolerância ao estresse traducional de *S. cerevisiae* através da capacidade de crescimento em meio YPD suplementado com higromicina B foi realizado conforme descrito em Materiais e Métodos (Item 3.1.3). As placas foram fotografadas após 48h de crescimento a 28°C. A Fig. é representativa de 4 replicatas biológicas.

4.2 O impacto da higromicina B na velocidade específica máxima de crescimento é acentuado na ausência de *LTN1*

Após a determinação da tolerância ao estresse traducional induzido por higromicina B através do ensaio de *Spot-Test* (Fig. 6), e sabendo que Ltn1p possui um papel central na resposta induzida por este antibiótico, decidimos quantificar o impacto da ausência do gene *LTN1* nas taxas de crescimento celular na presença de higromicina B através da determinação da velocidade específica máxima de

crescimento celular ($\mu^{\text{máx}}/\text{h}$) em meio YPD líquido em diferentes concentrações deste antibiótico. Foram realizadas curvas de crescimento celular de *S. cerevisiae* WT e *ltn1Δ* em que foram obtidos os valores de $\text{Abs}_{600\text{nm}}$ das culturas ao longo de 10 horas.

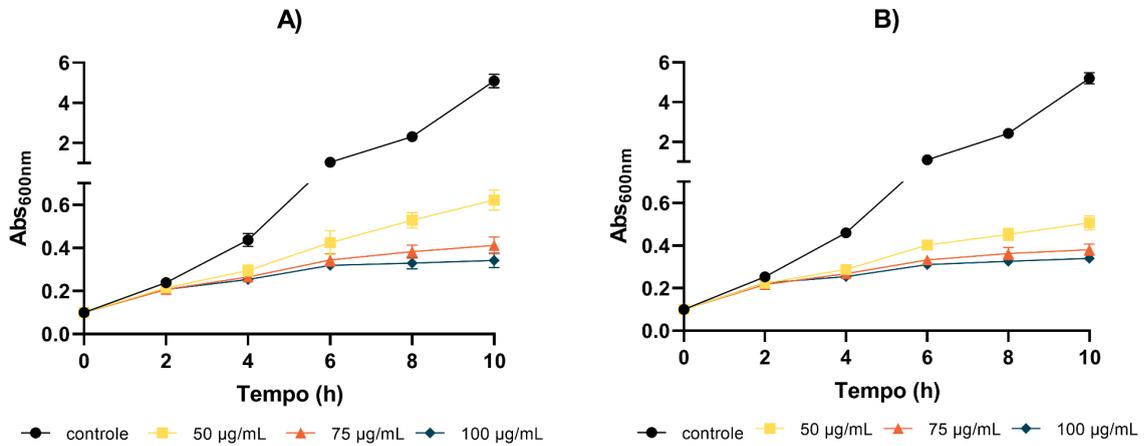


Fig. 7. Curva de crescimento celular de *S. cerevisiae* WT e *ltn1Δ* em diferentes concentrações de higromicina B. A determinação do crescimento celular de *S. cerevisiae* WT (painel A) e *ltn1Δ* (painel B) foi realizada através de medidas de $\text{Abs}_{600\text{nm}}$ conforme descrito em Materiais e Métodos (Item 3.1.4). Os dados são representativos de 3 replicatas biológicas.

De forma geral, observamos que a higromicina B inibe o crescimento de *S. cerevisiae* WT e *ltn1Δ* a partir de 2 h de crescimento celular já na concentração mínima utilizada de 50 $\mu\text{g/mL}$ (Fig. 7). Além disso, vemos que o antibiótico parece modular o crescimento celular de maneira distinta ao longo do tempo, já que a taxa de crescimento no intervalo de 2 a 6 h parece ser maior do que no intervalo de 6 a 10 h em ambas as linhagens. Para investigar o impacto da ação da higromicina B no crescimento celular nestes intervalos, calculamos a $\mu^{\text{máx}}/\text{h}$ em cada concentração do antibiótico utilizada.

Observamos então que, no intervalo de 2 a 6 h, as células desprovidas de *Ltn1p* possuem $\mu^{\text{máx}}/\text{h}$ menor quando cultivadas em meio contendo 100 $\mu\text{g/mL}$ de higromicina B (Fig. 8, painel D). Já no intervalo de 6 a 10 horas, vemos que as células *ltn1Δ* possuem $\mu^{\text{máx}}/\text{h}$ menor quando cultivadas em meio contendo 50 $\mu\text{g/mL}$ (Fig. 8, painel B).

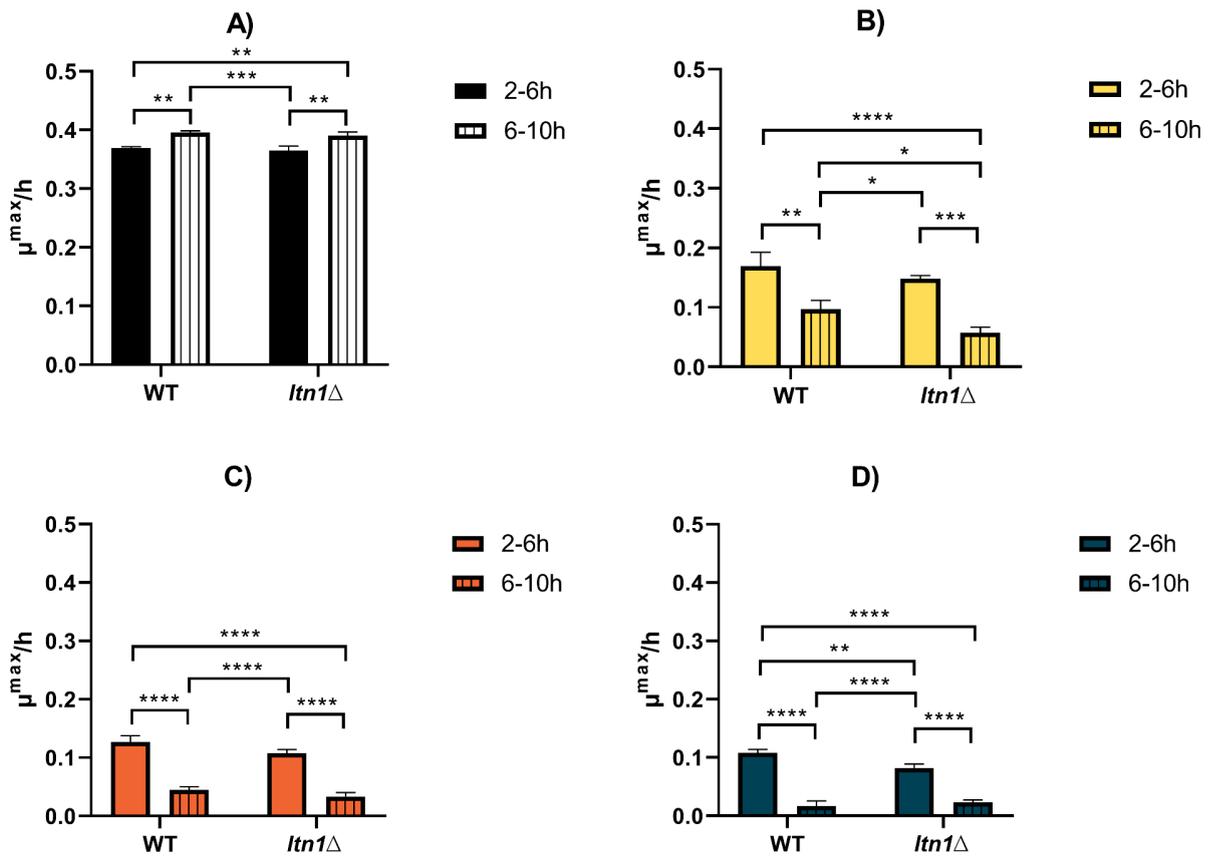


Fig. 8. Velocidade específica máxima de crescimento celular de *S. cerevisiae* WT e *ltn1\Delta* em higromicina B. Os cálculos para determinação da μ^{max}/h na ausência (painel A) e na presença de higromicina B 50 $\mu g/mL$ (painel B), 75 $\mu g/mL$ (painel C) e 100 $\mu g/mL$ (painel D) no intervalo de 2 a 6 horas e de 6 a 10 horas de crescimento foram realizados conforme descrito em Materiais e Métodos (Item 3.1.5). * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$ e **** $p < 0,0001$ (two-way ANOVA/Tukey).

4.3 A deleção de *RQC1* compromete a viabilidade celular de *S. cerevisiae* sob choque térmico enquanto a deleção concomitante de *LTN1* restaura o fenótipo selvagem

Sabendo que a HSR é a principal via ativada em condições de estresse celular, realizamos um ensaio de tolerância celular por *Spot-Test* para verificar o crescimento de *S. cerevisiae* WT, *rqc2\Delta*, *ltn1\Delta* e *rqc1\Delta* – mutantes desprovidos de Rqc2p, Ltn1p e Rqc1p, respectivamente, proteínas centrais na via RQC – em placas contendo meio de cultura YPD em condições-controle (28°C) e também em choque térmico (38°C).

Observamos que, dentre as três proteínas estudadas, somente a Rqc1p está envolvida com a adequada resposta celular de *S. cerevisiae* ao choque térmico, uma vez que a sua ausência promoveu marcante sensibilidade à essa condição (Fig. 9).

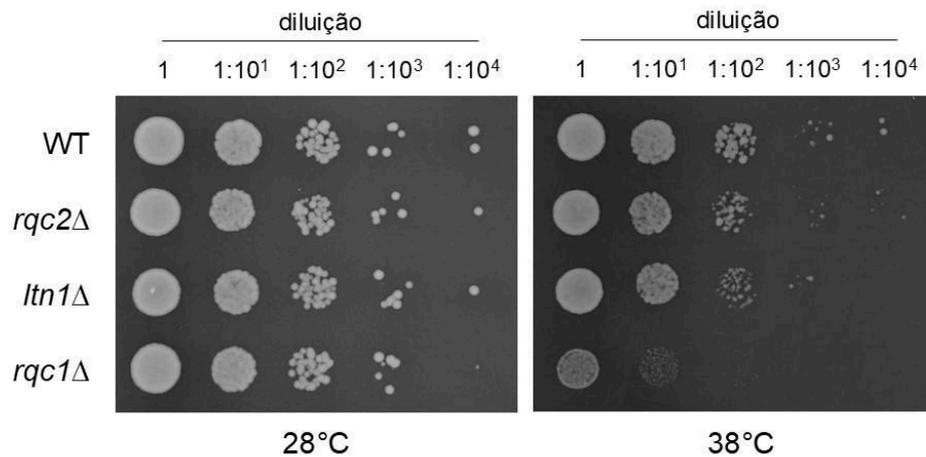


Fig. 9. Tolerância ao choque térmico de *S. cerevisiae* WT, *rqc2Δ*, *ltn1Δ* e *rqc1Δ*. A determinação da tolerância ao choque térmico de *S. cerevisiae* através da capacidade de crescimento em meio YPD sob choque térmico foi realizado conforme descrito em Materiais e Métodos (Item 3.1.3). As placas foram fotografadas após 44h de incubação à 28°C e à 38°C. A Fig. é representativa de 3 replicatas biológicas.

Diante deste resultado, decidimos também submeter o duplo mutante *rqc1Δltn1Δ* ao mesmo teste, na tentativa de verificar a existência de uma possível interação gênica entre *RQC1* e *LTN1* na resposta ao choque térmico. Curiosamente, o aumento da sensibilidade ao choque térmico em células *rqc1Δ* é revertido na ausência de *LTN1*, sendo o fenótipo selvagem restaurado nestes mutantes (Fig. 10).

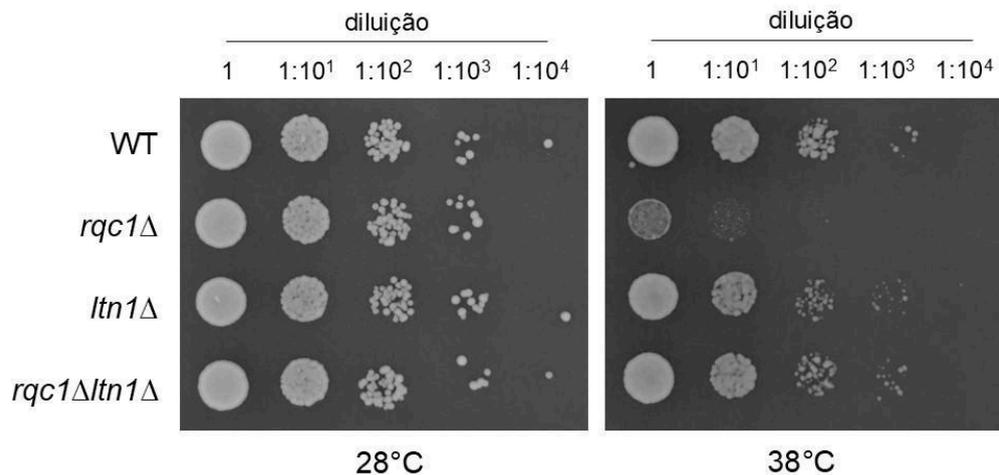


Fig. 10. Tolerância ao choque térmico de *S. cerevisiae* WT, *rqc1Δ*, *ltn1Δ* e *rqc1Δltn1Δ*. A determinação da tolerância ao choque térmico de *S. cerevisiae* através da capacidade de crescimento em meio YPD sob choque térmico foi realizado conforme descrito em Materiais e Métodos (Item 3.1.3). As placas foram fotografadas após 44h de incubação à 28°C e à 38°C. A Fig. é representativa de 3 replicatas biológicas.

4.4 A deleção do gene *RQC1* reduz a velocidade específica máxima de crescimento celular em choque térmico

Após observar que a proteína Rqc1p apresenta um papel relevante para a viabilidade celular de *S. cerevisiae* em choque térmico através do ensaio do *Spot-Test* (Fig. 9), buscamos investigar se o cultivo celular em meio líquido, em condições de choque térmico, altera a $\mu^{\text{máx}}/\text{h}$ de *S. cerevisiae* WT, *rqc1* Δ , *ltn1* Δ e *rqc1* Δ *ltn1* Δ .

Podemos observar que a biomassa gerada após 24 h de cultivo em choque térmico é menor do que aquela observada em condição-controle (Fig. 11, painéis A e B); de fato, o choque térmico reduziu significativamente o $\mu^{\text{máx}}/\text{h}$ de todas as células quando o cultivo foi conduzido a 38°C (Fig. 11, painéis C-F). No entanto, verificamos que, em condições de choque térmico, somente o mutante desprovido de Rqc1p apresentou uma diminuição significativa na $\mu^{\text{máx}}/\text{h}$ em relação às demais células estudadas (Fig. 11, painel H), resultado que se encontra em concordância com aqueles obtidos através da verificação do crescimento celular em meio sólido (Figs. 9 e 10).

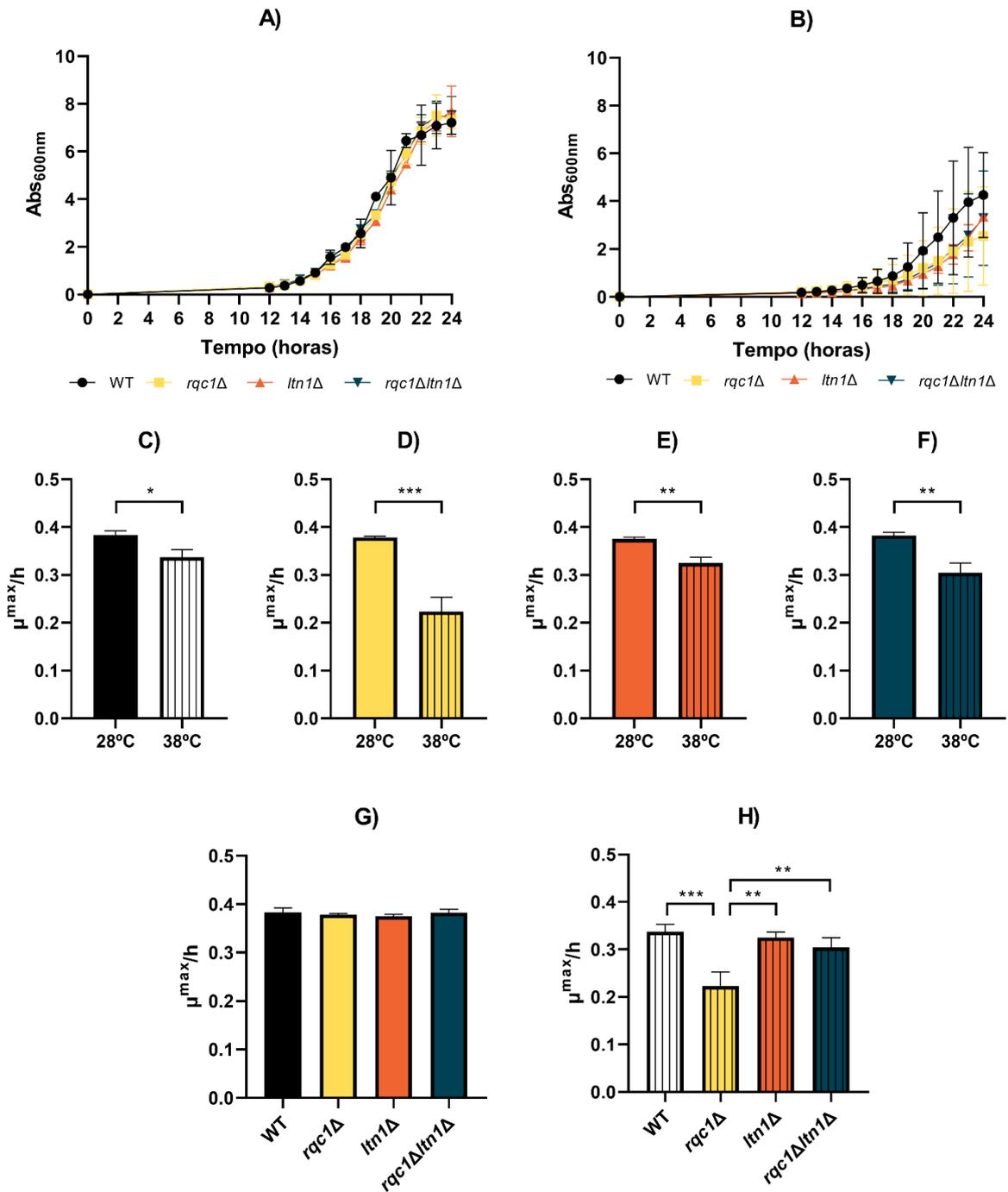


Fig. 11. Curva de crescimento e velocidade específica máxima de crescimento celular de *S. cerevisiae* WT, *rqc1*Δ, *ltn1*Δ e *rqc1*Δ*ltn1*Δ em choque térmico. A determinação do crescimento celular em 28°C (painel A) e 38°C (painel B) foi realizada através de medidas de Abs₆₀₀ conforme descrito em Materiais e Métodos (Item 3.1.4). Os cálculos para determinação da velocidade específica máxima de crescimento celular de WT (painel C), *rqc1*Δ (painel D), *ltn1*Δ (painel E), *rqc1*Δ*ltn1*Δ (painel F) e a 28°C (painel G) e 38°C (painel H) foram realizados conforme descrito em Materiais e Métodos (Item 3.1.5). Os dados são representativos de 3 replicatas. Painéis C a F: *p < 0,05; **p < 0,01; ***p < 0,001 vs. 28°C (teste t de Student não-pareado). Painéis G e H: (one-way ANOVA/Tukey).

4.5 Rqc1p apresenta uma região polibásica N-terminal e um domínio C-terminal conservado em eucariotos

A utilização de abordagens computacionais para o estudo de proteínas mal caracterizadas vem sendo amplamente utilizada nos últimos anos, e um de seus objetivos é o de delinear estudos experimentais futuros. Sabendo-se que a sequência de resíduos de aminoácidos de uma proteína – ou seja, a sua estrutura primária – determina sua estrutura tridimensional e, conseqüentemente, sua possível função biológica, torna as análises computacionais da estrutura primária, secundária e terciária essenciais para inferir funções biológicas de domínios proteicos (CHAKRABORTY; CHAKRABARTI, 2014).

O alinhamento de duas ou mais sequências de resíduos de aminoácidos de proteínas homólogas pode auxiliar na identificação de regiões de conservação. Sendo assim, a sequência de Rqc1p de *S. cerevisiae* e suas proteínas homólogas de organismos modelos foram alinhadas e identificadas as regiões de maior conservação entre os eucariotos (Apêndice I).

A primeira região, localizada próximo ao N-terminal da proteína, é caracterizada pela presença de resíduos de lisinas e argininas, denominada região polibásica (Fig. 12, painel A). Já a segunda região, localizada próxima ao C-terminal da proteína, é caracterizada pela presença de resíduos hidrofóbicos e polares altamente conservados (Fig. 12, painel B). Essa região compreende um domínio mal caracterizado anotado como TCF25.

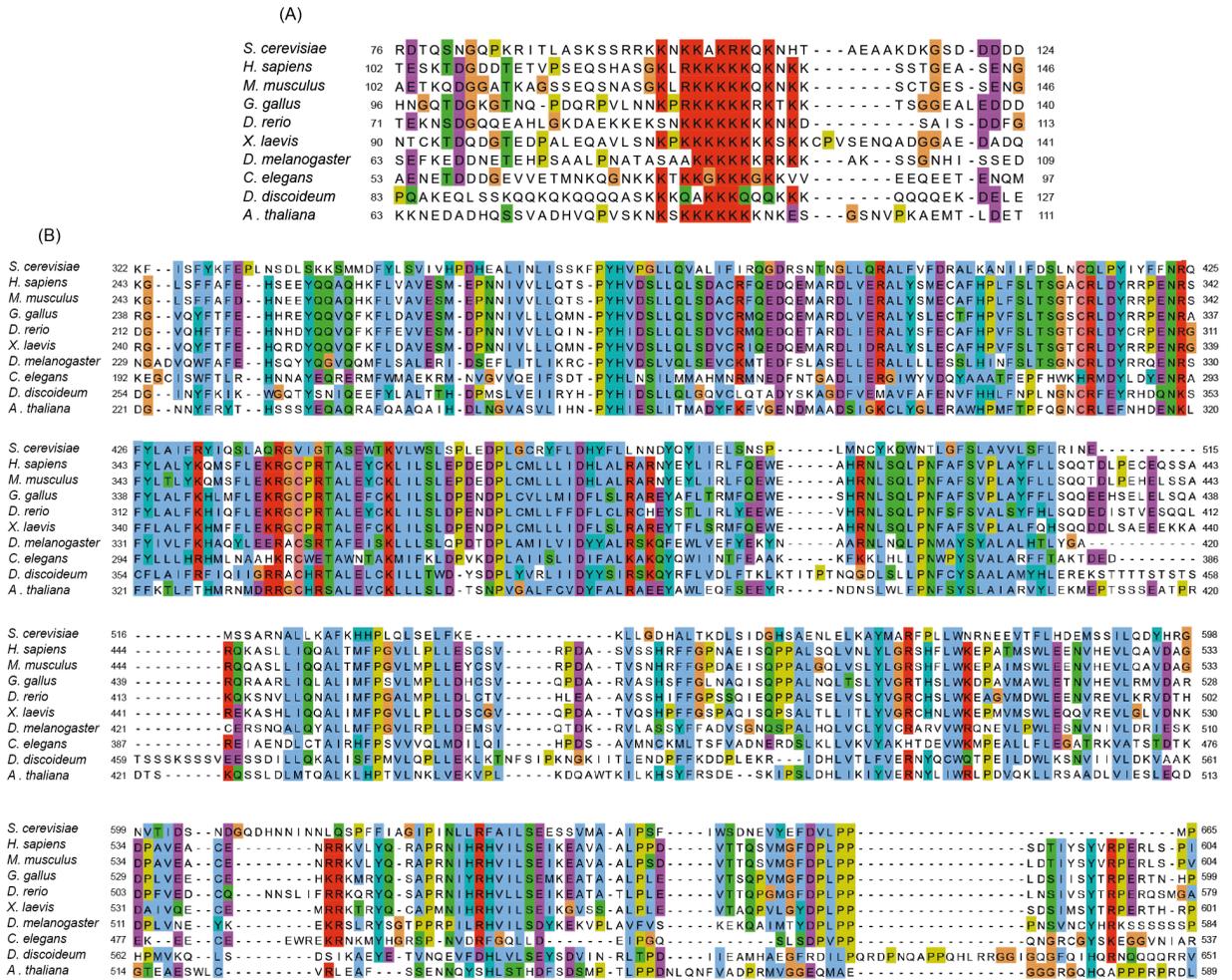


Fig. 12. Alinhamento múltiplo de seqüências de proteínas homólogas de Rqc1p com enfoque nas regiões de maior conservação. Os resíduos de aminoácidos estão destacados de acordo com sua classe sendo eles os hidrofóbicos (azul), com carga positiva (vermelho), com carga negativa (magenta), polares (verde) e aromáticos (ciano). Painel A: Alinhamento da porção N-terminal destacando a região polibásica altamente conservada entre os eucariotos. Painel B: Alinhamento da porção C-terminal destacando o domínio TCF25.

4.6 A predição de NLSs e a presença de sítios de fosforilação sugerem que Rqc1p pode ser translocada para o núcleo

A localização subcelular de uma proteína está diretamente relacionada à sua função; desta maneira, a identificação de sinais de localização e sítios de modificação pós-traducional são essenciais para entender sua dinâmica de translocação através dos compartimentos celulares (LANGE et al., 2007).

Utilizando a ferramenta DeepLoc 2.1, identificamos que Rqc1p e suas proteínas ortólogas são predominantemente localizadas no citoplasma. Além disso, a predição indicou uma probabilidade de localização parcial no núcleo para TCF25 de humano e camundongo (Apêndice II). Embora a probabilidade de Rqc1p para localização nuclear não ter atingido o limiar (0.5014), foi predito um NLS, sugerindo

que a proteína possa ser transportada para o núcleo sob certas condições (Fig. 13, painel A). Também foi observado que as regiões da sequência que contribuem para a previsão da localização e sinais de classificação estão localizadas próximas ao N-terminal, incluindo a região polibásica, que é conservada entre os eucariotos (Fig. 13, painel B).

A)

Espécie	Citoplasma	Núcleo	Localização	Sinais de classificação
<i>S. cerevisiae</i>	0.6625	0.4957	Citoplasma	Sinal de localização nuclear

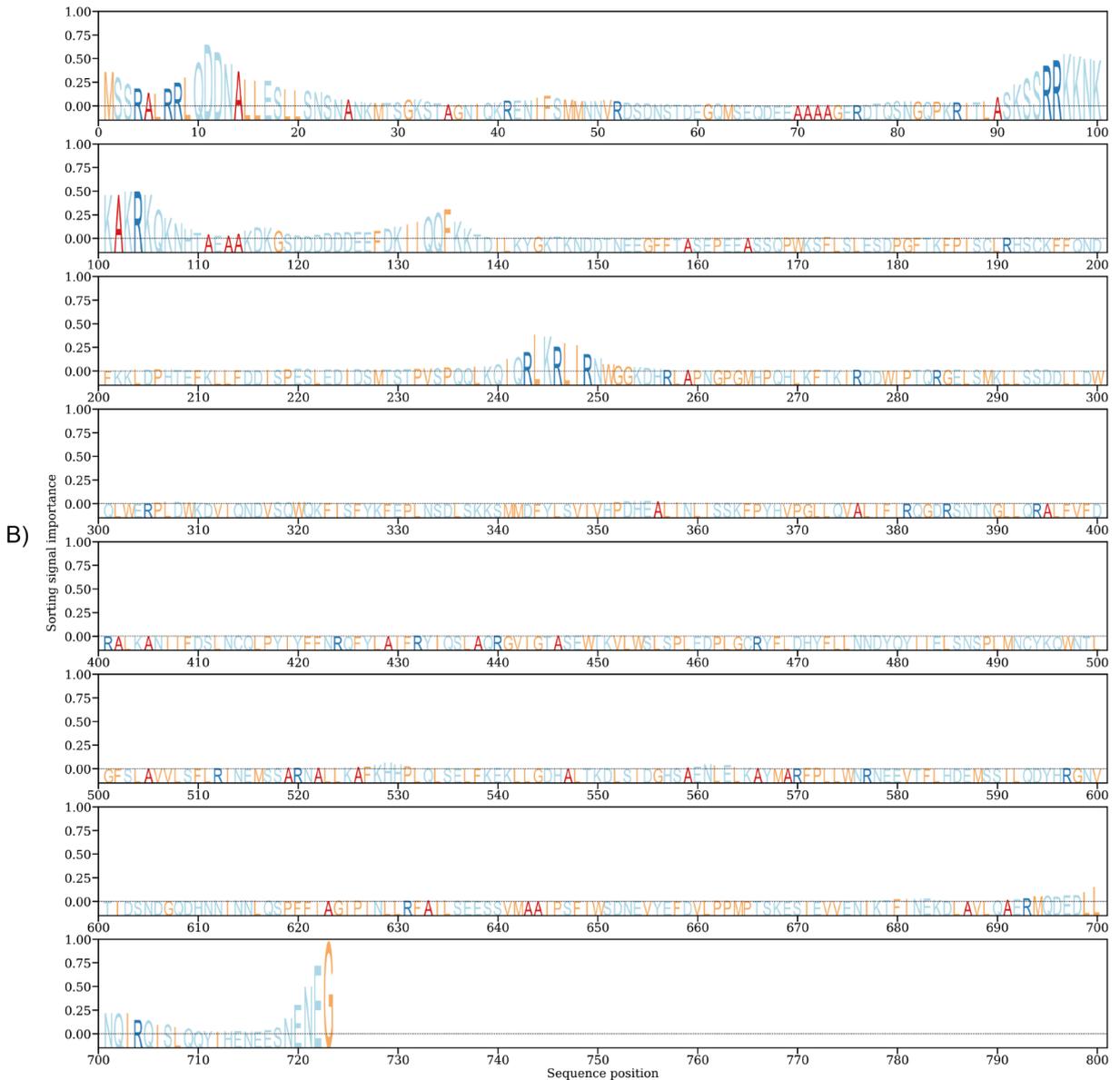


Fig. 13. Previsão da localização subcelular de Rqc1p de *S. cerevisiae* e sinais de classificação previstos pela ferramenta DeepLoc-2.1. A tabela mostra as pontuações de previsão de localização citoplasmática e nuclear e os sinais de classificação preditos (painel A). Os limiares de probabilidade acima de 0.4761 indica provável localização citoplasmática e acima de 0.5014 provável localização nuclear. As células em verde destacam os resultados das previsões e a intensidade da cor indica o nível de certeza. No gráfico, aminoácidos que possuem valores altos correspondem a regiões da sequência que contribuem para a previsão da localização subcelular e potencialmente estão relacionadas aos sinais de classificação (painel B).

Buscando aumentar a robustez das análises de caracterização dos sinais encontrados, utilizamos outra ferramenta, o cNLS Mapper, que prediz e classifica tipos de NLSs que são clássicos da via importina α/β . A análise identificou múltiplos NLSs em Rqc1p, incluindo um do tipo monopartido (Fig. 14, painel B) e diversos do tipo bipartido (Fig. 14, painel C). Notavelmente, o NLS predito como do tipo monopartido coincide com a sequência da região polibásica previamente identificada, reforçando sua potencial relevância para a translocação desta proteína para o núcleo.

Ademais, mapeamos os nove sítios de modificações pós-traducionais anotados para a proteína de acordo com o SGD. Notamos que dois resíduos de serina passíveis de fosforilação se encontram dentro de NLSs preditos como bipartidos (Fig. 14, painel A).

A) NLSs preditos na sequência		
MSSRALRRRLQDDNALLESLLSNSNANKMTSGK	STAGNIQKRENI	FSMMNN 50
VRDSNSTDEGQMS	EQDEEAAAAGERDTSNGQPKRITLASKSSRRKKNK	100
KAKRKQKNHTAEAAKDKGS	DDDDDEEFDKIIQQFKKTDILKYGKTKNDD	150
TNEEGFFT	ASEPEEASSQPWKSFLSLESDPGFTKFPISCLRHSCKFFQND	200
FKKLDPHTEFKLLFDDISPE	LEDIDSMTSTPVSPQQLKQIQRLKRLIRN	250
WGGKDHRLAPNGPGMHPQHLKFTKIRDDW	IPTQRGELSMKLLSSDDLDDW	300
QLWERPLDWKVDVIQNDV	SQWQKFISFYKFEPLNSDLSSKSMDFYLSVIV	350
HPDHEALINLIS	SKFPYHVPGLLQVALIFIRQGDRSNTNGLLQRALFVFD	400
RALKANII	FDLSLNCQLPYIYFFNRQFYLAIFRYIQSLAQRGVIGTASEWT	450
KVLWSLSPLEDPLGCRYFLDHYFLLNNDYQYIIELSNSPLMNCYQWNTL		500
GFSLAVVLSFLRINEMSSARNALLKAFKHHPLQLSELFKEKLLGDHALTK		550
DLSID	GHSAENLELKAYMARFPLLWNRNEEVTFLHDEMSSILQDYHRGNV	600
TIDSNDGQDHNINNLQSPFFIAGIPINLLRFAILSEESSVMAAIPSIW		650
SDNEVYEFDVLPPMPTSKESIEVVENIKTFINEKDLAVLQAERMQDELL		700
NQIRQISLQYIHENEES	NENEG	723

B) Monopartido		
Pos.	Sequência	Score
94	SRRKKNKA	5

C) Bipartido		
Pos.	Sequência	Score
5	ALRRLQDDNALLESLLSNSNANKMTSGK	3.3
75	ERDTSNGQPKRITLASKSSRRKKNKKA	3.7
83	QPKRITLASKSSRRKKNK	8.2
82	GQPKRITLASKSSRRKKNKKA	4.5
82	GQPKRITLASKSSRRKKNKAKRKQKNHTAE	3.5
103	KRKQKNHTAEAAKDKGSDDDDDEEFDKII	3.7
112	EAAKDKGSDDDDDEEFDKIIQQFKKTDI	5.1
112	EAAKDKGSDDDDDEEFDKIIQQFKKTDILK	4.5
241	IQRLKRLIRNWGGKDHRLAPNGPGMHPQHL	3.1
243	RLKRLIRNWGGKDHRLAPNGPGMHPQHLKFTKI	5.6
246	RLIRNWGGKDHRLAPNGPGMHPQHLKFTKIRDDW	3.5
520	RNALLKAFKHHPLQLSELFKEKLLGDHALTKDLSID	4.3

Fig. 14. Previsão de NLSs pela ferramenta cNLS Mapper. Os NLSs preditos do tipo monopartido (painel B) e bipartido (painel C) estão marcados em vermelho na sequência de entrada (painel A). Os resíduos passíveis de fosforilação estão destacados em azul (painel A).

4.7 A predição da estrutura secundária de Rqc1p indica a presença de regiões desordenadas no N-terminal e formação de múltiplas α -hélices no domínio TCF25

A estrutura secundária das proteínas compreende o conjunto de estruturas regulares que se formam a partir de ligações de hidrogênio entre os resíduos de aminoácidos de uma estrutura primária. As estruturas mais comuns formadas são α -hélices, folhas- β e voltas (BERG; TYMOCZKO; STRYER, 2013).

A predição da estrutura secundária realizada pelo NetSurfP-3.0 mostrou que Rqc1p é composta por regiões desordenadas, especialmente nas extremidades N- e C-terminal. Além disso, a proteína apresenta a formação de várias α -hélices e uma

folha- β , concentradas principalmente na região que corresponde ao domínio TCF25 (resíduos 328-663). A análise da RSA indicou que grande parte dos resíduos localizados na região N-terminal e próximos à extremidade C-terminal possuem valores elevados de RSA (>0.25), indicando resíduos que estão expostos ao solvente. Essas regiões coincidem com regiões preditas como altamente desordenadas, sugerindo flexibilidade estrutural (Fig. 15).

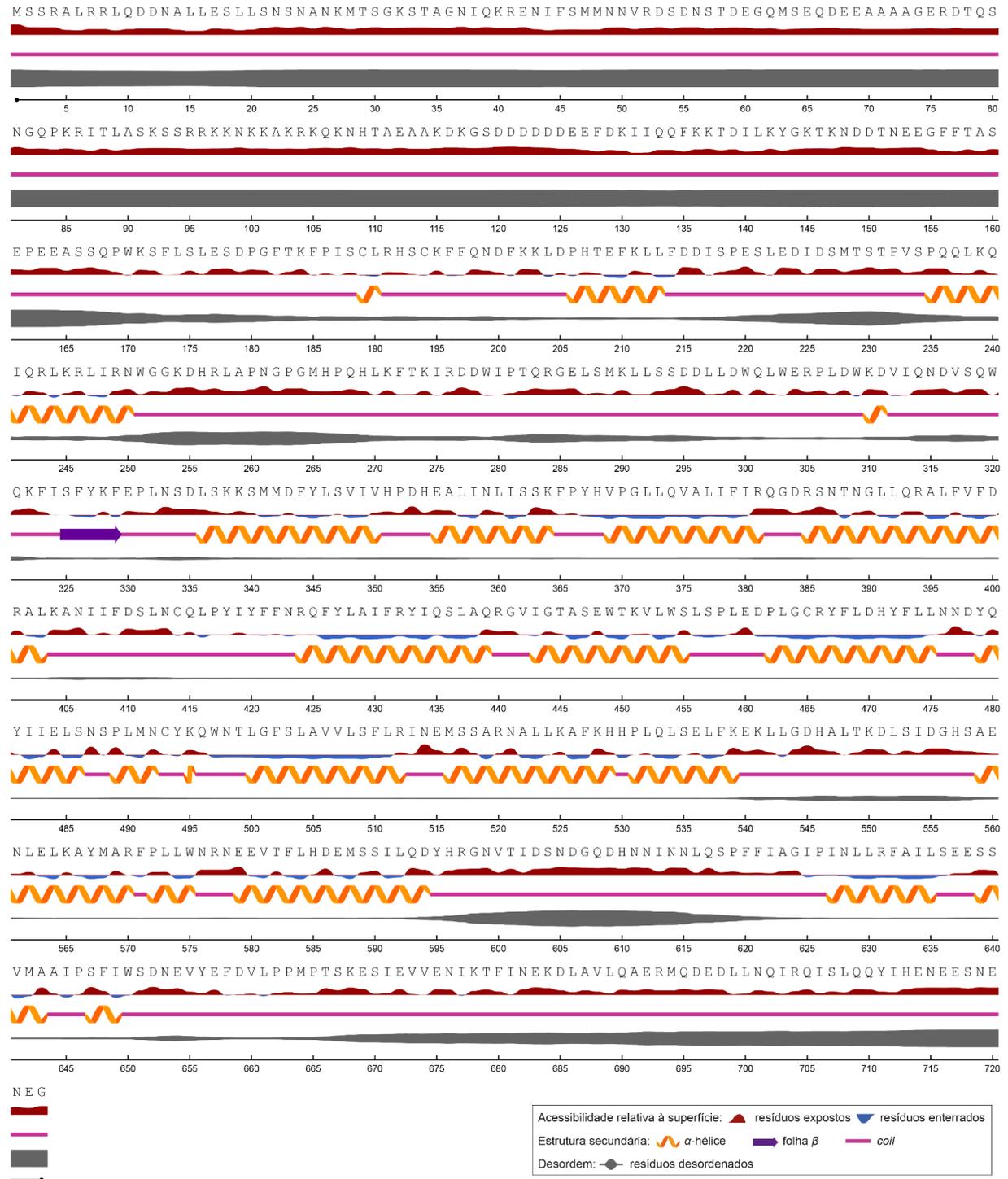


Fig. 15. Predição da estrutura secundária e de características estruturais de Rqc1p de *S. cerevisiae* utilizando a ferramenta NetSurfP-3.0. A Fig. representa a estrutura secundária podendo ser α -hélice (laranja), folha- β (roxo) e *coil* (rosa), além de informações sobre acessibilidade relativa à superfície sendo >0.25 para resíduos expostos (vermelho) e <0.25 resíduos enterrados (azul). As regiões desordenadas estão representadas em cinza com a espessura da linha indicando a probabilidade de desordem.

4.8 Modelos gerados *in silico* e validação estrutural

Uma vez que não há estrutura resolvida experimentalmente de Rqc1p ou suas proteínas ortólogas, utilizamos abordagens de modelagem molecular que não dependem exclusivamente de homologia, como o I-TASSER e o ColabFold, uma implementação otimizada do AlphaFold. O ColabFold foi utilizado para predição estrutural baseado em aprendizado de máquina, enquanto o I-TASSER foi empregado para modelagem baseada em *threading* e montagem de fragmentos. A utilização de abordagens independentes é importante, visto que minimiza as limitações de cada método e aumenta a robustez da predição estrutural dos modelos gerados.

Após a escolha dos melhores modelos preditos para Rqc1p e suas versões truncadas, Rqc1p¹⁻³²⁷ e Rqc1p³²⁸⁻⁷²³, realizamos a validação estrutural utilizando as ferramentas disponibilizadas pela plataforma SAVES (tabela 7).

Tabela 7. Descrição das métricas utilizadas para validação dos modelos gerados por abordagens computacionais.

Modelo	Obtenção	Regiões no gráfico de Ramachandran				ERRAT	Verify3D
		Mais favorecidas	Adicionalmente permitidas	Generosamente permitidas	Não permitidas		
Rqc1p	ColabFold	83.5%	13.2%	2.3%	1.1%	94.15	67.63%
	I-TASSER	78.0%	16.7%	3.0%	2.3%	90.74	44.40%
Rqc1p ¹⁻³²⁷	ColabFold	69.4%	22.2%	6.1%	2.4%	83.59	16.82%
	I-TASSER	87.9%	10.1%	1.3%	0.7%	92.16	76.15%
Rqc1p ³²⁸⁻⁷²³	ColabFold	89.6%	10.1%	0.3%	0.0%	96.23	75.51%
	I-TASSER	52.2%	33.3%	10.4%	4.1%	77.40	53.54%

De modo geral, os modelos gerados pelo ColabFold obtiveram melhores escores em todas as métricas analisadas, exceto para a modelagem de Rqc1p¹⁻³²⁷. No entanto, ao inspecionar visualmente os modelos selecionados, observamos que aqueles gerados pelo I-TASSER eram altamente heterogêneos entre si e apresentavam baixos valores de confiança (*C-score*), indicando baixa confiabilidade nas predições obtidas (Apêndice III). Dessa forma, seguimos as próximas análises utilizando os modelos gerados pelo ColabFold. Os modelos selecionados, assim

como as métricas obtidas pelo ERRAT e Ramachandran foram analisados separadamente (Fig. 16).

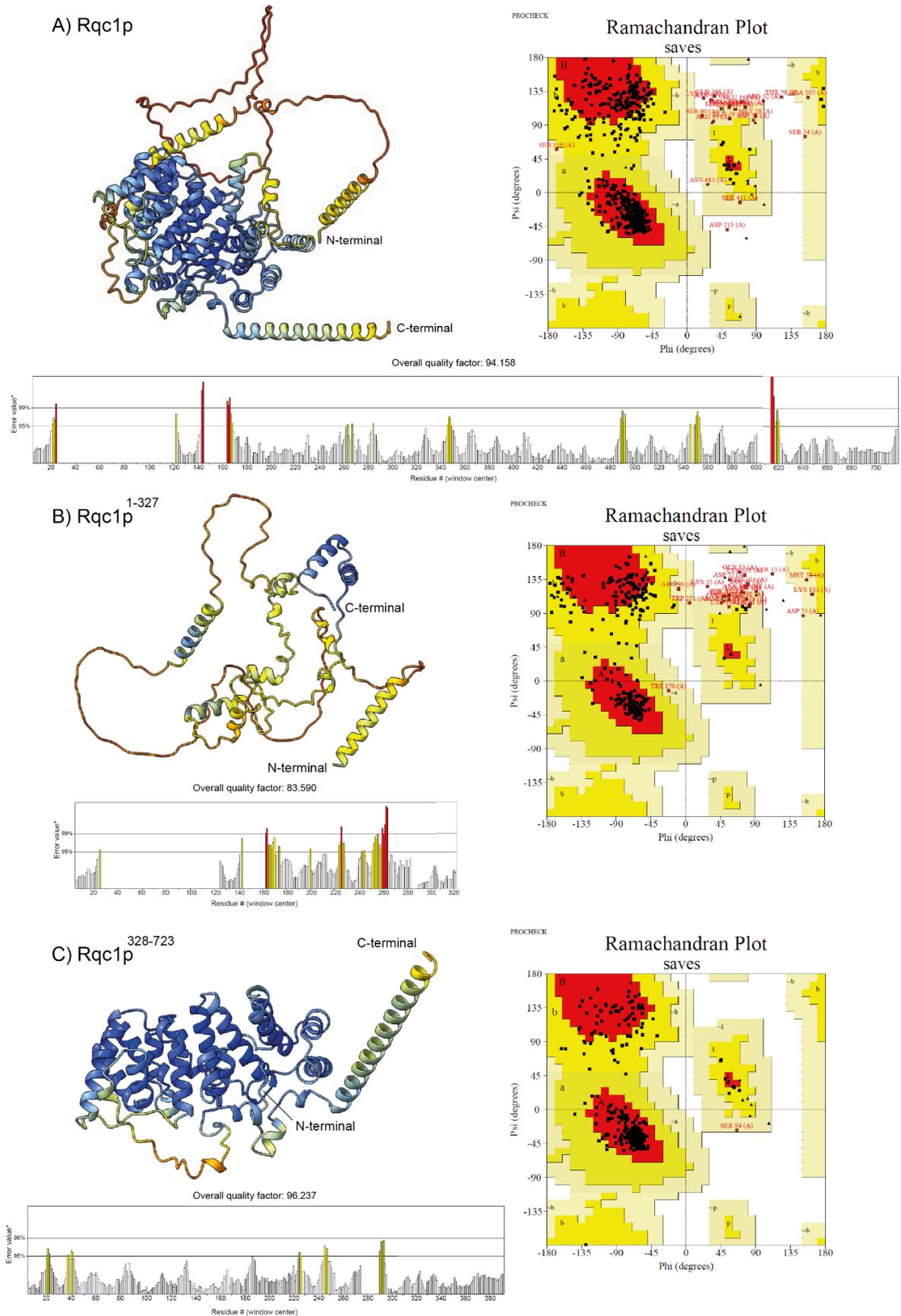


Fig. 16. Modelos gerados pelo ColabFold e gráficos de ERRAT e Ramachandran. Os modelos Rqc1p (painel A), Rqc1p¹⁻³²⁷ (painel B) e Rqc1p³²⁸⁻⁷²³ (painel C) estão coloridos de acordo com os valores de pLDDT que indica confiabilidade da predição da região através de uma escala de 0 a 100, sendo: pLDDT > 90, altíssima confiança (azul escuro), 90 > pLDDT > 70, alta confiança (azul claro), 70 > pLDDT > 50, baixa confiança (amarelo) e pLDDT < 50, baixíssima confiança (laranja). Os gráficos de ERRAT estão representados à direita inferior de cada painel destacando as regiões que se encontram acima do intervalo de 95% (amarelo) e 99% (vermelho) de confiança de rejeição. Já os gráficos de Ramachandran se encontram à esquerda de cada painel, onde os pontos pretos representam cada resíduo de aminoácido. As regiões mais favorecidas (vermelho), adicionalmente permitidas (amarelo), generosamente permitidas (amarelo claro) e não permitidas (branco) também estão representadas nos gráficos.

A análise das estruturas obtidas pelo ColabFold mostrou que, de acordo com os valores de pLDDT, a região N-terminal da proteína apresenta menor confiabilidade, caracterizada pela ausência de estruturas secundárias definidas (Fig. 16, painel A e B). Em contrapartida, a região C-terminal, que contém o domínio TCF25, apresentou regiões modeladas com alta confiabilidade que possuíam predominantemente a formação de múltiplas α -hélices (Fig. 16, painel A e C).

Em relação às métricas utilizadas para validação estrutural, foi observado que dentre as estruturas modeladas, o modelo gerado para Rqc1p³²⁸⁻⁷²³ foi o único que não apresentou nenhum resíduo que ultrapassou o intervalo de 99% de confiança e foi o que obteve o melhor valor de qualidade geral (Fig. 16, painel C). Além disso, ao analisar os gráficos de Ramachandran, vimos que as estruturas geradas para Rqc1p e Rqc1p¹⁻³²⁷ possuíam resíduos localizados em regiões não permitidas e generosamente permitidas, especialmente os resíduos que estão em regiões terminais da proteína (Fig. 16, painel A e B). Por outro lado, o modelo para Rqc1p³²⁸⁻⁷²³ apresentou a maior porcentagem dos seus resíduos em regiões mais favorecidas e nenhum em regiões não permitidas em comparação aos outros modelos. Desta forma, utilizamos este modelo para análises posteriores.

4.9 Identificação de resíduos de Rqc1p que coevoluíram

Utilizando o *software* PFstats, realizamos análises de conservação e coevolução de resíduos a partir do MSA de proteínas que pertencem a família TCF25.

Através da análise de conservação, identificamos três resíduos altamente conservados em mais de 80% das 740 sequências analisadas: N423, D461 e P531, com taxas de conservação de 83%, 95% e 81%, respectivamente (Fig. 17).

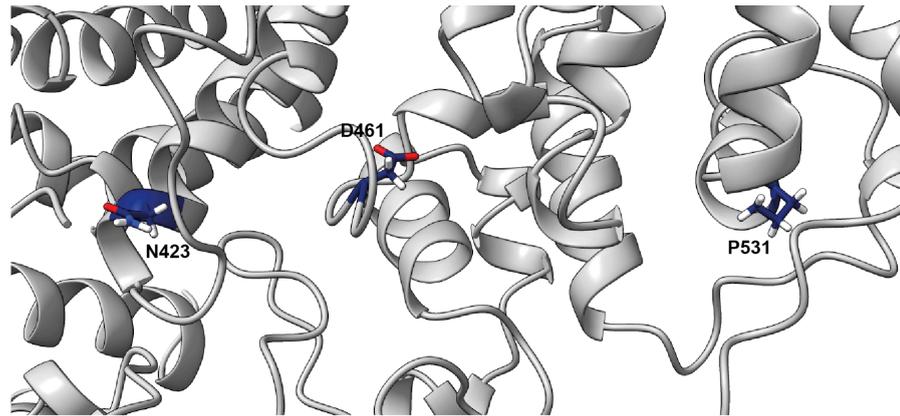


Fig. 17. Resíduos altamente conservados na família TCF25. Os resíduos N423, D461 e P531 estão destacados no modelo Rqc1p³²⁸⁻⁷²³. A numeração dos resíduos corresponde à proteína de *S. cerevisiae* (código acesso Uniprot Q05468).

Através da análise de coevolução, identificamos 9 comunidades decompostas da rede de correlação de resíduos (Fig. 18). Uma comunidade é definida como um conjunto de resíduos que são altamente correlacionados entre si, mas com baixa correlação com os outros da rede (FONSECA-JÚNIOR et al., 2018).

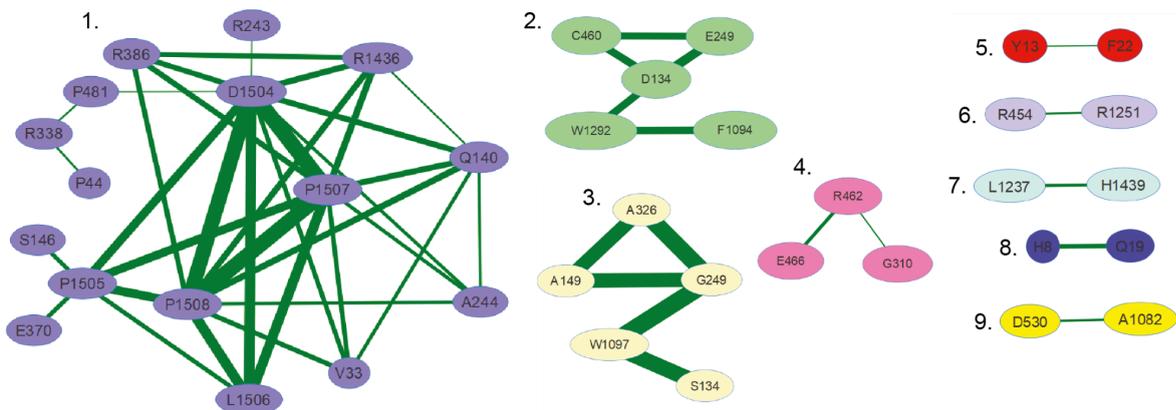


Fig. 18. Decomposição em comunidades da rede de correlação gerada para a família TCF25. A decomposição resultou em 9 comunidades separadas por cores. O número dos resíduos estão de acordo com a posição no alinhamento global. A espessura das arestas indicam a pontuação de correlação entre os resíduos.

Dentre as comunidades identificadas, a comunidade 1 é a maior, composta por 16 resíduos correlacionados. Observamos que a maioria desses resíduos é conservada na mesma posição, desde levedura até humano (Fig. 19, painel A). Além disso, verificamos que os resíduos da comunidade 1 estão localizados próximos na estrutura tridimensional de Rqc1p (Fig. 19, painel B), o que sugere que podem desempenhar um papel funcional importante na proteína.

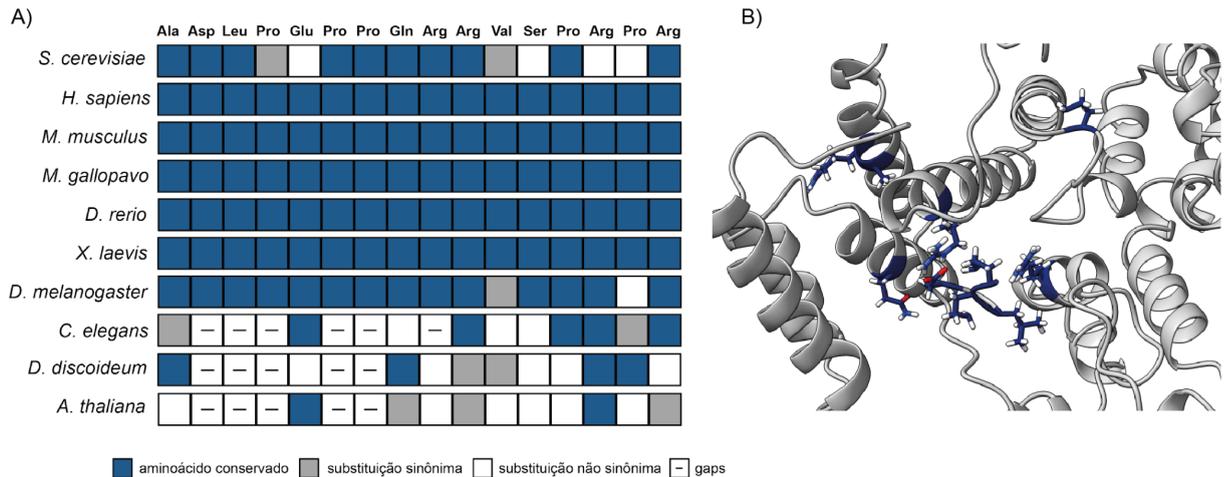


Fig. 19. Representação dos resíduos da comunidade 1 decomposta da rede de correlação gerada para a família TCF25. Painel A: resíduos que estão conservados entre as proteínas homólogas dos organismos modelo (azul) e os que sofreram substituições sinônimas, mantendo a classe do aminoácido (cinza). Painel B: Os resíduos A395, D659, L661, P662, P663, Q374, R631, R424, P458 E R394 da comunidade 1 conservados em *S.cerevisiae* estão destacados no modelo modelo Rqc1p³²⁸⁻⁷²³. A numeração dos resíduos corresponde à proteína de *S. cerevisiae* (código acesso Uniprot Q05468).

4.10 Construção dos plasmídeos YCplac111+Rqc1p, YCplac111+Rqc1p¹⁻³²⁷ e YCplac111+Rqc1p³²⁸⁻⁷²³

Desta forma, para compreender o papel de Rqc1p na resposta ao estresse celular induzido por choque térmico, foram construídos plasmídeos contendo a sequência completa do gene *RQC1* (YCplac111+Rqc1p), bem como as sequências do gene que codificam para a proteína truncada que contém a região polibásica (YCplac111+Rqc1p¹⁻³²⁷) e o domínio TCF25 (YCplac111+Rqc1p³²⁸⁻⁷²³). Também foram construídos vetores intermediários (YCplac111+terminador e YCplac111+promotor+terminador), que foi utilizado ao final para clonar os insertos correspondentes às regiões codificantes (Fig. 20, painel A).

Os fragmentos foram obtidos através de PCRs utilizando como molde o DNA genômico de *S. cerevisiae* S288c juntamente com seus respectivos *primers* (Tabela 2). Os *amplicons* obtidos (Fig. 20, painel B) e o plasmídeo YCplac111 foram purificados e posteriormente digeridos com as enzimas de restrição correspondentes da seguinte forma: promotor (HindIII e PstI), terminador (KpnI e EcoRI) e as regiões codificantes (KpnI e PstI).

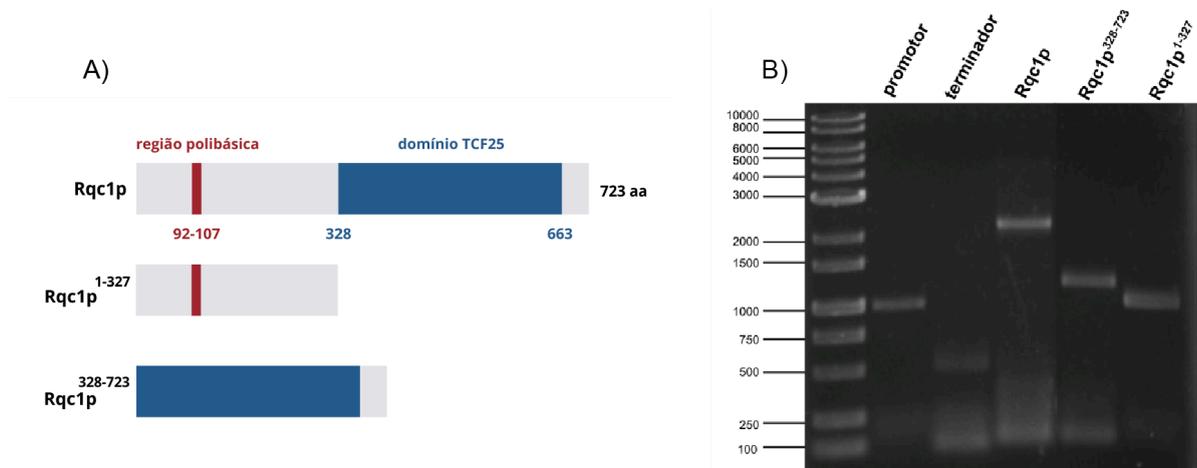


Fig. 20. Construção dos plasmídeos. O modelo esquemático (painel A) representa a estrutura da proteína gerada a partir da sequência do gene completo (Rqc1p) e das sequências que codificam para as proteínas truncadas que contêm a região polibásica em vermelho (Rqc1p¹⁻³²⁷) e o domínio TCF25 em azul (Rqc1p³²⁸⁻⁷²³). A análise da amplificação dos insertos por PCR foi através da observação das bandas esperadas no gel de agarose 1%. O tamanho das bandas dos insertos esperados e observados foram: promotor (1000 pb), terminador (488 pb), Rqc1p (2169 pb), Rqc1p¹⁻³²⁷ (981 pb) e Rqc1p³²⁸⁻⁷²³ (1188 pb).

Os produtos obtidos após a digestão foram purificados, sendo cada inserto ligado ao vetor linearizado. Em seguida, os produtos de ligação foram utilizados para transformar *E. coli* RR1 competente. Após a observação das colônias transformantes em placas contendo meio LB suplementado com ampicilina, foi feita a extração do DNA plasmidial (*miniprep* de colônia) o qual foi utilizado para a condução da análise de restrição. Os clones positivos foram selecionados e cultivados para a extração do DNA plasmidial, agora em maior concentração, para realização de quatro digestões diagnósticas, para mapeamento do plasmídeo e confirmação da clonagem. Em todos os clones positivos selecionados foram observadas as bandas que correspondem ao tamanho esperado para cada corte de enzimas utilizadas nas digestões diagnósticas (Fig. 21, painel D e Apêndice V).

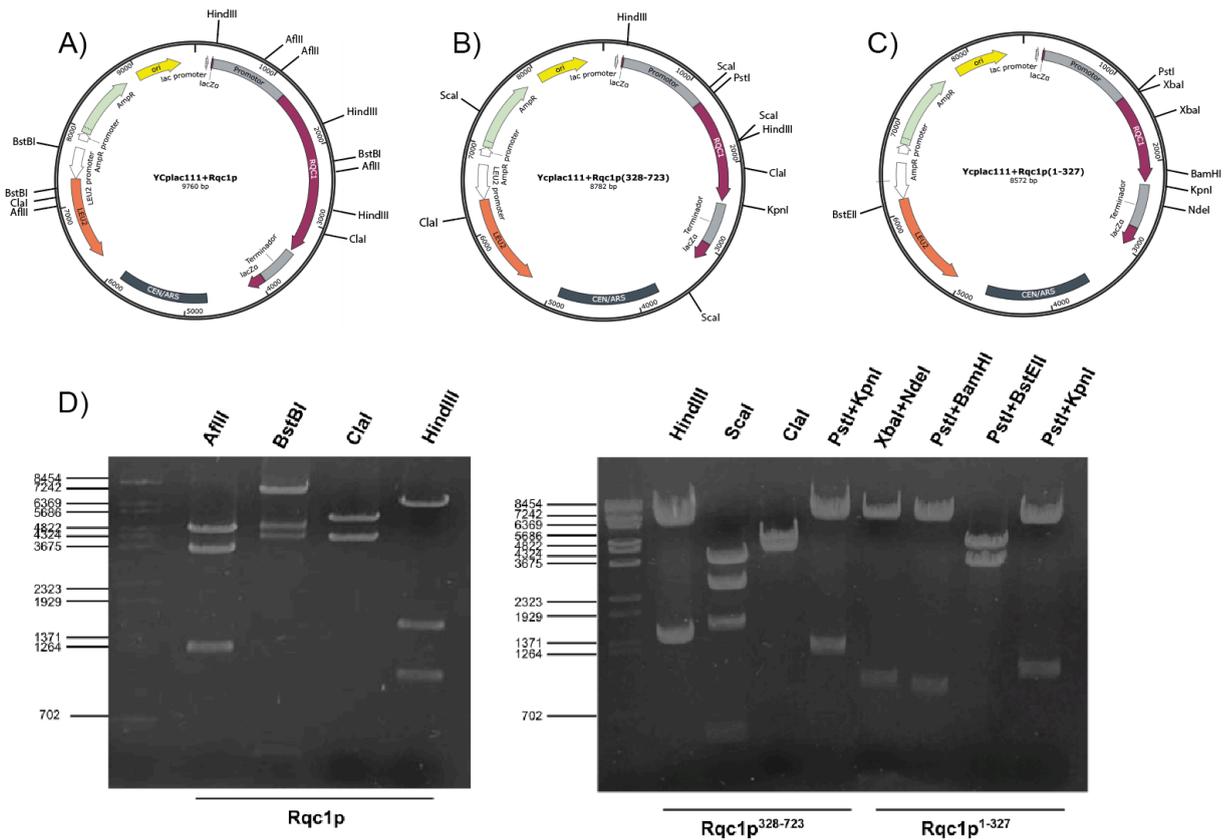


Fig. 21. Mapa das construções e confirmação da clonagem por digestão diagnóstica. Os sítios das enzimas de restrição utilizadas para as digestões diagnósticas estão destacadas nos mapas das construções de YCplac111+Rqc1p (painel A), YCplac111+Rqc1p¹⁻³²⁷ (painel C) e YCplac111+Rqc1p³²⁸⁻⁷²³ (painel B). A análise das digestões diagnósticas das construções foi feita através da observação das bandas esperadas em gel de agarose 1% (painel D). O tamanho das bandas dos insertos esperados e observados foram: YCplac111+Rqc1p: AflIII (4742 pb, 3540 pb, 1265 pb e 213 pb*), BstBI (4986 pb, 4283 pb e 491 pb), ClaI (5623 pb e 4137 pb); YCplac111+Rqc1p³²⁸⁻⁷²³: HindIII (7225 pb e 1557 pb), Scal (3862 pb, 2725 pb, 1636 pb e 556 pb), ClaI (4645 pb e 4137 pb) e PstI+KpnI (7558 pb e 1224 pb) e YCplac111+Rqc1p¹⁻³²⁷: XbaI+NdeI (7701 pb e 871 pb), PstI+BamHI (7736 pb e 831 pb), PstI+BstEII (4870 pb e 3702 pb) e PstI+KpnI (7558 pb e 1014 pb).

Após a confirmação do sucesso da clonagem através das digestões diagnósticas (Fig. 21, painel D), foi realizado o sequenciamento Sanger de todas as construções, conforme indicado em Materiais e Métodos (Item 3.1.12).

4.11 A reintrodução de *RQC1* em *S. cerevisiae rqc1Δ* complementa o fenótipo selvagem em choque térmico

Uma vez que Rqc1p se mostrou importante para a resposta ao choque térmico (Figs. 9 e 10), realizamos novamente um ensaio de tolerância celular para verificar o crescimento de *S. cerevisiae rqc1Δ* complementada com *RQC1* (*rqc1Δ*+Rqc1p) e com as sequências codificantes para as duas variantes truncadas (*rqc1Δ*+Rqc1p¹⁻³²⁷ e *rqc1Δ*+Rqc1p³²⁸⁻⁷²³). Para este ensaio, monitoramos o

crescimento das células em placas contendo meio sintético sólido desprovido de leucina (SD-Leu), que é o marcador auxotrófico do plasmídeo YCplac111, sob choque térmico.

Observamos que as células *rqc1*Δ complementadas com *RQC1* restauraram o fenótipo selvagem, recuperando a tolerância ao choque térmico (Fig. 22). Além disso, observamos que não houve restauração do fenótipo em nenhum dos transformantes contendo os plasmídeos que codificam ou a porção da proteína que contém a região polibásica (*rqc1*Δ+Rqc1p¹⁻³²⁷), ou o domínio TCF25 (*rqc1*Δ+Rqc1p³²⁸⁻⁷²³). Esta observação sugere que as duas regiões da proteína são importantes para a manutenção da viabilidade celular em condições de choque térmico.

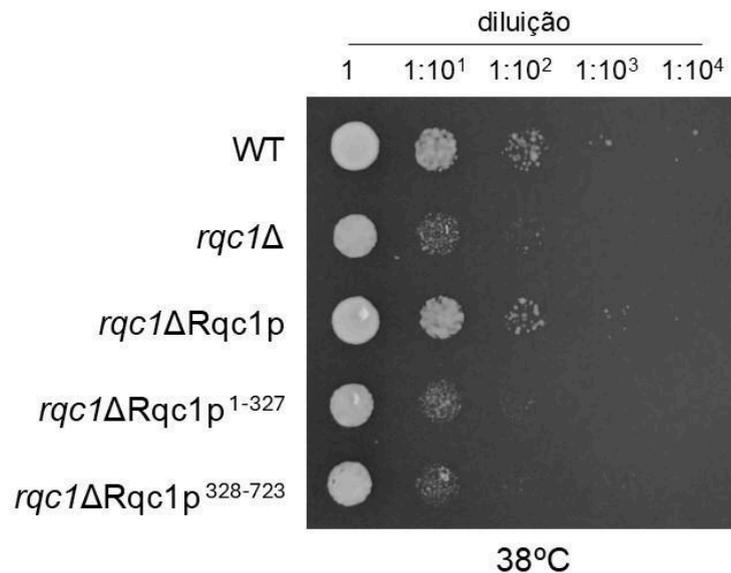


Fig. 22. Tolerância ao choque térmico de *S. cerevisiae* WT, *rqc1*Δ, *rqc1*Δ+Rqc1p, *rqc1*Δ+Rqc1p¹⁻³²⁷ e *rqc1*Δ+Rqc1p³²⁸⁻⁷²³. A determinação da tolerância ao choque térmico de *S. cerevisiae* através da capacidade de crescimento SD-Leu sob choque térmico foi realizado conforme descrito em Materiais e Métodos (Item 3.1.3). As placas foram fotografadas após 48h de incubação a 38°C. A Fig. é representativa de 4 replicatas biológicas.

4.12 Bibliotecas selecionadas, mapeamento, contagem de genes e análise de componentes principais

Com o objetivo de investigar o impacto do choque térmico na expressão dos genes associados a HSR e ao controle da síntese de proteínas pelo RQC, analisamos dados públicos de sequenciamento de RNA de *S. cerevisiae* WT tratadas nesta condição.

Após a curadoria e filtragem manual, quatro estudos foram selecionados para as análises de expressão diferencial de genes (Tabela 8 e Apêndice VI).

Tabela 8. Descrição dos dados de sequenciamento obtidos no SRA. Todos os estudos selecionados possuem ao mínimo 3 replicatas biológicas e sequenciamento *paired-end*. Os critérios para seleção dos estudos estão descritos em Materiais e Métodos (Item 3.3.1)

ID do projeto no SRA	Plataforma de sequenciamento	Profundidade das bibliotecas	Tamanho das reads	Linhagem de <i>S. cerevisiae</i>	Temperatura de cultivo	Tempo de cultivo em choque térmico	Artigo do estudo
PRJNA532914	Illumina HiSeq 4000	22,8 -31,5M	150 pb	WT BY4741	30°C (controle) e 37°C (tratados)	15 minutos	Peffer, 2019
PRJNA559331	Illumina HiSeq 2000	26,8 - 33,3M	101 pb	WT BY4741	30°C (controle) e 39°C (tratados)	20 e 120 minutos	Nuño-Cabanes, 2020
PRJNA312727	Illumina HiSeq 2500	8,3 - 10,6M	101 pb	WT CAY1015	25°C (controle) e 37°C (tratados)	30 minutos	Masser, 2020
PRJNA498270	Illumina HiSeq 2500	9,9 - 12,0M	50 pb	WT YPH499	23°C (controle) e 37°C (tratados)	120 minutos	Poveda-Huertes, 2020

É importante ressaltar que foram selecionados transcriptomas de estudos independentes que investigaram a resposta ao choque térmico em diferentes linhagens de *S. cerevisiae*. Embora existam diferenças genéticas entre as linhagens, todas são comumente utilizadas em estudos de estresse celular permitindo uma análise comparativa, já que os dados foram processados utilizando os mesmos parâmetros (Apêndice VII).

Após a seleção dos estudos, as bibliotecas foram mapeadas no genoma de referência e foi feita a contagem do número de *reads* mapeadas. Os percentuais de mapeamento das bibliotecas foram considerados satisfatórios em todos os estudos, sendo em sua maioria acima de 80% (Apêndice VIII).

Para a visualização dos padrões de variação entre as condições experimentais de cada estudo individualmente, foi realizada a Análise de Componentes Principais (PCA). Os gráficos de dispersão mostram que em todos os estudos analisados houve um agrupamento entre as amostras dentro da mesma condição, o que indica consistência entre as condições experimentais. Também vimos que houve separação entre as replicatas controles e tratadas em todos os estudos, destacando a diferença entre a ausência e a presença do choque térmico. (Fig. 23).

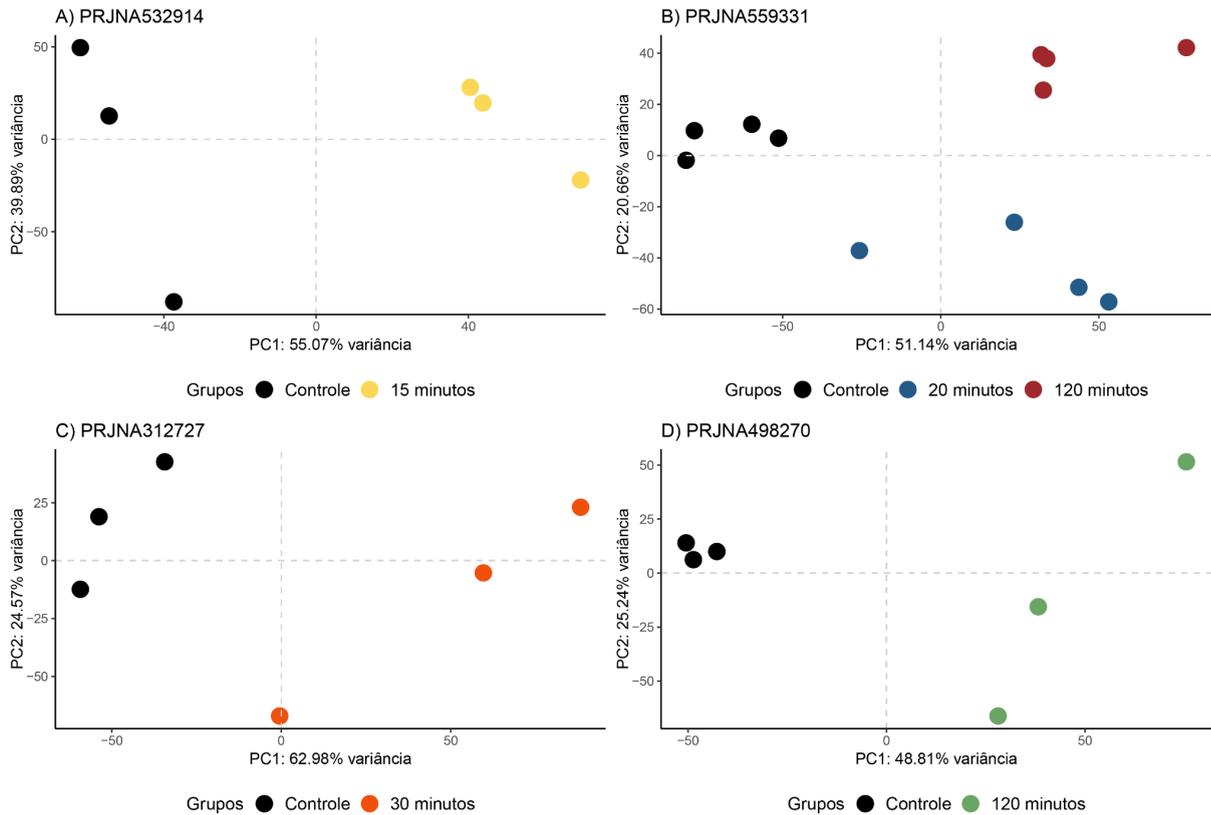


Fig. 23. Análise de componentes principais (PCA). Gráficos de dispersão representando os resultados do PCA para cada estudo analisado. Cada ponto representa uma amostra de cada uma das condições experimentais. Os componentes principais PC1 e PC2 explicaram para cada estudo, respectivamente: 55,07% e 39,89% para PRJNA532914 (painel A); 51,14% e 20,66% para PRJNA559331 (painel B), 62,98% e 24,57% para PRJNA312727 (painel C) e 48,81% e 25,24% para PRJNA498270 (painel D).

4.13 Análise da expressão diferencial de genes

Considerando que *S. cerevisiae* possui aproximadamente 6.000 genes, observamos que a maioria daqueles que são diferencialmente expressos (DEGs, *differentially expressed genes*) apresentaram perfil de superexpressão em todas as condições de choque térmico analisadas (Fig. 24). Curiosamente, o tratamento das células por 15 min em choque térmico teve o maior impacto na regulação da expressão gênica, envolvendo 1.146 genes (538 sub-expressos e 608 superexpressos) (Fig. 24, painel A). Por outro lado, o cultivo das células sob choque térmico por 120 min apresentou o menor impacto na regulação da expressão nos dois estudos analisados nesta condição (Fig. 24, painel D e E). Nestes casos, o total de DEGs foram de 470 (158 sub-expressos e 312 superexpressos) para o estudo em que as células foram tratadas a 39°C (Fig. 24, painel D) e 311 (115 sub-expressos e 196 superexpressos) no estudo em que as células foram cultivadas a 37°C (Fig. 24, painel E).

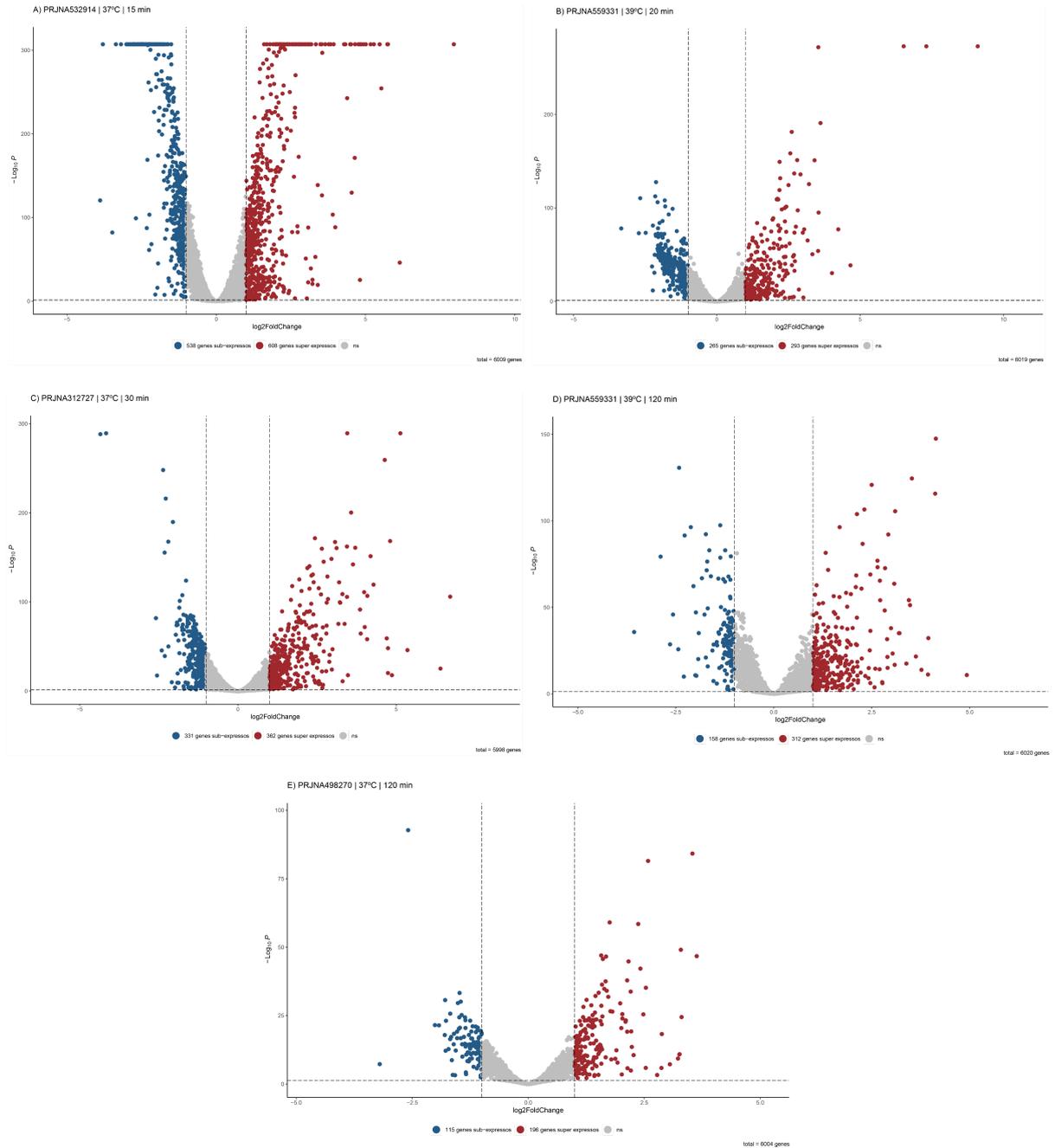


Fig. 24. Genes diferencialmente expressos em choque térmico. Volcano plots representando os genes diferencialmente expressos para cada estudo analisado. Os pontos representam os genes não diferencialmente expressos (cinza), genes sub-expressos (azul) e superexpressos (vermelho) de acordo com o valor de $\log_2\text{FoldChange} \geq |1|$ e $\text{padj} < 0,05$. Os estudos estão ordenados de acordo com o tempo de cultivo em condição de choque térmico: PRJNA532914, 15 minutos a 37°C (painel A), PRJNA559331, 20 minutos a 39°C (painel B), PRJNA312727, 30 minutos a 37°C (painel C), PRJNA559331, 120 minutos a 39°C (painel D) e PRJNA498270, 120 minutos a 37°C (painel E).

Com o objetivo de investigar o impacto do choque térmico na expressão dos genes associados a HSR e ao controle da síntese de proteínas pelo complexo RQC, selecionamos ao total 31 genes envolvidos nestas vias. Além dos genes que codificam para proteínas integrantes do RQC, também selecionamos genes

integrantes da via que fazem parte da rede de interação para Rqc1p obtida no STRING (Fig. 25).

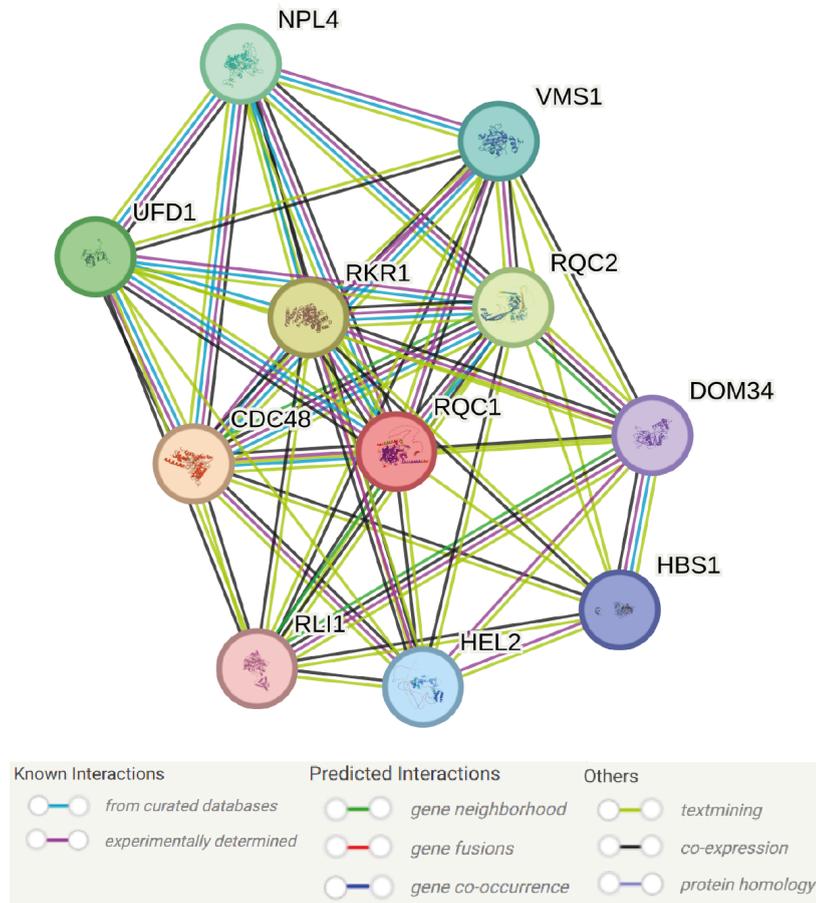


Fig. 25. Rede de interação de Rqc1p obtida pelo STRING v.12. As proteínas que interagem com Rqc1p estão destacadas nos nós e as conexões indicam interações baseadas em diferentes fontes de evidência como bancos de dados curados (azul claro), determinação experimental (roxo), predições computacionais e coexpressão (preto).

Observamos que a maioria dos genes que codificam proteínas envolvidas na resposta ao choque térmico (HSPs, do inglês *heat shock proteins*) estão diferencialmente expressos em todos os estudos analisados (Fig. 26). Além disso, como observado anteriormente, o tratamento prolongado por 120 min sob choque térmico levou à redução da expressão desses genes (Fig. 26, painel D e E). Por outro lado, entre os genes que codificam para as proteínas integrantes do RQC, somente *UFD1* foi diferencialmente expresso nas células expostas a 39°C por 20 min (Fig. 26. painel B).

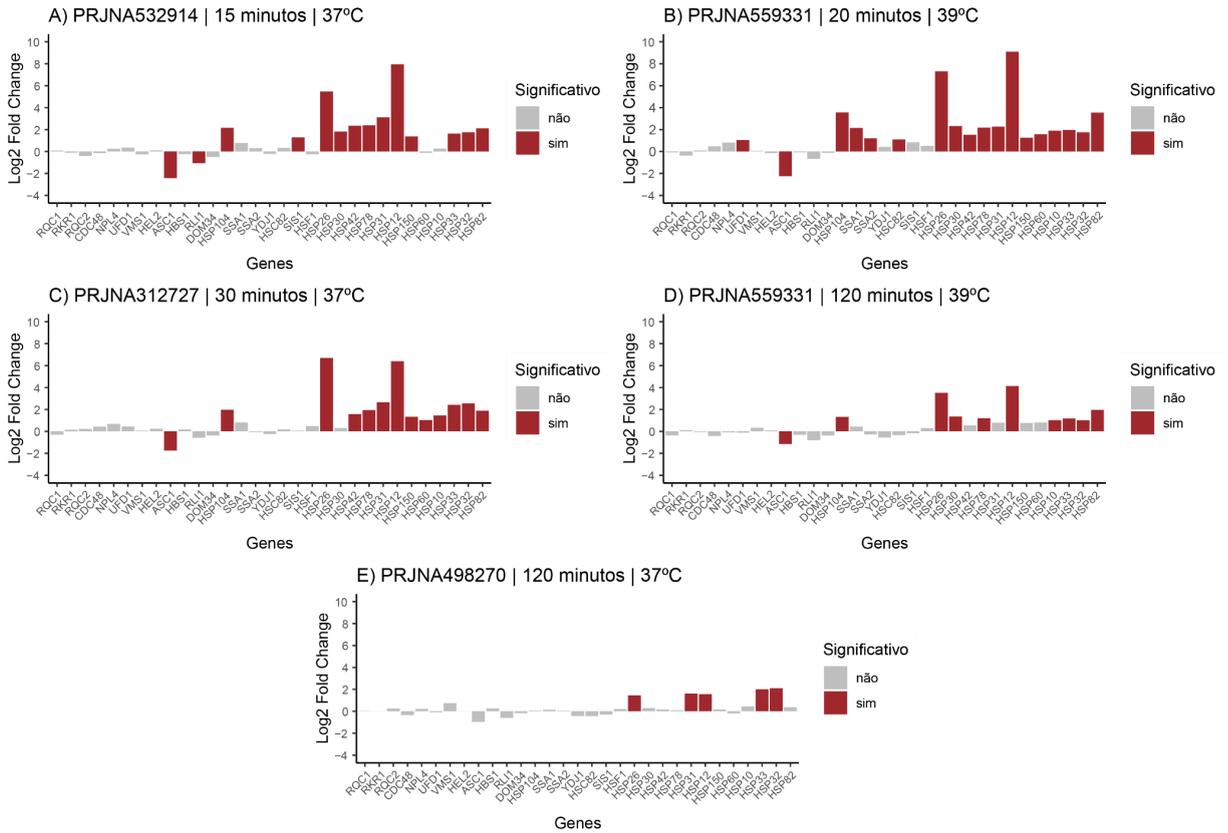


Fig. 26. Expressão dos genes associados a resposta ao choque térmico e ao controle da síntese de proteínas. As barras vermelhas representam os genes que são diferencialmente expressos de acordo com o valor de $\log_2\text{FoldChange} \geq |1|$ e $\text{padj} < 0,05$. Os estudos estão ordenados de acordo com o tempo de cultivo em condição de choque térmico: PRJNA532914, 15 minutos a 37°C (painel A), PRJNA559331, 20 minutos a 39°C (painel B), PRJNA312727, 30 minutos a 37°C (painel C), PRJNA559331, 120 minutos a 39°C (painel D) e PRJNA498270, 120 minutos a 37°C (painel E).

5 DISCUSSÃO

Neste trabalho, realizamos a caracterização estrutural e funcional da proteína Rqc1 de *S. cerevisiae* em resposta a estresses celulares. A caracterização estrutural foi conduzida por meio de abordagens computacionais, incluindo análise de conservação e coevolução entre ortólogos, predição da estrutura secundária e modelagem molecular. Além disso, investigamos a influência do estresse traducional e do choque térmico na viabilidade de mutantes de *S. cerevisiae* desprovidos de genes envolvidos no controle de qualidade em síntese de proteínas, através de ensaios de tolerância celular. Também avaliamos o papel de Rqc1p na resposta ao choque térmico por meio de ensaios de complementação gênica total ou parcial, utilizando versões truncadas da proteína contendo apenas os domínios N-terminal ou C-terminal. Por fim, analisamos a expressão de genes relacionados à via do complexo RQC, a partir de dados públicos de sequenciamento de RNA de linhagens WT de *S. cerevisiae* cultivadas em choque térmico.

Situações de estresses ambientais promovem remodelamento do proteoma para permitir que as células se adaptem às novas condições; neste contexto, ocorrem processos de re-enovelamento e degradação de proteínas defeituosas. Desta maneira, o controle de qualidade em síntese de proteínas é fundamental para a manutenção da proteostase celular nesses cenários (MÜHLHOFER et al., 2019). Para investigar o papel do complexo RQC na resposta a estresses, realizamos ensaios fenotípicos de tolerância celular em linhagens mutantes de *S. cerevisiae* para genes que codificam proteínas do complexo RQC.

Inicialmente, investigamos a capacidade de crescimento de *S. cerevisiae* desprovida dos genes *RQC1* e *LTN1* para avaliar a tolerância ao estresse induzido por higromicina B. A higromicina B é um antibiótico aminoglicosídeo que atua principalmente inibindo a translocação do tRNA entre os sítios do ribossomo durante a fase de alongamento da tradução, ocasionando a parada da síntese de proteínas (BOROVINSKAYA et al., 2008). Nossos resultados confirmam a elevada sensibilidade de células *ltn1Δ* à esse antibiótico (BENGTSON; JOAZEIRO, 2010) e adicionalmente mostram que a deleção de *RQC1* também reduz a viabilidade de *S. cerevisiae* em sua presença (Fig. 6).

Isso pode ser explicado já que a poliubiquitinação ocorre sinergicamente com a adição da *CAT tail* por Rqc2p, o que facilita a exposição dos resíduos de lisina que serão poliubiquitinados e podem permanecer enterrados no túnel de saída do

ribossomo (KOSTOVA et al., 2017). No entanto, na ausência de Ltn1p, os polipeptídeos que possuem *CAT tail* tendem a se agregar, já que não são direcionados para degradação eficiente, o que pode levar ao sequestro de chaperonas, induzindo estresse proteotóxico (CHOE et al., 2016; DEFENOUILLE et al., 2016; YONASHIRO et al., 2016). Além disso, sabe-se que Hsf1p é ativado quando ocorre o comprometimento da função do RQC, uma ativação dependente de Rqc2p, possivelmente mediada pela formação de agregados contendo *CAT tail* (BRANDMAN et al., 2012).

A partir disso, considerando que uma das principais vias de sinalização de estresse é a HSR, mediada por Hsf1p, investigamos o efeito do choque térmico na tolerância celular em linhagens nocautes de *S. cerevisiae* para os genes que codificam Rqc1p, Ltn1p e Rqc2p. Notavelmente, nossos resultados revelaram que a redução da resistência celular ao choque térmico somente ocorreu na ausência de *RQC1* (Fig. 9). Curiosamente, esse fenótipo é revertido quando *LTN1* é deletado concomitantemente (Fig. 10).

Para investigar se o mesmo fenótipo se repetia, utilizando agora um método quantitativo, verificamos o crescimento celular de *S. cerevisiae* WT, *rqc1Δ*, *ltn1Δ* e *rqc1Δltn1Δ* e calculamos a velocidade específica máxima de crescimento de todas as linhagens (Fig. 11). Corroborando os dados anteriores, somente *rqc1Δ* apresentou uma redução na velocidade específica máxima de crescimento, enquanto *rqc1Δltn1Δ* manteve sua velocidade similar a WT e *ltn1Δ* (Fig. 11, painel H).

Dessa forma, direcionamos nossos estudos para a proteína Rqc1p e, apesar de os esforços realizados nas últimas décadas terem contribuído para descrever a diversidade funcional de Rqc1p na célula, pouco se sabe sobre sua estrutura tridimensional, já que as tentativas experimentais de estabelecer a estrutura e posição no complexo RQC associado ao ribossomo não foram bem sucedidas (LYUMKIS et al., 2014; SHAO et al., 2015; TESINA et al., 2023). Com o objetivo de fornecer *insights* sobre sua função biológica e guiar experimentos futuros, realizamos a caracterização estrutural *in silico* de Rqc1p.

As proteínas ortólogas de camundongo e humano foram inicialmente previstas como fatores de transcrição que pertencem à família bHLH (OLSSON et al., 2002; CAI et al., 2006). As proteínas bHLH contêm um domínio hélice-volta-hélice (HLH), geralmente precedido por um trecho de resíduos básicos,

que são responsáveis pela dimerização e ligação ao DNA, respectivamente. Além disso, essas proteínas podem conter domínios adicionais que determinam sua especificidade na regulação transcricional (MURRE et al., 1994; MURRE, 2019). Os fatores de transcrição da família bHLH estão presentes em todos os eucariotos e desempenham papéis essenciais na regulação de redes transcricionais associadas a genes envolvidos na diferenciação e no desenvolvimento celular (ATCHLEY; FITCH, 1997).

De fato, o alinhamento múltiplo de sequências revelou que as regiões de maior conservação entre os eucariotos compreendem a região polibásica – localizada próximo ao N-terminal – e um domínio pouco caracterizado, anotado como TCF25, próximo ao C-terminal (Fig. 12). Já a predição da estrutura secundária e a modelagem molecular mostrou que Rqc1p é composta por regiões desordenadas em sua extremidade N-terminal e possui a formação de múltiplas α -hélices concentradas principalmente na região correspondente ao domínio TCF25 (Fig.s 15 e 16).

Sabe-se que regiões intrinsecamente desordenadas (IDR, *intrinsically disordered regions*) são comuns em fatores de transcrição e já foram relatadas como importantes para funcionalidade dentro da família bHLH (TARCZEWSKA; GREB-MARKIEWICZ, 2019). Essas regiões, geralmente também flexíveis, podem conter sequências específicas que permitem múltiplas interações com co-reguladores da transcrição, o que pode explicar a diversidade funcional descrita na literatura para TCF25 (STEEN; LINDHOLM, 2008; ZENG et al., 2018).

Dessa forma, sabendo da prevalência de IDRs em fatores de transcrição bHLH e considerando que os algoritmos de modelagem estrutural geralmente não modelam bem essas regiões – já que não adotam uma única conformação estável em solução e dependem do ambiente e das interações moleculares –, era esperado que estas regiões apresentassem baixa confiabilidade na predição realizada pelo I-TASSER e pelo ColabFold (Tabela 7). Isso também pode explicar o motivo dos modelos que continham o N-terminal terem obtido baixos escores após a análise da estabilidade da conformação de resíduos e a presença de resíduos em regiões desfavoráveis e não permitidas no gráfico de Ramachandran (Fig. 16, painel A e B).

Posteriormente, buscando compreender a relevância funcional da conservação do domínio TCF25 entre os eucariotos, utilizamos o alinhamento múltiplo de 740 sequências de proteínas ortólogas para análises de conservação e

decomposição de redes de coevolução de resíduos em comunidades. Identificamos uma comunidade de 16 resíduos correlacionados que estão conservados na mesma posição, desde leveduras a humanos, sugerindo que estes resíduos possam desempenhar um papel chave na função biológica exercida pela proteína e são fortes candidatos para ensaios futuros de mutações sítio-dirigidas (Fig. 19).

Além disso, através do alinhamento estrutural do modelo Rqc1p³²⁸⁻⁷²³ com estruturas de proteínas que foram resolvidas experimentalmente depositadas no PDB100, observamos que o domínio TCF25 possui similaridade estrutural com proteínas que contêm motivos de repetição tetratricopeptídeo (TPR, *tetratricopeptide repeat*) (Apêndice IV). Os domínios TPRs são amplamente difundidos na natureza e são constituídos de motivos de 34 resíduos de aminoácidos repetidos em que as repetições se compactam dando origem a estruturas super-helicoidais que interagem na formação de complexos em vários processos celulares como transcrição, controle do ciclo celular e respostas ao estresse (ALLAN; RATAJCZAK, 2011; PEREZ-RIBA; ITZHAKI, 2019). Dessa forma, esses achados podem sugerir que o domínio TCF25 pode ser responsável pela interação de Rqc1p com proteínas do complexo RQC ou de proteínas ribossomais.

É importante pontuar que, recentemente, Abaeva e colaboradores propuseram um modelo estrutural de interação de TCF25, Listerin (E3 ligase), Ube2D1 (E2 conjugase) e moléculas de ubiquitina, utilizando o Alphafold3 (ABAEVA et al., 2024, *preprint*). De acordo com o modelo, a região mais estruturada de TCF25 de humano (175-665) interage com o domínio RING de Listerin e com uma ubiquitina, orientando-a de modo que o resíduo K48 é posicionado para atacar a ligação tioéster do conjugado formado pela E2 e outra ubiquitina. No entanto, esse modelo não considerou a interação do complexo RQC associado a subunidade 60S, o que pode influenciar na interação das proteínas e estudos estruturais futuros são necessários para preencher essa lacuna.

Sabendo que as proteínas ortólogas de mamíferos foram descritas como fatores de transcrição e que há evidências de sua presença no núcleo em células humanas (STEEN; LINDHOLM, 2008), buscamos investigar a presença de sinais de localização e sítios de modificações pós-traducionais na sequência de *S. cerevisiae*, o organismo modelo deste trabalho. O transporte nucleocitoplasmático compreende atividades de importação e exportação nuclear, mediadas por modificações e interações entre os NLSs e NESs com proteínas específicas (LANGE et al., 2007).

Observamos que todas as proteínas analisadas foram preditas como presentes majoritariamente no citoplasma e as homólogas de camundongo e humano também apresentaram probabilidade de localização parcialmente nuclear, como esperado (Apêndice II). No entanto, apesar de a probabilidade de localização nuclear de Rqc1p não ter alcançado o limiar, foi predita um NLS associada à região da sequência próxima ao N-terminal (Fig. 13, painel A e B). Para comparar os resultados obtidos e aumentar a robustez das análises, utilizamos outra ferramenta que prevê NLSs clássicos da via importina α/β . Foi identificado um NLS do tipo monopartido em Rqc1p, localizado no N-terminal dentro da região polibásica, além de vários NLSs bipartidos. De fato, os primeiros estudos de caracterização de TCF25 de camundongo e humano, identificaram um NLS monopartido (KKKKKQK) na posição 127-133, semelhante ao NLS do antígeno T grande do vírus símio 40 (SV40) (OLSSON et al., 2002; CAI et al., 2006; KALDERON et al., 1984).

Além disso, também buscamos mapear os sítios de fosforilação anotados para Rqc1p disponíveis no SGD e identificamos dois resíduos de serina passíveis de fosforilação localizados dentro dos NLSs bipartidos preditos (Fig. 14, painel A). Sabendo que a fosforilação dentro ou ao redor de um NLS é comumente utilizada para regular o transporte de uma proteína – já que a fosforilação ao redor de NLS dependente de importina α mimetiza o efeito inibitório de aminoácidos ácidos (HARREMAN et al., 2004) –, esses resultados sugerem que a translocação pode ser mediada por este tipo de regulação.

Após a observação de que Rcp1p pode desempenhar um papel relevante na resposta induzida por choque térmico, e considerando que sua diversidade funcional possivelmente é mediada por seus domínios (como discutido anteriormente), realizamos ensaios de complementação gênica total ou parcial, utilizando *RQC1* selvagem e as variantes truncadas da proteína contendo a região polibásica (*rqc1Δ*+Rqc1p¹⁻³²⁷) ou domínio TCF25 (*rqc1Δ*+Rqc1p³²⁸⁻⁷²³). Nossas análises mostraram que a reinserção do gene *RQC1* em leveduras *rqc1Δ* restaurou o fenótipo selvagem, recuperando a tolerância ao choque térmico (Fig. 22). Por outro lado, o fenótipo não foi revertido quando houve a reinserção dos genes codificantes das sequências truncadas, indicando que ambos os domínios são essenciais para manutenção da função biológica de Rcp1p na resposta ao choque térmico. A partir disso, a realização de ensaios de localização subcelular utilizando leveduras cultivadas em choque térmico se torna essencial para determinar se a resposta ao

estresse está relacionada à atuação de Rqc1p no núcleo ou no citoplasma. Dessa forma, de acordo com a predição e identificação de NLSs presentes na porção N-terminal, é esperado que somente Rqc1p e Rqc1p¹⁻³²⁷ sejam translocadas para o núcleo.

Por último, a partir dos fenótipos observados que indicam que *RQC1* desempenha um papel importante na resposta ao choque térmico, buscamos investigar se os genes associados à via do complexo RQC apresentam expressão diferencial sob esta condição, através de análises de dados públicos de RNA-seq. É importante ressaltar que as análises foram realizadas com estudos que utilizaram linhagens WT para os ensaios experimentais, uma vez que não há dados publicados para linhagens *rqc1Δ*.

A análise de dados de RNA-seq de quatro estudos independentes revelou que os genes associados à resposta clássica ao choque térmico estavam diferencialmente expressos, indicando que a resposta foi capturada nos experimentos. A maioria dos genes diferencialmente expressos apresentou perfil de superexpressão, e o maior impacto ocorreu após 15 minutos de tratamento em choque térmico (Fig. 24, painel A). Já o tratamento prolongado de 120 minutos levou a uma redução na expressão de genes que codificam HSPs, indicando que a resposta ao estresse diminuiu ao longo do tempo (Fig. 24, painel D e E).

Por outro lado, os genes que codificam proteínas integrantes do complexo RQC, incluindo *RQC1*, não apresentaram diferença na expressão de acordo com os filtros utilizados em nenhuma condição, exceto *UFD1* (Fig. 26, painel B). Uma hipótese que pode explicar essa observação é que a regulação dos componentes do RQC pode ser independente da transcrição, ocorrendo possivelmente por mecanismos de modificações pós-traducionais. De fato, Barros e colaboradores propuseram um mecanismo para regulação positiva dos níveis de Rqc1p em leveduras na ausência de Ltn1p, onde Rqc1p pode ser degradado pós-traducionalmente por Ltn1p (BARROS et al., 2021).

Vale ressaltar que a ausência de mudanças transcricionais na linhagem WT não exclui o papel de *RQC1* no choque térmico, especialmente considerando que as células *rqc1Δ* apresentam fenótipo de sensibilidade e esse fenótipo foi complementado com a reinserção do gene, como demonstramos neste trabalho. No entanto, o desbalanço gênico na ausência de *RQC1* parece ser importante na condução da resposta frente ao choque térmico. Dessa forma, faz-se interessante a

obtenção de dados de sequenciamento de RNA das linhagens utilizadas neste trabalho para inferência de informações mais precisas sobre o papel dos genes da via do complexo RQC na regulação da expressão gênica associada ao choque térmico.

6 CONCLUSÕES

Nossos achados, em conjunto aos dados disponíveis na literatura, indicam que a proteína Rqc1 é conservada entre os eucariotos e que a região polibásica e o domínio TCF25 podem ser determinantes para sua função biológica no núcleo e no citoplasma. Enquanto a região polibásica faz parte do domínio bHLH, descrito por realizar interação com ácidos nucleicos, o domínio TCF25 apresenta estrutura predominantemente α -helicoidal e possui similaridade estrutural com domínios TPR, envolvidos em interação proteína-proteína. Além disso, nossos dados de predição de localização subcelular sugerem que a região polibásica medeia a translocação da proteína Rqc1 para o núcleo, possivelmente para exercer sua função de fator de transcrição.

A análise fenotípica de *S. cerevisiae* desprovida de proteínas do complexo RQC revelou que Rqc1p tem um papel central na resposta ao estresse induzido por choque térmico. A deleção de *RQC1* resultou em maior sensibilidade ao choque térmico e redução na velocidade específica máxima de crescimento em cultivo em placa e em meio líquido, respectivamente. Notavelmente, a complementação de células desprovidas do gene *RQC1* com vetor de expressão contendo o gene foi capaz de restaurar o fenótipo selvagem.

Por fim, a análise de dados públicos de RNA-seq de *S. cerevisiae* WT sob choque térmico não demonstrou diferenças significativas na expressão dos genes do complexo RQC, incluindo *RQC1*. No entanto, a ausência de mudanças transcricionais na linhagem WT não descarta a importância de *RQC1* na resposta ao choque térmico, especialmente considerando que as células *rqc1* Δ apresentaram fenótipo de sensibilidade e esse fenótipo foi revertido pela reinserção do gene, como demonstramos neste trabalho.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ABAEVA, I. S. et al. The ribosome-associated quality control factor TCF25 imposes K48 specificity on Listerin-mediated ubiquitination of nascent chains by binding and specifically orienting the acceptor ubiquitin. *bioRxiv* [preprint] 18 out. 2024. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/2024.10.17.618946>>.

AMBERG, D.C.; BURKE D.J.; STRATHERN J.N. Yeast DNA isolation: miniprep. **CSH Protocols**, v. 2006, n. 1, jun. 2006.

ALLAN, R. K.; RATAJCZAK, T. Versatile TPR domains accommodate different modes of target protein recognition and function. **Cell Stress and Chaperones**, v. 16, n. 4, p. 353–367, jul. 2011.

ATCHLEY, W. R.; FITCH, W. M. A natural classification of the basic helix–loop–helix class of transcription factors. **Proceedings of the National Academy of Sciences**, v. 94, n. 10, p. 5172–5176, may 1997.

BABRAHAM BIOINFORMATICS. FastQC A Quality Control tool for High Throughput Sequence Data.

BARROS, G. C. et al. Rqc1 and other yeast proteins containing highly positively charged sequences are not targets of the RQC complex. **Journal of Biological Chemistry**, v. 296, n. 100586, jan. 2021.

BENGTSON, M. H.; JOAZEIRO, C. A. P. Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. **Nature**, v. 467, n. 7314, p. 470–473, set. 2010.

BERG, J. M.; TYMOCZKO, J. L.; STRYER, L. **Bioquímica**. Editora Guanabara Koogan Ltda, 2013.

BLEICHER, L.; LEMKE, N.; GARRATT, R. C. Using Amino Acid Correlation and Community Detection Algorithms to Identify Functional Determinants in Protein Families. **PLoS ONE**, v. 6, n. 12, art. e27786, dez. 2011.

BOROVINSKAYA, M. A. et al. Structural basis for hygromycin B inhibition of protein biosynthesis. **RNA**, v. 14, n. 8, p. 1590–1599, ago. 2008.

BRACHMANN, C. B. et al. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. **Yeast**, v. 14, n. 2, p. 115–132, jan. 1998.

BRANDMAN, O. et al. A Ribosome-Bound Quality Control Complex Triggers Degradation of Nascent Peptides and Signals Translation Stress. **Cell**, v. 151, n. 5, p. 1042–1054, nov. 2012.

BRANDMAN, O.; HEGDE, R. S. Ribosome-associated protein quality control. **Nature Structural & Molecular Biology**, v. 23, n. 1, p. 7–15, jan. 2016.

BRAUN, M. A et al. Identification of Rkr1, a nuclear RING domain protein with functional connections to chromatin modification in *Saccharomyces cerevisiae*. **Molecular and cellular biology**, v. 27, n. 8, p. 2800–11, 2007.

BRITO QUERIDO, J.; DÍAZ-LÓPEZ, I.; RAMAKRISHNAN, V. The molecular basis of translation initiation and its regulation in eukaryotes. **Nature Reviews Molecular Cell Biology**, v. 25, n. 3, p. 168–186, mar. 2024.

CAI, Z. et al. hnulp1, a basic helix-loop-helix protein with a novel transcriptional repressive domain, inhibits transcriptional activity of serum response factor. **Biochemical and Biophysical Research Communications**, v. 343, n. 3, p. 973–981, 13 mar. 2006.

CHAKRABORTY, A.; CHAKRABARTI, S. A survey on prediction of specificity-determining sites in proteins. **Briefings in Bioinformatics**, v. 16, n. 1, p. 71–88, jan. 2014.

CHANDRASEKARAN, V. et al. Mechanism of ribosome stalling during translation of a poly(A) tail.

Nature Structural & Molecular Biology, v. 26, n. 12, p. 1132–1140, dez. 2019.

CHOE, Y.-J. et al. Failure of RQC machinery causes protein aggregation and proteotoxic stress. **Nature**, v. 531, n. 7593, p. 191–195, mar. 2016.

CHU, J. et al. A mouse forward genetics screen identifies LISTERIN as an E3 ubiquitin ligase involved in neurodegeneration. **Proceedings of the National Academy of Sciences**, v. 106, n. 7, p. 2097–2103, fev. 2009.

COLOVOS, C.; YEATES, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. **Protein Science**, v. 2, n. 9, p. 1511–1519, set. 1993.

DANECEK, P. et al. Twelve years of SAMtools and BCFtools. **GigaScience**, v. 10, n. 2, art. giab008, jan. 2021.

DEFENOUILLE, Q. et al. Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. **Proceedings of the National Academy of Sciences**, v. 110, n. 13, p. 5046–5051, mar. 2013.

DEFENOUILLE, Q. et al. Rqc1 and Ltn1 Prevent C-terminal Alanine-Threonine Tail (CAT-tail)-induced Protein Aggregation by Efficient Recruitment of Cdc48 on Stalled 60S Subunits. **Journal of Biological Chemistry**, v. 291, n. 23, p. 12245–12253, jun. 2016.

DOBIN, A. et al. STAR: ultrafast universal RNA-seq aligner. **Bioinformatics**, v. 29, n. 1, p. 15–21, jan. 2013.

DORAN, P. M. Bioprocess engineering principles. Reprint ed. Amsterdam: **Elsevier/Acad. Press**, 2004.

EISENBERG, D.; LÜTHY, R.; BOWIE, J. U. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. **Methods in Enzymology**, v. 277, p. 396–404, 1997.

EWELS, P. et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. **Bioinformatics**, v. 32, n. 19, p. 3047–3048, out. 2016.

FONSECA-JÚNIOR, N. J. et al. PFstats: A Network-Based Open Tool for Protein Family Analysis. **Journal of Computational Biology**, v. 25, n. 5, p. 480–486, maio 2018.

GOWDA, N. K. C. et al. Hsp70 nucleotide exchange factor Fes1 is essential for ubiquitin-dependent degradation of misfolded cytosolic proteins. **Proceedings of the National Academy of Sciences**, v. 110, n. 15, p. 5975–5980, abr. 2013.

GRAILLE, M.; SÉRAPHIN, B. Surveillance pathways rescuing eukaryotic ribosomes lost in translation. **Nature Reviews Molecular Cell Biology**, v. 13, n. 11, p. 727–735, nov. 2012.

HAHN, J.-S.; THIELE, D. J. Activation of the *Saccharomyces cerevisiae* Heat Shock Transcription Factor Under Glucose Starvation Conditions by Snf1 Protein Kinase. **Journal of Biological Chemistry**, v. 279, n. 7, p. 5169–5176, fev. 2004.

HANAHAHAN, D. Studies on transformation of *Escherichia coli* with plasmids. **J. Mol. Biol.** v. 166, n. 4 p. 557-580, jun. 1983.

HARJU, S.; FEDOSYUK, H.; PETERSON, K. R. Rapid isolation of yeast genomic DNA: Bust n' Grab. **BMC Biotechnology**, v. 4, n. 1, p. 8, abr. 2004.

HARREMAN, M. T. et al. Regulation of Nuclear Import by Phosphorylation Adjacent to Nuclear Localization Signals. **Journal of Biological Chemistry**, v. 279, n. 20, p. 20613–20621, mar. 2004.

HELLEN, C. U. T. Translation Termination and Ribosome Recycling in Eukaryotes. **Cold Spring Harbor Perspectives in Biology**, v. 10, n. 10, art. a032656, out. 2018.

- HØIE, M. H. et al. NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. **Nucleic Acids Research**, v. 50, n. W1, p. 510–515, jul. 2022.
- INADA, T.; BECKMANN, R. Mechanisms of Translation-coupled Quality Control. **Journal of Molecular Biology**, v. 436, n. 6, p. 168496, mar. 2024.
- JOAZEIRO, C. A. P. Ribosomal Stalling During Translation: Providing Substrates for Ribosome-Associated Protein Quality Control. **Annual Review of Cell and Developmental Biology**, v. 33, n. 1, p. 343–368, out. 2017.
- JOAZEIRO, C. A. P. Mechanisms and functions of ribosome-associated protein quality control. **Nature Reviews Molecular Cell Biology**, v. 20, n. 6, p. 368–383, jun. 2019.
- JUMPER, J. et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, ago. 2021.
- KALDERON, D. et al. A short amino acid sequence able to specify nuclear location. **Cell**, v. 39, n. 3, p. 499–509, dez. 1984.
- KIBBE, W. A. OligoCalc: an online oligonucleotide properties calculator. **Nucleic Acids Research**, v. 35, p. 43–46, maio 2007.
- KOSTOVA, K. K. et al. CAT-tailing as a fail-safe mechanism for efficient degradation of stalled nascent polypeptides. **Science**, v. 357, n. 6349, p. 414–417, jul. 2017.
- KOSUGI, S. et al. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. **Proceedings of the National Academy of Sciences**, v. 106, n. 25, p. 10171–10176, jun. 2009.
- KUROHA, K. et al. Release of Ubiquitinated and Non-ubiquitinated Nascent Chains from Stalled Mammalian Ribosomal Complexes by ANKZF1 and Pth1. **Molecular Cell**, v. 72, n. 2, p. 286–302, out. 2018.
- KWOLEK-MIREK, M.; ZADRAG-TECZA, R. Comparison of methods used for assessing the viability and vitality of yeast cells. **FEMS Yeast Research**, n. 14, n. 7, p. 1068–1079, set. 2014.
- LANGE, A. et al. Classical Nuclear Localization Signals: Definition, Function, and Interaction with Importin α . **Journal of Biological Chemistry**, v. 282, n. 8, p. 5101–5105, fev. 2007.
- LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, abr. 1993.
- LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923–930, abr. 2014.
- LIU, X.-D. Conservation of a stress response: human heat shock transcription factors functionally substitute for yeast HSF. **The EMBO Journal**, v. 16, n. 21, p. 6466–6477, nov. 1997.
- LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 550, dez. 2014.
- LYUMKIS, D. et al. Structural basis for translational surveillance by the large ribosomal subunit-associated protein quality control complex. **Proceedings of the National Academy of Sciences**, v. 111, n. 45, p. 15981–15986, nov. 2014.
- MADEIRA, F. et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. **Nucleic Acids Research**, v. 52, n. W1, p. 521–525, jul. 2024.
- MASSER, A. E. et al. Cytoplasmic protein misfolding titrates Hsp70 to activate nuclear Hsf1. **eLife**, v. 8, p. e47791, set. 2019.

- MATSUO, Y. et al. RQT complex dissociates ribosomes collided on endogenous RQC substrate SDD1. **Nature Structural & Molecular Biology**, v. 27, n. 4, p. 323–332, abr. 2020.
- MENG, E. C. et al. UCSF ChimeraX: Tools for structure building and analysis. **Protein Science**, v. 32, n. 11, art. e4792, nov. 2023.
- MIRDITA, M. et al. ColabFold: making protein folding accessible to all. **Nature Methods**, v. 19, n. 6, p. 679–682, jun. 2022.
- MORANO, K. A.; GRANT, C. M.; MOYE-ROWLEY, W. S. The Response to Heat Shock and Oxidative Stress in *Saccharomyces cerevisiae*. **Genetics**, v. 190, n. 4, p. 1157–1195, abr. 2012.
- MÜHLHOFER, M. et al. The Heat Shock Response in Yeast Maintains Protein Homeostasis by Chaperoning and Replenishing Proteins. **Cell Reports**, v. 29, n. 13, p. 4593–4607, dez. 2019.
- MURRE, C. et al. Structure and function of helix-loop-helix proteins. **Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression**, v. 1218, n. 2, p. 129–135, jun. 1994.
- MURRE, C. Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. **Genes & Development**, v. 33, n. 1–2, p. 6–25, jan. 2019.
- NUÑO-CABANES, C. et al. A multi-omics dataset of heat-shock response in the yeast RNA binding protein Mip6. **Scientific Data**, v. 7, n. 1, art. 69, fev. 2020.
- NUSSBAUM, I. et al. Deteriorated Stress Response in Stationary-Phase Yeast: Sir2 and Yap1 Are Essential for Hsf1 Activation by Heat Shock and Oxidative Stress, Respectively. **PLoS ONE**, v. 9, n. 10, art. e111505, out. 2014.
- ØDUM, M. T. et al. DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. **Nucleic Acids Research**, v. 52, n. 1, p. 215–220, jul. 2024.
- OLSSON, M. et al. Nulp1, a novel basic helix-loop-helix protein expressed broadly during early embryonic organogenesis and prominently in developing dorsal root ganglia. **Cell and Tissue Research**, v. 308, n. 3, p. 361–370, jun. 2002.
- PEFFER, S.; GONÇALVES, D.; MORANO, K. A. Regulation of the Hsf1-dependent transcriptome via conserved bipartite contacts with Hsp70 promotes survival in yeast. **Journal of Biological Chemistry**, v. 294, n. 32, p. 12191–12202, ago. 2019.
- PEREZ-RIBA, A.; ITZHAKI, L. S. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. **Current Opinion in Structural Biology**, v. 54, p. 43–49, fev. 2019.
- POVEDA-HUERTES, D. et al. An Early mtUPR: Redistribution of the Nuclear Transcription Factor Rox1 to Mitochondria Protects against Intramitochondrial Proteotoxic Aggregates. **Molecular Cell**, v. 77, n. 1, p. 180–188, jan. 2020.
- RICHTER, K.; HASLBECK, M.; BUCHNER, J. The Heat Shock Response: Life on the Verge of Death. **Molecular Cell**, v. 40, n. 2, p. 253–266, out. 2010.
- ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature Protocols**, v. 5, n. 4, p. 725–738, abr. 2010.
- SANTIAGO, A. M.; GONÇALVES, D. L.; MORANO, K. A. Mechanisms of sensing and response to proteotoxic stress. **Experimental Cell Research**, v. 395, n. 2, art. 112240, out. 2020.
- SCHULLER, A. P.; GREEN, R. Roadblocks and resolutions in eukaryotic translation. **Nature Reviews Molecular Cell Biology**, v. 19, n. 8, p. 526–541, ago. 2018.
- SHAO, S. et al. Structure and Assembly Pathway of the Ribosome Quality Control Complex.

Molecular Cell, v. 57, p. 433–444, fev. 2015.

SHAO, S. et al. Decoding Mammalian Ribosome-mRNA States by Translational GTPase Complexes. **Cell**, n. 167, p. 1229–1240, set. 2016.

SHEN, P. S. et al. Rqc2p and 60 S ribosomal subunits mediate mRNA-independent elongation of nascent chains. **Science**, v. 347, n. 6217, p. 75–78, jan. 2015.

SIKORSKI, R.S.; HIETER, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. **Genetics**, v. 122, n. 1, p. 19–27, may 1989.

SITRON, C. S.; BRANDMAN, O. CAT tails drive degradation of stalled polypeptides on and off the ribosome. **Nature Structural & Molecular Biology**, v. 26, n. 6, p. 450–459, jun. 2019.

STEEN, H.; LINDHOLM, D. Nuclear localized protein-1 (Nulp1) increases cell death of human osteosarcoma cells and binds the X-linked inhibitor of apoptosis protein. **Biochemical and Biophysical Research Communications**, v. 366, n. 2, p. 432–437, fev. 2008.

TARCZEWSKA, A.; GREB-MARKIEWICZ, B. The Significance of the Intrinsically Disordered Regions for the Functions of the bHLH Transcription Factors. **International Journal of Molecular Sciences**, v. 20, n. 21, p. 5306, out. 2019.

TESINA, P. et al. Molecular basis of eIF5A-dependent CAT tailing in eukaryotic ribosome-associated quality control. **Molecular Cell**, v. 83, n. 4, p. 607–621, fev. 2023.

THE UNIPROT CONSORTIUM et al. UniProt: the Universal Protein Knowledgebase in 2025. **Nucleic Acids Research**, v. 53, n. D1, p. D609–D617, jan. 2025.

TSUCHIYA, H. et al. In Vivo Ubiquitin Linkage-type Analysis Reveals that the Cdc48-Rad23/Dsk2 Axis Contributes to K48-Linked Chain Specificity of the Proteasome. **Molecular Cell**, v. 66, n. 4, p. 488–502, maio 2017.

VALDEZ, R. M. S. Impacto da interrupção traducional no destino metabólico de polipeptídeos recém-sintetizados em *Saccharomyces cerevisiae*. **Dissertação** (Mestrado em Bioquímica e Imunologia) - Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais. Minas Gerais, 2018.

VAN KEMPEN, M. et al. Fast and accurate protein structure search with Foldseek. **Nature Biotechnology**, v. 42, n. 2, p. 243–246, fev. 2024.

VERGHESE, J. et al. Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (*Saccharomyces cerevisiae*) as a Model System. **Microbiology and Molecular Biology Reviews**, v. 76, n. 2, p. 115–158, jun. 2012.

VERMA, R. et al. Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. **eLife**, v. 2, art. e00308, jan. 2013.

VERMA, R. et al. Vms1 and ANKZF1 peptidyl-tRNA hydrolases release nascent chains from stalled ribosomes. **Nature**, v. 557, n. 7705, p. 446–451, may 2018.

WATERHOUSE, A. M. et al. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. **Bioinformatics**, v. 25, n. 9, p. 1189–1191, jan. 2009.

YONASHIRO, R. et al. The Rqc2/Tae2 subunit of the ribosome-associated quality control (RQC) complex marks ribosome-stalled nascent polypeptide chains for aggregation. **eLife**, v. 5, art. e11794, mar. 2016.

ZAHA, A.; FERREIRA, H.; PASSAGGLIA, L. *Biologia molecular básica*. **Artmed**, 2014.

ZENG, Q. et al. The bHLH Protein Nulp1 is Essential for Femur Development Via Acting as a Cofactor

in Wnt Signaling in Drosophila. **Current Molecular Medicine**, v. 17, n. 7, p. 509–517, mar. 2018.

ZHANG, X. et al. NULP1 Alleviates Cardiac Hypertrophy by Suppressing NFAT3 Transcriptional Activity. **Journal of the American Heart Association**, v. 9, n. 16, art. e016419, ago. 2020.

8 APÊNDICES

Apêndice I. Alinhamento múltiplo de sequências de proteínas homólogas de Rqc1p. A porcentagem de identidade foi destacada em tons de azul e os códigos de acesso do UniProt são: *Saccharomyces cerevisiae* (Q05468), *Homo sapiens* (Q9BQ70), *Mus musculus* (Q8R3L2), *Gallus gallus* (A0A8V0Y0J9), *Danio rerio* (Q68EJ4), *Xenopus laevis* (A3KMT4), *Drosophila melanogaster* (Q9Y109), *Caenorhabditis elegans* (P90919), *Dictyostelium discoideum* (Q8T2A4) e *Arabidopsis thaliana* (O80734).

	320	330	340	350	360	370
<i>S._cerevisiae/1-723</i>	DDW I P T Q R G E L S M K L L S S D D L L D W Q L W E R P L D W K D V I Q N D V S - - - - - Q W Q					
<i>H._sapiens/1-676</i>	STW P R Y S K P G L S M R L L - - - - - E S K					
<i>M._musculus/1-676</i>	STW P R Y S K P G L S M R L L - - - - - E S K					
<i>G._gallus/1-661</i>	NTW P R Y S K T G I A M Q L V - - - - - E T R					
<i>D._rerio/1-650</i>	DTW P R F S R P G I S M S L V - - - - - E S K					
<i>X._laevis/1-675</i>	STW P R Y S K T G I S M S L I - - - - - E T K					
<i>D._melanogaster/1-702</i>	DSW P A L T K N T I T M K L L P A P D S P S V S S K S - - - - - I T D					
<i>C._elegans/1-647</i>	PRW F P D R D Y G L S M V - - - - - E T K					
<i>D._discoideum/1-740</i>	PEW P D V M G G F M V M E I C N D D K V P S S P T N A T T T T T T T T T T T T N T V S P K L K S Q Q Q S L					
<i>A._thaliana/1-627</i>	ENW S R W D R S - F S M E F L - - - - - E T K					
	380	390	400	410	420	
<i>S._cerevisiae/1-723</i>	K F - - I S F Y K F E P L N S D L S K K S M M D F Y L S V I V H P D H E A L I N L I S S K F P Y H V P G L					
<i>H._sapiens/1-676</i>	K G - - L S F F A F E - - H S E E Y Q Q A Q H K F L V A V E S M - E P N N I V V L L Q T S - P Y H V D S L					
<i>M._musculus/1-676</i>	K G - - L S F F A F D - - H N E E Y Q Q A Q H K F L V A V E S M - E P N N I V V L L Q T S - P Y H V D S L					
<i>G._gallus/1-661</i>	R G - - V Q Y F T F E - - H H R E Y Q Q V Q F K F L D A V E S M - D P N N I V L L L Q M N - P Y H V D S L					
<i>D._rerio/1-650</i>	D G - - V Q H F T F E - - H N H D Y Q Q V Q F K F F E V V E S M - D P N N I V V L L Q I N - P Y H I D S L					
<i>X._laevis/1-675</i>	R G - - V Q Y F T F E - - H Q R D Y Q Q V Q F K F L D A V E S M - D P N N I V L L L Q M N - P Y H V D S L					
<i>D._melanogaster/1-702</i>	N G A D V Q W F A F E - - H S Q Y Y Q G V Q Q M F L S A L E R I - D S E F L I T L I K R C - P Y H V D S L					
<i>C._elegans/1-647</i>	K E G C I S W F T L R - - H N N A Y E Q R E R M F W M A E K R M - N V G V V Q E I F S D T - P Y H L N S I					
<i>D._discoideum/1-740</i>	D G - - I N Y F K I K - - W G Q T Y S N I Q E E F Y L A L T T H - D P M S L V E I I R Y H - P Y H I D S L					
<i>A._thaliana/1-627</i>	D G - - N N Y F R Y T - - H S S S Y E Q A Q R A F Q A A Q A I H - D L N G V A S V L I H N - P Y H I E S L					
	430	440	450	460	470	
<i>S._cerevisiae/1-723</i>	L Q V A L I F I R Q G D R S N T N G L L Q R A L F V F D R A L K A N I I F D S L N C Q L P Y I Y F F N R Q					
<i>H._sapiens/1-676</i>	L Q L S D A C R F Q E D Q E M A R D L V E R A L Y S M E C A F H P L F S L T S G A C R L D Y R R P E N R S					
<i>M._musculus/1-676</i>	L Q L S D A C R F Q E D Q E M A R D L I E R A L Y S M E C A F H P L F S L T S G T C R L D Y R R P E N R S					
<i>G._gallus/1-661</i>	L Q L S D V C R M Q E D Q E M A R D L V E R A L Y S L E C T F H P V F S L T S G S C R L D Y R R P E N R A					
<i>D._rerio/1-650</i>	L Q L S D V C R M Q E D Q E T A R D L I E R A L Y S F E C A F H P V F S L T S G T C R L D Y C R P E N R G					
<i>X._laevis/1-675</i>	L Q L S D V C R I Q E D Q E M A R D L I D R A L Y S L E C A F H P V F S L T S G T C R L D Y R R P E N R G					
<i>D._melanogaster/1-702</i>	V Q L S E V C K M T E D F S L A S E L L E R A L L L E S S L H I N F S L T S G N C R L D Y R R P E N R S					
<i>C._elegans/1-647</i>	L M M A H M N R M N E D F N T G A D L I E R G I W Y V D Q Y A A A T F E P F H W K H R M D Y L D Y E N R A					
<i>D._discoideum/1-740</i>	L Q L G Q V C L Q T A D Y S K A G D F V E M A V F A F E N V F H H L F N P L N G N C R F E Y R H D Q N K S					
<i>A._thaliana/1-627</i>	I T M A D Y F K F V G E N D M A A D S I G K C L Y G L E R A W H P M F T P F Q G N C R L E F N H D E N K L					
	480	490	500	510	520	
<i>S._cerevisiae/1-723</i>	F Y L A I F R Y I Q S L A Q R G V I G T A S E W T K V L W S L S P L E D P L G C R Y F L D H Y F L L N N D					
<i>H._sapiens/1-676</i>	F Y L A L Y K Q M S F L E K R G C P R T A L E Y C K L I L S L E P D E D P L C M L L L I D H L A L R A R N					
<i>M._musculus/1-676</i>	F Y L T L Y K Q M S F L E K R G C P R T A L E Y C K L I L S L E P D E D P L C M L L L I D H L A L R A R N					
<i>G._gallus/1-661</i>	F Y L A L F K H L M F L E K R G C P R T A L E F C K L I L S L D P E N D P L C V L L M I D F L S L R A R E					
<i>D._rerio/1-650</i>	F Y L A L F K H I Q F L E K R G C P R T A L E Y C K L I L S L D P E N D P L C M L L F F D F L C L R C H E					
<i>X._laevis/1-675</i>	F F L A L F K H M F L E K R G C P R T A L E F C K L I L S L D P E N D P L C M L L I D F L S L R A R E					
<i>D._melanogaster/1-702</i>	F Y I V L F K H A Q Y L E E R A C S R T A F E I S K L L L S L Q P D T D P L A M I L V I D Y Y A L R S K Q					
<i>C._elegans/1-647</i>	F Y L L L H R H M L N A A H K R O W E T A W N T A K M I F K L D P V K D P L A I I S L I D I F A L K A K Q					
<i>D._discoideum/1-740</i>	C F L A I F R F I Q I I G R R A C H R T A L E L C K I L L T W D - Y S D P L Y V R L I I D Y Y S I R S K Q					
<i>A._thaliana/1-627</i>	F F K T L F T H M R N M D R R G C H R S A L E V C K L L L S L D - T S N P V G A L F C V D Y F A L R A E E					
	540	550	560	570	580	
<i>S._cerevisiae/1-723</i>	Y Q Y I I E L S N S P - - - - - L M N C Y K Q W N T L G F S L A V V L S F L R I N E - - - - -					
<i>H._sapiens/1-676</i>	Y E Y L I R L F Q E W E - - - - - A H R N L S Q L P N F A F S V P L A Y F L L S Q Q T D L P E C E Q S S A					
<i>M._musculus/1-676</i>	Y E Y L I R L F Q E W E - - - - - A H R N L S Q L P N F A F S V P L A Y F L L S Q Q T D L P E H L S S A					
<i>G._gallus/1-661</i>	Y A F L T R M F Q E W E - - - - - S H R N L S Q L P N F A F S V P L A Y F F L S Q Q E H S E L E L S Q A					
<i>D._rerio/1-650</i>	Y S T L I R L Y E E W E - - - - - V H R N L S Q L P N F S F S V A L S Y F H L S Q D E D I S T V E S Q Q L					
<i>X._laevis/1-675</i>	Y T F L S R M F Q E W E - - - - - A H R N L S Q L P N F A F S V P L A L F Q H S Q Q D D L S A E E E K K A					
<i>D._melanogaster/1-702</i>	F E W L V E F Y E K Y N - - - - - A A R N L N Q L P N M A Y S Y A L A L H T L Y G A - - - - -					
<i>C._elegans/1-647</i>	Y Q W I I N T F E A A K - - - - - K F K K L H L L P N W P Y S V A L A R F F T A K T D E D - - - - -					
<i>D._discoideum/1-740</i>	Y R F L V D L F T K L K T I T P T N Q G D L S L L P N F C Y S A A L A M Y H L E R E K S T T T T S T S T S					
<i>A._thaliana/1-627</i>	Y A W L E Q F S E E Y R - - - - - N D N S L W L F P N F S Y S L A I A R V Y L E K M E P T S S S E A T P R					
	590	600	610	620	630	
<i>S._cerevisiae/1-723</i>	- - - - - M S S A R N A L L K A F K H H P L Q L S E L F K E - - - - - K L L G D					
<i>H._sapiens/1-676</i>	- - - - - R Q K A S L L I Q Q A L T M F P G V L L P L L E S C S V - - - - - R P D A - - S V S S H					
<i>M._musculus/1-676</i>	- - - - - R Q Q A S L L I Q Q A L T M F P G V L M P L L E Y C S V - - - - - R P D A - - T V S N H					
<i>G._gallus/1-661</i>	- - - - - R Q R A A R L I Q L A L I M F P S V L M P L L D H C S V - - - - - Q P D A - - R V A S H					
<i>D._rerio/1-650</i>	- - - - - K Q K S N V L L Q N A L I M F P G A L M P L L D L C T V - - - - - H L E A - - A V S S H					
<i>X._laevis/1-675</i>	- - - - - R E K A S H L I Q Q A L I M F P G V L L P L L D S C G V - - - - - Q P D A - - T V Q S H					
<i>D._melanogaster/1-702</i>	- - - - - C E R S N Q A L Q Y A L L M F P G V L R P L L D E M S V - - - - - Q T D K - - R V L A S					
<i>C._elegans/1-647</i>	- - - - - R E I A E N D L C T A I R H F P S V V V Q L M D I L Q I - - - - - H P D S - - A V M N C					
<i>D._discoideum/1-740</i>	T S S S K S S S V E E S S D I L L Q K A L I S F P M V L Q P L L E K L K T N F S I P K N G K I I T L E N D					
<i>A._thaliana/1-627</i>	D T S - - - - - K Q S S L D L M T Q A L K L H P T V L N K L V E K V P L - - - - - K D Q A W T K I L K H					

	960	970
<i>S._cerevisiae/1-723</i>	- - ENEESNENEG - - - - -	
<i>H._sapiens/1-676</i>	- - EAPHEDDAEGEGEW - D	
<i>M._musculus/1-676</i>	- - EVPREDNPEGEGDW - D	
<i>G._gallus/1-661</i>	- - PVPGTPTRRQS - - - - -	
<i>D._rerio/1-650</i>	- - QWEDNQERDEE - EW - D	
<i>X._laevis/1-675</i>	- - PREDNPDEEADA EW - D	
<i>D._melanogaster/1-702</i>	AQSTSSNDEEEGSSDY I D	
<i>C._elegans/1-647</i>	- - EEDDVDEEQELP - - - - -	
<i>D._discoideum/1-740</i>	- - DFSDEDENEN - - - - -	
<i>A._thaliana/1-627</i>	- - DAHHQPDNNG - - - - -	

Apêndice II. Previsão da localização subcelular de Rqc1p e suas proteínas homólogas previstas pela ferramenta DeepLoc-2.1. A tabela mostra as pontuações de previsão de localização das proteínas da família TCF25. Os limiares de probabilidade acima de 0.4761 indica provável localização citoplasmática e acima de 0.5014 provável localização nuclear. As células em verde destacam os resultados das previsões e a intensidade da cor indica o nível de certeza.

Espécie	Citoplasma	Núcleo	Extracelular	Membrana celular	Mitocôndria	Plastídio	Reticulo endoplasmático	Lisossomo/ Vacúolo	Aparato de Golgi	Peroxisomo
<i>S. cerevisiae</i>	0.6625	0.4957	0.0353	0.0339	0.1014	0.0024	0.2229	0.1593	0.3561	0.0646
<i>H. sapiens</i>	0.6098	0.5215	0.0533	0.0843	0.2261	0.0034	0.2799	0.3314	0.3167	0.0246
<i>M. musculus</i>	0.6081	0.5295	0.0475	0.0832	0.2228	0.0036	0.2718	0.3437	0.3301	0.0205
<i>G. gallus</i>	0.6181	0.4832	0.0498	0.0772	0.1871	0.0021	0.3186	0.2865	0.2931	0.0318
<i>D. rerio</i>	0.6753	0.4831	0.0543	0.0699	0.1333	0.0025	0.3421	0.2377	0.3404	0.0115
<i>X. laevis</i>	0.6289	0.4754	0.0563	0.0875	0.1720	0.0037	0.3525	0.2603	0.3261	0.0207
<i>D. melanogaster</i>	0.6951	0.4603	0.0554	0.0910	0.1113	0.0041	0.4268	0.1931	0.4460	0.0450
<i>C. elegans</i>	0.6065	0.4351	0.0541	0.0672	0.1082	0.0021	0.2025	0.0919	0.2377	0.0021
<i>D. discoideum</i>	0.6389	0.5258	0.0313	0.0556	0.1021	0.0022	0.1474	0.1774	0.1359	0.0045
<i>A. thaliana</i>	0.6304	0.5167	0.0351	0.0394	0.1703	0.0051	0.3049	0.1006	0.2630	0.0146

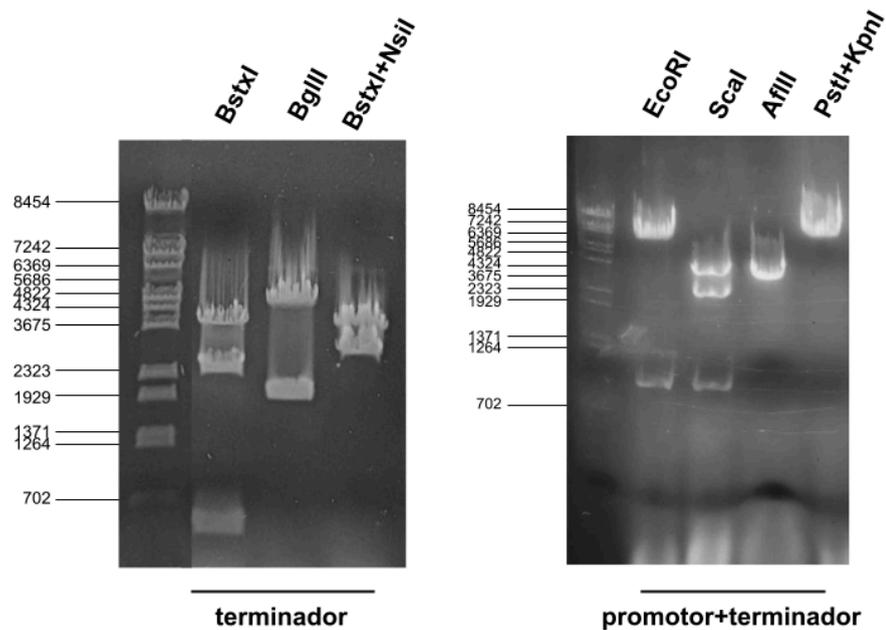
Apêndice III. Valores de C-score obtidos pela modelagem utilizando o I-TASSER. O C-score indica confiabilidade e varia entre -5 a 2, sendo que valores mais altos indicam maior confiança do modelo gerado.

Proteína	Modelo	C-score
Rqc1p	1	-2.51
	2	-2.03
	3	-2.39
	4	-1.75
	5	-4.59
Rqc1p ¹⁻³²⁷	1	-3.02
	2	-2.40
	3	-4.33
	4	-4.23
	5	-4.43
Rqc1p ³²⁸⁻⁷²³	1	-2.53
	2	-2.46
	3	-2.38
	4	-1.61
	5	-4.49

Apêndice IV. Resultados da busca por similaridade estrutural pelo FoldSeek. A tabela mostra a descrição dos principais *hits* obtidos pela busca por proteínas depositadas no PDB100 que possuem similaridade estrutural com o modelo gerado.

PDB	Descrição	Espécie	Identidade de sequência	RMSD	E-Value	Score
6SWU	Crystal structure of the TPR domain of KLC1 in complex with an engineered high-affinity cargo peptide	<i>Mus musculus</i>	11.9	11.31	1.64e-3	125
3NF1	Crystal structure of the TPR domain of kinesin light chain 1	<i>Homo sapiens</i>	12	9.3	4.42e-3	122
8BU0	Crystal structure of an 8 repeat consensus TPR superhelix with calcium	Construção sintética	15	4.31	6.19e-2	121
3CEQ	The TPR domain of Human Kinesin Light Chain 2 (hKLC2)	<i>Homo sapiens</i>	13.4	11.99	3.60e-3	120
3EDT	Crystal structure of the mutated S328N hKLC2 TPR domain	<i>Homo sapiens</i>	12.8	11.46	5.66e-3	117
3VTX	Crystal structure of MamA protein	<i>Candidatus Magnetobacterium bavaricum</i>	10.3	5.88	6.19e-2	115
5OJF	Crystal Structure of KLC2-TPR domain (fragment [A1-B6])	<i>Mus musculus</i>	12.8	10.51	8.21e-3	115
6EJN	The KLC2 TPR domain bound to the JIP3 leucine zipper domain	<i>Mus musculus</i>	11.5	10.68	1.14e-2	112
6FUZ	Crystal structure of the TPR domain of KLC1 in complex with the C-terminal peptide of JIP1	<i>Mus musculus</i>	12.4	11.04	5.01e-3	112
6EOU	O-GlcNAc transferase TPR domain with the intellectual disability associated mutation L254F	<i>Mus musculus</i>	10.7	11.42	8.91e-3	112

Apêndice V. Confirmação da clonagem por digestão diagnóstica. A análise das digestões diagnósticas das construções foi feita através da observação das bandas esperadas em gel de agarose 1%. O tamanho das bandas dos insertos esperados e observados foram: YCplac111+terminador: BstXI (3676 pb, 2389 pb e 526 pb), BgIII (4596 pb e 1998 pb) e BstXI+NsiI (3742 pb e 2849 pb) e YCplac111+promotor+terminador: EcoRI (6602 pb e 983 pb), Scal (3862 pb, 2725 pb e 995 pb), AflII (3832 pb, 3540 pb e 213 pb) e PstI e KpnI (7558 pb e 27 pb).



Apêndice VI. Descrição detalhada das bibliotecas dos estudos selecionados.

ID do projeto no SRA	ID das bibliotecas	Nome da biblioteca	Temperatura de cultivo e tempo choque térmico
PRJNA532914	SRR8902390	control_1	30°C
	SRR8902391	control_2	30°C
	SRR8902392	control_3	30°C
	SRR8902402	15min_1	37°C 15'
	SRR8902403	15min_2	37°C 15'
	SRR8902404	15min_3	37°C 15'
PRJNA559331	SRR9929263	control_1	30°C
	SRR9929264	control_2	30°C
	SRR9929273	control_3	30°C
	SRR9929282	control_4	30°C
	SRR9929265	20min_1	39°C 20'
	SRR9929271	20min_2	39°C 20'
	SRR9929274	20min_3	39°C 20'
	SRR9929280	20min_4	39°C 20'
	SRR9929266	120min_1	39°C 120'
	SRR9929272	120min_2	39°C 120'
	SRR9929279	120min_3	39°C 120'
	SRR9929281	120min_4	39°C 120'
PRJNA312727	SRR3180780	control_1	25°C
	SRR3180781	control_2	25°C
	SRR3180782	control_3	25°C
	SRR9007023	30min_1	37°C 30'
	SRR9007024	30min_2	37°C 30'
	SRR9007025	30min_3	37°C 30'
PRJNA498270	SRR9993596	control_1	23°C
	SRR9993588	control_2	23°C
	SRR9993587	control_3	23°C
	SRR9993591	2h_1	37°C 2"
	SRR9993590	2h_2	37°C 2"
	SRR9993589	2h_3	37°C 2"

Apêndice VII. Genótipo das linhagens utilizadas nos estudos.

Linhagem	Genótipo	Referência
BY4741	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	BRACHMANN et al., 1998
CAY1015	<i>MATa his3Δ1 leu2Δ0 ura3Δ0</i>	GOWDA et al., 2013
YPH499	<i>MATa, ade2-101, his3-Δ200, leu2-Δ1, ura3-52, trp1 Δ63, lys2-801</i>	SIKORSKI AND HIETER, 1989

Apêndice VIII. Percentual de mapeamento das reads utilizando o programa STAR.

ID do projeto no SRA	Bibliotecas	Reads de mapeamento único (%)	Reads de mapeamento em múltiplos loci (%)	Reads não mapeadas (%)
PRJNA559331	SRR9929263	71.82	18.96	6.00
	SRR9929264	80.57	8.78	9.21
	SRR9929265	80.57	8.64	9.12
	SRR9929266	77.50	11.77	7.78
	SRR9929271	81.44	9.01	7.73
	SRR9929272	76.82	8.05	12.62
	SRR9929273	78.54	10.99	8.19
	SRR9929274	76.65	8.40	12.86
	SRR9929279	82.31	8.53	7.42
	SRR9929280	78.65	9.93	9.64
	SRR9929281	79.21	11.35	7.18
	SRR9929282	75.45	12.49	9.99
PRJNA312727	SRR9007023	90.44	4.40	4.08
	SRR3180782	89.54	4.28	5.49
	SRR9007025	90.07	4.00	4.89
	SRR3180781	90.04	4.53	4.71
	SRR9007024	90.75	3.97	4.27
	SRR3180780	88.79	4.54	6.02
	SRR8902390	91.14	3.82	4.44

PRJNA532914	SRR8902391	92.06	3.80	3.51
	SRR8902392	91.66	3.85	3.83
	SRR8902402	90.43	4.01	4.81
	SRR8902403	90.84	4.16	4.27
	SRR8902404	91.03	4.02	4.19
PRJNA498270	SRR9993587	91.67	4.52	3.06
	SRR9993588	92.09	4.65	2.67
	SRR9993589	91.20	4.15	3.22
	SRR9993590	91.61	4.01	3.16
	SRR9993591	91.19	4.63	3.11
	SRR9993596	91.99	4.80	2.46

Apêndice XIX. Percentual de reads anotadas utilizando o FeatureCounts.

ID do projeto no SRA	Bibliotecas	Reads anotadas FeatureCounts (%)
PRJNA559331	SRR9929263	66
	SRR9929264	81.3
	SRR9929265	77.8
	SRR9929266	74.7
	SRR9929271	77.3
	SRR9929272	81.6
	SRR9929273	78.2
	SRR9929274	80
	SRR9929279	78
	SRR9929280	75
	SRR9929281	74.3
	SRR9929282	73
PRJNA312727	SRR9007023	93.5
	SRR3180782	94.4
	SRR9007025	93.7
	SRR3180781	93.9

	SRR9007024	93.8
	SRR3180780	93.7
PRJNA532914	SRR8902390	96
	SRR8902391	96.1
	SRR8902392	96.1
	SRR8902402	93.8
	SRR8902403	93.5
	SRR8902404	93.5
PRJNA498270	SRR9993587	90.4
	SRR9993588	88.4
	SRR9993589	89.5
	SRR9993590	90
	SRR9993591	88
	SRR9993596	91