

UNIVERSIDADE FEDERAL DE UBERLÂNDIA – UFU  
INSTITUTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA – PPGQUI

**William Oliveira Soté**

**Estudos *in silico* da proteína TcP21  
no contexto da doença de Chagas**

Uberlândia  
2026

**William Oliveira Soté**

**Estudos *in silico* da proteína TcP21  
no contexto da doença de Chagas**

Tese de Doutorado apresentada ao Programa de Pós-graduação em Química do Instituto de Química da Universidade Federal de Uberlândia, como parte dos requisitos para a obtenção do título de Doutor em Química.

*Área de concentração:* Físico-química.

*Orientador:* Prof. Dr. Moacyr Comar Junior.

Uberlândia

2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

S717 Soté, William Oliveira, 1995-  
2026 Estudos in silico da proteína TcP21 no contexto da doença de Chagas [recurso eletrônico] / William Oliveira Soté. - 2026.

Orientador: Moacyr Comar Junior.

Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Química.

Modo de acesso: Internet.

DOI <http://doi.org/10.14393/ufu.te.2026.378>

Inclui bibliografia.

Inclui ilustrações.

1. Química. I. Comar Junior, Moacyr ,1974-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Química. III. Título.

CDU: 54

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

Nelson Marcos Ferreira - CRB6/3074



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
 Coordenação do Programa de Pós-Graduação em Química  
 Av. João Naves de Ávila, 2121, Bloco 5I - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902  
 Telefone: (34) 3239-4385 - www.cpgquimica.iq.ufu.br - cpgquimica@ufu.br



### ATA

|                                    |                                                                                                              |                 |    |                       |       |
|------------------------------------|--------------------------------------------------------------------------------------------------------------|-----------------|----|-----------------------|-------|
| Programa de Pós-Graduação em:      | Química                                                                                                      |                 |    |                       |       |
| Defesa de:                         | Tese de Doutorado Acadêmico, 186, PPGQUI                                                                     |                 |    |                       |       |
| Data:                              | Vinte e sete de julho de dois mil e vinte e seis                                                             | Hora de início: | 8h | Hora de encerramento: | 12:00 |
| Matrícula do Discente:             | 12223QMI006                                                                                                  |                 |    |                       |       |
| Nome do Discente:                  | William Oliveira Soté                                                                                        |                 |    |                       |       |
| Título do Trabalho:                | "Estudo in silico da proteína TcP21 no contexto da doença de Chagas"                                         |                 |    |                       |       |
| Área de concentração:              | Química                                                                                                      |                 |    |                       |       |
| Linha de pesquisa:                 | Química Teórica e Físico-Química de Macromoléculas e Coloides                                                |                 |    |                       |       |
| Projeto de Pesquisa de vinculação: | Análise in silico de sistemas de detecção de compostos bioquímicos                                           |                 |    |                       |       |
| <u>ODS 3</u>                       | ODS 3 - Saúde e bem-estar: assegurar uma vida saudável e promover o bemestar para todos, em todas as idades. |                 |    |                       |       |

Reuniu-se, por webconferência, no *link*: <https://conferenciaweb.rnp.br/sala/wosote>, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Química, assim composta: Professores Doutores: **Francisco das Chagas Alves Lima**, da Universidade Estadual do Piauí-UESPI; **Teodorico de Castro Ramalho**, da Universidade Federal de Lavras-UFLA; **Karen Cacilda Weber**, Universidade Federal da Paraíba-UFPB; **Wagner Fernando Delfino Angelotti**, Universidade Federal do Triângulo Mineiro-UFTM, e **Moacyr Comar Júnior**, orientador(a) do candidato(a).

Iniciando os trabalhos o(a) presidente da mesa, **Dr. Moacyr Comar Júnior**, apresentou a Comissão Examinadora e o(a) candidato(a), agradeceu a presença do público, e concedeu ao(à) Discente a palavra para a exposição do seu trabalho. A duração da apresentação do(a) Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.



O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.



Documento assinado eletronicamente por **Moacyr Comar Junior, Professor(a) do Magistério Superior**, em 28/07/2026, às 08:35, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Wagner Fernando Delfino Angelotti, Usuário Externo**, em 28/07/2026, às 08:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Francisco das Chagas Alves Lima, Usuário Externo**, em 28/07/2026, às 09:08, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Karen Cacilda Weber, Usuário Externo**, em 30/07/2026, às 15:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Teodorico de Castro Ramalho, Usuário Externo**, em 03/08/2026, às 12:52, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **7344964** e o código CRC **B2A9F561**.

## Dedicatória

---

Dedico esta tese aos inúmeros mestres que tive ao longo da vida. Saibam que, se em algum momento duvidaram de sua contribuição para a ciência, vocês foram primordiais para que eu chegasse até aqui.

## Agradecimentos

---

Agradeço aos meus pais e irmã por todo o amor, incentivo e suporte ao longo dessa longa e misteriosa trajetória que é a jornada acadêmica. A segurança que sempre me ofereceram permitiu que eu me dedicasse integralmente à minha formação e buscasse alcançar o meu máximo potencial. Sou profundamente grato por todos os sacrifícios, pela compreensão e por acreditarem em mim.

Agradeço ao Prof. Dr. Moacyr Comar Junior pela dedicação, confiança e apoio constante ao longo de uma orientação que se iniciou no mestrado, em 2020, e que continuou por seis anos e meio. Sou profundamente grato por todas as oportunidades que me foram concedidas, pelas valiosas conversas e pelas inúmeras contribuições para a minha formação acadêmica, profissional e pessoal. Finalizo esta etapa importante da minha carreira levando comigo não apenas o conhecimento adquirido sob sua orientação, mas também a admiração e o sincero reconhecimento pelo tempo dedicado à minha formação.

Agradeço ao Instituto de Química da Universidade Federal de Uberlândia, unidade acadêmica da qual fiz parte ao longo dos últimos treze anos, trajetória iniciada durante minha graduação, em 2013. Registro meu sincero reconhecimento pelo compromisso com o ensino, pela infraestrutura e pelas oportunidades que tornaram possível essa jornada acadêmica. Estendo também meu agradecimento aos inúmeros colegas com quem compartilhei esse ambiente acadêmico e com os quais pude aprender muito.

Agradeço à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro concedido por meio de bolsa de doutorado (Convênio nº 5.10/2022).

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido no doutorado-sanduíche, por meio do Programa de Doutorado Sanduíche no Exterior (PDSE) (Processo nº 88881.982886/2024-01).

Agradeço ao Prof. Dr. Rafael C. Bernardi, da Auburn University (EUA), em nome do *Computacional Biophysics Group*, pela receptividade e orientação durante a realização do doutorado-sanduíche. Agradeço também ao Dr. Diego E. B. Gomes pela generosidade genuína com que me ensinou tanto ao longo desse curto período.

*Uma vez eliminado o impossível, o que restar, por mais improvável que seja, deve ser a verdade.*

*– O Signo dos Quatro (1890)*

*Não vim aqui para lhe dizer como isso vai terminar.  
Vim para lhe dizer como vai começar.*

*– The Matrix (1999)*

## Resumo

---

A doença de Chagas, causada pelo *Trypanosoma cruzi*, permanece como um importante problema de saúde pública, com opções terapêuticas ainda limitadas. Nesse contexto, a proteína TcP21, secretada pelas formas amastigotas extracelulares, configura-se como um alvo molecular de interesse, associado a vias alternativas de infecção pouco exploradas. A presente tese teve como objetivo caracterizar a TcP21 *in silico* sob duas perspectivas complementares: a identificação e caracterização de sítios de ligação em sua superfície e a elucidação de possíveis interfaces de interação no complexo TcP21–CXCR4. A integração de simulações de dinâmica molecular em solventes mistos, ancoragem molecular baseada em fragmentos e ancoragem molecular de sonda farmacofórica seguida de dinâmica molecular, conduzida por meio de mapas volumétricos de ocupação, permitiu a identificação convergente de um sítio de ligação na TcP21. Esse sítio foi caracterizado pela participação dos resíduos-chave GLU55, ASP52, VAL70, ILE62 e TRP77. As interações observadas, incluindo pontes salinas, ligações de hidrogênio, interações carga-carga e contatos  $\pi$ -alquil, configuraram uma rede intermolecular consistente com o reconhecimento de grupos funcionais amino protonado, acetamido e fenil. Modos de ligação do complexo TcP21–CXCR4 foram gerados por ancoragem molecular e subsequentemente refinados por simulações de dinâmica molecular multirréplicas em ambiente biomimético. A análise de consenso das trajetórias, baseada em uma função de pontuação multimétrica fundamentada em  $\Delta$ SASA e na análise de contatos intermoleculares (número, frequência e estabilidade temporal), revelou um segmento contínuo na TcP21 (LYS23–PHE39) com propriedades típicas de interfaces proteína–proteína. Na CXCR4, os domínios transmembrana TM5, TM6 e TM7 constituíram um núcleo estrutural estabilizado no interior do vestíbulo extracelular, enquanto regiões como ECL2 e N-terminal exibiram maior flexibilidade conformacional. A interface resultante foi caracterizada por uma organização modular, predominantemente estabilizada por interações de van der Waals, com contribuições eletrostáticas e polares conferindo especificidade direcional ao processo de reconhecimento molecular. A análise integrada dos estudos indicou a existência de duas regiões funcionalmente especializadas na TcP21, potencialmente associadas, respectivamente, à interação com ligantes farmacofóricos e ao reconhecimento do receptor do hospedeiro.

## Abstract

---

Chagas disease, caused by *Trypanosoma cruzi*, remains a significant public health concern, with still limited therapeutic options. In this context, the TcP21 protein, secreted by extracellular amastigote forms, has emerged as a molecular target of interest, involved in alternative and still poorly explored infection pathways. This thesis aimed to characterize TcP21 in silico from two complementary perspectives: the identification and characterization of binding sites on its surface, and the elucidation of potential interaction interfaces in the TcP21–CXCR4 complex. The integration of mixed-solvent molecular dynamics simulations, fragment-based molecular docking, and pharmacophore-based molecular docking followed by molecular dynamics, driven by occupancy-based volumetric maps, enabled the convergent identification of a binding site on TcP21. This site involved the key residues GLU55, ASP52, VAL70, ILE62, and TRP77. The observed interactions, including salt bridges, hydrogen bonds, charge–charge interactions, and  $\pi$ –alkyl contacts, formed an intermolecular network consistent with the recognition of protonated amino, acetamido, and phenyl functional groups. Binding modes of the TcP21–CXCR4 complex were generated by molecular docking and subsequently refined through multireplica molecular dynamics simulations in a biomimetic environment. Consensus analysis of the trajectories, based on a  $\Delta$ SASA-driven multimetric scoring function and on the evaluation of intermolecular contacts (number, frequency, and temporal stability), revealed a continuous segment in TcP21 (LYS23–PHE39) with typical protein–protein interface properties. In CXCR4, the transmembrane domains TM5, TM6, and TM7 formed a stabilized structural core within the extracellular vestibule, whereas regions such as ECL2 and the N-terminal domain exhibited greater conformational flexibility. The resulting interface displayed a modular organization, predominantly stabilized by van der Waals interactions, with additional electrostatic and polar contributions conferring directional specificity to the molecular recognition process. The integrated analysis of the studies indicated the presence of two functionally specialized regions in TcP21, potentially associated with pharmacophore ligand binding and host receptor recognition, respectively.

## Lista de Figuras

---

|                                                                                                                                                                                                                  |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1.1</b> – Estruturas químicas dos fármacos tripanossomicidas. De cima para baixo: nifurtimox e benznidazol. ....                                                                                       | 23 |
| <b>Figura 1.2</b> – Representação esquemática da proteína CXCR4, segundo distribuição dos aminoácidos e orientações das hélices transmembranares. ....                                                           | 28 |
| <b>Figura 2.1</b> – Fluxograma geral de trabalho do Estudo 1, dividido entre obtenção do modelo conformacional (A) e identificação de sítios de ligação (B). ....                                                | 36 |
| <b>Figura 2.2</b> – Modelos preditos da proteína TcP21 gerados por RoseTTAFold (vermelho) e AlphaFold 3 (azul). ....                                                                                             | 38 |
| <b>Figura 2.3</b> – Solventes orgânicos utilizados como sondas moleculares nas simulações de MSMD. ....                                                                                                          | 41 |
| <b>Figura 2.4</b> – Representação esquemática da configuração dos sistemas. O solvente orgânico é mostrado em malva e as moléculas de água foram omitidas para maior clareza. ....                               | 42 |
| <b>Figura 2.5</b> – Representação do modelo da sonda farmacofórica utilizado no estudo. ....                                                                                                                     | 44 |
| <b>Figura 2.6</b> – Pontuações 3D-1D por resíduo para os modelos gerados por AlphaFold 3 e RoseTTAFold. ....                                                                                                     | 48 |
| <b>Figura 2.7</b> – Perfil de erro por resíduo (ERRAT) para os modelos gerados por AlphaFold 3 e RoseTTAFold, indicando regiões normais (branco), de alerta (amarelo) e de rejeição estatística (vermelho). .... | 49 |
| <b>Figura 2.8</b> – Alinhamentos estruturais parciais dos modelos gerados por RoseTTAFold (vermelho) e AlphaFold 3 (azul). Da esquerda para a direita: resíduos 58–154 e 1–20. ....                              | 51 |
| <b>Figura 2.9</b> – Alinhamento estrutural dos modelos gerados por RoseTTAFold. Resíduos não alinhados são mostrados de forma transparente. Da esquerda para a direita: resíduos 57–154 e 1–57. ....             | 52 |
| <b>Figura 2.10</b> – Perfil de erro estrutural estimado dos resíduos preditos para os modelos gerados por RoseTTAFold. ....                                                                                      | 53 |
| <b>Figura 2.11</b> – Representação esquemática do modelo 2 gerado por RoseTTAFold. Da esquerda para a direita: vistas frontal e posterior. ....                                                                  | 54 |
| <b>Figura 2.12</b> – Perfil temporal médio de RMSD da cadeia principal da TcP21. ....                                                                                                                            | 55 |

|                                                                                                                                                                                                                                                                                                                          |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 2.13</b> – Perfil de RMSF dos resíduos da cadeia principal da TcP21.....                                                                                                                                                                                                                                       | 56 |
| <b>Figura 2.14</b> – Representação esquemática dos modelos da TcP21. De cima para baixo: modelo 2 e estrutura representativa. Da esquerda para a direita: vistas frontal e posterior.....                                                                                                                                | 57 |
| <b>Figura 2.15</b> – Mapas volumétricos dos respectivos solventes orgânicos: acetamida (ACM), acetonitrila (ACN), íon acetato (ACT), benzeno (BNZ), etanol (ETH) e íon metilamônio (MAM).....                                                                                                                            | 58 |
| <b>Figura 2.16</b> – Representação das regiões de maior correlação entre os solventes orgânicos e sobreposição das conformações da TcP21 para cada réplica. ....                                                                                                                                                         | 59 |
| <b>Figura 2.17</b> – Distribuição percentual dos modos de ligação obtidos pelo AutoDock Vina com base na energia de afinidade.....                                                                                                                                                                                       | 60 |
| <b>Figura 2.18</b> – Representações esquemáticas dos modos de ligação. Em vermelho, modos com energia superior a $-5,0 \text{ kcal mol}^{-1}$ ; em azul, entre $-5,0$ e $-5,9 \text{ kcal mol}^{-1}$ ; e em roxo, inferior a $-5,9 \text{ kcal mol}^{-1}$ . Da esquerda para a direita: vistas frontal e posterior. .... | 61 |
| <b>Figura 2.19</b> – Modos de ligação de menor energia (a) e diagramas 2D das interações proteína-sonda farmacofórica (b). ....                                                                                                                                                                                          | 62 |
| <b>Figura 2.20</b> – Representação dos complexos proteína-sonda farmacofórica. Da esquerda para a direita: mapa volumétrico da sonda e superimposição das conformações dos complexos. ....                                                                                                                               | 63 |
| <b>Figura 2.21</b> – Representação das conformações centrais do complexo sonda farmacofórica-proteína para cada réplica e seus respectivos diagramas 2D de interação.....                                                                                                                                                | 64 |
| <b>Figura 2.22</b> – Sítios de ligação preditos pelo servidor FTSite, representados por malhas em cor (salmão, verde e azul). ....                                                                                                                                                                                       | 66 |
| <b>Figura 2.23</b> – Modelo de sítio de ligação proposto, representado como superfície de van der Waals. De cima para baixo: vistas frontal e lateral. ....                                                                                                                                                              | 69 |
| <b>Figura 3.1</b> – Fluxograma geral de trabalho do Estudo 2. ....                                                                                                                                                                                                                                                       | 72 |
| <b>Figura 3.2</b> – Representação dos resíduos considerados atrativos (verde) e repulsivos (cinza) durante a ancoragem molecular. ....                                                                                                                                                                                   | 74 |
| <b>Figura 3.3</b> – Representação do sistema TcP21–CXCR4 em membrana lipídica, com CXCR4 em vermelho e lipídios em cinza. Água e íons omitidos. ....                                                                                                                                                                     | 76 |
| <b>Figura 3.4</b> – Medóides dos clusters mais populosos obtidos por ancoragem molecular. ....                                                                                                                                                                                                                           | 83 |



|                                                                                                                                                                                                           |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figura 3.5</b> – Perfis temporais de RMSD da CXCR4.....                                                                                                                                                | 85  |
| <b>Figura 3.6</b> – Perfis temporais de RMSD da TcP21.....                                                                                                                                                | 87  |
| <b>Figura 3.7</b> – Perfis temporais do número de ligações de hidrogênio entre TcP21 e CXCR4.....                                                                                                         | 90  |
| <b>Figura 3.8</b> – Perfis temporais do número de pontes salinas entre TcP21 e CXCR4.....                                                                                                                 | 92  |
| <b>Figura 3.9</b> – Resíduos de interface mais prováveis para as proteínas CXCR4 e TcP21. ....                                                                                                            | 94  |
| <b>Figura 3.10</b> – Região de interface de interação da TcP21. Átomos C $\alpha$ dos resíduos de interface são mostrados como esferas. ....                                                              | 96  |
| <b>Figura 3.11</b> – Pontos de contato interfaciais da CXCR4. Átomos C $\alpha$ dos resíduos de interface são mostrados como esferas. Da esquerda para a direita: vistas frontal e superior. ....         | 98  |
| <b>Figura 3.12</b> – Interface de interação do complexo TcP21–CXCR4. Átomos C $\alpha$ dos resíduos de interface são mostrados como esferas. Da esquerda para a direita: $R_1(M_1 - M_4)$ .....           | 99  |
| <b>Figura 3.13</b> – Frequências de ocupação de ligações de hidrogênio dos resíduos da TcP21, por sistema. ....                                                                                           | 102 |
| <b>Figura 3.14</b> – Resíduos de maior número de contatos da proteína TcP21, por sistema e agrupado. ....                                                                                                 | 104 |
| <b>Figura 3.15</b> – Top 30 pares de resíduos entre TcP21 e CXCR4 de maior recorrência agrupada. ....                                                                                                     | 105 |
| <b>Figura 3.16</b> – Distribuição dos perfis temporais dos pares de resíduos entre TcP21 e CXCR4, por sistema e agrupado. ....                                                                            | 108 |
| <b>Figura 3.17</b> – Distribuição das forças intermoleculares de estabilização entre os pares de resíduos TcP21–CXCR4, por sistema e agrupado. ....                                                       | 110 |
| <b>Figura 3.18</b> – Complexo interativo TcP21–CXCR4. Em azul escuro: domínios TM5, TM6 e TM7. Em azul claro: domínios ECL2 e TM1/N-terminal. Da esquerda para a direita: vistas frontal e superior. .... | 112 |
| <b>Figura A.1</b> – Fluxograma geral da rotina de identificação de resíduos interfaciais. ....                                                                                                            | 138 |
| <b>Figura B.1</b> – Fluxograma geral da rotina de caracterização temporal de contatos. ....                                                                                                               | 145 |
| <b>Figura C.1</b> – Gráfico de elbow para a determinação da distância intercluster de corte da réplica $R_1(M_1)$ .....                                                                                   | 154 |
| <b>Figura C.2</b> – Gráfico de elbow para a determinação da distância intercluster de corte da réplica $R_1(M_2)$ .....                                                                                   | 154 |

|                                                                                                                         |     |
|-------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figura C.3</b> – Gráfico de elbow para a determinação da distância intercluster de corte da replica $R_1(M_3)$ ..... | 155 |
| <b>Figura C.4</b> – Gráfico de elbow para a determinação da distância intercluster de corte da replica $R_1(M_4)$ ..... | 155 |

## Lista de tabelas

---

|                                                                                                                                                                        |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Tabela 1.1</b> – Número de casos confirmados de doença de Chagas aguda (DCA) e óbitos confirmados segundo Unidade Federada no Brasil, entre 2020–2024.....          | 21 |
| <b>Tabela 1.2</b> – Casos confirmados no Brasil segundo formas de transmissão, entre 2020–2024.....                                                                    | 22 |
| <b>Tabela 1.3</b> – Alvos moleculares, relevância biológica e candidatos terapêuticos em <i>Trypanosoma cruzi</i> . ....                                               | 24 |
| <b>Tabela 2.1</b> – Número de moléculas adicionadas para a concentração de 5% (V/V.) e suas respectivas massas molares, para cada sistema. ....                        | 43 |
| <b>Tabela 2.2</b> – Distribuição percentual de resíduos no diagrama de Ramachandran dos modelos gerados por AlphaFold 3 e RoseTTAFold.....                             | 50 |
| <b>Tabela 2.3</b> – Valores de RMSD dos modelos gerados por RoseTTAFold, calculados para toda a proteína, resíduos 1–57 e resíduos 58–154, relativos ao Modelo 2. .... | 52 |
| <b>Tabela 2.4</b> – Distribuição percentual de resíduos no diagrama de Ramachandran dos modelos gerados por RoseTTAFold. ....                                          | 53 |
| <b>Tabela 2.5</b> – Resíduos interativos da TcP21, classificados quanto à recorrência entre as regiões de interação previstas pelo FTSite.....                         | 67 |
| <b>Tabela 2.6</b> – Resultados sumariados de resíduos interativos recorrentes nos respectivos ensaios realizados.....                                                  | 68 |
| <b>Tabela 3.1</b> – Siglas, nomenclaturas e número de moléculas inseridas dos constituintes da bicamada lipídica modelada.....                                         | 75 |
| <b>Tabela 3.2</b> – Índices <i>g</i> de Hedges e respectivos descritores de magnitude de tamanho de efeito.....                                                        | 79 |
| <b>Tabela 3.3</b> – Classificação dos clusters obtidos por ancoragem molecular, segundo número de membros e pontuação ponderada. ....                                  | 82 |
| <b>Tabela 3.4</b> – Estatística descritiva de RMSD da CXCR4.....                                                                                                       | 84 |
| <b>Tabela 3.5</b> – Estatística descritiva de RMSD da TcP21.....                                                                                                       | 86 |
| <b>Tabela 3.6</b> – Estatística descritiva do número de ligações de hidrogênio entre CXCR4 e TcP21.....                                                                | 89 |
| <b>Tabela 3.7</b> – Estatística descritiva do número de pontes salinas entre CXCR4 e TcP21. ....                                                                       | 91 |

|                                                                                                                                   |     |
|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Tabela 3.8</b> – Top 10 resíduos da TcP21 de maior frequência de ocupação de ligações de hidrogênio, por sistema. ....         | 101 |
| <b>Tabela B.1</b> – Matriz de categorização temporal baseada nas transições entre janelas consecutivas. ....                      | 147 |
| <b>Tabela B.2</b> – Consolidação das categorias temporais em classes funcionais e suas respectivas interpretações físicas. ....   | 147 |
| <b>Tabela B.3</b> – Categorização dos aminoácidos em classes funcionais segundo suas propriedades físico-químicas. ....           | 150 |
| <b>Tabela B.4</b> – Correspondência entre combinações de classes funcionais e respectivas forças de interação predominantes. .... | 150 |
| <b>Tabela D.1</b> – Análise post hoc para RMSD da TcP21. ....                                                                     | 157 |
| <b>Tabela D.2</b> – Análise post hoc para número de ligações de hidrogênio entre TcP21 e CXCR4. ....                              | 157 |
| <b>Tabela D.3</b> – Análise post hoc para número de pontes salinas entre TcP21 e CXCR4. ....                                      | 158 |

## Lista de siglas

---

|       |                                                    |
|-------|----------------------------------------------------|
| ACM   | Acetamida                                          |
| ACN   | Acetonitrila                                       |
| ACT   | Íon acetato                                        |
| ALA   | Alanina                                            |
| ARG   | Arginina                                           |
| ASN   | Asparagina                                         |
| ASP   | Ácido aspártico                                    |
| BNZ   | Benzeno                                            |
| CXCR4 | C-X-C <i>chemokine receptor type 4</i>             |
| CYS   | Cisteína                                           |
| DM    | Dinâmica Molecular                                 |
| ECL   | <i>Extracellular loop</i>                          |
| EPM   | Erro padrão da média                               |
| ETH   | Etanol                                             |
| GLN   | Glutamina                                          |
| GLU   | Ácido glutâmico                                    |
| GLY   | Glicina                                            |
| GPCR  | <i>G protein-coupled receptor</i>                  |
| HIS   | Histidina                                          |
| ILE   | Isoleucina                                         |
| LEU   | Leucina                                            |
| LYS   | Lisina                                             |
| MAM   | Íon metilamônio                                    |
| MET   | Metionina                                          |
| MSCS  | <i>Multiple Solvent Crystal Structures</i>         |
| MSMD  | <i>Mixed-solvent Molecular Dynamics</i>            |
| OPLS  | <i>Optimized Potentials for Liquid Simulations</i> |
| PHE   | Fenilalanina                                       |
| PRO   | Prolina                                            |
| RMSD  | <i>Root-mean-squared Deviation</i>                 |
| RMSF  | <i>Root-mean-squared Fluctuation</i>               |
| SASA  | <i>Solvent-Accessible Surface Area</i>             |
| SER   | Serina                                             |
| THR   | Treonina                                           |
| TM    | Transmembrana                                      |
| TRP   | Triptofano                                         |
| TYR   | Tirosina                                           |
| VAL   | Valina                                             |

## Sumário

---

|         |                                                                |    |
|---------|----------------------------------------------------------------|----|
| 1       | <b>INTRODUÇÃO</b>                                              | 20 |
| 2       | <b>ESTUDO 1</b>                                                | 35 |
| 2.1     | <b>METODOLOGIA</b>                                             | 36 |
| 2.1.1   | <i>Fluxograma geral de trabalho</i>                            | 36 |
| 2.1.2   | <i>Obtenção da estrutura inicial da TcP21</i>                  | 37 |
| 2.1.3   | <i>Parâmetros de simulação de DM</i>                           | 39 |
| 2.1.4   | <i>Identificação dos sítios de ligação</i>                     | 40 |
| 2.1.4.1 | <b>Simulações de MSMD</b>                                      | 41 |
| (a)     | <i>Solventes orgânicos e descrição do Campo de Força</i>       | 41 |
| (b)     | <i>Configuração dos sistemas</i>                               | 42 |
| (c)     | <i>Parâmetros das simulações de MSMD</i>                       | 43 |
| 2.1.4.2 | <b>Ancoragem da sonda farmacofórica</b>                        | 43 |
| (a)     | <i>Sonda farmacofórica e descrição do Campo de Força</i>       | 44 |
| (b)     | <i>Ancoragem molecular da sonda farmacofórica</i>              | 45 |
| (c)     | <i>Parâmetros das simulações de DM</i>                         | 45 |
| 2.1.4.3 | <b>Ancoragem molecular baseada em fragmentos</b>               | 46 |
| 2.1.5   | <i>Análise dos sítios de ligação preditos</i>                  | 46 |
| 2.2     | <b>RESULTADOS E DISCUSSÃO</b>                                  | 48 |
| 2.2.1   | <i>Avaliação preliminar dos modelos preditos</i>               | 48 |
| 2.2.2   | <i>Conformação inicial da TcP21 e equilíbrio termodinâmica</i> | 51 |
| 2.2.3   | <i>Identificação de sítios de ligação</i>                      | 57 |
| 2.2.3.1 | <b>Simulações de MSMD</b>                                      | 57 |
| 2.2.3.2 | <b>Ancoragem da sonda farmacofórica</b>                        | 60 |
| 2.2.3.3 | <b>Ancoragem molecular baseada em fragmentos</b>               | 65 |
| 2.2.4   | <i>Correlação dos sítios de ligação preditos</i>               | 67 |
| 2.3     | <b>CONCLUSÃO</b>                                               | 70 |

|          |                                                                            |            |
|----------|----------------------------------------------------------------------------|------------|
| <b>3</b> | <b>ESTUDO 2</b>                                                            | <b>71</b>  |
| 3.1      | <b>METODOLOGIA</b>                                                         | <b>72</b>  |
| 3.1.1    | <i>Fluxograma geral de trabalho</i>                                        | 72         |
| 3.1.2    | <i>Ancoragem molecular</i>                                                 | 73         |
| 3.1.3    | <i>Construção dos sistemas</i>                                             | 74         |
| 3.1.4    | <i>Parâmetros das simulações de DM</i>                                     | 76         |
| 3.1.5    | <i>Análises dos sistemas</i>                                               | 78         |
| 3.1.5.1  | <b>Estabilidade conformacional e avaliação dos modos de ligação</b>        | 78         |
| 3.1.6    | <b><i>Determinação por consenso da interface de interação</i></b>          | <b>79</b>  |
| 3.1.6.1  | <b>Caracterização temporal dos contatos entre TcP21 e CXCR4</b>            | <b>80</b>  |
| 3.2      | <b>RESULTADOS E DISCUSSÃO</b>                                              | <b>82</b>  |
| 3.2.1    | <b><i>Obtenção dos modos de ligação</i></b>                                | <b>82</b>  |
| 3.2.2    | <b><i>Estabilidade conformacional e avaliação dos modos de ligação</i></b> | <b>83</b>  |
| 3.2.3    | <b><i>Determinação por consenso da interface de interação</i></b>          | <b>94</b>  |
| 3.2.3.1  | <b>Caracterização temporal dos contatos entre TcP21 e CXCR4</b>            | <b>103</b> |
| 3.2.3.2  | Variabilidade de parceiros de interação da TcP21                           | 103        |
| 3.2.3.3  | Recorrência de pares de resíduos TcP21–CXCR4                               | 105        |
| 3.2.3.4  | Distribuição do perfil temporal de contatos                                | 107        |
| 3.2.3.5  | Distribuição qualitativa das forças de interação                           | 110        |
| 3.2.4    | <b><i>Correlação das interfaces de interação identificadas</i></b>         | <b>112</b> |
| 3.3      | <b>CONCLUSÃO</b>                                                           | <b>114</b> |
| <b>4</b> | <b>CONCLUSÃO GERAL</b>                                                     | <b>115</b> |
|          | <b>REFERÊNCIAS</b>                                                         | <b>116</b> |
|          | <b>APÊNDICE A – ROTINA DE IDENTIFICAÇÃO DE RESÍDUOS INTERFACIAIS</b>       | <b>138</b> |
|          | <b>APÊNDICE B – ROTINA DE CARACTERIZAÇÃO TEMPORAL DE CONTATOS</b>          | <b>145</b> |
|          | <b>APÊNDICE C – ROTINA DE CLUSTERIZAÇÃO E GRÁFICOS DE <i>ELBOW</i></b>     | <b>151</b> |
|          | <b>APÊNDICE D – DADOS DE ANÁLISE <i>POST HOC</i></b>                       | <b>157</b> |

# 1 INTRODUÇÃO

---

Há 117 anos, a Tripanossomíase americana, doença tropical mais conhecida como doença de Chagas, causada pelo parasita protozoário *Trypanosoma cruzi*, foi descoberta e formalmente descrita pelo Dr. Carlos Ribeiro Justiniano das Chagas (Chagas, 1909). Trata-se de uma doença que não somente afeta populações socioeconomicamente vulneráveis, como também agrava a desigualdade econômica, contribui para a marginalização social e reduz ainda mais a já baixa visibilidade política daqueles afligidos por ela (De Oliveira Junior *et al.*, 2022; Kropf; Lima, 2022; Ramos-Junior; Sousa, 2017). Apesar disso — ou talvez até por esse motivo — a doença de Chagas ainda não possui cura ou tratamento eficaz.

Embora seja encontrada majoritariamente em países da América Latina, onde se estima que entre 6 e 8 milhões de pessoas estejam infectadas, cerca de 70 milhões de pessoas estejam expostas ao risco de infecção e cause 12 mil mortes por ano, essa doença acarreta prejuízos econômicos anuais de aproximadamente 8 bilhões de dólares, impactando substancialmente os sistemas de saúde e comprometendo a produtividade e o tempo de contribuição dos infectados na sociedade (Conteh *et al.*, 2010; Lee, Bruce Y *et al.*, 2013; Olivera; Buitrago, 2020).

No Brasil, dados do Ministério da Saúde obtidos a partir do sistema DATASUS e apresentados na Tabela 1.1, indicaram a ocorrência de 2.009 casos confirmados de doença de Chagas aguda no período de 2020 a 2024. A distribuição geográfica dessas notificações evidenciou forte concentração na região Norte, com destaque para o estado do Pará, responsável por 81% do total de casos registrados no país.

Em relação à mortalidade, foram registrados 19.887 óbitos no mesmo intervalo, valor aproximadamente dez vezes superior ao número de casos confirmados. A região Sudeste concentrou 46,2% desses óbitos, com destaque para os estados de Minas Gerais (4.553 mortes, 22,9%) e São Paulo (4.475 mortes, 22,5%). Diferentemente da distribuição dos casos confirmados, a mortalidade apresentou padrão mais homogêneo entre as regiões brasileiras, incluindo contribuições substanciais das regiões Nordeste (24,7%) e Centro-Oeste (22,8%).



**Tabela 1.1** – Número de casos confirmados de doença de Chagas aguda (DCA) e óbitos confirmados segundo Unidade Federada no Brasil, entre 2020–2024.

| <b>Região/UF de notificação</b> | <b>Casos confirmados</b> | <b>Óbitos confirmados</b> |
|---------------------------------|--------------------------|---------------------------|
| <b>Norte</b>                    | <b>1.923</b>             | <b>367</b>                |
| Acre                            | 35                       | 11                        |
| Amapá                           | 189                      | 16                        |
| Amazonas                        | 61                       | 4                         |
| Pará                            | 1.631                    | 89                        |
| Rondônia                        | 5                        | 73                        |
| Roraima                         | 0                        | 1                         |
| Tocantins                       | 2                        | 173                       |
| <b>Nordeste</b>                 | <b>49</b>                | <b>4.911</b>              |
| Alagoas                         | 0                        | 411                       |
| Bahia                           | 11                       | 3.026                     |
| Ceará                           | 7                        | 258                       |
| Maranhão                        | 4                        | 35                        |
| Paraíba                         | 17                       | 172                       |
| Pernambuco                      | 1                        | 562                       |
| Piauí                           | 6                        | 267                       |
| Rio Grande do Norte             | 3                        | 96                        |
| Sergipe                         | 0                        | 84                        |
| <b>Centro-Oeste</b>             | <b>10</b>                | <b>4.538</b>              |
| Distrito Federal                | 2                        | 1.044                     |
| Goiás                           | 0                        | 3.138                     |
| Mato Grosso                     | 4                        | 220                       |
| Mato Grosso do Sul              | 4                        | 136                       |
| <b>Sudeste</b>                  | <b>16</b>                | <b>9.197</b>              |
| Espírito Santo                  | 1                        | 27                        |
| Minas Gerais                    | 2                        | 4.553                     |
| Rio de Janeiro                  | 10                       | 142                       |
| São Paulo                       | 3                        | 4.475                     |
| <b>Sul</b>                      | <b>11</b>                | <b>874</b>                |
| Paraná                          | 5                        | 724                       |
| Rio Grande do Sul               | 6                        | 129                       |
| Santa Catarina                  | 0                        | 21                        |
| <b>Brasil</b>                   | <b>2.009</b>             | <b>19.887</b>             |

Fontes: Adaptado de Sistema de Informação de Agravos de Notificação (Disponível em: <https://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/chagasbr.def>) e Sistema de Informações sobre Mortalidade (Disponível em: <https://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def>), 2026.

O impacto da doença de Chagas, entretanto, se estende além dos grupos primariamente afetados, emergindo como um problema global que já demanda vigilância em muitos países não endêmicos, incluindo Estados Unidos, Canadá, Austrália, Japão e países Europeus (Agudelo Higueta *et al.*, 2024; Manne-Goehler *et*

*al.*, 2016; Ramos-Rincon *et al.*, 2024; Requena-Méndez *et al.*, 2015; Schmunis; Yadon, 2010).

Nesse contexto, as estratégias de enfrentamento das doenças tropicais, incluindo a doença de Chagas, estão atualmente alinhadas com a Agenda 2030 de Objetivos de Desenvolvimento Sustentável (WHO, 2020), impulsionadas por fatores como o aumento dos fluxos migratórios e as mudanças climáticas. Considerando que insetos vetores são organismos ectotérmicos, cuja fisiologia é regulada pela temperatura, antecipa-se que alterações climáticas, incluindo o aumento da temperatura média e alterações de regimes de precipitação, favorecerão a expansão de suas áreas de distribuição (Souza, Rita de Cássia Moreira De *et al.*, 2022).

Adicionalmente, a migração associada à urbanização, frequentemente desacompanhada de políticas públicas adequadas de saneamento e infraestrutura, vem modificando o paradigma da forma de transmissão da doença. Observa-se, atualmente, a transição de um cenário predominantemente dominado por transmissão vetorial para um aumento expressivo da transmissão oral (Agudelo Higueta *et al.*, 2026; Alarcón de Noya *et al.*, 2022). Esta forma de transmissão é usualmente associada ao consumo de alimentos contaminados com o *T. cruzi* durante seu processamento ou manipulação, incluindo produtos como açaí, cana-de-açúcar, vinho de palma e polpas de frutos silvestres (Beatty *et al.*, 2024; De Sousa *et al.*, 2023; Franco-Paredes *et al.*, 2020; Monsalve-Lara *et al.*, 2021; Sampaio *et al.*, 2020). Dados do Ministério da Saúde, apresentados na Tabela 1.2, corroboram essa tendência no Brasil, indicando que a transmissão oral foi responsável por 87,8% dos casos notificados no período de 2020 a 2024.

**Tabela 1.2** – Casos confirmados no Brasil segundo formas de transmissão, entre 2020–2024.

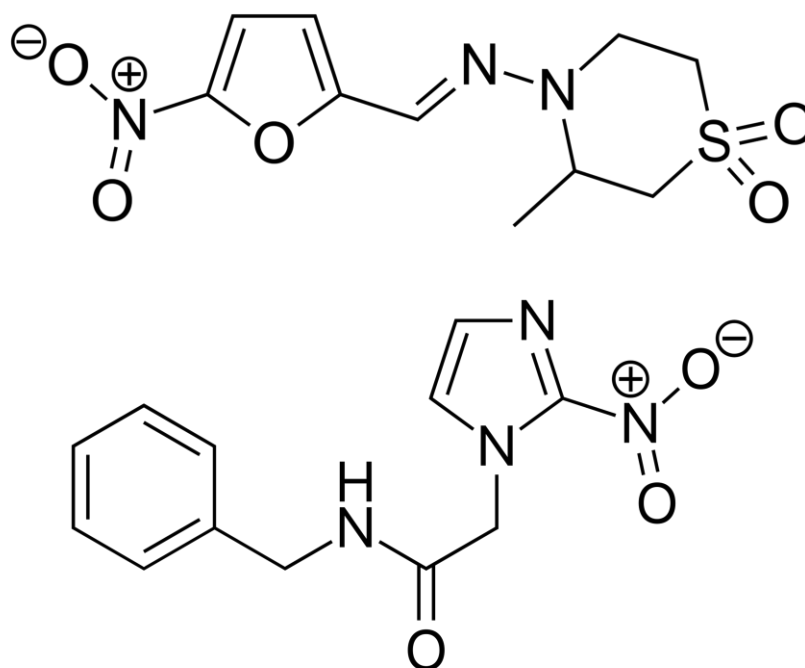
| Formas de transmissão | Casos confirmados |
|-----------------------|-------------------|
| Transfusional         | 0                 |
| Vetorial              | 114               |
| Vertical              | 5                 |
| Acidental             | 5                 |
| Oral                  | 1.764             |
| <b>Total</b>          | <b>1.888</b>      |

Fonte: Adaptado de Sistema de Informação de Agravos de Notificação (Disponível em: <https://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/chagasbr.def>), 2026.

Diversas estratégias de controle voltadas à interrupção da transmissão da doença de Chagas foram desenvolvidas ao longo das últimas décadas, incluindo o controle vetorial domiciliar (Tian *et al.*, 2024), melhorias nas condições habitacionais (Bates *et al.*, 2023), fortalecimento da vigilância epidemiológica (Rafael *et al.*, 2023) e implementação de medidas de segurança alimentar (Fujita *et al.*, 2019). Esses avanços resultaram em um panorama de diminuição da incidência da doença, sobretudo em países endêmicos (Cousin *et al.*, 2026). Ainda assim, a erradicação da doença não parece ser uma realidade a curto e médio prazos.

Do ponto de vista terapêutico, o tratamento disponível para a doença de Chagas permanece dependente, exclusivamente, de dois fármacos nitroheterocíclicos, o nifurtimox (US3262930A, 1966) (Bocca Tourres, 1969; Bock *et al.*, 1969; Rubio; Donoso, 1969) e o benznidazol (US3505349A, 1970) (Polak; Richle, 1978; Raaflaub; Ziegler, 1979) (Figura 1.1), ambos desenvolvidos há mais de cinco décadas.

**Figura 1.1** – Estruturas químicas dos fármacos tripanossomicidas. De cima para baixo: nifurtimox e benznidazol.



Fonte: Autor, 2026.

Esses fármacos apresentam eficiência relativa predominantemente na fase aguda da infecção e são associados a elevada toxicidade, manifestada por reações adversas como dermatite alérgica, distúrbios gastrointestinais, incluindo náusea e perda de peso, e distúrbios do sistema nervoso central, como insônia e depressão, frequentemente resultando na interrupção do tratamento por parte do paciente (Berenstein *et al.*, 2021; Jackson *et al.*, 2020). Do ponto de vista bioquímico, esses compostos não atuam sobre alvos moleculares definidos, mas induzem a formação de intermediários reativos e espécies reativas de oxigênio, promovendo danos inespecíficos a macromoléculas celulares (Sales Junior *et al.*, 2017).

Em resposta a essas limitações, esforços em pesquisas têm se concentrado na identificação de novos alvos moleculares e no desenvolvimento de candidatos terapêuticos mais seletivos e eficazes (Gomes *et al.*, 2023). A Tabela 1.3 apresenta alguns dos principais alvos moleculares investigados, juntamente com a sua relevância biológica e seus respectivos candidatos terapêuticos.

**Tabela 1.3** – Alvos moleculares, relevância biológica e candidatos terapêuticos em *Trypanosoma cruzi*.

| Alvo Molecular                               | Relevância biológica                                              | Candidato terapêutico                           | Referência                                                                                            |
|----------------------------------------------|-------------------------------------------------------------------|-------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| Proteassoma de cinetoplastídeos              | Degrada proteínas essenciais (homeostase proteica)                | LXE408                                          | (Lima <i>et al.</i> , 2022)                                                                           |
| N-miristoiltransferase                       | Modificação lipídica essencial para função e localização proteica | DDD85646                                        | (González García <i>et al.</i> , 2026)                                                                |
| Cruzipaína                                   | Protease envolvida na interação parasito–hospedeiro               | Inibidores de cruzapaína                        | (Burtoloso <i>et al.</i> , 2017; Santos <i>et al.</i> , 2023)                                         |
| CYP51                                        | Síntese de ergosterol (membrana)                                  | Posaconazol, ravuconazol, derivados de piridina | (Almeida-Silva <i>et al.</i> , 2022; Gunatilleke <i>et al.</i> , 2012; Hargrove <i>et al.</i> , 2013) |
| Via do ergosterol (geral)                    | Produção de esteróis de membrana                                  | Azóis (posaconazol) e inibidores                | (Gunatilleke <i>et al.</i> , 2012; Villalta; Rachakonda, 2019)                                        |
| Metabolismo redox / estresse nitro-oxidativo | Geração de espécies reativas tóxicas                              | Fexinidazol                                     | (Zuma; De Souza, 2022)                                                                                |
| Tripanotiona Redutase                        | Manutenção do equilíbrio redox                                    | Dissulfiram                                     | (Almeida-Silva <i>et al.</i> , 2022)                                                                  |

Fonte: Autor, 2026.

Entretanto, uma vez estabelecida a infecção do hospedeiro, o foco no combate à doença de Chagas se desloca de uma perspectiva sanitária macroscópica para a investigação da complexidade biológica do parasita em nível celular. Nesse contexto, a elucidação dos mecanismos de manutenção e progressão da infecção no organismo hospedeiro assume papel central.

O ciclo de vida do *Trypanosoma cruzi* ainda vem sendo amplamente estudado (Martín-Escolano *et al.*, 2022). Em hospedeiros mamíferos, o ciclo se inicia com a introdução da forma tripomastigota metacíclica na corrente sanguínea, seja por meio do contato com fezes infectadas dos vetores triatomíneos durante o repasto sanguíneo ou pela ingestão oral de parasitas (Pereira, Karen Signori *et al.*, 2009). Após a entrada na corrente sanguínea, o parasita invade células fagocíticas (macrófagos) e não-fagocíticas (musculares e epiteliais) do hospedeiro, diferenciando-se na forma amastigota e estabelecendo um ciclo intracelular. As amastigotas replicam-se por divisão binária no interior da célula, diferenciam-se na forma tripomastigota circulante e, então, são liberadas na corrente sanguínea, possibilitando a infecção de novas células. Cada etapa deste ciclo biológico compreende múltiplas vias de sinalização molecular e os mecanismos específicos de infecção dependem não somente da expressão morfológica do parasita, mas ainda da ampla variedade de cepas existentes e dos tipos de células hospedeiras a serem infectadas (Rodríguez-Bejarano *et al.*, 2021; Romano *et al.*, 2012; Zingales *et al.*, 2014).

Embora as expressões morfológicas de tripomastigotas (metacíclica e circulante) utilizem mecanismos distintos para a invasão celular no hospedeiro, estudos demonstram que as vias de sinalização molecular convergem para um processo comum de internalização. Esse processo envolve, inicialmente, a participação ativa de mucinas e trans-sialidases na modulação da superfície do parasita, promovendo sua adesão à célula hospedeira. Em seguida, proteínas parasitárias, como a gp82, induzem a mobilização citoplasmática de íons  $\text{Ca}^{2+}$ , desencadeando dois processos paralelos na região de contato: a desorganização do citoesqueleto cortical, por meio da despolimerização local de actina, e a exocitose lisossomal, que promove a incorporação da membrana lisossomal à membrana plasmática. Em conjunto, esses processos reduzem a rigidez mecânica da membrana celular na região de contato e aumentam a área de membrana disponível, tornando-a

permissiva à deformação e ao envolvimento progressivo do parasita. Por fim, esse processo culmina na formação de um vacúolo parasitóforo, permitindo a internalização do parasita na célula hospedeira (Burleigh; Andrews, 1998; Maeda *et al.*, 2012).

Entretanto, observou-se que, após a invasão e subsequente lise celular, a forma tripomastigota circulante não foi a única expressão morfológica liberada na corrente sanguínea; a forma amastigota também estava presente, agora denominada amastigota extracelular ou circulante (Tyler; Engman, 2001). Esta forma apresentou capacidade de induzir um estado metabolicamente inativo (quiescência metabólica), fator importante para a resistência a medicamentos contra a doença de Chagas (Francisco *et al.*, 2017; Sánchez-Valdéz *et al.*, 2018). Além disso, contrariando o que se assumia anteriormente, também apresentou capacidade infecciosa além da função meramente replicativa (Ley *et al.*, 1988; Martín-Escolano *et al.*, 2022; Mortara *et al.*, 2005), característica que pode ser determinante para compreender a dinâmica de infecção do parasita e sua persistência nos organismos infectados, sobretudo em tecidos musculares (Desale *et al.*, 2024; Mortara *et al.*, 2008; Won *et al.*, 2023).

Nesse contexto, da Silva *et al.* (Da Silva *et al.*, 2009) identificaram uma proteína ubíqua de 21 kDa, denominada P21 ou TcP21, secretada por amastigotas extracelulares, caracterizada por elevada afinidade para adesão à membrana celular e por se ligar a células hospedeiras de forma dose-dependente. Estudos subsequentes associaram a presença da forma recombinante desta proteína no meio extracelular à inibição da replicação de amastigotas intracelulares, também de forma dose-dependente, e à indução da polimerização de actina (Teixeira, Thaise L. *et al.*, 2019). Adicionalmente, a TcP21 foi implicada na modulação da internalização do parasita, promovendo a entrada de tripomastigotas metacíclicos após pré-tratamento das células hospedeiras, bem como aumentando ou inibindo a de amastigotas extracelulares, a depender da sua presença durante ou antes da infecção (Da Silva *et al.*, 2009), além da participação ativa na retenção intracelular do parasita sob condições de estresse, favorecendo um estado compatível com quiescência metabólica, previamente associado a formas amastigotas extracelulares (Martins *et al.*, 2020; Silveira *et al.*, 2024).

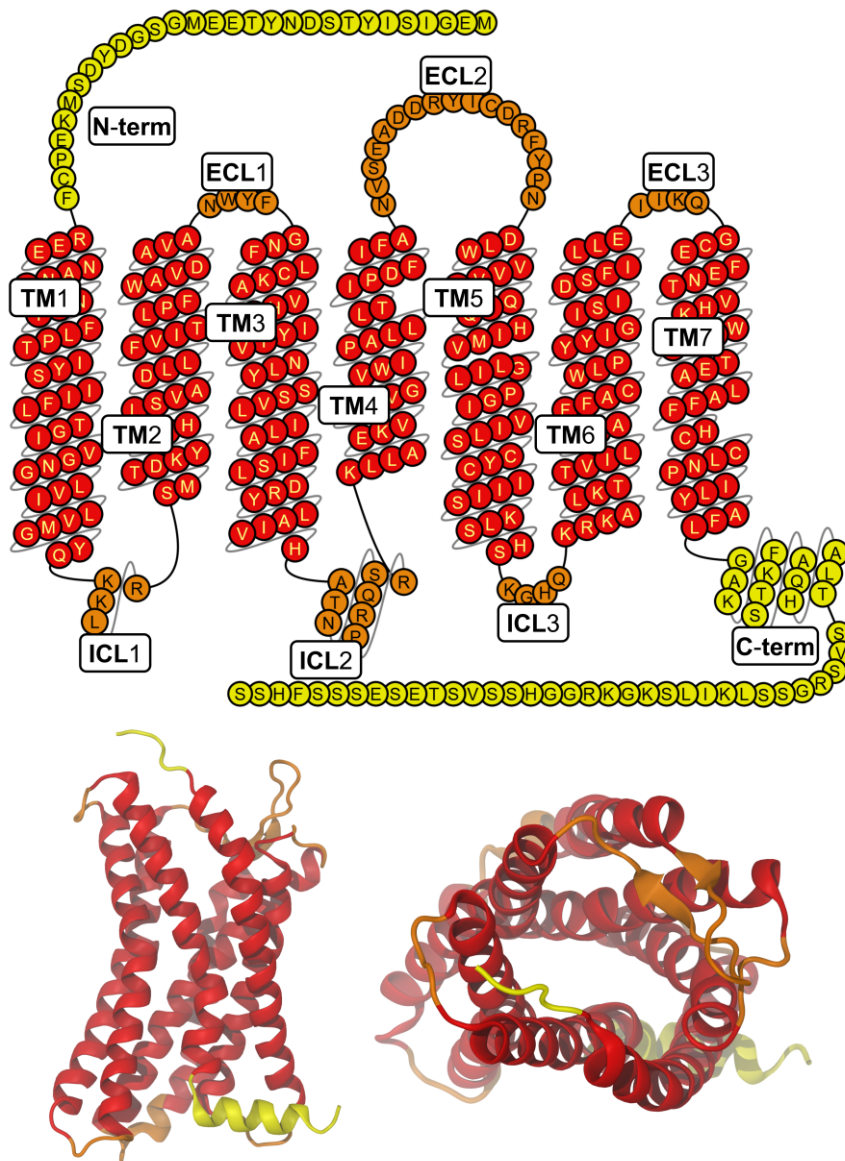
Somado a isso, a proteína TcP21 demonstrou capacidade de modular processos celulares, incluindo a indução de fagocitose em células de macrófagos, por meio da interação direta com o receptor de membrana CXCR4 (do inglês, C-X-C

*chemokine receptor type 4*) (Borges *et al.*, 2020; Rodrigues *et al.*, 2012; Teixeira, Samuel Cota *et al.*, 2017). A proteína CXCR4 consiste em um receptor de quimiocina amplamente expresso no organismo humano, cuja ativação pelo ligante nativo CXCL12 desencadeia múltiplas vias de sinalização intracelular associadas a processos como migração celular, adesão, reparo tecidual e fagocitose (Bianchi; Mezzapelle, 2020; Guo *et al.*, 2016; Pawig *et al.*, 2015; Zhang, Zhiying *et al.*, 2025).

Estruturalmente, o CXCR4 pertence à superfamília de receptores transmembranares acoplados à proteína G (GPCR, do inglês *G protein-coupled receptors*), caracterizados por sua topologia de sete hélices transmembranares (TM1–7), três voltas intracelulares (ICL1–3) e três voltas extracelulares (ECL1–3) (Figura 1.2).

Essa superfamília representa um dos maiores conjuntos de proteínas de membrana em eucariotos e é responsável pela transdução de uma ampla variedade de sinais extracelulares, incluindo neurotransmissores e quimiocinas, desempenhando papel fundamental na fisiologia e na patogênese de diversas doenças (Griffith *et al.*, 2014; Proudfoot, 2002; Sarma *et al.*, 2023). Estima-se que aproximadamente dois terços dos hormônios humanos (Foster *et al.*, 2019) e cerca de um terço dos fármacos comercialmente disponíveis (Hauser *et al.*, 2017; Lorente *et al.*, 2025) atuem por meio de receptores GPCR.

**Figura 1.2** – Representação esquemática da proteína CXCR4, segundo distribuição dos aminoácidos e orientações das hélices transmembranares.



Fonte: Adaptado de <https://gpcrdb.org>, 2026.

Apesar do potencial significativo da TcP21 para a compreensão e o enfrentamento da doença de Chagas (Teixeira, Thaise Lara *et al.*, 2015), bem como de sua multifuncionalidade em outros contextos biológicos, como a modulação da proliferação de células tumorais (Azevedo Silveira *et al.*, 2023), de respostas inflamatórias para a recuperação tecidual (Barbosa *et al.*, 2021), e da infecção de outros parasitas como o *Toxoplasma gondii* (De Souza, Guilherme *et al.*, 2023), diversos aspectos estruturais e funcionais dessa proteína permanecem desconhecidos. Destacam-se, entre eles, a ausência de estrutura tridimensional



experimentalmente determinada, a falta de identificação de sítios de ligação em sua superfície e a inexistência de descrição detalhada de seu mecanismo de interação com o receptor CXCR4 em nível molecular.

Os avanços recentes em arquiteturas de *deep learning*, exemplificados por métodos como AlphaFold 3 (Abramson *et al.*, 2024) e RoseTTAFold (Baek *et al.*, 2021), redefiniram os paradigmas da predição *in silico* de estruturas proteicas. Enquanto abordagens experimentais clássicas, incluindo a cristalografia de raios X, a criomicroscopia eletrônica (*cryo*-EM) e a espectroscopia de ressonância magnética nuclear (RMN), dependem de condições experimentais altamente específicas para a obtenção de estruturas de alta resolução, tais como cristalização, preparo de amostra e estabilidade conformacional (Kühlbrandt, 2014; Wüthrich, 2001), métodos baseados em *deep learning* passaram a inferir a organização tridimensional diretamente a partir da sequência primária de aminoácidos, explorando padrões estruturais implicitamente aprendidos durante o treinamento do modelo (Jumper *et al.*, 2021). Dessa forma, essas abordagens possibilitaram a obtenção de modelos estruturais com acurácia próxima à experimental, mesmo na ausência de homólogos estruturais conhecidos, além de viabilizarem a predição integrada de complexos biomoleculares envolvendo proteínas, ácidos nucleicos e ligantes (Jänes; Beltrao, 2024; Wang *et al.*, 2026; Zhang, Jintong *et al.*, 2026). Como consequência, observou-se uma expansão sem precedentes do número de estruturas disponíveis, resultando na geração de milhões de modelos estruturais (atualmente próximo de 250 milhões) e superando em várias ordens de magnitude o número de estruturas determinadas por abordagens experimentais (Varadi *et al.*, 2024; Yang *et al.*, 2026).

No contexto da identificação de sítios de ligação, técnicas experimentais, como o método MSCS (do inglês, *Multiple Solvent Crystal Structures*) (Allen *et al.*, 1996; Dechene *et al.*, 2009; English *et al.*, 1999; Mattos *et al.*, 2006; Mattos; Ringe, 1996), são consideravelmente laboriosas, caras e complexas, pois apresentam uma série de restrições experimentais, incluindo a integração de informações estruturais da proteína alvo e limitações relacionadas à solubilidade em certos solventes orgânicos (English *et al.*, 2001; Köhler *et al.*, 2020; Lisi; Loria, 2016; Ngan, Kevin Chun Ho *et al.*, 2023; Romanuka *et al.*, 2023). Nesse contexto, as abordagens computacionais direcionadas a uma avaliação preliminar de possíveis sítios de ligação têm se tornado

uma prática comum, já demonstrando sucesso considerável (Feher *et al.*, 2014; Wakefield *et al.*, 2022).

Há uma ampla variedade de métodos e ferramentas computacionais disponíveis para a predição de sítios alostéricos e seus tipos de modulação, avaliação de mudanças conformacionais em proteínas, amostragem de curvas de energia para diferentes complexos proteína-ligante e análise da dinâmica de proteínas em solvente (Amamuddy *et al.*, 2020). Contudo, apesar de inúmeras metodologias (Alvarez-Garcia; Barril, 2014; Kimura *et al.*, 2017; MacKerell *et al.*, 2020; Sabanés Zariquey *et al.*, 2019; Smith; Carlson, 2021), ainda não há um consenso sobre qual abordagem computacional seja a mais adequada para investigar sítios de ligação, dado que essas metodologias são frequentemente combinadas e estão em constante aperfeiçoamento (Wakefield *et al.*, 2022).

Duas abordagens computacionais comumente empregadas e com relativa correspondência ao MSCS são a Dinâmica Molecular em Solventes Mistos (MSMD; do inglês, *Mixed Solvent Molecular Dynamics*) (Arcon *et al.*, 2017; Chan *et al.*, 2023; DasGupta *et al.*, 2022; Ilie *et al.*, 2023; Lal Gupta; Carlson, 2022) e a ancoragem molecular baseada em fragmentos (Ahmed; Brooks, 2023; Buchholz *et al.*, 2023; Grant *et al.*, 2011; Pereira, Samille Henriques *et al.*, 2024).

A metodologia de MSMD se baseia na amostragem estatística de eventos moleculares de proteínas em um ambiente misto de solvente orgânico/água de forma explícita, empregando simulações por Dinâmica Molecular clássica (DM). Embora as informações obtidas por essa abordagem sejam de natureza estatística, ela oferece as vantagens de ser mais eficiente, econômica e menos trabalhosa que a abordagem experimental, permitindo uma alocação mais eficiente de recursos experimentais (Ghanakota; Carlson, 2016; Hertig *et al.*, 2016; Ung *et al.*, 2016). Além disso, a MSMD fornece informações sobre os tipos mais frequentes de interações intermoleculares entre a proteína e as sondas orgânicas, assim como sobre a dinâmica dessas interações ao longo do tempo (DasGupta *et al.*, 2024; Feng *et al.*, 2024; Kudo *et al.*, 2023).

De forma alternativa, a metodologia de ancoragem molecular baseada em fragmentos se fundamenta em modelos de corpo rígido e solvatação contínua para amostrar, de maneira livre, inúmeras conformações interativas — denominadas

modos de ligação — de moléculas orgânicas ao redor da proteína (Kozakov *et al.*, 2015), categorizando-as de acordo com funções de pontuação baseadas em energias de afinidade (Brenke *et al.*, 2009). Esta abordagem apresenta ainda mais eficiência que a MSMD, o que permite sua aplicação a uma maior variedade de sondas orgânicas. No entanto, isso ocorre em detrimento da desconsideração de elementos dinâmicos das interações intermoleculares e a presença explícita de moléculas de solvente (Kozakov *et al.*, 2015).

Por fim, a elucidação de interfaces de interação proteína–proteína configura-se como um desafio adicional, dado que propriedades emergentes, como complementaridade conformacional, efeitos alostéricos e reorganizações dinâmicas, desempenham papéis centrais na determinação do reconhecimento molecular (Weis; Kobilka, 2018). Nesse contexto, abordagens baseadas exclusivamente em modelos estruturais estáticos, como a ancoragem molecular, são frequentemente insuficientes para capturar, de forma realista, os determinantes moleculares mais prováveis que governam o reconhecimento e a estabilidade do complexo. Essa limitação está associada, sobretudo, à simplificação inerente ao tratamento de efeitos de solvatação, temperatura, flexibilidade estrutural, força iônica e à limitada extensão da amostragem conformacional (Huang, Sheng-You; Zou, 2010).

Dessa forma, a estratégia metodológica consolidada para mitigar essas limitações é a integração entre ancoragem molecular e simulações de dinâmica molecular. Nesse contexto, a ancoragem molecular é empregada como ferramenta inicial para a obtenção de conjuntos de modos de ligação plausíveis do complexo de interesse, que são então amostrados dinamicamente por meio de simulações de dinâmica molecular. Essa estratégia possibilita identificar padrões recorrentes de interação, tais como contatos intermoleculares persistentes ou redes de ligações de hidrogênio, bem como distinguir interações transitórias e pouco frequentes de elementos estruturais robustos, potencialmente relevantes para a formação e manutenção do complexo (Hollingsworth; Dror, 2018; Karplus; McCammon, 2002; Keskin *et al.*, 2016).

Em se tratando de sistemas moleculares de elevada complexidade, como é o caso de proteínas GPCR — incluindo a CXCR4 —, cuja dinâmica e funcionalidade são intrinsecamente dependentes do ambiente de membrana, impõe-se, também, o desafio da representatividade estatística em simulações envolvendo centenas de

milhares de átomos (Latorraca *et al.*, 2017; Lee, Ji Young; Lyman, 2012; Mondal *et al.*, 2013). A subamostragem do sistema pode introduzir vieses significativos na caracterização dos estados conformacionais com maior probabilidade de serem termodinamicamente acessados e, conseqüentemente, comprometer a reprodutibilidade das interfaces de interação identificadas (Gapsys; De Groot, 2020; Latorraca *et al.*, 2017). Nesse caso, estratégias que permitem uma amostragem mais abrangente, como a combinação sistemática de múltiplos modos de ligação e múltiplas réplicas independentes de simulação para cada modo, reduzem a dependência de estados iniciais específicos e aumentam a robustez estatística e a confiabilidade das inferências estruturais realizadas (Knapp *et al.*, 2018; Nigam *et al.*, 2021).

## OBJETIVOS

---

Dada a originalidade da proteína TcP21 e seu papel potencialmente relevante na infecção por *Trypanosoma cruzi* por meio de uma via alternativa ainda pouco explorada, a presente tese teve como objetivo geral investigar a TcP21 sob duas perspectivas computacionais distintas e complementares: (i) sua dinâmica em água, com a identificação, em sua superfície, de sítios potenciais de ligação a ligantes farmacofóricos; e (ii) a análise exploratória do complexo TcP21–CXCR4, com vista à predição de possíveis interfaces de interação. Para tanto, o trabalho foi estruturado em dois estudos, cada qual com objetivos específicos.

No Estudo 1, sobre o mapeamento de sítios de ligação para ligantes farmacofóricos, foram definidos os seguintes objetivos específicos:

(i) Obter uma conformação inicial da TcP21 por meio de ferramentas preditivas baseadas em Inteligência Artificial;

(ii) Empregar a metodologia de Dinâmica Molecular em Solventes Mistos, utilizando como sondas moleculares acetamida, acetonitrila, íon acetato, benzeno, etanol e íon metilamônio;

(iii) Construir um modelo de sonda farmacofórica e aplicá-lo em estudos de ancoragem molecular, seguidos por simulações de dinâmica molecular;

(iv) Aplicar a metodologia de ancoragem molecular baseada em fragmentos, por meio de um servidor automatizado;

(v) Integrar os resultados das três abordagens computacionais, propor possíveis sítios de interação na superfície da TcP21 e caracterizá-los quanto aos resíduos-chave e aos tipos de interações intermoleculares.

No Estudo 2, sobre a predição de possíveis interfaces de interação entre TcP21 e CXCR4, foram definidos os seguintes objetivos específicos:

- (i) Obter modos de ligação do complexo por meio de ancoragem molecular;
- (ii) Construir os sistemas moleculares, incluindo uma bicamada lipídica de composição semelhante à da célula de macrófago;
- (iii) Realizar simulações de dinâmica molecular multirréplicas para cada modo de ligação;
- (iv) Identificar os resíduos interfaciais e os contatos intermoleculares mais prováveis para a estabilização do complexo;
- (v) Caracterizar a dinâmica estrutural dos domínios envolvidos na interação;
- (vi) Propor um mecanismo geral de interação para o complexo, incluindo a identificação das interações intermoleculares que governam sua estabilização.

## 2 ESTUDO 1

---

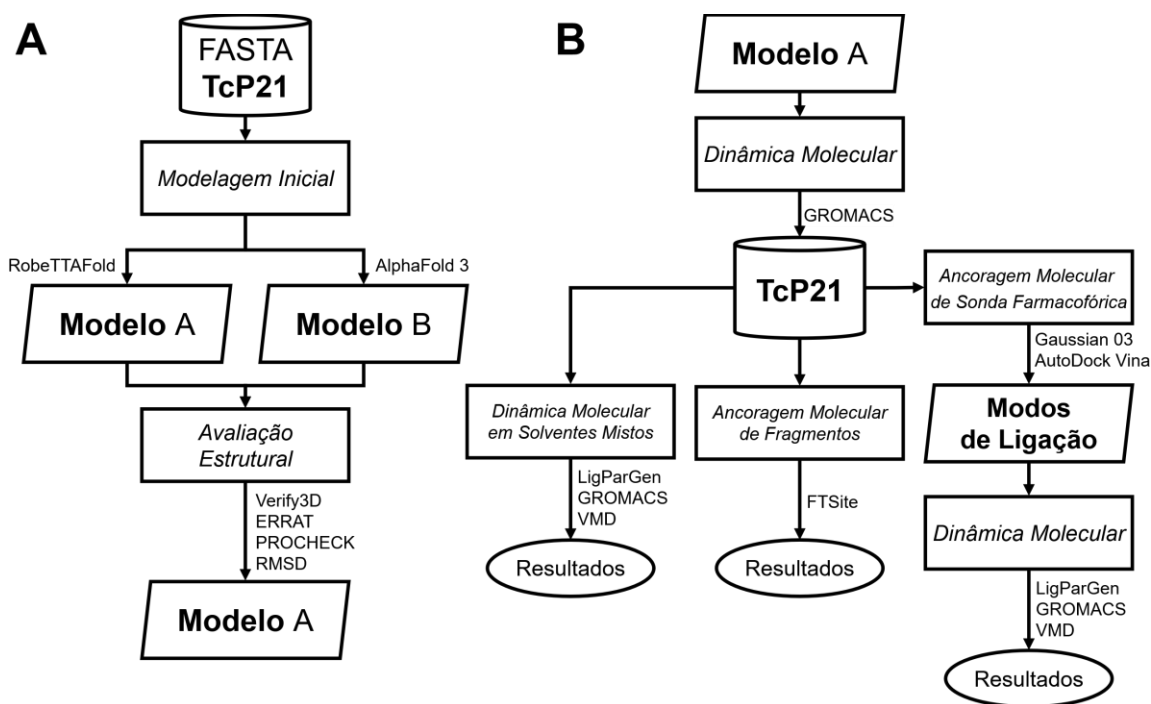
### **Mapeamento de sítios de ligação na TcP21 empregando abordagens integradas baseadas em sondas moleculares**

## 2.1 METODOLOGIA

### 2.1.1 Fluxograma geral de trabalho

A metodologia adotada neste Capítulo, voltada à obtenção de um modelo conformacional da TcP21 e à identificação de potenciais sítios de interação em sua estrutura, foi organizada como um fluxograma de trabalho e apresentada de forma geral na Figura 2.1.

**Figura 2.1** – Fluxograma geral de trabalho do Estudo 1, dividido entre obtenção do modelo conformacional (A) e identificação de sítios de ligação (B).



Fonte: Autor, 2026.

Em termos gerais, a metodologia se iniciou com a utilização da sequência FASTA da proteína TcP21 para a obtenção de possíveis modelos conformacionais por meio de modelagem baseada em Inteligência Artificial. A partir dos modelos preditos, foi feita uma avaliação estrutural rigorosa com a finalidade de verificar a qualidade dos modelos e identificar o mais adequado para os estudos subsequentes. O modelo



selecionado foi utilizado como ponto de partida para os ensaios de (i) dinâmica molecular em solventes mistos, (ii) ancoragem molecular de sonda farmacofórica seguida de dinâmica molecular e (iii) ancoragem molecular baseada em fragmentos.

As amostragens realizadas foram analisadas segundo um único eixo analítico principal: mapas volumétricos de frequência de ocupação. O processamento e as análises das trajetórias foram realizados utilizando as ferramentas do pacote computacional GROMACS (Abraham *et al.*, 2015), versão 2019.2, além de módulos do programa VMD (Humphrey *et al.*, 1996).

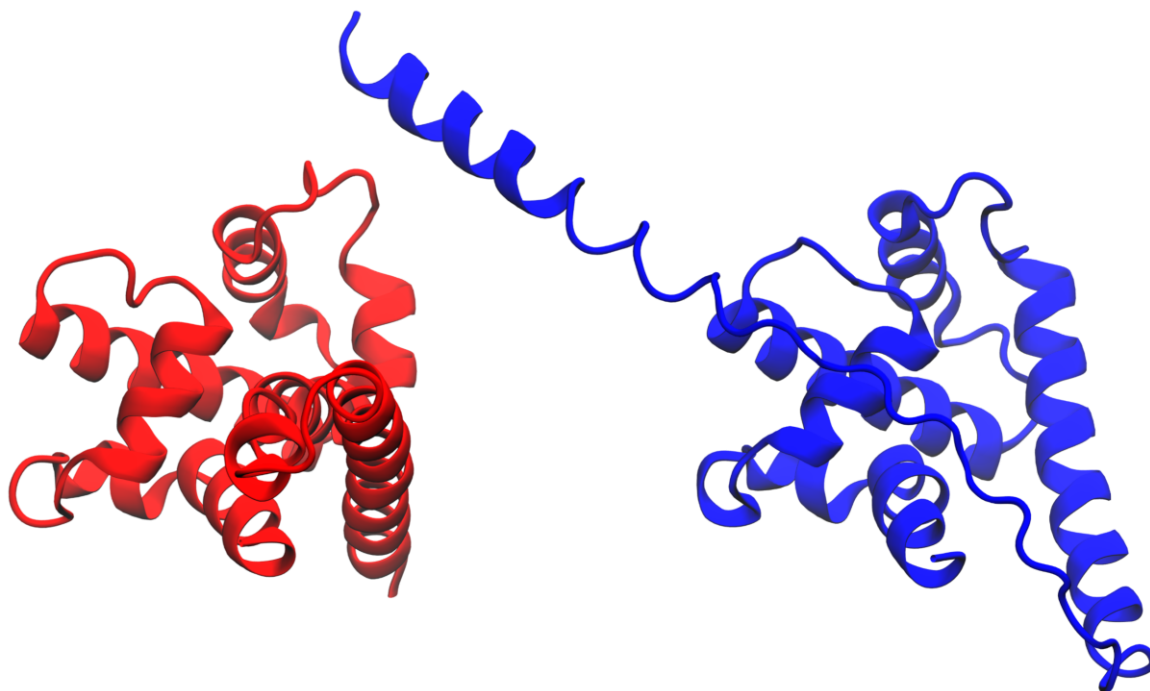
Os procedimentos metodológicos específicos, assim como os parâmetros e protocolos computacionais empregados em cada etapa, foram descritos nas subseções subsequentes.

### 2.1.2 **Obtenção da estrutura inicial da TcP21**

Dado que a estrutura da proteína TcP21 ainda não foi determinada experimentalmente, empregaram-se os métodos RoseTTAFold (Baek *et al.*, 2021) e AlphaFold 3 (Abramson *et al.*, 2024), ambos fundamentados em arquiteturas de *deep learning*, para a predição de conformações iniciais. Como ponto de partida para a modelagem, foi utilizada a sequência FASTA da proteína disponível no GenBank (ABS31351.1).

Para o método RoseTTAFold, foram obtidos cinco modelos candidatos, enquanto para o método AlphaFold 3, um modelo candidato. Os modelos preditos foram submetidos a uma análise comparativa de qualidade estrutural baseada nos programas Verify3D (Bowie *et al.*, 1991; Lüthy *et al.*, 1992), ERRAT (Colovos; Yeates, 1993) e PROCHECK (Laskowski *et al.*, 1993). Em função da alta similaridade entre os cinco modelos gerados por RoseTTAFold, apenas um deles foi utilizado como representante para a comparação com o modelo por AlphaFold 3. Os dois modelos considerados são mostrados na Figura 2.2.

**Figura 2.2** – Modelos preditos da proteína TcP21 gerados por RoseTTAFold (vermelho) e AlphaFold 3 (azul).



Fonte: Autor, 2026.

Quando comparados os cinco modelos candidatos gerados por RoseTTAFold entre si, foi selecionado para experimentos subsequentes aquele que apresentou melhor qualidade estrutural global, com base no perfil de erro estrutural por resíduo fornecido pelo próprio método e na análise do diagrama de Ramachandran.

Para refinamento estrutural, o modelo selecionado foi submetido a simulações de DM, visando à incorporação de efeitos de solvatação explícita, temperatura e pressão. A conformação inicial foi posicionada no centro de uma caixa de simulação retangular com dimensões de  $74,53 \times 69,79 \times 82,54$  Å, assegurando-se uma distância mínima de 15 Å entre a macromolécula e as faces da caixa. O sistema resultante compreendeu 41.732 partículas. Após a realização das simulações, a estrutura representativa central da TcP21 foi adotada como conformação inicial para os experimentos subsequentes.

### 2.1.3 *Parâmetros de simulação de DM*

Ao longo deste estudo, simulações de DM clássica foram utilizadas com frequência para amostrar conformações moleculares, e as trajetórias geradas foram empregadas na análise e caracterização de potenciais sítios de ligação. Os seguintes parâmetros de simulação foram aplicados para todas as simulações de DM. Detalhes específicos de cada etapa serão apresentados em suas respectivas seções.

Inicialmente, o sistema foi submetido a uma minimização de energia utilizando o algoritmo de *steepest descent*, com um passo de minimização de 0,1 Å e um critério de tolerância para força máxima inferior a 50,0 kJ mol<sup>-1</sup> Å<sup>-1</sup>. Após atingir um mínimo de energia local que satisfizesse o critério, foi realizado um tratamento térmico em *ensemble* isocórico-isotérmico (NVT). Esse tratamento consistiu em um incremento linear de temperatura de 0 a 310 K durante 400 ps, seguido de uma manutenção a 310 K durante 200 ps. A temperatura foi controlada pelo método *velocity-rescaling* (Bussi *et al.*, 2007) com uma constante de tempo de acoplamento de 0,1 ps. Este processo foi iterado três vezes para relaxar gradualmente a estrutura da proteína no ambiente do solvente, ajustando-se a constante de força aplicada ao movimento dos átomos, exceto hidrogênios, a cada iteração. A constante de força foi inicialmente de 10 kJ mol<sup>-1</sup> Å<sup>-2</sup>, depois reduzida para 5 kJ mol<sup>-1</sup> Å<sup>-2</sup> e, por fim, 1 kJ mol<sup>-1</sup> Å<sup>-2</sup>. Após o tratamento térmico, uma pressão de 1,0 bar foi aplicada isotropicamente durante 1.000 ps, em *ensemble* isobárico-isotérmico (NPT). A pressão foi controlada pelo método de Berendsen (Berendsen *et al.*, 1984), com uma constante de tempo de acoplamento de 2,0 ps e uma compressibilidade isotérmica de  $4,5 \times 10^{-5}$  bar<sup>-1</sup>. Durante o acoplamento de pressão, a constante de força foi mantida em 1 kJ mol<sup>-1</sup> Å<sup>-2</sup>.

Após a convergência de temperatura e da pressão nos valores requeridos, todas as restrições de movimento das partículas foram removidas e o sistema foi amostrado por 100 ns, no mesmo *ensemble*. Nesse momento, a pressão foi controlada pelo método de Parrinello-Rahman (Nosé; Klein, 1983; Parrinello; Rahman, 1981). Interações de van der Waals de curta distância foram limitadas por um raio de corte de 11 Å. Interações eletrostáticas de longo alcance foram calculadas empregando o algoritmo SPME (do inglês *smooth Particle-mesh Ewald*) (Darden *et al.*, 1993; Essmann *et al.*, 1995) e correções de longo alcance do termo repulsivo da

equação de Lennard-Jones para interações dispersivas foram aplicadas à energia e à pressão. Movimentos de estiramento e flexão das ligações foram restritos pelo algoritmo P-LINCS (do inglês *Parallel Linear Constraint Solver*) (Hess, 2008). Condições periódicas de contorno foram aplicadas em todas as direções e as equações de movimento clássicas foram integradas utilizando o algoritmo *leapfrog*.

Os dados do sistema simulado foram armazenados a cada 2 ps de simulação. A equilibração e amostragem foram conduzidas com o pacote computacional GROMACS (Abraham *et al.*, 2015), versão 2019.2, e o campo de força *all-atom* OPLS (Jorgensen *et al.*, 1996) foi utilizado para descrever os parâmetros mecânico-moleculares das partículas simuladas. O modelo de água TIP4P (do inglês, *transferable intermolecular potential with 4 points*) (Jorgensen *et al.*, 1983) foi escolhido como solvente, e íons sódio e cloreto foram adicionados para neutralizar a carga total do sistema. Quando foi necessário obter uma estrutura representativa do sistema, as conformações amostradas na trajetória foram agrupadas por meio do método de clusterização hierárquica aglomerativa, baseada na matriz de distâncias RMSD par-a-par, empregando o critério de *average linkage*, e a estrutura medóide foi selecionada.

#### 2.1.4 **Identificação dos sítios de ligação**

Para a identificação de sítios de ligação potenciais e de resíduos críticos na superfície da TcP21, foram utilizadas três abordagens independentes (seções 2.1.4.1., 2.1.4.2. e 2.1.4.3.). A escolha dessa estratégia visou minimizar os vieses metodológicos associados a cada abordagem, bem como mitigar a limitação de acurácia decorrente da ausência de validação experimental. Os resultados provenientes dessas abordagens foram analisados de forma integrada, com o intuito de aumentar a robustez e a confiabilidade das predições.

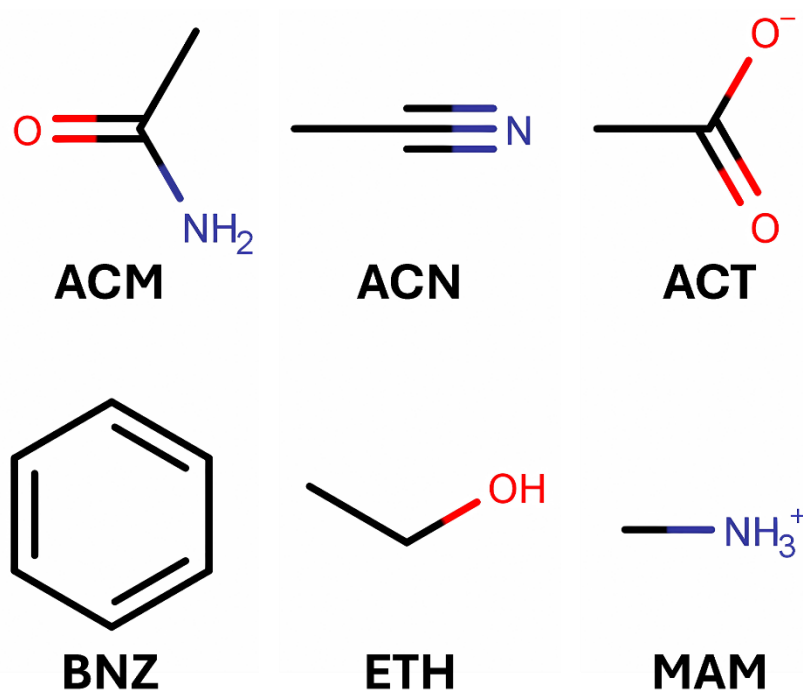
### 2.1.4.1 Simulações de MSMD

#### (a) Solventes orgânicos e descrição do Campo de Força

Nas simulações de MSMD, foram utilizados seis solventes orgânicos: acetamida (ACM), acetonitrila (ACN), íon acetato (ACT), benzeno (BNZ), etanol (ETH) e íon metilamônio (MAM), conforme ilustrado na Figura 2.3. A escolha desses solventes levou em consideração a miscibilidade em água, a diversidade de interações intermoleculares (incluindo contribuições polares, apolares e iônicas) e a frequência de uso observada na literatura (Arcon *et al.*, 2017; Graham *et al.*, 2018; Raman *et al.*, 2013).

As conformações iniciais, bem como os parâmetros mecânico-moleculares dos solventes, foram obtidos a partir do servidor de parametrização automática OPLS para ligantes orgânicos (Dodda *et al.*, 2017). Os estados de protonação dos solventes foram definidos com base no pH fisiológico ( $\approx 7,40$ ).

**Figura 2.3** – Solventes orgânicos utilizados como sondas moleculares nas simulações de MSMD.



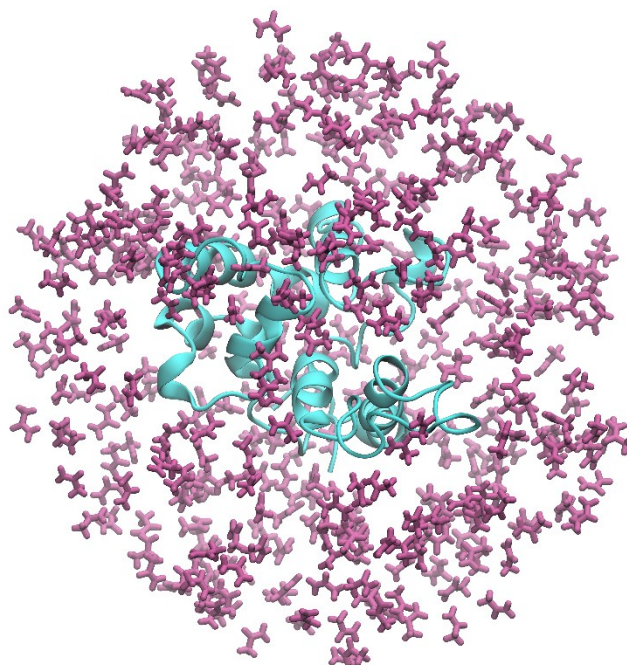
Fonte: Autor, 2026.

### (b) Configuração dos sistemas

A construção das configurações iniciais dos sistemas foi realizada utilizando a ferramenta PACKMOL (Martínez *et al.*, 2009). A concentração dos solventes orgânicos foi definida como 5% (V/V.), conforme recomendado na literatura para favorecer a identificação de sítios de ligação verdadeiros e minimizar a incidência de sítios de ligação falso positivos (Chan *et al.*, 2021; Ung *et al.*, 2016).

Inicialmente, a conformação da proteína TcP21, obtida na seção 2.1.2., foi posicionada no centro de uma caixa de simulação. Os solventes orgânicos foram então inseridos até atingir a proporção volumétrica definida, distribuídos aleatoriamente em uma região esférica de 40 Å de raio ao redor da proteína, conforme ilustrado na Figura 2.4, de modo a favorecer os contatos com a superfície proteica.

**Figura 2.4** – Representação esquemática da configuração dos sistemas. O solvente orgânico é mostrado em malva e as moléculas de água foram omitidas para maior clareza.



Fonte: Autor, 2026.

O número de moléculas adicionadas de cada componente, assim como suas massas molares, são descritos na Tabela 2.1. Em cada sistema, apenas um tipo de

solvente orgânico foi considerado em solução aquosa, sem a inclusão de misturas entre solventes distintos.

**Tabela 2.1** – Número de moléculas adicionadas para a concentração de 5% (V/V.) e suas respectivas massas molares, para cada sistema.

| <b>Solvente Orgânico</b> | <b>Nº de moléculas (solvente)</b> | <b>Nº de moléculas (água)</b> | <b>Massa Molar (g mol<sup>-1</sup>)</b> |
|--------------------------|-----------------------------------|-------------------------------|-----------------------------------------|
| Acetamida (ACM)          | 371                               | 20.034                        | 59,07                                   |
| Acetonitrila (ACN)       | 364                               | 20.020                        | 41,05                                   |
| Íon acetato (ACT)        | 339                               | 20.001                        | 59,04                                   |
| Benzeno (BNZ)            | 213                               | 20.022                        | 78,11                                   |
| Etanol (ETH)             | 328                               | 20.008                        | 46,07                                   |
| Íon metilamônio (MAM)    | 393                               | 20.043                        | 32,07                                   |

*Nota:* Água: 18,015 g mol<sup>-1</sup>.

Fonte: Autor, 2026.

### (c) *Parâmetros das simulações de MSMD*

As simulações de MSMD foram realizadas em triplicata para cada sistema contendo solvente orgânico em solução aquosa, resultando em um total de 18 sistemas simulados e um tempo de simulação acumulado de 1,8  $\mu$ s. O número de partículas simuladas por sistema variou entre 62.528 e 66.143.

#### 2.1.4.2 **Ancoragem da sonda farmacofórica**

Os solventes orgânicos empregados nas simulações de MSMD foram utilizados como representações simplificadas de grupos farmacofóricos frequentemente encontrados em candidatos a fármacos (Muegge, 2002; Zuccotto, 2003). Contudo, essa representação apresenta limitações, uma vez que não incorpora efeitos estéricos nem a interação competitiva entre diferentes grupos funcionais presentes em uma mesma molécula, importantes para o reconhecimento molecular.

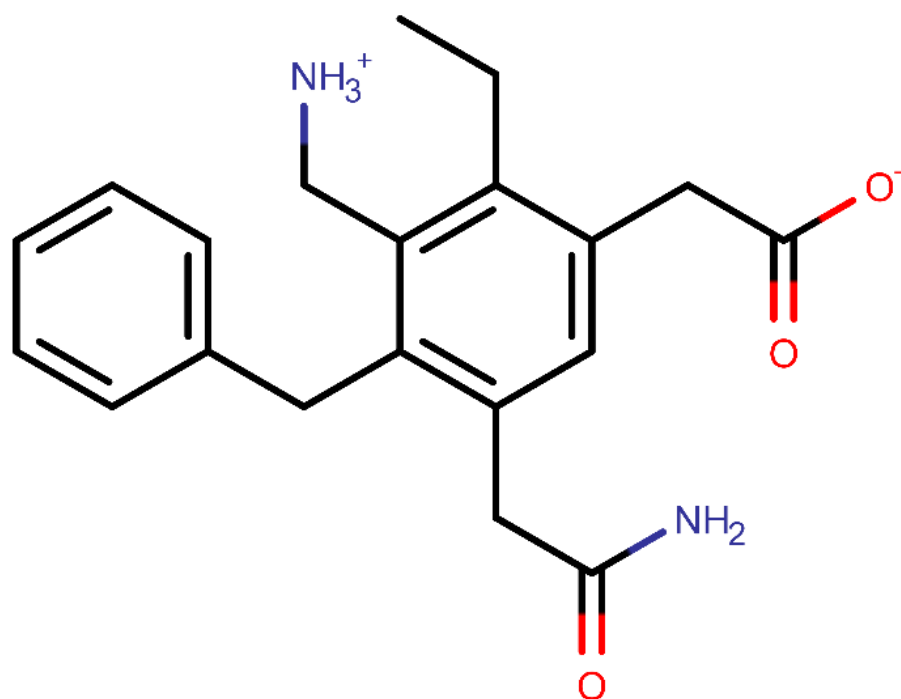
Com o objetivo de suprir essas limitações, foi proposta uma sonda farmacofórica hipotética. Embora não represente um ligante real, essa sonda permitiu investigar potenciais sítios de interação da TcP21 sob uma perspectiva alternativa,

aproximando-se do comportamento de um ligante farmacofórico. Essa estratégia se baseou em metodologias consolidadas no contexto de design de fármacos, como as descritas por Giordano *et al.* (Giordano *et al.*, 2022).

(a) *Sonda farmacofórica e descrição do Campo de Força*

A sonda farmacofórica, apresentada na Figura 2.5, foi modelada de forma a contemplar diferentes tipos de grupos funcionais, incluindo regiões hidrofóbicas (aromática e alifática), além de grupos carboxilato, acetamido e amino protonado, em analogia aos solventes empregados nas simulações de MSMD. Os estados de protonação foram definidos com base no pH fisiológico ( $\approx 7,40$ ).

**Figura 2.5** – Representação do modelo da sonda farmacofórica utilizado no estudo.



Fonte: Autor, 2026.

A geometria molecular da sonda foi inicialmente otimizada no programa Gaussian 03 (Gaussian, Inc., Wallingford CT, 2016), utilizando o modelo teórico composto pelo funcional B3LYP (Becke, 1988; Lee, Chengteh *et al.*, 1988; Vosko *et al.*, 1980) e o conjunto de base 6-31G (Hariharan; Pople, 1973; Hehre *et al.*, 1972). As



cargas parciais atômicas foram obtidas pelo método CHELPG (Breneman; Wiberg, 1990), enquanto os demais parâmetros de campo de força foram obtidos por meio do servidor OPLS para ligantes orgânicos.

Posteriormente à otimização geométrica, a sonda farmacofórica foi submetida a simulações de DM em meio aquoso para equilíbrio termodinâmica, em um sistema contendo 12.157 partículas. Todas as conformações amostradas foram agrupadas por análise de cluster, sendo selecionada uma conformação representativa para os estudos de ancoragem molecular.

#### *(b) Ancoragem molecular da sonda farmacofórica*

A ancoragem molecular da sonda farmacofórica foi realizada por meio do programa AutoDock Vina (Eberhardt *et al.*, 2021; Trott; Olson, 2010). O espaço de busca conformacional foi definido em uma região com dimensões de  $120 \times 100 \times 126$  Å ao redor da proteína, assegurando-se uma ancoragem às cegas.

O parâmetro de exaustividade foi estabelecido em 32, e o valor de corte da energia de afinidade baseada em função de pontuação foi definido em  $7 \text{ kcal mol}^{-1}$ . Foram realizadas vinte rodadas independentes de ancoragem, cada uma utilizando uma *seed* aleatória independente. Em cada rodada, foram obtidos vinte modos de ligação, resultando em um total de 400 conformações amostradas.

#### *(c) Parâmetros das simulações de DM*

A conformação correspondente ao modo de ligação de menor energia de afinidade foi selecionada para as simulações de DM. As simulações foram conduzidas em três réplicas independentes, resultando em um tempo de simulação acumulado de 300 ns. O sistema simulado apresentou um total de 63.543 partículas.

#### 2.1.4.3 **Ancoragem molecular baseada em fragmentos**

Como complemento às abordagens anteriores, foi utilizada uma metodologia automatizada para predição de sítios de ligação, por meio do servidor FTSite (Ngan, Chi-Ho *et al.*, 2012). Os cálculos foram realizados a partir da estrutura da TcP21 descrita na seção 3.2.

O método adotado pelo servidor baseia-se na ancoragem de 16 sondas orgânicas de baixo peso molecular (etano, etanol, álcool isopropílico, 2-metilpropan-2-ol, acetonitrila, metilamina, N,N-dimetilformamida, metil éter, benzaldeído, benzeno, cicloexano, fenol, acetamida, acetona, acetaldeído e ureia) na superfície rígida da proteína.

Após a ancoragem, as sondas são agrupadas de acordo com critérios energéticos, e os sítios de ligação são determinados com base na concentração de interações intermoleculares entre a proteína e as sondas moleculares em cada agrupamento. Três regiões de maior probabilidade foram fornecidas como resultado.

#### 2.1.5 **Análise dos sítios de ligação preditos**

Os resultados obtidos a partir das simulações de MSMD, da ancoragem molecular de sonda farmacofórica seguida de DM e da predição automatizada por servidor foram analisados por meio de mapas volumétricos de frequência de ocupação. Esses mapas, derivados de funções de distribuição espacial, permitem representar, de forma global, as regiões da superfície da proteína mais frequentemente acessadas pelas sondas.

Com exceção dos mapas gerados pela predição por servidor, todos os demais foram obtidos utilizando o programa VMD (Humphrey *et al.*, 1996). As frequências de ocupação foram calculadas considerando a totalidade de *frames* das trajetórias amostradas, correspondendo a um total de 50.000 conformações por sistema.

As interações intermoleculares entre a sonda farmacofórica e a TcP21, assim como os resíduos-chave associados a essas interações, foram identificados por meio

do programa Discovery Studio (BIOVIA, Dassault Systèmes, Discovery Studio, v.24, San Diego, 2023).

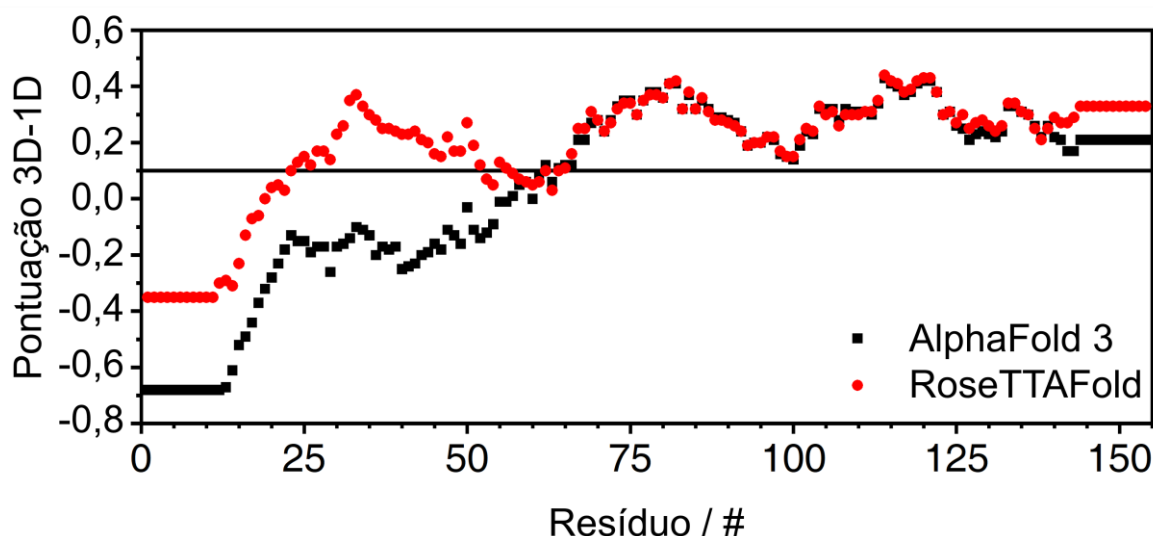
## 2.2 RESULTADOS E DISCUSSÃO

### 2.2.1 Avaliação preliminar dos modelos preditos

Os modelos estruturais da TcP21 gerados por RoseTTAFold e AlphaFold 3 foram submetidos a uma avaliação comparativa de qualidade utilizando os programas Verify3D, ERRAT e PROCHECK. Em função da alta similaridade entre os cinco modelos gerados por RoseTTAFold, apenas um deles foi utilizado como representante para a comparação com o modelo por AlphaFold 3.

A avaliação pelo Verify3D, apresentada na Figura 2.6, indicou que o modelo por RoseTTAFold apresentou 81% dos resíduos com pontuação 3D-1D  $\geq 0,1$ , enquanto 31 resíduos situaram-se abaixo do limite de qualidade, concentrando-se principalmente na região N-terminal (resíduos 1–22). Em contraste, o modelo por AlphaFold 3 apresentou apenas 60% dos resíduos acima do critério aceitável, com 62 resíduos abaixo do limite de qualidade, compreendendo a região N-terminal até o resíduo 62. Baseando-se somente no perfil de pontuação 3D-1D, apenas o modelo por RoseTTAFold satisfaz o critério mínimo ( $\geq 80\%$ ) de qualidade estrutural.

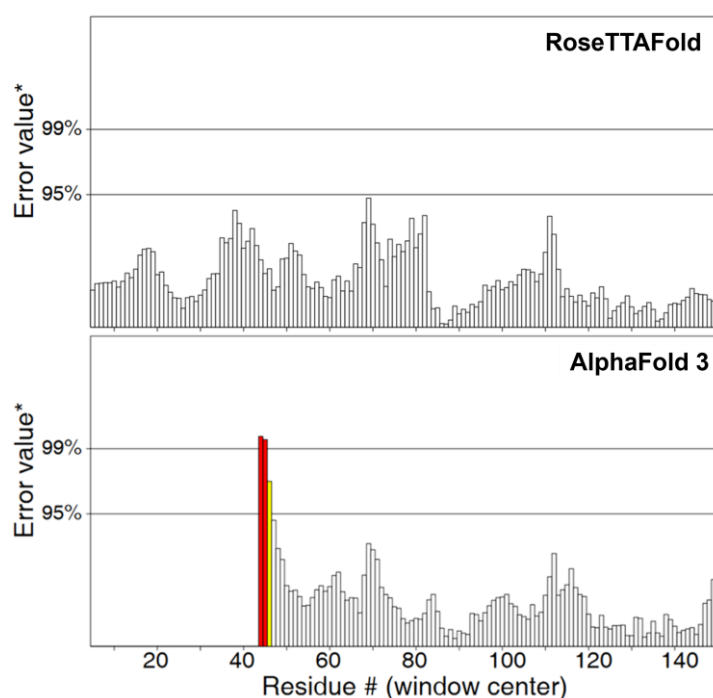
**Figura 2.6** – Pontuações 3D-1D por resíduo para os modelos gerados por AlphaFold 3 e RoseTTAFold.



Fonte: Autor, 2026.

Os resultados do ERRAT, apresentados na Figura 2.7, evidenciaram diferenças significativas na qualidade das interações não-covalentes. O modelo por RoseTTAFold apresentou um fator de qualidade de 100%, com todos os resíduos abaixo do limite de confiança do erro de 95%. Já o modelo por AlphaFold 3 apresentou um fator de 81% e um perfil heterogêneo, correspondente a resíduos com diferentes níveis de confiabilidade. O resíduo 46 foi classificado como potencialmente questionável (entre 95 e 99%), enquanto os resíduos 44–45 excederam o limite de 99% de confiança do erro, indicando possíveis erros estruturais. Adicionalmente, a região entre os resíduos 5–43 apresentou baixa densidade de interações não-covalentes, sugerindo instabilidade estrutural local.

**Figura 2.7** – Perfil de erro por resíduo (ERRAT) para os modelos gerados por AlphaFold 3 e RoseTTAFold, indicando regiões normais (branco), de alerta (amarelo) e de rejeição estatística (vermelho).



Fonte: Adaptado de ERRAT, 2026.

A análise qualitativa de distribuição percentual de resíduos no diagrama de Ramachandran, mostrada na Tabela 2.2, mostrou que ambos os modelos apresentaram distribuição adequada de ângulos diedros, sem ocorrência de resíduos

em regiões proibidas. Os resultados indicam que, do ponto de vista estereoquímico, ambos os modelos são aceitáveis.

**Tabela 2.2** – Distribuição percentual de resíduos no diagrama de Ramachandran dos modelos gerados por AlphaFold 3 e RoseTTAFold.

|                                                    | RoseTTAFold |       | AlphaFold 3 |       |
|----------------------------------------------------|-------------|-------|-------------|-------|
| Resíduos em regiões mais favorecidas               | 131         | 96,3% | 130         | 95,6% |
| Resíduos em regiões adicionalmente permitidas      | 4           | 2,9%  | 6           | 4,4%  |
| Resíduos em regiões generosamente permitidas       | 1           | 0,7%  | 0           | 0,0%  |
| Resíduos em regiões não permitidas                 | 0           | 0,0%  | 0           | 0,0%  |
| <hr/>                                              |             |       |             |       |
| Número de resíduos não-glicina e não-prolina       | 136         | 100%  | 136         | 100%  |
| Número de resíduos terminais (excluindo Gly e Pro) | 2           |       | 2           |       |
| Número de resíduos de glicina                      | 10          |       | 10          |       |
| Número de resíduos de prolina                      | 6           |       | 6           |       |
| <hr/>                                              |             |       |             |       |
| Número total de resíduos                           | 154         |       | 154         |       |

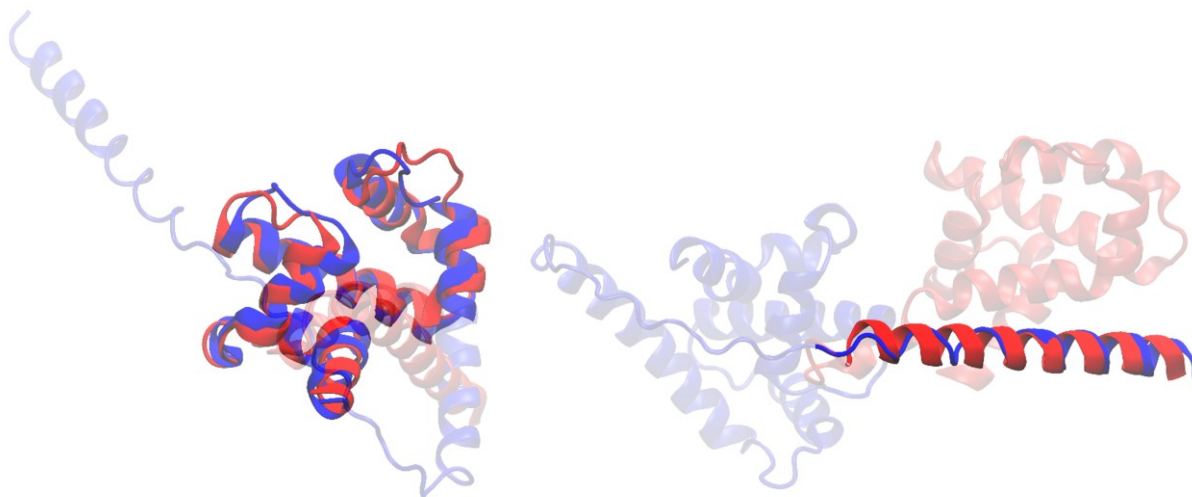
Fonte: Adaptado de PROCHECK, 2026.

A análise integrada dos resultados revelou que as principais discrepâncias estruturais estão associadas à região N-terminal, ainda que em magnitudes diferentes. O modelo gerado por AlphaFold 3 apresentou uma maior extensão de regiões com baixa qualidade estrutural, conforme indicado tanto pelo Verify3D quanto pelo ERRAT, enquanto o modelo por RoseTTAFold apresentou desvios mais localizados.

O alinhamento estrutural entre os modelos, apresentado na Figura 2.8, mostrou alta similaridade entre os resíduos 58–154 (RMSD de 2,13 Å), correspondendo a aproximadamente 63% da proteína. Em contraste, a região N-terminal expôs a divergência estrutural apontada anteriormente.

Enquanto o modelo por RoseTTAFold manteve uma organização helicoidal definida, o modelo por AlphaFold 3 apresentou perda gradual de estrutura ao longo dos resíduos 21–57. Embora ambos os modelos apresentem similaridade entre os resíduos 1–20 (RMSD de 2,10 Å), essa organização não foi mantida ao longo da região N-terminal no modelo por AlphaFold 3.

**Figura 2.8** – Alinhamentos estruturais parciais dos modelos gerados por RoseTTAFold (vermelho) e AlphaFold 3 (azul). Da esquerda para a direita: resíduos 58–154 e 1–20.



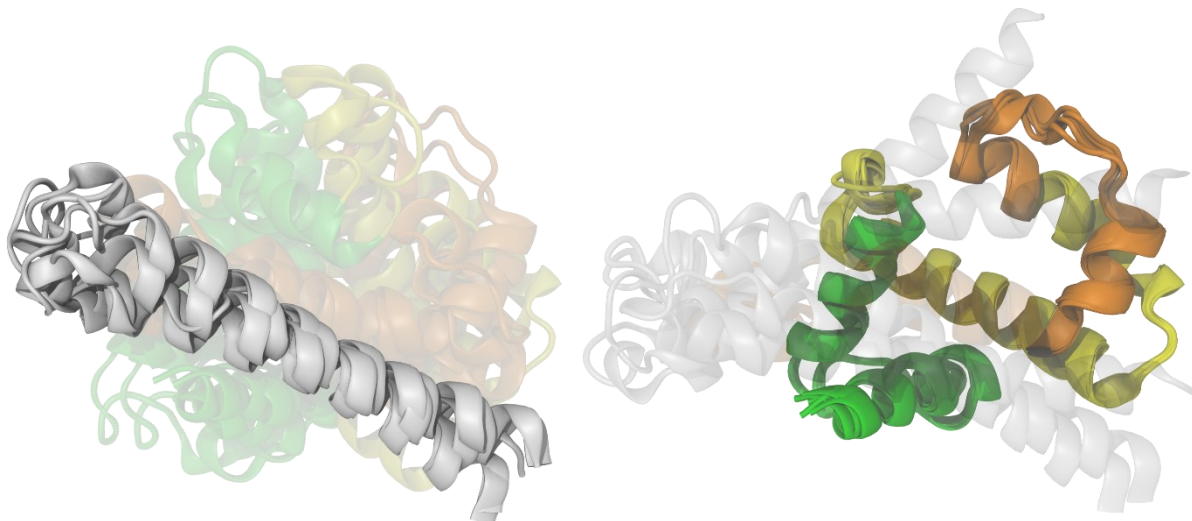
Fonte: Autor, 2026.

De forma geral, os resultados indicaram que, embora ambos os modelos apresentassem qualidade estereoquímica adequada, o modelo gerado por RoseTTAFold apresentou maior confiabilidade estrutural, particularmente na região N-terminal. Em função disso, optou-se pela utilização exclusiva desse modelo nas simulações subsequentes.

### 2.2.2 *Conformação inicial da TcP21 e equilibração termodinâmica*

Uma vez determinada a utilização do modelo gerado por RoseTTAFold, os cinco modelos candidatos foram sobrepostos a fim de avaliar a similaridade conformacional global entre as estruturas. A Figura 2.9 apresenta duas perspectivas de alinhamento, destacando as duas regiões de maior divergência, correspondentes aos resíduos 1–57 e 58–154. O esquema de cores adotado na Figura 2.9 foi adotado unicamente para clareza visual e não corresponde a nenhuma característica molecular específica. O mesmo esquema foi adotado para representar a TcP21 ao longo do texto. Os valores de RMSD foram determinados por meio de comparações par-a-par entre os modelos, sendo os resultados apresentados na Tabela 2.3.

**Figura 2.9** – Alinhamento estrutural dos modelos gerados por RoseTTAFold. Resíduos não alinhados são mostrados de forma transparente. Da esquerda para a direita: resíduos 57–154 e 1–57.



Fonte: Autor, 2026.

**Tabela 2.3** – Valores de RMSD dos modelos gerados por RoseTTAFold, calculados para toda a proteína, resíduos 1–57 e resíduos 58–154, relativos ao modelo 2.

| Modelo | Proteína | RMSD / Å      |                 |
|--------|----------|---------------|-----------------|
|        |          | Resíduos 1–57 | Resíduos 58–154 |
| 1      | 2,79     | 1,06          | 0,74            |
| 2      | 0,00     | 0,00          | 0,00            |
| 3      | 1,63     | 1,13          | 0,68            |
| 4      | 2,49     | 1,67          | 0,70            |
| 5      | 2,24     | 0,94          | 0,89            |

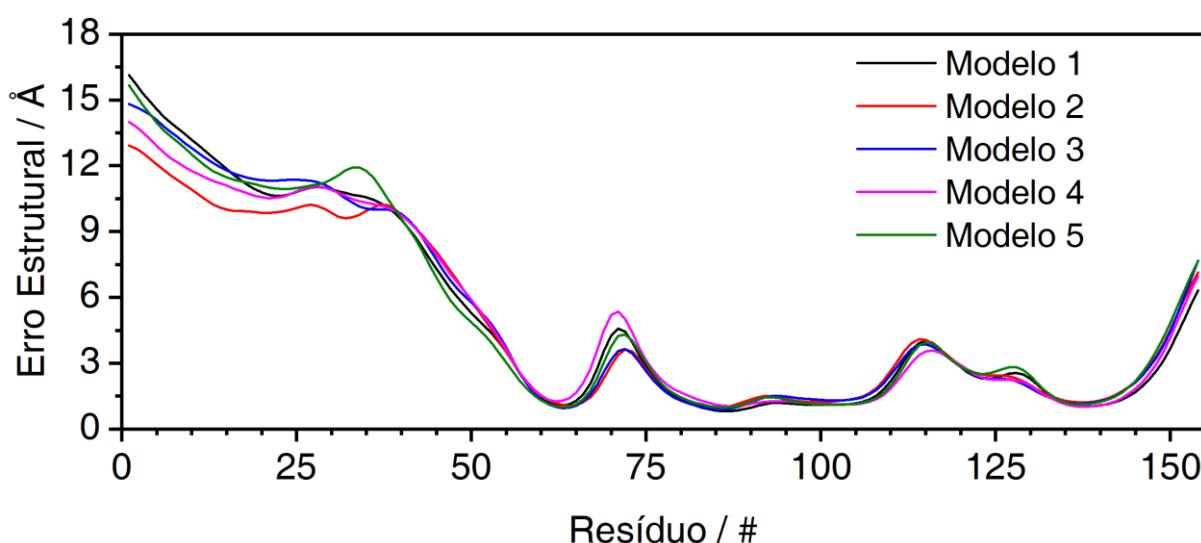
Fonte: Autor, 2026.

A análise dos valores de RMSD revelou desvios moderados para a estrutura completa (1,63–2,79 Å), ainda que compatíveis com a preservação de um mesmo enovelamento global. A região correspondente aos resíduos 58–154 apresentou baixos valores de RMSD (0,68–0,89 Å), sugerindo equivalência estrutural. Em contraste, a região N-terminal exibiu maior variabilidade (0,94 e 1,67 Å), em concordância com os resultados previamente obtidos na análise comparativa com o modelo gerado por AlphaFold 3. Em conjunto, esses resultados indicaram que os modelos gerados compartilham um arranjo estrutural global comum, com diferenças predominantemente localizadas na porção N-terminal.



Considerando a elevada similaridade observada, a seleção da conformação inicial da TcP21 para os experimentos subsequentes foi realizada considerando-se os critérios de erro estrutural estimado por resíduo (fornecido pelo próprio método RoseTTAFold) e a distribuição dos ângulos diedros no diagrama de Ramachandran, obtida com o PROCHECK. Os resultados correspondentes são apresentados na Figura 2.10 e Tabela 2.4.

**Figura 2.10** – Perfil de erro estrutural estimado dos resíduos preditos para os modelos gerados por RoseTTAFold.



Fonte: Autor, 2026.

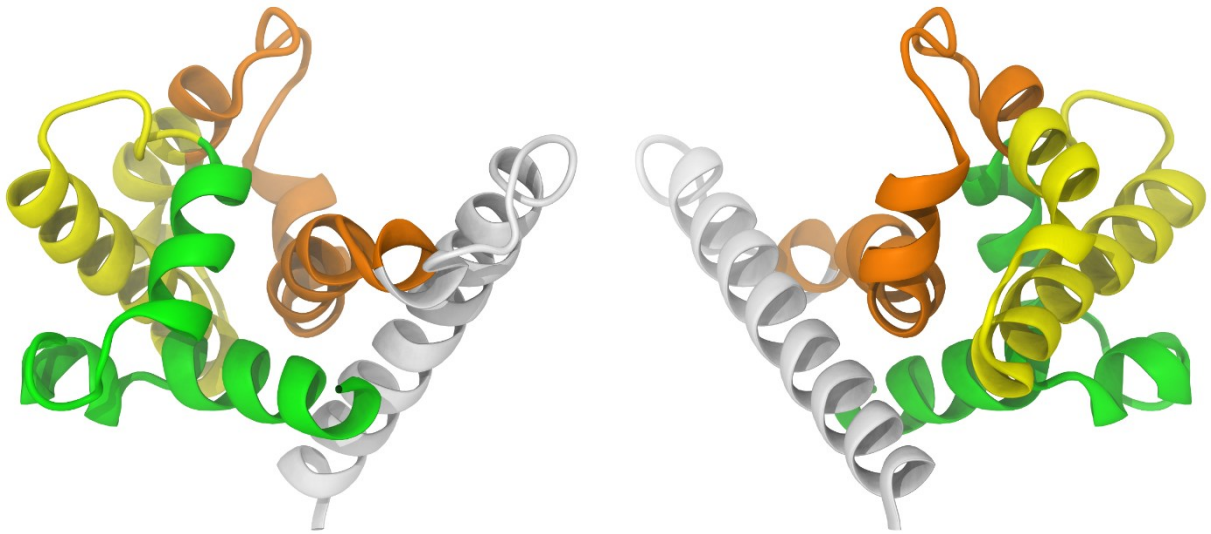
**Tabela 2.4** – Distribuição percentual de resíduos no diagrama de Ramachandran dos modelos gerados por RoseTTAFold.

|                                                    | Modelo 1 |       | Modelo 2 |       | Modelo 3 |       | Modelo 4 |       | Modelo 5 |       |
|----------------------------------------------------|----------|-------|----------|-------|----------|-------|----------|-------|----------|-------|
| Resíduos em regiões mais favorecidas               | 127      | 93,4% | 131      | 96,3% | 130      | 95,6% | 130      | 95,6% | 121      | 89,0% |
| Resíduos em regiões adicionalmente permitidas      | 8        | 5,9%  | 4        | 2,9%  | 5        | 3,7%  | 3        | 2,2%  | 8        | 5,9%  |
| Resíduos em regiões generosamente permitidas       | 1        | 0,7%  | 1        | 0,7%  | 0        | 0,0%  | 0        | 0,0%  | 5        | 3,7%  |
| Resíduos em regiões não permitidas                 | 0        | 0,0%  | 0        | 0,0%  | 1        | 0,7%  | 3        | 2,2%  | 2        | 1,5%  |
| Número de resíduos não-glicina e não-prolina       | 136      | 100%  | 136      | 100%  | 136      | 100%  | 136      | 100%  | 136      | 100%  |
| Número de resíduos terminais (excluindo Gly e Pro) | 2        |       | 2        |       | 2        |       | 2        |       | 2        |       |
| Número de resíduos de glicina                      | 10       |       | 10       |       | 10       |       | 10       |       | 10       |       |
| Número de resíduos de prolina                      | 6        |       | 6        |       | 6        |       | 6        |       | 6        |       |
| Número total de resíduos                           | 154      |       | 154      |       | 154      |       | 154      |       | 154      |       |

Fonte: Adaptado de PROCHECK, 2026.

Considerando o menor erro estrutural estimado por resíduos—sobretudo para os resíduos 1 a 40—e o maior percentual de resíduos em regiões favorecidas no diagrama de Ramachandran (96,3%), o modelo 2, mostrado na Figura 2.11, foi selecionado como conformação inicial.

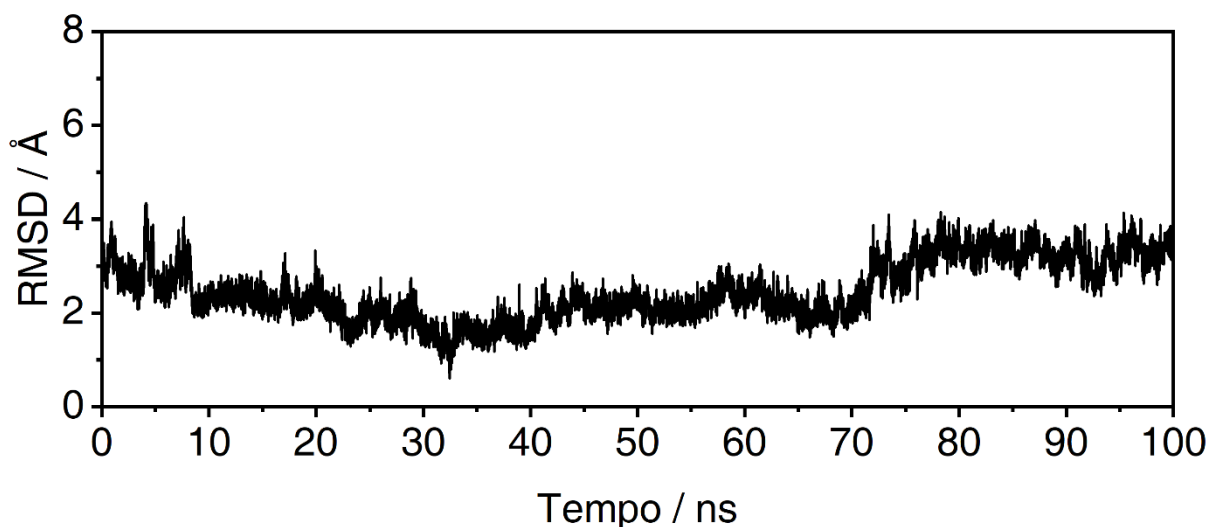
**Figura 2.11** – Representação esquemática do modelo 2 gerado por RoseTTAFold. Da esquerda para a direita: vistas frontal e posterior.



Fonte: Autor, 2026.

Após a realização das simulações de DM, a estabilidade estrutural do modelo foi inicialmente avaliada por meio do perfil médio de RMSD da cadeia principal da TcP21, apresentado na Figura 2.12.

**Figura 2.12** – Perfil temporal médio de RMSD da cadeia principal da TcP21.

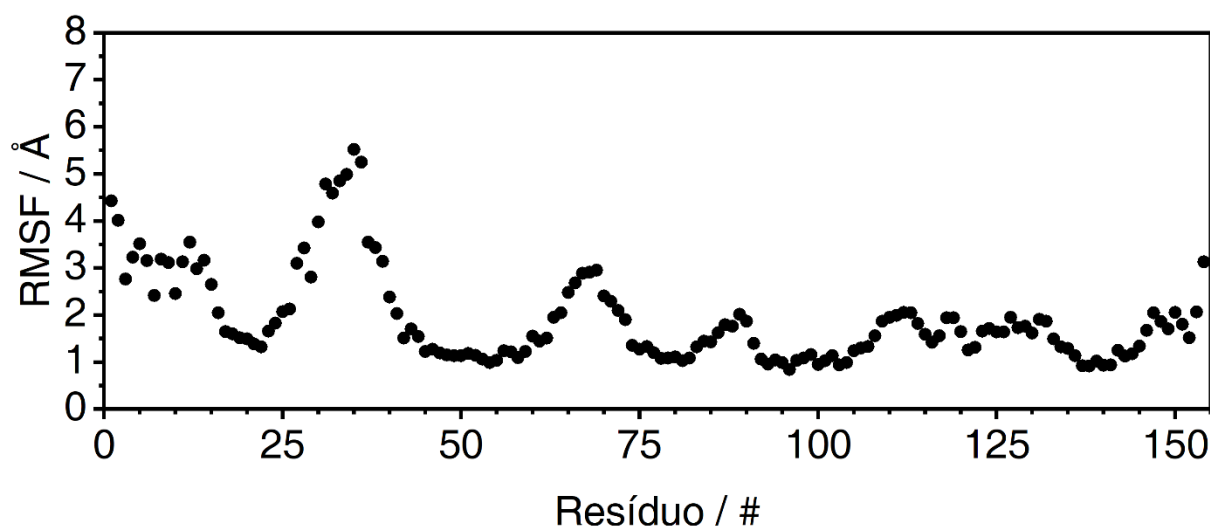


Fonte: Autor, 2026.

Conforme observado, a conformação inicial passa por rearranjos estruturais ao longo da trajetória, evidenciando a influência do ambiente químico na conformação da proteína. O valor médio de RMSD foi de  $2,46 \pm 0,63$  Å. Os resultados observados são consistentes com a natureza idealizada da estrutura gerada pelo RoseTTAFold, dado que o método empregado não considera explicitamente fatores químicos, como solvatação, e termodinâmicos, como temperatura e pressão.

Adicionalmente, foi calculado o perfil médio de RMSF a nível residual, apresentado na Figura 2.13, com o objetivo de caracterizar a flexibilidade estrutural da proteína.

**Figura 2.13** – Perfil de RMSF dos resíduos da cadeia principal da TcP21.

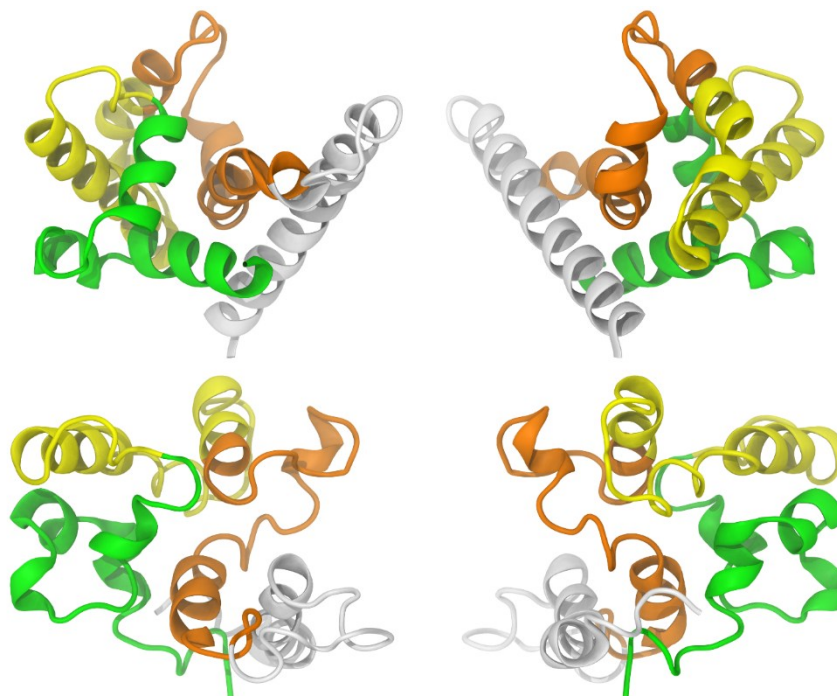


Fonte: Autor, 2026.

Os resultados indicaram regiões de flexibilidade consideravelmente superiores a outras, notadamente nos intervalos dos resíduos 1–15, 25–45 e 60–75. Inseridos nessas regiões, os resíduos 1–5, 25–44 e 65–73 foram identificados como voltas, caracterizando estruturas secundárias não definidas. A elevada flexibilidade dessas regiões contribui para os valores de RMSD observados, uma vez que esse parâmetro representa uma média global da proteína. Nesse contexto, os perfis de RMSD e RMSF mostraram-se coerentes e sugeriram uma convergência estrutural relativa ao longo da simulação.

Para a obtenção de uma estrutura representativa da TcP21, todas as conformações amostradas foram agrupadas segundo suas similaridades geométricas, sendo selecionada a estrutura central do conjunto, apresentada na Figura 2.14 em comparação com o modelo 2.

**Figura 2.14** – Representação esquemática dos modelos da TcP21. De cima para baixo: modelo 2 e estrutura representativa. Da esquerda para a direita: vistas frontal e posterior.



Fonte: Autor, 2026.

A inspeção qualitativa dessa estrutura revelou a perda parcial de algumas estruturas secundárias inicialmente previstas, em decorrência da flexibilidade estrutural da proteína em solução e de suas interações com o solvente. Apesar dessas alterações, a forma globular e alguns outros elementos conformacionais, como determinadas hélices, permaneceram preservados ao longo da simulação.

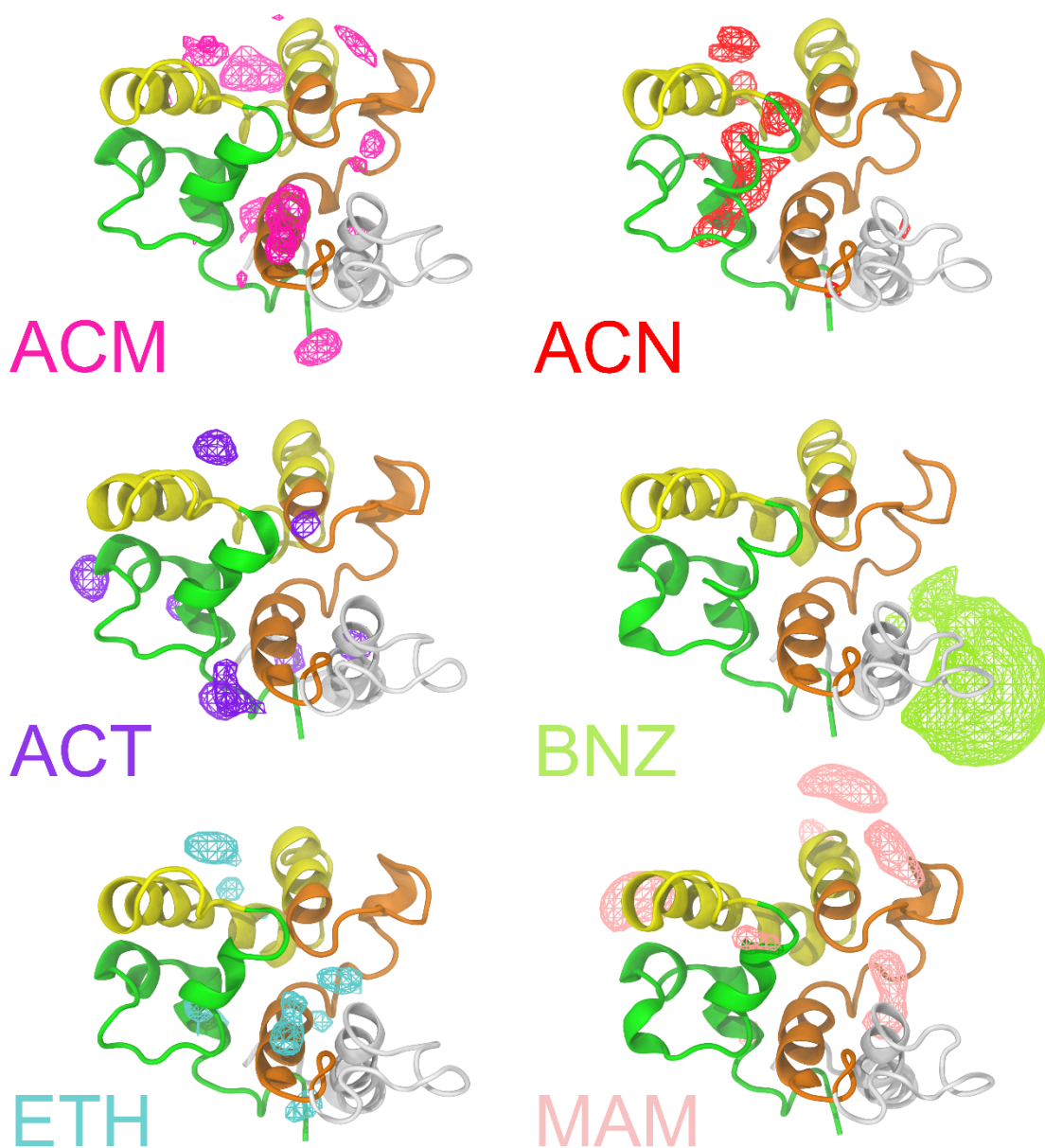
### 2.2.3 Identificação de sítios de ligação

#### 2.2.3.1 Simulações de MSMD

As simulações em solventes mistos foram utilizadas como etapa inicial na identificação de sítios de ligação. Dado o seu mapeamento não enviesado, resultante da maior liberdade conformacional e translacional dos solventes orgânicos, essa abordagem favoreceu a identificação de múltiplas áreas potenciais de interação.

Os sítios de ligação foram inferidos a partir de mapas volumétricos construídos com base nas regiões de maior frequência de ocupação de cada solvente orgânico, considerando suas três réplicas. Os resultados são apresentados na Figura 2.15, organizados por tipo de sonda molecular utilizada. Cada mapa volumétrico representa a combinação das triplicatas, incorporando tanto a sobreposição quanto as recorrências das regiões ocupadas.

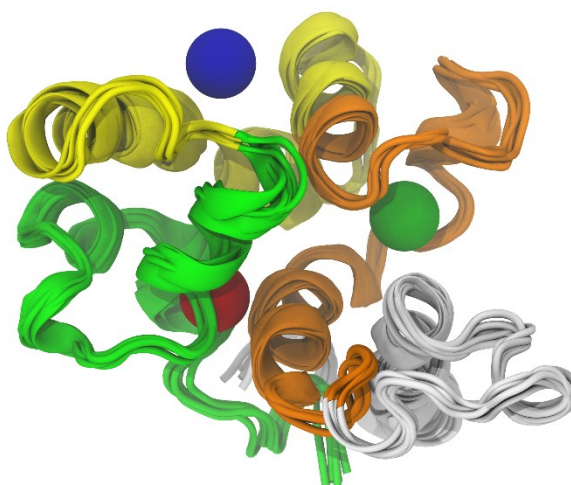
**Figura 2.15** – Mapas volumétricos dos respectivos solventes orgânicos: acetamida (ACM), acetonitrila (ACN), íon acetato (ACT), benzeno (BNZ), etanol (ETH) e íon metilamônio (MAM).



A análise da Figura 2.15 indicou que as sondas moleculares, em sua maioria, interagiram com regiões próximas à proteína, tanto em áreas superficiais quanto internas. O BNZ constituiu uma exceção, agregando-se em uma única região externa, associada a um fragmento hidrofóbico da TcP21. Considerando a heterogeneidade das propriedades físico-químicas dos solventes orgânicos utilizados, não se esperava um padrão uniforme de ocupação. Ainda assim, foram observadas regiões de convergência entre diferentes sondas.

Com base em uma avaliação qualitativa, três regiões principais de correlação foram identificadas e representadas na Figura 2.16 por esferas coloridas, juntamente com a sobreposição das conformações centrais da TcP21 obtidas nos diferentes sistemas. Ressalta-se que as regiões identificadas não estabelecem uma ordem de probabilidade, mas indicam apenas sobreposição espacial entre as sondas.

**Figura 2.16** – Representação das regiões de maior correlação entre os solventes orgânicos e sobreposição das conformações da TcP21 para cada réplica.



Fonte: Autor, 2026.

As esferas azul e vermelha correspondem à sobreposição dos solventes ACM, ACN, ACT e ETH, sendo que a esfera vermelha se encontra completamente inserida no interior da proteína. No caso do solvente ACT, a região correspondente na Figura 2.15 apresenta posicionamento ligeiramente posterior à hélice verde. A esfera verde representa a sobreposição dos solventes ACM, ETH e MAM. Em função das

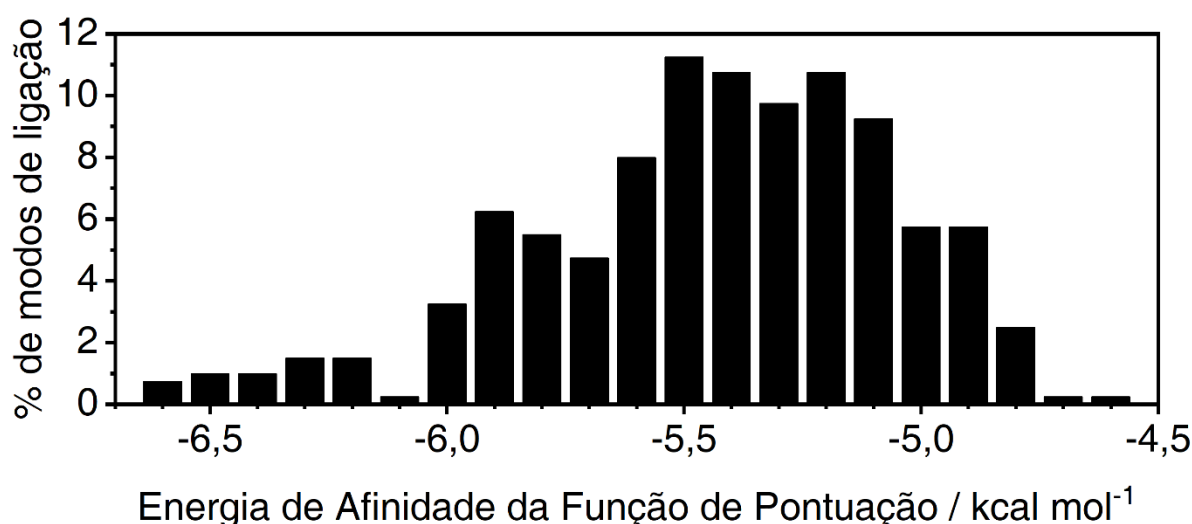
propriedades físico-químicas dos solventes, as regiões identificadas são consistentes com interações polares, apolares e iônicas (positiva ou negativa).

Dada a ausência de informações experimentais sobre sítios de ligação na TcP21, não foi possível validar diretamente as regiões identificadas. Dessa forma, a comparação com resultados obtidos por metodologias baseadas em diferentes princípios mostrou-se fundamental, de modo a aumentar a robustez das conclusões por meio da correlação entre elas.

### 2.2.3.2 *Ancoragem da sonda farmacofórica*

A ancoragem molecular da sonda farmacofórica foi empregada como segunda abordagem para a identificação de sítios de ligação na TcP21. Dos 400 modos de ligação obtidos originalmente, a energia de afinidade baseada em função de pontuação variou entre -4,6 e -6,6 kcal mol<sup>-1</sup>. A distribuição percentual dos modos de ligação em relação à energia de afinidade é apresentada na Figura 2.17.

**Figura 2.17** – Distribuição percentual dos modos de ligação obtidos pelo AutoDock Vina com base na energia de afinidade.

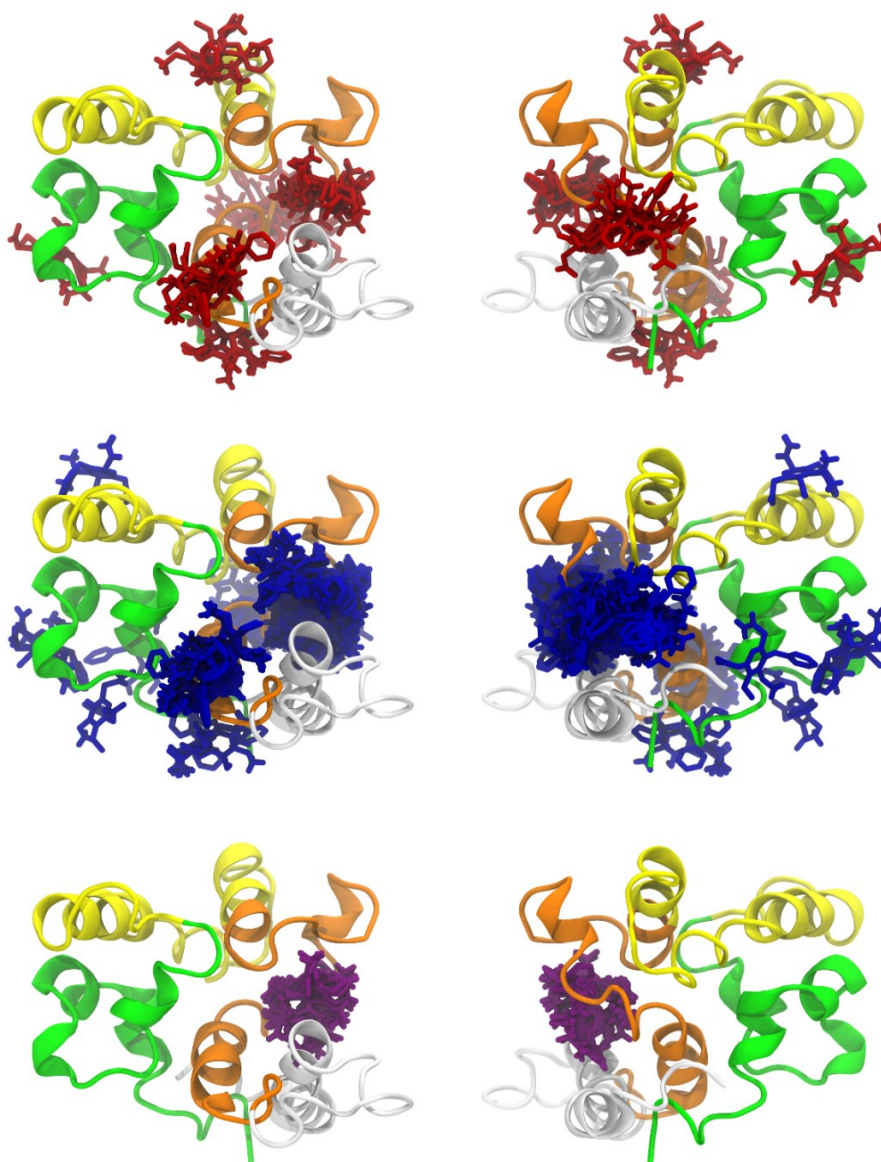


Fonte: Autor, 2026.



Verifica-se que a maior parte dos modos de ligação se concentrou no intervalo de  $-5,0$  e  $-5,9$  kcal mol<sup>-1</sup>, correspondendo a 82% do total amostrado, sendo  $-5,5$  kcal mol<sup>-1</sup> o valor mais frequente. A energia de afinidade mais negativa foi de  $-6,6$  kcal mol<sup>-1</sup>, observada em apenas três modos de ligação (0,75%), oriundos de rodadas distintas. Com o objetivo de analisar a distribuição espacial, os modos foram classificados em três grupos energéticos, conforme ilustrado na Figura 2.18, sendo estes: (1) superior a  $-5,0$  kcal mol<sup>-1</sup>, (2) entre  $-5,0$  e  $-5,9$  kcal mol<sup>-1</sup>, e (3) inferior a  $-5,9$  kcal mol<sup>-1</sup>.

**Figura 2.18** – Representações esquemáticas dos modos de ligação. Em vermelho, modos com energia superior a  $-5,0$  kcal mol<sup>-1</sup>; em azul, entre  $-5,0$  e  $-5,9$  kcal mol<sup>-1</sup>; e em roxo, inferior a  $-5,9$  kcal mol<sup>-1</sup>. Da esquerda para a direita: vistas frontal e posterior.

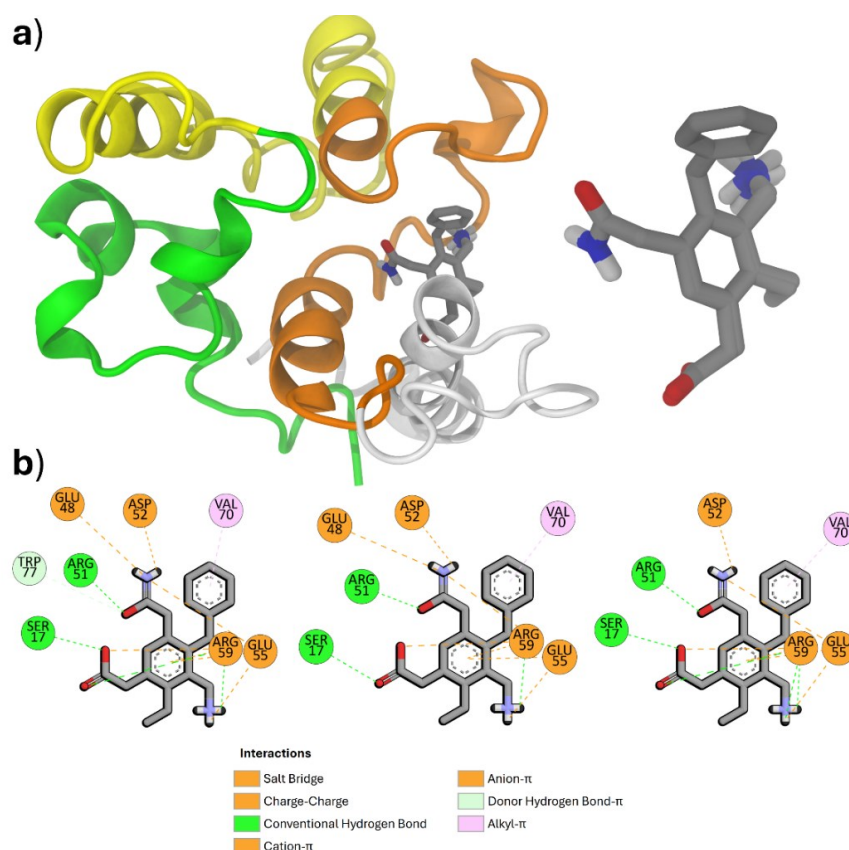


Os modos de ligação pertencentes aos grupos de maior energia (1) e (2), apresentaram distribuição mais difusa ao redor da proteína, enquanto aqueles do grupo de menor energia (3), concentraram-se em uma região específica. Embora essa região não seja exclusiva do grupo (3), a variância observada foi significativamente menor para os modos de energia de afinidade mais negativa.

A análise comparativa com os resultados das simulações de MSMD indicou correspondência espacial entre essa região e a região representada pela esfera verde previamente identificada (sobreposição dos solventes ACM, ETH e MAM), sugerindo consistência entre as metodologias empregadas. Com base nesses resultados, os três modos de menor energia ( $-6,6 \text{ kcal mol}^{-1}$ ) foram selecionados para análise detalhada.

A sobreposição desses modos revelou elevada similaridade, conforme mostrado na Figura 2.19a, com um valor máximo de RMSD de  $0,57 \text{ \AA}$ , com diferenças estruturais restritas à orientação do grupo amino protonado.

**Figura 2.19** – Modos de ligação de menor energia (a) e diagramas 2D das interações proteína-sonda farmacofórica (b).

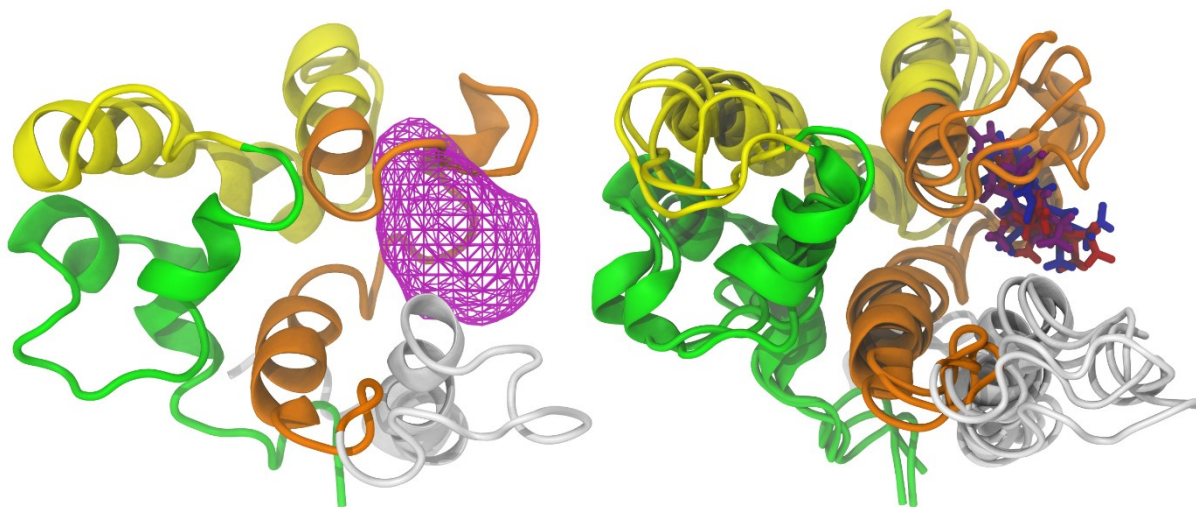


Fonte: Adaptado de Discovery Studio, 2026.

As interações intermoleculares entre as sondas farmacofóricas e a TcP21, mostradas na Figura 2.19b, foram classificadas por meio do programa Discovery Studio. As interações foram predominantemente eletrostáticas (incluindo ponte salina, carga-carga,  $\pi$ -cátion e  $\pi$ -ânion) e ligações de hidrogênio. Diferenças pontuais envolveram a participação dos resíduos TRP77 (interações do tipo  $\pi$ -doador) e GLU48 (interações do tipo carga-carga). Dada a semelhança global entre os modos, esses foram considerados equivalentes, e a seleção do sistema inicial para as simulações de DM foi realizada de forma arbitrária.

Após as simulações de DM, a estabilidade das interações entre a sonda farmacofórica e a TcP21 foi avaliada por meio de mapas volumétricos de ocupação e análise estrutural das réplicas, apresentados na Figura 2.20.

**Figura 2.20** – Representação dos complexos proteína-sonda farmacofórica. Da esquerda para a direita: mapa volumétrico da sonda e superimposição das conformações dos complexos.

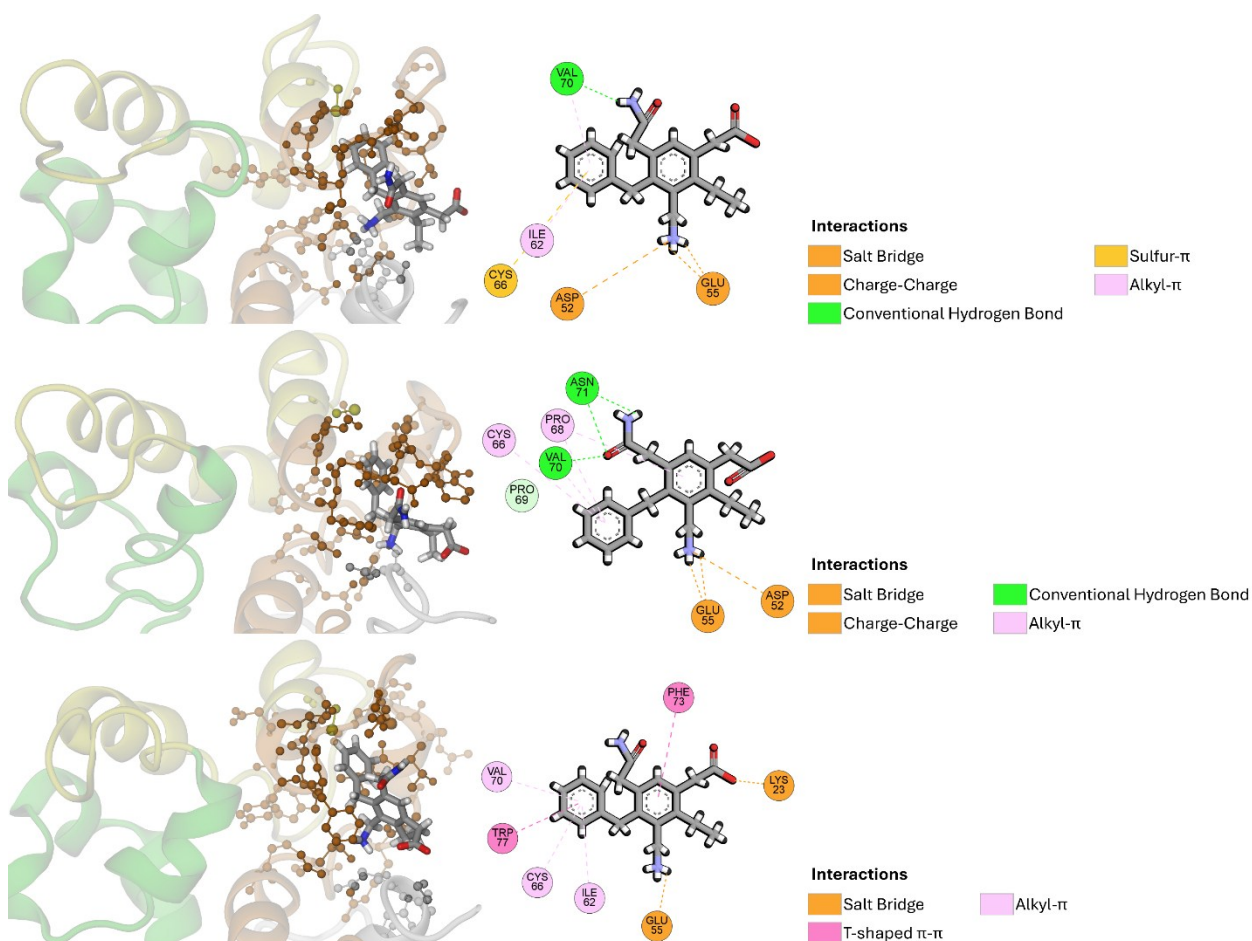


Fonte: Autor, 2026.

Em contraste aos mapas volumétricos observados nas simulações de MSMD, que consideraram a frequência de ocupação de múltiplas moléculas simultaneamente, a distribuição espacial observada para a sonda farmacofórica é substancialmente mais restrita. A permanência da sonda próxima ao sítio inicial ao longo das simulações sugere interações intermoleculares de relativa estabilidade. Nota-se que a correspondência espacial da região de interação permaneceu a mesma para aquela observada nas simulações de MSMD.

A análise de interações intermoleculares foi conduzida para as conformações centrais do complexo, em cada uma das três réplicas. Os diagramas 2D das interações, apresentados na Figura 2.21, revelaram variações entre as réplicas, mantendo, contudo, um núcleo comum de interações.

**Figura 2.21** – Representação das conformações centrais do complexo sonda farmacofórica-proteína para cada réplica e seus respectivos diagramas 2D de interação.



Fonte: Adaptado de Discovery Studio, 2026.

Os resíduos GLU55, CYS66 e VAL70 foram identificados de forma consistente, sugerindo um papel relevante na estabilização do complexo. Interações adicionais envolveram os resíduos ASP52, ILE62 e ASN71, dependendo da réplica analisada.

O resíduo GLU55 apresentou interações do tipo carga-carga e ponte salina com o grupo amino protonado em todas as réplicas. O resíduo CYS66 participou de interações com o grupo fenil, sendo estas do tipo  $\pi$ -enxofre na primeira réplica e  $\pi$ -

alquil nas últimas duas. Da mesma forma, o resíduo VAL70 interagiu com o grupo fenil por meio de interações  $\pi$ -alquil em ambas as réplicas, além de ligações de hidrogênio, atuando como aceptor para a amina do grupo acetamido ou doador para a carbonila do mesmo grupo. Nota-se que, quando o resíduo VAL70 atuou como doador para o grupo acetamido, o resíduo ASN71 também interagiu com o mesmo grupo, participando tanto como aceptor quanto doador nas ligações de hidrogênio, formando uma rede interativa entre eles. O resíduo ASP52 interagiu predominantemente por meio de interações do tipo carga-carga com o grupo amino protonado, enquanto o resíduo ILE62 participou de interações do tipo  $\pi$ -alquil com o grupo fenil.

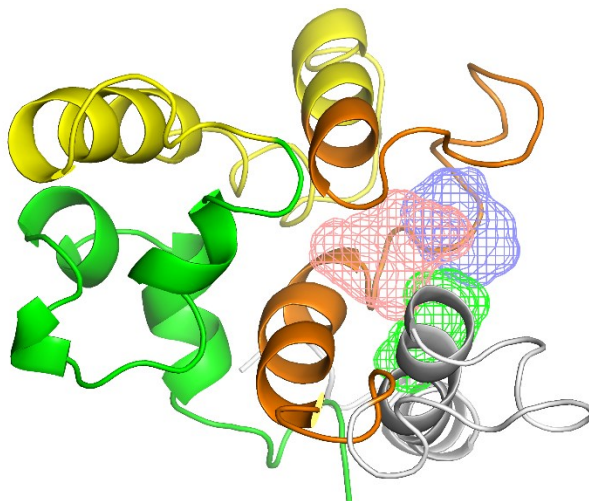
De forma geral, os resultados foram consistentes entre si, e sugeriram que os grupos amino protonado, acetamido e fenil atuaram como os principais elementos de interação com a proteína, enquanto os grupos carboxilato e alifático permaneceram orientados ao solvente, com menor participação direta nas interações.

### 2.2.3.3 *Ancoragem molecular baseada em fragmentos*

Como complemento às simulações de simulações de MSMD e à ancoragem da sonda farmacofórica, a terceira ferramenta preditiva empregada no estudo dos sítios de ligação na TcP21 foi a ancoragem molecular baseada em fragmentos, realizada por meio do servidor FTSite. Os resultados correspondentes são apresentados na Figura 2.22.



**Figura 2.22** – Sítios de ligação preditos pelo servidor FTSite, representados por malhas em cor (salmão, verde e azul).



Fonte: Autor, 2026.

Conforme apresentado na seção 2.1.4.3, os três sítios de ligação preditos pelo servidor FTSite correspondem aos clusters de maior correlação, segundo a concentração de interações intermoleculares entre a proteína e as sondas moleculares empregadas. As regiões são apresentadas nas cores salmão, verde e azul, em ordem decrescente do número de interações, correspondendo a uma correlação de 93,7%, 87,5%, e 62,5% das 16 sondas utilizadas, respectivamente. Embora apresentem diferenças quantitativas, essas regiões exibiram elevada similaridade espacial com aqueles identificadas pelas abordagens anteriores. A Tabela 2.5 lista os resíduos interativos da TcP21 associados a cada região, diferenciando-os entre resíduos recorrentes (presentes em múltiplas regiões) e resíduos exclusivos (presentes em apenas uma região).

**Tabela 2.5** – Resíduos interativos da TcP21, classificados quanto à recorrência entre as regiões de interação preditas pelo FTSite.

| Região Salmão               | Região Verde | Região Azul |
|-----------------------------|--------------|-------------|
| <i>Resíduos Recorrentes</i> |              |             |
| ASP52                       | ASP52        | –           |
| GLU55                       | GLU55        | GLU55       |
| ARG59                       | ARG59        | ARG59       |
| VAL70                       | –            | VAL70       |
| TRP77                       | –            | TRP77       |
| <i>Resíduos Exclusivos</i>  |              |             |
| –                           | ALA13        | –           |
| –                           | CYS14        | –           |
| –                           | SER15        | –           |
| –                           | VAL16        | –           |
| –                           | SER17        | –           |
| –                           | GLU20        | –           |
| ARG24                       | –            | –           |
| GLU48                       | –            | –           |
| ARG51                       | –            | –           |
| THR54                       | –            | –           |
| –                           | –            | CYS58       |
| –                           | –            | GLU60       |
| –                           | –            | GLU61       |
| –                           | –            | ILE62       |
| –                           | –            | THR63       |
| PHE73                       | –            | –           |

Fonte: Autor, 2026.

Entre os resíduos recorrentes, GLU55 e ARG59 destacaram-se por estarem presentes em todas as regiões preditas, enquanto ASP52, VAL70 e TRP77 foram observados em duas das três regiões. Entre os exclusivos, foi observada maior heterogeneidade, com resíduos distribuídos ao longo do segmento 13–73, contudo em uma proporção semelhante de 5 a 6 resíduos interativos por região.

#### 2.2.4 **Correlação dos sítios de ligação preditos**

A comparação com os dados obtidos por simulações de DM, apresentados na Tabela 2.6, evidenciou o GLU55 como o único resíduo recorrente em todas as abordagens consideradas, com frequência de ocorrência de 100%. Os resíduos ASP52 e VAL70 também apresentaram elevada recorrência, com frequências de

ocorrência de 67 e 83%, respectivamente. Por outro lado, CYS66 foi observado exclusivamente nas simulações de DM (todas as réplicas), enquanto ARG59 foi identificado apenas pelo FTSite (todos os clusters). Os resíduos TRP77 e ILE62 apresentaram ocorrência intermediária entre os métodos, e PHE73, baixa ocorrência.

**Tabela 2.6** – Resultados sumariados de resíduos interativos recorrentes nos respectivos ensaios realizados.

| DM (sonda farmacofórica) |                |                | Ancoragem (FTSite) |                |                | Frequência de ocorrência |
|--------------------------|----------------|----------------|--------------------|----------------|----------------|--------------------------|
| R <sub>1</sub>           | R <sub>2</sub> | R <sub>3</sub> | C <sub>1</sub>     | C <sub>2</sub> | C <sub>3</sub> |                          |
| ASP52                    | ASP52          | –              | ASP52              | ASP52          | –              | 0,67                     |
| GLU55                    | GLU55          | GLU55          | GLU55              | GLU55          | GLU55          | 1,00                     |
| –                        | –              | –              | ARG59              | ARG59          | ARG59          | 0,50*                    |
| ILE62                    | –              | ILE62          | –                  | –              | ILE62          | 0,50                     |
| CYS66                    | CYS66          | CYS66          | –                  | –              | –              | 0,50*                    |
| VAL70                    | VAL70          | VAL70          | VAL70              | –              | VAL70          | 0,83                     |
| –                        | –              | PHE73          | PHE73              | –              | –              | 0,33                     |
| –                        | –              | TRP77          | TRP77              | –              | TRP77          | 0,50                     |

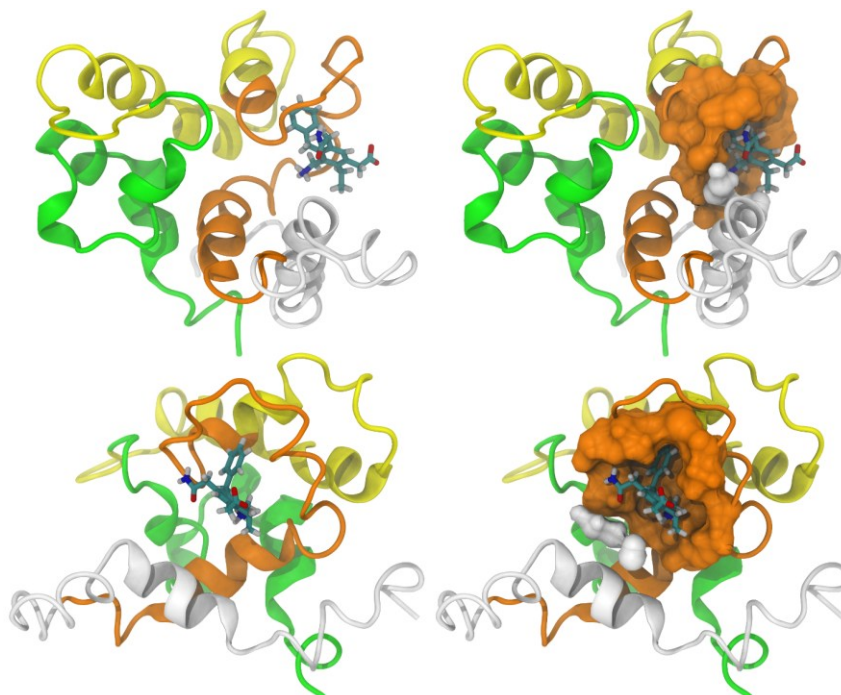
*Nota.* Os códigos de letra R e C representam “Réplica” e “Cluster”, respectivamente.

Fonte: Autor, 2026.

A evidente correspondência entre as simulações de MSMD, a ancoragem molecular da sonda farmacofórica seguida de simulações de DM e a ancoragem molecular baseada em fragmentos do servidor FTSite, possibilitou a proposição de um modelo qualitativo do sítio de ligação da TcP21, apresentado na Figura 2.23. A estrutura de referência utilizada para a ilustração do modelo corresponde a uma conformação representativa obtida a partir das simulações de DM com a sonda farmacofórica.



**Figura 2.23** – Modelo de sítio de ligação proposto, representado como superfície de van der Waals. De cima para baixo: vistas frontal e lateral.



Fonte: Autor, 2026.

O modelo proposto inclui os resíduos recorrentes observados entre as abordagens utilizadas, assim como aqueles localizados a um raio de 5 Å da sonda farmacofórica, permitindo uma caracterização mais abrangente do microambiente do sítio de ligação. Esse sítio se localiza na superfície da proteína TcP21 e não apresenta impedimentos estéricos significativos, sugerindo potencial acessibilidade a ligantes com características estruturais semelhantes às da sonda farmacofórica.

Na ausência de dados experimentais, não é possível validar diretamente a existência do sítio proposto nem inferir o tipo de modulação alostérica associada, caso existente. No entanto, a elevada convergência entre as metodologias confere robustez aos resultados observados, podendo fornecer detalhes moleculares relevantes para estudos futuros sobre a elucidação da TcP21 e sua participação na doença de Chagas.

Ademais, ressalta-se que os resultados deste estudo foram publicados na revista *Journal of Chemical Information and Modeling* (FI 5,3), sob o título “*Exploring Binding Sites in Chagas Disease Protein TcP21 Using Integrated Mixed Solvent Molecular Dynamics Approaches*” (Oliveira Soté; Comar Junior, 2025).

## 2.3 CONCLUSÃO

---

No presente estudo, foi inicialmente obtida uma conformação estrutural da proteína TcP21 por meio de ferramentas preditivas baseadas em Inteligência Artificial, sendo a estrutura gerada pelo método RoseTTAFold selecionada como modelo inicial por apresentar menor incerteza estrutural em relação à obtida pelo método AlphaFold 3. Foram integradas três abordagens computacionais independentes — simulações de dinâmica molecular em solventes mistos, ancoragem molecular baseada em fragmentos e ancoragem molecular de sonda farmacofórica seguida de simulações de dinâmica molecular — com o objetivo de identificar e caracterizar potenciais sítios de ligação na TcP21. A convergência dos resultados obtidos pelas diferentes metodologias permitiu a identificação de um sítio comum, cujos principais resíduos envolvidos foram GLU55 e ASP52 (hidrofílicos de caráter ácido), VAL70 e ILE62 (hidrofóbicos alifáticos) e TRP77 (hidrofóbico aromático). Esse sítio, localizado na superfície da TcP21, apresentou interações consistentes com os fragmentos amino protonado, acetamido e fenil da sonda farmacofórica empregada. As interações intermoleculares identificadas — incluindo pontes salinas, ligações de hidrogênio, interações carga-carga e  $\pi$ -alquil — estabeleceram uma rede de interações consistentes entre a sonda e a proteína, sugerindo a relevância desses grupos funcionais no reconhecimento molecular nesse sítio.

### 3 ESTUDO 2

---

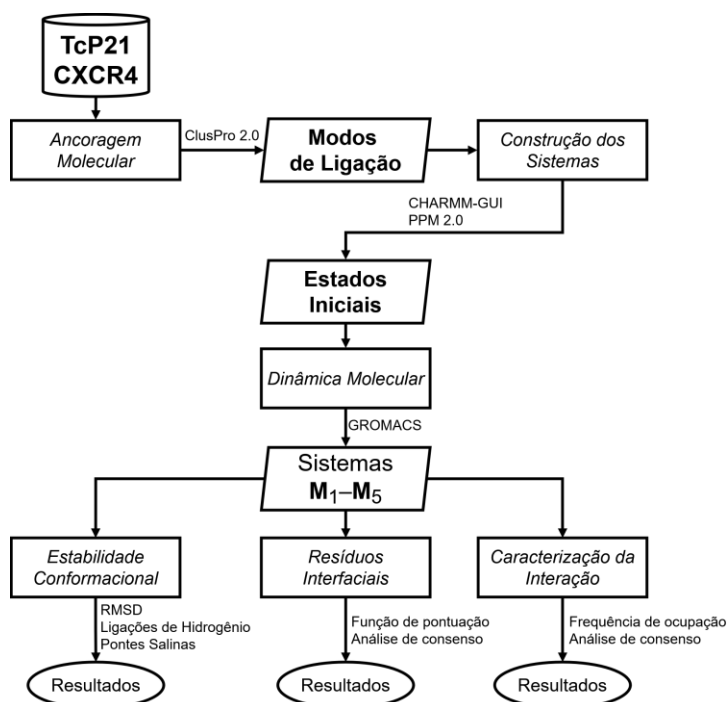
**Elucidação molecular da interface de interação  
TcP21–CXCR4 empregando análise de consenso  
intersistema**

## 3.1 METODOLOGIA

### 3.1.1 Fluxograma geral de trabalho

A metodologia adotada neste Capítulo, voltada à obtenção do complexo proteico e à caracterização dinâmica do reconhecimento molecular e dos contatos entre as proteínas TcP21 e CXCR4, foi organizada como um fluxograma de trabalho e apresentada de forma geral na Figura 3.1.

**Figura 3.1** – Fluxograma geral de trabalho do Estudo 2.



Fonte: Autor, 2026.

Em termos gerais, a metodologia se iniciou com a utilização das estruturas isoladas das proteínas de interesse para a obtenção de modos de ligação potenciais por meio de simulações de ancoragem molecular. A partir dessas configurações, cinco modos de ligação representativos foram utilizados na construção dos sistemas completos, incluindo solvatação, íons e bicamada lipídica, sendo então empregados como pontos de partida para simulações multirréplicas de dinâmica molecular. A

amostragem temporal das interações intermoleculares possibilitou investigar o comportamento dinâmico dos complexos proteicos em condições próximas às fisiológicas.

As trajetórias geradas foram analisadas segundo três eixos analíticos principais: (i) avaliação da estabilidade conformacional, (ii) identificação de resíduos-chave na interface de interação e (iii) caracterização do padrão de interações intermoleculares.

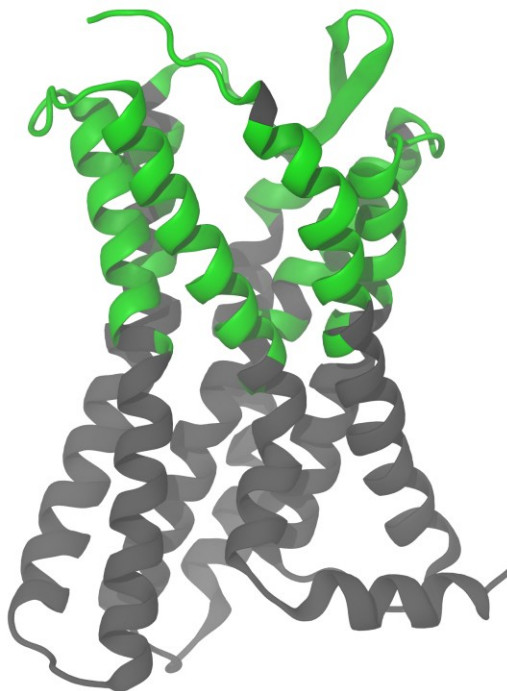
Os procedimentos metodológicos específicos, assim como os parâmetros e protocolos computacionais empregados em cada etapa, foram descritos nas subseções subsequentes.

### 3.1.2 **Ancoragem molecular**

Para o estudo da interação entre CXCR4 e TcP21, foi necessária a modelagem estrutural do complexo por meio da obtenção de modos de ligação, uma vez que não havia estruturas disponíveis na literatura. A TcP21 foi representada pela estrutura equilibrada obtida na seção 2.2.2 (Oliveira Soté; Comar Junior, 2025). A estrutura utilizada da CXCR4 correspondeu ao único modelo experimental disponível de seu estado ativo, complexada com CXCL12 (PDB ID: 8K3Z).

A ancoragem molecular foi realizada no servidor ClusPro 2.0 (Ashizawa *et al.*, 2026; Kozakov *et al.*, 2017). Os resíduos da CXCR4 previamente localizados a até 10 Å do ligante CXCL12 foram utilizados no servidor como referência para a seleção de resíduos classificados como “*Atrativos*”, servindo como restrições espaciais para a orientação da TcP21, levando em consideração o ambiente extracelular da membrana lipídica. Todos os demais resíduos foram utilizados como referências para a seleção de resíduos classificados como “*Repulsivos*”. A representação dos resíduos considerados “*Atrativos*” e “*Repulsivos*” é apresentada na Figura 3.2.

**Figura 3.2** – Representação dos resíduos considerados atrativos (verde) e repulsivos (cinza) durante a ancoragem molecular.



Fonte: Autor, 2026.

A determinação dos modos de ligação mais representativos, correspondentes aos medóides de cada cluster, foi realizada pelo servidor com base na função de pontuação classificada como “*Balanceada*”. As medóides dos cinco clusters mais populosos foram selecionadas para a construção dos sistemas.

### 3.1.3 **Construção dos sistemas**

Com base nos modos de ligação obtidos para o complexo TcP21–CXCR4, as configurações iniciais dos sistemas foram construídas na plataforma CHARMM-GUI (Jo *et al.*, 2008; Lee, Jumin *et al.*, 2016). A bicamada lipídica foi modelada com composição heterogênea, fundamentada em dados de lipidômica de macrófagos (Khedri *et al.*, 2022). Os constituintes utilizados são apresentados segundo a Tabela 3.1.

**Tabela 3.1** – Siglas, nomenclaturas e número de moléculas inseridas dos constituintes da bicamada lipídica modelada.

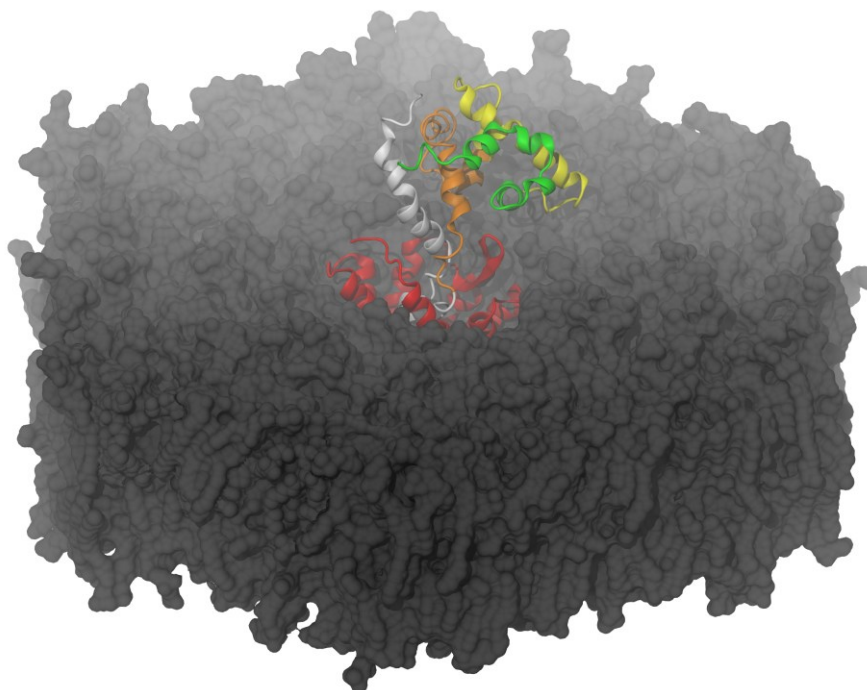
| <b>Sigla</b> | <b>Nomenclatura</b>                           | <b>Nº de moléculas</b> |
|--------------|-----------------------------------------------|------------------------|
| <i>CHL</i>   | Colesterol                                    | 272                    |
| <i>POPC</i>  | 1-palmitoil-2-oleoil-glicero-fosfocolina      | 220                    |
| <i>PSM</i>   | Palmitoil-esfingomielina                      | 80                     |
| <i>POPE</i>  | 1-palmitoil-2-oleoil-glicero-fosfoetanolamina | 76                     |
| <i>POPG</i>  | 1-palmitoil-2-oleoil-glicero-fosfoglicerol    | 60                     |
| <i>POPI</i>  | 1-palmitoil-2-oleoil-glicero-fosfoinositol    | 40                     |
| <i>POPS</i>  | 1-palmitoil-2-oleoil-glicero-fosfoserina      | 24                     |

Fonte: Autor, 2026.

A orientação do complexo em relação ao eixo da membrana, considerando as regiões intra e extracelulares, foi determinada com o auxílio do servidor PPM 2.0 (Lomize *et al.*, 2012).

Foi adotada uma caixa de simulação hexagonal, com solvatação realizada pelo modelo de água TIP3P (do inglês *transferable intermolecular potential with 3 points*) (Jorgensen *et al.*, 1983). Íons  $K^+$  e  $Cl^-$  foram adicionados para neutralizar a carga total do sistema e em concentração fisiológica de 0,15 M. As interações foram descritas pelo campo de força CHARMM36m (Huang, Jing *et al.*, 2017). Os sistemas resultantes apresentaram entre 290.783 e 302.757 partículas. A Figura 3.3 ilustra um dos sistemas construídos. Moléculas de água e íons foram omitidos para maior clareza visual.

**Figura 3.3** – Representação do sistema TcP21–CXCR4 em membrana lipídica, com CXCR4 em vermelho e lipídios em cinza. Água e íons omitidos.



Fonte: Autor, 2026.

#### 3.1.4 **Parâmetros das simulações de DM**

Todas as simulações por dinâmica molecular clássica foram conduzidas usando o pacote GROMACS (Abraham *et al.*, 2015), versão 2024.5, com base em um protocolo adaptado do CHARMM-GUI.

Cada sistema foi inicialmente submetido a duas minimizações de energia sucessivas utilizando o algoritmo *steepest descent*, com um passo de minimização de 0,1 Å. As duas minimizações foram diferenciadas apenas pela precisão numérica empregada nos cálculos (dupla precisão seguida de precisão simples). Ambas as etapas foram realizadas com até 5000 passos e critério de convergência correspondente à força máxima inferior a 100 kJ mol<sup>-1</sup> Å<sup>-1</sup>.

A etapa de equilibração termodinâmica foi composta por seis fases consecutivas ( $E_1 - E_6$ ), nas quais as restrições de posição aplicadas ao sistema foram progressivamente reduzidas. O algoritmo de *leapfrog* foi empregado em todas as fases, com passo de tempo de 1 fs nas três primeiras fases ( $E_1 - E_3$ ) e 2 fs nas



três últimas ( $E_4 - E_6$ ). As três primeiras fases tiveram duração de 125 ps cada, enquanto as fases subsequentes tiveram duração de 500 ps cada.

As interações não-ligadas foram tratadas com o esquema de Verlet, adotando-se um raio de corte de 12 Å para interações de van der Waals e eletrostáticas. As interações de van der Waals foram suavizadas por um modificador do tipo *force-switch* entre 10 e 12 Å. As interações eletrostáticas foram descritas pelo algoritmo SPME (do inglês *smooth Particle-mesh Ewald*) (Darden *et al.*, 1993; Essmann *et al.*, 1995).

O controle de temperatura foi realizado pelo método de *velocity-rescaling* (Bussi *et al.*, 2007), com uma constante de tempo de acoplamento de 1 ps e temperatura de 310 K. Todas as ligações envolvendo hidrogênio foram restringidas pelo algoritmo P-LINCS (do inglês *Parallel Linear Constraint Solver*) (Hess, 2008). As velocidades iniciais dos átomos foram atribuídas apenas na fase  $E_1$ , de acordo com a distribuição de Maxwell-Boltzmann para a temperatura de 310 K, sendo posteriormente propagadas.

As fases  $E_1 - E_2$  foram conduzidas em *ensemble* isocórico-isotérmico (NVT), enquanto as demais utilizaram o *ensemble* isobárico-isotérmico (NPT). O controle de pressão foi realizado utilizando o método de *C-rescale* (Bernetti; Bussi, 2020), com uma constante de tempo de acoplamento de 5 ps, compressibilidade isotérmica de  $4,5 \times 10^{-5} \text{ bar}^{-1}$  e pressão de 1,0 bar, aplicada de forma semi-isotrópica.

Após a equilibração, foi conduzida, para cada sistema, uma simulação de produção inicial de 1 ns, sem restrições de posição, e, posteriormente, cinco réplicas independentes com duração de 100 ns cada, identificadas como  $R_1 - R_5$ . Nessas etapas, foram mantidos os parâmetros de simulação previamente estabelecidos. O registro das coordenadas amostradas foi feito a cada 10 ps. No total, foram simulados 25 sistemas, correspondendo a um tempo de produção acumulado de 2,5  $\mu\text{s}$ . Para fins de padronização, cada conjunto de cinco réplicas de um mesmo modo de ligação foi identificado por  $M_1 - M_5$ , de acordo com a numeração do respectivo modo de ligação.

Quando foi necessário obter uma estrutura representativa do sistema, as conformações amostradas na trajetória foram agrupadas por meio do método de clusterização hierárquica aglomerativa, baseada na matriz de distâncias RMSD par-

a-par, empregando o critério de *average linkage*, e a estrutura medóide foi selecionada. A descrição do método é apresentada no Apêndice C.

### 3.1.5 **Análises dos sistemas**

#### 3.1.5.1 **Estabilidade conformacional e avaliação dos modos de ligação**

Após a conclusão das etapas de equilibração e amostragem, cada sistema foi analisado quanto à estabilidade conformacional do complexo, com o propósito de ordenar os modos de ligação quanto às suas respectivas relevâncias analíticas. Para isso, foram consideradas três métricas globais: raiz quadrada do desvio quadrático médio (RMSD) da cadeia principal das proteínas individualmente; o número de ligações de hidrogênio e o número de pontes salinas entre as proteínas do complexo.

As métricas utilizadas na avaliação dos modos de ligações foram analisadas por meio de testes de significância estatística e medidas de tamanho de efeito. Inicialmente, foi realizada a estatística descritiva para todos os sistemas. Para cada sistema ( $M_1 - M_5$ ), considerando-se as cinco réplicas ( $R_1 - R_5$ ), calcularam-se as médias individuais por réplica. Essas médias foram então utilizadas para compor um novo conjunto de dados, a partir do qual se estimaram a média global e o erro padrão da média (EPM), refletindo a variabilidade entre as réplicas.

A variabilidade global foi estimada por meio do desvio-padrão agrupado, calculado a partir de todos os *frames* de todas as réplicas de um mesmo sistema. Adicionalmente, a mediana global também foi obtida a partir do conjunto agrupado de observações, como medida robusta de tendência central.

Para a análise inferencial, assumiu-se a não normalidade dos dados e foram empregados testes robustos à heterogeneidade de variâncias. A análise de variância de Welch (Welch, 1951) foi utilizada para testar diferenças globais entre sistemas independentes. Quando aplicável, realizaram-se comparações múltiplas *post hoc* utilizando o teste de Games-Howell (Games; Howell, 1976).

O tamanho de efeito foi estimado pelo  $g$  de Hedges (Hedges, 1981), devido à sua correção para vieses em amostras pequenas. A magnitude dos efeitos foi

interpretada conforme os descritores de magnitude de Cohen-Sawilowsky (Cohen, 2013; Sawilowsky, 2009), apresentados na Tabela 3.2. Todos os testes consideraram um nível de confiança de 95%.

**Tabela 3.2** – Índices  $g$  de Hedges e respectivos descritores de magnitude de tamanho de efeito.

| $g$ de Hedges | Magnitude             |
|---------------|-----------------------|
| 0,20          | <i>Insignificante</i> |
| 0,50          | <i>Fraca</i>          |
| 0,80          | <i>Moderada</i>       |
| 1,20          | <i>Forte</i>          |
| 2,00          | <i>Muito Forte</i>    |
| >2,00         | <i>Enorme</i>         |

Fonte: Adaptado de Sawilowsky, 2009.

Nos cálculos estatísticos, os primeiros 20 ns de cada trajetória foram desconsiderados, de modo a eliminar efeitos associados à acomodação estrutural inicial.

### 3.1.6 **Determinação por consenso da interface de interação**

A identificação dos resíduos interfaciais mais prováveis do complexo TcP21–CXCR4 foi conduzida por meio de um método original, fundamentado em uma análise de consenso. Essa abordagem foi derivada diretamente de cálculos de área superficial acessível ao solvente (SASA) ao longo das trajetórias amostradas, métrica amplamente consolidada como descritor estrutural na caracterização de interfaces proteína-proteína (De Vries; Bonvin, 2008; Parvathy *et al.*, 2024; Xue *et al.*, 2015). A descrição detalhada da rotina de análise, incluindo etapas de implementação e validação, foi apresentada no Apêndice A. Todas as análises foram realizadas por meio de *scripts* em Python, executados em ambiente Jupyter Notebook.

Os cálculos de SASA foram realizados utilizando o algoritmo de Shrake–Rupley (Shrake; Rupley, 1973), empregando um raio de sonda de 1,4 Å e discretização de superfície atômica em 960 pontos por átomo, conforme implementado na biblioteca MDTraj 1.10 (McGibbon *et al.*, 2015). Para cada resíduo, foram extraídas as

estatísticas descritivas (média, mediana e desvio-padrão) e as frequências de ocupação em três intervalos de  $\Delta\text{SASA}$  ( $10 \leq \Delta\text{SASA} < 30 \text{ \AA}^2$ ;  $30 \leq \Delta\text{SASA} < 50 \text{ \AA}^2$  e  $\Delta\text{SASA} \geq 50 \text{ \AA}^2$ ).

Os resíduos foram avaliados por uma função de pontuação multimétrica composta por três descritores independentes: (i) magnitude de exposição (baseada na mediana de  $\Delta\text{SASA}$ ); (ii) persistência de interação (derivada de frequências de ocupação total e parciais); e (iii) confiabilidade estatística (estimada pela relação sinal-ruído). Todos os descritores foram normalizados no intervalo  $[0, 1]$  e combinados com pesos iguais.

Resíduos com pontuação final  $\geq 0,7$  foram considerados para a etapa de análise de consenso. A robustez dos candidatos foi então avaliada por consenso intrassistema, classificando-se os resíduos como “*Conservados*” quando presentes em 100% das réplicas ou “*Recorrentes*” quando presentes em pelo menos 60%.

#### 3.1.6.1 **Caracterização temporal dos contatos entre TcP21 e CXCR4**

A caracterização da interação entre TcP21 e CXCR4 foi realizada por meio da análise de contatos intermoleculares formados ao longo das trajetórias, considerando-se de forma integrada o número de contatos, suas frequências de ocupação e sua estabilidade temporal.

Os contatos intermoleculares foram definidos a partir de uma distância de corte de 4 Å. Para a avaliação da estabilidade temporal desses contatos, as trajetórias foram segmentadas em três janelas de mesmo número de *frames* (início, meio e fim), sendo calculadas, para cada uma, as frequências de ocupação parciais. Um critério de variação relativa de  $\pm 25\%$  foi empregado para classificar a evolução temporal dos contatos em termos de aumento, redução ou sem alteração significativa. Os contatos foram categorizados em quatro classes funcionais: “*Persistente*”, “*Emergente*”, “*Transiente*” ou “*Desestabilizante*”.

A partir desse processamento, foram estabelecidas quatro métricas principais: (i) a variabilidade de parceiros de interação por resíduo da TcP21; (ii) a recorrência dos pares de resíduos TcP21–CXCR4 ao longo da estabilização do complexo; (iii) a

distribuição do perfil temporal dos contatos mais prováveis; e (iv) a distribuição qualitativa das forças de interação. Para a métrica (iv), os aminoácidos foram agrupados segundo suas propriedades físico-químicas e as combinações de propriedades entre os pares observados foram utilizadas para inferir o tipo de força intermolecular predominante, categorizadas em: “*Interação Hidrofóbica*”, “*Atração Eletrostática*”, “*Interação Polar*” e “*Interação de van der Waals*”.

A robustez dos contatos foi avaliada por meio de análise de consenso intrassistema, calculando-se para cada par a frequência de ocorrência correspondente ao número de réplicas em que o contato foi observado. A descrição detalhada da rotina de análise, incluindo etapas de implementação e validação, é apresentada no Apêndice B. Todas as análises foram realizadas por meio de *scripts* em Python, executados em ambiente Jupyter Notebook.

## 3.2 RESULTADOS E DISCUSSÃO

---

### 3.2.1 *Obtenção dos modos de ligação*

A ancoragem molecular realizada no servidor ClusPro 2.0 resultou em 10 modos de ligação distintos, cada um representando o medóide dos clusters mais populosos.

Os clusters foram classificados segundo seus tamanhos, sendo os mais populosos considerados os mais representativos do espaço conformacional amostrado. Os resultados obtidos são apresentados na Tabela 3.3, incluindo as pontuações ponderadas de cada medóide, calculadas a partir da descrição da função equilibrada.

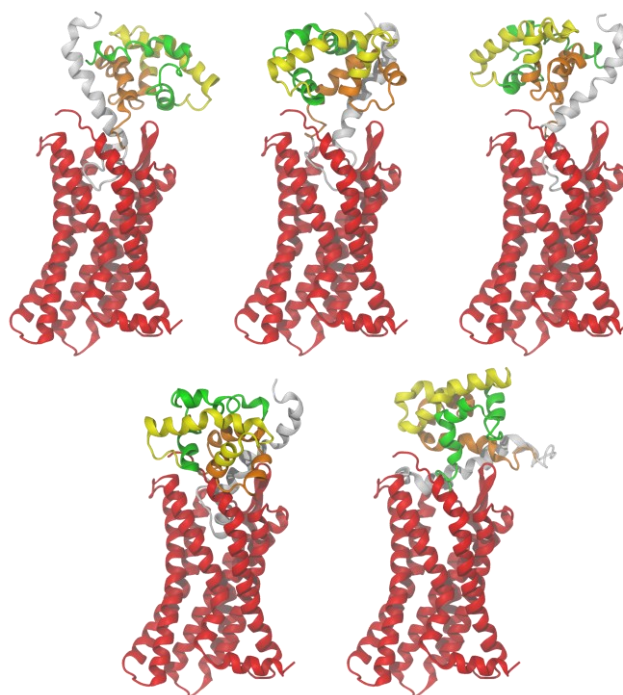
**Tabela 3.3** – Classificação dos clusters obtidos por ancoragem molecular, segundo número de membros e pontuação ponderada.

| Cluster | Membros | Pontuação |
|---------|---------|-----------|
| 1       | 164     | -1752,7   |
| 2       | 119     | -1647,3   |
| 3       | 98      | -1601,8   |
| 4       | 96      | -1513,5   |
| 5       | 73      | -1589,7   |
| 6       | 72      | -1777,4   |
| 7       | 58      | -1833,9   |
| 8       | 49      | -1539,0   |
| 9       | 48      | -1545,6   |
| 10      | 43      | -1510,9   |

Fonte: Autor, 2026.

Com base no critério de tamanho do cluster, os cinco modos de ligação mais representativos (clusters 1–5) foram selecionados para as simulações por DM. Os medóides dos respectivos clusters são apresentados individualmente na Figura 3.4.

**Figura 3.4** – Medóides dos clusters mais populosos obtidos por ancoragem molecular.



Fonte: Autor, 2026.

### 3.2.2 ***Estabilidade conformacional e avaliação dos modos de ligação***

Dado que os modos de ligação obtidos por ancoragem molecular representam configurações estruturais estáticas, suas estabilidades conformacionais foram avaliadas por meio de simulações de DM.

Os sistemas amostrados foram agrupados de acordo com os modos de ligação correspondentes, entretanto, em função do caráter exploratório do presente estudo, todos os modos de ligação foram considerados inicialmente equivalentes quanto à sua relevância analítica.

A estabilidade conformacional do complexo proteico foi avaliada utilizando três métricas globais: RMSD das cadeias individuais, o número de ligações de hidrogênio e o número de pontes salinas entre as proteínas do complexo. Para os cálculos estatísticos, os primeiros 20 ns de cada trajetória foram desconsiderados, de modo a excluir o período inicial de acomodação estrutural.

Para cada métrica analisada, cada trajetória foi resumida por seu valor médio ao longo do tempo de produção considerado. Esses valores médios foram então

utilizados para compor um novo conjunto de dados, no qual cada réplica representou uma observação independente dentro de um respectivo modo de ligação. A partir desse conjunto foram estimados a média geral e o erro padrão da média (EPM). O desvio-padrão agrupado ( $S_P$ ), por sua vez, foi calculado a partir do conjunto de valores observados ao longo das trajetórias individuais pertencentes a um mesmo modelo, permitindo estimar a dispersão dos estados amostrados durante as simulações.

Os cálculos de RMSD foram realizados separadamente para as proteínas CXCR4 e TcP21, considerando apenas os átomos diferentes de hidrogênio da cadeia principal. Além da análise estatística, os perfis temporais de RMSD também foram avaliados qualitativamente por meio de curvas de cada réplica, bem como das curvas de média e média suavizada. Os resultados estatísticos obtidos para a proteína CXCR4 são apresentados na Tabela 3.4 e os perfis temporais são mostrados individualmente na Figura 3.5.

**Tabela 3.4** – Estatística descritiva de RMSD da CXCR4.

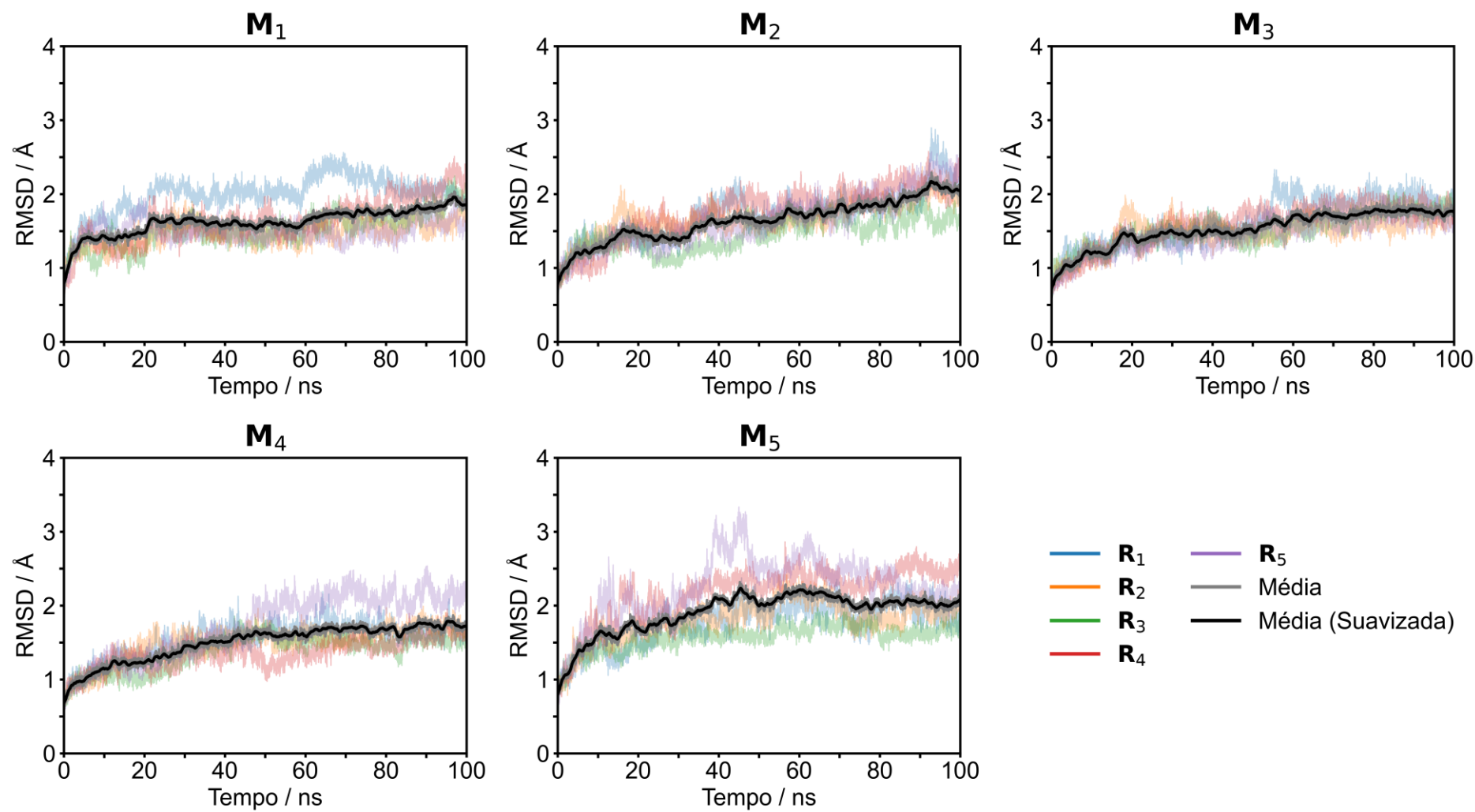
| Sistema        | RMSD / Å        |       |
|----------------|-----------------|-------|
|                | Média $\pm$ EPM | $S_P$ |
| M <sub>1</sub> | 1,69 $\pm$ 0,11 | 0,16  |
| M <sub>2</sub> | 1,73 $\pm$ 0,07 | 0,24  |
| M <sub>3</sub> | 1,62 $\pm$ 0,04 | 0,18  |
| M <sub>4</sub> | 1,60 $\pm$ 0,08 | 0,18  |
| M <sub>5</sub> | 2,02 $\pm$ 0,14 | 0,19  |

*Notas.* Valores expressos como média  $\pm$  erro padrão da média (EPM) de cinco réplicas independentes por sistema. Nenhuma diferença estatisticamente significativa foi observada entre os sistemas de acordo com o teste de ANOVA de Welch ( $\alpha = 0,05$ ).

Fonte: Autor, 2026.



Figura 3.5 – Perfis temporais de RMSD da CXCR4.



Fonte: Autor, 2026.

O teste de hipótese empregado (ANOVA de Welch) não indicou diferenças estatisticamente significativas entre os sistemas quando considerado o RMSD da cadeia principal da proteína CXCR4. Assim, dentro do conjunto de estados amostrados, os resultados indicam que os diferentes modos de ligação não produziram alterações relevantes na estabilidade conformacional global da proteína ao longo das simulações. Esse resultado é consistente com o contexto estrutural da CXCR4, uma proteína de membrana cuja mobilidade conformacional tende a ser parcialmente restringida pela presença da bicamada lipídica.

Os baixos valores de EPM observados indicam alta reprodutibilidade entre as réplicas independentes para um mesmo modo de ligação. Por outro lado, os valores de  $S_P$  refletem a baixa dispersão estrutural observada ao longo das trajetórias amostradas. Em conjunto, os valores de média geral e de  $S_P$  obtidos sugerem que a proteína manteve um regime estrutural estável durante o período de produção analisado, indicando convergência estrutural satisfatória das simulações, condição necessária para a interpretação das análises subsequentes.

Da mesma forma, a estatística descritiva e os perfis temporais do RMSD da proteína TcP21 também foram obtidos, e são apresentados na Tabela 3.5 e Figura 3.6, respectivamente.

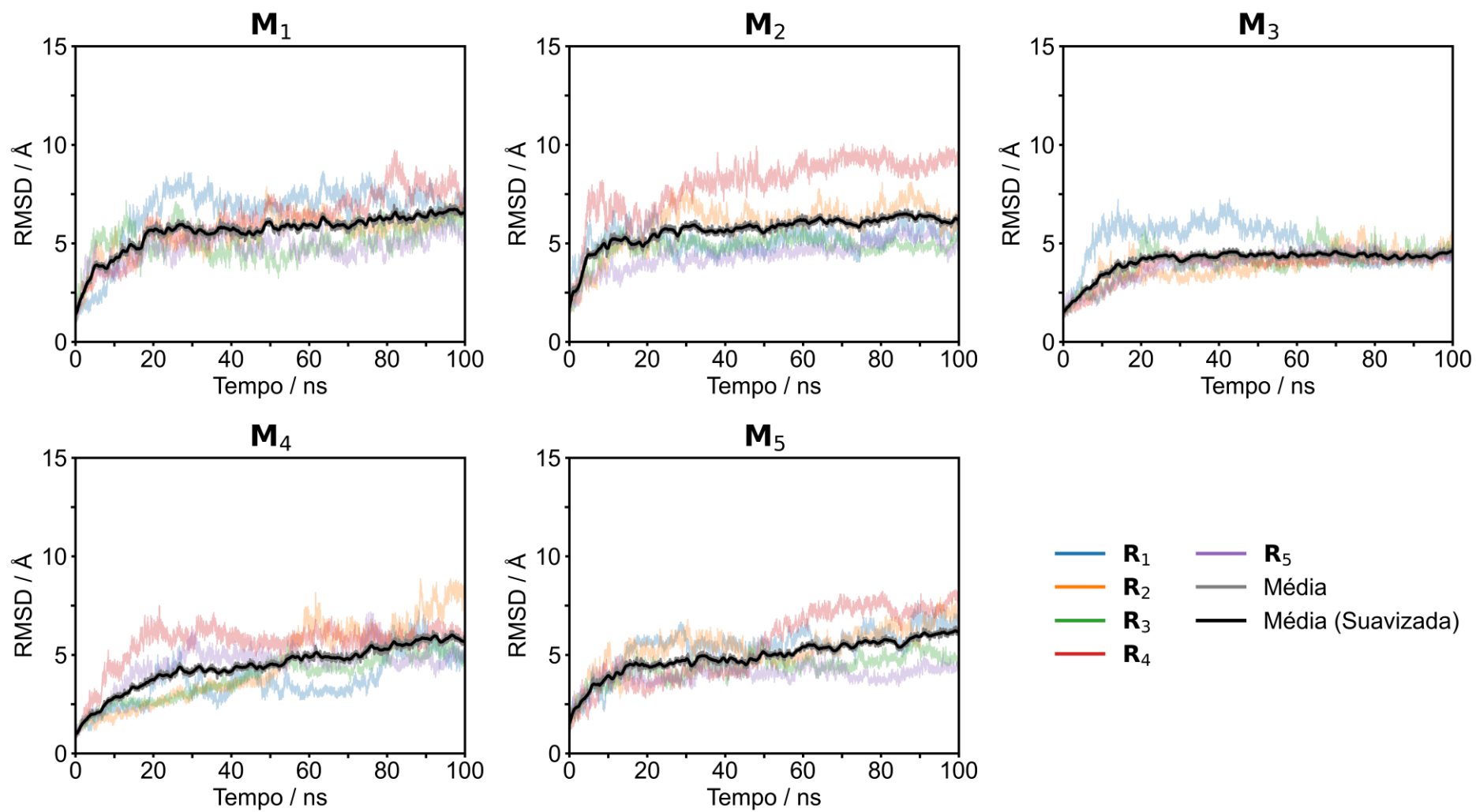
**Tabela 3.5** – Estatística descritiva de RMSD da TcP21.

| Sistema        | RMSD / Å        |       |
|----------------|-----------------|-------|
|                | Média $\pm$ EPM | $S_P$ |
| M <sub>1</sub> | 5,98 $\pm$ 0,43 | 0,66  |
| M <sub>2</sub> | 5,97 $\pm$ 0,71 | 0,53  |
| M <sub>3</sub> | 4,38 $\pm$ 0,16 | 0,49  |
| M <sub>4</sub> | 4,85 $\pm$ 0,34 | 0,88  |
| M <sub>5</sub> | 5,23 $\pm$ 0,38 | 0,72  |

*Notas.* Valores expressos como média  $\pm$  erro padrão da média (EPM) de cinco réplicas independentes por sistema. Diferenças estatisticamente significativas foram observadas entre os sistemas de acordo com o teste de ANOVA de Welch ( $\alpha = 0,05$ ). Comparações de pares e tamanhos de efeito são mostrados na análise *post hoc* (Apêndice D).

Fonte: Autor, 2026.

**Figura 3.6** – Perfis temporais de RMSD da TcP21.



Fonte: Autor, 2026.

Em contraste com o comportamento observado para a CXCR4, a proteína TcP21 apresentou valores de RMSD consideravelmente mais elevados ao longo das simulações, indicando maior flexibilidade estrutural. Esse perfil pode ser parcialmente explicado pela ausência dos efeitos estéricos impostos pela bicamada lipídica, além de características estruturais intrínsecas da própria proteína (pequeno tamanho e ausência de empacotamento terciário rígido), cuja flexibilidade já foi discutida anteriormente neste trabalho (seção 2.2.2).

Essa maior mobilidade conformacional também se manifesta na maior dispersão dos valores amostrados e nos perfis qualitativos observados entre réplicas pertencentes a um mesmo modo de ligação. Ainda assim, quando considerados os valores das médias gerais e os valores de  $S_P$  obtidos, observa-se que as trajetórias mantiveram um regime estrutural relativamente consistente ao longo dos estados amostrados. Para o sistema  $M_2$ , em particular, a magnitude do EPM, relativamente ao  $S_P$  de cada réplica, indica a presença de heterogeneidade entre as simulações. A inspeção gráfica das trajetórias e das médias individuais de cada réplica sugeriu que a réplica  $R_4$  foi a principal responsável por esse comportamento distinto. No entanto, a ocorrência pontual desse fenômeno não invalidou o estado amostrado, especialmente dada a natureza de ensemble estatístico da amostragem em simulações de Dinâmica Molecular clássica.

A análise do tamanho de efeito ( $g$  de Hedges) indicou que as comparações entre os pares  $(M_1, M_3)$  e  $(M_2, M_3)$  apresentaram as maiores magnitudes, ambas classificadas como “*Muito Forte*”, contribuindo para a diferença estatisticamente significativa observada no teste global. Por outro lado, a comparação entre  $(M_1, M_2)$  foi a única considerada insignificante, indicando equivalência estatística entre esses dois modos de ligação.

Apesar dessas diferenças, os valores absolutos de RMSD da cadeia principal da proteína TcP21 permanecem relativamente próximos entre si. Assim, em termos estruturais, os resultados não indicam diferenças suficientemente pronunciadas para justificar uma distinção acentuada entre os modos de ligação.

Além da estabilidade conformacional avaliada por meio do RMSD, também foi investigado o número de interações intermoleculares dos tipos ligação de hidrogênio e ponte salina formadas entre as proteínas TcP21–CXCR4. A análise de número de interações foi conduzida de forma análoga às análises anteriores, com a inclusão adicional dos valores de mediana e máximo para cada métrica. Iniciando-se pelo número de ligações de hidrogênio, os resultados são apresentados na Tabela 3.6 e Figura 3.7.

**Tabela 3.6** – Estatística descritiva do número de ligações de hidrogênio entre CXCR4 e TcP21.

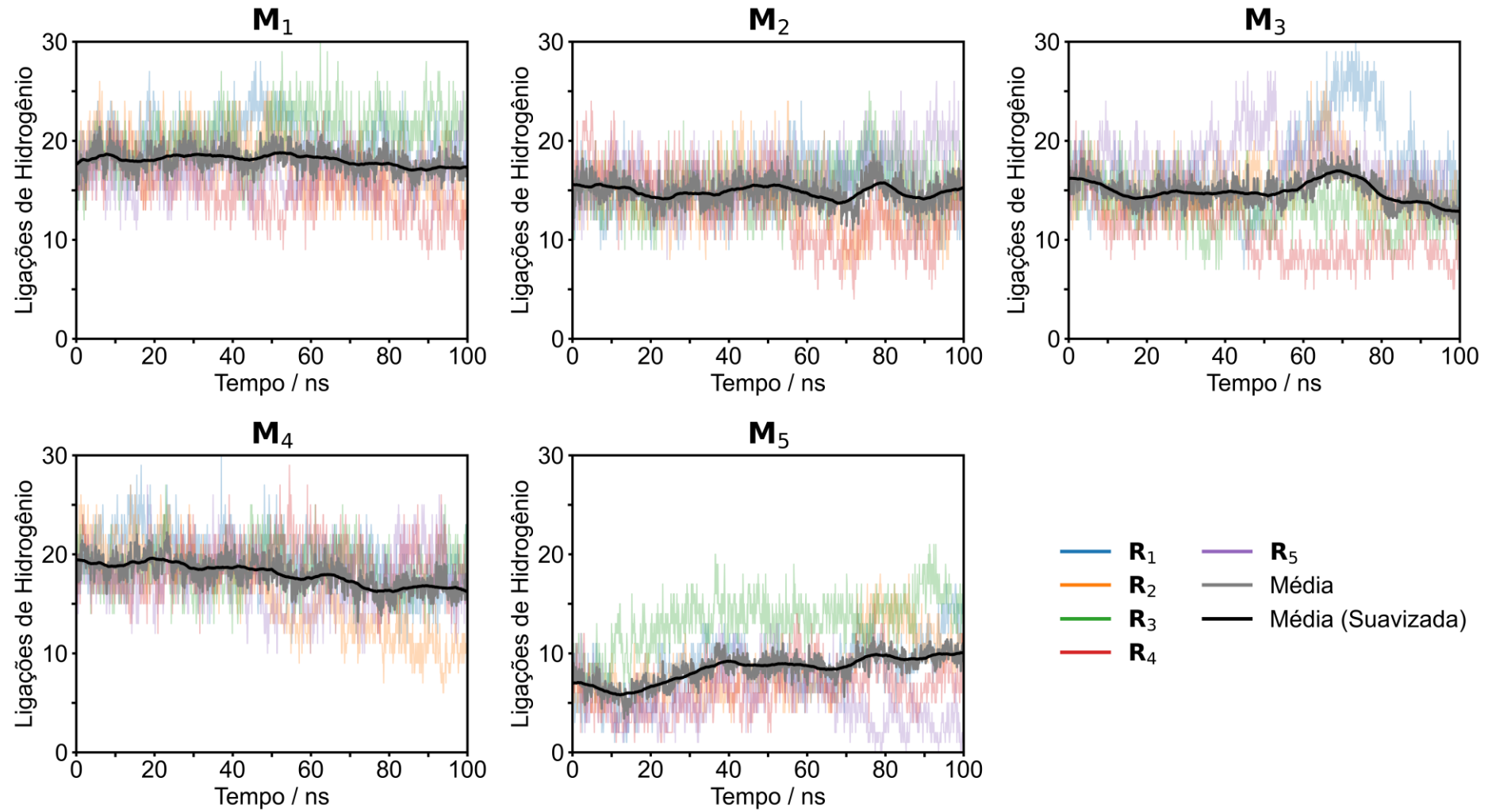
| Sistema        | Número de ligações de hidrogênio |                |         |        |
|----------------|----------------------------------|----------------|---------|--------|
|                | Média ± EPM                      | S <sub>P</sub> | Mediana | Máximo |
| M <sub>1</sub> | 18,0 ± 1,1                       | 2,6            | 18      | 28     |
| M <sub>2</sub> | 14,7 ± 0,7                       | 2,9            | 15      | 24     |
| M <sub>3</sub> | 14,8 ± 1,4                       | 3,0            | 15      | 27     |
| M <sub>4</sub> | 17,7 ± 0,8                       | 2,9            | 18      | 28     |
| M <sub>5</sub> | 8,8 ± 1,5                        | 2,5            | 8       | 19     |

*Notas.* Valores expressos como média ± erro padrão da média (EPM) de cinco réplicas independentes por sistema. Diferenças estatisticamente significativas foram observadas entre os sistemas de acordo com o teste de ANOVA de Welch ( $\alpha = 0,05$ ). Comparações de pares e tamanhos de efeito são mostrados na análise *post hoc* (Apêndice D).

Fonte: Autor, 2026.

O teste de hipótese empregado rejeitou a equivalência estatística entre os modos de ligação. A análise de tamanho de efeito indicou diferenças expressivas entre a maioria dos modos de ligação, exceto entre o par (M<sub>1</sub>, M<sub>4</sub>), considerado estatisticamente equivalente. Os modelos M<sub>2</sub> e M<sub>3</sub> também foram classificados como equivalentes entre si, embora suas diferenças em relação aos demais tenham sido consideradas expressivas.

**Figura 3.7** – Perfis temporais do número de ligações de hidrogênio entre TcP21 e CXCR4.



Fonte: Autor, 2026.

A magnitude das diferenças variou entre “*Forte*” ( $M_1, M_3$ ) e “*Enorme*” ( $M_1, M_5$ ). Em contraste com os resultados observados para o RMSD, o número de ligações de hidrogênio indicou distinções mais acentuadas entre os modos de ligação, particularmente para o sistema  $M_5$ . Considerando as medianas observadas, esse modo apresentou entre 7–10 ligações de hidrogênio a menos do que os demais, correspondendo a uma redução entre aproximadamente 47–56% no número de interações. Mesmo quando considerados os estados de maior número de ligações de hidrogênio observadas ao longo das trajetórias (valor máximo), essa diferença permanece significativa, variando entre 5–9 interações, o equivalente a uma redução entre aproximadamente 21–32% do número de interações.

Como etapa final do estudo de estabilidade conformacional e avaliação dos modos de ligações simulados, também foi investigado o número de pontes salinas. Os resultados são apresentados na Tabela 3.7 e na Figura 3.8, e seguem a mesma estrutura adotada nas análises anteriores.

**Tabela 3.7** – Estatística descritiva do número de pontes salinas entre CXCR4 e TcP21.

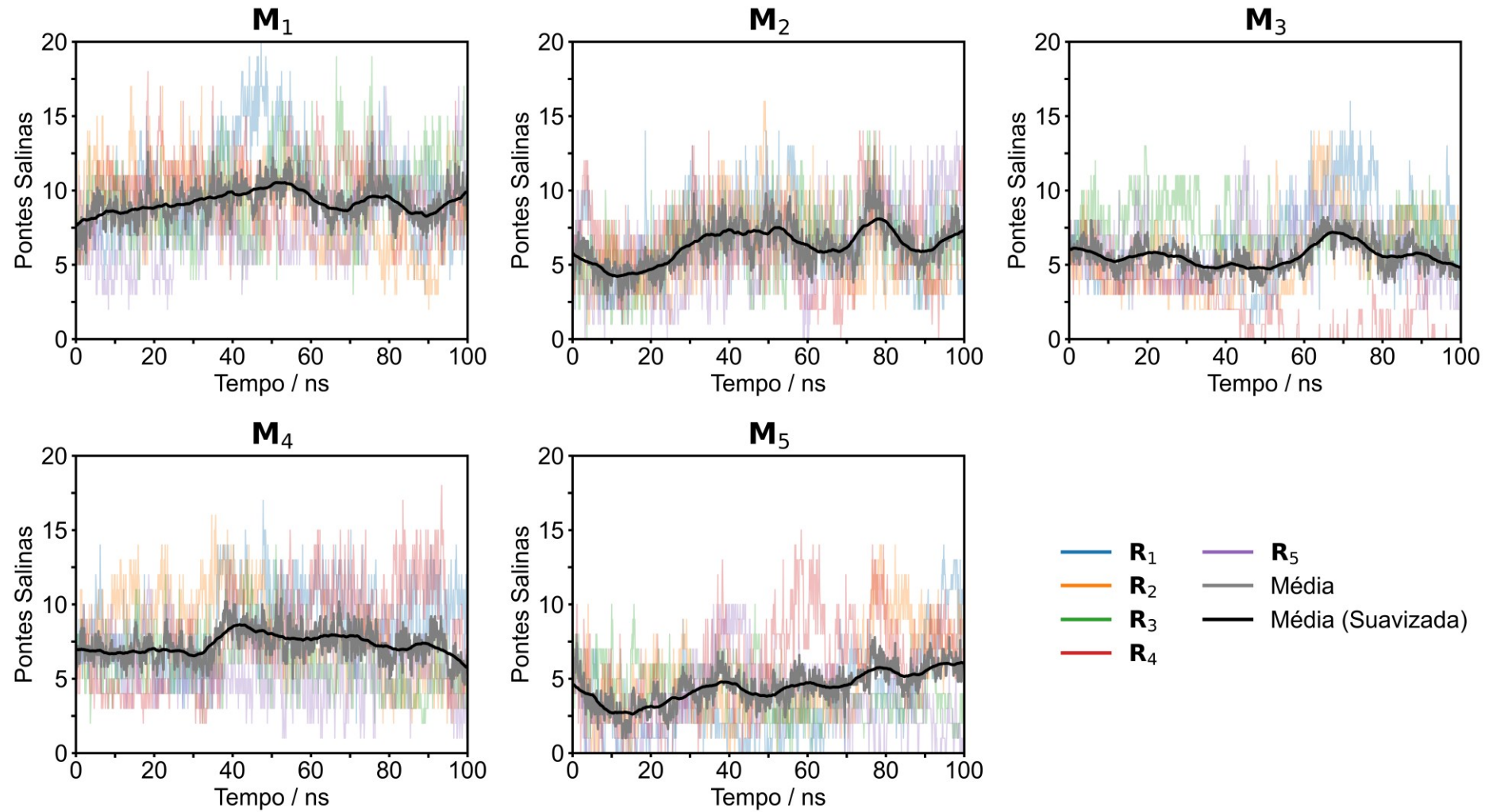
| Sistema | Número de pontes salinas |       |         |        |
|---------|--------------------------|-------|---------|--------|
|         | Média $\pm$ EPM          | $S_p$ | Mediana | Máximo |
| $M_1$   | 9,4 $\pm$ 0,5            | 2,7   | 10      | 21     |
| $M_2$   | 6,7 $\pm$ 0,1            | 2,6   | 7       | 16     |
| $M_3$   | 5,6 $\pm$ 1,1            | 2,2   | 6       | 14     |
| $M_4$   | 7,4 $\pm$ 0,8            | 2,5   | 8       | 16     |
| $M_5$   | 4,7 $\pm$ 0,5            | 2,6   | 4       | 16     |

*Notas.* Valores expressos como média  $\pm$  erro padrão da média (EPM) de cinco réplicas independentes por sistema. Diferenças estatisticamente significativas foram observadas entre os sistemas de acordo com o teste de ANOVA de Welch ( $\alpha = 0,05$ ). Comparações de pares e tamanhos de efeito são mostrados na análise *post hoc* (Apêndice D).

Fonte: Autor, 2026.

O teste de hipótese novamente rejeitou a equivalência estatística entre os modos de ligação. A análise de tamanho de efeito não indicou equivalência estatística entre nenhum par. A menor diferença observada ocorreu entre o par ( $M_3, M_5$ ), classificada como “*Fraca*”, enquanto as maiores diferenças foram observadas entre os pares ( $M_1, M_2$ ) e ( $M_1, M_5$ ), ambas classificadas como “*Enorme*”.

**Figura 3.8** – Perfis temporais do número de pontes salinas entre TcP21 e CXCR4.



Fonte: Autor, 2026.



Os valores de EPM indicaram elevada reprodutibilidade entre as réplicas para um mesmo modo de ligação, enquanto os valores de  $S_P$  refletiram uma dispersão substancial dos valores observados ao longo das trajetórias simuladas. De maneira semelhante ao observado para o número de ligações de hidrogênio, os resultados referentes ao número de pontes salinas revelaram distinções pronunciadas entre os modos de ligação, particularmente para o sistema  $M_1$ , que apresentou o maior número de interações desse tipo.

Considerando as medianas observadas, a diferença efetiva entre os modos variou entre 2–6 pontes salinas, correspondendo a um aumento aproximado de 20–60% do número de interações. Quando considerados os estados de maior número de interações observados ao longo das trajetórias (valor máximo), essa diferença varia entre 5–7 interações, equivalente a um aumento entre aproximadamente 24–33% do número de pontes salinas.

De forma geral, os resultados obtidos indicaram que, embora os modos de ligação tenham sido inicialmente considerados analiticamente equivalentes, diferenças relevantes foram observadas entre eles, particularmente no número de ligações de hidrogênio e de pontes salinas. Nesse contexto, o sistema  $M_5$  apresentou o comportamento mais discrepante em relação aos demais. Tal observação é consistente com a classificação obtida na etapa de ancoramento molecular, na qual esse modo de ligação correspondeu à estrutura central do cluster menos populoso dentre aqueles considerados para as simulações.

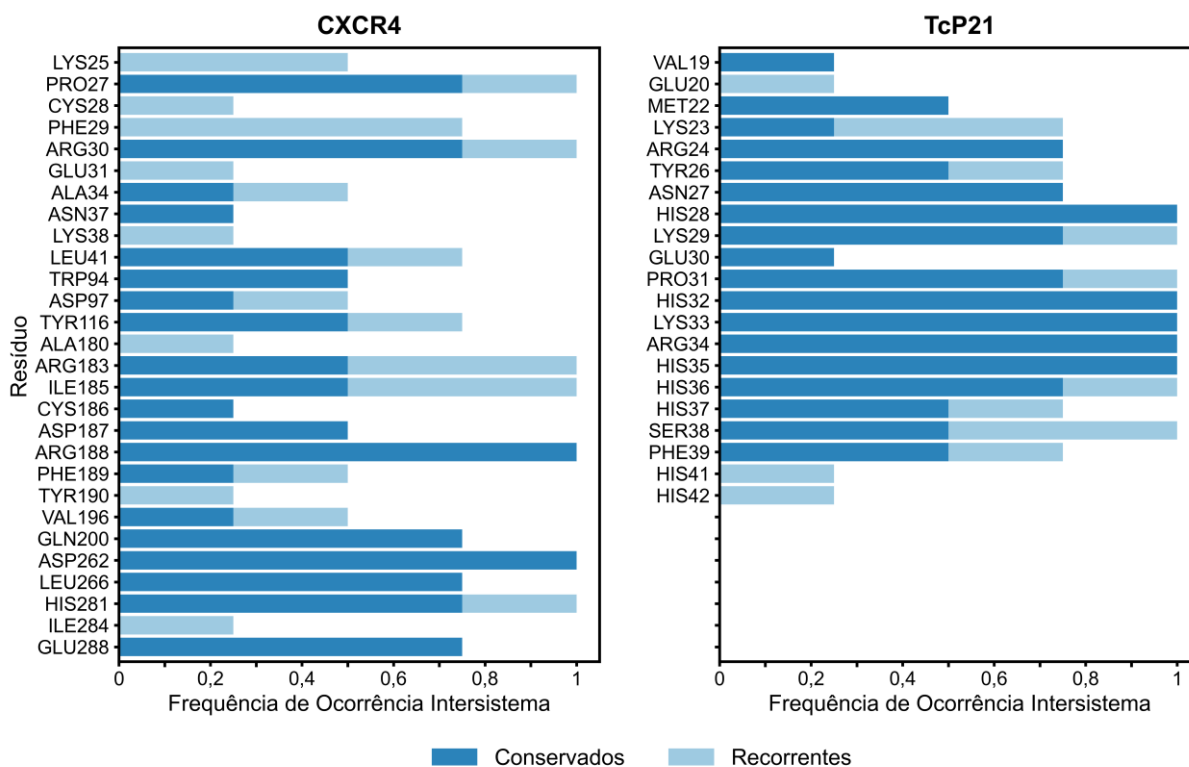
Assim, ainda que  $M_5$  permaneça estruturalmente plausível, os resultados de estabilidade conformacional dos modos de ligação sugerem que ele possa representar um estado de interação de menor probabilidade de ocorrência quando comparado aos demais modos investigados. Dessa forma, embora o modo de ligação tenha sido mantido nas análises subsequentes, os resultados derivados desse sistema foram interpretados com a devida ressalva.

### 3.2.3 Determinação por consenso da interface de interação

Após a avaliação dos modos de ligação por meio de análises preliminares de estabilidade conformacional, os resíduos de interface mais prováveis foram propostos com base em determinação por consenso fundamentada na função de pontuação multimétrica baseada em  $\Delta SASA$ , detalhada no Apêndice A.

Os resíduos de interface propostos para as proteínas CXCR4 e TcP21 são exibidos na Figura 3.9, organizados em ordem crescente do número do resíduo. A distribuição entre resíduos “*Conservados*” e “*Recorrentes*” foi incluída como informação complementar sobre a persistência e a recorrência intrassistema dos resíduos avaliados.

**Figura 3.9** – Resíduos de interface mais prováveis para as proteínas CXCR4 e TcP21.



Fonte: Autor, 2026.

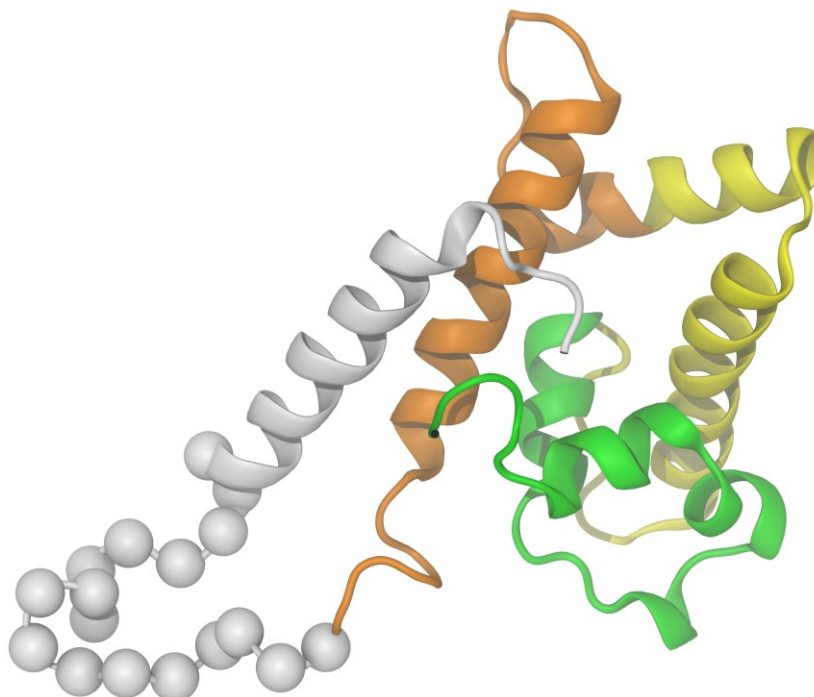
A função de pontuação empregada neste trabalho identificou um total de 21 e 28 possíveis resíduos de interface para as proteínas TcP21 e CXCR4,

respectivamente. Para a TcP21, esses resíduos concentraram-se em uma região predominantemente contínua, correspondente ao intervalo dos resíduos 19–42, com interrupções pontuais, observadas nos resíduos 21, 25 e 40. Em contraste, a CXCR4 apresentou uma distribuição mais fragmentada, com quatro possíveis regiões de interface localizadas nos intervalos 25–41, 94–116, 180–200 e 262–288, compostas por 10, 3, 10 e 5 resíduos, respectivamente, considerando-se as interrupções pontuais dentro de cada intervalo. Essa distribuição mais dispersa é consistente com a organização estrutural da CXCR4, uma proteína transmembranar caracterizada por um canal formado por sete alfa-hélices, capaz de acomodar a região de interação da proteína TcP21 no interior do vestíbulo extracelular.

Considerando a frequência de ocorrência agrupada como métrica de priorização dos resíduos de interface, observou-se que determinados resíduos apresentaram maior destaque. Na TcP21, observou-se alta recorrência para os resíduos no intervalo MET22–PHE39. Entre estes, ARG24, ASN27, HIS28, HIS32, LYS33, ARG34 e HIS35 destacaram-se por sua alta conservação, indicando um papel essencial na interação. Resíduos como LYS23, TYR26, LYS29, PRO31, HIS36, HIS37, SER38 e PHE39 apresentaram perfil intermediário, com distribuição entre “*Conservados*” e “*Recorrentes*”, indicando um papel auxiliar na interação. Em contraste, VAL19, HIS41 e HIS42 apresentaram baixa frequência e predominância de “*Recorrentes*”, possivelmente por corresponderem a posições periféricas da interface.

Esse segmento contínuo da TcP21 foi caracterizado por elevada densidade de resíduos carregados (ARG, LYS) e polares (HIS, ASN), sugerindo possível complementaridade eletrostática com os receptores da CXCR4 e indicando uma interface coesa, com alta consistência intrassistema e variações locais compatíveis com flutuações conformacionais. A Figura 3.10, que corresponde à estrutura medóide de  $M_1(R_1)$ , destaca o segmento identificado, caracterizado estruturalmente como uma volta. Ressalta-se que a estrutura utilizada foi selecionada somente com finalidade ilustrativa.

**Figura 3.10** – Região de interface de interação da TcP21. Átomos Ca dos resíduos de interface são mostrados como esferas.



Fonte: Autor, 2026.

Na CXCR4, a interface apresentou organização descontínua do ponto de vista topológico, porém estruturalmente convergente. Resíduos distribuídos entre a região N-terminal (MET1–PHE29) e o domínio TM1 (ARG30–GLN66) compuseram um primeiro conjunto funcional de interação. PRO27 e ARG30 mostraram alta frequência e comportamento consistente, enquanto PHE29, LEU41 e ALA34 apresentaram caráter auxiliar, com contatos mais transientes. Esse padrão indicou um estágio inicial de reconhecimento dinâmico e flexível.

No domínio ECL2 (ASN176–ASN192), observou-se um agrupamento de resíduos frequentes, incluindo ARG188, CYS186 e ASP187, predominantemente classificados como “*Conservados*”, além de ARG183 e ILE185, de comportamento misto. A predominância de resíduos “*Conservados*” nessa região sugeriu a formação de contatos persistentes entre o loop e a TcP21.

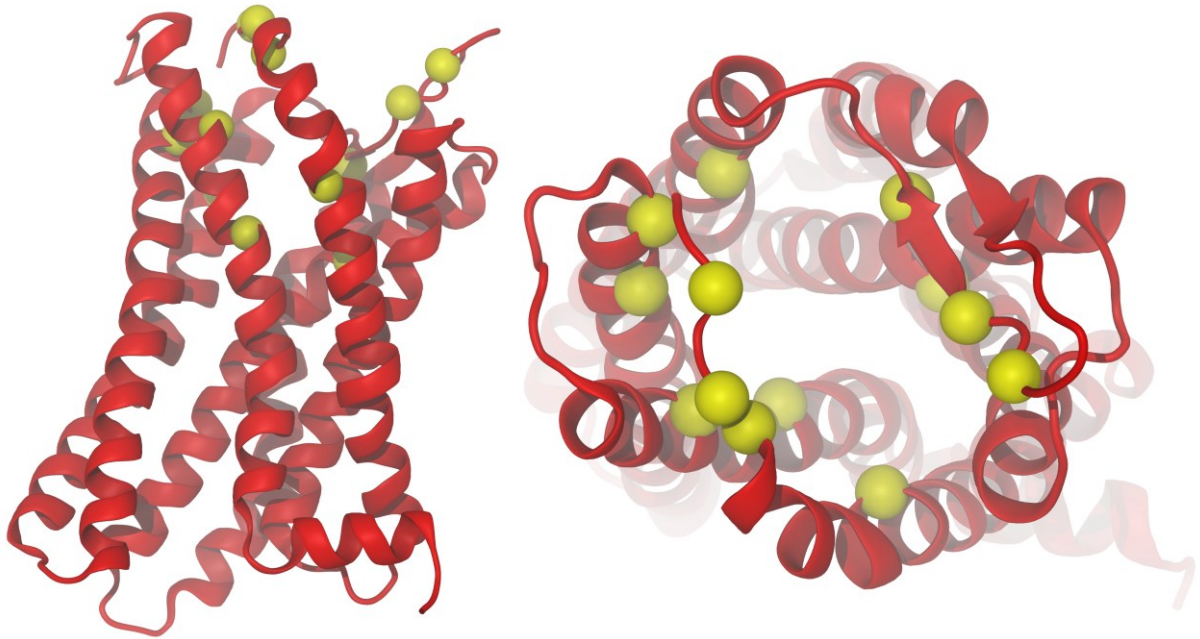
Adicionalmente, resíduos localizados nos domínios TM5 (ASP193–SER229) e TM7 (GLY273–LEU305), como GLN200, ASP262, LEU266, GLU288 e HIS281, apresentaram alta recorrência. A elevada proporção de classificação “*Conservados*” nesses domínios sugeriu a formação de uma região de estabilização do complexo

mais interna na CXCR4. Nesse contexto, esses resíduos foram associados ao núcleo funcional da interação, enquanto os demais contribuíram para o posicionamento e estabilização inicial da TcP21 na abertura do canal.

Ressalta-se que o sistema  $M_5$  não foi considerado para a determinação por consenso. Dados os resultados obtidos anteriormente nas análises de estabilidade conformacional dos modos de ligação, foi considerado que a inclusão desse modo poderia enviesar a identificação dos resíduos de interface ao incorporar contribuições associadas a um estado de menor probabilidade de acesso. Ainda assim, realizou-se uma análise complementar dos resíduos identificados nesse sistema. Para a TcP21, destacaram-se MET1, ARG2, LEU8, LEU12, CYS14, PHE150 e ARG153, enquanto, para a CXCR4, foram observados SER178, ASP181, PHE189, TYR190, ASP193, LEU266, GLU268 e GLU277. A integração desses resíduos com os demais sistemas revelou que a maioria apresentou frequência de 20%, indicando ocorrência exclusiva em  $M_5$ . Apenas PHE189 e LEU266 foram observados em outros sistemas. Em conjunto, esses resultados reforçam que a inclusão do sistema  $M_5$  no consenso introduziria principalmente contribuições específicas de um estado pouco representativo, sem impacto significativo sobre os padrões gerais previamente observados.

Assim, a análise baseada na função de pontuação de resíduos de interface indicou a existência de uma região predominantemente contínua de interação na TcP21, abrangendo o segmento LYS23–PHE39, com até 15 resíduos envolvidos. Em contraste, na CXCR4, não foi observada uma região contínua equivalente, mas sim um conjunto de pontos de contato distribuídos ao longo da abertura e do interior do vestíbulo extracelular. Esses pontos de contato foram PRO27, PHE29, ARG30, LEU41, TYR116, ARG183, ILE185, ARG188, GLN200, ASP262, LEU266, HIS281 e GLU288, refletindo uma organização espacialmente dispersa da interface. Os pontos de contato são mostrados na Figura 3.11.

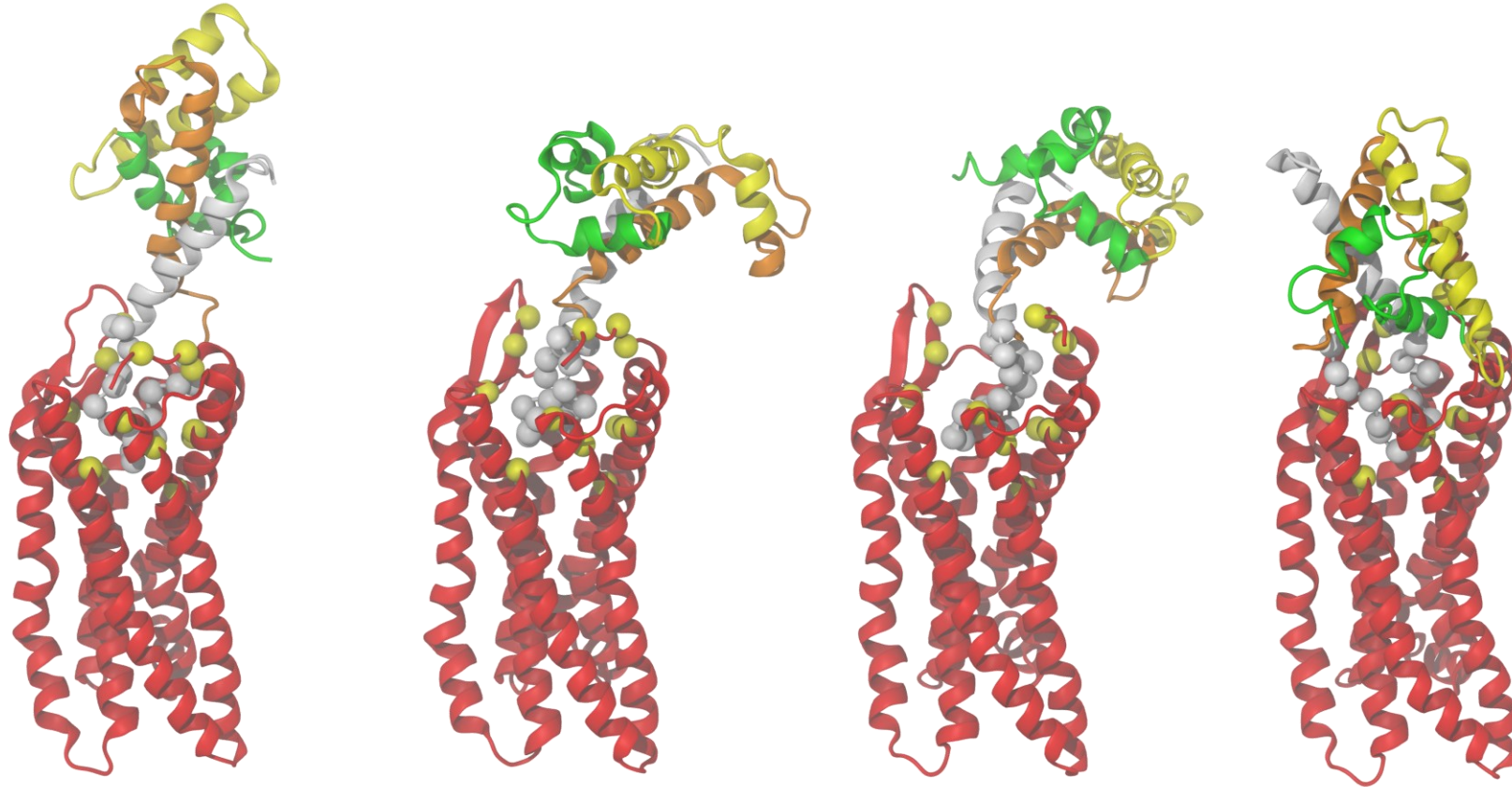
**Figura 3.11** – Pontos de contato interfaciais da CXCR4. Átomos C $\alpha$  dos resíduos de interface são mostrados como esferas. Da esquerda para a direita: vistas frontal e superior.



Fonte: Autor, 2026.

No geral, esse arranjo sugere um mecanismo de interação em múltiplas etapas. Em um primeiro estágio, contatos transientes na região N-terminal e no domínio TM1 promovem um reconhecimento inicial, dinâmico e flexível, possivelmente associado ao alinhamento inicial da TcP21. Em seguida, ocorre a estabilização do complexo, envolvendo os domínios ECL2 e TM5/TM7, com formação de interações mais internas e persistentes no interior do vestíbulo extracelular. A Figura 3.12 ilustra as estruturas medóides das réplicas  $R_1$ , com exceção do sistema  $M_5$ , com destaque para os resíduos identificados como de interface para ambas as proteínas.

**Figura 3.12** – Interface de interação do complexo TcP21–CXCR4. Átomos C $\alpha$  dos resíduos de interface são mostrados como esferas. Da esquerda para a direita:  $R_1(M_1 - M_4)$ .



Fonte: Autor, 2026.

Com base na observação de uma região contínua de interação na TcP21, foram calculadas as frequências de ocupação de ligações de hidrogênio por resíduo da TcP21, visando avaliar a consistência temporal dessa interface. Os resultados estão apresentados na Figura 3.13. Dado que os sistemas  $M_1 - M_5$  não precisaram ser agrupados em um único conjunto para o cálculo das frequências de ocupação, o sistema  $M_5$  foi incluído.

Os sistemas  $M_1 - M_4$  apresentaram perfis qualitativamente semelhantes, com predominância de ligações de hidrogênio na região compreendida entre os resíduos 26–39. Conforme mostrado na Tabela 3.8, os dez resíduos com maior ocupação nesses sistemas exibiram, majoritariamente, valores superiores a 60% do tempo simulado, todos pertencentes à interface de interação previamente definida. Adicionalmente, vários resíduos apresentaram ocupação acima de 90%, evidenciando elevada persistência dessas interações ao longo do tempo.

Em contrapartida, o sistema  $M_5$  apresentou apenas dois resíduos com ocupação acima de 60%, ambos não observados nos demais sistemas analisados. Além disso, os valores de desvios padrão foram consistentemente mais elevados, refletindo maior dispersão dos dados e, conseqüentemente, menor estabilidade das interações.



**Tabela 3.8** – Top 10 resíduos da TcP21 de maior frequência de ocupação de ligações de hidrogênio, por sistema.

| M <sub>1</sub> |              | M <sub>2</sub> |              | M <sub>3</sub> |              |
|----------------|--------------|----------------|--------------|----------------|--------------|
| Resíduo        | Ocupação / % | Resíduo        | Ocupação / % | Resíduo        | Ocupação / % |
| 29             | 99,9 ± 0,1   | 27             | 99,8 ± 0,2   | 33             | 98,19 ± 3,26 |
| 34             | 99,9 ± 0,2   | 35             | 90,8 ± 12,7  | 37             | 97,64 ± 3,23 |
| 28             | 99,4 ± 0,9   | 28             | 90,8 ± 14,7  | 26             | 94,64 ± 6,24 |
| 32             | 96,3 ± 3,1   | 36             | 80,0 ± 10,6  | 34             | 86,8 ± 22,5  |
| 33             | 92,2 ± 13,8  | 31             | 74,2 ± 15,9  | 32             | 74,0 ± 26,9  |
| 35             | 80,5 ± 15,4  | 29             | 65,8 ± 14,8  | 35             | 72,7 ± 13,8  |
| 27             | 79,7 ± 39,0  | 33             | 62,5 ± 23,0  | 29             | 70,6 ± 37,4  |
| 42             | 63,0 ± 38,6  | 32             | 59,2 ± 32,6  | 23             | 58,5 ± 10,6  |
| 31             | 53,3 ± 20,0  | 38             | 53,8 ± 12,8  | 24             | 57,5 ± 34,2  |
| 38             | 53,0 ± 29,2  | 34             | 52,6 ± 18,7  | 38             | 48,6 ± 42,3  |

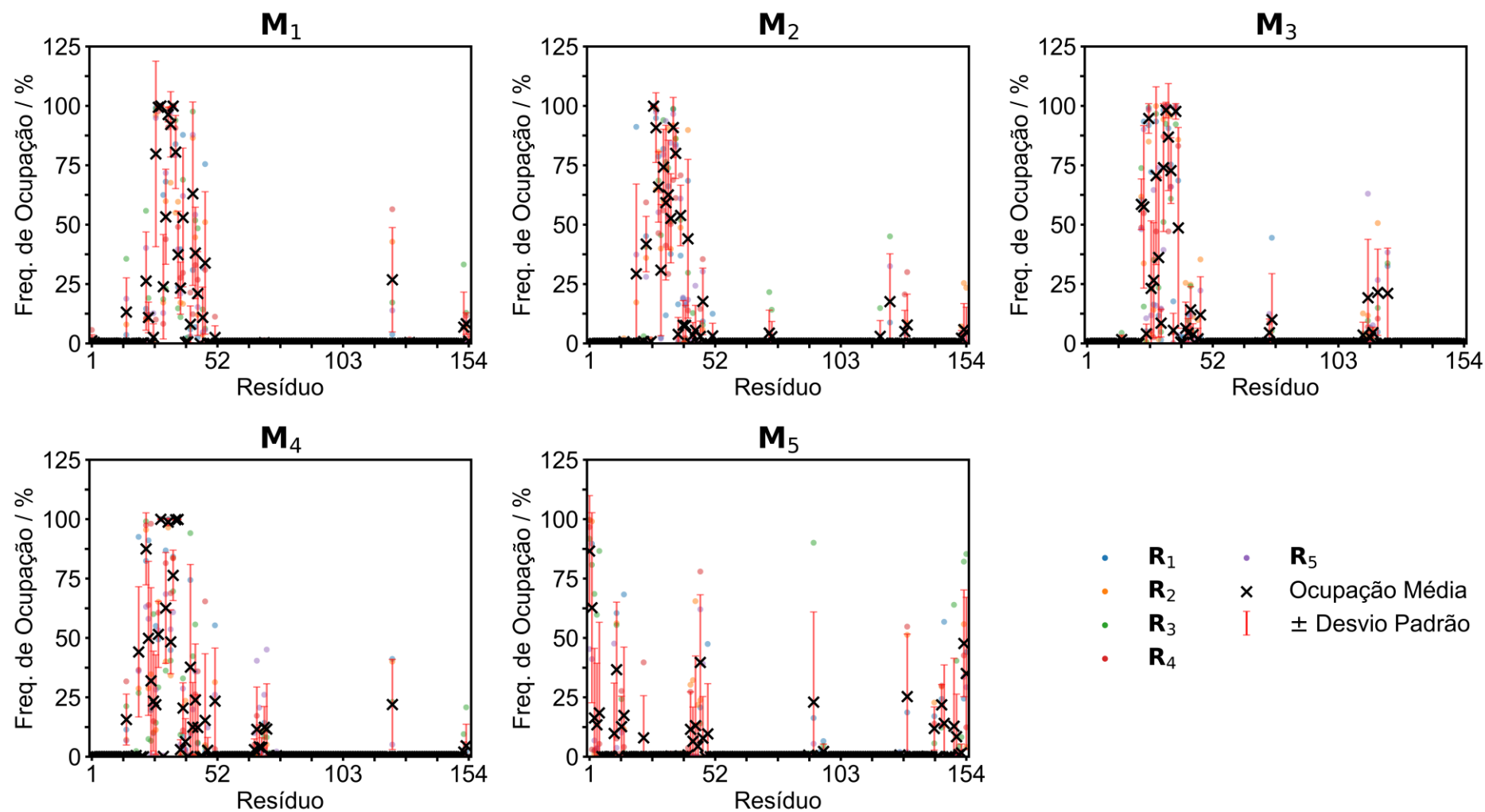
  

| M <sub>4</sub> |              | M <sub>5</sub> |              |
|----------------|--------------|----------------|--------------|
| Resíduo        | Ocupação / % | Resíduo        | Ocupação / % |
| 29             | 99,99 ± 0,01 | 1              | 86,6 ± 23,3  |
| 36             | 99,91 ± 0,07 | 2              | 62,7 ± 40,0  |
| 35             | 99,79 ± 0,20 | 153            | 47,7 ± 22,4  |
| 32             | 98,77 ± 1,51 | 46             | 39,7 ± 28,4  |
| 23             | 87,5 ± 15,1  | 12             | 36,7 ± 28,4  |
| 34             | 76,3 ± 10,6  | 154            | 35,0 ± 32,1  |
| 31             | 62,6 ± 23,3  | 130            | 25,3 ± 26,3  |
| 28             | 51,5 ± 13,9  | 92             | 23,0 ± 38,0  |
| 24             | 49,8 ± 32,4  | 144            | 21,8 ± 8,58  |
| 33             | 48,4 ± 13,4  | 5              | 18,4 ± 38,2  |

*Notas.* Valores expressos como média ± desvio-padrão de cinco réplicas independentes por sistema. As frequências de ocupação foram calculadas independentemente do número de parceiros ligantes da proteína CXCR4.

Fonte: Autor, 2026.

**Figura 3.13** – Frequências de ocupação de ligações de hidrogênio dos resíduos da TcP21, por sistema.



Fonte: Autor, 2026.

### 3.2.3.1 *Caracterização temporal dos contatos entre TcP21 e CXCR4*

Após a identificação dos resíduos mais prováveis de interface da TcP21, foi realizada uma análise exploratória das interações com a CXCR4, considerando o número, a frequência e a estabilidade dos contatos. Essa análise não se restringiu aos resíduos previamente definidos como de interface, uma vez que resíduos auxiliares também poderiam contribuir para a estabilização do complexo.

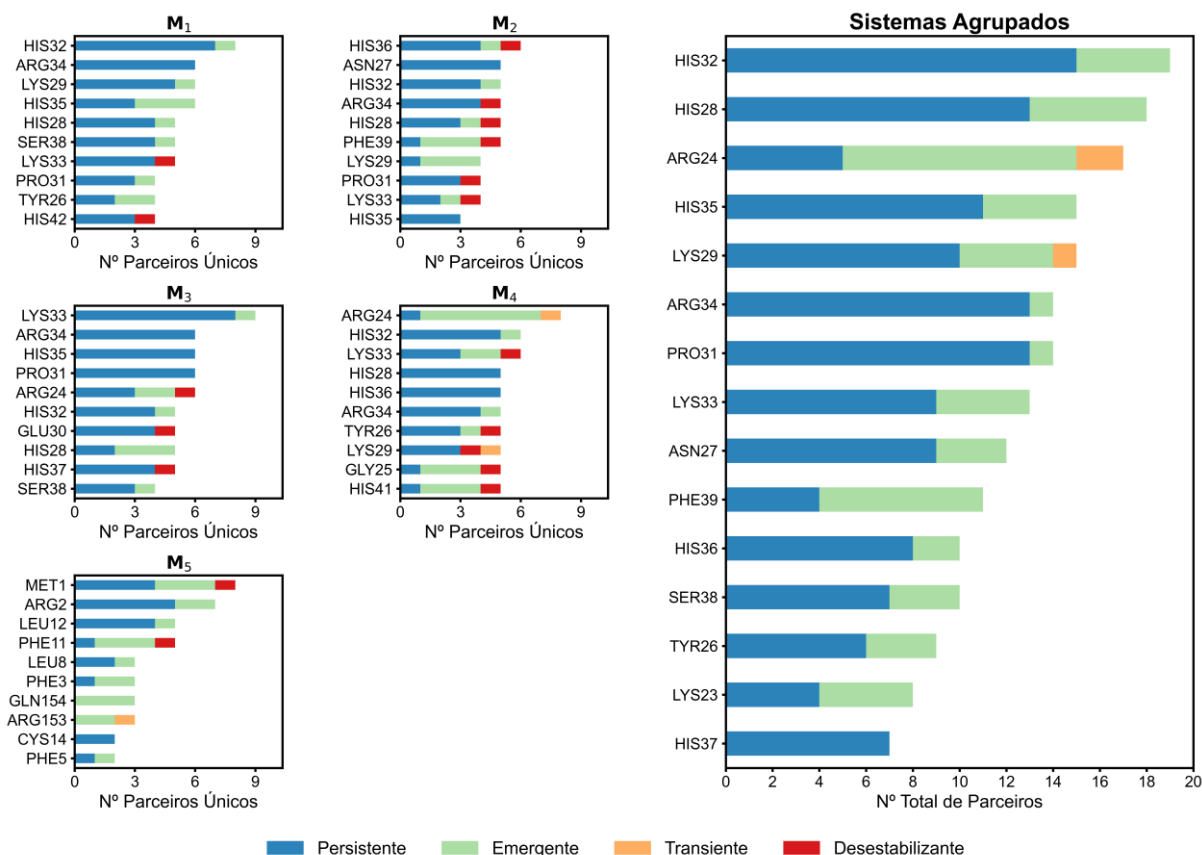
A caracterização dos contatos foi conduzida sob uma perspectiva temporal, contemplando: (i) a variabilidade de parceiros de interação por resíduo da TcP21; (ii) a identificação dos pares TcP21–CXCR4 mais recorrentes durante a estabilização do complexo; (iii) a distribuição do perfil temporal dos contatos mais prováveis; e (iv) a distribuição qualitativa dos tipos de forças intermoleculares estabilizantes do complexo.

### 3.2.3.2 *Variabilidade de parceiros de interação da TcP21*

Inicialmente, os resíduos da proteína TcP21 foram avaliados quanto ao número de parceiros únicos e totais, considerando consensos em regimes intrassistema e intersistema, e classificados adicionalmente de acordo com a evolução dinâmica dos contatos. Para cada sistema, foram selecionados os dez resíduos com maior número de parceiros únicos, enquanto na análise agrupada se considerou os 15 resíduos com maior número total de parceiros. Os resultados são apresentados na Figura 3.14.

A análise do conjunto agrupado indicou que a estabilização do complexo TcP21–CXCR4 se concentrou na região contínua previamente descrita, evidenciando sua relevância funcional. Observou-se a formação de um núcleo de interação predominantemente estável, centrado nos resíduos HIS28–HIS35, ao redor do qual se organizaram resíduos com maior plasticidade, incluindo ARG24 e PHE39. Esses resíduos apresentaram maior frequência de contatos emergentes, sugerindo participação em processos de adaptação estrutural e ajuste induzido dos modos de ligação ao longo das simulações.

**Figura 3.14** – Resíduos de maior número de contatos da proteína TcP21, por sistema e agrupado.



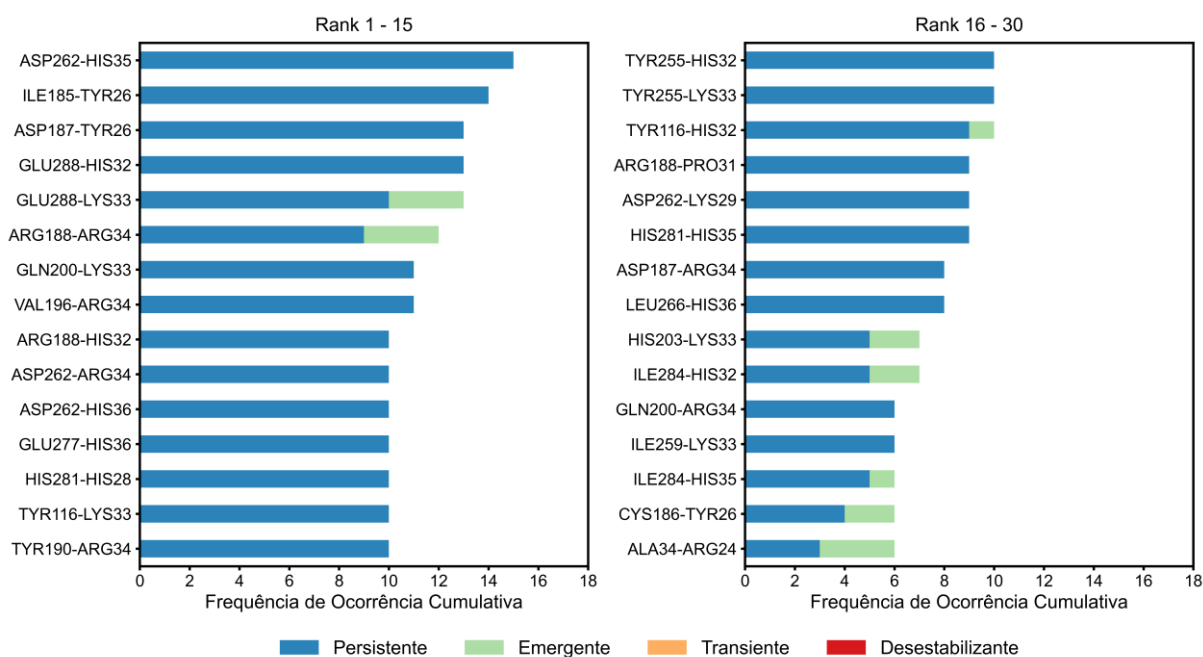
Fonte: Autor, 2026.

Nos sistemas individuais,  $M_1 - M_4$  apresentaram um padrão convergente, caracterizado pela manutenção do núcleo de interações composto pela tríade HIS32, HIS28 e HIS35, o conjunto de resíduos carregados positivamente ARG34, LYS29 e LYS33, e o resíduo PRO31, todos com elevado grau de conectividade. Paralelamente, foi identificado um comportamento recorrente de resíduos periféricos com maior variabilidade do perfil dinâmico, como ARG24, PHE39, TYR26 e LYS23, sugerindo um papel na modulação da interface sem comprometer a estabilidade do seu núcleo. Em contraste, o sistema  $M_5$  apresentou um perfil único, com predomínio de interações envolvendo resíduos na região N-terminal da TcP21, como MET1, ARG2, PHE11 e LEU12, implicando um modo de ligação divergente dos demais.

### 3.2.3.3 Recorrência de pares de resíduos TcP21–CXCR4

A análise dos pares mais recorrentes entre CXCR4 e TcP21, mostrada na Figura 3.15, indicou regiões de interação espacialmente distribuídas e funcionalmente coordenadas. A TcP21 não se limitou a uma única região do receptor, estabelecendo interações simultâneas com diferentes domínios transmembranares e loops extracelulares. Esse arranjo resultou em uma rede de contatos envolvendo os domínios ECL2, TM3, TM5, TM6 e TM7. Em termos globais, a rede foi dominada por contatos consistentemente estáveis, sem a ocorrência de interações transientes ou desestabilizantes, indicando elevada reprodutibilidade e persistência dinâmica desses pares.

**Figura 3.15** – Top 30 pares de resíduos entre TcP21 e CXCR4 de maior recorrência agrupada.



Fonte: Autor, 2026.

O mapeamento dos resíduos da TcP21 mostrou que todos os resíduos envolvidos nas interações estavam contidos na região de interface delimitada por ARG24–HIS36, em concordância com a proposta anterior. Essa região apresentou uma organização funcional que pôde ser subdividida em diferentes contribuições. O núcleo da interface foi definido pelos resíduos HIS32, LYS33 e ARG34, que atuaram

como centros de conectividade ao estabelecerem múltiplas interações com domínios distintos da CXCR4. Especificamente, ARG34 interagiu com ECL2, TM5 e TM6; HIS32 com ECL2, TM3, TM6 e TM7; e LYS33 com TM3, TM5, TM6 e TM7.

Em contraste, os resíduos TYR26 e PRO31, situados na porção N-terminal da interface, apresentaram interações preferenciais com o domínio ECL2, indicando possível contribuição para o ancoramento inicial da TcP21 ao receptor. Por outro lado, HIS35 e HIS36 interagiram majoritariamente com os domínios TM6 e TM7, sugerindo participação na estabilização do complexo em regiões mais profundas do vestíbulo extracelular.

O mapeamento dos resíduos na CXCR4 indicou uma organização hierárquica das regiões de interação, com domínios centrais e contribuições periféricas bem definidas. O domínio TM6 destacou-se como principal região de interação, especialmente por meio do resíduo ASP262, que apresentou elevada recorrência de contatos com o núcleo da interface da TcP21. Esse padrão sugeriu que TM6 atuou como um centro estrutural de interação, capaz de acomodar diferentes cadeias laterais. Tal observação foi consistente com dados estruturais prévios de GPCR (Hauser *et al.*, 2021; Hilger *et al.*, 2018; Zhou *et al.*, 2019), ainda que este estudo não tenha abordado diretamente mecanismos de ativação ou inibição do receptor.

O domínio ECL2, por sua vez, apresentou interações altamente reprodutíveis que caracterizaram uma superfície de ancoramento periférica. A combinação de propriedades hidrofóbicas e polares nessa região sugeriu flexibilidade local, permitindo acomodação conformacional sem perda de estabilidade, o que reforçou seu papel na estabilização da porção N-terminal da interface de interação da TcP21.

Além disso, os domínios TM5 e TM7 contribuíram para a expansão da rede de contatos em direção ao vestíbulo extracelular, atuando de forma cooperativa para aumentar a área de contato entre as proteínas. Ainda que menos recorrentes, essas contribuições reforçaram a ideia de que a estabilidade do complexo emergiu de múltiplas contribuições distribuídas. Por fim, resíduos aromáticos no domínio TM3 possivelmente desempenharam um papel na estabilização geométrica da interface, associado ao empacotamento local e à orientação das cadeias laterais.

A observação de interações “*Emergentes*” foi indicativa da ocorrência de rearranjos locais na interface de interação, contudo, como a distribuição majoritária de

contatos foi do tipo “*Persistentes*”, sugere-se que esses rearranjos possivelmente se restringiram a reorientações de cadeias laterais. Dado que a seleção dos pares se baseou na ocorrência cumulativa entre os sistemas, sem discriminação explícita entre diferentes modos de ligação, o sistema  $M_5$  foi incluído na análise.

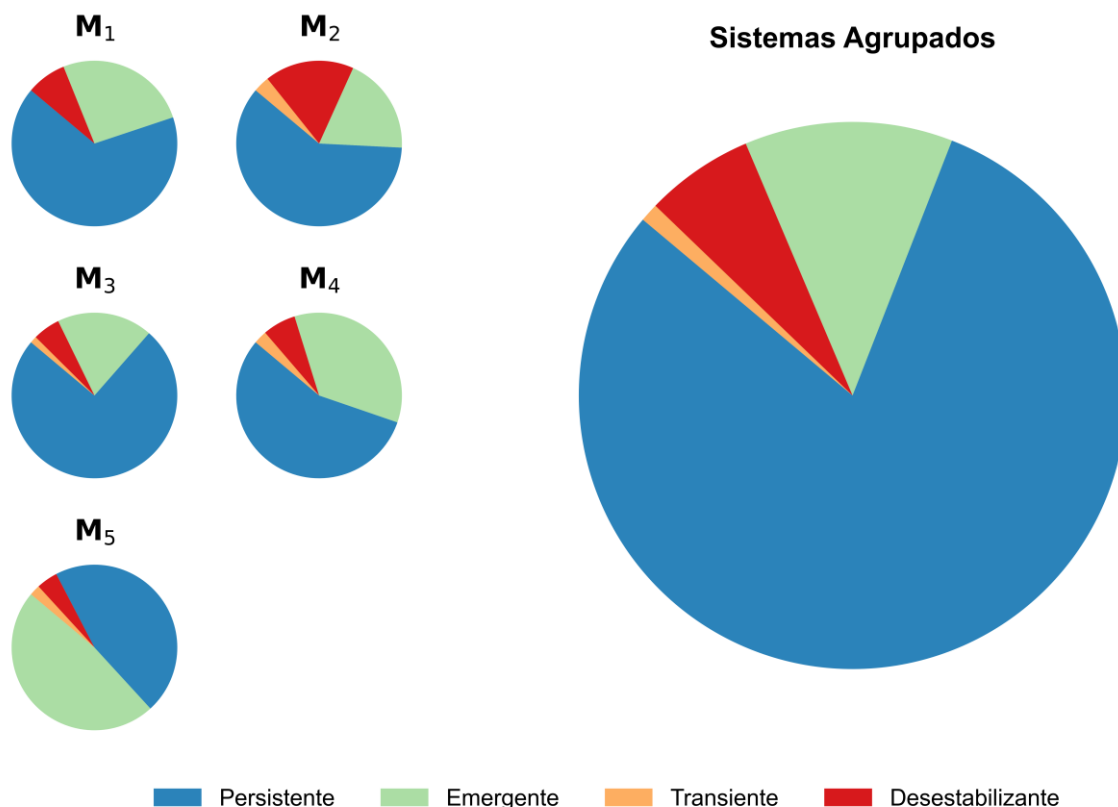
A alta reprodutibilidade de parte dos pares identificados foi compatível com a existência de um núcleo interativo conservado entre os sistemas. Por outro lado, a distribuição das interações por múltiplos domínios do receptor sugeriu que diferentes modos de ligação possam ter enfatizado subconjuntos específicos de contatos interfaciais, resultando em arranjos estruturalmente relacionados, porém distintos.

Assim, os dados sustentaram um modelo de interação no qual a TcP21 estabeleceu uma rede de contatos distribuída ao longo das regiões extracelulares e vestibulares da CXCR4, com interações predominantes nos domínios TM6 e ECL2 e estabilizada por contribuições complementares ao longo do receptor.

#### 3.2.3.4 *Distribuição do perfil temporal de contatos*

Em sequência, considerando-se agora todos os contatos observados e não apenas os 30 pares mais recorrentes, os perfis temporais foram distribuídos em regimes intrassistema e intersistema, conforme apresentado na Figura 3.16. A análise detalhada por sistema revelou heterogeneidade entre os sistemas, permitindo a caracterização individual de seus perfis de contatos.

**Figura 3.16** – Distribuição dos perfis temporais dos pares de resíduos entre TcP21 e CXCR4, por sistema e agrupado.



Fonte: Autor, 2026.

Os sistemas M<sub>1</sub>, M<sub>3</sub> e M<sub>4</sub> constituíram um grupo com comportamento temporal semelhante, marcado pela dominância de interações “*Persistentes*”.

O sistema M<sub>3</sub> destacou-se por apresentar a maior fração de contatos persistentes e a menor ocorrência de interações desestabilizantes, indicando que poucas reorganizações interfaciais ocorreram durante as simulações em relação ao estado inicial. O sistema M<sub>1</sub> seguiu essa tendência, embora com maior participação de interações “*Emergentes*”, sugerindo ajustes estruturais adicionais ao longo da simulação. Em contraste, o sistema M<sub>4</sub> apresentou o maior número de interações emergentes entre todos os sistemas, indicando maior flexibilidade da interface, ainda que mantendo um núcleo estável predominante.

De modo geral, esses resultados sugeriram que, para esse grupo, os modos de ligação iniciais já representaram configurações favoráveis para a interação entre TcP21 e CXCR4.



O sistema  $M_2$  ocupou uma posição intermediária no conjunto analisado. Apesar da predominância de interações “*Persistentes*”, esse sistema apresentou a maior ocorrência de contatos “*Desestabilizantes*”. Esse comportamento indicou uma redução relativa na estabilidade dinâmica, em vez de um processo predominantemente adaptativo. Além disso, ao contrário do observado para o sistema  $M_4$ , a menor contribuição de interações “*Emergentes*” sugeriu a ausência de compensação efetiva para as perdas de contato.

O sistema  $M_5$  apresentou o comportamento mais discrepante entre os sistemas analisados. Diferentemente dos demais, exibiu predominância de interações “*Emergentes*”, em detrimento das interações “*Persistentes*”, configurando, assim, um perfil dinâmico distinto. Esse padrão sugeriu que a estabilidade observada foi alcançada principalmente por meio da formação de novos contatos ao longo da trajetória, e não pela preservação de um núcleo interativo inicial. A reduzida ocorrência de interações “*Desestabilizantes*” indicou que a divergência observada não resultou de instabilidade acentuada, mas sim de uma configuração inicial com menor densidade de contatos. Assim, as interações emergentes passaram a desempenhar um papel central na estabilização do sistema. Em função desse comportamento, o sistema  $M_5$  não foi incluído na análise agrupada, de forma a evitar a introdução de viés associado a um modo de ligação possivelmente menos representativo.

A análise do conjunto agrupado evidenciou a predominância de contatos classificados como “*Persistentes*”. Esse resultado indicou que, ao longo das 20 trajetórias, a interação entre TcP21 e CXCR4 foi caracterizada por elevada persistência e por estabilidade estrutural, compatível com a existência de um núcleo interativo bem definido. Adicionalmente, foi observada uma contribuição secundária de contatos “*Emergentes*”, sugerindo que, embora os modos de ligação iniciais tenham desempenhado um papel central, houve adaptação estrutural com formação de novos contatos estáveis ao longo das simulações. Esse comportamento foi consistente com mecanismos de ajuste conformacional induzido.

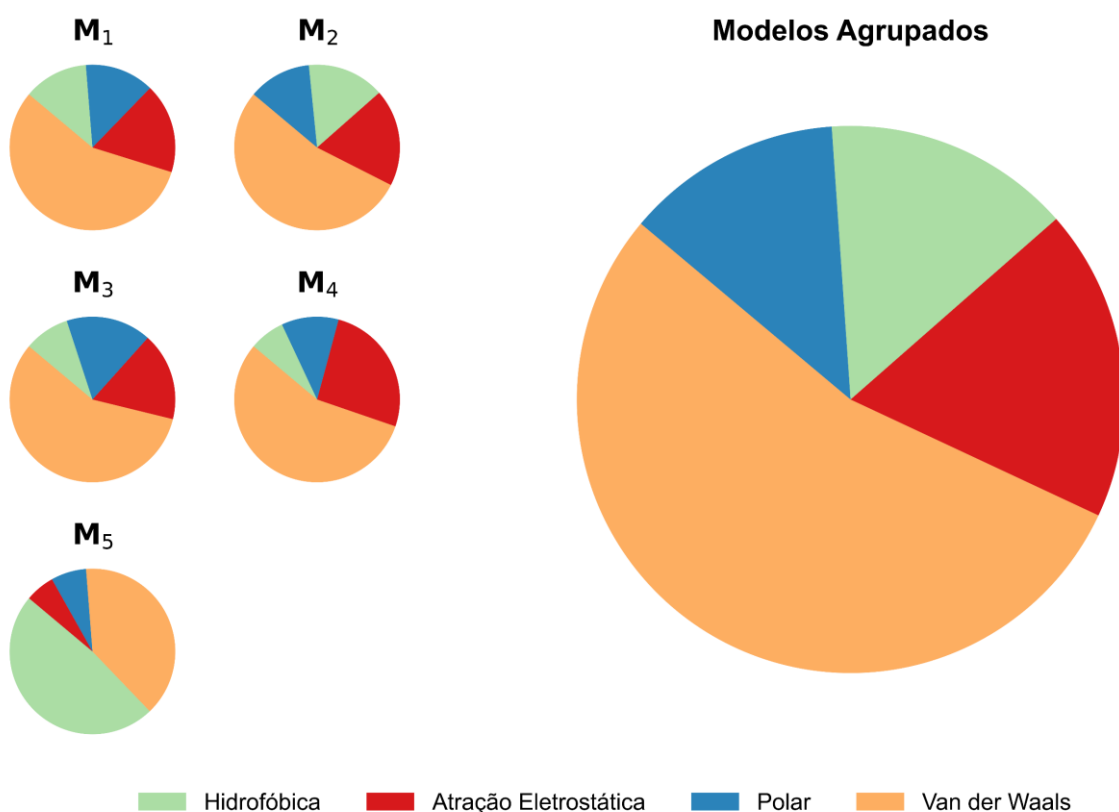
Por outro lado, contatos “*Desestabilizantes*” foram pouco frequentes, e interações “*Transientes*” praticamente não contribuíram para a distribuição global. Essa distribuição sugeriu que a dissociação das interações iniciais foi incomum e predominantemente persistente, não refletindo flutuações reversíveis. Em conjunto,

os resultados descreveram um mecanismo de interação governado por um núcleo interativo estável, complementado por contribuições adaptativas de menor magnitude.

### 3.2.3.5 Distribuição qualitativa das forças de interação

Por fim, realizou-se uma análise qualitativa da distribuição das forças intermoleculares estabilizantes do complexo. Conforme descrito na seção 3.1.6.1, a classificação baseou-se na inferência da natureza das interações a partir das propriedades físico-químicas dos resíduos envolvidos, ponderadas por seus valores relativos de ocorrência. Os resultados são apresentados na Figura 3.17.

**Figura 3.17** – Distribuição das forças intermoleculares de estabilização entre os pares de resíduos TcP21–CXCR4, por sistema e agrupado.



Fonte: Autor, 2026.

A avaliação individual dos sistemas  $M_1 - M_4$  revelou um perfil consistente, com predominância de interações de van der Waals, seguidas por contribuições eletrostáticas e por um balanço semelhante entre contatos polares e hidrofóbicos. Esse arranjo sugere que a estabilização do complexo seja majoritariamente conduzida por complementaridade estérica e empacotamento interfacial, enquanto as demais interações contribuam para o refinamento da estabilidade e da especificidade. As diferenças observadas entre esses sistemas foram essencialmente de natureza quantitativa, sugerindo ajustes conformacionais dentro de um mesmo regime de interação, e não a presença de mecanismos de interação fundamentalmente distintos.

Em contraste, o sistema  $M_5$  apresentou um perfil divergente, caracterizado pela predominância de interações hidrofóbicas e pela redução expressiva de interações de van der Waals, além de participação mínima de contribuições eletrostáticas e polares. Esse comportamento sugeriu um regime de estabilização distinto, associado a uma menor eficiência de empacotamento e à limitação na formação de interações direcionais. Dessa forma, o sistema  $M_5$  foi interpretado novamente como um estado menos representativo que os demais, não sendo incluído na análise agrupada e mantido apenas para avaliação individual.

Restrita ao subconjunto de sistemas  $M_1 - M_4$ , a análise agrupada sugeriu que interações do tipo van der Waals constituíram a principal força estabilizante do complexo TcP21–CXCR4. A magnitude dessa força reforçou a importância do empacotamento interfacial e da complementaridade geométrica na manutenção da interface.

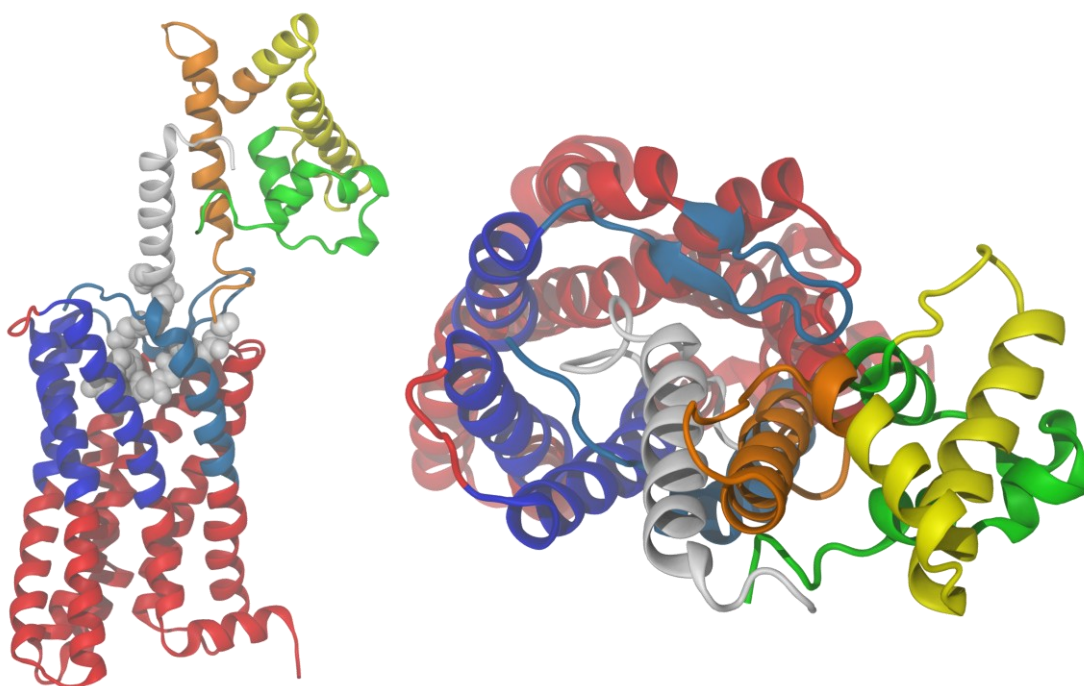
Adicionalmente, interações eletrostáticas, hidrofóbicas e polares contribuíram em proporções relativamente equilibradas, sugerindo atuação conjunta na modulação da estabilidade. Embora individualmente menos expressivas, essas interações contribuíram possivelmente para o refinamento da interface e para a especificidade dos modos de ligação. Assim, os dados sugeriram um modelo de interação em que o núcleo do complexo é definido por interações de van der Waals, sendo reforçado por contribuições auxiliares que aumentam a especificidade e a direcionalidade das interações.

### 3.2.4 *Correlação das interfaces de interação identificadas*

De maneira geral, os resultados indicaram um modelo de interação marcado por pronunciada assimetria estrutural entre TcP21 e CXCR4. Enquanto a TcP21 apresentou uma região de interface contínua e bem definida, centrada na região dos resíduos LYS23–PHE39, a CXCR4 exibiu uma interface distribuída, envolvendo múltiplos domínios extracelulares e transmembranares.

A interação foi organizada em dois níveis funcionais complementares. O primeiro correspondeu a um núcleo estabilizador, formado principalmente pelos domínios TM6, TM7 e TM5, responsável pela manutenção da estabilidade conformacional do complexo no interior do vestíbulo extracelular. O segundo nível envolveu os domínios ECL2 e TM1/N-terminal, que atuaram como regiões periféricas de ancoragem, contribuindo para ajustes conformacionais e acomodação estrutural na região extracelular. A Figura 3.18 ilustra o complexo TcP21–CXCR4, com os respectivos vestíbulos extracelulares dos domínios e região de interface destacadas.

**Figura 3.18** – Complexo interativo TcP21–CXCR4. Em azul escuro: domínios TM5, TM6 e TM7. Em azul claro: domínios ECL2 e TM1/N-terminal. Da esquerda para a direita: vistas frontal e superior.



Essa configuração heterogênea favoreceu a formação de uma interface ampla e dinâmica, consistente com um mecanismo de ajuste conformacional induzido. O núcleo interativo apresentou elevada estabilidade, enquanto resíduos periféricos, dotados de maior plasticidade conformacional, contribuíram para a modulação da interface por meio do estabelecimento de contatos emergentes.

Sob uma perspectiva mecanicista, a estabilidade das interações entre TcP21 e CXCR4 foi predominantemente governada por interações de van der Waals, indicando a importância da complementaridade geométrica entre as superfícies de contato das biomoléculas. Em paralelo, interações eletrostáticas e polares contribuíram para a especificidade e direcionalidade das interações.

### 3.3 CONCLUSÃO

---

No presente estudo, múltiplos modos de ligação do complexo TcP21–CXCR4 foram obtidos por ancoragem molecular e subsequentemente avaliados por simulações de dinâmica molecular multirréplicas em ambiente biomimético, incluindo bicamada lipídica com composição análoga à de células de macrófagos. A análise de consenso entre as trajetórias identificou um segmento contínuo na TcP21, compreendendo os resíduos LYS23–PHE39, com características típicas de interface de interação proteína-proteína. Na CXCR4, os domínios TM5, TM6 e TM7 constituíram um núcleo estrutural estabilizado no interior do vestíbulo extracelular, enquanto os domínios ECL2 e TM1/N-terminal apresentaram maior flexibilidade conformacional. Essa organização estrutural assimétrica favoreceu a formação de uma interface ampla e modular, sustentada por uma rede de contatos intermoleculares consistentes. A estabilização do complexo foi predominantemente governada por interações de van der Waals, indicativas de complementaridade geométrica, com contribuições adicionais de interações eletrostáticas e polares, conferindo direcionalidade ao reconhecimento molecular.

## 4 CONCLUSÃO GERAL

---

Decorridos 117 anos desde sua descoberta, a doença de Chagas permanece um mal que impacta profundamente o bem-estar e as condições socioeconômicas de milhares de indivíduos ao redor do mundo. Apesar dos avanços no entendimento de sua fisiopatologia, ainda não há um tratamento plenamente eficaz. O desenvolvimento de novas estratégias terapêuticas é limitado por fatores como a complexidade biológica do parasita *T. cruzi*, a elucidação incompleta de seus mecanismos de infecção e a dificuldade na identificação de alvos moleculares e sítios de ligação suscetíveis à modulação. Nesse contexto, abordagens computacionais configuram-se como ferramentas estratégicas de baixo custo para a prospecção inicial de alvos moleculares, contribuindo para o direcionamento de estudos experimentais.

Na presente tese, a proteína TcP21 foi caracterizada estrutural e funcionalmente *in silico* por meio de dois estudos complementares. Primeiramente, a identificação de um sítio de ligação consistente entre diferentes metodologias sugeriu a existência de uma região capaz de estabelecer interações dinamicamente estáveis com grupos químicos específicos. Em paralelo, a caracterização da interface TcP21–CXCR4 sugeriu a existência de um arranjo estrutural compatível com interface de interação proteína-proteína, sustentada predominantemente por complementaridade geométrica, com contribuições eletrostáticas e polares conferindo direcionalidade ao processo de associação.

A análise conjunta desses resultados sugeriu a existência de regiões funcionalmente especializadas na TcP21, no contexto do reconhecimento molecular de moléculas farmacofóricas e da interação com o receptor do hospedeiro. Nesse contexto, os resultados oferecem uma perspectiva direcionada sobre determinantes estruturais potencialmente associados à função dessa proteína.

Na ausência de validação experimental, os resultados devem ser interpretados como observações preliminares. Nesse sentido, ensaios de mutagênese sítio-dirigida poderiam ser empregados para a validação experimental das regiões propostas. Ademais, estratégias como o desenho racional de fármacos ou a engenharia de epítopos poderiam ser empregadas como abordagens voltadas à modulação seletiva das interações mediadas pela TcP21, sob um possível viés terapêutico.

## REFERÊNCIAS

---

ABRAHAM, Mark James; MURTOLA, Teemu; SCHULZ, Roland; PÁLL, Szilárd; SMITH, Jeremy C.; HESS, Berk; LINDAHL, Erik. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, sep. 2015. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>.

ABRAMSON, Josh et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. **Nature**, v. 630, n. 8016, p. 493–500, jun. 2024. DOI: <https://doi.org/10.1038/s41586-024-07487-w>.

AGUDELO HIGUITA, Nelson Iván et al. Chagas disease in the United States: a call for increased investment and collaborative research. **The Lancet Regional Health - Americas**, v. 34, p. 100768, jun. 2024. DOI: <https://doi.org/10.1016/j.lana.2024.100768>.

AGUDELO HIGUITA, Nelson Iván; HENAO-MARTÍNEZ, Andrés F; VIEIRA BATISTA, Marjorie; HERRERA BERNAL, Claudia Patricia; REZENDE FILHO, Joffre; RASSI, Anis. State-of-the-Art Review: Chagas Disease—an Enduring Challenge. **Clinical Infectious Diseases**, v. 82, n. 2, p. 187–205, feb. 2026. DOI: <https://doi.org/10.1093/cid/ciaf654>.

AHMED, Furyal; BROOKS, Charles L. FASTDock: A Pipeline for Allosteric Drug Discovery. **Journal of Chemical Information and Modeling**, v. 63, n. 22, p. 7219–7227, nov. 2023. DOI: <https://doi.org/10.1021/acs.jcim.3c00895>.

ALARCÓN DE NOYA, Belkisyole; DÍAZ-BELLO, Zoraida; RUIZ-GUEVARA, Raiza; NOYA, Oscar. Chagas Disease Expands Its Epidemiological Frontiers From Rural to Urban Areas. **Frontiers in Tropical Diseases**, v. 3, apr. 2022. DOI: <https://doi.org/10.3389/fitd.2022.799009>.

ALLEN, Karen N.; BELLAMACINA, Cornelia R.; DING, Xiaochun; JEFFERY, Constance J.; MATTOS, Carla; PETSKO, Gregory A.; RINGE, Dagmar. An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. **The Journal of Physical Chemistry**, v. 100, n. 7, p. 2605–2611, jan. 1996. DOI: <https://doi.org/10.1021/jp952516o>.

ALMEIDA-SILVA, Juliana; MENEZES, Diego Silva; FERNANDES, Juan Mateus Pereira; ALMEIDA, Márcio Cerqueira; VASCO-DOS-SANTOS, Deyvison Rhuan; SARAIVA, Roberto Magalhães; VIÇOSA, Alessandra Lifschitz; PEREZ, Sandra Aurora Chavez; ANDRADE, Sônia Gumes; SUAREZ-FONTES, Ana Márcia; VANNIER-SANTOS, Marcos André. The repositioned drugs disulfiram/diethyldithiocarbamate combined to benznidazole: Searching for Chagas disease selective therapy, preventing toxicity and drug resistance. **Frontiers in Cellular and Infection Microbiology**, v. 12, jul. 2022. DOI: <https://doi.org/10.3389/fcimb.2022.926699>.



ALVAREZ-GARCIA, Daniel; BARRIL, Xavier. Molecular Simulations with Solvent Competition Quantify Water Displaceability and Provide Accurate Interaction Maps of Protein Binding Sites. **Journal of Medicinal Chemistry**, v. 57, n. 20, p. 8530–8539, oct. 2014. DOI: <https://doi.org/10.1021/jm5010418>.

ARCON, Juan Pablo; DEFELIPE, Lucas A.; MODENUTTI, Carlos P.; LÓPEZ, Elias D.; ALVAREZ-GARCIA, Daniel; BARRIL, Xavier; TURJANSKI, Adrián G.; MARTÍ, Marcelo A. Molecular Dynamics in Mixed Solvents Reveals Protein–Ligand Interactions, Improves Docking, and Allows Accurate Binding Free Energy Predictions. **Journal of Chemical Information and Modeling**, v. 57, n. 4, p. 846–863, apr. 2017. DOI: <https://doi.org/10.1021/acs.jcim.6b00678>.

ASHIZAWA, Ryota et al. Modeling Protein–Protein and Protein–Ligand Interactions by the ClusPro Team in CASP16. **Proteins: Structure, Function, and Bioinformatics**, v. 94, n. 1, p. 183–191, jan. 2026. DOI: <https://doi.org/10.1002/prot.70066>.

AZEVEDO SILVEIRA, Anna Clara; CRISTINA DE OLIVEIRA, Rayane; RODRIGUES, Cassiano Costa; TEIXEIRA, Samuel Cota; BORGES, Bruna Cristina; VIEIRA DA SILVA, Claudio. Trypanosoma cruzi infection induces proliferation and impairs migration of a human breast cancer cell line. **Experimental Parasitology**, v. 245, p. 108443, feb. 2023. DOI: <https://doi.org/10.1016/j.exppara.2022.108443>.

BAEK, Minkyung et al. Accurate prediction of protein structures and interactions using a three-track neural network. **Science**, v. 373, n. 6557, p. 871–876, aug. 2021. DOI: <https://doi.org/10.1126/science.abj8754>.

BARBOSA, Jeranice Silva; MOURA, Francielle Borges Rosa De; FERREIRA, Bruno Antonio; MARTINS, Flávia Alves; MUNIZ, Elusca Helena; GOMIDE, José Augusto Leoncio; SILVA, Cláudio Vieira Da; RIBEIRO, Daniele Lisboa; ARAÚJO, Fernanda de Assis; TOMIOSSO, Tatiana Carla. Recombinant Protein rP21 from Trypanosoma cruzi has Effect on Inflammation, Angiogenesis and Fibrogenesis in Skin Wound Model C57BL/6 Mouse. **Advances in Research**, p. 28–37, mar. 2021. DOI: <https://doi.org/10.9734/air/2021/v22i130285>.

BATES, Benjamin R.; CARRASCO-TENEZACA, Majo; MENDEZ-TRIVINO, Angela M.; MENDOZA, Luis E.; NIETO-SANCHEZ, Claudia; BAUS, Esteban G.; GRIJALVA, Mario J. Identifying Barriers and Facilitators for Home Reconstruction for Prevention of Chagas Disease: An Interview Study in Rural Loja Province, Ecuador. **Tropical Medicine and Infectious Disease**, v. 8, n. 4, p. 228, apr. 2023. DOI: <https://doi.org/10.3390/tropicalmed8040228>.

BEATTY, Norman L.; ARANGO-FERREIRA, Catalina; GUAL-GONZALEZ, Lúdia; ZULUAGA, Sara; NOLAN, Melissa S.; CANTILLO-BARRAZA, Omar. Oral Chagas Disease in Colombia—Confirmed and Suspected Routes of Transmission. **Tropical Medicine and Infectious Disease**, v. 9, n. 1, p. 14, jan. 2024. DOI: <https://doi.org/10.3390/tropicalmed9010014>.

BECKE, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. **Physical Review A**, v. 38, n. 6, p. 3098–3100, sep. 1988. DOI: <https://doi.org/10.1103/PhysRevA.38.3098>.

BELAPURE, Jaydeep; SOROKINA, Marija; KASTRITIS, Panagiotis L. IRAA: A statistical tool for investigating a protein–protein interaction interface from multiple structures. **Protein Science**, v. 32, n. 1, jan. 2023. DOI: <https://doi.org/10.1002/pro.4523>.

BERENDSEN, H. J. C.; POSTMA, J. P. M.; VAN GUNSTEREN, W. F.; DINOLA, A.; HAAK, J. R. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, v. 81, n. 8, p. 3684–3690, oct. 1984. DOI: <https://doi.org/10.1063/1.448118>.

BERENSTEIN, A. J.; FALK, N.; MOSCATELLI, G.; MORONI, S.; GONZÁLEZ, N.; GARCIA -BOURNISSEN, F.; BALLERING, G.; FREILIJ, H.; ALTCHER, J. Adverse Events Associated with Nifurtimox Treatment for Chagas Disease in Children and Adults. **Antimicrobial Agents and Chemotherapy**, v. 65, n. 2, jan. 2021. DOI: <https://doi.org/10.1128/AAC.01135-20>.

BERNETTI, Mattia; BUSSI, Giovanni. Pressure control using stochastic cell rescaling. **The Journal of Chemical Physics**, v. 153, n. 11, sep. 2020. DOI: <https://doi.org/10.1063/5.0020514>.

BIANCHI, Marco E.; MEZZAPELLE, Rosanna. The Chemokine Receptor CXCR4 in Cell Proliferation and Tissue Regeneration. **Frontiers in Immunology**, v. 11, aug. 2020. DOI: <https://doi.org/10.3389/fimmu.2020.02109>.

BOCCA TOURRES, C L. [Acute period of Chagas' disease and its treatment with Bay 2502]. **Boletín chileno de parasitología**, v. 24, n. 1, p. 24–7, 1969.

BOCK, M; GÖNNERT, R; HABERKORN, A. Studies with Bay 2502 on animals. **Boletín chileno de parasitología**, v. 24, n. 1, p. 13–9, 1969.

BOGAN, Andrew A; THORN, Kurt S. Anatomy of hot spots in protein interfaces. **Journal of Molecular Biology**, v. 280, n. 1, p. 1–9, jul. 1998. DOI: <https://doi.org/10.1006/jmbi.1998.1843>.

BORGES, Bruna Cristina; UEHARA, Isadora Akemi; DOS SANTOS, Marlus Alves; MARTINS, Flávia Alves; DE SOUZA, Fernanda Carvalho; JUNIOR, Álvaro Ferreira; DA LUZ, Felipe Andrés Cordero; DA COSTA, Mylla Spirandelli; NOTÁRIO, Ana Flávia Oliveira; LOPES, Daiana Silva; TEIXEIRA, Samuel Cota; TEIXEIRA, Thaise Lara; DE CASTILHOS, Patrícia; DA SILVA, Claudio Vieira; SILVA, Marcelo José Barbosa. The Recombinant Protein Based on Trypanosoma cruzi P21 Interacts With CXCR4 Receptor and Abrogates the Invasive Phenotype of Human Breast Cancer Cells. **Frontiers in Cell and Developmental Biology**, v. 8, oct. 2020. DOI: <https://doi.org/10.3389/fcell.2020.569729>.

BOWIE, James U.; LÜTHY, Roland; EISENBERG, David. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. **Science**, v. 253, n. 5016, p. 164–170, jul. 1991. DOI: <https://doi.org/10.1126/science.1853201>.

BRENEMAN, Curt M.; WIBERG, Kenneth B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. **Journal of Computational Chemistry**, v. 11, n. 3, p. 361–373, apr. 1990. DOI: <https://doi.org/10.1002/jcc.540110311>.

BRENKE, Ryan; KOZAKOV, Dima; CHUANG, Gwo-Yu; BEGLOV, Dmitri; HALL, David; LANDON, Melissa R.; MATTOS, Carla; VAJDA, Sandor. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. **Bioinformatics**, v. 25, n. 5, p. 621–627, mar. 2009. DOI: <https://doi.org/10.1093/bioinformatics/btp036>.

BUCHHOLZ, Caroline R.; SNEDDON, Molly S.; MCPHERSON, Joseph E.; DAS, Sourav; GEE, Clifford T.; GRILLO, Michael J.; CHAI, Sergio C.; LEE, Richard E.; CHEN, Taosheng; HARKI, Daniel A.; SHELAT, Anang A.; POMERANTZ, William C. K. Fragment-Based NMR Screening of the BPTF PHD Finger Methyl Lysine Reader Leads to the First Small-Molecule Inhibitors. **ACS Medicinal Chemistry Letters**, v. 14, n. 10, p. 1441–1447, oct. 2023. DOI: <https://doi.org/10.1021/acsmedchemlett.3c00343>.

BURLEIGH, Barbara A; ANDREWS, Norma W. Signaling and host cell invasion by *Trypanosoma cruzi*. **Current Opinion in Microbiology**, v. 1, n. 4, p. 461–465, aug. 1998. DOI: [https://doi.org/10.1016/S1369-5274\(98\)80066-0](https://doi.org/10.1016/S1369-5274(98)80066-0).

BURTOLOSO, Antonio C. B.; DE ALBUQUERQUE, Sérgio; FURBER, Mark; GOMES, Juliana C.; GONÇALEZ, Cristiana; KENNY, Peter W.; LEITÃO, Andrei; MONTANARI, Carlos A.; QUILLES, José Carlos; RIBEIRO, Jean F. R.; ROCHA, Josmar R. Anti-trypanosomal activity of non-peptidic nitrile-based cysteine protease inhibitors. **PLOS Neglected Tropical Diseases**, v. 11, n. 2, p. e0005343, feb. 2017. DOI: <https://doi.org/10.1371/journal.pntd.0005343>.

BUSSI, Giovanni; DONADIO, Davide; PARRINELLO, Michele. Canonical sampling through velocity rescaling. **The Journal of Chemical Physics**, v. 126, n. 1, jan. 2007. DOI: <https://doi.org/10.1063/1.2408420>.

CHAGAS, Carlos. 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, n. 2, p. 159–218, aug. 1909. DOI: <https://doi.org/10.1590/S0074-02761909000200008>.

CHAN, Wallace K. B.; DASGUPTA, Debarati; CARLSON, Heather A.; TRAYNOR, John R. Mixed-solvent molecular dynamics simulation-based discovery of a putative allosteric site on regulator of G protein signaling 4. **Journal of Computational Chemistry**, v. 42, n. 30, p. 2170–2180, nov. 2021. DOI: <https://doi.org/10.1002/jcc.26747>.

CHAN, Wallace K.B.; CARLSON, Heather A.; TRAYNOR, John R. Application of Mixed-Solvent Molecular Dynamics Simulations for Prediction of Allosteric Sites on G Protein–Coupled Receptors. **Molecular Pharmacology**, v. 103, n. 5, p. 274–285, may 2023. DOI: <https://doi.org/10.1124/molpharm.122.000612>.

COHEN, Jacob. **Statistical Power Analysis for the Behavioral Sciences**. [S.l.]: Routledge, 2013.

COLOVOS, Chris; YEATES, Todd O. Verification of protein structures: Patterns of nonbonded atomic interactions. **Protein Science**, v. 2, n. 9, p. 1511–1519, sep. 1993. DOI: <https://doi.org/10.1002/pro.5560020916>.

CONTEH, Lesong; ENGELS, Thomas; MOLYNEUX, David H. Socioeconomic aspects of neglected tropical diseases. **The Lancet**, v. 375, n. 9710, p. 239–247, jan. 2010. DOI: [https://doi.org/10.1016/S0140-6736\(09\)61422-7](https://doi.org/10.1016/S0140-6736(09)61422-7).

COUSIN, Ewerton et al. Global, regional, and national burden of Chagas disease, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. **The Lancet Infectious Diseases**, v. 26, n. 3, p. 284–301, mar. 2026. DOI: [https://doi.org/10.1016/S1473-3099\(25\)00562-6](https://doi.org/10.1016/S1473-3099(25)00562-6).

DA SILVA, Claudio V.; KAWASHITA, Silvia Y.; PROBST, Christian M.; DALLAGIOVANNA, Bruno; CRUZ, Mário C.; DA SILVA, Erika A.; SOUTO-PADRÓN, Thaís C.B.S.; KRIEGER, Marco A.; GOLDENBERG, Samuel; BRIONES, Marcelo R.S.; ANDREWS, Norma W.; MORTARA, Renato A. Characterization of a 21 kDa protein from *Trypanosoma cruzi* associated with mammalian cell invasion. **Microbes and Infection**, v. 11, n. 5, p. 563–570, apr. 2009. DOI: <https://doi.org/10.1016/j.micinf.2009.03.007>.

DARDEN, Tom; YORK, Darrin; PEDERSEN, Lee. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089–10092, jun. 1993. DOI: <https://doi.org/10.1063/1.464397>.

DASGUPTA, Debarati; MEHRANI, Ramin; CARLSON, Heather A.; SHARMA, Sumit. Identifying Potential Ligand Binding Sites on Glycogen Synthase Kinase 3 Using Atomistic Cosolvent Simulations. **ACS Applied Bio Materials**, v. 7, n. 2, p. 588–595, feb. 2024. DOI: <https://doi.org/10.1021/acsabm.2c01079>.

DASGUPTA, Debarati; CHAN, Wallace K. B.; CARLSON, Heather A. Computational Identification of Possible Allosteric Sites and Modulators of the SARS-CoV-2 Main Protease. **Journal of Chemical Information and Modeling**, v. 62, n. 3, p. 618–626, feb. 2022. DOI: <https://doi.org/10.1021/acs.jcim.1c01223>.

DE OLIVEIRA JUNIOR, Wilson Alves; GÓMEZ I PRAT, Jordi; ALBAJAR-VIÑAS, Pedro; CARRAZZONE, Cristina; KROPF, Simone Petraglia; DEHOUSSE, Aurore; CAMARGO, Ana Maria de Arruda; ANSELM, Mariella; BARBA, Maria Cristina Parada; GUIU, Isabel Claveria; BARROS, Maria das Neves Dantas Silveira; CAVALVANTI, Maria da Glória Melo; CORREIA, Cassandra Barros; MARTINS, Silvia Marinho; FINDECHAGAS WORKGROUP. How people affected by Chagas disease have struggled with their negligence: history, associative movement and World Chagas Disease Day. **Memórias do Instituto Oswaldo Cruz**, v. 117, 2022. DOI: <https://doi.org/10.1590/0074-02760220066>.

DE SOUSA, Débora Raysa Teixeira; DE OLIVEIRA GUERRA, Jorge Augusto; ORTIZ, Jessica Vanina; DO NASCIMENTO COUCEIRO, Katia; DA SILVA E SILVA, Monica Regina Hosannah; JORGE BRANDÃO, Alba Regina; GUEVARA, Elsa; ARCANJO, Ana Ruth Lima; DE OLIVEIRA JÚNIOR, Edival Ferreira; SMITH-DORIA, Susan; MWANGI, Victor Irungu; MORAIS, Rômulo Freire; SILVA, George Allan Villarouco; MOLINA, Israel; SILVEIRA, Henrique; FERREIRA, João Marcos Bemfica Barbosa; GUERRA, Maria das Graças Vale Barbosa. Acute Chagas disease associated with ingestion of contaminated food in Brazilian western Amazon. **Tropical Medicine & International Health**, v. 28, n. 7, p. 541–550, jul. 2023. DOI: <https://doi.org/10.1111/tmi.13899>.

DE SOUZA, Guilherme; TEIXEIRA, Samuel Cota; FAJARDO MARTÍNEZ, Aryani Felix; SILVA, Rafaela José; LUZ, Luana Carvalho; LIMA JÚNIOR, Joed Pires De; ROSINI, Alessandra Monteiro; DOS SANTOS, Natália Carine Lima; DE OLIVEIRA, Rafael Martins; PASCHOALINO, Marina; BARBOSA, Matheus Carvalho; ALVES, Rosiane Nascimento; GOMES, Angelica Oliveira; DA SILVA, Claudio Vieira; FERRO, Eloisa Amália Vieira; BARBOSA, Bellisa Freitas. Trypanosoma cruzi P21 recombinant protein modulates Toxoplasma gondii infection in different experimental models of the human maternal–fetal interface. **Frontiers in Immunology**, v. 14, oct. 2023. DOI: <https://doi.org/10.3389/fimmu.2023.1243480>.

DE VRIES, Sjoerd; BONVIN, Alexandre. How Proteins Get in Touch: Interface Prediction in the Study of Biomolecular Complexes. **Current Protein & Peptide Science**, v. 9, n. 4, p. 394–406, aug. 2008. DOI: <https://doi.org/10.2174/138920308785132712>.

DECHENE, Michelle; WINK, Glenna; SMITH, Mychal; SWARTZ, Paul; MATTOS, Carla. Multiple solvent crystal structures of ribonuclease A: An assessment of the method. **Proteins: Structure, Function, and Bioinformatics**, v. 76, n. 4, p. 861–881, sep. 2009. DOI: <https://doi.org/10.1002/prot.22393>.

DESALE, Hans; HERRERA, Claudia; DUMONTEIL, Eric. Trypanosoma cruzi amastigote transcriptome analysis reveals heterogenous populations with replicating and dormant parasites. **Microbes and Infection**, v. 26, n. 1–2, p. 105240, jan. 2024. DOI: <https://doi.org/10.1016/j.micinf.2023.105240>.

DODDA, Leela S.; CABEZA DE VACA, Israel; TIRADO-RIVES, Julian; JORGENSEN, William L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. **Nucleic Acids Research**, v. 45, n. W1, p. W331–W336, jul. 2017. DOI: <https://doi.org/10.1093/nar/gkx312>.

EBERHARDT, Jerome; SANTOS-MARTINS, Diogo; TILLACK, Andreas F.; FORLI, Stefano. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. **Journal of Chemical**

**Information and Modeling**, v. 61, n. 8, p. 3891–3898, aug. 2021. DOI: <https://doi.org/10.1021/acs.jcim.1c00203>.

ENGLISH, Andrew C.; DONE, Sarah H.; CAVES, Leo S.D.; GROOM, Colin R.; HUBBARD, Roderick E. Locating interaction sites on proteins: The crystal structure of thermolysin soaked in 2% to 100% isopropanol. **Proteins: Structure, Function, and Genetics**, v. 37, n. 4, p. 628–640, dec. 1999. DOI: [https://doi.org/10.1002/\(SICI\)1097-0134\(19991201\)37:4<628::AID-PROT13>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0134(19991201)37:4<628::AID-PROT13>3.0.CO;2-G).

ENGLISH, Andrew C.; GROOM, Colin R.; HUBBARD, Roderick E. Experimental and computational mapping of the binding surface of a crystalline protein. **Protein Engineering, Design and Selection**, v. 14, n. 1, p. 47–59, jan. 2001. DOI: <https://doi.org/10.1093/protein/14.1.47>.

ESSMANN, Ulrich; PERERA, Lalith; BERKOWITZ, Max L.; DARDEN, Tom; LEE, Hsing; PEDERSEN, Lee G. A smooth particle mesh Ewald method. **The Journal of Chemical Physics**, v. 103, n. 19, p. 8577–8593, nov. 1995. DOI: <https://doi.org/10.1063/1.470117>.

FEHER, Victoria A; DURRANT, Jacob D; VAN WART, Adam T; AMARO, Rommie E. Computational approaches to mapping allosteric pathways. **Current Opinion in Structural Biology**, v. 25, p. 98–103, apr. 2014. DOI: <https://doi.org/10.1016/j.sbi.2014.02.004>.

FENG, Yanjun; GONG, Changda; ZHU, Jieyu; LIU, Guixia; TANG, Yun; LI, Weihua. Unraveling the Ligand-Binding Sites of CYP3A4 by Molecular Dynamics Simulations with Solvent Probes. **Journal of Chemical Information and Modeling**, v. 64, n. 8, p. 3451–3464, apr. 2024. DOI: <https://doi.org/10.1021/acs.jcim.4c00089>.

FOSTER, Simon R.; HAUSER, Alexander S.; VEDEL, Line; STRACHAN, Ryan T.; HUANG, Xi-Ping; GAVIN, Ariana C.; SHAH, Sushrut D.; NAYAK, Ajay P.; HAUGAARD-KEDSTRÖM, Linda M.; PENN, Raymond B.; ROTH, Bryan L.; BRÄUNER-OSBORNE, Hans; GLORIAM, David E. Discovery of Human Signaling Systems: Pairing Peptides to G Protein-Coupled Receptors. **Cell**, v. 179, n. 4, p. 895–908.e21, oct. 2019. DOI: <https://doi.org/10.1016/j.cell.2019.10.010>.

FRANCISCO, Amanda F.; JAYAWARDHANA, Shiromani; LEWIS, Michael D.; TAYLOR, Martin C.; KELLY, John M. Biological factors that impinge on Chagas disease drug development. **Parasitology**, v. 144, n. 14, p. 1871–1880, dec. 2017. DOI: <https://doi.org/10.1017/S0031182017001469>.

FRANCO-PAREDES, Carlos; VILLAMIL-GÓMEZ, Wilmer E.; SCHULTZ, Jonathan; HENAO-MARTÍNEZ, Andrés F.; PARRA-HENAO, Gabriel; RASSI, Anis; RODRÍGUEZ-MORALES, Alfonso J.; SUAREZ, José Antonio. A deadly feast: Elucidating the burden of orally acquired acute Chagas disease in Latin America – Public health and travel medicine importance. **Travel Medicine and Infectious Disease**, v. 36, p. 101565, jul. 2020. DOI: <https://doi.org/10.1016/j.tmaid.2020.101565>.

FUJITA, Dennis Minoru; NASCIMENTO, Marilda Savoia; DE ANDRADE JÚNIOR, Heitor Franco. The oral transmission of chagas disease in Brazil: New food supplies and travel experience. **Acta Tropica**, v. 197, p. 105038, sep. 2019. DOI: <https://doi.org/10.1016/j.actatropica.2019.05.029>.

GAMES, Paul A.; HOWELL, John F. Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study. **Journal of Educational Statistics**, v. 1, n. 2, p. 113, 1976. DOI: <https://doi.org/10.2307/1164979>.

GAPSYS, Vytautas; DE GROOT, Bert L. On the importance of statistics in molecular simulations for thermodynamics, kinetics and simulation box size. **eLife**, v. 9, aug. 2020. DOI: <https://doi.org/10.7554/eLife.57589>.

GHANAKOTA, Phani; CARLSON, Heather A. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics. **Journal of Medicinal Chemistry**, v. 59, n. 23, p. 10383–10399, dec. 2016. DOI: <https://doi.org/10.1021/acs.jmedchem.6b00399>.

GIORDANO, Deborah; BIANCANIELLO, Carmen; ARGENIO, Maria Antonia; FACCHIANO, Angelo. Drug Design by Pharmacophore and Virtual Screening Approach. **Pharmaceuticals**, v. 15, n. 5, p. 646, may 2022. DOI: <https://doi.org/10.3390/ph15050646>.

GOMES, Daniele Cavalcante; MEDEIROS, Thayse Silva; ALVES PEREIRA, Eron Lincoln; DA SILVA, João Felipe Oliveira; DE FREITAS OLIVEIRA, Johny W.; FERNANDES-PEDROSA, Matheus de Freitas; DE SOUSA DA SILVA, Marcelo; DA SILVA-JÚNIOR, Arnóbio Antônio. From Benzimidazole to New Drugs: Nanotechnology Contribution in Chagas Disease. **International Journal of Molecular Sciences**, v. 24, n. 18, p. 13778, sep. 2023. DOI: <https://doi.org/10.3390/ijms241813778>.

GONZÁLEZ GARCÍA, Diana; TORRES, Angel; TALEVI, Alan; ALBERCA, Lucas N.; BELTRAN, Miguel A.; LARA, Frida; DA SILVA FERREIRA, Marina; FARANI, Priscila S. G.; ALMEIDA, Igor C.; MALDONADO, Rosa A. Advancing Trypanosoma cruzi N-myristoyltransferase as a drug target for Chagas disease through in silico discovery and biochemical evaluation. **Frontiers in Molecular Biosciences**, v. 12, jan. 2026. DOI: <https://doi.org/10.3389/fmolb.2025.1666768>.

GRAHAM, Sarah E.; SMITH, Richard D.; CARLSON, Heather A. Predicting Displaceable Water Sites Using Mixed-Solvent Molecular Dynamics. **Journal of Chemical Information and Modeling**, v. 58, n. 2, p. 305–314, feb. 2018. DOI: <https://doi.org/10.1021/acs.jcim.7b00268>.

GRANT, Barry J.; LUKMAN, Suryani; HOCKER, Harrison J.; SAYYAH, Jaqueline; BROWN, Joan Heller; MCCAMMON, J. Andrew; GORFE, Alemayehu A. Novel Allosteric Sites on Ras for Lead Generation. **PLoS ONE**, v. 6, n. 10, p. e25711, oct. 2011. DOI: <https://doi.org/10.1371/journal.pone.0025711>.

GRIFFITH, Jason W.; SOKOL, Caroline L.; LUSTER, Andrew D. Chemokines and Chemokine Receptors: Positioning Cells for Host Defense and Immunity. **Annual Review of Immunology**, v. 32, n. 1, p. 659–702, mar. 2014. DOI: <https://doi.org/10.1146/annurev-immunol-032713-120145>.

GUNATILLEKE, Shamila S.; CALVET, Claudia M.; JOHNSTON, Jonathan B.; CHEN, Chiung-Kuang; ERENBURG, Grigori; GUT, Jiri; ENGEL, Juan C.; ANG, Kenny K. H.; MULVANEY, Joseph; CHEN, Steven; ARKIN, Michelle R.; MCKERROW, James H.; PODUST, Larissa M. Diverse Inhibitor Chemotypes Targeting *Trypanosoma cruzi* CYP51. **PLoS Neglected Tropical Diseases**, v. 6, n. 7, p. e1736, jul. 2012. DOI: <https://doi.org/10.1371/journal.pntd.0001736>.

GUO, F; WANG, Y; LIU, J; MOK, S C; XUE, F; ZHANG, W. CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks. **Oncogene**, v. 35, n. 7, p. 816–826, feb. 2016. DOI: <https://doi.org/10.1038/onc.2015.139>.

HARGROVE, Tatiana Y.; WAWRZAK, Zdzislaw; ALEXANDER, Paul W.; CHAPLIN, Jason H.; KEENAN, Martine; CHARMAN, Susan A.; PEREZ, Catherine J.; WATERMAN, Michael R.; CHATELAIN, Eric; LEPESHEVA, Galina I. Complexes of *Trypanosoma cruzi* Sterol 14 $\alpha$ -Demethylase (CYP51) with Two Pyridine-based Drug Candidates for Chagas Disease. **Journal of Biological Chemistry**, v. 288, n. 44, p. 31602–31615, nov. 2013. DOI: <https://doi.org/10.1074/jbc.M113.497990>.

HARIHARAN, P. C.; POPLE, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. **Theoretica Chimica Acta**, v. 28, n. 3, p. 213–222, 1973. DOI: <https://doi.org/10.1007/BF00533485>.

HAUSER, Alexander S.; KOOISTRA, Albert J.; MUNK, Christian; HEYDENREICH, Franziska M.; VEPRINTSEV, Dmitry B.; BOUVIER, Michel; BABU, M. Madan; GLORIAM, David E. GPCR activation mechanisms across classes and macro/microscales. **Nature Structural & Molecular Biology**, v. 28, n. 11, p. 879–888, nov. 2021. DOI: <https://doi.org/10.1038/s41594-021-00674-7>.

HAUSER, Alexander S.; ATTWOOD, Misty M.; RASK-ANDERSEN, Mathias; SCHIÖTH, Helgi B.; GLORIAM, David E. Trends in GPCR drug discovery: new agents, targets and indications. **Nature Reviews Drug Discovery**, v. 16, n. 12, p. 829–842, dec. 2017. DOI: <https://doi.org/10.1038/nrd.2017.178>.

HEDGES, Larry V. Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. **Journal of Educational Statistics**, v. 6, n. 2, p. 107, 1981. DOI: <https://doi.org/10.2307/1164588>.

HEHRE, W. J.; DITCHFIELD, R.; POPLE, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. **The Journal of Chemical Physics**, v. 56, n. 5, p. 2257–2261, mar. 1972. DOI: <https://doi.org/10.1063/1.1677527>.



HERTIG, Samuel; LATORRACA, Naomi R.; DROR, Ron O. Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. **PLOS Computational Biology**, v. 12, n. 6, p. e1004746, jun. 2016. DOI: <https://doi.org/10.1371/journal.pcbi.1004746>.

HESS, Berk. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. **Journal of Chemical Theory and Computation**, v. 4, n. 1, p. 116–122, jan. 2008. DOI: <https://doi.org/10.1021/ct700200b>.

HILGER, Daniel; MASUREEL, Matthieu; KOBILKA, Brian K. Structure and dynamics of GPCR signaling complexes. **Nature Structural & Molecular Biology**, v. 25, n. 1, p. 4–12, jan. 2018. DOI: <https://doi.org/10.1038/s41594-017-0011-7>.

HOLLINGSWORTH, Scott A.; DROR, Ron O. Molecular Dynamics Simulation for All. **Neuron**, v. 99, n. 6, p. 1129–1143, sep. 2018. DOI: <https://doi.org/10.1016/j.neuron.2018.08.011>.

HUANG, Jing; RAUSCHER, Sarah; NAWROCKI, Grzegorz; RAN, Ting; FEIG, Michael; DE GROOT, Bert L; GRUBMÜLLER, Helmut; MACKERELL, Alexander D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. **Nature Methods**, v. 14, n. 1, p. 71–73, jan. 2017. DOI: <https://doi.org/10.1038/nmeth.4067>.

HUANG, Sheng-You; ZOU, Xiaoqin. Advances and Challenges in Protein-Ligand Docking. **International Journal of Molecular Sciences**, v. 11, n. 8, p. 3016–3034, aug. 2010. DOI: <https://doi.org/10.3390/ijms11083016>.

HUMPHREY, William; DALKE, Andrew; SCHULTEN, Klaus. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33–38, feb. 1996. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).

ILIE, Ioana M.; EHRHARDT, Claus; CAFLISCH, Amedeo; WEITZ-SCHMIDT, Gabriele. Decrypting Integrins by Mixed-Solvent Molecular Dynamics Simulations. **Journal of Chemical Information and Modeling**, v. 63, n. 12, p. 3878–3891, jun. 2023. DOI: <https://doi.org/10.1021/acs.jcim.3c00480>.

JACKSON, Yves; WYSSA, Baptiste; CHAPPUIS, François. Tolerance to nifurtimox and benznidazole in adult patients with chronic Chagas' disease. **Journal of Antimicrobial Chemotherapy**, v. 75, n. 3, p. 690–696, mar. 2020. DOI: <https://doi.org/10.1093/jac/dkz473>.

JÄNES, Jürgen; BELTRAO, Pedro. Deep learning for protein structure prediction and design—progress and applications. **Molecular Systems Biology**, v. 20, n. 3, p. 162–169, jan. 2024. DOI: <https://doi.org/10.1038/s44320-024-00016-x>.

JO, Sunhwan; KIM, Taehoon; IYER, Vidyashankara G.; IM, Wonpil. CHARMM-GUI: A web-based graphical user interface for CHARMM. **Journal of Computational Chemistry**, v. 29, n. 11, p. 1859–1865, aug. 2008. DOI: <https://doi.org/10.1002/jcc.20945>.

JONES, Susan; THORNTON, Janet M. Analysis of protein-protein interaction sites using surface patches. **Journal of Molecular Biology**, v. 272, n. 1, p. 121–132, sep. 1997. DOI: <https://doi.org/10.1006/jmbi.1997.1234>.

JORGENSEN, William L.; CHANDRASEKHAR, Jayaraman; MADURA, Jeffry D.; IMPEY, Roger W.; KLEIN, Michael L. Comparison of simple potential functions for simulating liquid water. **The Journal of Chemical Physics**, v. 79, n. 2, p. 926–935, jul. 1983. DOI: <https://doi.org/10.1063/1.445869>.

JORGENSEN, William L.; MAXWELL, David S.; TIRADO-RIVES, Julian. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. **Journal of the American Chemical Society**, v. 118, n. 45, p. 11225–11236, nov. 1996. DOI: <https://doi.org/10.1021/ja9621760>.

JUMPER, John et al. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, aug. 2021. DOI: <https://doi.org/10.1038/s41586-021-03819-2>.

KARPLUS, Martin; MCCAMMON, J. Andrew. Molecular dynamics simulations of biomolecules. **Nature Structural Biology**, v. 9, n. 9, p. 646–652, sep. 2002. DOI: <https://doi.org/10.1038/nsb0902-646>.

KESKIN, Ozlem; TUNCBAG, Nurcan; GURSOY, Attila. Predicting Protein–Protein Interactions from the Molecular to the Proteome Level. **Chemical Reviews**, v. 116, n. 8, p. 4884–4909, apr. 2016. DOI: <https://doi.org/10.1021/acs.chemrev.5b00683>.

KHEDRI, Mohammad; AFSHARCHI, Fatemeh; SOUDERJANI, Amirhosein Hasanpour; REZVANTALAB, Sima; DIDANDEH, Mohsen; MALEKI, Reza; MUSAIE, Kiyan; SANTOS, Hélder A.; SHAHBAZI, Mohammad-Ali. Molecular scale study on the interactions of biocompatible nanoparticles with macrophage membrane and blood proteins. **Nano Select**, v. 3, n. 8, p. 1252–1263, aug. 2022. DOI: <https://doi.org/10.1002/nano.202200043>.

KIMURA, S. Roy; HU, Hai Peng; RUVINSKY, Anatoly M.; SHERMAN, Woody; FAVIA, Angelo D. Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics. **Journal of Chemical Information and Modeling**, v. 57, n. 6, p. 1388–1401, jun. 2017. DOI: <https://doi.org/10.1021/acs.jcim.6b00623>.

KNAPP, Bernhard; OSPINA, Luis; DEANE, Charlotte M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. **Journal of Chemical Theory and Computation**, v. 14, n. 12, p. 6127–6138, dec. 2018. DOI: <https://doi.org/10.1021/acs.jctc.8b00391>.

KÖHLER, C.; CARLSTRÖM, G.; GUNNARSSON, A.; WEININGER, U.; TÅNGEFJORD, S.; ULLAH, V.; LEPISTÖ, M.; KARLSSON, U.; PAPA VOINE, T.; EDMAN, K.; AKKE, M. Dynamic allosteric communication pathway directing differential activation of the glucocorticoid receptor. **Science Advances**, v. 6, n. 29, jul. 2020. DOI: <https://doi.org/10.1126/sciadv.abb5277>.

KOZAKOV, Dima; HALL, David R; XIA, Bing; PORTER, Kathryn A; PADHORN, Dmitry; YUEH, Christine; BEGLOV, Dmitri; VAJDA, Sandor. The ClusPro web server for protein–protein docking. **Nature Protocols**, v. 12, n. 2, p. 255–278, feb. 2017. DOI: <https://doi.org/10.1038/nprot.2016.169>.

KOZAKOV, Dima; GROVE, Laurie E; HALL, David R; BOHNUUD, Tanggis; MOTTARELLA, Scott E; LUO, Lingqi; XIA, Bing; BEGLOV, Dmitri; VAJDA, Sandor. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. **Nature Protocols**, v. 10, n. 5, p. 733–755, may 2015. DOI: <https://doi.org/10.1038/nprot.2015.043>.

KROPF, Simone Petraglia; LIMA, Nísia Trindade. The history of Chagas disease: reflections on science in action. **Memórias do Instituto Oswaldo Cruz**, v. 117, n. 1, 2022. DOI: <https://doi.org/10.1590/0074-02760200372>.

KUDO, Genki; YANAGISAWA, Keisuke; YOSHINO, Ryunosuke; HIROKAWA, Takatsugu. AAp-MSMD: Amino Acid Preference Mapping on Protein–Protein Interaction Surfaces Using Mixed-Solvent Molecular Dynamics. **Journal of Chemical Information and Modeling**, v. 63, n. 24, p. 7768–7777, dec. 2023. DOI: <https://doi.org/10.1021/acs.jcim.3c01677>.

KÜHLBRANDT, Werner. Cryo-EM enters a new era. **eLife**, v. 3, aug. 2014. DOI: <https://doi.org/10.7554/eLife.03678>.

LAL GUPTA, Pancham; CARLSON, Heather A. Cosolvent Simulations with Fragment-Bound Proteins Identify Hot Spots to Direct Lead Growth. **Journal of Chemical Theory and Computation**, v. 18, n. 6, p. 3829–3844, jun. 2022. DOI: <https://doi.org/10.1021/acs.jctc.1c01054>.

LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, apr. 1993. DOI: <https://doi.org/10.1107/S0021889892009944>.

LATORRACA, Naomi R.; VENKATAKRISHNAN, A. J.; DROR, Ron O. GPCR Dynamics: Structures in Motion. **Chemical Reviews**, v. 117, n. 1, p. 139–155, jan. 2017. DOI: <https://doi.org/10.1021/acs.chemrev.6b00177>.

LEE, Bruce Y; BACON, Kristina M; BOTTAZZI, Maria Elena; HOTEZ, Peter J. Global economic burden of Chagas disease: a computational simulation model. **The Lancet Infectious Diseases**, v. 13, n. 4, p. 342–348, apr. 2013. DOI: [https://doi.org/10.1016/S1473-3099\(13\)70002-1](https://doi.org/10.1016/S1473-3099(13)70002-1).

LEE, Chengteh; YANG, Weitao; PARR, Robert G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. **Physical Review B**, v. 37, n. 2, p. 785–789, jan. 1988. DOI: <https://doi.org/10.1103/PhysRevB.37.785>.

LEE, Ji Young; LYMAN, Edward. Predictions for Cholesterol Interaction Sites on the A2A Adenosine Receptor. **Journal of the American Chemical Society**, v. 134, n. 40, p. 16512–16515, oct. 2012. DOI: <https://doi.org/10.1021/ja307532d>.

LEE, Jumin; CHENG, Xi; SWAILS, Jason M.; YEOM, Min Sun; EASTMAN, Peter K.; LEMKUL, Justin A.; WEI, Shuai; BUCKNER, Joshua; JEONG, Jong Cheol; QI, Yifei; JO, Sunhwan; PANDE, Vijay S.; CASE, David A.; BROOKS, Charles L.; MACKERELL, Alexander D.; KLAUDA, Jeffery B.; IM, Wonpil. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. **Journal of Chemical Theory and Computation**, v. 12, n. 1, p. 405–413, jan. 2016. DOI: <https://doi.org/10.1021/acs.jctc.5b00935>.

LEVY, Emmanuel D. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. **Journal of Molecular Biology**, v. 403, n. 4, p. 660–670, nov. 2010. DOI: <https://doi.org/10.1016/j.jmb.2010.09.028>.

LEY, V; ANDREWS, N W; ROBBINS, E S; NUSSENZWEIG, V. Amastigotes of *Trypanosoma cruzi* sustain an infective cycle in mammalian cells. **The Journal of experimental medicine**, v. 168, n. 2, p. 649–659, aug. 1988. DOI: <https://doi.org/10.1084/jem.168.2.649>.

LIMA, Marta L.; TULLOCH, Lindsay B.; CORPAS-LOPEZ, Victoriano; CARVALHO, Sandra; WALL, Richard J.; MILNE, Rachel; RICO, Eva; PATTERSON, Stephen; GILBERT, Ian H.; MONIZ, Sonia; MACLEAN, Lorna; TORRIE, Leah S.; MORGILLO, Carmine; HORN, David; ZUCCOTTO, Fabio; WYLLIE, Susan. Identification of a Proteasome-Targeting Arylsulfonamide with Potential for the Treatment of Chagas' Disease. **Antimicrobial Agents and Chemotherapy**, v. 66, n. 1, jan. 2022. DOI: <https://doi.org/10.1128/AAC.01535-21>.

LISI, George P.; LORIA, J. Patrick. Solution NMR Spectroscopy for the Study of Enzyme Allostery. **Chemical Reviews**, v. 116, n. 11, p. 6323–6369, jun. 2016. DOI: <https://doi.org/10.1021/acs.chemrev.5b00541>.

LOMIZE, Mikhail A.; POGOZHEVA, Irina D.; JOO, Hyeon; MOSBERG, Henry I.; LOMIZE, Andrei L. OPM database and PPM web server: resources for positioning of proteins in membranes. **Nucleic Acids Research**, v. 40, n. D1, p. D370–D376, jan. 2012. DOI: <https://doi.org/10.1093/nar/gkr703>.

LORENTE, Javier Sánchez; SOKOLOV, Aleksandr V.; FERGUSON, Gavin; SCHIÖTH, Helgi B.; HAUSER, Alexander S.; GLORIAM, David E. GPCR drug discovery: new agents, targets and indications. **Nature Reviews Drug Discovery**, v. 24, n. 6, p. 458–479, jun. 2025. DOI: <https://doi.org/10.1038/s41573-025-01139-y>.

LÜTHY, Roland; BOWIE, James U.; EISENBERG, David. Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, n. 6364, p. 83–85, mar. 1992. DOI: <https://doi.org/10.1038/356083a0>.

MACKERELL, Alexander D.; JO, Sunhwan; LAKKARAJU, Sirish Kaushik; LIND, Christoffer; YU, Wenbo. Identification and characterization of fragment binding sites for allosteric ligand design using the site identification by ligand competitive saturation hotspots approach (SILCS-Hotspots). **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 1864, n. 4, p. 129519, apr. 2020. DOI: <https://doi.org/10.1016/j.bbagen.2020.129519>.

MAEDA, Fernando Y.; CORTEZ, Cristian; YOSHIDA, Nobuko. Cell signaling during *Trypanosoma cruzi* invasion. **Frontiers in Immunology**, v. 3, n. NOV, 2012. DOI: <https://doi.org/10.3389/fimmu.2012.00361>.

MANNE-GOEHLER, Jennifer; UMEH, Chukwuemeka A.; MONTGOMERY, Susan P.; WIRTZ, Veronika J. Estimating the Burden of Chagas Disease in the United States. **PLOS Neglected Tropical Diseases**, v. 10, n. 11, p. e0005033, nov. 2016. DOI: <https://doi.org/10.1371/journal.pntd.0005033>.

MARTÍN-ESCOLANO, Javier; MARÍN, Clotilde; ROSALES, María J.; TSAOUSIS, Anastasios D.; MEDINA-CARMONA, Encarnación; MARTÍN-ESCOLANO, Rubén. An Updated View of the *Trypanosoma cruzi* Life Cycle: Intervention Points for an Effective Treatment. **ACS Infectious Diseases**, v. 8, n. 6, p. 1107–1115, jun. 2022. DOI: <https://doi.org/10.1021/acscinfecdis.2c00123>.

MARTÍNEZ, L.; ANDRADE, R.; BIRGIN, E. G.; MARTÍNEZ, J. M. PACKMOL : A package for building initial configurations for molecular dynamics simulations. **Journal of Computational Chemistry**, v. 30, n. 13, p. 2157–2164, oct. 2009. DOI: <https://doi.org/10.1002/jcc.21224>.

MARTINS, Flávia Alves et al. The Recombinant Form of *Trypanosoma cruzi* P21 Controls Infection by Modulating Host Immune Response. **Frontiers in Immunology**, v. 11, jun. 2020. DOI: <https://doi.org/10.3389/fimmu.2020.01010>.

MATTOS, Carla; BELLAMACINA, Cornelia R.; PEISACH, Ezra; PEREIRA, Antonio; VITKUP, Dennis; PETSKO, Gregory A.; RINGE, Dagmar. Multiple Solvent Crystal Structures: Probing Binding Sites, Plasticity and Hydration. **Journal of Molecular Biology**, v. 357, n. 5, p. 1471–1482, apr. 2006. DOI: <https://doi.org/10.1016/j.jmb.2006.01.039>.

MATTOS, Carla; RINGE, Dagmar. Locating and characterizing binding sites on proteins. **Nature Biotechnology**, v. 14, n. 5, p. 595–599, may 1996. DOI: <https://doi.org/10.1038/nbt0596-595>.

MCGIBBON, Robert T.; BEAUCHAMP, Kyle A.; HARRIGAN, Matthew P.; KLEIN, Christoph; SWAILS, Jason M.; HERNÁNDEZ, Carlos X.; SCHWANTES, Christian R.; WANG, Lee-Ping; LANE, Thomas J.; PANDE, Vijay S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. **Biophysical Journal**, v. 109, n. 8, p. 1528–1532, oct. 2015. DOI: <https://doi.org/10.1016/j.bpj.2015.08.015>.

MONDAL, Sayan; JOHNSTON, Jennifer M.; WANG, Hao; KHELASHVILI, George; FILIZOLA, Marta; WEINSTEIN, Harel. Membrane Driven Spatial Organization of GPCRs. **Scientific Reports**, v. 3, n. 1, p. 2909, oct. 2013. DOI: <https://doi.org/10.1038/srep02909>.

MONSALVE-LARA, Jackeline; LILIOSO, Maurício; VALENÇA-BARBOSA, Carolina; THYSSEN, Patricia J; MIGUEL, Danilo C; LIMEIRA, Cleanne; GADELHA, Fernanda R; FONTES, Fernanda V H M; PIRES-SILVA, Dayane; DORNAK, L Lynnette; LIMA, Marli M; DONALISIO, Maria R; ALMEIDA, Carlos E. The risk of oral transmission in an area of a Chagas disease outbreak in the Brazilian northeast evaluated through entomological, socioeconomic and schooling indicators. **Acta Tropica**, v. 215, p. 105803, mar. 2021. DOI: <https://doi.org/10.1016/j.actatropica.2020.105803>.

MORTARA, Renato A.; ANDREOLI, Walter K.; FERNANDES, Maria Cecília D. C.; DA SILVA, Claudio V.; FERNANDES, Adriana B.; L'ABBATE, Carolina; DA SILVA, Solange. Host Cell Actin Remodeling in Response to *Trypanosoma cruzi*: Trypomastigote Versus Amastigote Entry. **Subcellular Biochemistry**. [S.l.: s.n.], 2008. v. 47. p. 101–109. Disponível em: <[http://link.springer.com/10.1007/978-0-387-78267-6\\_8](http://link.springer.com/10.1007/978-0-387-78267-6_8)>.

MORTARA, Renato A.; ANDREOLI, Walter K.; TANIWAKI, Noemi N.; FERNANDES, Adriana B.; SILVA, Claudio V. Da; FERNANDES, Maria Cecília D.C.; L'ABBATE, Carolina; SILVA, Solange Da. Mammalian cell invasion and intracellular trafficking by *Trypanosoma cruzi* infective forms. **Anais da Academia Brasileira de Ciências**, v. 77, n. 1, p. 77–94, mar. 2005. DOI: <https://doi.org/10.1590/S0001-37652005000100006>.

MUEGGE, Ingo. Pharmacophore Features of Potential Drugs. **Chemistry - A European Journal**, v. 8, n. 9, p. 1976, may 2002. DOI: [https://doi.org/10.1002/1521-3765\(20020503\)8:9<1976::AID-CHEM1976>3.0.CO;2-K](https://doi.org/10.1002/1521-3765(20020503)8:9<1976::AID-CHEM1976>3.0.CO;2-K).

NGAN, Chi-Ho; HALL, David R.; ZERBE, Brandon; GROVE, Laurie E.; KOZAKOV, Dima; VAJDA, Sandor. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. **Bioinformatics**, v. 28, n. 2, p. 286–287, jan. 2012. DOI: <https://doi.org/10.1093/bioinformatics/btr651>.

NGAN, Kevin Chun-Ho; HOENIG, Samuel M; KWOK, Hui Si; LUE, Nicholas Z; GOSAVI, Pallavi M; TANNER, David A; GARCIA, Emma M; LEE, Ceejay; LIAU, Brian B. Activity-based CRISPR scanning

uncovers allostery in DNA methylation maintenance machinery. **eLife**, v. 12, feb. 2023. DOI: <https://doi.org/10.7554/eLife.80640>.

NIGAM, AkshatKumar; POLLICE, Robert; HURLEY, Matthew F. D.; HICKMAN, Riley J.; ALDEGHI, Matteo; YOSHIKAWA, Naruki; CHITHRANANDA, Seyone; VOELZ, Vincent A.; ASPURU-GUZZIK, Alán. Assigning confidence to molecular property prediction. **Expert Opinion on Drug Discovery**, v. 16, n. 9, p. 1009–1023, sep. 2021. DOI: <https://doi.org/10.1080/17460441.2021.1925247>.

NOSÉ, Shuichi; KLEIN, M.L. Constant pressure molecular dynamics for molecular systems. **Molecular Physics**, v. 50, n. 5, p. 1055–1076, dec. 1983. DOI: <https://doi.org/10.1080/00268978300102851>.

OLIVEIRA SOTÉ, William; COMAR JUNIOR, Moacyr. Exploring Binding Sites in Chagas Disease Protein TcP21 Using Integrated Mixed Solvent Molecular Dynamics Approaches. **Journal of Chemical Information and Modeling**, v. 65, n. 1, p. 363–377, jan. 2025. DOI: <https://doi.org/10.1021/acs.jcim.4c01927>.

OLIVERA, Mario J.; BUITRAGO, Giancarlo. Economic costs of Chagas disease in Colombia in 2017: A social perspective. **International Journal of Infectious Diseases**, v. 91, p. 196–201, feb. 2020. DOI: <https://doi.org/10.1016/j.ijid.2019.11.022>.

PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied Physics**, v. 52, n. 12, p. 7182–7190, dec. 1981. DOI: <https://doi.org/10.1063/1.328693>.

PARVATHY, Jayadevan; YAZHINI, Arangasamy; SRINIVASAN, Narayanaswamy; SOWDHAMINI, Ramanathan. Interfacial residues in protein–protein complexes are in the eyes of the beholder. **Proteins: Structure, Function, and Bioinformatics**, v. 92, n. 4, p. 509–528, apr. 2024. DOI: <https://doi.org/10.1002/prot.26628>.

PAWIG, Lukas; KLASSEN, Christina; WEBER, Christian; BERNHAGEN, Jürgen; NOELS, Heidi. Diversity and Inter-Connections in the CXCR4 Chemokine Receptor/Ligand Family: Molecular Perspectives. **Frontiers in Immunology**, v. 6, aug. 2015. DOI: <https://doi.org/10.3389/fimmu.2015.00429>.

PEREIRA, Karen Signori; SCHMIDT, Flávio Luis; GUARALDO, Ana M.A.; FRANCO, Regina M.B.; DIAS, Viviane L.; PASSOS, Luiz A.C. Chagas' Disease as a Foodborne Illness. **Journal of Food Protection**, v. 72, n. 2, p. 441–446, feb. 2009. DOI: <https://doi.org/10.4315/0362-028X-72.2.441>.

PEREIRA, Samille Henriques; SÁ MAGALHÃES SERAFIM, Mateus; MORAES, Thaís de Fátima Silva; ZINI, Nathalia; ABRAHÃO, Jônatas Santos; NOGUEIRA, Maurício Lacerda; COELHO DOS REIS, Jordana Graziela Alves; BAGNO, Flávia Fonseca; DA FONSECA, Flávio Guimarães. Design, development, and validation of multi-epitope proteins for serological diagnosis of Zika virus infections

and discrimination from dengue virus seropositivity. **PLOS Neglected Tropical Diseases**, v. 18, n. 4, p. e0012100, apr. 2024. DOI: <https://doi.org/10.1371/journal.pntd.0012100>.

POLAK, Annemarie; RICHLE, R. Mode of action of the 2-nitroimidazole derivative benznidazole. **Annals of Tropical Medicine & Parasitology**, v. 72, n. 1, p. 45–54, feb. 1978. DOI: <https://doi.org/10.1080/00034983.1978.11719278>.

PROUDFOOT, Amanda E. I. Chemokine receptors: multifaceted therapeutic targets. **Nature Reviews Immunology**, v. 2, n. 2, p. 106–115, feb. 2002. DOI: <https://doi.org/10.1038/nri722>.

RAAFLAUB, J; ZIEGLER, W H. Single-dose pharmacokinetics of the trypanosomicide benznidazole in man. **Arzneimittel-Forschung**, v. 29, n. 10, p. 1611–4, 1979.

RAFAEL, Aline Ferreira; FERREIRA, Raquel Aparecida; MOTA, Ariela Ferreira; DAMASCENO, Renata Fiúza; MENEZES, Agna Soares da Silva; LOPES, Bartolomeu Teixeira; PAULO, Gustavo Libério De; SABINO, Ester Cerdeira; RIBEIRO, Antônio Luiz Pinho; QUINTINO, Nayara Domela; VIEIRA, Thallyta Maria. Overview of Chagas disease surveillance in an endemic region in Southeastern Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 65, 2023. DOI: <https://doi.org/10.1590/s1678-9946202365051>.

RAMAN, E. Prabhu; YU, Wenbo; LAKKARAJU, Sirish K.; MACKERELL, Alexander D. Inclusion of Multiple Fragment Types in the Site Identification by Ligand Competitive Saturation (SILCS) Approach. **Journal of Chemical Information and Modeling**, v. 53, n. 12, p. 3384–3398, dec. 2013. DOI: <https://doi.org/10.1021/ci4005628>.

RAMOS-JUNIOR, Alberto Novaes; SOUSA, Andréa Silvestre De. The continuous challenge of Chagas disease treatment: bridging evidence-based guidelines, access to healthcare, and human rights. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 50, n. 6, p. 745–747, dec. 2017. DOI: <https://doi.org/10.1590/0037-8682-0495-2017>.

RAMOS-RINCON, Jose-Manuel; TORRÚS-TENDERO, Diego; GARCÍA-MORANTE, Hilarion; GIMENO-GASCÓN, Adelina; MARCO, Francisco; GIL-ANGUITA, Concepción; WIKMAN-JORGENSEN, Philip; LUCAS-DATO, Ana; RODRIGUEZ-DIAZ, Juan-Carlos; AMADOR, Concepción; LLENAS-GARCÍA, Jara. Cytokine profile levels and their relationship with parasitemia and cardiomyopathy in people with Chagas disease in Spain. A prospective observational study. **Parasitology Research**, v. 123, n. 1, p. 66, jan. 2024. DOI: <https://doi.org/10.1007/s00436-023-08042-8>.

REQUENA-MÉNDEZ, Ana; ALDASORO, Edelweiss; DE LAZZARI, Elisa; SICURI, Elisa; BROWN, Michael; MOORE, David A. J.; GASCON, Joaquim; MUÑOZ, Jose. Prevalence of Chagas Disease in Latin-American Migrants Living in Europe: A Systematic Review and Meta-analysis. **PLOS Neglected Tropical Diseases**, v. 9, n. 2, p. e0003540, feb. 2015. DOI: <https://doi.org/10.1371/journal.pntd.0003540>.



RODRIGUES, Adele A.; CLEMENTE, Tatiana M.; DOS SANTOS, Marlus A.; MACHADO, Fabrício C.; GOMES, Rafael G. B.; MOREIRA, Heline Hellen T.; CRUZ, Mário C.; BRÍGIDO, Paula C.; DOS SANTOS, Paulo C. F.; MARTINS, Flávia A.; BAHIA, Diana; MARICATO, Juliana T.; JANINI, Luiz M. R.; REBOREDO, Eduardo H.; MORTARA, Renato A.; DA SILVA, Claudio V. A Recombinant Protein Based on *Trypanosoma cruzi* P21 Enhances Phagocytosis. **PLoS ONE**, v. 7, n. 12, p. e51384, dec. 2012. DOI: <https://doi.org/10.1371/journal.pone.0051384>.

RODRÍGUEZ-BEJARANO, Oscar Hernán; AVENDAÑO, Catalina; PATARROYO, Manuel Alfonso. Mechanisms Associated with *Trypanosoma cruzi* Host Target Cell Adhesion, Recognition and Internalization. **Life**, v. 11, n. 6, p. 534, jun. 2021. DOI: <https://doi.org/10.3390/life11060534>.

ROMANO, Patricia Silvia; CUETO, Juan Agustín; CASASSA, Ana Florencia; VANRELL, María Cristina; GOTTLIEB, Roberta A.; COLOMBO, María Isabel. Molecular and cellular mechanisms involved in the *Trypanosoma cruzi*/host cell interplay. **IUBMB Life**, v. 64, n. 5, p. 387–396, may 2012. DOI: <https://doi.org/10.1002/iub.1019>.

ROMANUKA, Julija; FOLKERS, Gert E.; GNIDA, Manuel; KOVAČIČ, Lidija; WIENK, Hans; KAPTEIN, Robert; BOELEN, Rolf. Genetic switching by the Lac repressor is based on two-state Monod–Wyman–Changeux allostery. **Proceedings of the National Academy of Sciences**, v. 120, n. 49, dec. 2023. DOI: <https://doi.org/10.1073/pnas.2311240120>.

RUBIO, M; DONOSO, F. [Chagas' disease in children and its treatment with Bay 2502]. **Boletín chileno de parasitología**, v. 24, n. 1, p. 43–8, 1969.

SABANÉS ZARIQUIEY, Francesc; DE SOUZA, João V.; BRONOWSKA, Agnieszka K. Cosolvent Analysis Toolkit (CAT): a robust hotspot identification platform for cosolvent simulations of proteins to expand the druggable proteome. **Scientific Reports**, v. 9, n. 1, p. 19118, dec. 2019. DOI: <https://doi.org/10.1038/s41598-019-55394-2>.

SALES JUNIOR, Policarpo Ademar; MOLINA, Israel; FONSECA MURTA, Silvana Maria; SÁNCHEZ-MONTALVÁ, Adrián; SALVADOR, Fernando; CORRÊA-OLIVEIRA, Rodrigo; CARNEIRO, Cláudia Martins. Experimental and Clinical Treatment of Chagas Disease: A Review. **The American Journal of Tropical Medicine and Hygiene**, v. 97, n. 5, p. 1289–1303, nov. 2017. DOI: <https://doi.org/10.4269/ajtmh.16-0761>.

SAMPAIO, George Harisson Felinto; SILVA, Andressa Noronha Barbosa Da; BRITO, Carlos Ramon do Nascimento; HONORATO, Nathan Ravi Medeiros; OLIVEIRA, Luara Musse De; CÂMARA, Antônia Claudia Jácome Da; GALVÃO, Lúcia Maria da Cunha. Epidemiological profile of acute Chagas disease in individuals infected by oral transmission in northern Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 53, 2020. DOI: <https://doi.org/10.1590/0037-8682-0088-2020>.

SÁNCHEZ-VALDÉZ, Fernando J; PADILLA, Angel; WANG, Wei; ORR, Dylan; TARLETON, Rick L. Spontaneous dormancy protects *Trypanosoma cruzi* during extended drug exposure. **eLife**, v. 7, mar. 2018. DOI: <https://doi.org/10.7554/eLife.34039>.

SANTOS, Viviane Corrêa; LEITE, Paulo Gaio; SANTOS, Lucianna Helene; PASCUTTI, Pedro Geraldo; KOLB, Peter; MACHADO, Fabiana Simão; FERREIRA, Rafaela Salgado. Structure-based discovery of novel cruzain inhibitors with distinct trypanocidal activity profiles. **European Journal of Medicinal Chemistry**, v. 257, p. 115498, sep. 2023. DOI: <https://doi.org/10.1016/j.ejmech.2023.115498>.

SARMA, Parishmita; CARINO, Carlo Marion C.; SEETHARAMA, Deeksha; PANDEY, Shubhi; DWIVEDI-AGNIHOTRI, Hemlata; RUI, Xue; CAO, Yubo; KAWAKAMI, Kouki; KUMARI, Poonam; CHEN, Yu-Chih; LUKER, Kathryn E.; YADAV, Prem N.; LUKER, Gary D.; LAPORTE, Stéphane A.; CHEN, Xin; INOUE, Asuka; SHUKLA, Arun K. Molecular insights into intrinsic transducer-coupling bias in the CXCR4-CXCR7 system. **Nature Communications**, v. 14, n. 1, p. 4808, aug. 2023. DOI: <https://doi.org/10.1038/s41467-023-40482-9>.

SAWILOWSKY, Shlomo S. New Effect Size Rules of Thumb. **Journal of Modern Applied Statistical Methods**, v. 8, n. 2, p. 597–599, nov. 2009. DOI: <https://doi.org/10.22237/jmasm/1257035100>.

SCHMUNIS, Gabriel A.; YADON, Zaida E. Chagas disease: A Latin American health problem becoming a world health problem. **Acta Tropica**, v. 115, n. 1–2, p. 14–21, jul. 2010. DOI: <https://doi.org/10.1016/j.actatropica.2009.11.003>.

SHEIK AMAMUDDY, Olivier; VELDMAN, Wayde; MANYUMWA, Colleen; KHAIRALLAH, Afrah; AGAJANIAN, Steve; OLUYEMI, Odeyemi; VERKHIVKER, Gennady M.; TASTAN BISHOP, Özlem. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. **International Journal of Molecular Sciences**, v. 21, n. 3, p. 847, jan. 2020. DOI: <https://doi.org/10.3390/ijms21030847>.

SHRAKE, A.; RUPLEY, J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. **Journal of Molecular Biology**, v. 79, n. 2, p. 351–371, sep. 1973. DOI: [https://doi.org/10.1016/0022-2836\(73\)90011-9](https://doi.org/10.1016/0022-2836(73)90011-9).

SILVEIRA, Anna Clara Azevedo; UOMBE, Nelsa Paula Inácio; VELIKKAKAM, Teresiamia; BORGES, Bruna Cristina; TEIXEIRA, Thaise Lara; DE ALMEIDA, Vitelhe Ferreira; TORRES, Jhoan David Aguilon; PEREIRA, Cecília Luiza; DE SOUZA, Guilherme; TEIXEIRA, Samuel Cota; SERVATO, João Paulo Silva; SILVA, Marcelo José Barbosa; MINEO, Tiago Wilson Patriarca; RIBAS, Rosineide Marques; MORTARA, Renato Arruda; DA SILVEIRA, José Franco; DA SILVA, Claudio Vieira. The *Trypanosoma cruzi* pleiotropic protein P21 orchestrates the intracellular retention and in-vivo parasitism control of virulent Y strain parasites. **Frontiers in Cellular and Infection Microbiology**, v. 14, jun. 2024. DOI: <https://doi.org/10.3389/fcimb.2024.1412345>.

SMITH, Richard D.; CARLSON, Heather A. Identification of Cryptic Binding Sites Using MixMD with Standard and Accelerated Molecular Dynamics. **Journal of Chemical Information and Modeling**, v. 61, n. 3, p. 1287–1299, mar. 2021. DOI: <https://doi.org/10.1021/acs.jcim.0c01002>.

SOUZA, Rita de Cássia Moreira De; GORLA, David E; CHAME, Marcia; JARAMILLO, Nicolas; MONROY, Carlota; DIOTAIUTI, Lileia. Chagas disease in the context of the 2030 agenda: global warming and vectors. **Memórias do Instituto Oswaldo Cruz**, v. 117, n. 1, 2022. DOI: <https://doi.org/10.1590/0074-02760200479>.

TEIXEIRA, Samuel Cota; LOPES, Daiana Silva; GIMENES, Sarah Natalie Cirilo; TEIXEIRA, Thaise Lara; DA SILVA, Marcelo Santos; BRÍGIDO, Rebecca Tavares e Silva; DA LUZ, Felipe Andrés Cordero; DA SILVA, Aline Alves; SILVA, Makswell Almeida; FLORENTINO, Pilar Veras; TAVARES, Paula Cristina Brígido; DOS SANTOS, Marlus Alves; ÁVILA, Veridiana de Melo Rodrigues; SILVA, Marcelo José Barbosa; ELIAS, Maria Carolina; MORTARA, Renato Arruda; DA SILVA, Claudio Vieira. Mechanistic Insights into the Anti-angiogenic Activity of Trypanosoma cruzi Protein 21 and its Potential Impact on the Onset of Chagasic Cardiomyopathy. **Scientific Reports**, v. 7, n. 1, p. 44978, mar. 2017. DOI: <https://doi.org/10.1038/srep44978>.

TEIXEIRA, Thaise L.; CASTILHOS, Patrícia; RODRIGUES, Cassiano C.; DA SILVA, Aline A.; BRÍGIDO, Rebecca TS.; TEIXEIRA, Samuel C.; BORGES, Bruna C.; DOS SANTOS, Marlus A.; MARTINS, Flávia A.; SANTOS, Paulo César F.; SERVATO, João Paulo S.; SILVA, M.S.; DA SILVA, M.J.B.; ELIAS, M.C.; DA SILVA, Claudio V. Experimental evidences that P21 protein controls Trypanosoma cruzi replication and modulates the pathogenesis of infection. **Microbial Pathogenesis**, v. 135, p. 103618, oct. 2019. DOI: <https://doi.org/10.1016/j.micpath.2019.103618>.

TEIXEIRA, Thaise Lara et al. Trypanosoma cruzi P21: a potential novel target for chagasic cardiomyopathy therapy. **Scientific Reports**, v. 5, n. 1, p. 16877, nov. 2015. DOI: <https://doi.org/10.1038/srep16877>.

TIAN, Yuexun; DURDEN, Cassandra; HAMER, Gabriel L. A scoping review of triatomine control for Chagas disease prevention: current and developing tools in Latin America and the United States. **Journal of Medical Entomology**, apr. 2024. DOI: <https://doi.org/10.1093/jme/tjae043>.

TROTT, Oleg; OLSON, Arthur J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of Computational Chemistry**, v. 31, n. 2, p. 455–461, jan. 2010. DOI: <https://doi.org/10.1002/jcc.21334>.

TYLER, K.M.; ENGMAN, D.M. The life cycle of Trypanosoma cruzi revisited. **International Journal for Parasitology**, v. 31, n. 5–6, p. 472–481, may 2001. DOI: [https://doi.org/10.1016/S0020-7519\(01\)00153-9](https://doi.org/10.1016/S0020-7519(01)00153-9).

UNG, Peter M. U.; GHANAKOTA, Phani; GRAHAM, Sarah E.; LEXA, Katrina W.; CARLSON, Heather A. Identifying binding hot spots on protein surfaces by mixed-solvent molecular dynamics: HIV-1

protease as a test case. **Biopolymers**, v. 105, n. 1, p. 21–34, jan. 2016. DOI: <https://doi.org/10.1002/bip.22742>.

VARADI, Mihaly et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. **Nucleic Acids Research**, v. 52, n. D1, p. D368–D375, jan. 2024. DOI: <https://doi.org/10.1093/nar/gkad1011>.

VILLALTA, Fernando; RACHAKONDA, Girish. Advances in preclinical approaches to Chagas disease drug discovery. **Expert Opinion on Drug Discovery**, v. 14, n. 11, p. 1161–1174, nov. 2019. DOI: <https://doi.org/10.1080/17460441.2019.1652593>.

VOSKO, S. H.; WILK, L.; NUSAIR, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. **Canadian Journal of Physics**, v. 58, n. 8, p. 1200–1211, aug. 1980. DOI: <https://doi.org/10.1139/p80-159>.

WAKEFIELD, Amanda E.; KOZAKOV, Dima; VAJDA, Sandor. Mapping the binding sites of challenging drug targets. **Current Opinion in Structural Biology**, v. 75, p. 102396, aug. 2022. DOI: <https://doi.org/10.1016/j.sbi.2022.102396>.

WANG, Wenkai; SU, Baoquan; XIANG, Chenxiao; YANG, Jianyi. Opportunities and challenges in protein structure prediction. **Quantitative Biology**, v. 14, n. 2, jun. 2026. DOI: <https://doi.org/10.1002/qub2.70035>.

WEIS, William I.; KOBILKA, Brian K. The Molecular Basis of G Protein–Coupled Receptor Activation. **Annual Review of Biochemistry**, v. 87, n. 1, p. 897–919, jun. 2018. DOI: <https://doi.org/10.1146/annurev-biochem-060614-033910>.

WELCH, B. L. On the Comparison of Several Mean Values: An Alternative Approach. **Biometrika**, v. 38, n. 3/4, p. 330, dec. 1951. DOI: <https://doi.org/10.2307/2332579>.

WHO. **Ending the neglect to attain the sustainable development goals: a road map for neglected tropical diseases 2021–2030**. Geneva, Switzerland: [s.n.], 2020. Disponível em: <<https://iris.who.int/handle/10665/338565>>.

WON, Madalyn M.; KRÜGER, Timothy; ENGSTLER, Markus; BURLEIGH, Barbara A. The Intracellular Amastigote of *Trypanosoma cruzi* Maintains an Actively Beating Flagellum. **mBio**, v. 14, n. 2, apr. 2023. DOI: <https://doi.org/10.1128/mbio.03556-22>.

WÜTHRICH, Kurt. The way to NMR structures of proteins. **Nature Structural Biology**, v. 8, n. 11, p. 923–925, nov. 2001. DOI: <https://doi.org/10.1038/nsb1101-923>.

XUE, Li C.; DOBBS, Drena; BONVIN, Alexandre M.J.J.; HONAVAR, Vasant. Computational prediction of protein interfaces: A review of data driven methods. **FEBS Letters**, v. 589, n. 23, p. 3516–3526, nov. 2015. DOI: <https://doi.org/10.1016/j.febslet.2015.10.003>.

YANG, Wanqing; WANG, Yanwei; WANG, Yang. Deep learning-driven protein structure prediction and design: Key model developments by Nobel laureates and multi-domain applications. **Biophysics Reviews**, v. 7, n. 1, mar. 2026. DOI: <https://doi.org/10.1063/5.0273394>.

ZHANG, Jintong; WANG, Shengjie; GAO, Le; LI, Qian. Deep learning revolutionizes protein research: Advances in structure prediction, functional annotation, and engineered design. **Journal of Biotechnology**, v. 414, p. 91–110, jun. 2026. DOI: <https://doi.org/10.1016/j.jbiotec.2026.03.012>.

ZHANG, Zhiying; ZHANG, Hongwei; ZHENG, Lyuqin; CHEN, Shihua; DU, Shuo; XIAO, Junyu; PATEL, Dinshaw J. CXCR4 mediated recognition of HIV envelope spike and inhibition by CXCL12. **Nature Communications**, v. 16, n. 1, p. 8653, sep. 2025. DOI: <https://doi.org/10.1038/s41467-025-63815-2>.

ZHOU, Qingtong; YANG, Dehua; WU, Meng; GUO, Yu; GUO, Wanjing; ZHONG, Li; CAI, Xiaoqing; DAI, Antao; JANG, Wonjo; SHAKHNOVICH, Eugene I; LIU, Zhi-Jie; STEVENS, Raymond C; LAMBERT, Nevin A; BABU, M Madan; WANG, Ming-Wei; ZHAO, Suwen. Common activation mechanism of class A GPCRs. **eLife**, v. 8, dec. 2019. DOI: <https://doi.org/10.7554/eLife.50279>.

ZINGALES, Bianca; MILES, Michael A; MORAES, Carolina B; LUQUETTI, Alejandro; GUHL, Felipe; SCHIJMAN, Alejandro G; RIBEIRO, Isabela. Drug discovery for Chagas disease should consider *Trypanosoma cruzi* strain diversity. **Memórias do Instituto Oswaldo Cruz**, v. 109, n. 6, p. 828–833, sep. 2014. DOI: <https://doi.org/10.1590/0074-0276140156>.

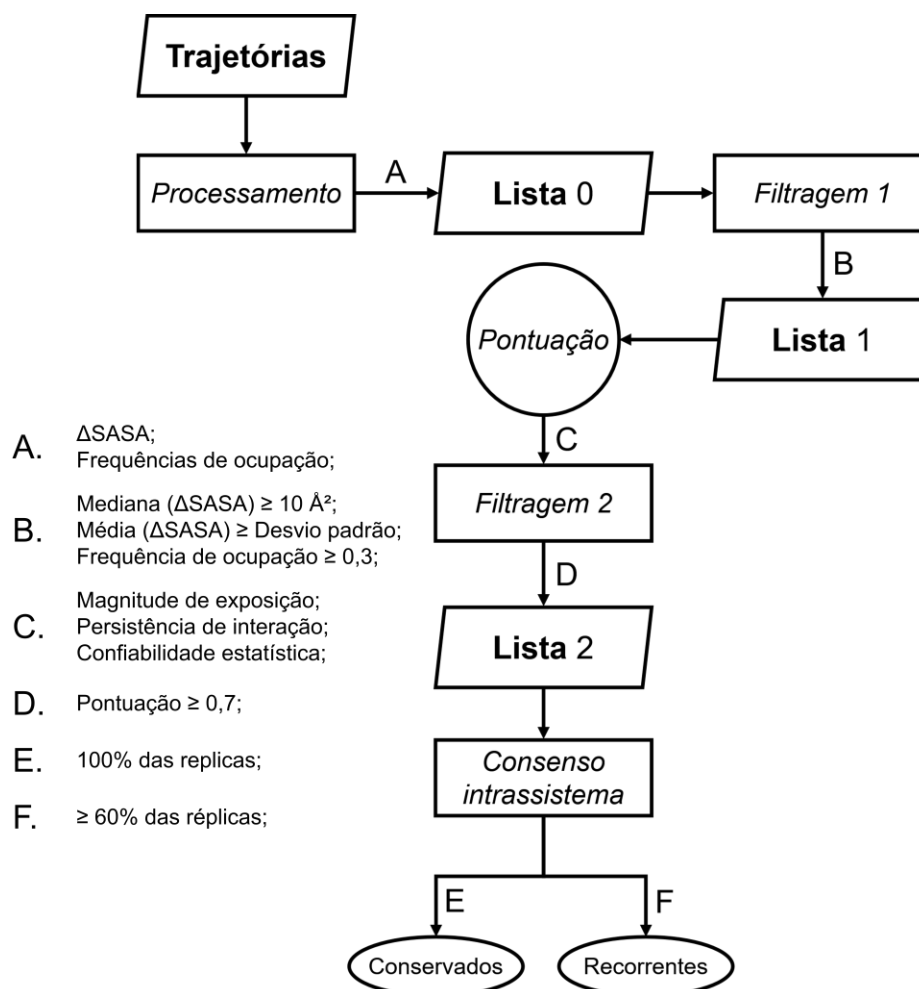
ZUCCOTTO, Fabio. Pharmacophore Features Distributions in Different Classes of Compounds. **Journal of Chemical Information and Computer Sciences**, v. 43, n. 5, p. 1542–1552, sep. 2003. DOI: <https://doi.org/10.1021/ci034068k>.

ZUMA, Aline Araujo; DE SOUZA, Wanderley. Fexinidazole interferes with the growth and structural organization of *Trypanosoma cruzi*. **Scientific Reports**, v. 12, n. 1, p. 20388, nov. 2022. DOI: <https://doi.org/10.1038/s41598-022-23941-z>.

## APÊNDICE A – ROTINA DE IDENTIFICAÇÃO DE RESÍDUOS INTERFACIAIS

O método utilizado nesta tese para a identificação dos resíduos interfaciais mais prováveis do complexo TcP21–CXCR4, fundamentado em uma análise de consenso derivada de cálculos de  $\Delta$ SASA, foi organizado como um fluxograma de trabalho e apresentado de forma geral na Figura A.1. A rotina de análise foi apresentada de forma detalhada nas seções seguintes, incluindo todas as etapas necessárias à sua implementação e validação.

**Figura A.1** – Fluxograma geral da rotina de identificação de resíduos interfaciais.



Fonte: Autor, 2026.

## i. ***Etapa de processamento***

A identificação preliminar dos resíduos interfaciais foi realizada a partir das trajetórias de dinâmica molecular amostradas, considerando-se cada réplica de forma independente.

### a. *Cálculo de $\Delta$ SASA*

Para cada *frame* amostrado, determinou-se a SASA de cada resíduo. A variação de acessibilidade ao solvente foi quantificada por meio de  $\Delta\text{SASA}_i t$ , definida como a diferença entre os estados não ligado e ligado:

$$\Delta\text{SASA}_i t = \text{SASA}_{i,\text{n\~ao ligado}} t - \text{SASA}_{i,\text{ligado}} t \quad 1$$

onde os termos representam, respectivamente, a área exposta ao solvente do resíduo *i* em cada estado, em cada *frame* *t*.

### b. *Cálculo de frequências de ocupação*

A partir dessas séries temporais de  $\Delta\text{SASA}$ , foram calculadas, para cada resíduo, estatísticas descritivas (média, mediana e desvio-padrão) e as frequências de ocupação parciais associadas a três faixas de área de corte: (i)  $10 \leq \Delta\text{SASA}_i < 30 \text{ \AA}^2$ ; (ii)  $30 \leq \Delta\text{SASA}_i < 50 \text{ \AA}^2$ ; e (iii)  $\Delta\text{SASA}_i \geq 50 \text{ \AA}^2$ . Os intervalos utilizados refletem diferentes intensidades de interação interfacial, estimados com base em topologias de interfaces de proteínas (Belapure *et al.*, 2023; Bogan; Thorn, 1998; Jones; Thornton, 1997; Levy, 2010).

A frequência de ocupação total foi expressa como:

$$\phi_i^{\text{total}} = \sum_{n=1}^3 \phi_i^{(n)} \quad 2$$

sendo cada componente calculada por:

$$\begin{aligned}
\phi_i^1 &= \frac{1}{T} \sum_{t=1}^T I(10 \leq \Delta \text{SASA}_i t < 30 \text{ \AA}^2) \\
\phi_i^2 &= \frac{1}{T} \sum_{t=1}^T I(30 \leq \Delta \text{SASA}_i t < 50 \text{ \AA}^2) \\
\phi_i^3 &= \frac{1}{T} \sum_{t=1}^T I(\Delta \text{SASA}_i t \geq 50 \text{ \AA}^2)
\end{aligned} \tag{3}$$

onde  $T$  representa o número total de *frames* e  $I(\cdot)$  é a função indicadora, que assume valor 1 quando a condição é satisfeita e 0 caso contrário.

Por fim, foram selecionados os resíduos que satisfizeram simultaneamente os seguintes critérios: (i) mediana ( $x_{\Delta \text{SASA}}$ ) maior ou igual a 10  $\text{\AA}^2$ ; (ii) média ( $x_{\Delta \text{SASA}}$ ) maior ou igual ao desvio-padrão ( $\sigma_{\Delta \text{SASA}}$ ); e (iii) frequência de ocupação total ( $\phi_i^{\text{total}}$ ) maior ou igual a 0,3. Esses critérios permitiram a seleção de resíduos com exposição consistente ao solvente e relevância potencial para a interface.

## ii. ***Etapas de Pontuação***

Os resíduos remanescentes após a triagem inicial foram pontuados por meio de uma função de pontuação multimétrica, construída a partir de três descritores complementares: (i) magnitude de exposição ao solvente; (ii) persistência temporal de interação; e (iii) confiabilidade estatística das estimativas.

A pontuação final foi definida como a média dos descritores normalizados:

$$W_i = \frac{1}{3} (S_i^{\text{mag}} + S_i^{\text{per}} + S_i^{\text{con}}) \tag{4}$$

em que  $W_i$  denota a pontuação do resíduo  $i$ , enquanto  $S_i^{\text{mag}}$ ,  $S_i^{\text{per}}$  e  $S_i^{\text{con}}$  representam, respectivamente, as contribuições normalizadas de magnitude, persistência e confiabilidade estatística.

A normalização foi aplicada individualmente a cada descritor, em intervalo  $[0,1]$ , permitindo sua consolidação direta na pontuação final. A escolha dos pesos iguais refletiu a ausência de evidência prévia que justificasse a priorização de um descritor em relação aos demais.



a. *Pontuação de magnitude*

A pontuação de magnitude de exposição ao solvente foi definida a partir da mediana ( $x$ ) de  $\Delta\text{SASA}_i$ , a qual foi utilizada como medida robusta de tendência central de exposição ao solvente do resíduo  $i$ . A pontuação não normalizada foi atribuída conforme as faixas previamente estabelecidas:

$$S_i^{mag} = \begin{cases} 3, & x_{\Delta\text{SASA}_i} \geq 50 \text{ \AA}^2 \\ 2, & 30 \leq x_{\Delta\text{SASA}_i} < 50 \text{ \AA}^2 \\ 1, & 10 \leq x_{\Delta\text{SASA}_i} < 30 \text{ \AA}^2 \end{cases} \quad 5$$

A forma normalizada da pontuação de magnitude foi obtida por:

$$S_i^{mag} = \frac{1}{3} S_i^{mag} \quad 6$$

b. *Pontuação de persistência*

A pontuação de persistência temporal de interação foi construída a partir de dois descritores: (i) a frequência de ocupação total e (ii) a distribuição das frequências parciais, associadas a intervalos mutuamente exclusivos de  $\Delta\text{SASA}_i$ .

A componente de frequência de ocupação total foi definida como:

$$S_i^{per,oc} = \begin{cases} 3, & \phi_i^{total} \geq 0,7 \\ 2, & 0,5 \leq \phi_i^{total} < 0,7 \\ 1, & 0,3 \leq \phi_i^{total} < 0,5 \end{cases} \quad 7$$

A componente de distribuição das frequências parciais foi definida com base na condição de alguma das frequências de ocupação parcial ( $\phi_i^{(n)}$ ) ter tido valor igual ou superior a 0,5 (50%), representando que o  $\Delta\text{SASA}_i$  do resíduo analisado permaneceu uma fração predominante de *frames* em alguma das faixas de área de corte definidas anteriormente. Nesses casos, a pontuação foi atribuída conforme o valor da faixa predominante:

$$S_i^{per,pred} = \begin{cases} 3, & \phi_i^{(3)} \geq 0,5 \\ 2, & \phi_i^{(2)} \geq 0,5 \\ 1, & \phi_i^{(1)} \geq 0,5 \end{cases} \quad 8$$

Quando nenhuma frequência de ocupação parcial apresentou predominância, atribuiu-se  $S_i^{per,pred}$  igual a 1,5, caracterizando um comportamento dinâmico com transições frequentes entre diferentes regimes de interação interfacial.

A forma normalizada da pontuação de persistência foi obtida por:

$$S_i^{pi} = \frac{1}{6} (S_i^{per,oc} + S_i^{per,pred}) \quad 9$$

### c. Pontuação de confiabilidade estatística

A pontuação de confiabilidade estatística foi baseada na relação sinal-ruído (SNR, do inglês *signal-to-noise ratio*) dos valores de  $\Delta SASA_i$ , definida como

$$SNR_i = \frac{x_{\Delta SASA_i}}{\sigma_{\Delta SASA_i}} \quad 10$$

onde  $x_{\Delta SASA_i}$  e  $\sigma_{\Delta SASA_i}$  representam, respectivamente, a média e o desvio-padrão da série temporal.

A pontuação não normalizada foi atribuída conforme:

$$S_i^{con} = \begin{cases} 3, & SNR_i \geq 3,0 \\ 2, & 2,0 \leq SNR_i < 3,0 \\ 1, & 1,0 \leq SNR_i < 2,0 \end{cases} \quad 11$$

A forma normalizada da pontuação de confiabilidade foi obtida por:

$$S_i^{con} = \frac{1}{3} S_i^{con} \quad 12$$

### iii. **Análise de consenso intrassistema**

Após a etapa de pontuação, apenas os resíduos com  $W_i \geq 0,7$  foram considerados para a análise de consenso, restringindo-se, assim, a análise a candidatos com maior confiabilidade de pertencerem à interface de interação.

Para cada resíduo, calculou-se sua frequência de ocorrência intrassistema, em intervalo  $[1, 5]$ , definida como

$$f_i^{intra} = \sum_{r=1}^5 I(C_i^r) \quad 13$$

em que  $C_i^r$  corresponde à condição de ocorrência do resíduo  $i$  na réplica  $r$ .

Com base nessa métrica, os resíduos foram classificados como “*Conservados*”, quando presentes em 100% das réplicas de um mesmo sistema, ou “*Recorrentes*”, quando presentes em pelo menos 60% das réplicas. Resíduos que satisfizeram o critério de corte de pontuação, mas não encaixaram nas classificações estabelecidas, foram desconsiderados na etapa de determinação por consenso global.

## REFERÊNCIAS DO APÊNDICE A

---

BELAPURE, Jaydeep; SOROKINA, Marija; KASTRITIS, Panagiotis L. IRAA: A statistical tool for investigating a protein–protein interaction interface from multiple structures. **Protein Science**, v. 32, n. 1, jan. 2023. DOI: <https://doi.org/10.1002/pro.4523>.

BOGAN, Andrew A; THORN, Kurt S. Anatomy of hot spots in protein interfaces. **Journal of Molecular Biology**, v. 280, n. 1, p. 1–9, jul. 1998. DOI: <https://doi.org/10.1006/jmbi.1998.1843>.

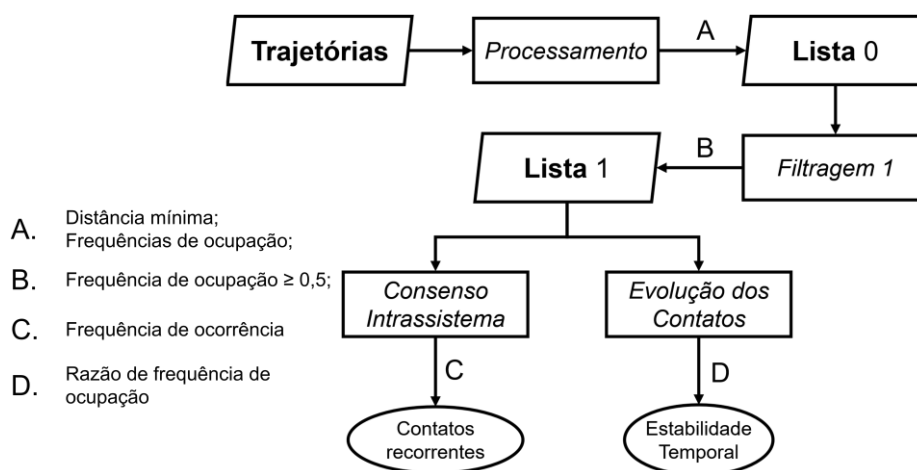
JONES, Susan; THORNTON, Janet M. Analysis of protein-protein interaction sites using surface patches. **Journal of Molecular Biology**, v. 272, n. 1, p. 121–132, sep. 1997. DOI: <https://doi.org/10.1006/jmbi.1997.1234>.

LEVY, Emmanuel D. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. **Journal of Molecular Biology**, v. 403, n. 4, p. 660–670, nov. 2010. DOI: <https://doi.org/10.1016/j.jmb.2010.09.028>.

## APÊNDICE B – ROTINA DE CARACTERIZAÇÃO TEMPORAL DE CONTATOS

O método utilizado nesta tese para a caracterização temporal dos contatos entre TcP21 e CXCR4 foi organizado como um fluxograma de trabalho e apresentado de forma geral na Figura B.1. A rotina de análise foi apresentada de forma detalhada nas seções seguintes, incluindo todas as etapas necessárias à sua implementação e validação.

**Figura B.1** – Fluxograma geral da rotina de caracterização temporal de contatos.



Fonte: Autor, 2026.

### i. ***Etapas de processamento***

A caracterização temporal dos contatos intermoleculares foi realizada a partir das trajetórias de dinâmica molecular amostradas, considerando-se cada réplica de forma independente. Inicialmente, foram avaliadas todas as combinações possíveis de resíduos. Contatos intramoleculares não foram considerados.

Para cada *frame*  $t$  amostrado, calcularam-se as distâncias entre todos os átomos diferentes de hidrogênio pertencentes a cada par de resíduos. A distância mínima observada para cada par  $ij$  foi utilizada como descritor de distância em cada

*frame*, originando uma nova série temporal de distâncias mínimas para cada par  $(d_{ij}^{min}(t))$ .

Com base nessas novas séries, adotou-se uma distância de corte de 4 Å para a definição de contato intermolecular, calculando-se a frequência de ocupação total, expressa como

$$\phi_{ij}^{total} = \frac{1}{T} \sum_{t=1}^T \mathbf{I}(d_{ij}^{min}(t) \geq 4 \text{ Å}) \quad 14$$

Paralelamente, para avaliar a estabilidade temporal desses contatos, cada trajetória foi segmentada em três janelas  $(w_n)$  de mesmo número de *frames*, e as frequências de ocupação parciais  $(\phi_{ij}^{w_n})$  foram recalculadas em cada janela, gerando frações correspondentes ao início, meio e fim da trajetória. Foram considerados para análises subsequentes os pares de resíduos cuja frequência de ocupação total ou qualquer uma das frequências de ocupação parcial tenha apresentado valor maior ou igual a 0,5.

## ii. **Análise de consenso intrassistema**

Dentre os pares restantes, a recorrência dos contatos entre réplicas de um mesmo sistema foi avaliada por meio de uma análise de consenso. Para cada par de resíduos, calculou-se sua frequência de ocorrência, em intervalo  $[1,5]$ , conforme definida em (13).

## iii. **Categorização da evolução temporal**

A estabilidade temporal dos contatos foi analisada a partir das frequências de ocupação parciais calculadas anteriormente. A evolução temporal entre janelas sequenciais foi quantificada pela razão de frequência de ocupação  $(\rho_{ij}^{(n \rightarrow n+1)})$ , definida por:

$$\rho_{ij}^{(n \rightarrow n+1)} = \frac{\phi_{ij}^{w_{n+1}}}{\phi_{ij}^{w_n} + \varepsilon} \quad 15$$

onde  $\varepsilon > 0$  evita divisão por zero.

As razões calculadas foram interpretadas segundo um critério de tolerância de  $\pm 25\%$ , de modo que aumentos significativos ( $\rho \geq 1,25$ ), reduções significativas ( $\rho \leq 0,75$ ) e variações não significativas ( $0,75 < \rho < 1,25$ ) pudessem ser distinguidos. A combinação das duas transições consecutivas ( $w_1 \rightarrow w_2$  e  $w_2 \rightarrow w_3$ ) resultou em nove categorias dinâmicas distintas (Tabela B.1), posteriormente consolidadas em quatro classes funcionais (Tabela B.2).

**Tabela B.1** – Matriz de categorização temporal baseada nas transições entre janelas consecutivas.

|                              |                            | $w_2 \rightarrow w_3$    |                      |                         |
|------------------------------|----------------------------|--------------------------|----------------------|-------------------------|
|                              |                            | <i>Sem variação</i>      | <i>Aumento</i>       | <i>Redução</i>          |
| $w_2$<br>$\uparrow$<br>$w_1$ | <b><i>Sem variação</i></b> | Consistentemente Estável | Estabilização Tardia | Desestabilização Tardia |
|                              | <b><i>Aumento</i></b>      | Estabilizante            | Emergente            | Pico Transiente         |
|                              | <b><i>Redução</i></b>      | Desestabilização Inicial | Vale Transiente      | Desestabilizante        |

Fonte: Autor, 2026.

**Tabela B.2** – Consolidação das categorias temporais em classes funcionais e suas respectivas interpretações físicas.

| <b>Classe Funcional</b> | <b>Categoria Temporal</b>                                                                  | <b>Interpretação Física</b>                    |
|-------------------------|--------------------------------------------------------------------------------------------|------------------------------------------------|
| <i>Persistente</i>      | Consistentemente Estável                                                                   | Persistência ao longo de toda a trajetória     |
| <i>Emergente</i>        | Estabilização Tardia<br>Emergente<br>Estabilizante                                         | Formação ou fortalecimento progressivo         |
| <i>Transiente</i>       | Vale Transiente                                                                            | Enfraquecimento seguido de fortalecimento      |
| <i>Desestabilizante</i> | Desestabilização Tardia<br>Pico Transiente<br>Desestabilização Inicial<br>Desestabilizante | Enfraquecimento ou desaparecimento progressivo |

Fonte: Autor, 2026.

#### iv. ***Caracterizações baseadas em contatos***

##### a. *Distribuição do perfil temporal dos contatos*

A distribuição do perfil temporal dos contatos foi obtida com base nas classes funcionais e frequências de ocorrências de pares, descritas anteriormente. A análise foi realizada em dois regimes distintos: (i) por sistema (intrassistema) e (ii) agrupada (intersistema).

Em cada sistema, foi contabilizado o número de pares de resíduos associados a cada classe funcional, permitindo descrever a composição de perfis temporal dentro de cada conjunto de réplicas. Para a análise agrupada, as contribuições de cada classe foram obtidas pela soma das frequências de ocorrência ao longo de todos os sistemas, de modo que cada contato contribuiu proporcionalmente à sua recorrência. Os valores cumulativos resultantes, em intervalo [1,25], foram normalizados e convertidos em porcentagens, possibilitando a comparação entre classes funcionais em um contexto global.

##### b. *Variabilidade de parceiros de interação por resíduo*

Os resíduos da TcP21 com maior conectividade foram identificados considerando duas perspectivas distintas: (i) diversidade de contatos e (ii) persistência de contatos.

Para cada sistema, a conectividade foi definida como o número de parceiros de interação únicos, permitindo identificar resíduos com maior diversidade de contatos. Nessa perspectiva, para cada resíduo da TcP21, cada par foi contabilizado uma única vez, independentemente de sua frequência de ocorrência. Na análise agrupada, a conectividade foi redefinida em termos de persistência de contatos. Para cada resíduo da TcP21, foram somadas as frequências de ocorrência de todos os pares, ao longo de todos os sistemas. Essa soma resultou em uma medida cumulativa na qual valores mais elevados indicam não somente maior número de possíveis



contatos, mas também maior recorrência de contatos. Independente da perspectiva considerada, os contatos também foram organizados segundo as classes funcionais temporais.

Os resíduos foram ordenados em ordem decrescente de conectividade. Em casos de empate, aplicou-se uma soma ponderada conforme a classificação temporal dos contatos, com pesos 3 (*Persistentes*), 2 (*Emergentes*), 1 (*Transientes*) e 0 (*Desestabilizantes*), seguida de ordenação alfabética quando necessário. Na análise agrupada, por se tratar de *hubs* estabilizantes, contatos classificados como *“Desestabilizantes”* não foram considerados na somatória cumulativa de ocorrência.

#### c. *Recorrência de pares de resíduos*

Os pares de resíduos mais recorrentes foram identificados somente a partir da análise agrupada dos sistemas, considerando a frequência de ocorrência desses contatos ao longo de todas as réplicas. Nesse contexto, cada par foi tratado como uma unidade independente. Para cada par de resíduos, foi calculada a frequência de ocorrência cumulativa, em intervalo [1, 25]. Adicionalmente, as contribuições foram organizadas segundo as classes funcionais temporais previamente definidas. Os pares foram classificados em ordem decrescente de frequência de ocorrência cumulativa. Em situações de empate, foi aplicado um critério de desempate consistente com o adotado anteriormente.

#### d. *Distribuição qualitativa das forças de interação*

As forças intermoleculares associadas aos contatos foram caracterizadas por meio de uma abordagem qualitativa aplicada individualmente a cada par de resíduos.

Inicialmente, os aminoácidos foram categorizados em quatro classes funcionais, de acordo com suas propriedades físico-químicas (Tabela B.3).

**Tabela B.3** – Categorização dos aminoácidos em classes funcionais segundo suas propriedades físico-químicas.

| <b>Classes Funcionais</b>      | <b>Aminoácidos Correspondentes</b>                    |
|--------------------------------|-------------------------------------------------------|
| <i>Hidrofóbico</i>             | ALA, VAL, ILE, LEU, MET, PHE, TYR, TRP, PRO, GLY, CYS |
| <i>Positivamente carregado</i> | ARG, LYS, HIS                                         |
| <i>Negativamente carregado</i> | ASP, GLU                                              |
| <i>Polar</i>                   | SER, THR, ASN, GLN                                    |

Fonte: Autor, 2026.

Com base nas combinações entre essas classes, cada par foi associado a um tipo predominante de interação intermolecular (Tabela B.4). A caracterização foi considerada qualitativa devido à inferência da predominância de um tipo específico de interação intermolecular por meio apenas da combinação das propriedades físico-químicas dos resíduos nos pares observados.

**Tabela B.4** – Correspondência entre combinações de classes funcionais e respectivas forças de interação predominantes.

| <b>Forças de Interação</b>        | <b>Combinações Correspondentes</b>             |
|-----------------------------------|------------------------------------------------|
| <i>Interação Hidrofóbica</i>      | Hidrofóbico + Hidrofóbico                      |
| <i>Atração Eletrostática</i>      | Cargas opostas                                 |
| <i>Interação Polar</i>            | Polar + Polar<br>Polar + Carregado             |
| <i>Interação de van der Waals</i> | Hidrofóbico + Polar<br>Hidrofóbico + Carregado |

Fonte: Autor, 2026.

Para quantificar a contribuição relativa de cada tipo de força intermolecular, somaram-se, em cada sistema, as frequências de ocorrência dos pares correspondentes a cada categoria. Na análise agrupada, as contribuições foram acumuladas ao longo de todos os sistemas e posteriormente expressas em porcentagem.

## APÊNDICE C – ROTINA DE CLUSTERIZAÇÃO E GRÁFICOS DE *ELBOW*

---

Neste trabalho, utilizou-se o método de clusterização hierárquica aglomerativa, construído a partir da matriz de distâncias RMSD calculadas par-a-par, empregando o critério de *average linkage*. Essa abordagem possibilitou a identificação de estados conformacionais estruturalmente consistentes sem a necessidade de definir, *a priori*, o número de clusters.

### a. *Análise de similaridade*

A similaridade conformacional entre duas configurações  $i$  e  $j$ , ambas contendo  $N$  átomos, foi estimada por meio do cálculo de RMSD, definido como:

$$\text{RMSD } i, j = \sqrt{\frac{1}{N} \sum_{k=1}^N \| r_k^i - r_k^j \|^2} \quad 16$$

onde  $r_k^i$  e  $r_k^j$  representam as coordenadas cartesianas do átomo  $k$ , respectivamente, após a superposição estrutural que minimizou o RMSD.

Com base nesses valores, foi construída uma matriz de distância simétrica contendo os valores de RMSD calculados para todos os pares de configurações.

### b. *Clusterização hierárquica aglomerativa*

A clusterização hierárquica aglomerativa foi conduzida por meio de um processo iterativo de fusão sucessiva de clusters, aplicado ao conjunto de configurações amostradas ( $i \in \{1, \dots, n\}$ ).

Inicialmente, cada configuração constituiu um cluster unitário, resultando em  $n$  clusters. A distância intercluster ( $d(A, B)$ ) entre cada par de clusters  $A$  e  $B$  foi definida segundo o critério de *average linkage*, dado por:

$$d_{A,B} = \frac{1}{n_A n_B} \sum_{i \in A} \sum_{j \in B} \text{RMSD}_{i,j} \quad 17$$

onde  $n_A$  e  $n_B$  correspondem ao número de configurações pertencentes aos respectivos clusters analisados.

A cada iteração, identificou-se o par de clusters que minimizou  $d_{A,B}$ , o qual foi então agrupado em um novo cluster. Esse processo foi reiterado até o completo agrupamento das configurações amostradas em um único cluster.

A sucessão de agrupamentos definiu uma estrutura hierárquica, representada por um dendrograma, no qual cada nível hierárquico está associado a um valor específico de distância intercluster, refletindo a similaridade relativa entre os agrupamentos formados.

### c. Determinação dos limites inferior e superior de distância intercluster

A definição da distância de corte otimizada foi conduzida de modo a evitar regimes extremos de clusterização, caracterizados por superfragmentação ou por agregação excessiva das configurações. Valores de corte reduzidos resultam em um número excessivo de clusters de baixa população, frequentemente associados a flutuações térmicas e, portanto, de relevância conformacional limitada. Para mitigar esse efeito e assegurar a confiabilidade estatística dos agrupamentos (Jain *et al.*, 1999), o limite inferior foi definido como:

$$N_C^{(t)} \leq \sqrt{n} \quad 18$$

onde  $N_C^{(t)}$  é o número de clusters pertencentes à distância de corte  $t$ , e  $n$  é o número total de configurações amostradas.

Em contrapartida, valores de corte elevados promovem o agrupamento excessivo de configurações estruturalmente distintas, resultando em um número reduzido de clusters altamente populados, comprometendo a resolução da heterogeneidade estrutural. Para restringir esse comportamento, o limite superior foi definido pelas condições:

$$\begin{aligned} P_{top5}^t &\geq 90\% \\ N_C^t &\geq 5 \end{aligned} \quad 19$$

em que  $P_{top5}^t$  corresponde à população cumulativa dos cinco clusters mais populosos.

A imposição conjunta desses critérios assegurou a representação adequada dos estados conformacionais mais prováveis, ao mesmo tempo em que preservou um nível mínimo de diversidade estrutural, evitando a degeneração das configurações amostradas em poucos clusters dominantes.

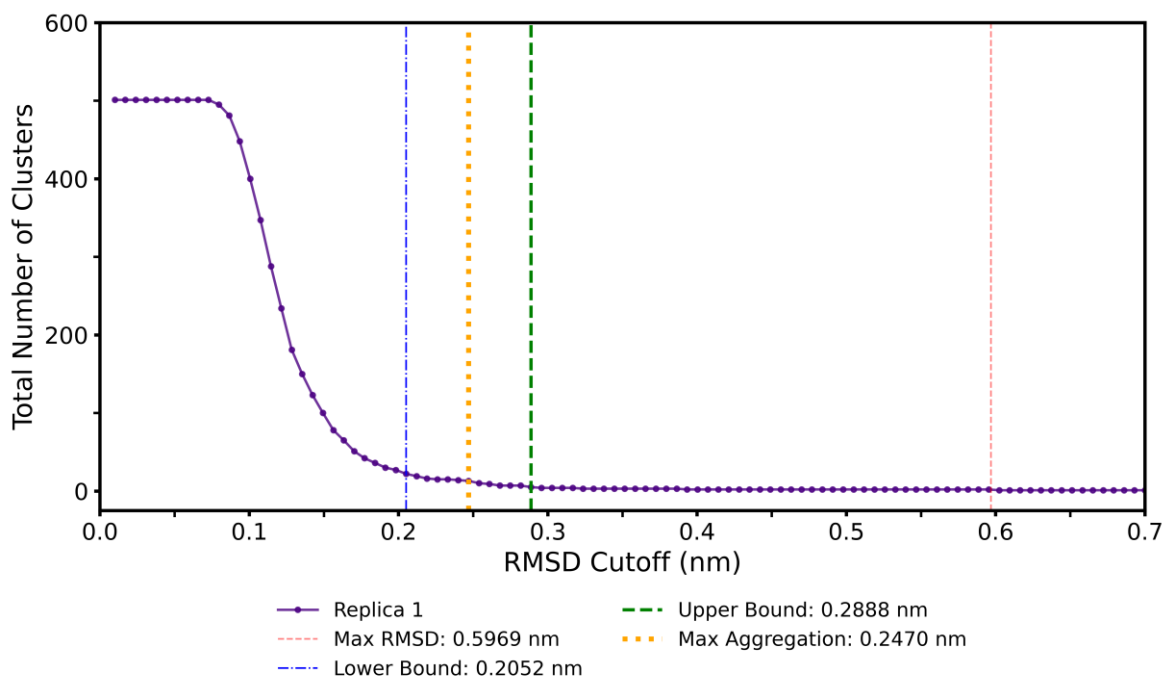
d. *Determinação da distância de corte otimizada*

Uma vez estabelecidos os limites inferior e superior, a distância de corte otimizada foi determinada com base em um critério análogo ao método *elbow*. Nesse contexto, identificou-se a distância intercluster correspondente ao ponto de inflexão na dinâmica de agrupamento dos clusters, caracterizado por uma variação máxima na distribuição de suas populações. Tal procedimento é conceitualmente equivalente ao método *elbow* clássico (Thorndike, 1953), sendo, contudo, adaptado para incorporar explicitamente a evolução das populações dos clusters ao longo do processo de clusterização hierárquica. A distância de corte assim obtida foi empregada na identificação das configurações medóides.

e. Gráficos de *elbow* para as réplicas  $R_1(M_1 - M_4)$

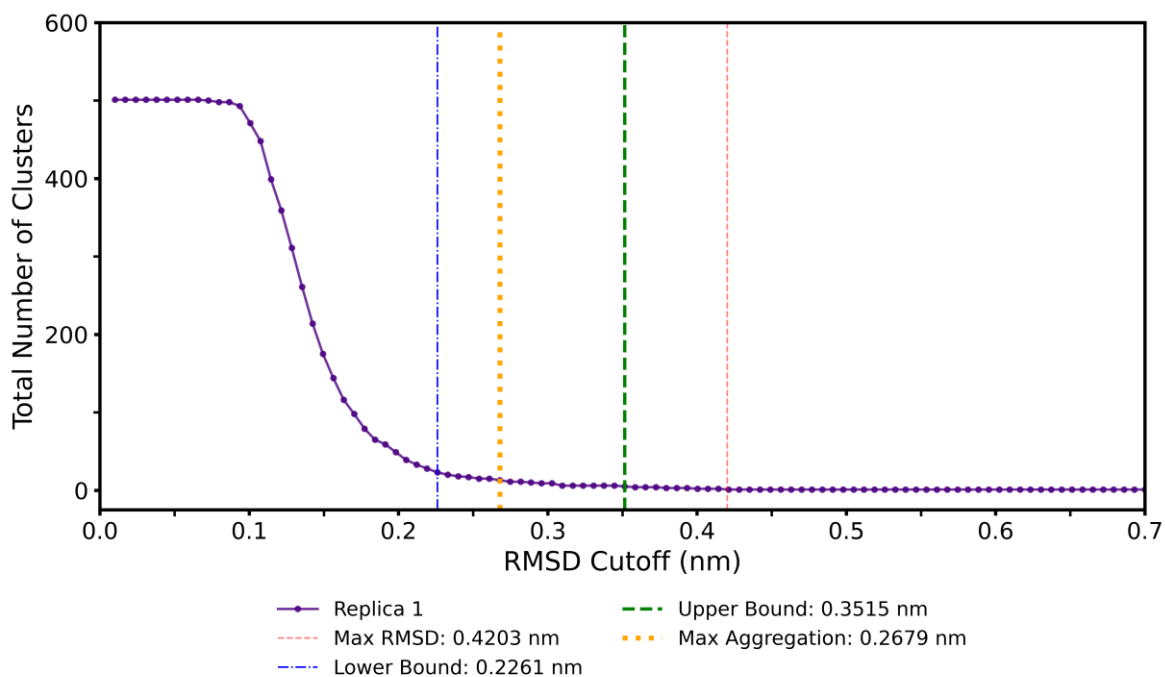
Os gráficos de *elbow* resultantes da clusterização das réplicas  $R_1(M_1 - M_4)$ , empregados na determinação das distâncias de corte intercluster, são mostrados nas Figuras C.1 a C.4. As configurações medóides obtidas por meio da análise de clusterização são mostradas na Figura 3.12.

**Figura C.1** – Gráfico de *elbow* para a determinação da distância intercluster de corte da réplica  $R_1(M_1)$ .



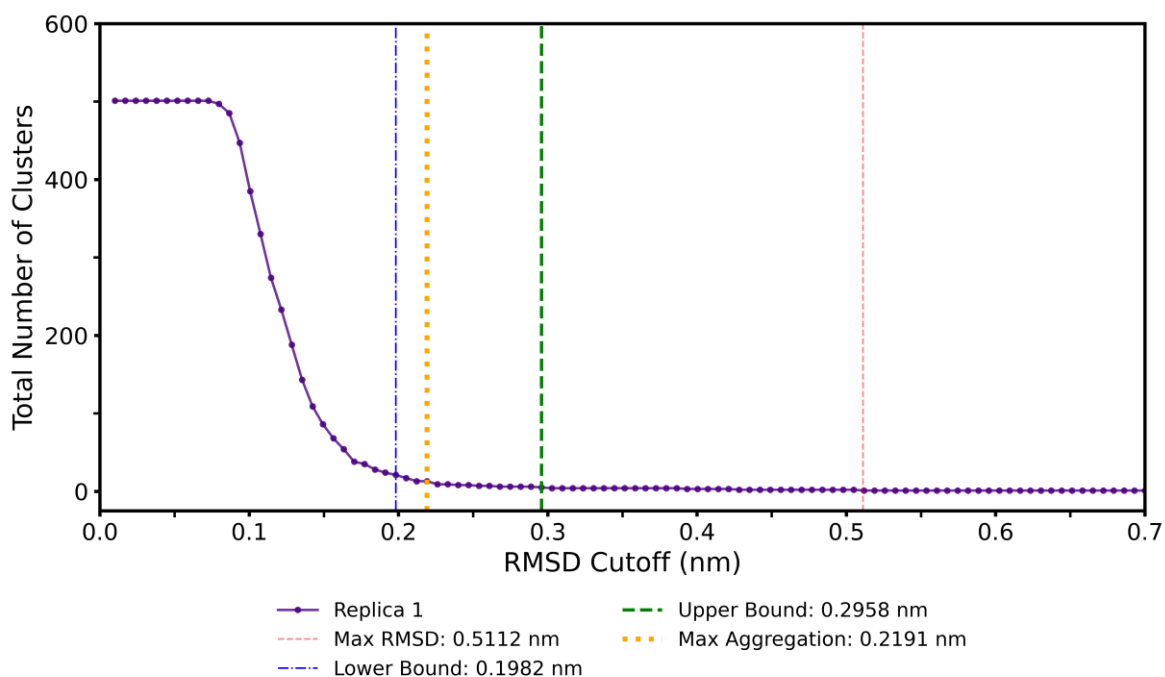
Fonte: Autor, 2026.

**Figura C.2** – Gráfico de *elbow* para a determinação da distância intercluster de corte da réplica  $R_1(M_2)$ .



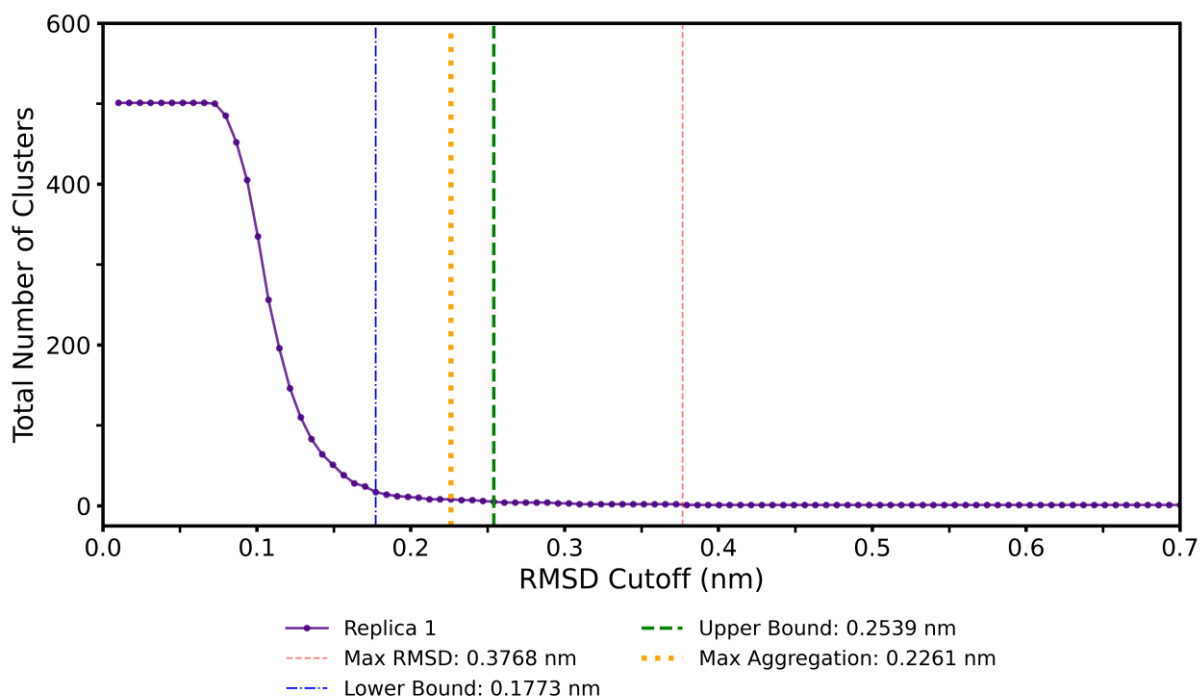
Fonte: Autor, 2026.

**Figura C.3** – Gráfico de *elbow* para a determinação da distância intercluster de corte da replica  $R_1(M_3)$ .



Fonte: Autor, 2026.

**Figura C.4** – Gráfico de *elbow* para a determinação da distância intercluster de corte da replica  $R_1(M_4)$ .



Fonte: Autor, 2026.

## REFERÊNCIAS DO APÊNDICE C

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering. **ACM Computing Surveys**, v. 31, n. 3, p. 264–323, sep. 1999. DOI: <https://doi.org/10.1145/331499.331504>.

THORNDIKE, Robert L. Who Belongs in the Family? **Psychometrika**, v. 18, n. 4, p. 267–276, dec. 1953. DOI: <https://doi.org/10.1007/BF02289263>.



## APÊNDICE D – DADOS DE ANÁLISE *POST HOC*

Os resultados das análises *post hoc* são apresentados nas Tabelas D.1 a D.3, incluindo as comparações de pares realizadas pelo teste de Games-Howell e as estimativas de tamanho de efeito obtidas teste de  $g$  de Hedges, para as métricas RMSD da TcP21, número de ligações de hidrogênio e número de pontes salinas entre TcP21 e CXCR4.

**Tabela D.1** – Análise *post hoc* para RMSD da TcP21.

| Par 1          | Par 2          | $\Delta_{\text{média}}$ | p-valor (ajust.) | Rejeita? | $g$ de Hedges | Tamanho de efeito     |
|----------------|----------------|-------------------------|------------------|----------|---------------|-----------------------|
| M <sub>1</sub> | M <sub>2</sub> | 0,002                   | 1,000            | Não      | 0,01          | <i>Insignificante</i> |
| M <sub>1</sub> | M <sub>3</sub> | 0,160                   | 0,080            | Não      | 2,00          | <i>Muito Forte</i>    |
| M <sub>1</sub> | M <sub>4</sub> | 0,113                   | 0,322            | Não      | 1,18          | <i>Forte</i>          |
| M <sub>1</sub> | M <sub>5</sub> | 0,075                   | 0,695            | Não      | 0,75          | <i>Moderado</i>       |
| M <sub>2</sub> | M <sub>3</sub> | 0,158                   | 0,330            | Não      | 1,24          | <i>Muito Forte</i>    |
| M <sub>2</sub> | M <sub>4</sub> | 0,111                   | 0,642            | Não      | 0,81          | <i>Forte</i>          |
| M <sub>2</sub> | M <sub>5</sub> | 0,073                   | 0,883            | Não      | 0,52          | <i>Moderado</i>       |
| M <sub>3</sub> | M <sub>4</sub> | -0,047                  | 0,725            | Não      | -0,72         | <i>Moderado</i>       |
| M <sub>3</sub> | M <sub>5</sub> | -0,085                  | 0,356            | Não      | -1,17         | <i>Forte</i>          |
| M <sub>4</sub> | M <sub>5</sub> | -0,038                  | 0,941            | Não      | -0,42         | <i>Fraco</i>          |

Fonte: Autor, 2026.

**Tabela D.2** – Análise *post hoc* para número de ligações de hidrogênio entre TcP21 e CXCR4.

| Par 1          | Par 2          | $\Delta_{\text{média}}$ | p-valor (ajust.) | Rejeita? | $g$ de Hedges | Tamanho de efeito     |
|----------------|----------------|-------------------------|------------------|----------|---------------|-----------------------|
| M <sub>1</sub> | M <sub>2</sub> | 3,28                    | 0,196            | Não      | 1,42          | <i>Muito Forte</i>    |
| M <sub>1</sub> | M <sub>3</sub> | 3,19                    | 0,435            | Não      | 1,03          | <i>Forte</i>          |
| M <sub>1</sub> | M <sub>4</sub> | 0,31                    | 0,999            | Não      | 0,13          | <i>Insignificante</i> |
| M <sub>1</sub> | M <sub>5</sub> | 9,17                    | 0,009            | Sim      | 2,84          | <i>Enorme</i>         |
| M <sub>2</sub> | M <sub>3</sub> | -0,08                   | 1,000            | Não      | -0,03         | <i>Insignificante</i> |
| M <sub>2</sub> | M <sub>4</sub> | -2,96                   | 0,136            | Não      | -1,56         | <i>Muito Forte</i>    |
| M <sub>2</sub> | M <sub>5</sub> | 5,90                    | 0,064            | Não      | 2,03          | <i>Enorme</i>         |
| M <sub>3</sub> | M <sub>4</sub> | -2,88                   | 0,450            | Não      | -1,02         | <i>Forte</i>          |
| M <sub>3</sub> | M <sub>5</sub> | 5,98                    | 0,102            | Não      | 1,68          | <i>Muito Forte</i>    |
| M <sub>4</sub> | M <sub>5</sub> | 8,86                    | 0,010            | Sim      | 3,00          | <i>Enorme</i>         |

Fonte: Autor, 2026.

**Tabela D.3** – Análise *post hoc* para número de pontes salinas entre TcP21 e CXCR4.

| Par 1          | Par 2          | $\Delta_{\text{média}}$ | p-valor<br>(ajust.) | Rejeita? | <i>g</i> de Hedges | Tamanho de efeito  |
|----------------|----------------|-------------------------|---------------------|----------|--------------------|--------------------|
| M <sub>1</sub> | M <sub>2</sub> | 2,71                    | 0,026               | Sim      | 2,87               | <i>Enorme</i>      |
| M <sub>1</sub> | M <sub>3</sub> | 3,81                    | 0,094               | Não      | 1,84               | <i>Muito Forte</i> |
| M <sub>1</sub> | M <sub>4</sub> | 1,95                    | 0,312               | Não      | 1,20               | <i>Muito Forte</i> |
| M <sub>1</sub> | M <sub>5</sub> | 4,67                    | 0,002               | Sim      | 3,60               | <i>Enorme</i>      |
| M <sub>2</sub> | M <sub>3</sub> | 1,10                    | 0,833               | Não      | 0,59               | <i>Moderado</i>    |
| M <sub>2</sub> | M <sub>4</sub> | -0,76                   | 0,856               | Não      | -0,56              | <i>Moderado</i>    |
| M <sub>2</sub> | M <sub>5</sub> | 1,96                    | 0,082               | Não      | 2,07               | <i>Enorme</i>      |
| M <sub>3</sub> | M <sub>4</sub> | -1,86                   | 0,637               | Não      | -0,81              | <i>Forte</i>       |
| M <sub>3</sub> | M <sub>5</sub> | 0,86                    | 0,943               | Não      | 0,41               | <i>Fraco</i>       |
| M <sub>4</sub> | M <sub>5</sub> | 2,71                    | 0,112               | Não      | 1,67               | <i>Muito Forte</i> |

Fonte: Autor, 2026.