

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

MURILO LEÃO PEREIRA

Estudo de ORC1/CDC6 de *Trypanosoma cruzi*

SÃO CARLOS
2023

MURILO LEÃO PEREIRA

Estudo de ORC1/CDC6 de *Trypanosoma cruzi*

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada
Opção Biomolecular

Orientador: Prof. Dr. Otavio Henrique Thiemann

Versão corrigida
(versão original disponível na Unidade que aloja o Programa)

SÃO CARLOS
2023

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Pereira, Murilo Leão

Estudo de ORC1/CDC6 de *Trypanosoma cruzi* / Murilo Leão Pereira; orientador Otávio Henrique Thiemann - versão corrigida -- São Carlos, 2023.

95 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Aplicada Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2023.

1. Replicação. 2. DNA. 3. Complexo pré-replicação. 4. Doença de Chagas. I. Thiemann, Otávio Henrique, orient. II. Título.

Dedico esta obra à minha família.

AGRADECIMENTOS

Fazer ciência pode levar a caminhos muito imprevisíveis e tortuosos. Esses anos de mestrado foram um exercício de paciência e perseverança que só está sendo superado por conta de várias pessoas que merecem toda a minha gratidão.

Aos meus pais, Nazaré e Jorge, que fazem da distância presença e me apoiam incondicionalmente. Aos meus irmãos, Camila e Matheus, que são meus maiores parceiros e professores na vida. À minha avó, Jacy, um exemplo de força, garra e amor que muito me inspira.

Ao meu orientador, Prof. Dr. Otávio Thiemann, pelas discussões, paciência, confiança e por participar de toda a minha formação.

Aos técnicos de laboratório, Susana, Andressa e Humberto por todo o auxílio na execução dos experimentos. Em especial, tenho que agradecer à Susana por todo o companheirismo no laboratório.

Aos meus colegas de laboratório que conhecem os desafios e batalham todos os dias.

Aos meus amigos por tornarem tudo mais leve. Em especial, quero agradecer ao Matheus por ter acompanhado cada reviravolta durante esse período.

À USP e ao CNPq por terem possibilitado este trabalho.

"Persistence is the shortest path to success"
(Charlie Chaplin)

RESUMO

PEREIRA, M. L. **Estudo de ORC1/CDC6 de *Trypanosoma cruzi***. 2023. 95p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2023.

A replicação do material genético é essencial para a manutenção e propagação da vida. No caso do DNA em eucariotos, esse processo se inicia pelo Complexo Pré-Replicação (pre-RC). Para sua formação, ocorre o reconhecimento de Origens de Replicação (ORIs) no genoma pelo ORC₁₋₆, seguido da ligação aos fatores de licenciamento CDC6 e CDT1 e recrutamento da helicase MCM₂₋₇. O parasita responsável pela doença de Chagas *Trypanosoma cruzi* apresenta apenas ORC1 do hexâmero de reconhecimento de ORI e essa proteína é intimamente relacionada em sequência a CDC6. A predição da estrutura de TcORC1/CDC6 por diferentes métodos computacionais, seu estudo filogenético em protozoários e preliminar ensaios biofísicos foram o foco do trabalho. Os modelos estruturais gerados por I-TASSER, Robbeta, AlphaFold2 e homologia compartilharam grande similaridade principalmente no provável sítio ativo da proteína, e a comparação dos modelos indicou AlphaFold2 como o método com melhor predição. Consultas comparativas de estruturas proteicas evidenciaram a qualidade da predição. As análises filogenéticas traçaram um histórico evolutivo para a proteína e apontaram em nível molecular divisões que são biologicamente válidas. Os experimentos de expressão e purificação da proteína apresentaram dificuldade em manter a sua estabilidade. Por fim, a identificação de TcORC1/CDC6 pelo anticorpo anti-ORC1 apontou incorretamente a Proteína de Ligação à Maltose (MBP), e isso culminou na caracterização biofísica ter sido realizada com uma molécula diferente da de interesse. Os resultados abrem caminho para uma futura caracterização biofísica e estrutural de TcORC1/CDC6 e para o melhor entendimento da via de replicação no parasita.

Palavras-chaves: Replicação. DNA. Complexo pré-replicação. Doença de Chagas.

ABSTRACT

PEREIRA, M. L. **Study of ORC1/CDC6 from *Trypanosoma cruzi***. 2023. 95p. Dissertation (Master in Science) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2023.

The replication of genetic material is essential for the maintenance and propagation of life. In the case of DNA in eukaryotes, this process begins with Pre-Replication Complex (pre-RC). For its formation, there is recognition of (Origins of Replication (ORIs) in the genome by ORC₁₋₆, followed by binding to the licensing factors CDC6 and CT1 and recruitment of the MCM₂₋₇ helicase. The parasite responsible for Chagas Disease *Trypanosoma cruzi* has only ORC1 of the ORI recognition hexamer and this protein is closely related in sequence to CDC6. The prediction of the structure of TcORC1/CDC6 by different computational methods, its phylogenetic study in Protozoa and preliminary biophysical assays were the focus of this work. The structural models generated by I-TASSER, Robbeta, AlphaFold2 and homology shared great similarity mainly in the probable active site of the protein, and the comparison of the models indicated AlphaFold2 as the method with the best prediction. Comparative queries of protein structures showed the quality of the prediction. Phylogenetic analyses traced an evolutionary history for the protein and pointed out divisions at the molecular level that are biologically valid. The protein expression and purification experiments showed difficulty in maintaining the stability of the protein. Finally, the identification of TcORC1/CDC6 by the anti-ORC1 antibody incorrectly identified Maltose Binding Protein (MPB), and this culminated in the biophysical characterization being performed with a different molecule than the one of interest. The results pave the way for a future biophysical and structural characterization of TcORC1/CDC6 and for a better understanding of the replication pathway in the parasite.

Keywords: Replication. DNA. Pre-replication complex. Chagas disease.

LISTA DE FIGURAS

- Figura 1 - Esquema representativo do ciclo da Doença de Chagas.....22
- Figura 2 - Esquema representativo da formação do CPR. Passo 1: Reconhecimento de ORI por ORC. Passo 2: Interação com CDC6. Passo 3: Ligação a CDT1. Passo 4: Recrutamento de MCM.....24
- Figura 3 - Esquema representativo de ORC em *Trypanosoma brucei*, indicando o grau de conservação das subunidades.....25
- Figura 4 - Representação da montagem da placa para o teste de concentração. Os espaços vazios representam cada poço.....36
- Figura 5 - Motivo AAA extraído do PDB 1IY2.....40
- Figura 6 - Diagramas de Ramachandran para diferentes modelagens computacionais. (a) I-TASSER. (b) Robbeta. (c) AlphaFold2. (d) Modelagem.....41
- Figura 7 - Representação em *cartoon* dos modelos gerados pelos diferentes métodos de predição de estrutura proteica: (a) I-TASSER, (b) Robbeta, (c) AlphaFold2 e (d) homologia.....46
- Figura 8 - Sobreposição da representação em *cartoon* obtida pelo AlphaFold2 contra (a) I-TASSER, (b) Robbeta e (c) homologia. (d) apresenta a estrutura do AlphaFold2 isolada.....48
- Figura 9 - Representação em *cartoon* do motivo AAA dos modelos gerados pelos diferentes métodos de predição de estrutura proteica. (a) I-TASSER. (b) Robbeta. (c) AlphaFold2. (d) Modelagem.....50
- Figura 10 - Sobreposição da representação em *cartoon* obtida pelo AlphaFold2 do motivo AAA contra (a) I-TASSER, (b) Robbeta e (c) homologia. (d) apresenta a estrutura do AlphaFold2 isolada.....52
- Figura 11 - Modelo de TcORC1/CDC6 gerado pelo I-TASSER em dois ângulos distintos (a) e (b) com regiões de *loops* mal resolvidos sendo indicadas por uma seta.....55
- Figura 12 - Árvore construída por método de Máxima Verossimilhança. Destaca-se em verde, azul e vermelho os 3 nós escolhidos para a predição da sequência ancestral.....60
- Figura 13 - Árvore construída pelo método de Agrupamento de Vizinhos.....61
- Figura 14 - Árvore construída pelo método de Máxima parcimônia.....62
- Figura 15 - Representação em *cartoon* dos modelos das sequências ancestrais dos nós: (a) Tryp_Leish, (b) Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.....66
- Figura 16 - Representação em *cartoon* da sobreposição de TcORC1 contra os modelos das sequências ancestrais dos nós: (a) Tryp_Leish, (b)

Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.....	68
Figura 17 - Representação em <i>cartoon</i> do motivo AAA dos modelos das sequências ancestrais dos nós: (a) Tryp_Leish, (b) Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.....	70
Figura 18 - Representação em <i>cartoon</i> da sobreposição do motivo AAA de TcORC1 contra o motivo AAA dos modelos das sequências ancestrais dos nós: (a) Tryp_Leish, (b) Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.....	72
Figura 19 - Purificação por Cromatografia de Afinidade de TcORC1/CDC6. (a) cromatograma da purificação. (b) gel 10% de SDS-PAGE das frações da expressão e purificação de TcORC1/CDC6.....	75
Figura 20 - Purificação por Cromatografia de Exclusão Molecular de TcORC1/CDC6. (a) cromatograma da purificação. (b) gel 10% de SDS-PAGE das frações da purificação.....	76
Figura 21 - Resultado do <i>Western-blot</i> das amostras de expressão, purificação por cromatografia de afinidade e por cromatografia de exclusão molecular de TcORC1/CDC6.....	77
Figura 22 - Cromatograma da Cromatografia de Exclusão Molecular com coluna analítica de TcORC1/CDC6.....	78
Figura 23 - Curvas de calibração da coluna Superdex™ 200 10/300 (GE) analítica de 24 mL. (a) calibração para massa molecular. (b) calibração para raio hidrodinâmico.....	79
Figura 24 - Resultado do Espalhamento Dinâmico da Luz de TcORC1/CDC6.....	80
Figura 25 - Desnaturação térmica de TcORC1/CDC6 na presença de Sypro Orange. Em destaque, está a condição escolhida. (a) gráfico da desnaturação em função da temperatura. (b) primeira derivada da curva de desnaturação em função da temperatura.....	81
Figura 26 - Histogramas comparativos da variação do Tm nas diferentes condições de teste.....	82
Figura 27 - Curva de desnaturação das condições do teste de tampão cuja variação relativa de Tm foi positiva.....	83
Figura 28 - Espectro de CD de TcORC1/CDC6.....	84
Figura 29 - Espectros de variação estrutural por temperatura pelo sinal a 222 nm. (a) espectro de sinal de CD por temperatura. (b) primeira derivada da curva de desnaturação pela temperatura.....	84
Figura 30 - Cromatograma resultado da análise por SEC-MALS de TcORC1/CDC6.....	85

LISTA DE TABELAS

Tabela 1 -	Composição de TcORC1/DCC6 em aminoácidos (AA) em porcentagem.....	39
Tabela 2 -	Análise estatística do Diagrama de Ramachandran construído a partir de cada preditor.....	44
Tabela 3 -	RMSD para a sobreposição dos modelos gerados por diferentes preditores contra TcORC1/CDC6 do AlphaFold2.....	54
Tabela 4 -	Moléculas relevantes do servidor DALI por código PBD contra a base de dados do PDB.....	56
Tabela 5 -	Moléculas relevantes do servidor DALI por espécie contra o AF-DB.....	57
Tabela 6 -	Organismos escolhidos para a análise filogenética com grau de identidade, número de aminoácidos e esquema representativo de motivo proteico. C é a cobertura; Id, a identidade; e aa, aminoácidos.....	59
Tabela 7 -	Dados biológicos e geográficos das espécies analisadas no estudo filogenético.....	63
Tabela 8 -	RMSD para a sobreposição dos modelos ancestrais contra TcORC1/CDC6.....	74
Tabela 9 -	Padrões para calibração da coluna. MM é massa molar. R_H é raio hidrodinâmico. V_e é volume de eluição. K_{av} é a razão de eluição.....	78
Tabela 10 -	Parâmetros biofísicos obtidos por diferentes técnicas para a proteína purificada no trabalho.....	87

LISTA DE SIGLAS

BLAST	<i>Basic Local Alignment Search Tool</i> (tradução: Ferramenta Básica de Pesquisa de Alinhamento Local)
CD	Dicroísmo Circular.
CDC6	Complexo de Controle de Divisão Celular 6.
CDC/DPDX	<i>Centers for Disease Control and Prevention/Laboratory Identification of Parasites of Public Health Concern</i> (tradução: Centros de Controle e Prevenção de Doenças/Laboratório de Identificação de Parasitas de Interesse para a Saúde Pública).
CDD/SPARCLE	The Conserved Domain Database (tradução: A Base de Dados de Domínios Conservados).
CDT1	Complexo do Fator de Replicação de DNA e Licenciamento de Cromatina 1.
DmORC1	Complexo de Reconhecimento de Origem Subunidade 1 de <i>Drosophila melanogaster</i> .
DNA	Ácido Desorribonucleico.
DNDi	<i>Drugs for Neglected Diseases initiative</i> (tradução: Iniciativa de Drogas para Doenças Negligenciadas).
DLS	Espalhamento Dinâmico da Luz.
DO	Densidade óptica.
DSF	Fluorimetria de Varredura Diferencial.
GRAVY	Índice de estabilidade e hidropaticidade média.
IPTG	Isopropil β -D-1-tiogalactopiranosida.
LALI	Número de resíduos estruturalmente equivalentes.
MBP	Proteína de ligação à maltose.
MCM	Complexo de de Manutenção de Minicromossomos.
NRES	Número de resíduos na estrutura alvo.
OMS	Organização Mundial de Saúde.
ORC	Complexo de Reconhecimento de Origem.
ORI	Origem de replicação.
PBS	Tampão Fosfato Salino.
PDB	<i>Protein Data Bank</i> (tradução: Banco de Dados de Proteínas).
pI _{teórico}	Ponto isoelétrico teórico.

PMSF	Fluoreto de fenilmetilsulfonil.
pre-RC	Complexo de Pré-Replicação.
NTDs	<i>Neglected Tropical Diseases</i> (tradução: Doenças Tropicais Negligenciadas).
RMSC	Raiz quadrada do desvio quadrático médio.
RNA	Ácido Ribonucleico.
Sarcosil	Lauroil sarcosinato de sódio.
SEC	Cromatografia de Exclusão Molecular.
SEC-MALS	Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho.
SDS-PAGE	Dodecil sulfato de sódio - Eletroforese com gel de poliacrilamida.
TcORC1/CDC6	Complexo de Reconhecimento de Origem Subunidade 1/Proteína de Controle de Divisão Celular 6 de <i>Trypanosoma cruzi</i> .
UCLA-DOE LAB - Saves v.6.0: University of California-The U.S. Department of Energy Laboratory - Saves v.6.0 (tradução: Laboratório da Universidade da Califórnia-Departamento de Energia dos Estados Unidos – Saves v.6.0).	

SUMÁRIO

1	INTRODUÇÃO	21
1.1	TRIPANOSOMATÍDEOS	21
1.2	DOENÇAS DE CHAGAS	21
1.3	REPLICAÇÃO	23
1.4	REPLICAÇÃO EM TRIPANOSOMATÍDEOS	24
2	OBJETIVOS	27
2.1	OBJETIVOS GERAIS	27
2.2	OBJETIVOS ESPECÍFICOS	27
3	MATERIAIS E MÉTODOS	29
3.1	CARACTERIZAÇÃO BIOC COMPUTACIONAL	29
3.1.1	Análise da sequência	29
3.1.2	Predição da estrutura	29
3.1.3	Filogenética	30
3.2	PRODUÇÃO DA PROTEÍNA RECOMBINANTE	31
3.2.1	Expressão	31
3.2.2	Lise	32
3.2.3	Purificação	32
3.2.3.1	Cromatografia de afinidade	32
3.2.3.2	Cromatografia de exclusão molecular	33
3.2.3.3	<i>Western-blot</i>	33
3.3	CARACTERIZAÇÃO BIOFÍSICA	34
3.3.1	Curva de calibração da Exclusão Molecular	34
3.3.2	Espalhamento Dinâmico da Luz (DLS)	35
3.3.3	Fluorimetria de Varredura Diferencial (DSF)	35
3.3.3.1	Teste de concentração	35
3.3.3.2	<i>Screen</i> de tampão	36
3.3.4	Dicroísmo Circular (CD)	36
3.3.5	Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho (SEC-MAL)	37
3.3.6	Espectrometria de Massas	37
4	RESULTADOS E DISCUSSÃO	39
4.1	CARACTERIZAÇÃO BIOC COMPUTACIONAL	39
4.1.1	Análise da sequência	39
4.1.2	Predição da estrutura	41

4.1.3	Filogenética	58
4.2	PRODUÇÃO DA PROTEÍNA RECOMBINANTE	74
4.2.1	Purificação.....	74
4.2.1.1	Cromatografia de afinidade	74
4.2.1.2	Cromatografia de exclusão molecular	75
4.2.2	Western-blot	76
4.3	CARACTERIZAÇÃO BIOFÍSICA	77
4.3.1	Curva de calibração da Exclusão Molecular	77
4.3.2	Espalhamento Dinâmico da Luz (DLS).....	79
4.3.3	Fluorimetria de Varredura Diferencial (DSF)	80
4.3.3.1	Teste de concentração	80
4.3.3.2	Screen de tampão	81
4.3.4	Dicroísmo Circular (CD)	83
4.3.5	Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho (SEC-MAL) 85	
4.3.6	Espectrometria de massas	85
5	CONCLUSÃO E PERSPECTIVAS	87
	REFERÊNCIAS	89
	ANEXOS	93
	ANEXO A - SCRIPT PYTHON PARA MODELAGEM	93
	ANEXO B - SEQUÊNCIAS ANCESTRAIS.....	95

1 INTRODUÇÃO

1.1 TRIPANOSOMATÍDEOS

Tripanosomatídeos são protozoários unicelulares flagelados pertencentes à ordem Kinetoplastida, cujos integrantes patogênicos humanos de maior relevância são *Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania spp.* Esses parasitas são, respectivamente, os agentes etiológicos da Doença de Chagas, da Tripanossomíase Africana e de diversas Leishmanioses.¹

O ciclo de vida dessas doenças é díxênico, possuindo como hospedeiro intermediário um inseto e hospedeiro final um mamífero. Por conta da característica do hospedeiro intermediário, essas tripanosomíases concentram-se em países tropicais, em especial nos pertencentes à América Latina, África e Ásia.²

Essas doenças estão entre as categorizadas pela Fundação Rockefeller em 1970 como Doenças Tropicais Negligenciadas (NTDs, do inglês *Neglected Tropical Diseases*). Essa classificação fez parte de um esforço global para incentivar o estudo de doenças infecciosas que afetam principalmente a população em pobreza extrema (menos de U\$1,90 por dia).³

A espécie foco do trabalho é *Trypanosoma cruzi*, e seus primeiros exemplares foram isolados por Carlos Chagas em 1909 no estado de Minas Gerais do intestino posterior de um inseto hematófago do gênero Reduviidae, subfamília Triatominae, popularmente denominado barbeiro.⁴

1.2 DOENÇAS DE CHAGAS

O ciclo da Doença de Chagas apresenta variações morfológicas e funcionais e se inicia pelo triatoma realizando uma refeição de sangue de um animal infectado com *Trypanosoma cruzi*. A forma epimastigota é replicada no intestino do vetor e diferencia-se para a forma com alta mobilidade tripomastigota metacíclico. Esta é excretada nas fezes do inseto durante uma próxima refeição e entra em contato com a corrente sanguínea do hospedeiro por mucosas ou ferimentos na pele.

No hospedeiro, as células dos mais diversos tecidos podem ser infectadas e, dentro delas, ocorre a transição para a forma replicativa amastigota. Com uma população parasitária alta dentro da célula, a diferenciação para tripomastigota metacíclico acontece e ocorre a lise celular. Uma vez na corrente sanguínea, o

protozoário busca novas células para infectar⁵. O ciclo da doença está apresentado na Figura 1.

De acordo com a *Drugs for Neglected Diseases initiative* (DNDi)⁶ em dezembro de 2022, a Doença de Chagas é um problema de saúde global que afeta mais de 6 milhões de pessoas, sendo que anualmente ocorre cerca de 30 mil novas infecções e de 12 mil mortes. Estima-se que há mais de 75 milhões de pessoas em risco de infecção e que apenas 30% das pessoas infectadas são diagnosticadas. Em termos geográficos, é endêmica em 21 países na América Latina e está presente na América do Europa, Europa, Japão e Austrália.

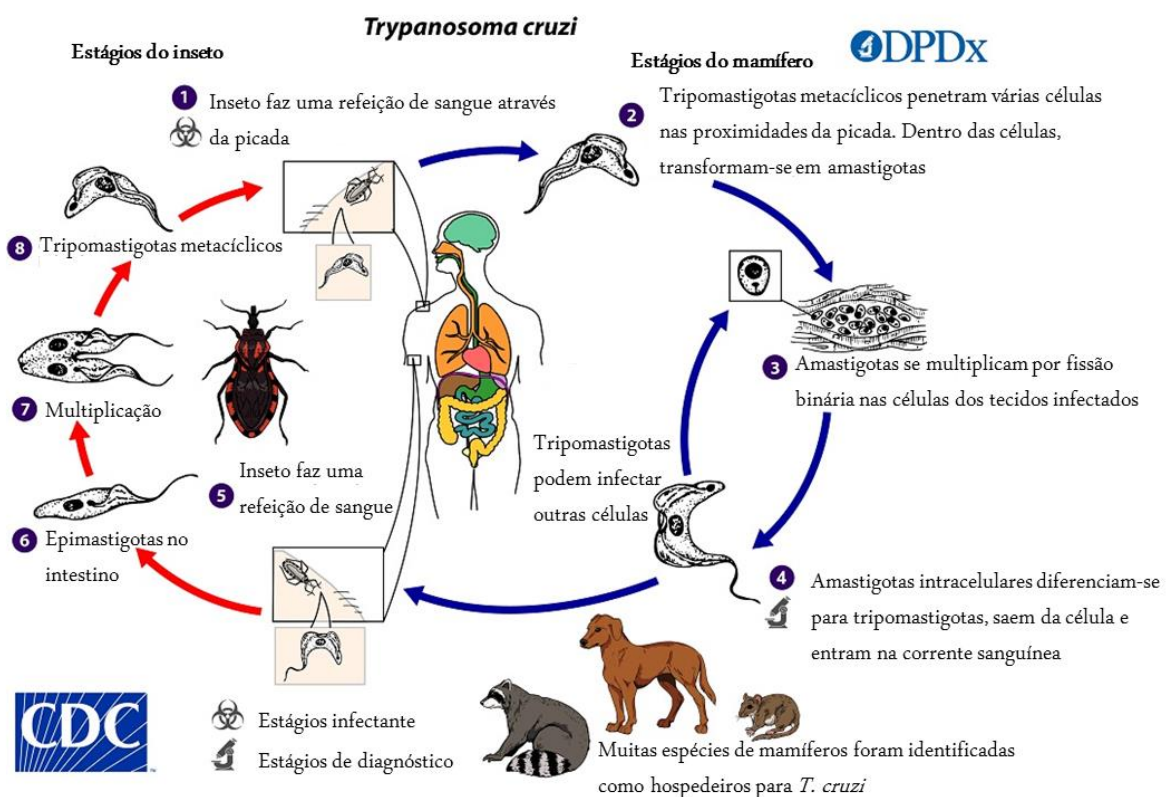


Figura 1 - Esquema representativo do ciclo da Doença de Chagas.

Fonte: Adaptada de CDC/DPDx.⁷

O diagnóstico dessa NTD pode ser realizado em qualquer fase e consiste na análise clínica, epidemiológica e laboratorial, podendo envolver na fase aguda testes parasitológicos com o sangue do paciente e na fase crônica ensaios imunológicos para detectar anticorpos contra o parasito. O tratamento envolve os medicamentos antiparasitários nifurtimox e benznidazole e é frequentemente interrompido por ser prologando (60 a 90 dias) e pelos efeitos adversos.⁸

Para nifurtimox, os efeitos adversos mais observados são anorexia, perda de peso, alterações psíquicas, sono, náusea e vômito. Para benznidazole, as reações adversas mais graves são depressão da medula óssea e agranulocitose, sendo também notado irritações manifestações na pele como hipersensibilidade, dermatite com erupção cutânea e dor nas articulações e muscular.⁹

É importante ressaltar que hoje no Brasil a obtenção do benznidazol é assegurada de forma gratuita pelo Ministério da Saúde, enquanto o nifurtimox não é mais comercializado desde a década de 1980, sendo seu acesso possível apenas através de um acordo entre a Organização Mundial de Saúde (OMS) e a Bayer, não existindo o fornecimento direto.¹⁰

1.3 REPLICAÇÃO

O genoma dos organismos, com exceção de vírus de ácido ribonucleico (RNA), é constituído por ácido desoxirribonucleico (DNA), o qual deve ser replicado durante o ciclo celular para a propagação da vida. O início da replicação se dá pelo reconhecimento de regiões específicas nomeadas de Origens de Replicação (ORIs) e, em eucariotos, é um processo com múltiplas etapas e dependente de ATP. Esse reconhecimento é feito pelo Complexo de Pré-Replicação (pre-RC), composto por: Complexo de Reconhecimento de Origem (ORC₁₋₆), Proteína de Controle de Divisão Celular 6 (CDC6), Fator de Replicação de DNA e Licenciamento de Cromatina 1 (CDT1) e Complexo de Manutenção de Minicromossomos (MCM₂₋₇).¹¹⁻¹²

A formação do pre-RC, de forma simplificada, começa com ORC reconhecendo as ORIs presentes no DNA e se anelando a elas. Em seguida, ORC interage com CDC6, o que permite a ligação à CDT1. CDC6 e CDT1 são fatores de replicação essenciais para o licenciamento de origem. Por fim, o complexo transiente ORC-CDC6-CDT1 recruta a helicase MCM e daí se inicia a replicação.¹³ Esses passos estão esquematizados na Figura 2.

O processo de replicação é extremamente sofisticado e coordena a unicidade de duplicação de cada segmento de DNA por célula. E, por mais que esse seja bem conservado, há divergências significativas na estrutura de ORIs e na formação do pre-RC ao longo da evolução.¹⁴

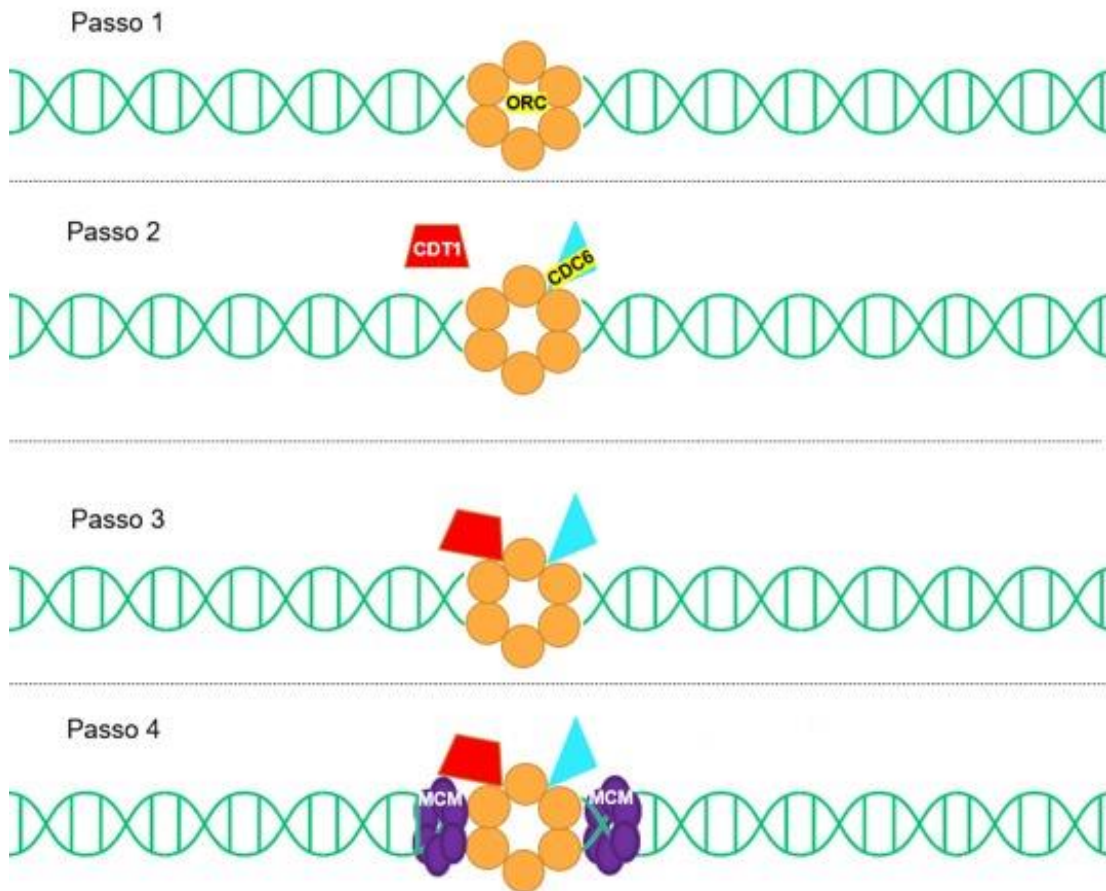


Figura 2 - Esquema representativo da formação do CPR. Passo 1: Reconhecimento de ORI por ORC. Passo 2: Interação com CDC6. Passo 3: Ligação a CDT1. Passo 4: Recrutamento de MCM.

Fonte: Adaptada de LIMA *et al.*¹³

1.4 REPLICAÇÃO EM TRIPANOSOMATÍDEOS

O sequenciamento do genoma de *Trypanosoma cruzi* identificou inicialmente apenas uma proteína relacionada a ORC que apresenta similaridade estrutural com ORC1/CDC6 de Archaea e, por isso, foi nomeada TcORC1/CDC6 (ORC1/CDC6 de *Trypanosoma cruzi*).¹⁵

A espécie *Trypanosoma brucei* é o exemplar do gênero que primeiramente teve elementos do pre-RC caracterizados e nele, além de ORC1/CDC6, há a presença TbORC1B, TbORC4, TbORC2 e TbORC5 atuando na replicação de DNA, sendo que TbORC4 aparenta ser uma ortóloga distante de TbORC4 e TbORC2 e TbORC5 apresentam baixa similaridades com suas homólogas. Há ainda indícios de que TbORC1/CDC6, TbORC4, TbORC2 e TbORC5 se apresentam em um complexo de alta massa molecular no parasita, o que estaria apontando para uma estrutura similar a ORC. Não se conhece ao certo se TbORC1B participa efetivamente do ORC. A

identificação dessas proteínas foi feita por RNA de interferência (RNAi), imunolocalização e co-expressão em *T. brucei* e purificação de complexo.¹⁶⁻¹⁷

A visão atual de ORC em *Trypanosoma brucei* está apresentada na Figura 3.

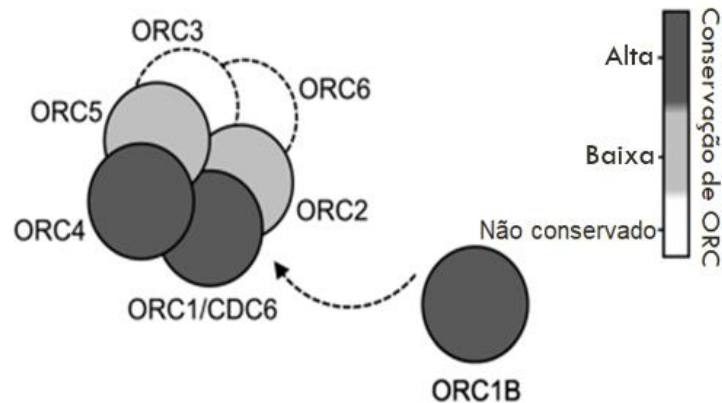


Figura 3 - Esquema representativo de ORC em *Trypanosoma brucei*, indicando o grau de conservação das subunidades.

Fonte: Adaptada de DAMASCENO *et al.*¹⁷

A partir da identificação dessas proteínas em *T. brucei*, foi possível identificar homólogas no genoma de *T. cruzi*.

O estudo de TcORC1/CDC6 é de suma importância por auxiliar no entendimento de como a replicação acontece no organismo. Essa via é essencial ao parasita e é um possível alvo para desenvolvimento de fármacos uma vez que a via seja conhecida plenamente. Em vista disso, o presente trabalho busca a caracterização biocomputacional de TcORC1/CDC6 através de ensaios comparativos de diferentes preditores e de estudos de filogenética e, em seguida, a expressão, purificação e caracterização de TcORC1/CDC6.

2 OBJETIVOS

2.1 OBJETIVOS GERAIS

Caracterizar TcORC1/CDC6 computacional e biofísicamente para melhor elucidar a maquinaria de replicação em Tripanosomatídeos.

2.2 OBJETIVOS ESPECÍFICOS

- Analisar a sequência de aminoácidos em busca de motivos conservados;
- Modelar a proteína usando como molde homóloga presente em bancos de dados na tentativa de prever prováveis comportamentos durante o enovelamento;
- Comparar a modelagem com o modelo gerado por preditores online (I-TASSER, Robbeta, AlphaFold2);
- Analisar filogeneticamente a proteína;
- Produzir a proteína recombinante em sistema bacteriano (*Escherichia coli*);
- Analisar massa molecular e raio hidrodinâmico da proteína por Cromatografia de Exclusão Molecular, Espalhamento Dinâmico da Luz e Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho;
- Verificar enovelamento da macromolécula por Dicroísmo Circular;
- Medir termoestabilidade da proteína por Fluorimetria de Varredura Diferencial.

3 MATERIAIS E MÉTODOS

3.1 CARACTERIZAÇÃO BIOCOMPUTACIONAL

3.1.1 Análise da sequência

Para o estudo, utilizou-se a proteína TcORC1/CDC6 (XP_808490.1). Inicialmente foram preditos parâmetros físicos e químicos, como a composição de aminoácidos, peso molecular, ponto isoelético teórico ($pI_{\text{teórico}}$), coeficiente de extinção molar, índice de estabilidade e hidropaticidade média (GRAVY) pelo uso do servidor ProtParam tool do ExPASy¹⁸ e raio hidrodinâmico teórico pela relação $Rh = 0,475n^{0,29}$ nm, onde n é o número de resíduos.¹⁹ Além disso, o servidor DiANNA 1.1 do Boston College²⁰⁻²¹ foi utilizado para analisar a predisposição de pares de cisteínas formarem cistinas.

Na busca de motivos proteicos e domínios conservados, foi utilizado respectivamente o PfamScan²² e *The Conserved Domain Database* CDD/SPARCLE.²³

3.1.2 Predição da estrutura

A predição estrutural foi realizada inicialmente com a modelagem de TcORC1/CDC6, usando DmORC1 (ORC1 de *Drosophila melanogaster*), PDB: 4XGC, como base por esta já possuir estrutura cristalográfica resolvida²⁴ e maior similaridade com a sequência do presente estudo. Utilizou-se o software Modeller²⁵ associado a um *script* em Python (Anexo A) para a modelagem. DmORC1 possui identidade de 31,51% e cobertura de 49%.

A fim de se obter o melhor modelo possível, a sequência da proteína foi submetida aos seguintes preditores online: I-TASSER,²⁶ Robetta²⁷ e AlphaFold2.²⁸

Para melhor analisar os modelos computacionais gerados, confeccionou-se o Diagrama de Ramachandran pelo servidor UCLA-DOE LAB - SAVES v6.0²⁹ para cada um deles.

Uma vez preditas as estruturas, gerou-se os *cartoons* representativos dos modelos pelo PyMOL *Molecular Graphic System*³⁰ e utilizou-se o servidor DALI³¹ para comparação de estrutura dos modelos gerados com proteínas da base de dados do PDB e do AlphaFold2 (AF-DB). Para analisar os resultados do DALI, determinou-se que moléculas relevantes deveriam ter um *Z-score* acima do limite definido

empiricamente por $n/10 - 4$, onde n é a quantidade de resíduos da sequência consultada, e identidade acima de 20%.³²

3.1.3 Filogenética

O estudo filogenético de TcORC1/CDC6 centrou-se em traçar indícios evolutivos em Kinetoplastida, selecionando-se sequências com identidade de pelo menos 55% na consulta pelo *Basic Local Alignment Search Tool* BLAST.³³ Em seguida, foram adicionados *Plasmodium vivax*, *Schistosoma bovis*, *Drosophila melanogaster* e *Homo sapiens* como o grupo externo.

O material selecionado passou por um alinhamento múltiplo das sequências de proteínas pelo ClustalW, seleção de blocos interessantes pelo programa 0,91b Gblocks³⁴⁻³⁵ e então usado para a construção de árvores filogenéticas no MEGA pelos métodos de Agrupamento de Vizinhos, Máxima Parcimônia e Máxima Verossimilhança.

A escolha dos métodos foi realizada com a intenção de se ter uma análise que fosse complementar em nível biológico e estatístico. Enquanto a Máxima Verossimilhança é um modelo explícito de evolução que realiza inferências probabilísticas para construção da árvore, sendo mais voltado para a estatística; a Máxima Parcimônia agrupa dados pela menor quantidade de eventos evolutivos que os separam, sendo mais voltado para biologia. O Agrupamento de Vizinhos foi usado como meio termo entre estatística e biologia por ser um agrupamento iterativo que define quais sequências compartilham um nó pai interno.

Para o alinhamento múltiplo pelo ClustalW, utilizou-se os parâmetros de penalidade para *gaps* como 10 por abertura, 0,2 por extensão, -1 por fechamento e 4 por separação. A matriz utilizada para o alinhamento foi a de Gonnet.

Para a configuração do gblocks, definiu-se o número mínimo de sequências para posição de flanqueamento como 85% do número de sequências, máxima posições contínuas não-conservadas como 8, tamanho mínimo de bloco como 10 e não permitir *gaps* no final de blocos.

Em todas as árvores, utilizou-se para o teste de filogenética o método *Bootstrap* com número de réplicas 1000 e calculou-se o melhor de modelo para a construção da árvore. O que teve melhor resultado foi com a substituição LG e distribuição Gamma.

Para Máxima Verossimilhança, o modelo de substituição foi o LG, determinou-se os padrões como uma distribuição *Gamma* com 5 categorias discretas, as opções de inferência de árvore foram obtidas pelo método heurístico de vizinho mais próximo.

Para Agrupamento de Vizinhos, o modelo de substituição foi o de Poisson, os padrões foram definidos como uma distribuição *Gamma* com 5 parâmetros acompanhado do tratamento de *gaps*/dados faltantes como exclusão por pares.

Para Máxima Parcimônia, as opções de inferência de árvore foram: o método de busca foi de poda e re-enxerto, com o número inicial de árvores como 10, o nível de busca como 1 e o máximo número de árvores a reter como 100.

Com base na árvore construída pelo método de Máxima Verossimilhança, determinou-se alguns nós de interesse, a sequência ancestral foi predita,³⁶ e ela teve sua estrutura predita pelo AlphaFold2 para fins comparativos.

3.2 PRODUÇÃO DA PROTEÍNA RECOMBINANTE

3.2.1 Expressão

A construção utilizada no estudo foi pMAL+TcORC1/CDC6 em *Escherichia coli* Rosetta DE3 (Novagen). Esta foi cedida pelo Laboratório de Ciclo Celular do Instituto Butantan.

Para iniciar a expressão da proteína, inoculou-se a bactéria ainda congelada do meio de crio-conservação estoque em tubos de ensaio com 5 mL de meio LB (10 gL⁻¹ de triptona, 5 gL⁻¹ de extrato de levedura, 5 gL⁻¹ de NaCl, pH 7) com o acréscimo dos antibióticos para se atingir as concentrações de 50 µg/mL de ampicilina e 34 µg/mL de cloranfenicol. A partir daí, incubou-se a cultura a 37°C sob agitação 100 rpm durante 16 h.

Passado esse período, foi feita a diluição 1:100 em meio LB novo com os devidos antibióticos em volume de 500 mL em erlenmeyers de 2 L, e foi retomada a incubação a 37°C sob agitação a 150 rpm. Monitorou-se o crescimento da cultura por espalhamento de luz em espectrofotômetro DO_{600nm} até ser alcançado o valor 0,6, ponto no qual foi introduzido o isopropil β-D-1-tiogalactopiranosida (IPTG) para uma concentração de 1 mM a fim de induzir a expressão heteróloga da proteína.

Após 3 h de indução, centrifugou-se a cultura a 4.658 g a 4 °C por 45 min. Lavou-se o precipitado de células com PBS 1x, transferiu-se o ressuspendido para tubo cônico de 50 mL por praticidade de armazenamento e foi feita nova centrifugação

a 4.000 g a 4 °C por 60 min. O sobrenadante foi descartado, e o sedimento foi congelado a -20 °C.

3.2.2 Lise

O sedimento foi ressuspensionado em 50 mL de Tampão A (50 mM de Tris-HCl pH 7,4 e 300 mM de NaCl) com 1 mM de DTT, 1 mM de PMSF, 1 µg/mL de Leupeptina e 1 µg/mL de Pepstatina.

Em seguida, a lise das células foi iniciada com sonificador Branson Sonic Power Co utilizando-se 14 ciclos, com 60 s de pulso por 30 s de descanso cada e amplitude de 40 %. Centrifugou-se o extrato obtido a 20.000 g por 45 min a 4 °C. No final, o sobrenadante foi recuperado para uso nos passos de purificação.

O sedimento foi ressuspensionado em 50 mL de tampão A. Alíquotas das frações solúvel e insolúvel foram coletadas, adicionou-se tampão de amostra 4x e, após desnaturação das proteínas, elas foram analisadas por SDS-PAGE, sendo o gel corado com Azul de Comassie.

3.2.3 Purificação

3.2.3.1 Cromatografia de afinidade

Na construção pMAL+TcORC1/CDC6, a proteína é expressa fusionada a uma extensão de Proteína de Ligação à Maltose (MBP). Esta extensão interage com dextrina principalmente por ligações de hidrogênio, e essa interação pode ser usada para a purificação da proteína. Para liberá-la, é feita a eluição com um tampão rico em maltose, composto que compete pela ligação à matriz.

Inicialmente, a coluna MBPTrapTM HP 5 mL (GE Healthcare) foi equilibrada com 10 volumes de Tampão A, 5 volumes de Tampão A com 10 mM de maltose e 3 volumes de Tampão A novamente.

Em seguida, o sobrenadante da lise foi filtrado a 0,45 µm e aplicado à coluna com o auxílio de um cromatógrafo ÄKTA Purifier 10 (Amersham Pharmacia Bioscience). Esse equipamento permitiu o acompanhamento de todos os passos de purificação pela absorbância no comprimento de onda 280 nm, correspondente ao sinal dos elétrons ressonantes nos anéis dos resíduos de fenilalanina, tirosina e triptofano.

Após aplicado o sobrenadante, injetou-se Tampão A até o sinal a 280 nm voltar ao inicial no intuito de retirar proteínas que não interagem com a matriz. Depois foi feita a eluição com 1 mM de maltose em Tampão A e, para finalizar, injetou-se tampão A com 10 mM de maltose para retirar qualquer resquício de proteína que pudesse ter continuado ligado à matriz. Alíquotas foram coletadas de cada etapa para verificação em SDS-PAGE.

As frações referentes à eluição foram coletadas e dialisadas com Tampão A por 18 h a 4 °C sob leve agitação para a retirada da maltose da amostra.

3.2.3.2 Cromatografia de exclusão molecular

A Cromatografia de Exclusão Molecular (SEC) tem como finalidade a separação de moléculas pelas diferenças de massa em solução. Para tal, a coluna de SEC é composta por um polímero com poros de diversos tamanhos. Desse modo, quanto maior a molécula, menos ela é capaz de entrar nos poros e mais rapidamente ela é eluída; enquanto, quanto menor a molécula, mais ela é capaz de entrar e migrar dentro dos poros, percorrendo um maior caminho e, conseqüentemente, mais tardiamente ela é eluída.

Utilizou-se uma Superdex™ 200 16/60 (GE) preparativa e não-analítica de 120 mL para a SEC vinculada ao cromatógrafo ÄKTA.

Inicialmente, passou-se 180 mL de água pela coluna e depois equilibrou-se com 180 mL de Tampão A. Daí injetou-se a amostra devidamente concentrada pelo sistema de ultrafiltração Amicon com corte de 30 kDa e acompanhou-se mais 180 mL de tampão A. As frações dos picos apresentados foram coletadas para posterior análise em SDS-PAGE.

3.2.3.3 *Western-blot*

Com as amostras das Cromatografias de Afinidade e de Exclusão Molecular, foi realizado um *western-blot* com anticorpo específico a TcORC1/CDC6. Para isso, foi realizada a separação das proteínas por massa por SDS-PAGE, o conteúdo do gel foi transferido para uma membrana de nitrocelulose, e esta foi revelada com Ponceau para confirmar se a transferência havia sido bem-sucedida.

Bloqueou-se a membrana com leite desnatado 5% em PBST 1x (Tampão Fosfato Salino com Tween 1%) por 2 h a temperatura ambiente sob agitação. Foi

realizada uma série de três lavagens de 5 min sob agitação com PBST 1x. Incubou-se o anticorpo primário (contra TcORC1/CDC6) em leite 2 % em PBST 1x por 16 h na proporção 1:150 sob agitação a 4 °C. Em seguida, fez-se nova série de três lavagens de 5 min sob agitação com PBST 1x. Incubou-se com anticorpo secundário (*anti-rabbit*) na proporção 1:5.000 em leite 2% em PBST 1x por 2 h. Foi realizada uma série de três lavagens de 5 min sob agitação com PBST 1x. E, por fim, revelou-se a membrana com solução reveladora (1,1% luminol, 0,48% ácido P-cumárico, 11,1% Tris 1 M pH 8,5), adição de peróxido de hidrogênio na 1:300 (v/v) e com o auxílio do scanner *C-Digit Blot* (LI-COR Biosciences UK Ltd) e adquiriu-se as imagens pelo software *Image Studio*TM (LI-COR Biosciences UK Ltd).

3.3 CARACTERIZAÇÃO BIOFÍSICA

3.3.1 Curva de calibração da Exclusão Molecular

Como apontado anteriormente, SEC baseia sua separação pelo tamanho. Portanto, dada uma proteína qualquer, esta terá um volume de eluição característico proporcional à sua massa e ao seu raio hidrodinâmico. Para se calcular esses parâmetros com maior precisão, é utilizada uma coluna de SEC especial chamada de analítica. Através dela, é estimado o volume de eluição para diferentes amostras de tamanho conhecido e é montada a curva de calibração em função da razão de eluição k_{av} , que é definida por:

$$k_{av} = \frac{V_e - V_0}{V_T - V_0}$$

Sendo V_e o volume de eluição da proteína de interesse, V_0 o volume morto da coluna e V_T o volume total da coluna.

Construindo o gráfico de $k_{av} \times \ln(MW)$, estima-se a massa da proteína de interesse. E, pelo gráfico de $\sqrt{-\log(k_{av})} \times R_H$, estima-se o raio hidrodinâmico da molécula.

Para a curva de calibração, utilizou-se a coluna SuperdexTM 200 10/300 (GE) analítica de 24 mL com protocolo semelhante ao apresentado em 3.2.3.2, substituindo os volumes de 180 mL por 36 mL. Foi feita a SEC de TcORC1/CDC6 e realizada a calibração da coluna com os padrões do laboratório (ribonuclease, anidrase carbônica bovina, ovalbumina, conalbumina, β -amilase de batata doce, ferritina, tiroglobulina e blue dextran).

3.3.2 Espalhamento Dinâmico da Luz (DLS)

No Espalhamento Dinâmico da Luz, é analisada a dispersão da luz causada por partículas na fase líquida em movimento Browniano. Esse espalhamento está relacionado ao Raio Hidrodinâmico da molécula em solução e à sua Massa Molar. Ainda é possível, estimar a homogeneidade de uma amostra pela técnica, já que espécies diferentes dispersam a luz de forma distinta. Pode-se observar esta homogeneidade ao se observar a porcentagem de polidispersividade.

Para a leitura, utilizou-se a proteína a 1 mg/mL em 50 mM de Tris-HCl pH 7,4. Esta foi primeiramente centrifugada a 10.000g por 10 min e aplicada em cubeta de quartzo. A leitura foi realizada em Zetasizer μ V (Malvern).

3.3.3 Fluorimetria de Varredura Diferencial (DSF)

Na Fluorimetria de Varredura Diferencial (DSF), é realizada a desnaturação térmica controlada da proteína para analisar sua termoestabilidade. Para isso, a técnica faz uso de uma molécula repórter, um fluoróforo, que apresenta fluorescência baixa em ambientes polares e alta fluorescência em ambientes apolares.

O sinal de emissão obtido forma uma sigmoide, cujo ponto de inflexão significa a temperatura de *melting* (T_m), que corresponde à temperatura na qual 50% das proteínas estão desenoveladas. Para a determinação desse ponto, basta calcular a primeira derivada da curva e verificar o ponto de inflexão.

3.3.3.1 Teste de concentração

É necessário primeiramente estimar a proporção ideal de proteína/fluoróforo. Para isso, montou-se uma placa conforme esquematizado na Figura 4. Escolheu-se o fluoróforo Sypro™ Orange (Fermentas), cujo máximo de excitação se encontra em 488 nm e máximo de emissão em 570 nm, sendo vendido em uma concentração indicada de 5.000 X. A solução proteína e fluorófora foi realizada em 50 mM de Tris-HCl pH 7,4.

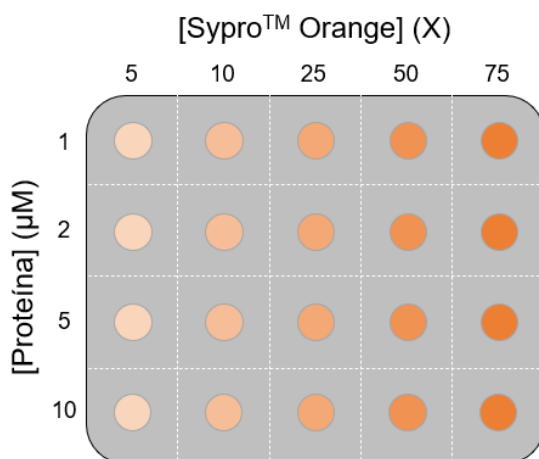


Figura 4 - Representação da montagem da placa para o teste de concentração. Os espaços vazios representam cada poço.

Fonte: Elaborada pelo autor

Após a montagem, submeteu-se a placa a uma centrifugação 1.000 g, por 3 min a 4 °C. E, por fim, utilizou-se o aparelho de PCR em tempo real CFX96 (Bio-rad) para a leitura varrendo as temperaturas de 5 a 100 °C em passos de 1 °C por minuto.

3.3.3.2 *Screen* de tampão

Uma placa de 96 poços com a quantidade ideal de proteína/Sypro Orange por poço (conforme obtido em 3.3.3.1) foi montada, e utilizou-se o kit Stability & Solubility Screen™ (Hampton Research) na busca da melhor condição de agente tamponante e sal para a proteína com base na solução de proteína e fluoróforo inicial em 50 mM de Tris-HCl pH 7,4.

Finalizada a montagem, o protocolo se manteve como explicitado em 3.3.3.1. Para facilitar a análise, foi calculado o T_m de cada condição e subtraído o valor de T_m obtido em 3.3.3.1. A partir disso, montou-se um histograma no qual valores negativos são de menor T_m e valores positivos são de maior T_m .

3.3.4 Dicroísmo Circular (CD)

A espectroscopia de Dicroísmo Circular (CD) para se verificar o enovelamento da proteína. Os espectros de CD foram obtidos utilizando-se um espectropolarímetro JASCO J-815 (JASCO), equipado com um sistema de controle de temperatura do tipo PELTIER PTC 423S/15 no Laboratório de Biofísica Molecular do IFSC-USP.

Os espectros foram coletados na região do UV distante, no intervalo de comprimento de onda de 270 a 205 nm, com resolução de 0,1 nm e taxa de 100 nm

min⁻¹, com resposta de dois segundos, utilizando-se uma cubeta de quartzo de caminho óptico de 0,5 cm para amostras com concentração de 0,8 mg mL⁻¹ em 50 mM de Tris-HCl pH 7,4. O programa Spectra Manager (JASCO) foi usado para registro dos dados.

Ademais, foram feitas leituras de 20 a 90 °C, com passos de 2 °C, e a variação do sinal a 222 nm foi acompanhado para se traçar uma curva de desnaturação da proteína e estimar sua termoestabilidade.

3.3.5 Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho (SEC-MAL)

O Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho (SEC-MALS) é uma combinação de Cromatografia de Exclusão Molecular com análises de espalhamento de luz em multi-ângulo. No espalhamento, um feixe de luz colimado e polarizado atravessa uma solução com macromoléculas. Com isso, é possível medir a intensidade e a dependência angular. A primeira está relacionada à massa molar da partícula. A segunda, ao tamanho.

Para essas análises, utilizou-se a proteína a 2 mg mL⁻¹ em 50 mM de Tris-HCl pH 7,4 e uma Superdex™ 200 10/300 (GE) analítica de 24 mL.

3.3.6 Espectrometria de Massas

Para o preparo das amostras, foi realizado SDS-PAGE da amostra visualmente mais pura de TcORC1/CDC6, e o gel resultante foi corado com Comassie. A banda a ser analisada foi recortada do gel e fragmentada. Em seguida, foram realizadas etapas de descoloração com solução de 50% (v/v) metanol e 2,5 (v/v) de ácido acético até a retirada total da coloração azul. Passou-se para a desidratação com acetonitrila (100%) por 5 min. Descartou-se a acetonitrila e repetiu-se o passo de acetonitrila. O remanescente foi evaporado em *speed-vac* por 5 min a 45 °C.

Uma vez com a amostra de gel descorada e seca, foi realizada a redução com DTT 10 mM por 30 min e a fase líquida foi removida ao final. A alquilação dos fragmentos foi realizada por 30 min com solução de iodoacetamida 50 mM e bicarbonato de amônio 100 mM. A fase sólida foi lavada com solução de 100 mM de bicarbonato de amônio por 10 min.

Os fragmentos da etapa anterior foram desidratados com acetonitrila (100%) por 5 min e reidratados com solução de bicarbonato de amônio 100 mM por 10 min. A etapa de desidratação foi repetida por mais duas vezes, e o líquido remanescente foi evaporado em *speed-vac* por 5 min a 45 °C.

Para a digestão das bandas, foi adicionado 50 µL de solução 20 ng/µL de tripsina e bicarbonato de amônio 50 mM aos fragmentos de gel, em banho de gelo, por 30 min, e o sobrenadante foi retirado ao final. Adicionou-se a solução de bicarbonato de amônio 50 mM e incubou-se por 16 h a 37 °C. Terminada as 16 h, a extração foi iniciada com solução de ácido fórmico 5% em água por 10 min. O sobrenadante foi coletado, e um segundo passo de extração com solução de ácido fórmico 5% em acetonitrila 50% (v/v). O sobrenadante foi coletado no mesmo tubo da etapa anterior.

Após isso, a amostra foi analisada no Espectrofotômetro micrOTOF-QII (Bruker Daltonics) com corrida associada ao cromatógrafo Dionex Ultimate 3000, RSLCbaby system (Thermo Scientific) com a coluna Acclaim PepMap® RSCL, 75 µm x 150 nm, C18, 2µm, 100 Å.

4 RESULTADOS E DISCUSSÃO

4.1 CARACTERIZAÇÃO BIOCOMPUTACIONAL

4.1.1 Análise da sequência

TcORC1/CDC6 é uma proteína composta por 450 aminoácidos, e a distribuição desses aminoácidos está apresentada na Tabela 1. Ela possui massa molecular de 50,15 kDa e $pI_{teórico}$ de 8,89. O raio hidrodinâmico teórico é de 2,79 nm.

Observando a composição dos aminoácidos, vale ressaltar a presença de Fenilalanina, Triptofano e Tirosina, o que facilitará possíveis análises a 280 nm da molécula, como a estimativa de concentração da proteína em solução por espectrofotometria, pela característica da cadeia lateral desses aminoácidos ter na composição um anel aromático, que possui elétrons ressonantes emitem nesse comprimento de onda.

Também é importante visualizar que não há um número alto de cisteínas, contabilizando em toda a sequência dez desses aminoácidos. Ao se fazer a predição de possíveis ligações dissulfeto pelo servidor DiANNA 1.1 (Boston College), foi encontrado apenas um par de cisteínas com *score* maior que 0,13. Este possuía *score* de ~0,63. Quanto mais próximo de 1 é esse *score*, maior é a probabilidade que o par de cisteínas formem uma ligação. Esses dados indicam que ligações dissulfeto não são muito favorecidas na sequência, o que pode ser indício positivo a uma possível solubilidade da proteína em

Tabela 1 - Composição de TcORC1/CDC6 em aminoácidos (AA) em porcentagem.

AA	%	AA	%
Ala (A)	9,1	Lys (K)	5,8
Arg (R)	6,2	Met (M)	2,0
Asn (N)	3,6	Phe (F)	4,9
Asp (D)	4,7	Ser (S)	6,7
Cys (C)	2,2	Pro (P)	4,0
Gln (Q)	3,3	Thr (T)	5,8
Glu (E)	5,6	Trp (W)	0,7
Gly (G)	4,9	Tyr (Y)	2,2
His (H)	2,7	Val (V)	8,4
Ile (I)	6,0	Pyl (O)	0,0
Leu (L)	11,3	Sec (U)	0,0

Fonte: Elaborada pelo autor.

O coeficiente de extinção molar calculado foi de $32.025 \text{ M}^{-1}\text{cm}^{-1}$ no caso de todos os pares de cisteína formarem cistina e de $31.400 \text{ M}^{-1}\text{cm}^{-1}$ no caso de todas as cisteínas estarem reduzidas. O coeficiente de estabilidade da proteína calculado foi de 35,61, o que classifica a proteína como estável. E o GRAVY foi estimado em 0,048. Isso indica que a proteína é levemente mais hidrofóbica, algo que pode impactar a solubilidade.

Em relação a motivos proteicos, em TcORC1/CDC6 só foi encontrado o motivo AAA. Ele é definido como um *P-loop*_NTPase (Figura 5), domínio prevalente em distintos enovelamentos de ligação a nucleotídeos.²² O motivo AAA é característico de uma série de ATPases com as mais diversas funções e é comum em proteínas iniciadoras de replicação de DNA.^{16,37} A interação de ORC1 com ssDNA é dependente de ATP.³⁴ Dessa forma, a presença de um motivo de ATPases está de acordo ao esperado.

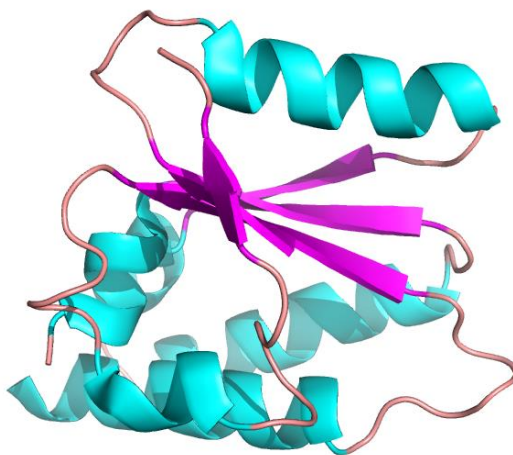


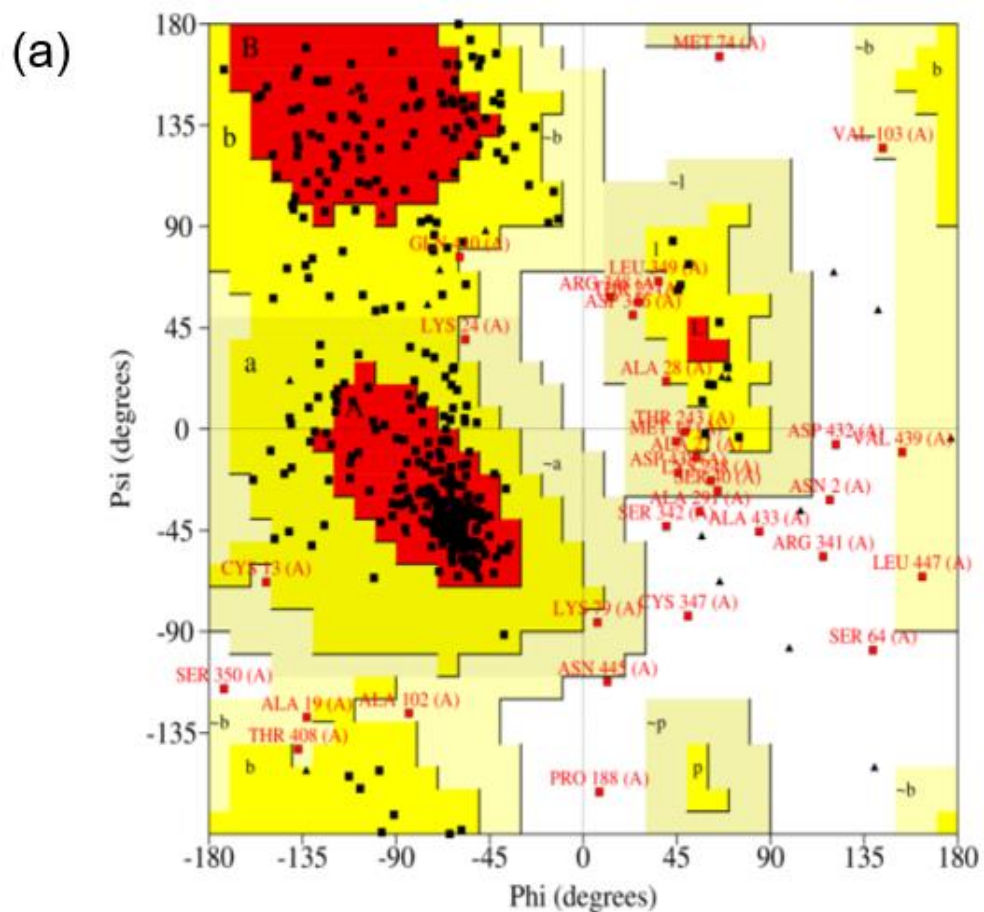
Figura 5 - Motivo AAA extraído do PDB 1IY2.

Fonte: Elaborada pelo autor.

Ao se estudar famílias de proteínas, TcORC1/CDC6 foi apontada a pertencer à super família PTZ00112, com código de acesso cl36513. A descrição dessa família é de ser de proteína do tipo ORC1. Pela análise, esse domínio encontra-se entre os aminoácidos 45-269.

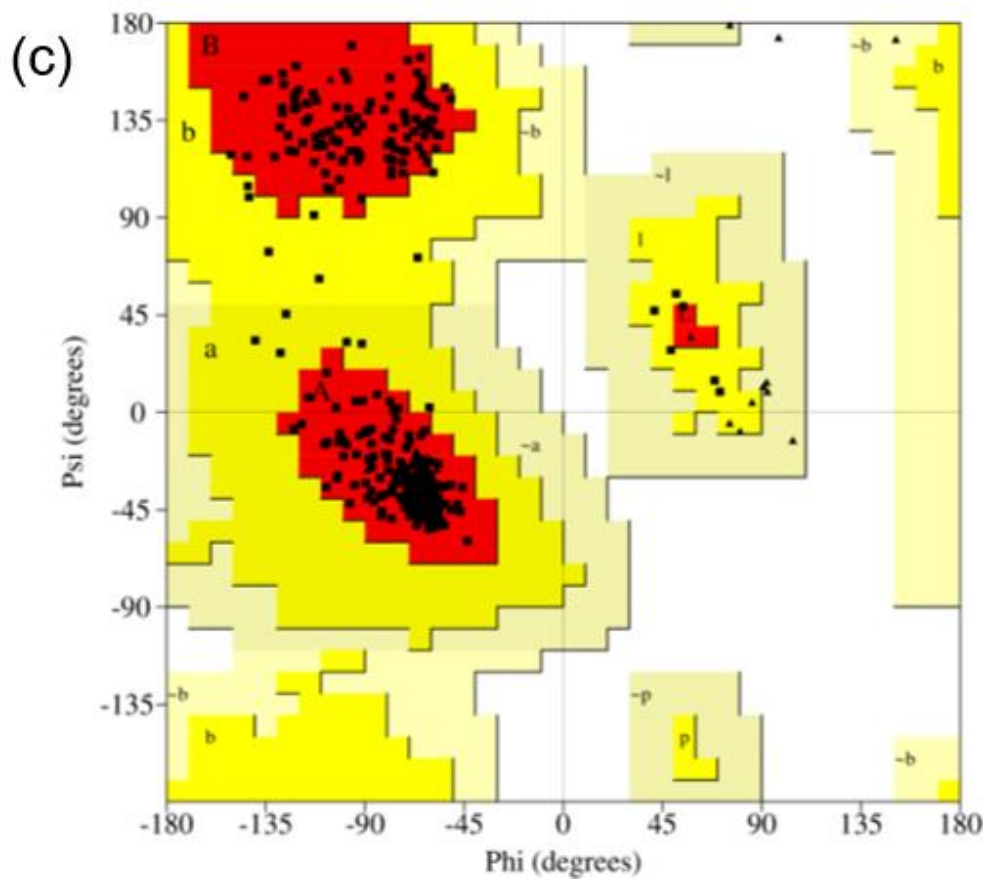
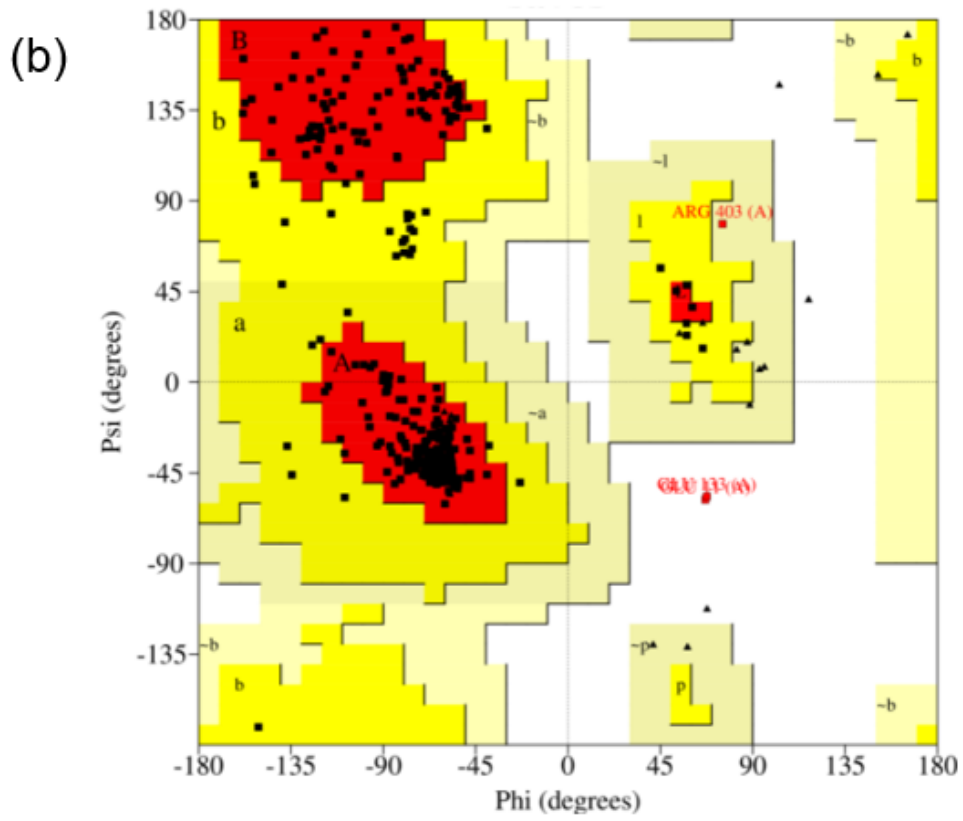
4.1.2 Predição da estrutura

Para auxiliar na análise das predições de estruturas proteicas, foram construídos os Diagramas de Ramachandran de cada modelo (Figura 6) e a Tabela 2. Esta mostra a disposição dos resíduos no Diagrama de Ramachandran. Nos diagramas, há a definição de 4 regiões: as mais favoráveis, que é representada pela cor vermelha e indicada por A, B e L; as adicionais mais permissivas, que é representada pela cor amarela mais escura e indicada por a, b, l e p; as permissivas, que é representada pela cor amarela mais clara e indicada por ~a, ~b, ~l e ~p; e as não permissivas, representada pela cor branca.



(continua)

(continuação)



(continua)

(continuação)

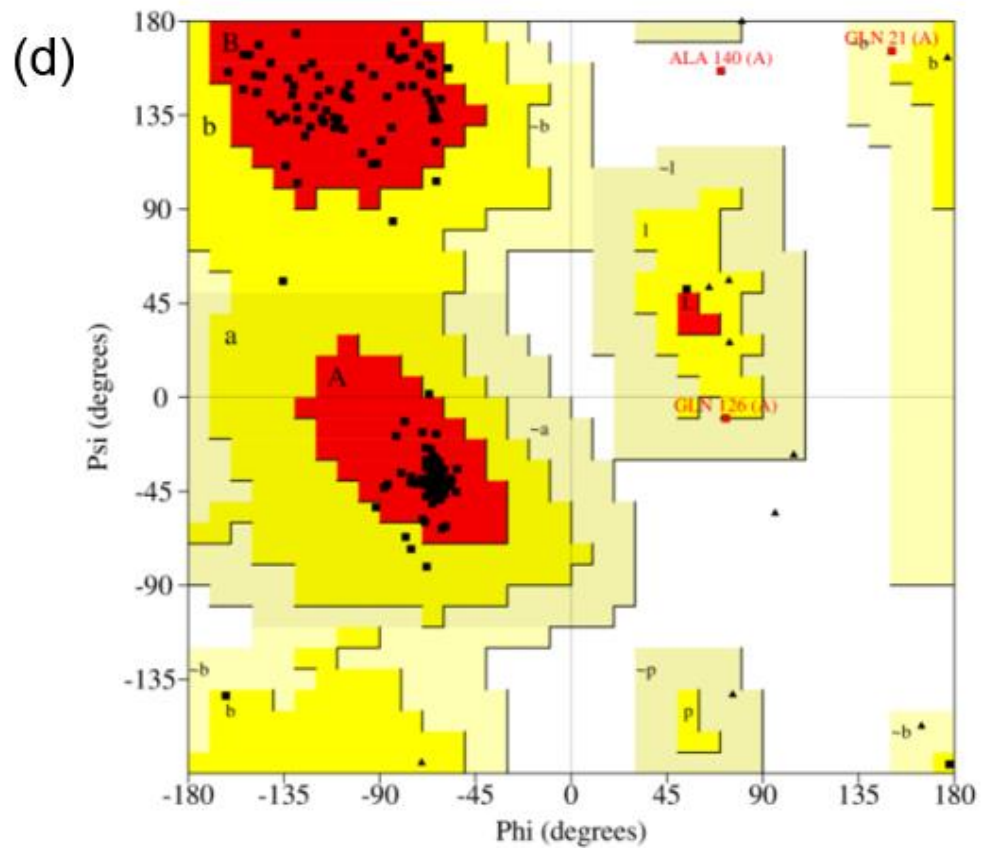


Figura 6 - Diagramas de Ramachandran para diferentes modelagens computacionais. (a) I-TASSER. (b) Robbeta. (c) AlphaFold2. (d) Modelagem.

Fonte: Elaborada pelo autor.

Tabela 2 - Análise do Diagrama de Ramachandran construído a partir de cada preditor evidenciando a quantidade (Qtd) de resíduos e percentual (%).

	I-TASSER		Robbeta		AlphaFold2		Homologia	
	Qtd	%	Qtd	%	Qtd	%	Qtd	%
Resíduos em regiões mais favoráveis (A, B, L)	278	68,1	373	91,4	389	95,3	292	91,8
Resíduos em regiões adicionais mais permissivas (a, b, l, p)	98	24,0	32	7,8	19	4,7	20	6,3
Resíduos em regiões permissivas (~a, ~b, ~l, ~p)	21	5,1	1	0,2	0	0,0	4	1,3
Resíduos em regiões não permissivas	11	2,7	2	0,5	0	0,0	2	0,6
Número de resíduos que não são glicinas nem prolinas	408	100,0	408	100,0	408	100,0	318	100,0
Números de resíduos finais (excluindo glicina e prolina)	2	—	2	—	2	—	1	—
Número de glicinas	22	—	22	—	22	—	20	—
Número de prolinas	18	—	18	—	18	—	12	—
Número total de resíduos	450	—	450	—	450	—	351	—

Fonte: Elaborada pelo autor.

Com base na Figura 6 e Tabela 2, observa-se que o melhor modelo obtido foi através AlphaFold2 (Figura 6c) considerando como métrica de seleção possuir mais aminoácidos em regiões com maior permissividade. Ele é o preditor mais recente analisado e utiliza um modelo baseado em rede neural para estimar com precisão atômica até mesmo proteínas que não possuem estrutura homóloga resolvida²⁸. No modelo do AlphaFold2, 100% dos resíduos estão nas regiões mais favoráveis ou nas regiões adicionais mais permissivas, possuindo a maior porcentagem entre os 4 modelos em regiões mais favoráveis – 95,3% dos resíduos.

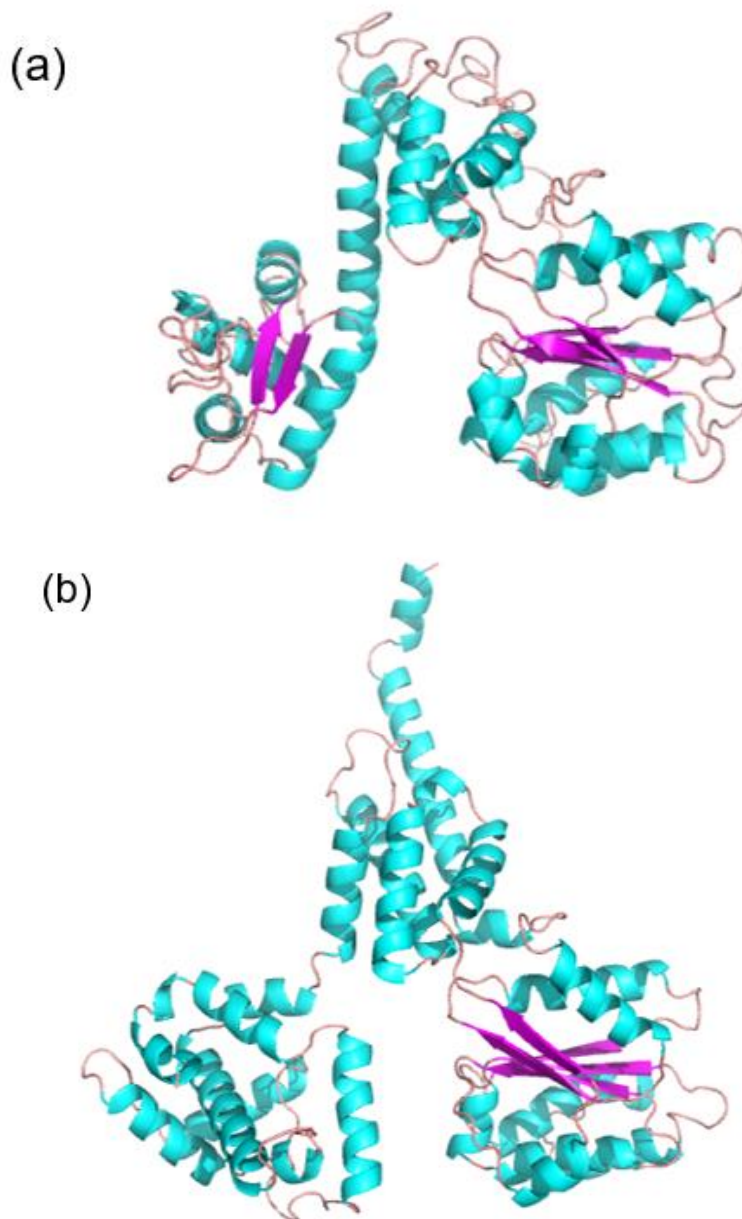
Pela mesma métrica de seleção de modelo pela quantidade de aminoácidos com regiões com maior permissividade, o gerado pelo Robetta foi o segundo melhor (Figura 6b). Esse preditor transforma a sequência consultada em domains putativos e gera modelos estruturais por modelagem comparativa caso encontre um *match* confiável. Caso contrário, utiliza o método de predição *de novo*. Este gera a predição através de princípios e energética gerais de enovelamento proteico obtido de estruturas nativas²⁷. No modelo do Robetta, 99,2% dos resíduos estão nas regiões mais favoráveis ou nas regiões adicionais mais permissivas e consta 0,5% dos resíduos em regiões não permissivas.

Em terceiro, tem-se o modelo gerado por homologia a DmORC1 (Figura 6d). Vale lembrar que as sequências possuem 31% de identidade, que esse método não cobriu a sequência completa de TcORC1/CDC6 e que modelagem por homologia tem sua qualidade intrinsecamente vinculada ao grau de similaridade das estruturas. No modelo gerado por modelagem por homologia, 99,2% dos resíduos estão nas regiões mais favoráveis ou nas regiões adicionais mais permissivas e há 0,6% dos resíduos em regiões não permissivas.

O modelo que mais apresentou aminoácidos em regiões menos permissivas foi o gerado pelo I-TASSER (Figura 6a). Este utiliza uma abordagem de modelagem proteína hierárquica baseada no alinhamento de Perfil-Perfil aprimorado de estrutura secundária (PPA) e a implementação iterativa do programa *Threading ASSEmbly Refinement* (TASSER), tendo como primeiro passo a busca em uma biblioteca estrutural representativa do PDB por sequências com identidade acima de 70%^{26,39,40}. A não existência de sequências com esse grau de similaridade depositadas deve justificar a qualidade do modelo. Este apresenta 92,1% dos resíduos em regiões mais favoráveis ou nas regiões adicionais mais permissivas, mas apenas 68,1% estão nas

regiões mais favoráveis, e há 2,7% dos resíduos em regiões não permissivas. Este último valor é 4,5 vezes maior do apresentado no modelo gerado por homologia.

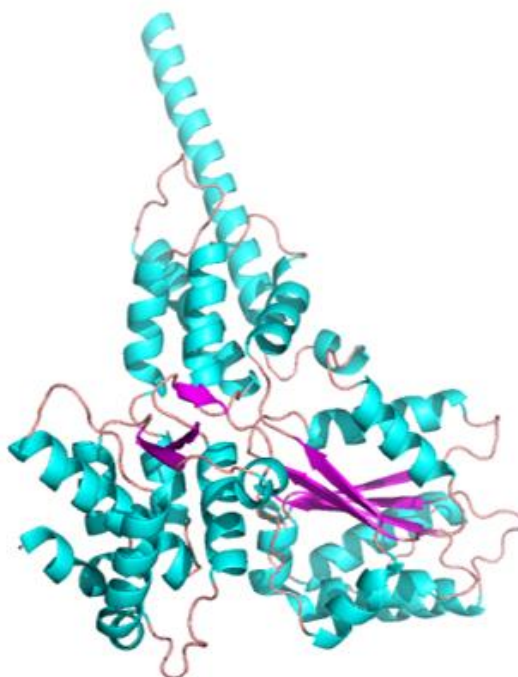
Para se obter uma percepção mais visual dos modelos gerados, construiu-se as representações em *cartoons* para cada preditor separadamente e depois sobrepondo contra o melhor modelo analisado (AlphaFold2) tanto considerando o modelo inteiro quanto considerando apenas a região do motivo AAA, conforme apontando em 4.1.1. Isso pode ser visto nas Figuras 7 a 10.



(continua)

(continuação)

(c)



(d)

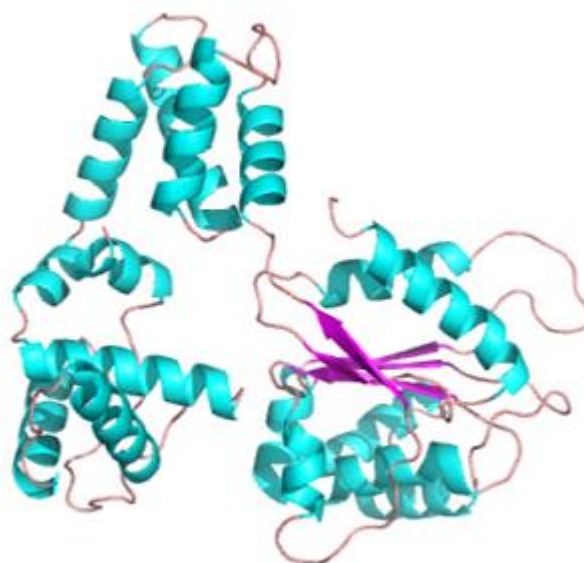
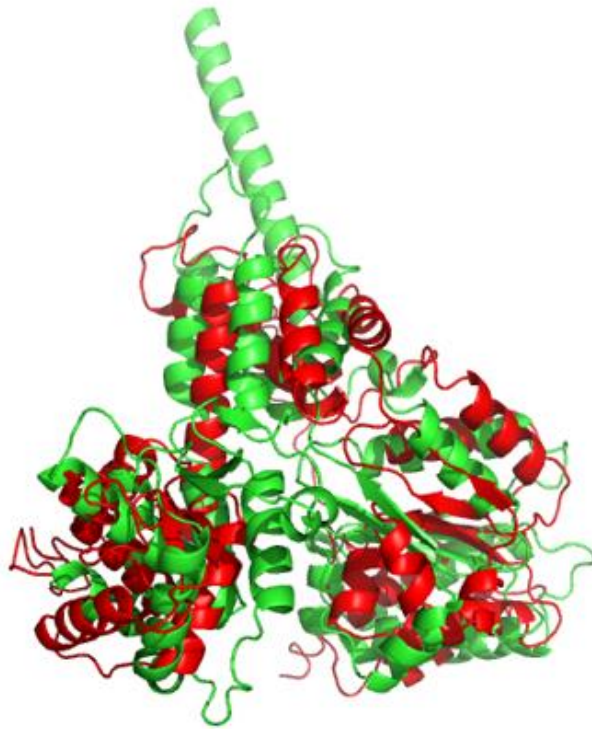


Figura 7 - Representação em *cartoon* dos modelos gerados pelos diferentes métodos de predição de estrutura proteica: (a) I-TASSER, (b) Robbeta, (c) AlphaFold2 e (d) homologia.

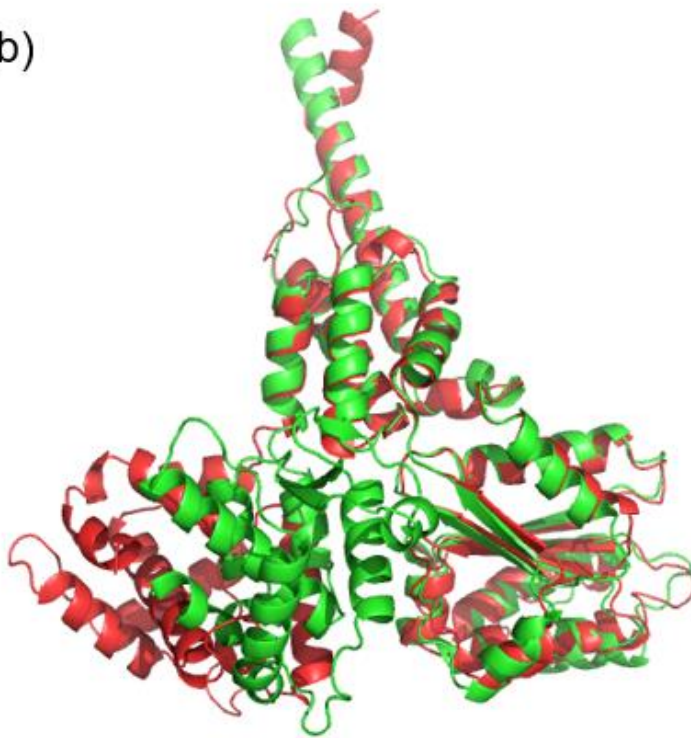
Fonte: Elaborada pelo autor.

Pela figura 7, é notória uma maior similaridade entre os modelos gerados pelo I-TASSER e modelagem por homologia e os gerados pelo AlphaFold2 e Robbeta. Ademais, há uma clara majoritariedade de hélices α nos modelos em comparação a fitas β .

(a)



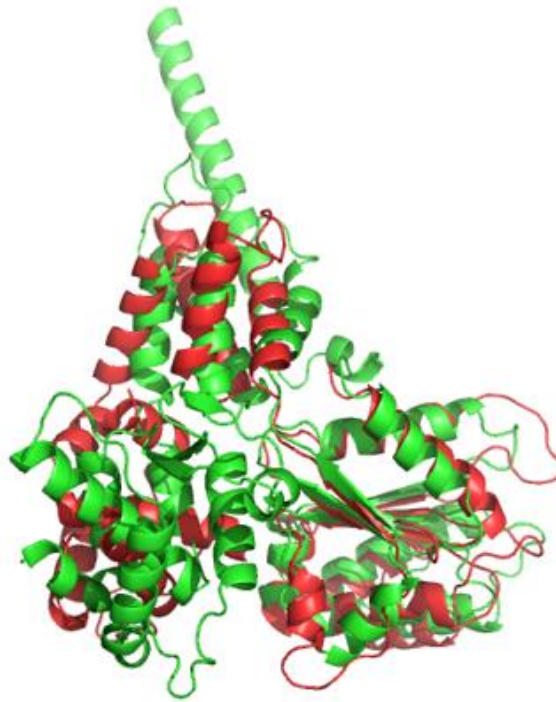
(b)



(continua)

(continuação)

(c)



(d)

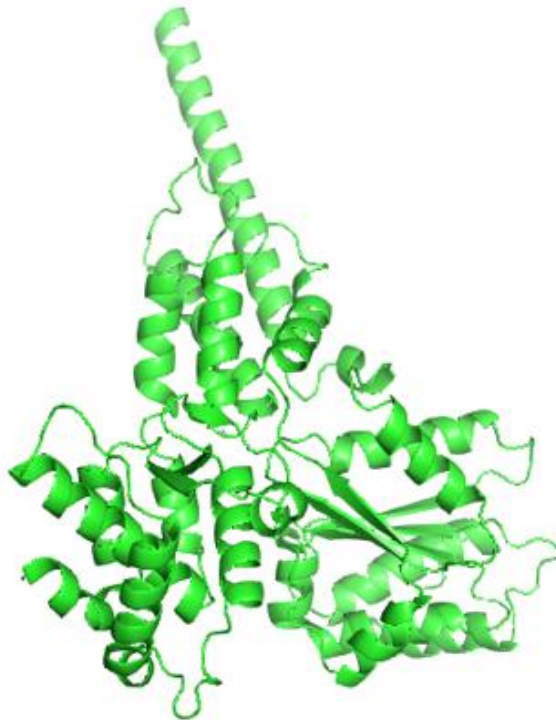
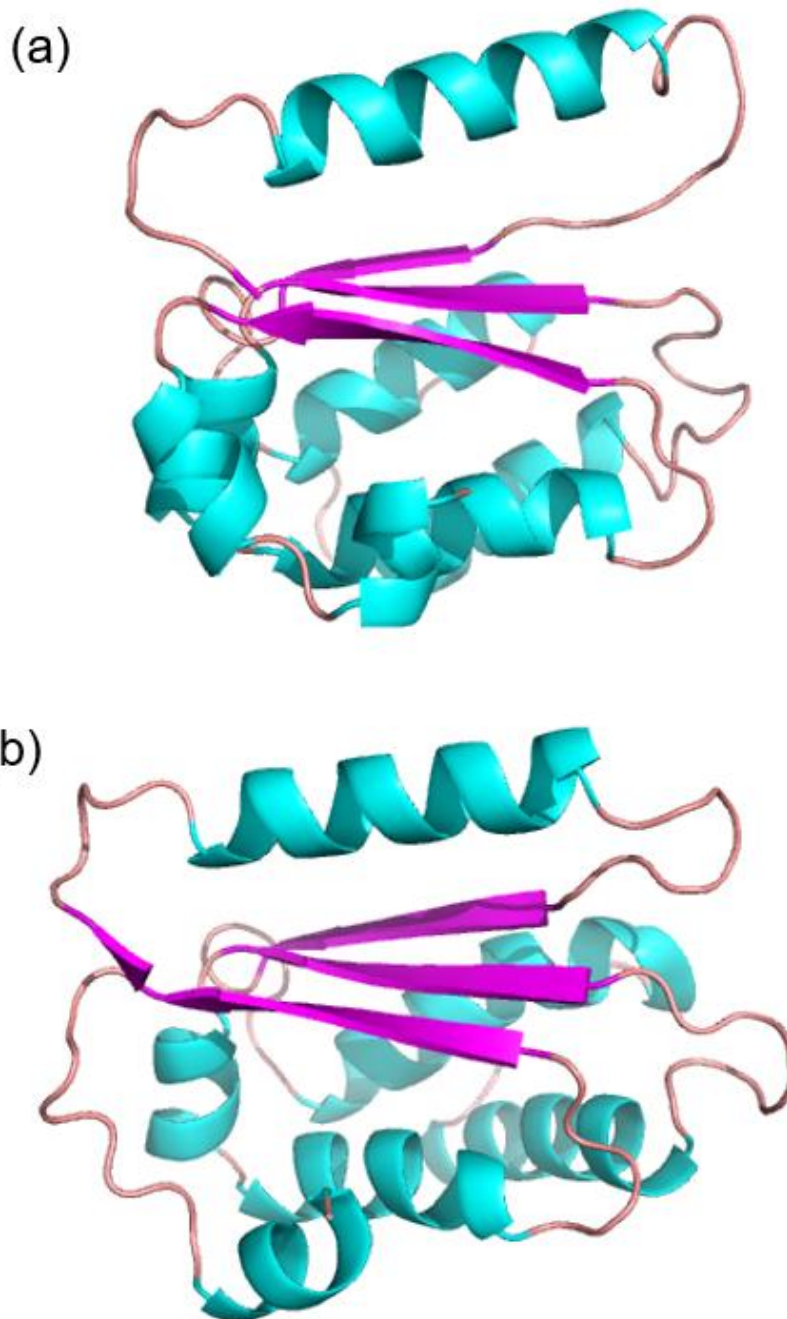


Figura 8 - Sobreposição da representação em *cartoon* obtida pelo AlphaFold2 contra (a) I-TASSER, (b) Robbeta e (c) homologia. (d) apresenta a estrutura do AlphaFold2 isolada.

Fonte: Elaborada pelo autor.

Pela figura 8, observa-se que a estrutura completa pelos diferentes métodos possui melhor sobreposição na região que corresponde ao motivo AAA. Isso é evidenciado nas Figuras 9 e 10.



(continua)

(continuação)

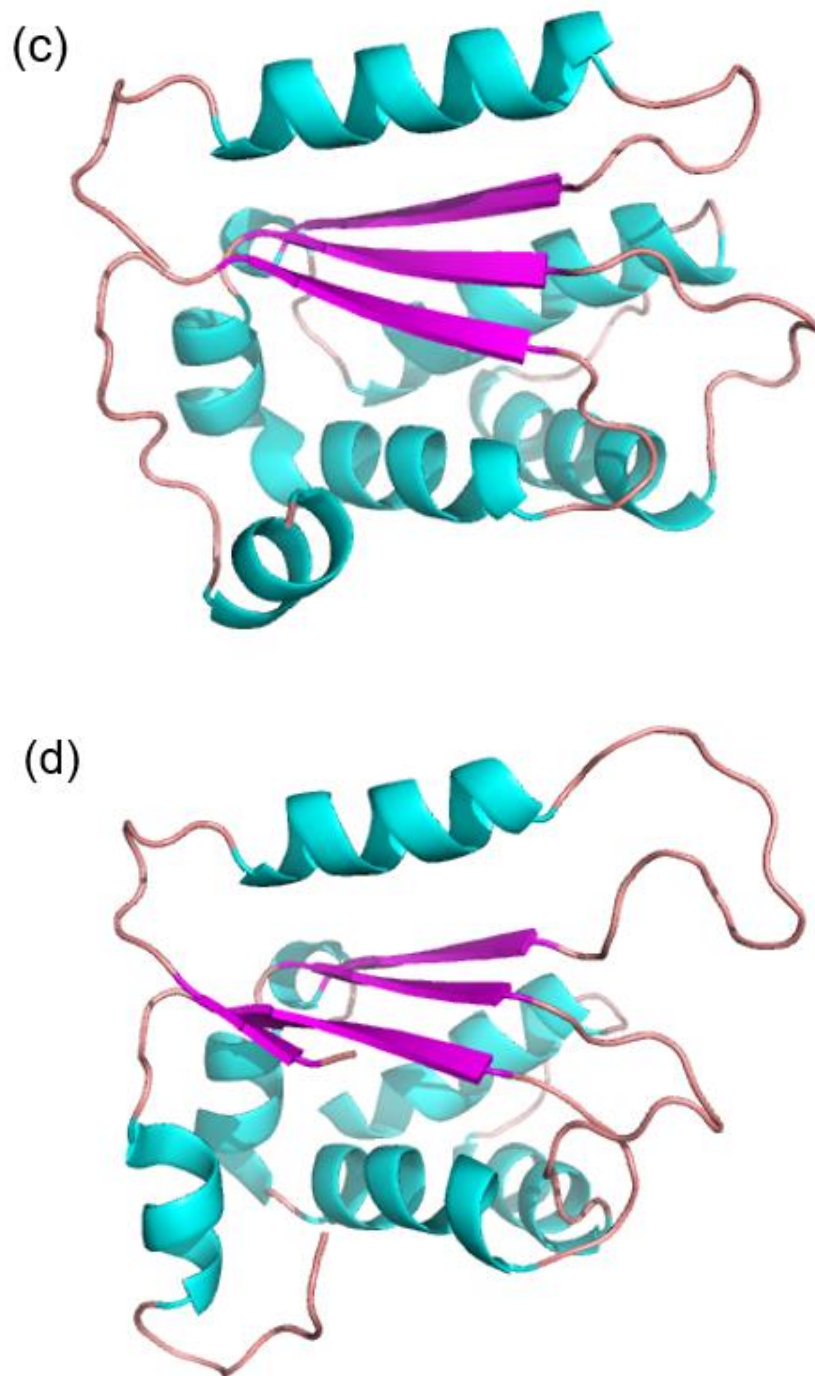
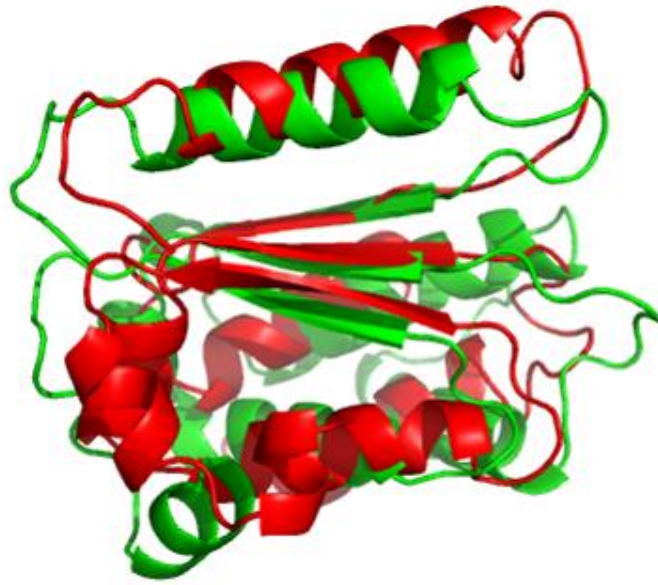


Figura 9 - Representação em *cartoon* do motivo AAA dos modelos gerados pelos diferentes métodos de predição de estrutura proteica. (a) I-TASSER. (b) Robbeta. (c) AlphaFold2. (d) Modelagem.

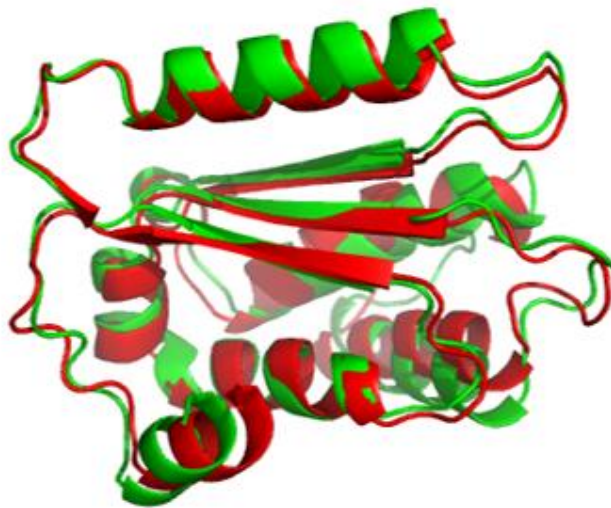
Fonte: Elaborada pelo autor.

Analisando a Figura 9, o motivo AAA gerado pelos diferentes métodos possuem uma topologia similar, principalmente ao se comparar os gerados por Robbeta e AlphaFold2. Essa semelhança fica evidente ao sobrepor os modelos na Figura 10.

(a)



(b)



(continua)

(continuação)

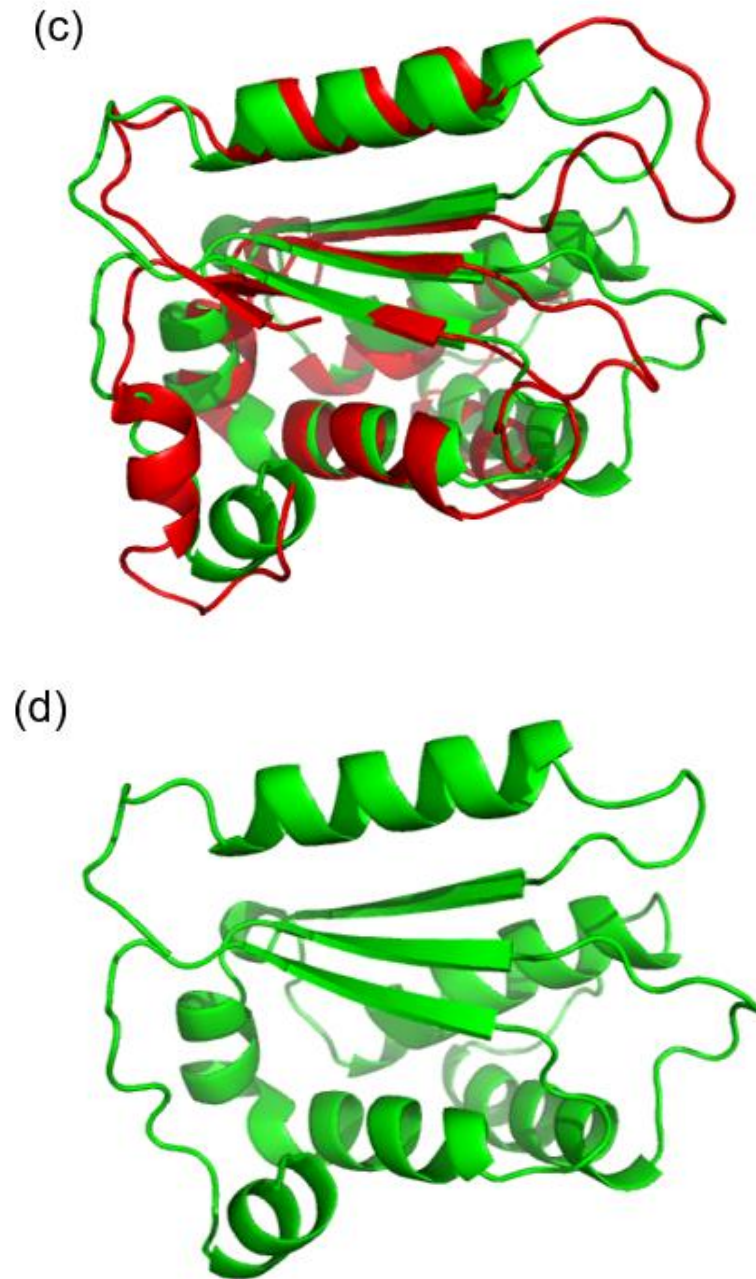


Figura 10 - Sobreposição da representação em *cartoon* obtida pelo AlphaFold2 do motivo AAA contra (a) I-TASSER, (b) Robbeta e (c) homologia. (d) apresenta a estrutura do AlphaFold2 isolada

Fonte: Elaborada pelo autor.

Para auxiliar na comparação dos modelos, foi construída a Tabela 3 com os valores de RMSD para as Figuras 8 e 10.

Tabela 3 - RMSD para a sobreposição dos modelos gerados por diferentes preditores contra TcORC1/CDC6 do AlphaFold2

Sequência	Preditor	RMSD (Å)
Inteira	I-TASSER	8,785
	Robetta	10,875
	Homologia	22,913
Motivo AAA	I-TASSER	1,513
	Robetta	1,143
	Homologia	2,021

Fonte: Elaborada pelo autor.

Com base nas Figuras 7 a 10 e Tabela 3, percebe-se uma maior similaridade entre os modelos gerados pelo Robetta e AlphaFold2 de forma geral por mais que os RMSD da sequência inteira tenha sido bem alto (de 8,785 até 22,913 Å). É muito perceptível que o modelo gerado pelo I-TASSER tem uma região significativa de *loops* mal resolvidos, que é evidenciada pela Figura 11, e que vai ao encontro do que o Diagrama de Ramachandran apontou (Figura 6a). Apesar disso, quando é analisado de forma isolada o motivo AAA, percebe-se que os modelos se sobrepõem bem (RMSD variando de 1,143 até 2,021 Å entre modelos), indicando que os preditores conseguiram ter resultados bem semelhantes quando se tratava de um motivo proteico comum.

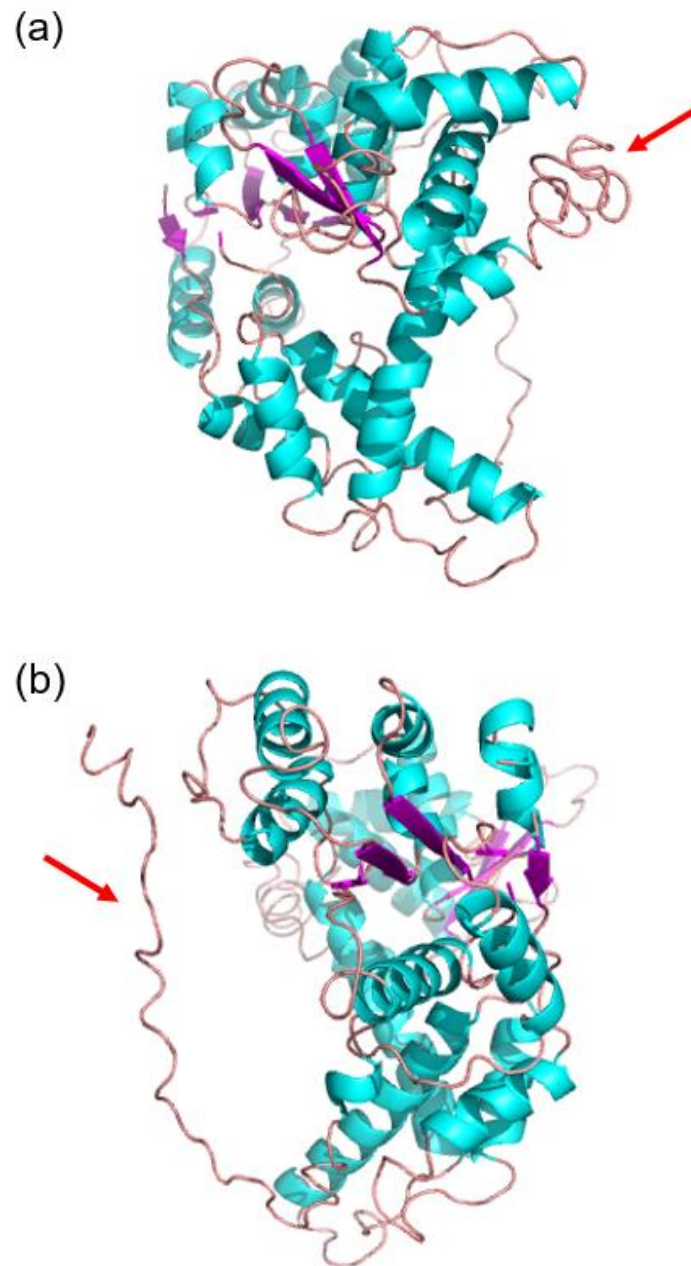


Figura 11 - Modelo de TcORC1/CDC6 gerado pelo I-TASSER em dois ângulos distintos (a) e (b) com regiões de loops mal resolvidos sendo indicadas por uma seta.

Fonte: Elaborada pelo autor.

Por fim, foi feita a submissão do motivo AAA (145 aminoácidos) do modelo gerado por AlphaFold2 ao servidor DALI utilizando como amostragem a base de dados do PDB e o AF-DB por se tratar da região conservada da proteína. Os resultados estão nas Tabelas 4 e 5.

Tabela 4 - Moléculas relevantes do servidor DALI por código PDB contra a base de dados do PDB.

PDB	Z	RMSD	LALI	NRES	%Id	Proteína	Espécie
6wgg-A	15,6	1,9	129	424	26	CDT1	<i>Saccharomyces cerevisiae</i>
6wgc-A	14,6	2,0	129	425	26	CDC6	<i>Saccharomyces cerevisiae</i>
5zr1-A	14,0	2,3	138	494	24	ORC1	<i>Saccharomyces cerevisiae</i>
5v8f-A	14,0	2,1	129	425	26	MCM2	<i>Saccharomyces cerevisiae</i>
2qby-A	13,5	2,3	130	366	25	CDC6	<i>Saccharolobus solfataricus</i>
2v1u-A	12,7	2,8	131	382	23	CDC6	<i>Aeropyrum pernix</i>
5ujm-A	12,6	2,3	120	337	27	ORC1	<i>Homo sapiens</i>
5uj7-A	12,6	2,4	120	337	27	ORC1	<i>Homo sapiens</i>
4xgc-A	11,9	2,3	111	310	31	ORC1	<i>Drosophila melanogaster</i>
5uj7-B	10,6	2,4	120	337	27	ORC4	<i>Homo sapiens</i>

Fonte: Elaborada pelo autor.

Tabela 5 - Moléculas relevantes do servidor DALI por espécie contra o AF-DB.

Espécie	Z	RMSD	LALI	NRES	%Id	Proteína	Código
<i>Trypanosoma brucei</i>	19,1	0,9	144	436	85	ORC1	TRYB2/AFDB:AF-Q384U5-F1
<i>Leishmania infantum</i>	23,3	1,0	143	431	65	ORC1	LEIIN:AF-A4I376-F1
<i>Plasmodium falciparum</i>	16,1	2,4	138	1189	27	ORC1	PLAF7:AF-A0A144A1W6-F1
<i>Schistosoma mansoni</i>	14,7	2,3	135	569	27	ORC1	SCHMA/AFDB:AF-A0A3Q0KUC0-F1
<i>Schistosoma mansoni</i>	14,0	2,5	134	524	25	CDC6	SCHMA/AFDB:AF-A0A3Q0KP41-F1
<i>Drosophila melanogaster</i>	16,1	2,2	136	924	29	ORC1	DROME:AF-O16810-F1
<i>Drosophila melanogaster</i>	12,6	2,5	130	662	23	CDC6	DROME:AF-Q9VSM9-F1
<i>Homo sapiens</i>	15,7	2,4	136	861	26	ORC1	HUMAN:AF-Q13415-F1

Fonte: Elaborada pelo autor.

Pela análise das Tabelas 4 e 5, nota-se que o motivo AAA de TcORC1/CDC6 foi modelado por AlphaFold2 de forma satisfatória, pois, ao se buscar similaridade de enovelamento contra a base do PDB, foram resgatados principalmente ORC1 e CDC6 de vários organismos e com valores de RMSD próximos de 2. As demais proteínas apresentadas são pertencentes ao pre-RC, e isso é indicativo que esse motivo é comum a mais elementos do pre-RC em alguns organismos. E, ao se buscar similaridade de enovelamento contra a AF-DB de organismos diversos, só se obteve ORC1 e CDC6.














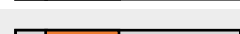












4.1.3 Filogenética

Ao iniciar o estudo filogenético, logo se percebe que, mesmo entre Tripanossomatídeos, é marcante uma divergência evolutiva na sequência, o que já havia sido discutido por OCAÑA-PALLARÈS.⁴¹ Analisando as sequências, observa-se um único domínio conservado entre todas as sequências analisadas e identidades baixas em especial ao se sair do gênero *Trypanosoma*. No gênero *Leishmania*, já se vê como maior identidade 57,11%. Isso está representado na Tabela 6.







O único domínio conservado é o AAA. Este é característico de uma Família de ATPases PF00004. É interessante ter esse único domínio conservado, já que isso é um indício que o mais fundamental elemento da sequência é realmente a habilidade de hidrolisar ATP, o que vai ao encontro da ligação de TcORC1/CDC6 à ORI é dependente da hidrólise de ATP.

As árvores filogenéticas com base no alinhamento múltiplo das sequências apresentadas na Tabela 6 foram construídas e são apresentadas nas Figuras 12 a 14. Na Figura 12, são evidenciados os nós utilizados para a predição das sequências ancestrais. A escolha foi baseada na separação entre gêneros a partir do nó que separa o gênero *Trypanosoma* do *Leishmania*. Para auxiliar na discussão das árvores filogenéticas, foi elaborada a Tabela 7 com informações biológicas de cada espécie analisada. Dessa forma, será possível constatar se as divisões em nível molecular fazem sentido biológico.

Tabela 6 - Organismos escolhidos para a análise filogenética com grau de identidade, número de aminoácidos e esquema representativo de motivo proteico. C é a cobertura; Id, a identidade; e aa, aminoácidos.

Organismo	C (%)	Id (%)	aa	Motivos proteicos
<i>T. cruzi</i>	-	-	450	
<i>T. conorhini</i>	97	87,41	437	
<i>T. rangeli</i>	97	85,81	437	
<i>T. grayi</i>	97	84,90	437	
<i>T. theileri</i>	97	82,61	437	
<i>T. brucei</i>	96	78,85	436	
<i>T. congolense</i>	96	73,62	436	
<i>T. vivax</i>	96	69,27	435	
<i>L. pyrrocoris</i>	95	57,11	431	
<i>L. donovani</i>	95	55,48	431	
<i>L. tarentolae</i>	95	55,48	431	
<i>L. mexicana</i>	95	55,24	431	
<i>L. orientalis</i>	95	55,01	431	
<i>L. major</i>	95	55,01	431	
<i>L. seymouri</i>	95	55,01	431	
<i>L. enriettii</i>	95	54,55	431	
<i>L. guyanensis</i>	95	54,42	431	
<i>L. martiniquensis</i>	95	54,31	431	
<i>L. panamensis</i>	96	54,13	457	
<i>L. braziliensis</i>	96	54,13	457	
<i>S. culicis</i>	54	53,69	431	
<i>P. hertigi</i>	95	52,91	431	
<i>D. melanogaster</i>	49	31,51	924	
<i>S. bovis</i>	54	29,1	597	
<i>H. sapiens</i>	63	26,01	861	
<i>P. vivax</i>	64	25,5	1162	

Legenda:

-  AAA: PF00004, Família de ATPase
-  AAA_lid_10: PF17872, domínio lid AAA
-  Nop: PF01798, domínio de ligação a snoRNA
-  Cdc6_C: PF09079, CDC6, domínio hélice alada
-  BAH: PF01426, domínio BAH
-  Band_7_1: PF13421, família do domínio Band_7 SPFH

Fonte: Elaborada pelo autor.

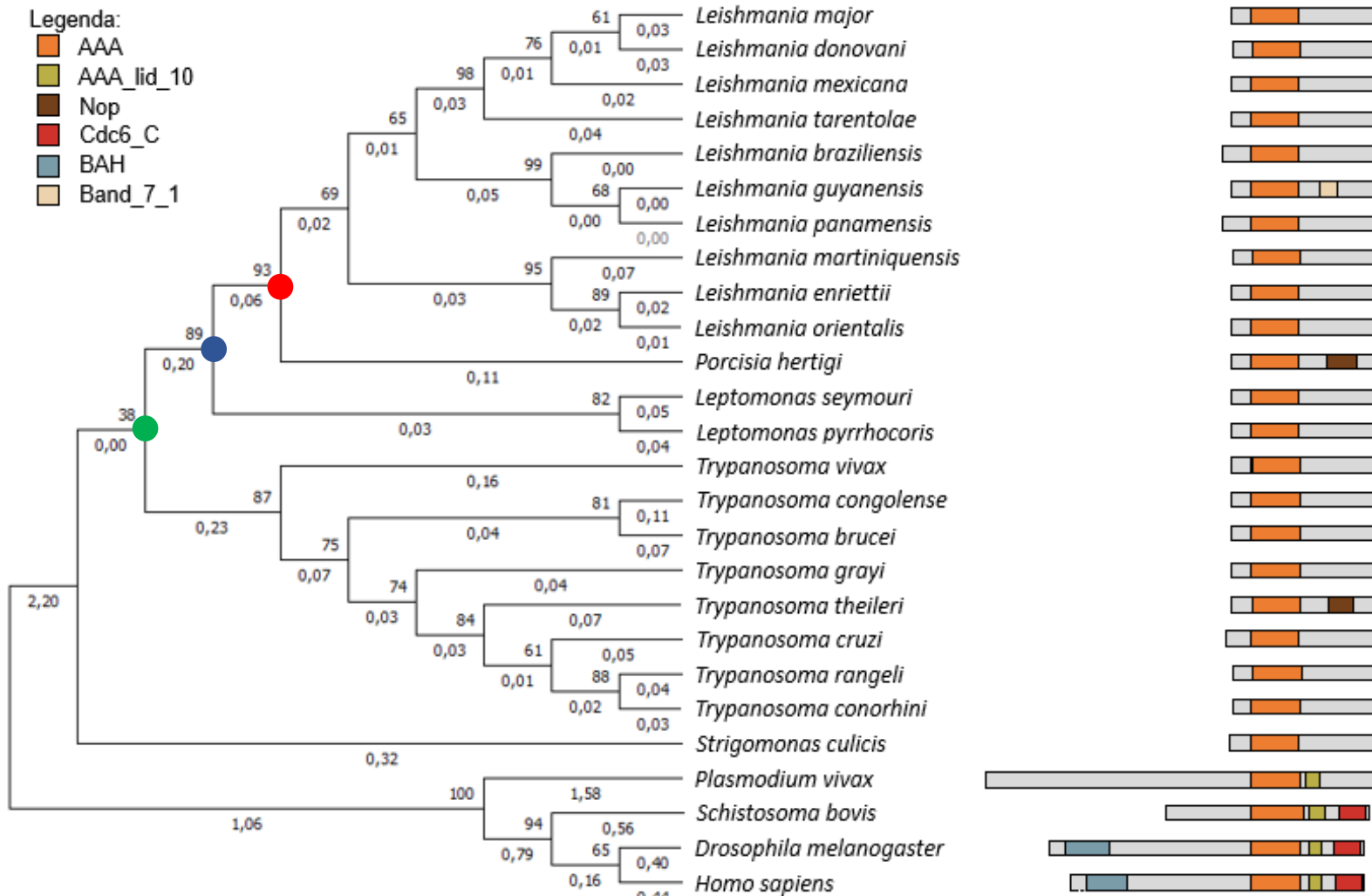


Figura 12 - Árvore construída pelo método de Máxima Verossimilhança. Destaca-se em verde, azul e vermelho os 3 nós escolhidos para a predição da sequência ancestral.

Fonte: Elaborada pelo autor.

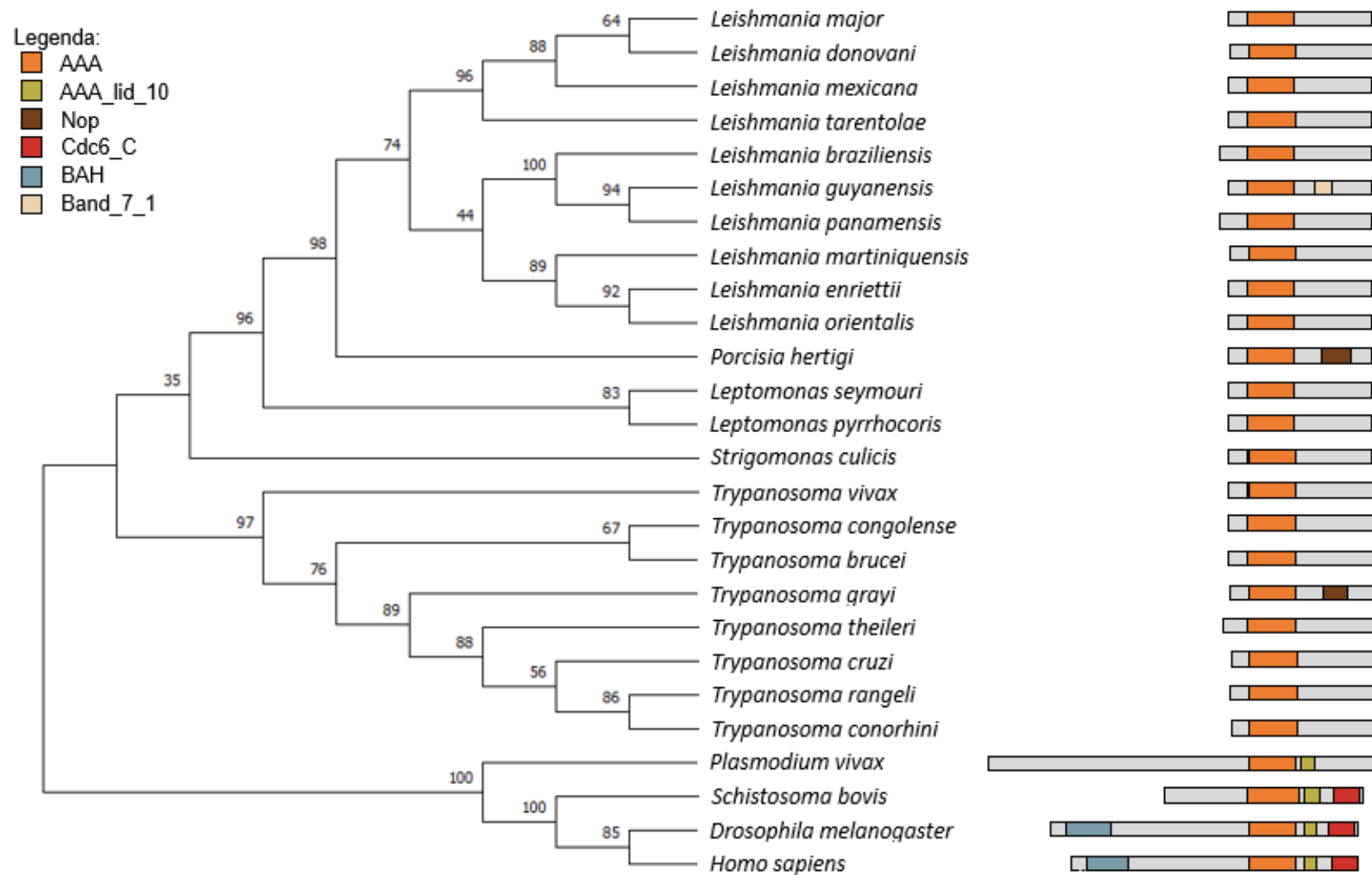


Figura 14 - Árvore construída pelo método de Máxima Parcimônia.

Fonte: Elaborada pelo autor.

Tabela 7 - Dados biológicos e geográficos das espécies analisadas no estudo filogenético.

Espécie	Mundo	Localização	Ciclo de vida	Localização celular	Nº de acesso
<i>T. rangeli</i>	Novo	América do Sul e Central	Dixênico	Extracelular	XP029239059.1
<i>T. conorhini</i>	Cosmopolita	Cosmopolita	Dixênico	Extracelular	XP029228900.1
<i>T. cruzi</i>	Novo	América do Sul e Central	Dixênico	Intracelular	XP808490.1
<i>T. grayi</i>	Velho	África	Dixênico	Extracelular	XP009308441.1
<i>T. theileri</i>	Cosmopolita	Cosmopolita	Dixênico	Intracelular	XP028882694.1
<i>T. congolense</i>	Velho	África	Dixênico	Extracelular	CCC95343.1
<i>T. brucei</i>	Velho	África	Dixênico	Extracelular	XP011779967.1
<i>T. vivax</i>	Velho	Cosmopolita	Dixênico	Extracelular	KAG8347316.1
<i>L. enriettii</i>	Novo	América do Sul e Austrália	Dixênico	Intracelular	KAG5474902.1
<i>L. orientalis</i>	Velho	Ásia	Dixênico	Intracelular	KAG5474565.1
<i>L. martiniquensis</i>	Cosmopolita	Cosmopolita	Dixênico	Intracelular	KAG5474236.1
<i>L. guyanensis</i>	Novo	América do Sul	Dixênico	Intracelular	CCM16840.1
<i>L. panamensis</i>	Novo	América do Sul	Dixênico	Intracelular	XP010700348.1
<i>L. braziliensis</i>	Novo	América do Sul e Central	Dixênico	Intracelular	XP001566040.1
<i>L. major</i>	Velho	África e Ásia	Dixênico	Intracelular	XP001684281.1
<i>L. donovani</i>	Velho	África e Ásia	Dixênico	Intracelular	AEP37338.1
<i>L. mexicana</i>	Novo	América Central e do Norte	Dixênico	Intracelular	XP003876836.1
<i>L. tarentolae</i>	Cosmopolita	Sul da Europa e África	Dixênico	Intracelular	GET89918.1
<i>P. hertigi</i>	Novo	América Central e do Sul	Dixênico	Intracelular	KAG5500327.1
<i>L. seymouri</i>	Cosmopolita	Cosmopolita	Monoxênico	Extracelular	KPI88290.1
<i>L. pyrrhocoris</i>	Cosmopolita	Cosmopolita	Monoxênico	Extracelular	XP015664020.1
<i>S. culicis</i>	Cosmopolita	Cosmopolita	Monoxênico	Extracelular	EPY20124.1

Fonte: Elaborada pelo autor.

Ao analisar as árvores filogenéticas das Figuras 12 a 14 e a Tabela 7, nota-se que a topologia das árvores é bem semelhante. Como já era esperado, houve o agrupamento dos gêneros *Trypanosoma* e *Leishmania* e, analisando a separação nos clados, é percebido que de forma geral espécies do Novo Mundo ficaram próximas e o mesmo ocorreu para as do Velho Mundo com as cosmopolitas entre elas.

De forma mais detalhada, será realizada a análise do gênero *Trypanosoma* a partir do clado que apresenta *Trypanosoma cruzi*, que é o foco do trabalho. Nesse clado, também se encontra *Trypanosoma rangeli* e *Trypanosoma conorhini*, as duas últimas espécies formam um clado. Biologicamente isso faz sentido por serem parasitas com ciclo de vida extracelular e não pertencerem ao Velho Mundo, enquanto *Trypanosoma cruzi* é intracelular. Essa maior proximidade evolutiva dessas duas espécies com *Trypanosoma cruzi* já foi reportado por BRADWELL.⁴²

Em seguida, tem-se *Trypanosoma theileri*. Este é o único outro *Trypanosoma* cosmopolita analisado, fazendo com que sua localização na árvore entre as espécies de Velho e Novo Mundo seja válida. Além disso, é um parasita intracelular, o que é característico de espécies do Novo Mundo.⁴³

A partir daí, estão presentes as espécies do gênero *Trypanosoma* do Velho Mundo. A separação de *Trypanosoma grayi* das demais espécies pode ser justificado pelo seu hospedeiro. Diferente dos outros que infectam mamíferos, *Trypanosoma grayi* infecta répteis, mais especificamente crocodilos.⁴⁴ Essa diferença de hospedeiros justifica em nível biológico a divisão filogenética observada.

Sobre *Leishmania orientalis*, esta é uma espécie descrita recentemente cujo vetor natural permanece desconhecido. A separação dela das demais espécies do Velho Mundo pode ser justificada por ela ser presente na Tailândia, enquanto as demais são encontradas na África e principalmente sudoeste da Ásia.⁴⁵

A proximidade filogenética de *Porcisia hertigi* ao gênero *Leishmania* é justificável já que até 2016 a espécie era nomeada *Leishmania hertigi*.⁴⁶

Há ainda na análise filogenética 3 espécies monoxênicas, sendo que esse tipo de ciclo de vida é ancestral ao dixênico em tripanosomatídeos. Entre esses, constam dois representantes do gênero *Leptomonas*, que é reportado por ser intimamente relacionado ao gênero *Leishmania*,⁴⁷ o que vai ao encontro da disposição obtida nas árvores filogenéticas.

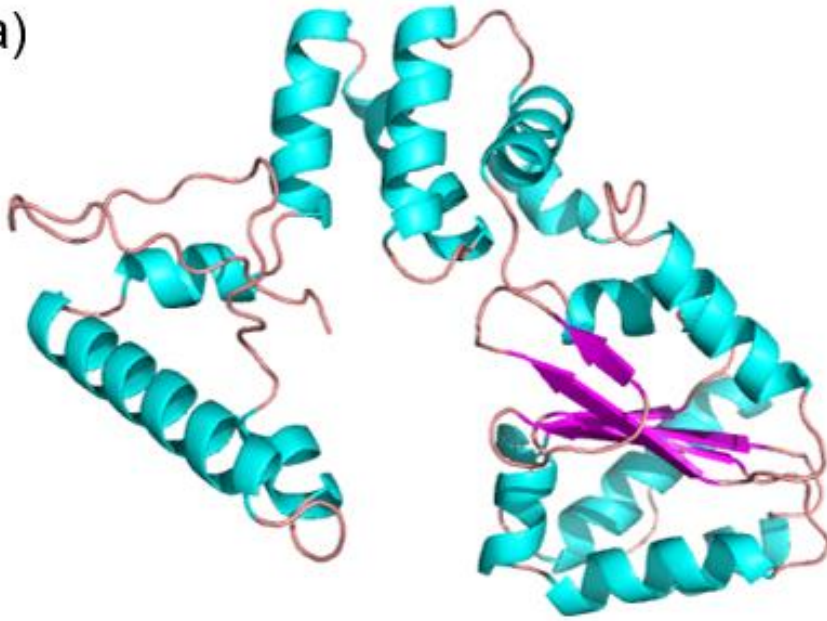
O terceiro representante monoxênico é *Strigomonas culicis*. Esta é a espécie que mais sofreu alteração de acordo com o método utilizado para gerar a árvore

filogenética, ora se agrupando com Trypanosoma (Agrupamento de Vizinhos), ora com Leishmania (Máxima Parcimônia) e ora com nenhum dos dois (Máxima Verossimilhança). Ela é modelo de estudos sobre a teoria endossimbiótica da vida por possuir uma bactéria simbiótica que mantém uma relação obrigatória com o protozoário⁴⁸. Por essa questão de suposta ancestralidade do organismo, biologicamente seria esperada a distribuição apresentada pelo método de Máxima Verossimilhança.

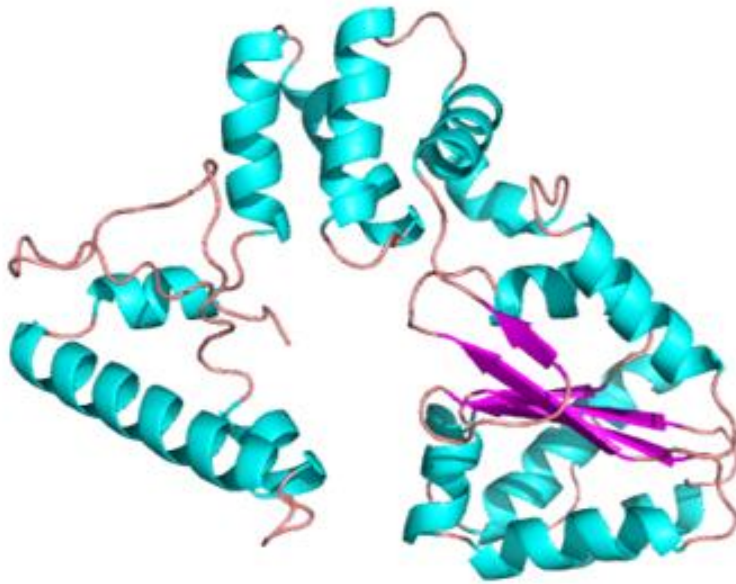
Em vista da discussão acima e por possuir *scores* de *bootstrap* mais altos, considera-se que o método de Máxima Verossimilhança gerou a melhor árvore filogenética para os dados analisados e nele foram escolhidos 3 nós de interesse para determinação de sequência ancestral, assinalados na Figura 12. Por questão de praticidade, será nomeado o nó verde, que separa o gênero Trypanosoma de Leishmania, de Tryp_Leish. O nó azul, que separa Leishmania de Leptomonas, será intitulado Leish_Lepto. E, por fim, o nó vermelho, que separa Leishmania de Porcisia, será denominado Leish_Porc.

Com as sequências ancestrais preditas (Anexo B), encontrou-se que estas também apresentam o motivo AAA e foi realizada a submissão ao AlphaFold2 com a intenção de comparar com o modelo gerado para TcORC1/CDC6 por AlphaFold2. As Figuras 15 a 18 mostram representações dos resultados como feito em 4.1.2.

(a)



(b)



(continua)

(continuação)

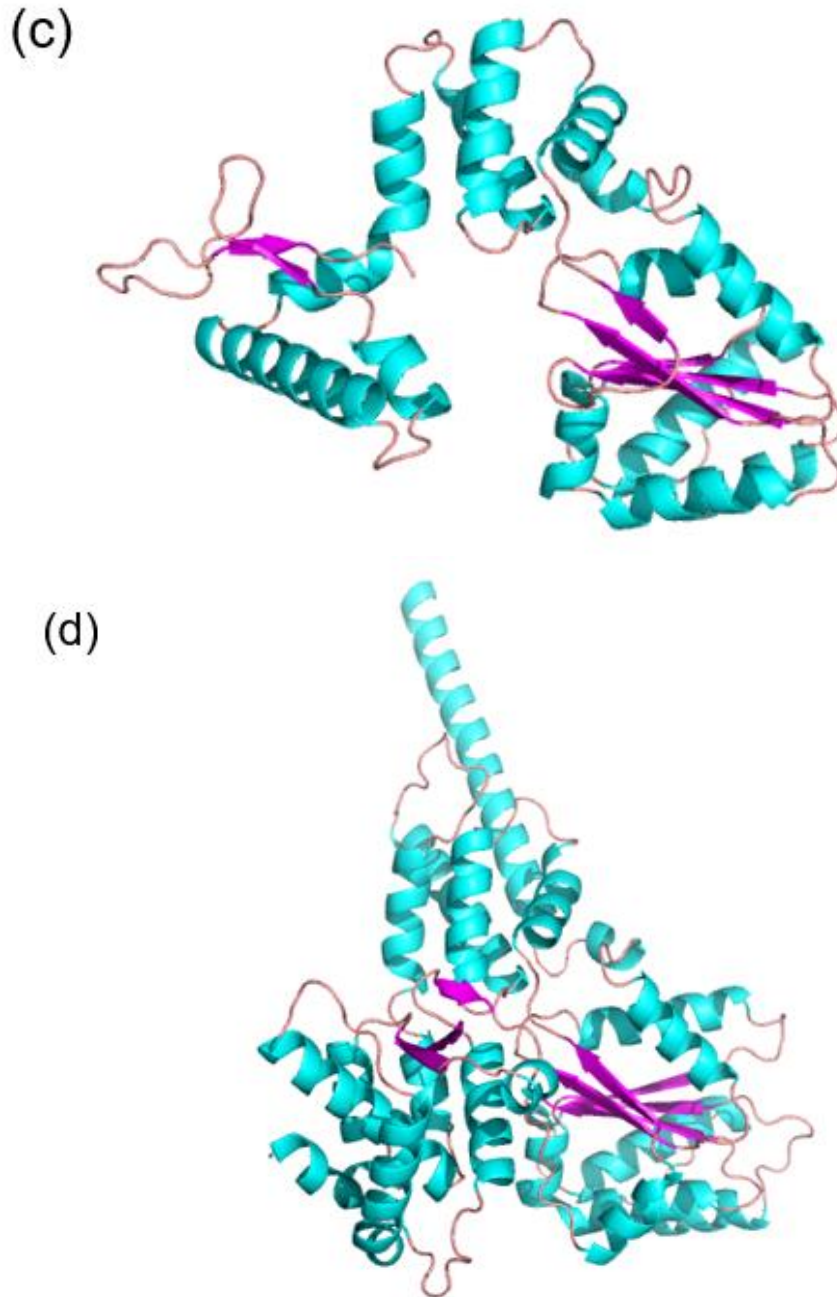
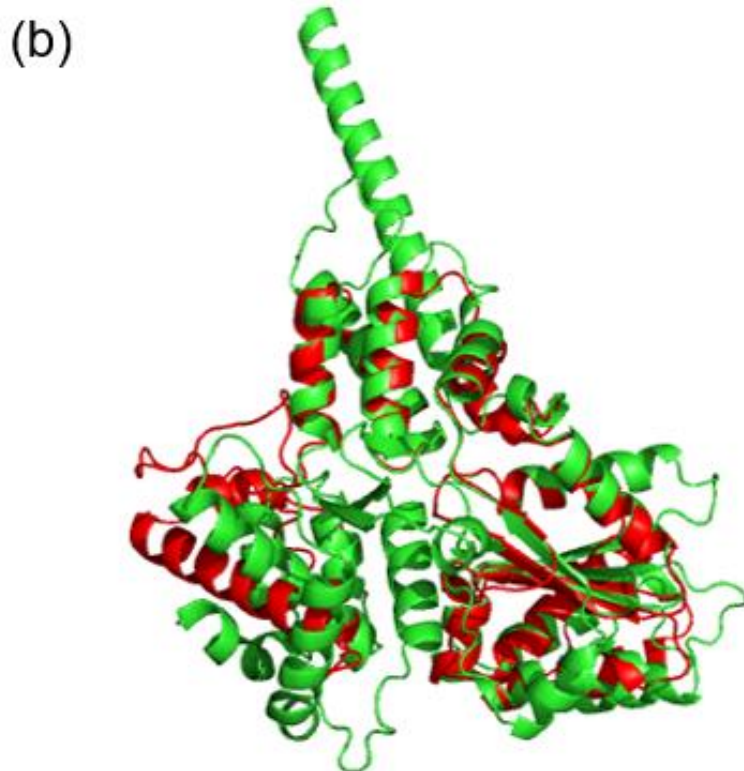
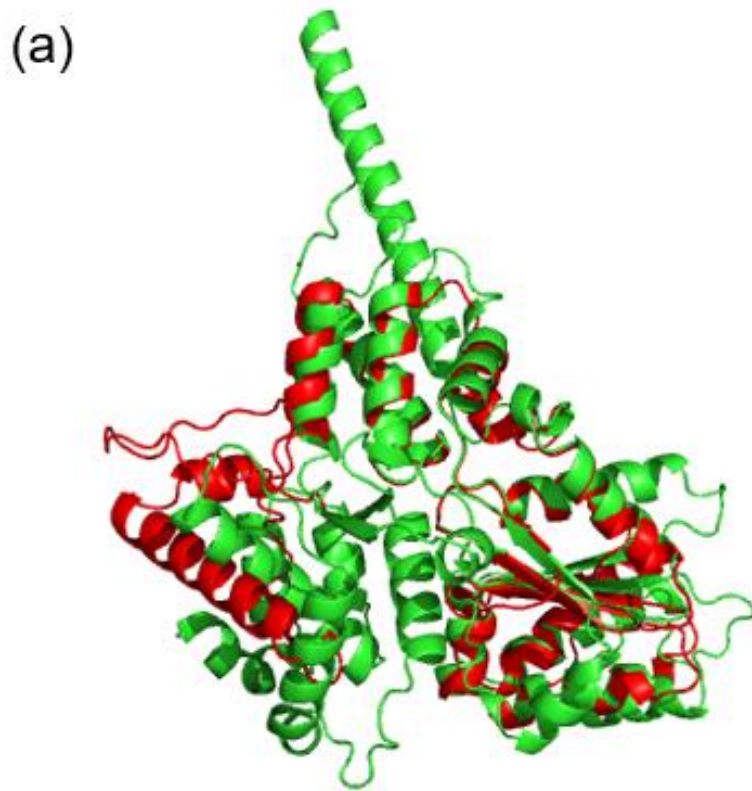


Figura 15 - Representação em *cartoon* dos modelos das seqüências ancestrais dos nós: (a) Tryp_Leish, (b) Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.

Fonte: Elaborada pelo autor.

Pela análise da Figura 15, nota-se que os modelos gerados para os ancestrais Tryp_Leish, Leish_Lepto e Leish_Porc são bem parecidos entre si, mas aparentam ser diferentes do modelo de TcORC1/CDC6. Isso pode ser explicado por conta de que, durante o processo de análise filogenética, ocorre a seleção de blocos de maior

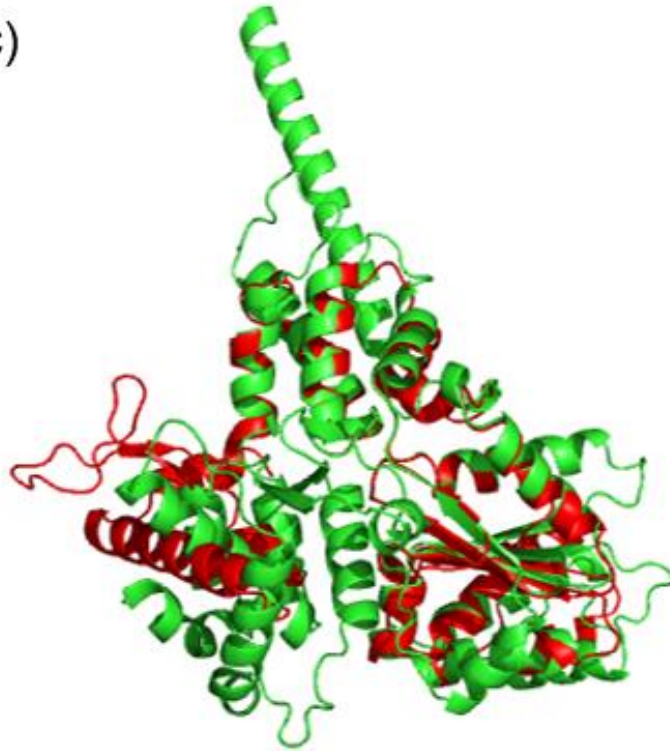
relevância para a construção das árvores, e as sequências ancestrais foram obtidas a partir dessas sequências truncadas.



(continua)

(continuação)

(c)



(d)

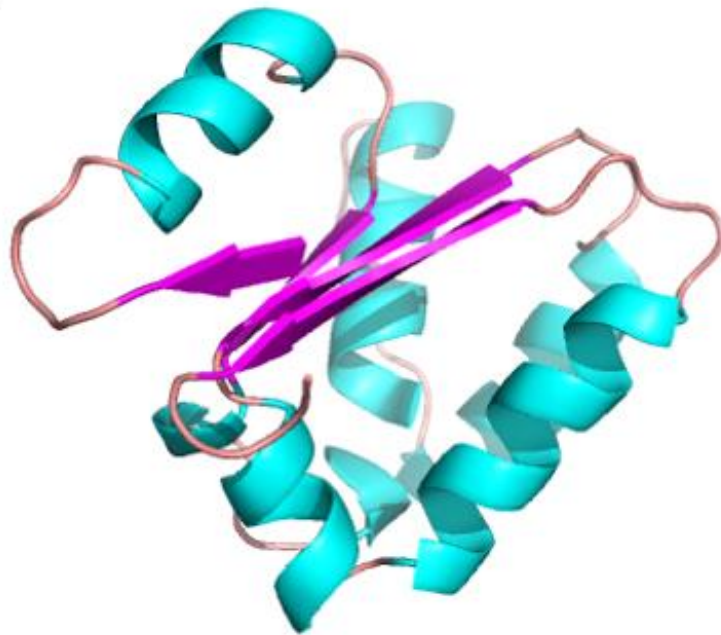


Figura 16 - Representação em *cartoon* da sobreposição de TcORC1/CDC6 contra os modelos das seqüências ancestrais dos nós: (a) Tryp_Leish, (b) Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.

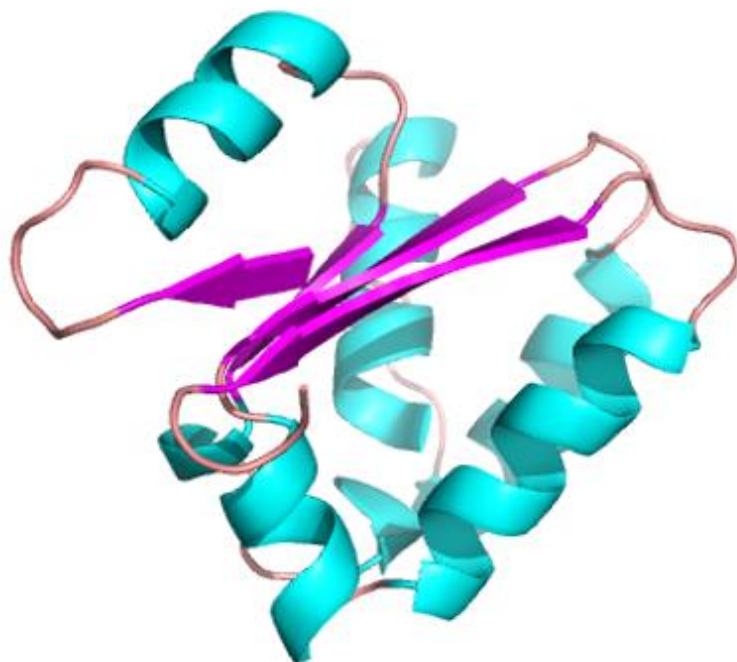
Fonte: Elaborada pelo autor.

Na Figura 16, a sobreposição dos modelos das sequências ancestrais com a gerada para TcORC1/CDC6 mostra de forma mais clara que há uma distinção entre os modelos gerados e novamente é percebida novamente uma região com melhor sobreposição aparente. Isso é evidenciado nas Figuras 17 e 18.

(a)



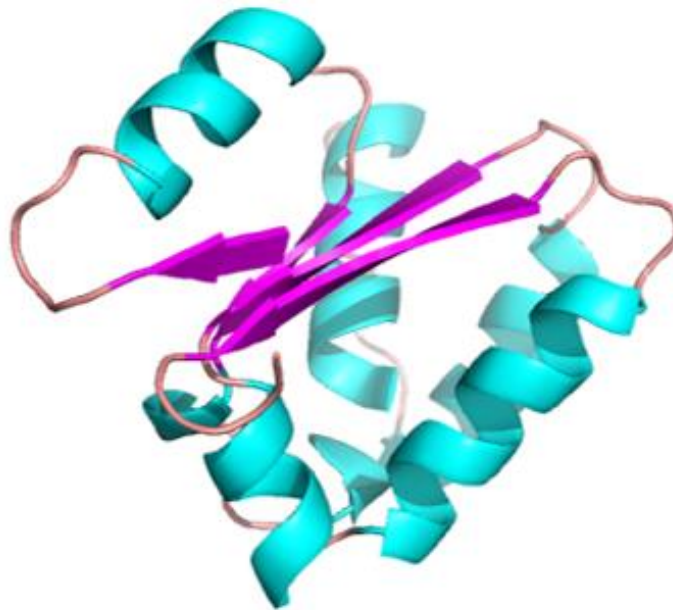
(b)



(continua)

(continuação)

(c)



(d)

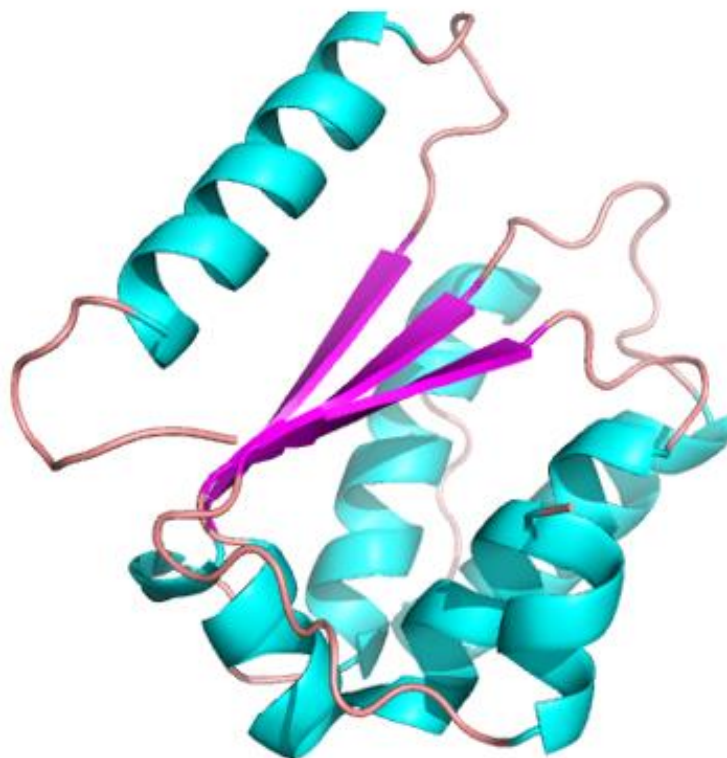


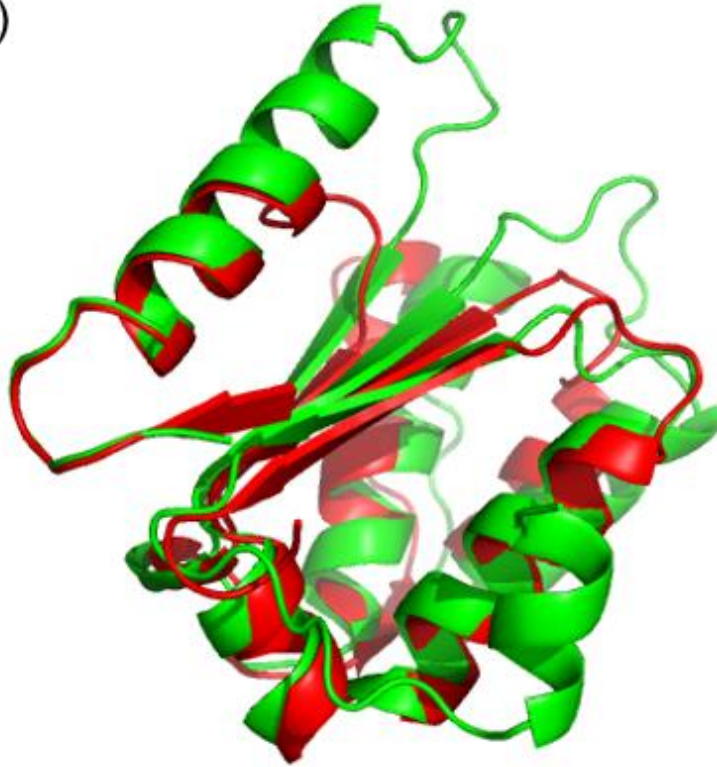
Figura 17 - Representação em *cartoon* do motivo AAA dos modelos das sequências ancestrais dos nós: (a) Tryp_Leish, (b) Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.

Fonte: Elaborada pelo autor.

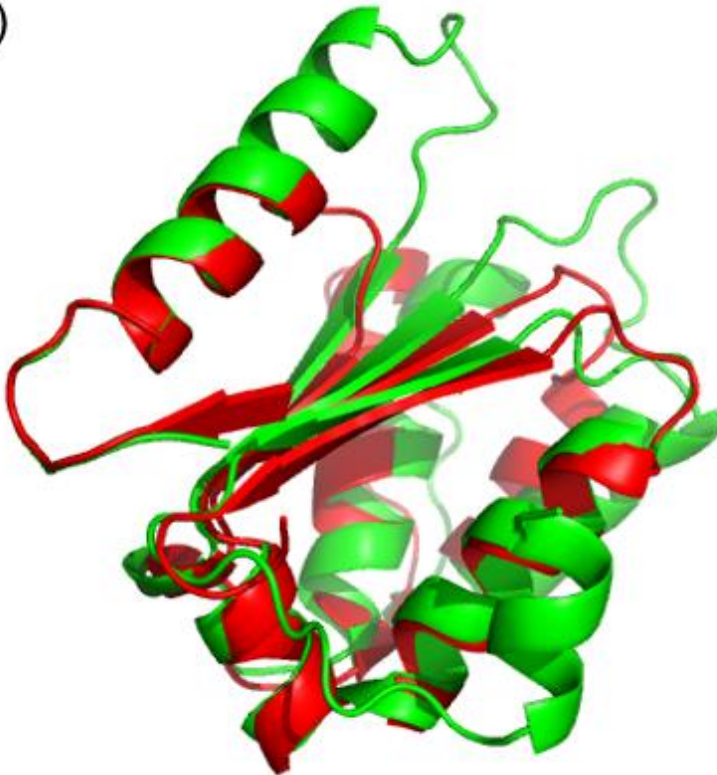
Pela análise da Figura 17, é observada topologia similar entre todos os modelos. Como eles representam o motivo conservado por todas as sequências da

análise filogenética, é coerente que nas sequências ancestrais essa região seja também conservada.

(a)



(b)



(continua)

(continuação)

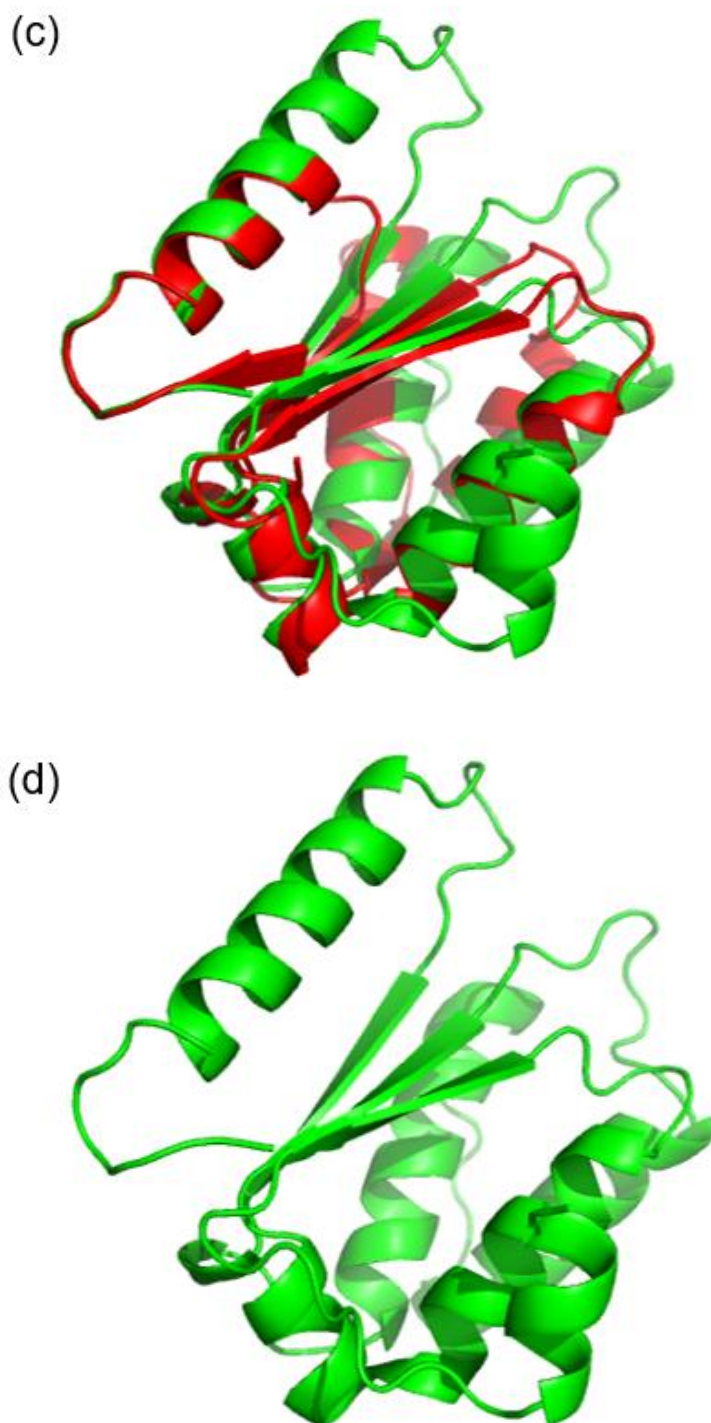


Figura 18 - Representação em *cartoon* da sobreposição do motivo AAA de TcORC1/CDC6 contra o motivo AAA dos modelos das sequências ancestrais dos nós: (a) Tryp_Leish, (b) Leish_Lepto, (c) Leish_Porc. (d) TcORC1/CDC6.

Fonte: Elaborada pelo autor.

Para auxiliar na comparação dos modelos, foi construída a Tabela 8 com os valores de RMSD para as Figuras 16 e 18.

Tabela 8 - RMSD para a sobreposição dos modelos ancestrais contra TcORC1/CDC6.

Sequência	Preditor	RMSD (Å)
Inteira	Tryp_Leish	5,714
	Leish_Lepto	5,421
	Leish_Porc	5,817
Motivo AAA	Tryp_Leish	1,003
	Leish_Lepto	1,168
	Leish_Porc	1,289

Fonte: Elaborada pelo autor.

Ao analisar as Figuras 15 a 18 e a Tabela 8, nota-se que os modelos ancestrais gerados são bem similares entre si e, por mais que a sobreposição contra TcORC1/CDC6 inteira tenha apresentado regiões com baixa sobreposição estrutural, média de 6,65 Å, a sobreposição contra o motivo AAA apresentou RMSD próximo de 1 Å, demonstrando que esse motivo é conservado evolutivamente, coerente com seu papel importante para a função da proteína.

4.2 PRODUÇÃO DA PROTEÍNA RECOMBINANTE

4.2.1 Purificação

4.2.1.1 Cromatografia de afinidade

Pela cromatografia de afinidade, foi observado apenas um pico de eluição correspondente a 1 mM de Maltose. O “ombro” no cromatograma da afinidade foi obtido durante a lavagem com 0 mM de maltose e apresentou conteúdo proteico muito parecido com o não-ligado da coluna. Isso está apresentado na Figura 19.

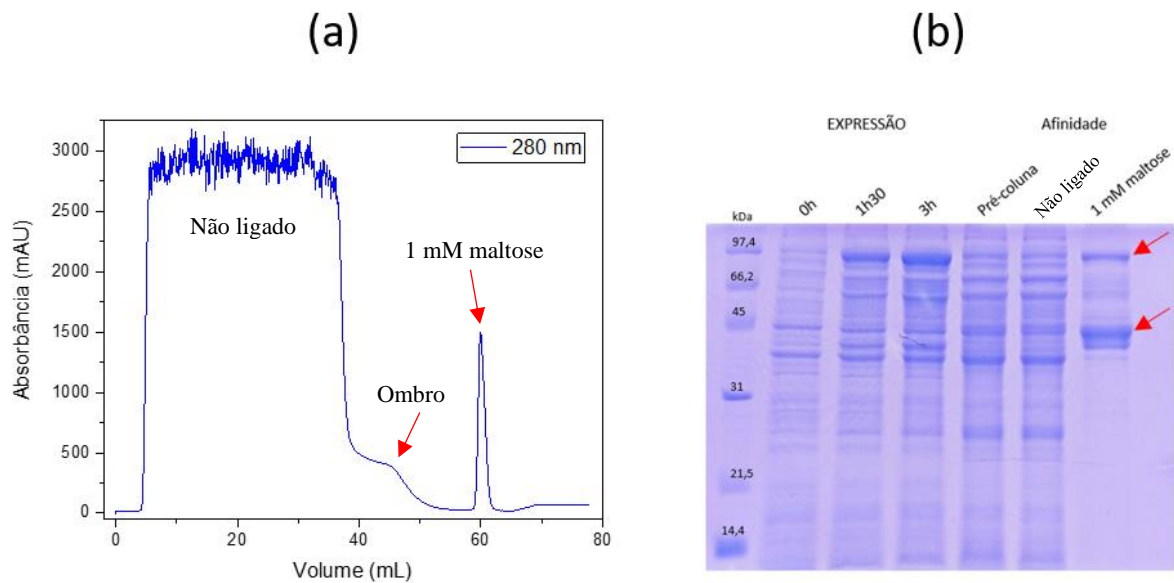


Figura 19 - Purificação por Cromatografia de Afinidade de TcORC1/CDC6. (a) cromatograma da purificação. (b) Eletroforese em gel de poliacrilamida (PAGE) com dodecil-sulfato de sódio (SDS) 10% das frações da expressão e purificação de TcORC1/CDC6.

Fonte: Elaborada pelo autor.

Pela análise da Eletroforese em gel de poliacrilamida (PAGE) com dodecil-sulfato de sódio (SDS) 10% da Figura 19b, percebe-se que a eluição apresenta duas bandas majoritárias: uma logo abaixo de 97,4 kDa, que poderia ser TcORC1/CDC6 + MBP, e outra próximo de 45 kDa, que precisou ser estudada mais a fundo.

4.2.1.2 Cromatografia de exclusão molecular

Para se obter maior grau de pureza, foi realizado a etapa de Cromatografia de Exclusão Molecular (SEC). Pelo cromatograma desta, pode-se observar a formação de dois picos (Figura 20a).

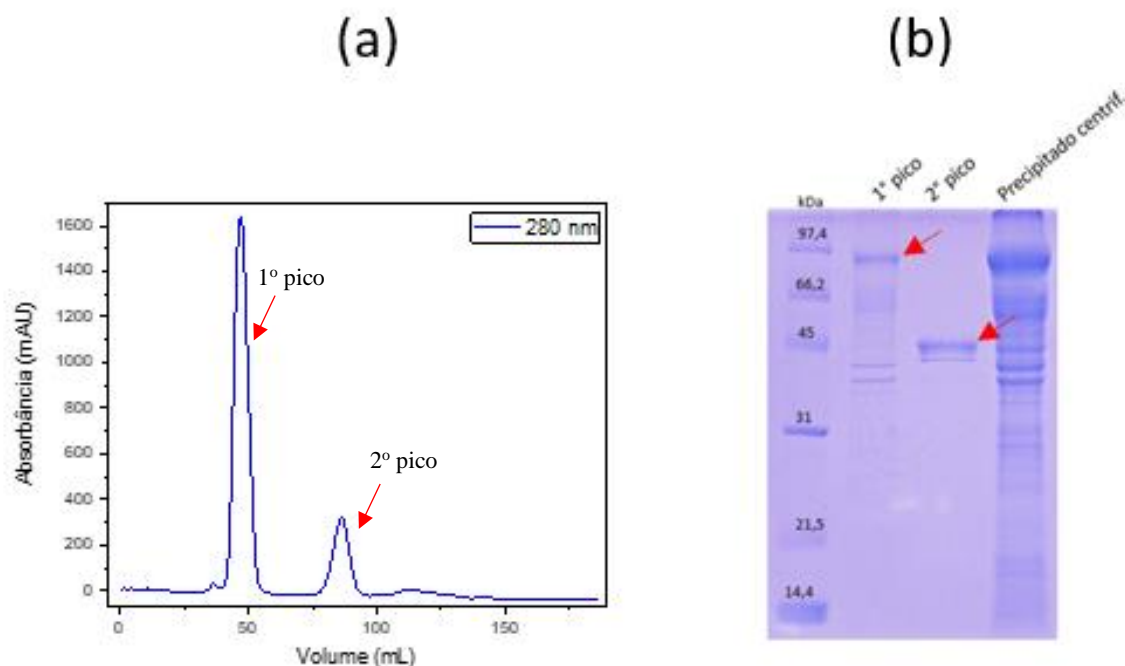


Figura 20 - Purificação por Cromatografia de Exclusão Molecular de TcORC1/CDC6. (a) cromatograma da purificação. (b) Eletroforese em gel de poliacrilamida (PAGE) com dodecil-sulfato de sódio (SDS) 10% das frações da purificação.

Fonte: Elaborada pelo autor.

Ao submeter as frações à análise de SDS-PAGE (Figura 20b), notou-se que o primeiro pico do cromatograma, consistia majoritariamente de uma proteína com massa molecular próximo de 97,4 kDa, esperado para o complexo proteico TcORC/CDC6 + MBP. No segundo pico, eluído em um volume de próximo de 86 mL, observou-se a predominância de uma proteína com massa próxima de 45 kDa, o que poderia corresponder ao 50 kDa de TcORC1/CDC6. Houve muita perda de proteína durante a etapa de centrifugação por precipitação, o que é evidente na Figura 20b.

4.2.2 Western-blot

O *western-blot* foi realizado com o extrato proteico, a eluição da cromatografia de afinidade e os dois picos da cromatografia de exclusão molecular (Figura 21). A primeira informação a ser observada no gel é que a banda de massa molecular próxima a 97,5 kDa, que possivelmente refere-se a TcORC/CDC6 + MBP, está presente em todas as três etapas do processo. Desde a etapa da expressão, percebe-se bandas de menor massa que poderiam ser resultado de uma proteólise.

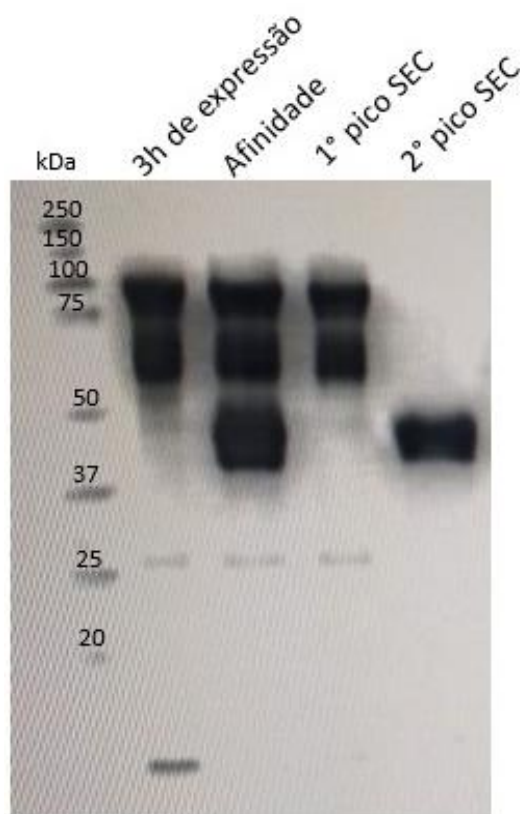


Figura 21 - Resultado do *western-blot* das amostras de expressão, purificação por Cromatografia de Afinidade e Cromatografia de Exclusão Molecular de TcORC1/CDC6.

Fonte: Elaborada pelo autor.

Pela análise do *western-blot*, decidiu-se que seria utilizado o segundo pico da cromatografia de exclusão molecular para as etapas posteriores. A decisão foi tomada pelo fato dela ter demonstrado interação com o anticorpo anti-TcORC1/CDC6 e possuir peso molecular próximo a 45 kDa.

4.3 CARACTERIZAÇÃO BIOFÍSICA

4.3.1 Curva de calibração da Exclusão Molecular

Ao se submeter a amostra escolhida em 4.2.2 à Cromatografia de Exclusão Molecular com Superdex™ 200 10/300 (GE) analítica de 24 mL, obteve-se um volume de eluição de 17,59 mL, evidenciado pelo cromatograma da Figura 22.

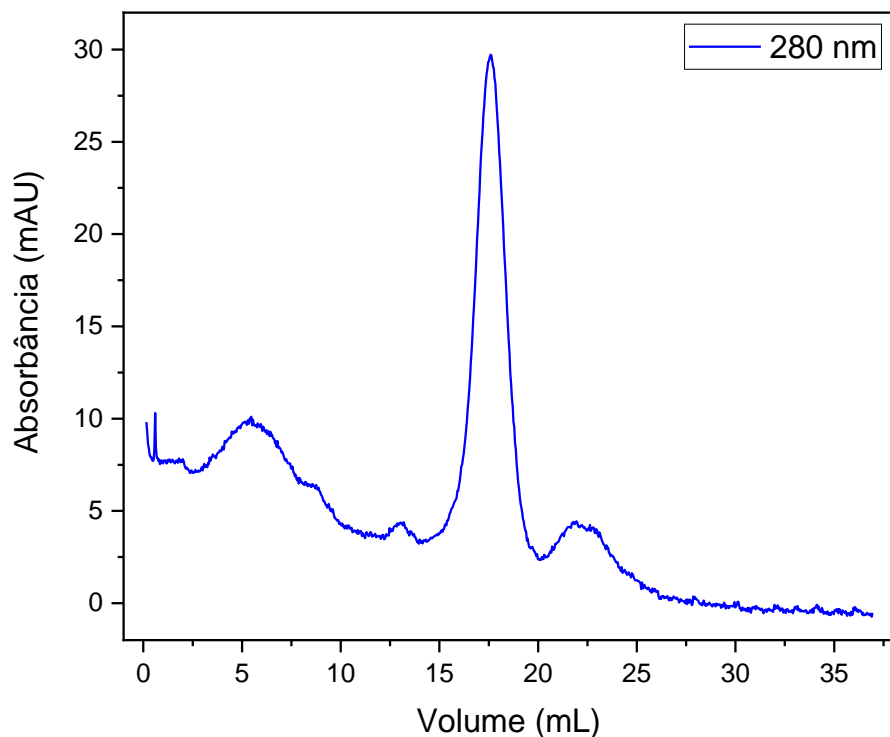


Figura 22 - Cromatograma da Cromatografia de Exclução Molecular com coluna analítica de TcORC1/CDC6.

Fonte: Elaborada pelo autor.

Os volumes de eluição para os padrões de massa utilizados estão apresentados na Tabela 9.

Tabela 9 - Padrões para calibração da coluna. MM é Massa Molar. R_H é raio hidrodinâmico. V_e é volume eluição. K_{av} é a razão de eluição.

Padrão	MM (kDa)	R_H (Å)	V_e (mL)	k_{av}
Ribonuclease	13,7	16,4	19,69	0,706
Anidrase carbônica bovina	29,0	21,0	18,50	0,625
Ovalbumina	43,0	30,5	17,25	0,540
Conalbumina	75,0	36,4	16,19	0,468
β -amilase (batata doce)	200,0	54,0	14,08	0,324
Ferritina	440,0	67,1	12,61	0,224
Tiroglobulina	669,0	86,0	11,11	0,122
Blue dextran	20.000	—	9,32	—

Fonte: Elaborada pelo autor.

Por meio dos dados apresentados na Tabela 9, construiu-se os gráficos de Calibração da Coluna para Massa Molar e para Raio hidrodinâmico (Figura 23).

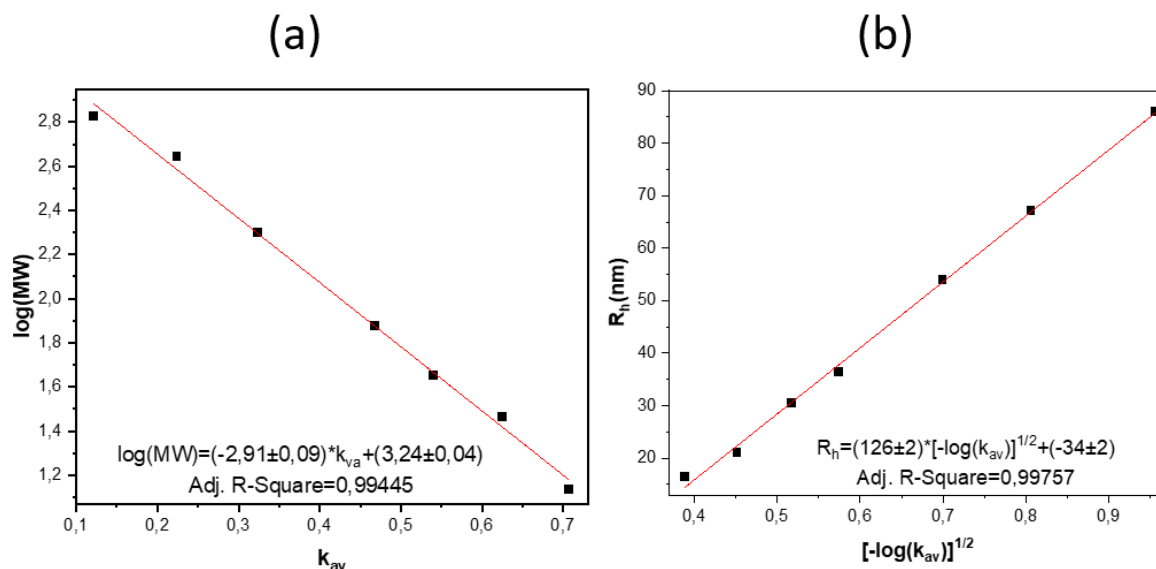


Figura 23 - Curvas de calibração da coluna Superdex™ 200 10/300 (GE) analítica de 24 mL, (a) calibração para massa molecular. (b) calibração para raio hidrodinâmico.

Fonte: Elaborada pelo autor.

Pela análise das curvas de calibração, estimou-se que a proteína de trabalho possuía massa molecular de (40 ± 8) kDa, indicando um desvio relativo de 20% comparado ao teórico de TcORC1/CDC6, e raio hidrodinâmico de $(2,9 \pm 0,3)$ nm, indicando um erro relativo de 4% comparado ao teórico. O desvio relativo alto para a massa poderia ser causado por conta do grau de precisão do método utilizado.

4.3.2 Espalhamento Dinâmico da Luz (DLS)

Pela análise de Espalhamento Dinâmico da Luz (Figura 24), observou-se correspondente a 100% em massa um único pico monodisperso com massa molecular de (47 ± 3) kDa, desvio relativo de 6% ao teórico de TcORC1/CDC6. Em relação ao raio hidrodinâmico, obteve $(6,2 \pm 0,7)$ nm, um desvio muito grande do teórico. Talvez isso esteja associado ao baixo sinal de intensidade do pico (14,8%). Vale ressaltar que o DLS também mostra que há formação de agregados na amostra, sendo estes os picos com maior intensidade.

Distribution Results

	Mean \pm SD (nm)	%Pd	Est. MW (KDa) (Mean \pm SD)*	% Intensity	% Mass	Peak Polydispersity
Peak 1:	6,169 \pm 0,6614	10,7	47,0 \pm 3,0	14,8	100,0	Monodisperse
Peak 2:	133,8 \pm 10,87	8,1	6,29e+4 \pm 5,11e+3	39,2	0,0	Monodisperse
Peak 3:	560,6 \pm 35,83	6,4	1,80e+6 \pm 1,92e+5	46,0	0,0	Monodisperse

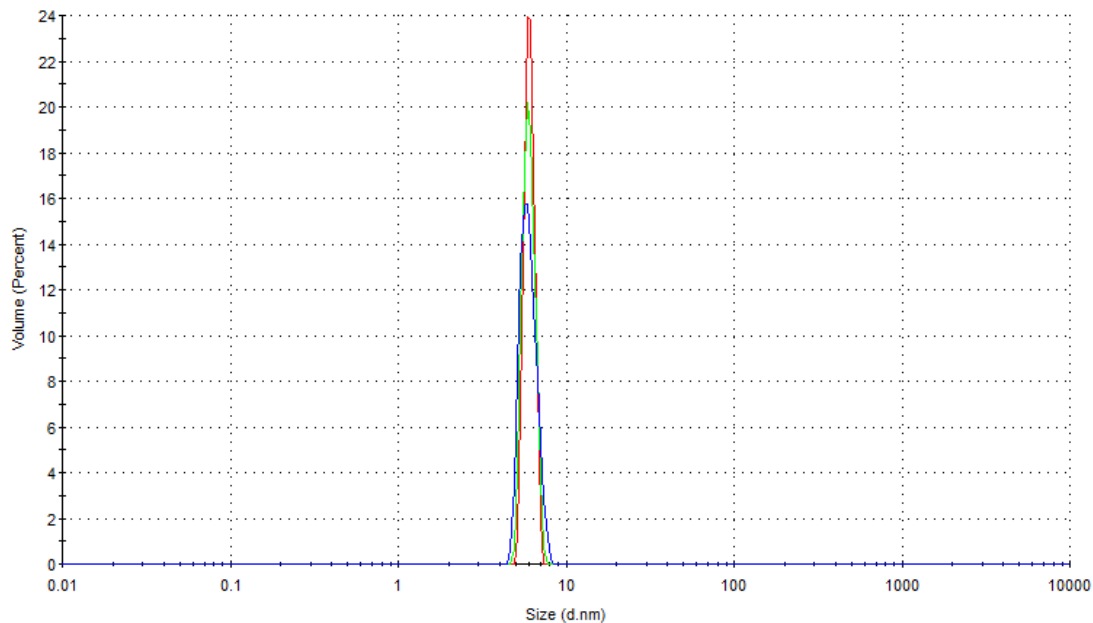


Figura 24 - Resultado do Espalhamento Dinâmico da Luz de TcORC1/CDC6.

Fonte: Elaborada pelo autor.

4.3.3 Fluorimetria de Varredura Diferencial (DSF)

4.3.3.1 Teste de concentração

Das condições testadas, a que apresentou uma curva de desnaturação mais homogênea e com melhor relação sinal/ruído foi a de 1 μ M de proteína e 25 x de Sypro™ Orange. Pela derivação da curva, foi possível calcular o ponto de inflexão da curva correspondente ao T_m. Este medido em 57 °C. A Figura 25 apresenta a curva de desnaturação e sua respectiva derivação.

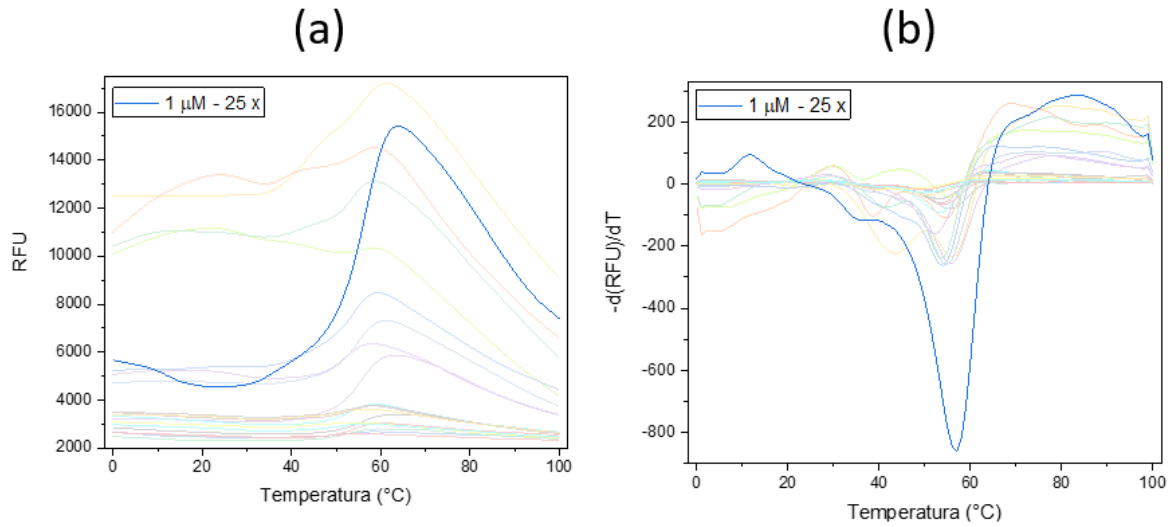


Figura 25 - Desnaturação térmica de TcORC1/CDC6 na presença de Sypro Orange. Em destaque, está a condição escolhida. (a) gráfico da desnaturação em função da temperatura. (b) primeira derivada da curva de desnaturação em função da temperatura.

Fonte: Elaborada pelo autor.

4.3.3.2 *Screen* de tampão

Como o T_m calculado em 4.3.3.1 foi de 57 °C, condições interessantes de tampão seriam as que apresentassem T_m acima desse valor. Nos gráficos da Figura 26, estão apresentados os T_m obtidos para as 96 condições testadas, sendo o zero correspondente ao ponto de T_m 57 °C.

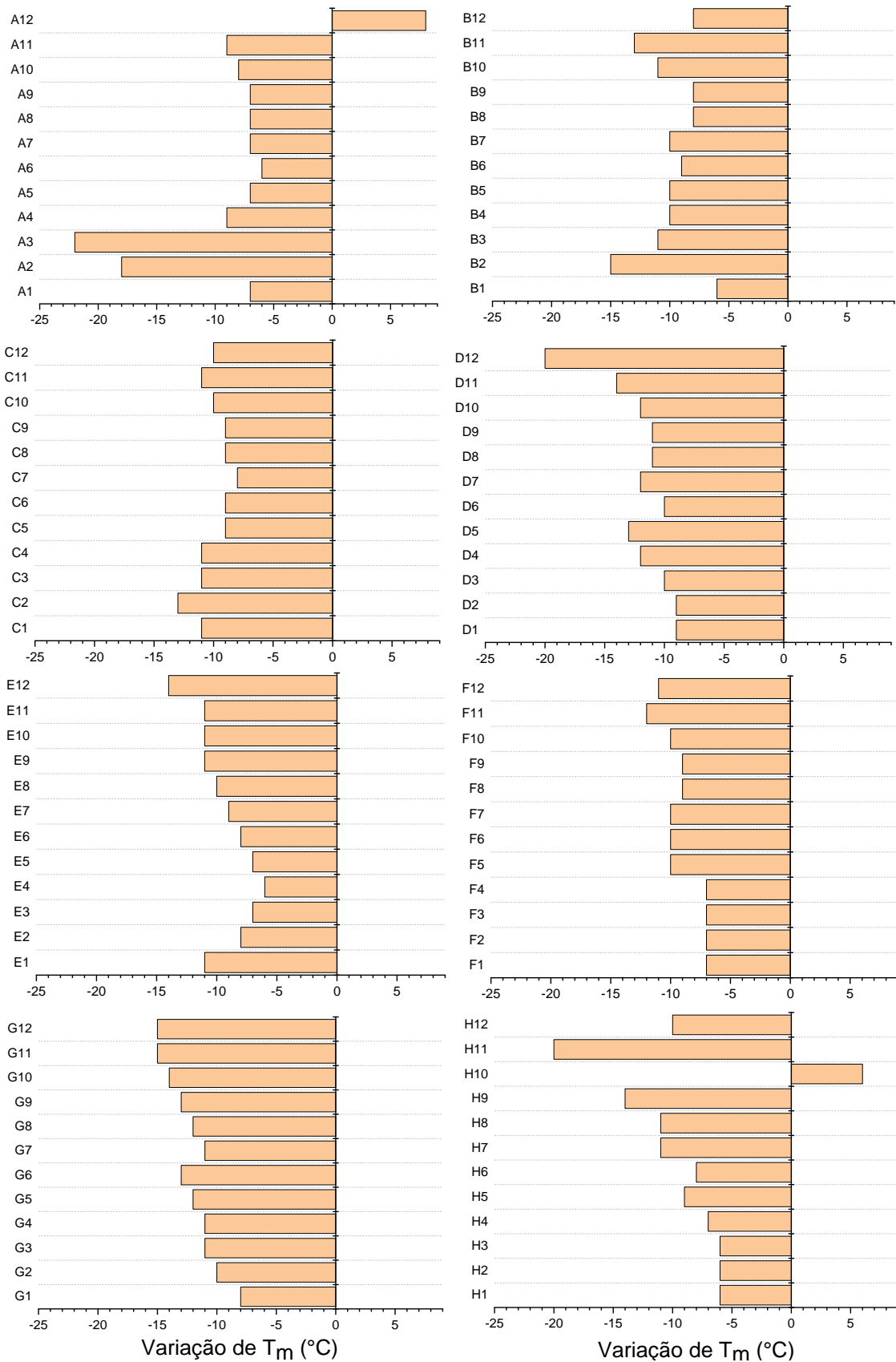


Figura 26 - Histogramas comparativos da variação do T_m nas diferentes condições de teste.

Fonte: Elaborada pelo autor.

Pela análise da Figura 26, percebe-se que há apenas duas condições cujo T_m foi maior. No entanto, ao observar as curvas de desnaturação dessas condições (Figura 27), percebe-se que a desnaturação não é homogênea e, por esse motivo, optou-se por continuar trabalhando com a proteína na condição tamponante inicial.

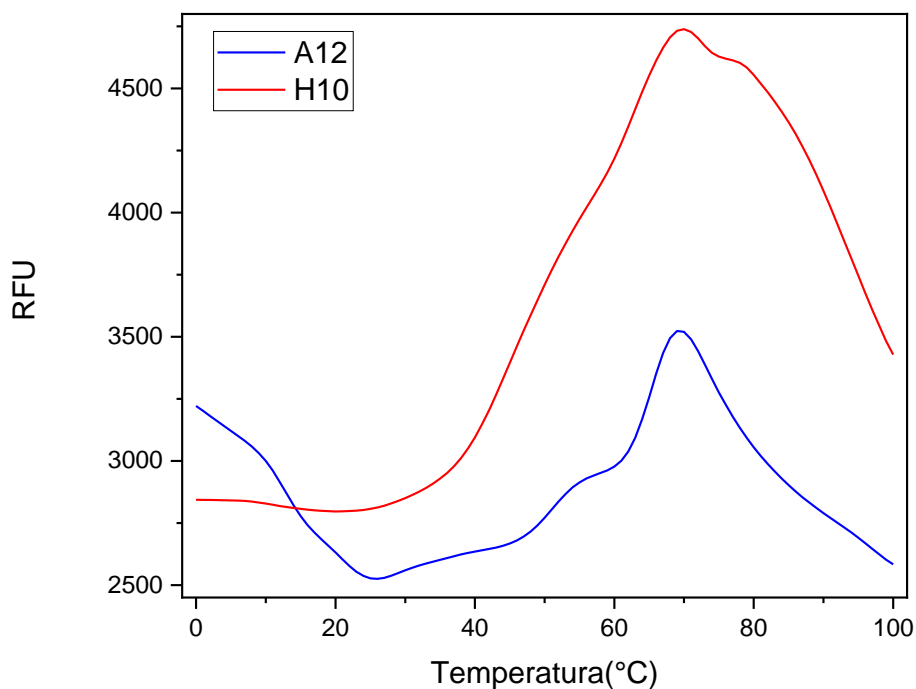


Figura 27 - Curva de desnaturação das condições do teste de tampão cuja variação relativa de T_m foi positiva

Fonte: Elaborada pelo autor.

4.3.4 Dicroísmo Circular (CD)

Por mais que a concentração de proteína para a análise tenha sido alta ($0,8 \text{ mg mL}^{-1}$ em 50 mM de Tris-HCl pH 7,4), o espectro de Dicroísmo Circular obtido ficou muito ruidoso, impossibilitando a visualização de mínimos em comprimentos específicos e, conseqüente, indicação do conteúdo de estrutura secundária da proteína. Por mais que a análise final tenha sido realizada com proteína fresca e em solução sem NaCl, o ruído de manteve. Seria interessante se repetir a análises em um tampão que não possua íons Cl^- livres.

Este resultado se apresenta na Figura 28.

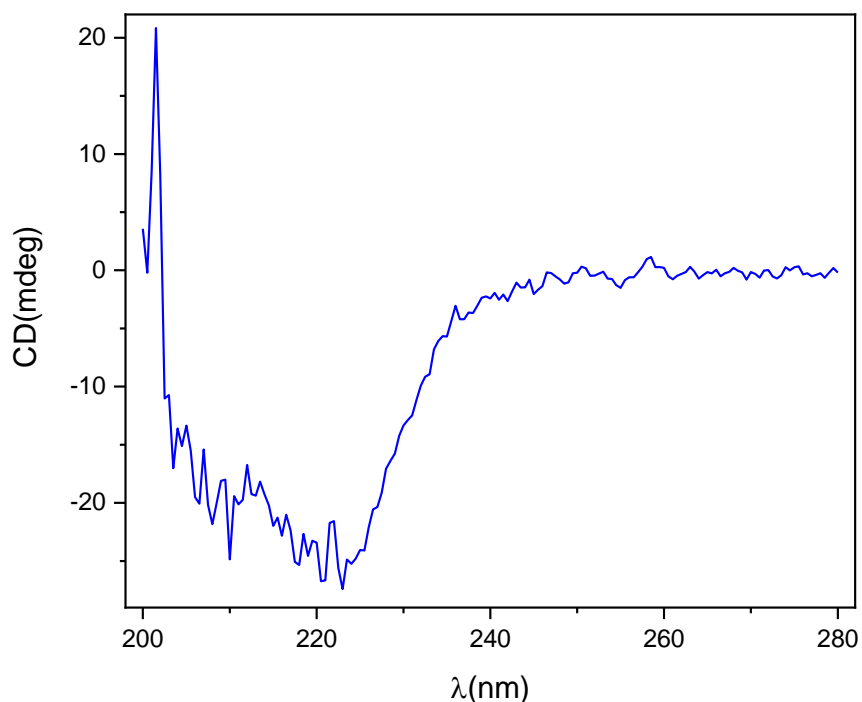


Figura 28 - Espectro de CD de TcORC1/CDC6.

Fonte: Elaborada pelo autor.

No entanto, foi possível medir a termoestabilidade da proteína pelos espectros de Dicroísmo circular em diferentes temperaturas e é visível uma transição de estados na proteína. Ao se derivar a curva, obtém-se o $T_m = 58 \pm 2 \text{ } ^\circ\text{C}$, que vai ao encontro do resultado dado pela técnica de DSF de $57 \pm 1 \text{ } ^\circ\text{C}$. Este resultado encontra-se na Figura 29.

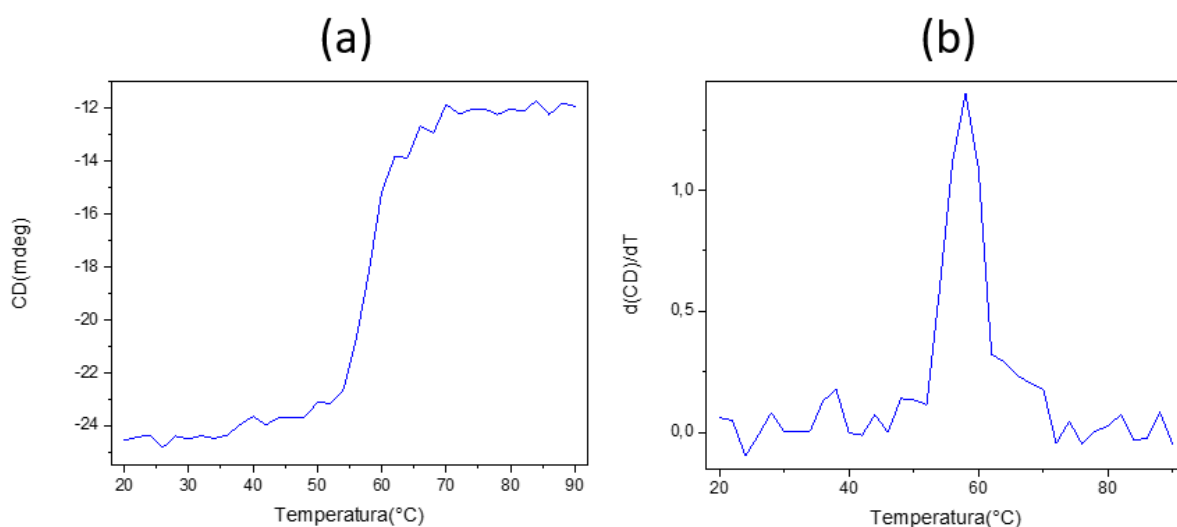


Figura 29 – Espectro de variação estrutural por temperatura pelo sinal a 222 nm. (a) espectro de sinal de CD por temperatura. (b) primeira derivada da curva de desnaturação pela temperatura.

Fonte: Elaborada pelo autor.

4.3.5 Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho (SEC-MAL)

Ao se submeter a amostra ao Espalhamento de Luz em Multi-Ângulo para Massa Molar e Tamanho, obteve o cromatograma apresentado na Figura 30.

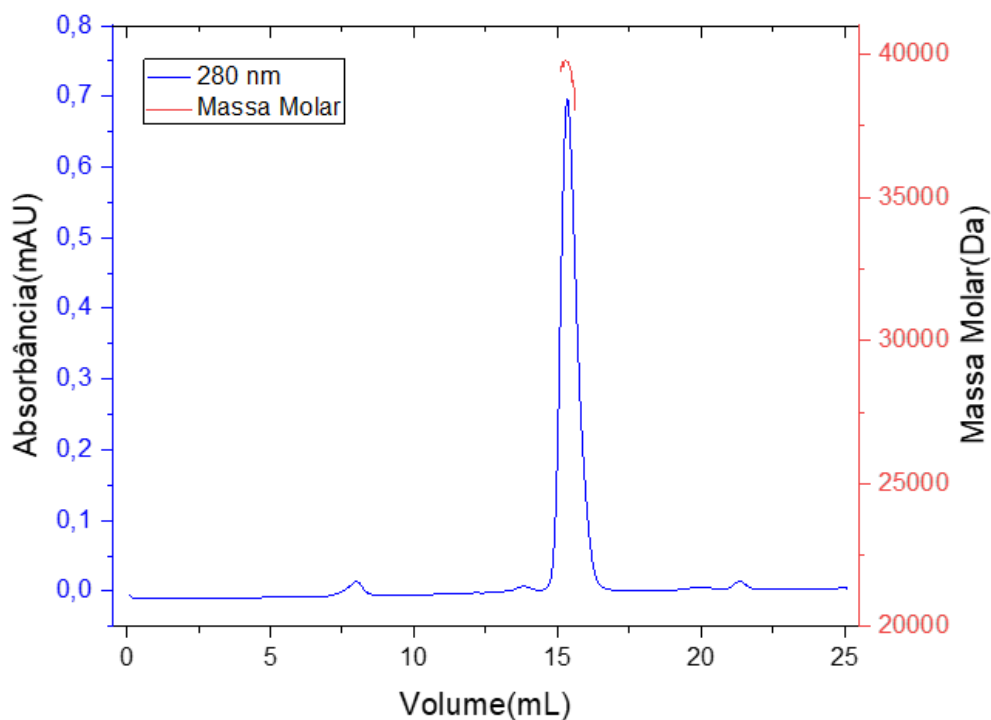


Figura 30 - Cromatograma resultado da análise por SEC-MALS de TcORC1/CDC6.

Fonte: Elaborada pelo autor.

A análise apontou uma proteína com massa de (40 ± 2) kDa, desvio relativo de 20% comparado ao teórico de TcORC1/CDC6, e raio hidrodinâmico de (11 ± 1) nm, desvio relativo ainda maior do que o obtido no DLS. Isso novamente talvez seja explicado por conta de sinal. Houve muita perda durante o preparo da amostra para essa técnica, e o sinal de absorbância demonstra que a intensidade do sinal captado realmente foi bem baixa.

4.3.6 Espectrometria de massas

A Espectrometria de Massas indicou que a banda com a qual se optou por prosseguir com as análises não se tratava de TcORC1/CDC6, e sim MBP, a cauda com a qual a proteína estava sendo expressa. Então, provavelmente houve algum problema com o anticorpo que marcou inicialmente a banda como TcORC1/CDC6 e as análises seguintes de massa e raio faziam referência a MBP.

Enquanto estava sendo aguardado o resultado dessa análise que as caracterizações biofísicas foram analisadas. Por isso, foram realizadas as diversas análises biofísicas descritas anteriormente.

5 CONCLUSÃO E PERSPECTIVAS

A predição da estrutura de TcORC1/CDC6 com diferentes métodos computacionais possibilitou análises comparativas que auxiliaram na validação dos modelos gerados na ausência de uma estrutura resolvida homóloga a do estudo com alta identidade. Como resultado, obteve-se um modelo com *score* significativo a homólogos depositados no PDB, sendo um ótimo indício de sua qualidade.

O estudo filogenético de TcORC1/CDC6 foi capaz de traçar indícios evolutivos centrados em protozoários com real sentido biológico e reforçou a importância do motivo AAA para a atividade da proteína.

A análise biofísica da proteína gerada no trabalho permitiu-se a criação da Tabela 10 com os parâmetros obtidos por diferentes técnicas.

Tabela 10 - Parâmetros biofísicos obtidos por diferentes técnicas para a proteína purificada no trabalho.

Técnica	Massa Molar (kDa)	Raio Hidrodinâmico (nm)	T_m (°C)
SEC	40 ± 8	2,9 ± 0,3	—
DLS	47 ± 3	6,2 ± 0,7	—
SEC-MALS	40 ± 2	11 ± 1	—
DSF	—	—	57 ± 1
CD	—	—	58 ± 2

Fonte: Elaborada pelo autor.

Pela Tabela 10, nota-se que todos os valores obtidos para massa molar são equivalentes entre si ao se levar em conta os erros associados, assim como os de T_m. No entanto, há uma grande diferença relacionada ao raio hidrodinâmico. Mais análises seriam necessárias para entender o motivo dessa diferença e seria importante repetir os experimentos para se ter sinais mais altos e, conseqüentemente, melhor resolução das medidas.

Na expressão, purificação e caracterização da proteína, foram encontrados desafios que impossibilitaram a obtenção de resultados para complementar o trabalho biocomputacional realizado de TcORC1/CDC6, já que a proteína que foi purificada e caracterizada foi MBP. Seria necessária uma revisão sobre o anticorpo utilizado para entender o motivo de ter interagido com MBP, sobre o processo de expressão e purificação da proteína para melhorar sua estabilidade e ser possível separá-la de forma eficiente.

REFERÊNCIAS

- 1 BHATTACHARYA, A. *et al.* Of Drugs and Trypanosomatids: new tools and knowledge to reduce bottlenecks in drugs discovery. *Genes (Basel)*, v. 11, n. 7, p. 722, July 2020.
- 2 DESQUESNES, M. *et al.* A review on the diagnosis of animal trypanosomoses. *Parasitic Vectors*, v. 15, n. 1, p. 64, Feb. 2022
- 3 MOLYNEUX, D.H. *et al.* The history of the neglected tropical disease movement. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, v. 115, n. 2, p. 169-175, Feb. 2021.
- 4 CHAGAS, C. Nova tripanozomíaze humana: estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen, n sp, agente etiológico de nova entidade morbida do homem. *Memórias do Instituto Oswaldo Cruz*, v. 1, p. 159–218, Aug. 1909.
- 5 PÉREZ-MAZILIAH, D.; WARD, A.I.; LEWIS, M.D. Host-parasite dynamics in Chagas disease from systemic to hyper-local scales. *Parasite Immunology*, v. 43, n. 2, p. e12786, Feb. 2021.
- 6 DRUGS FOR NEGLECTED DISEASES INITIATIVE. *Diseases & projects - Chagas disease*. Disponível em: <https://www.dndi.org/diseases-projects/chagas/>. Acesso em 23 mar. 2023.
- 7 CENTERS FOR DISEASE CONTROL AND PREVENTION. *DPDx - laboratory identification of parasites of public health concern. American Trypanosomiasis [Trypanosoma cruzi]*. Disponível em: <https://www.cdc.gov/dpdx/trypanosomiasisamerican/>. Acesso em: 24 mar. 2023.
- 8 LIDANI, K.C.F. *et al.* Chagas disease: from discovery to a worldwide health problem. *Frontiers in Public Health*, v. 7, p. 166, July 2019.
- 9 CASTRO, J. A.; MECCA, M. M.; BARTEL, L.C. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Human and Experimental Toxicology*, v. 25, n. 8, p. 471-479, Aug. 2006.
- 10 FERREIRA, A.M. *et al.* Reações adversas ao benzonidazol no tratamento da Doença de Chagas: revisão sistemática de ensaios clínicos randomizados e controlados. *Cadernos Saúde Coletiva*, v. 27, n.3, p. 354-362, July 2019.
- 11 DUNCKER, B. P.; CHESNOKOV, I. N.; McCONKEY, B. J. The origin recognition complex protein Family. *Genome Biology*, v. 10, n. 3, p. 214, Mar. 2009.
- 12 YEELES, J.T.P. *et al.* Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*, v. 519, n. 7544, p. 431-435, Mar. 2015.
- 13 LIMA, L.P. *et al.* Ortholog of the polymerase theta helicase domain modulates DNA replication in *Trypanosoma cruzi*. *Scientific Reports*, v. 9, n. 1, p. 2888, Feb. 2019.

- 14 SUN, J.; KONG, D. DNA replication origins, ORC/DNA interaction, and assembly of pre-replication complex in eukaryotes. *Acta Biochimica et Biophysica Sinica* (Shanghai), v. 42, n. 7, p. 433-439, July 2010.
- 15 EL-SAYED, N.M. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, v. 309, n. 5733, p. 409-415, July 2005.
- 16 MARQUES, C.A. *et al.* Diverged composition and regulation of the *Trypanosoma brucei* origin recognition complex that mediates DNA replication initiation. *Nucleic Acids Research*, v. 44, n. 10, p. 4763-4784, June 2016.
- 17 DAMASCENO, J.D. *et al.* Read, write, adapt: challenges and opportunities during kinetoplastid genome replication. *Trends in Genetics*, v. 37, n. 1, p. 21-34, Jan. 2021.
- 18 GASTEIGER, E. *et al.* Protein identification and analysis tools on the expasy server. In: WALKER, J. M. (ed): *The proteomics protocols handbook*. Totowa: Humana Press, 2005. p. 571-607.
- 19 WILKINS, D.K. *et al.* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, v. 38, n. 50, p. 16424-43431, Dec. 1999.
- 20 FERRÈ, F.; CLOTE, P. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Research*, v. 33, p. W230–W232, July 2005.
- 21 FERRÈ, F.; CLOTE, P. DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Research*, v. 34, p. W182–W185, July 2006.
- 22 FINN, R.D. *et al.* Pfam: the protein families database. *Nucleic Acids Research*, v. 42, Jan. 2014. DOI: 10.1093/nar/gkt1223.
- 23 LU, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research*, v. 48, n. D1, p. 265-268, Jan. 2020.
- 24 BLEICHERT, F.; BOTCHAN, M. R.; BERGER, J. M. Crystal structure of the eukaryotic origin recognition complex. *Nature*, v. 519, n. 7543, p. 321-326, Mar. 2015.
- 25 ESWAR, N. *et al.* Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, Oct. 2006. DOI: 10.1002/0471250953.bi0506s15.
- 26 ZHANG, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, v. 23, p. 9-40, Jan. 2008.
- 27 KIM, D.E.; CHIVIAN, D.; BAKER, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, v. 32, p. W526–W531, July 2004.
- 28 JUMPER, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*, v. 596, n. 7873, p. 583-589, Aug. 2021.

- 29 UCLA-DOE LAB - Saves v6.0. *To run any or all programs*: upload your structure, in PDB format only. Disponível em: <https://saves.mbi.ucla.edu/>. Acesso em 10 mar. 2023.
- 30 SCHRÖDINGER, L.; DELANO, W. *PyMOL*. 2020. Disponível em: <http://www.pymol.org/pymol>. Acesso em: 23 jan. 2022
- 31 HOLM, L. Dali server: structural unification of protein families. *Nucleic Acids Research*, v. 50, n. W1, p. W210-W215, May 2022.
- 32 HOLM, L *et al.* Searching protein structure databases with DaliLite v.3. *Bioinformatics*, v. 24, n. 23, p. 2780-2781, Dec. 2008.
- 33 MADDEN, T. The BLAST sequence analysis tool. outubro de 2002 [atualizado em agosto de 2003]. In: McENTYRE, J.; OSTELL, J. (ed.). *The NCBI handbook*. Bethesda: National Center for Biotechnology Information, 2002. Capítulo 16.
- 34 CASTRESANA, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, v. 17, n. 4, p. 540–552, Apr. 2000.
- 35 TALAVERA, G.; CASTRESANA, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, v. 56, n. 4, p. 564-577, Aug. 2007.
- 36 JOY, J. B. *et al.* Ancestral reconstruction. *PLOS Computational Biology*, v. 12, n. 7, p. e1004763, 2016.
- 37 GIRALDO, R. Common domains in the initiators of DNA replication in bacteria, archaea and eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiology Reviews*, v. 26, p. 5, p. 533-554, Jan. 2003.
- 38 BOWERS, J. L. *et al.* ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Molecular Cell*, v. 16, n. 6, p. 967-978, Dec. 2004.
- 39 WU, S.; ZHANG, Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Research*, v. 35, n. 10, p. 3375-3382, May 2007.
- 40 ZHANG, Y; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. *Proteins*, v. 57, n. 4, p. 702-710, Dec. 2004.
- 41 OCAÑA-PALLARÈS, E. *et al.* Origin recognition complex (ORC) evolution is influenced by global gene duplication/loss patterns in eukaryotic genomes. *Genome Biology and Evolution*, v. 12, n. 2, p. 3878–3889, Feb. 2020.
- 42 BRADWELL, K. R. *et al.* Genomic comparison of *Trypanosoma conorhini* and *Trypanosoma Rangeli* to *Trypanosoma cruzi* strains of high and low virulence. *BMC Genomics*, v. 19, n. 1, p. 770, Oct. 2018.

- 43 BROTKOVÁ, A. *et al.* Trypanosomes of the *Trypanosoma theileri* group: phylogeny and new potential vectors. *Microorganisms*, v. 10, n. 2, p. 294, Jan. 2022.
- 44 KELLY, S. *et al.* A draft genome for the African crocodylian trypanosome *Trypanosoma grayi*. *Scientific Data*, v. 1, n. 14, p. 140024, Aug. 2014.
- 45 CHANMOL, W. *et al.* Development of *Leishmania orientalis* in the sand fly *Lutzomyia longipalpis* (Diptera: Psychodidae) and the biting midge *Culicoides soronensis* (Diptera: Ceratopogonidae). *Acta Tropica*, v. 199, n. 10, p. 105157, Nov. 2019.
- 46 ESPINOSA, O. A. *et al.* An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as *Leishmania* and *Endotrypanum*. *Parasitology*, v. 145, n. 4, p. 430-442, Apr. 2018.
- 47 FLEGONTOV, P. *et al.* Genome of *Leptomonas pyrrocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Scientific Reports*, v. 6, p. 23704, Mar. 2016.
- 48 BRUM, F.L. *et al.* Structural characterization of the cell division cycle in *Strigomonas culicis*, an endosymbiont-bearing trypanosomatid. *Microscopy and Microanalysis*, v. 20, n. 1, p. 228-237, Feb. 2014.

ANEXOS

ANEXO A - SCRIPT PYTHON PARA MODELAGEM

```
from modeller import *
from modeller.automodel import *

# Criação de um novo ambiente do modeller
env = environ()
# Definição de parâmetros físicos
env.schedule_scale = physical.values(default=1.0, soft_sphere=0.7)
env.io.hetatm = True

# Geração do modelo inicial
a = automodel(env, alnfile='orc1.pir',
              knowns=('best'), sequence='TcORC1',
              assess_methods=(assess.DOPE, assess.GA341))
a.starting_model = 1
a.ending_model = 100
a.deviation = 4.0

# Grau de randomização entre modelos
a.generate_method = generate.transfer_xyz
a.rand_method = randomize.xyz

# Otimização pelo método da função alvo variável
a.library_schedule = autosched.slow
a.max_var_iterations = 300

# Otimização por dinâmica molecular
a.md_level = refine.very_slow
a.final_malign3d = True
a.make()

# Gerando a lista de todos os modelos construídos com sucesso
ok_models = filter(lambda x: x['failure'] is None, a.outputs)

# Rankeando os modelos pelo score DOPE
key = 'DOPE score'
ok_models.sort(lambda a,b: cmp(a[key], b[key]))
# Definição do melhor modelo
m = ok_models[0]
```

```
print "Top model: %s (DOPE score %.3f)" % (m['name'], m[key])
```


ANEXO B - SEQUÊNCIAS ANCESTRAIS

> ORC1/CDC6_Tryp_Leish

ALRKGVAALSVSLVCRDEHVKAIIDFLNDNAHHTMQIFGMPGTGKTATVNYALALLA
 VFLNGYVVQKSSDIYWTLYSHLTKARLGTVENCPPDQCAANIEKRFRHGWVIIIDEV
 DKILKKHSAKALFKIVDWLTLPANCKLITISNSMELQLVFSPYNTQELKEILLHRVGPPLF
 ADQAINLLCHQTASHYGDVRRLLQICGVLMIQDGVKEIHAVVRQIFHDRFVEFIKTIR
 KPVLFIKAVLAKETEELFKRNASDCRELLRQVSLIDVSVGEERIPPLESTEDVYVSLQ
 P

> ORC1/CDC6_Leish_Lepto

ALKKGVAALSVSLVCREEHAKAIQDFLEDDKHHTMQIFGMPGTGKTATVNYALAQL
 AVFLNGFVVQKSSDIYYTLHHHLTKARLGTVEQCPVAQCASNIEKRFRHGWVIIIDEV
 DKILEKHAKGLFKVVDWLTLPANCKLITISNSMELQLVFSSYGTQELREILLHRVGPPL
 FADQAVNQLCTQTASHYGDVRRLLQICSVLMKIQDGLREIHAVVRQIFHDRFVEFITR
 LRKPVLFIKAVLKGKETEELFKRREVDCRELLRQVSLIDVSVGEDRIPLLESTEDVYVSL
 LQP

> ORC1/CDC6_Leish_Porc

ALKEGVQALSVSLVCREDHARAIQDFLEDDKHHTMQIFGMPGTGKTATVNFALAQL
 AVFLNGFVVQKSSDIYYTLHHHLTKARLGVVEQCPVAQCASNIEKRFRHGWVIVVDE
 VDKILEKHAKGLFKVVDWLTLPANCKLITISNSMELQLVFSSYGTQELREILLHRVGP
 KLFADQAVNQLCTQTASHYGDVRRLLQICSVLMRIQDGLREIHSVVRQIFHDRFVEFI
 TMRKPVLFIKAVLKGKETEELIRKREVDCRELLRQVSLIEVCVGEDRIPLLDSTEDVYV
 LLLQP