



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE ENGENHARIA DA PESCA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA DE RECURSOS
NATURAIS

ALICE SOARES DE QUEIROZ

DESENVOLVIMENTO DE PROTOCOLO, *IN SILICO*, VISANDO AVALIAÇÃO DA
INTERFACE ENTRE UM FRAGMENTO DE ANTICORPO E O CD22, PARA
APLICAÇÃO EM CAR-T CELL

FORTALEZA
2024

ALICE SOARES DE QUEIROZ

DESENVOLVIMENTO DE PROTOCOLO, *IN SILICO*, VISANDO AVALIAÇÃO DA
INTERFACE ENTRE UM FRAGMENTO DE ANTICORPO E O CD22, PARA
APLICAÇÃO EM CAR-T *CELL*

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biotecnologia de Recursos Naturais da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Biotecnologia de Recursos Naturais. Área de concentração: Biologia estrutural de proteínas.

Orientador: Prof. Dr. Marcos Roberto Lourenzoni

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

Q42d Queiroz, Alice Soares de.
Desenvolvimento de protocolo, in silico, visando avaliação da interface entre um fragmento de anticorpo e o CD22, para aplicação em CAR-T CELL / Alice Soares de Queiroz. – 2024.
192 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de Pós-Graduação em Biotecnologia de Recursos Naturais, Fortaleza, 2024.
Orientação: Prof. Dr. Marcos Roberto Lourenzoni.

1. CARs anti-CD22. 2. scFvs de M971. 3. Dinâmica Molecular. 4. MM/PBSA. I. Título.

CDD 660.6

ALICE SOARES DE QUEIROZ

DESENVOLVIMENTO DE PROTOCOLO, *IN SILICO*, VISANDO AVALIAÇÃO DA
INTERFACE ENTRE UM FRAGMENTO DE ANTICORPO E O CD22, PARA
APLICAÇÃO EM CAR-T CELL

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biotecnologia de Recursos Naturais da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Biotecnologia de Recursos Naturais. Área de concentração: Biologia estrutural de proteínas.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Marcos Roberto Lourenzoni (Orientador)
Fundação Oswaldo Cruz Ceará (Fiocruz-CE)

Prof. Dr. Gilvan Pessoa Furtado
Fundação Oswaldo Cruz Ceará (Fiocruz-CE)

Prof. Dr. Davi Serradella Vieira
Universidade Federal do Rio Grande do Norte (UFRN)

Dedico este trabalho à Deus, à minha mãe,
ao meu pai "*In Memoriam*", aos meus irmãos
e demais amigos.

AGRADECIMENTOS

Agradeço, primeiramente, à **Deus**, pelo dom da vida e da sabedoria, por guiar meu caminho ao longo dessa jornada.

Aos meus pais, **Carmem e José** “*In Memoriam*”, por desde a minha infância investirem na minha educação, mesmo com todas as adversidades que surgiram no meio do caminho. Por sempre acreditarem que o caminho da educação transforma vidas.

Aos meus irmãos, **Adriana, Alexandre e Daniela**, que sempre vibraram com minhas conquistas por menores que fossem. Por investirem na minha educação em conjunto com meus pais.

Aos demais **familiares**, que contribuíram de alguma forma na minha trajetória acadêmica.

À **Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP)**, pelo apoio financeiro com a manutenção da bolsa de Mestrado Acadêmico.

À **Fundação Oswaldo Cruz/CE** e ao **Grupo de Pesquisa em Engenharia de Proteínas & Soluções para Saúde (GPeSS)**, por terem me cedido infraestrutura necessária para o desenvolvimento desse trabalho.

Ao Prof. **Dr. Marcos Roberto Lourenzoni**, por desde minha Iniciação Científica acreditar no meu potencial e me incentivar a superar os desafios. Por me estimular e instigar o meu interesse na área da Biotecnologia e Bioinformática. Pela paciência, ensinamentos e me corrigir nos momentos necessários. Por ser além de um orientador, ser uma pessoa a quem recorrer para meus anseios pessoais.

As professoras participantes da banca examinadora do exame de Qualificação de Mestrado **Dra. Dayane Alves Costa (Fiocruz-CE)** e **Dra. Andréa Queiroz Maranhão (UnB)** pelo tempo, pelas valiosas colaborações e sugestões. Aos professores participantes da banca examinadora de defesa de Dissertação **Dr. Gilvan Pessoa Furtado (Fiocruz-CE)** e **Dr. Davi Serradella Vieira (UFRN)** por aceitar fazerem parte da banca, pelas valiosas colaborações e sugestões.

À **Pós-Graduação em Biotecnologia de Recursos Naturais** pela excelente grade curricular que possibilitou conhecimentos necessários ao engrandecimento intelectual.

Aos **professores do curso de Pós-Graduação em Biotecnologia de Recursos Naturais**, da Universidade Federal do Ceará, pelas excelentes aulas ministradas e dedicação com a pesquisa científica.

À **Universidade Federal do Ceará**, por ter me recebido de braços abertos durante meu período de mestrado na instituição. Agradecimento em especial, a coordenação do programa de Pós-Graduação em Biotecnologia de Recursos Naturais por sanar minhas dúvidas quando necessário.

Aos colegas de trabalho do Grupo de Pesquisa em Engenharia de Proteínas & Soluções para Saúde (GEPeSS) na Fundação Oswaldo Cruz/Ceará, **Alison Rebouças, Ana Júlia Ferreira, Ana Virgínia Frota, Cássio Pinheiro, Mathias Coelho, Natália Fernandes, Karoline Vieira, Lucas Almeida, José Samuel e Daniel Oliveira** por deixarem o ambiente de trabalho mais descontraído e por sempre estarem dispostos a ajudar quando requisitados. À minha amiga **Dra. Agnes Magri**, que apesar do seu curto período na FIOCRUZ/CE, foi uma amizade valiosa que levarei pelo resto da vida.

Aos **funcionários da Fundação Oswaldo Cruz/CE**, pelo suporte, pela simpatia e por terem ajudado em tudo que estivesse ao alcance.

Aos amigos que fiz no programa de Pós-Graduação em Biotecnologia de Recursos Naturais, **Jéssica Duarte, João Pedro Queiroz, José Gabriel e Lucas Almeida**, que me auxiliaram de alguma forma para permanecer firme e forte nessa caminhada.

Aos meus colegas de graduação do Curso de Licenciatura em Química do IFCE Maracanaú, em especial ao **Kalleu Alencar e Anderson Victor**, pelas palavras de apoio moral nos momentos necessários. Em breve também irão se tornar mestres e possivelmente doutores, quem sabe nos tornaremos até colegas de profissão.

Aos demais **amigos**, principalmente os de infância, por estarem presentes mesmo que virtualmente em muitos momentos da minha vida.

“Só fazemos melhor aquilo que repetidamente insistimos em melhorar. A busca da excelência não deve ser um objetivo, e sim um hábito”, Aristóteles.

RESUMO

A terapia por células CAR-T é uma imunoterapia que objetiva alcançar a remissão de diversos tipos de câncer, como a Leucemia Linfóide Aguda. A terapia inclui a etapa de expressão de um Receptor Quimérico de Antígeno (CAR) em Célula T do paciente, o qual possui a função de redimensionar a ação celular, tornando-a capaz de reconhecer antígenos específicos na superfície de células tumorais, como o CD22. O fragmento variável de cadeia única (scFv) é a estrutura do CAR responsável pelo reconhecimento antigênico e é formado pelos domínios VL e VH de um anticorpo, ligados por uma sequência de resíduos denominada *linker*. Foram publicados resultados de ensaios clínicos de dois CARs anti-CD22, denominados de CARs de scFv de *linker* curto e longo, ambos usando o VH e VL de M971, sendo identificado que o scFv curto tem maior afinidade ao CD22 que o scFv longo. O objetivo deste trabalho foi compreender as diferenças estruturais do scFv e CD22, na interface scFv/CD22 com o CD22 conectado à modelo de membrana lipídica, desenvolvendo um protocolo, *in silico* com capacidade preditiva, para ser utilizado na proposição de mutações efetivas para desenvolver novos CARs. Simulações de Dinâmica Molecular foram executadas e a trajetória analisada, para investigar a energia livre de ligação entre o CD22 e os scFvs. O fragmento Fab de M971 e os domínios D6-D7 do CD22 estão depositados no PDB, cód. 7O52 e foram usados para gerar os modelos 3D, submetidos às simulações em solução aquosa e bicamada lipídica. Os complexos scFv/CD22 foram construídos, simulados e avaliados estrutura e energeticamente. Os sistemas scFv/CD22 em bicamada foram simulados em triplicata. Como resultado, foi observado que scFv/CD22 de linker curto e longo mostraram estabilidade estrutural diferentes entre si e em solução aquosa, mas semelhantes quando em ambiente de membrana celular. Os $\Delta G_{\text{ligação}}$ médios entre scFv/CD22 estão próximos do $\Delta G_{\text{ligação}}$ experimental, somente quando em modelo membrana lipídica. Os resultados alcançados geraram um protocolo que pode ser usado para proposição de mutações, a serem testadas *in silico*, para prosseguir com maior velocidade e qualidade aos testes *in vitro* e *in vivo*, propondo novos CARs, mais efetivos para terapia celular *CAR-T Cell*.

Palavras-chave: CARs anti-CD22; scFvs de M971; Dinâmica Molecular; MM/PBSA.

ABSTRACT

CAR-T cell therapy is an immunotherapy aimed at achieving remission of various types of cancer, such as Acute Lymphoblastic Leukemia (ALL). The therapy includes the step of expressing a Chimeric Antigen Receptor (CAR) in the patient's T cell, which has the function of resizing cellular action, making it capable of recognizing specific antigens on the surface of tumor cells, such as CD22. The single-chain variable fragment (scFv) is the CAR structure responsible for antigenic recognition and is formed by the VL and VH domains of an antibody, linked by a sequence of residues called a linker. Results from clinical trials of two anti-CD22 CARs, named short and long linker scFvs, both using the VH and VL of M971, have been published, identifying that the short scFv has a higher affinity to CD22 than the long scFv. This work aimed to understand the structural differences of scFv and CD22 at the scFv/CD22 interface with CD22 connected to a lipid membrane model, developing a predictive capacity in silico protocol to propose effective mutations for developing new CARs. Molecular Dynamics simulations were executed, and the trajectory was analyzed to investigate the free binding energy between CD22 and the scFvs. The Fab fragment of M971 and the D6-D7 domains of CD22 are deposited in the PDB, code 7O52, and were used to generate 3D models submitted to simulations in an aqueous solution and lipid bilayer. The scFv/CD22 complexes were constructed, simulated, and evaluated structurally and energetically. The scFv/CD22 systems in the lipid bilayer were simulated in triplicate. As a result, it was observed that short and long linkers scFv/CD22 showed different structural stability from each other in aqueous solutions but were similar when in the cellular membrane environment. The average $\Delta G_{\text{binding}}$ between scFv/CD22 is close to the experimental $\Delta G_{\text{binding}}$ only when in the lipid membrane model. The results generated a protocol that can be used for the proposition of mutations to be tested in silico to proceed with incredible speed and qualities to the in vitro and in vivo tests, proposing new CARs that are more effective for CAR-T cell therapy.

Keywords: CARs anti-CD22; scFvs from M971; Molecular Dynamics; MM/PBSA.

LISTA DE FIGURAS

Figura 1	– Modelos estruturais de CARs.....	32
Figura 2	– Representações dos CARs anti-CD22 de linker curto e longo..	35
Figura 3	– Classificação dos mAbs de acordo com a sua estrutura e imunogenicidade.....	39
Figura 4	– Representação estrutural de um anticorpo da classe IgG.....	40
Figura 5	– Representação estrutural de um scFv.....	42
Figura 6	– Representação esquemática da glicoproteína de membrana CD22 na superfície celular interagindo com anticorpo terapêutico Epratuzumab.....	45
Figura 7	– Representação da interface de interação Fab-CD22.....	47
Figura 8	– (A) e (B) – Representação das estruturas de CD22 α e CD22 β , em cartoon, realizadas no Pymol, modeladas no programa de modelagem molecular Robetta.....	49
Figura 9	– Organização celular do CD22 e interações “cis” e “trans” com glicoproteínas de células B e de outras células.....	51
Figura 10	– Tipos de N-glicosilação mais comuns em seres eucarióticos....	52
Figura 11	– Ciclo termodinâmico para cálculo de energia livre de ligação por MM/PBSA.....	58
Figura 12	– Alinhamento entre as sequências de aminoácidos dos domínios extracelulares D6-D7 do CD22 obtidas pelo Protein Data Bank e pelo UniProt.....	71
Figura 13	– Esquema da árvore de glicosilação adicionada aos resíduos N70 e N130 do CD22 obtida através do Modeller 10.1.....	72
Figura 14	– Fluxograma que esquematiza a definição dos sistemas simulados.....	75
Figura 15	– Estrutura tridimensional dos modelos de scFv curto e longo na orientação VH-VL.....	84

Figura 16	– Estrutura tridimensional do CD22, obtido a partir de modelagem por homologia.....	87
Figura 17	– Estrutura tridimensional do modelo completo de CD22 glicosilado, obtido a partir da adição das ligações dissulfeto e glicosilações previstas nos resíduos N70 e N130, utilizando o programa CHARMM-GUI.....	88
Figura 18	– Estrutura tridimensional do modelo de domínio transmembrana obtido através do programa I-TASSER.....	90
Figura 19	– Gráfico Ramachandran gerado pelo programa MolProbity para o modelo do domínio transmembrana obtido pelo programa I-TASSER.....	91
Figura 20	– Estrutura tridimensional do modelo de domínio transmembrana inserido na membrana.....	92
Figura 21	– Perfil de RMSD dos C α dos componentes do scFv curto em solução aquosa durante 600 ns de simulação.....	93
Figura 22	– Curvas de RMSD relativo entre as estruturas tridimensionais dos domínios VH e VL do scFv curto submetido à 600 ns de simulação em solução aquosa.....	94
Figura 23	– Conformações do VH+VL coletadas em 50 ns e em 450 ns de simulação evidenciando as diferenças estruturais no domínio VH que foram observadas no gráfico do perfil de RMSD relativo	95
Figura 24	– Curvas de Rg dos domínios VH+VL, VH e VL do scFv de linker curto em solução aquosa ao longo de 600 ns de simulação.....	96
Figura 25	– Curvas de Análise de Componentes Principais.(A) PC1 e PC2 em função do tempo durante 600 ns de simulação. (B) Projeções da simulação do scFv de linker curto nos subespaços comuns definidos por PC1/PC2.....	97
Figura 26	– Resultado análise de cluster para a estrutura VH+VL do scFv com linker curto.....	98

Figura 27	– Representação estrutural das configurações do scFv curto após a sobreposição do domínio VH+VL do scFv de linker curto	99
Figura 28	– Perfil de RMSD dos C α dos componentes do scFv longo em solução aquosa, durante 600 ns de simulação.....	100
Figura 29	– Curvas de RMSD relativo entre as estruturas tridimensionais dos domínios VH e VL do scFv longo submetido à 600 ns de simulação em solução aquosa.....	101
Figura 30	– Conformações do VH+VL coletadas em 0 e 500 ns de simulação evidenciando as diferenças estruturais no domínio VL que foram observadas no gráfico do perfil de RMSD relativo.....	102
Figura 31	– Curvas de Rg dos domínios VH+VL, VH e VL do scFv de linker longo em solução aquosa ao longo de 600 ns de simulação.....	103
Figura 32	– Curvas de Análise de Componentes Principais. (A) PC1 e PC2 em função do tempo durante 600 ns de simulação. (B) Projeções da simulação do scFv de linker longo nos subespaços comuns definidos por PC1/PC2.....	104
Figura 33	– Resultado análise de cluster para a estrutura VH+VL do scFv com linker longo.....	105
Figura 34	– Representação estrutural das configurações do scFv longo, após a sobreposição do domínio VH+VL, do scFv de linker longo.....	106
Figura 35	– Perfis de RMSD dos C α das configurações estruturais do CD22, ao longo de 500 ns de simulação.....	106
Figura 36	– Conformações de CD22 coletadas em 0 e 310 ns de simulação evidenciando que as árvores de glicosilação estão bem distintas da posição inicial.....	107
Figura 37	– Perfil de RMSD para os C α do CD22-Glicanos dos domínios D6 e D7, comparadas com a estrutura de referência inicial.....	108

Figura 38	– Perfil de RMSD para os C α do CD22-Loops-Glicanos, considerando RMSD de D6 e sobreposição de D7 e vice-versa	109
Figura 39	– Clusters de configurações do CD22 ao longo de 500 ns simulação, utilizando raio de corte de 0,36 nm.....	110
Figura 40	– Perfil de RMSD dos C α das configurações estruturais do domínio transmembrana de CD22, ao longo de 200 ns de simulação, comparadas a estrutura inicial como referência.....	111
Figura 41	– Raio de Giro do domínio transmembrana de CD22, ao longo do tempo de simulação, comparadas a estrutura inicial como referência, nos sistemas 12 x 12 e 16 x 16 nm.....	112
Figura 42	– Perfil de RMSD do CD22-t considerando a sobreposição dos C α das configurações em comparação com a estrutura inicial de referência.....	113
Figura 43	– Perfil de RMSD dos C α das configurações estruturais do CD22 e Transmembrana, ao longo de 1000 ns de simulação, comparadas a estrutura inicial como referência.....	114
Figura 44	– Representação estrutural do CD22-t após a sobreposição do domínio transmembrana.....	115
Figura 45	– Perfis de RMSD para os domínios do scFv nos sistemas scFv/CD22, sendo que o scFv pode ser de linker curto ou longo	118
Figura 46	– Perfis de RMSD do CD22 e do VH+VL com sobreposição de CD22 dos sistemas scFv/CD22.....	120
Figura 47	– Curvas de PCA para os sistemas scFv/CD22 curto e longo. A) e C) PC1 e PC2 em função do tempo durante 600 ns de simulação. B) e D) Projeções das curvas PC1/PC2.....	121
Figura 48	– Curvas de Δ Gligação dos complexos durante 600 ns de simulação entre o (A) scFv e o CD22 de linker curto e (B) scFv de linker longo.....	122

Figura 49	– PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22, A) curto. B) longo.....	123
Figura 50	– Perfis de RMSD para os domínios do scFv nos sistemas scFv/CD22, sendo que o scFv pode ser de linker curto ou longo.....	125
Figura 51	– Perfis de RMSD do CD22 e do VH+VL com sobreposição de CD22 dos sistemas scFv/CD22.....	127
Figura 52	– Curvas de PCA para os sistemas scFv/CD22 curto e longo. A) e C) PC1 e PC2 em função do tempo durante 600 ns de simulação. B) e D) Projeções das curvas PC1/PC2.....	128
Figura 53	– Curvas de Δ Gligação dos complexos durante 600 ns de simulação entre o (A) scFv e o CD22 de linker curto e (B) scFv de linker longo.....	129
Figura 54	– PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22, A) curto. B) longo.....	130
Figura 55	– Perfis de RMSD para os domínios do scFv nos sistemas scFv/CD22, sendo que o scFv pode ser de linker curto ou longo	132
Figura 56	– Perfis de RMSD do CD22 e do VH+VL com sobreposição de CD22 dos sistemas scFv/CD22.....	133
Figura 57	– Curvas de PCA para os sistemas scFv/CD22 curto e longo. A) e C) PC1 e PC2 em função do tempo durante 400 ns de simulação. B) e D) Projeções das curvas PC1/PC2.....	134
Figura 58	– Curvas de Δ Gligação dos complexos durante 400 ns de simulação entre o (A) scFv e o CD22 de linker curto e (B) scFv de linker longo.....	135
Figura 59	– PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22, A) curto. B) longo.....	136

Figura 60	– Representação estrutural da configuração inicial dos sistemas scFv/CD22-t curto e longo.....	137
Figura 61	– Perfis de RMSD para os domínios do scFv e CD22 do sistema scFv/CD22-t, scFv de linker curto. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas os RMSDs VH+VL, VH e VL (A, B e C) e CD22 (D, E e F).....	139
Figura 62	– Perfis de RMSD para os domínios do scFv e CD22 do sistema scFv/CD22-t, scFv de linker longo. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas os RMSDs VH+VL, VH e VL (A, B e C) e CD22 (D, E e F).....	140
Figura 63	– Perfis de RMSD de VH+VL após sobreposição do CD22 dos sistemas scFv/CD22-t curto e longo.....	142
Figura 64	– Curvas de PCA de VH+VL/CD22 para os sistemas scFv/CD22-t curto em triplicata de 400 ns de simulação. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas PC1 e PC2 em função do tempo (A, B e C) e Projeções das curvas PC1/PC2 (D, E e F).....	143
Figura 65	– Curvas de PCA de VH+VL/CD22 para os sistemas scFv/CD22-t longo em triplicata de 400 ns de simulação. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas PC1 e PC2 em função do tempo (A, B e C) e Projeções das curvas PC1/PC2 (D, E e F).....	144
Figura 66	– PII entre os scFvs e a membrana, dentro de um raio de corte de 1,3 nm, nos sistemas scFv/CD22-t.....	145
Figura 67	– PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a membrana, dentro de um raio de corte de 1,3 nm, nos sistemas scFv/CD22-t.....	147
Figura 68	– PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3), dentro de um raio de corte de 0,5 nm, nos sistemas scFv/CD22-t.....	148

Figura 69	– PII entre a glicosilação, no domínio D7 de CD22, e a membrana, dentro de um raio de corte de 1,3 nm, nos sistemas scFv/CD22-t.....	149
Figura 70	– Curvas de ΔG ligação das triplicatas de 400 ns de simulação dos complexos scFv/CD22-t curto e longo.....	150
Figura 71	– PII entre resíduos do complexo scFv/CD22-t curto dentro de um raio de corte de 0,5 nm.....	152
Figura 72	– PII entre resíduos do complexo scFv/CD22-t longo dentro de um raio de corte de 0,5 nm.....	153
Figura 73	– Representação em cartoon da interface de interação scFv-CD22-t curto e longo evidenciando os resíduos (em sticks) que contribuem para interação atrativa mostrado pelas Figuras 71 e 72.....	154
Figura 74	– Representação em cartoon da sobreposição dos scFvs curto e longo evidenciando as diferenças de posicionamento das CDRs, em especial, da CDR-H3.....	156
Figura 75	– Perfil de RMSD dos C α das configurações estruturais da CDR-H3 nos sistemas scFv curto e longo, no decorrer de 600 ns de simulação, comparadas a estrutura inicial e final como referência.....	157
Figura 76	– Comparação dos experimentos realizados por Singh e colaboradores (2021) e Ereno-orbea e colaboradores (2021) com a situação realista de inserção dos scFvs em uma célula CAR-T anti-CD22.....	161

LISTA DE QUADROS

Quadro 1	– Resíduos que fazem parte dos tipos de linker utilizados em construções de scFvs.....	43
Quadro 2	– CDRs dos domínios VH e VL do M971 determinadas pelo método de Kabat utilizando o programa AbRSA.....	46
Quadro 3	– Em negrito e sublinhado os resíduos do M971 que constituem a área de superfície de contato descritos por Ereño-Orbea e colaboradores (2021).....	47
Quadro 4	– Sequências de aminoácidos que foram utilizados para a modelagem dos dois modelos de scFv. As sequências dos domínios variáveis (VH e VL) foram derivadas do anticorpo M971.....	70
Quadro 5	– Identificação da sequência de aminoácidos e da numeração dos loops faltantes na estrutura cristalográfica que foram modelados por meio do programa Modeller 10.1.....	71
Quadro 6	– Sequência de aminoácidos do domínio transmembrana da proteína de membrana CD22.....	73
Quadro 7	– Quantidade dos tipos de fosfolípido utilizado para a construção da membrana 12 x 12 nm de linfócitos B humana no CHARMM-GUI.....	74
Quadro 8	– Quantidade dos tipos de fosfolípido utilizado para a construção da membrana 16 x 16 nm de linfócitos B humana no CHARMM-GUI.....	74
Quadro 9	– Quantidade de íons Na ⁺ e Cl ⁻ na concentração 0.15 M, bem como de moléculas de água adicionadas em cada sistema scFv-água.....	76
Quadro 10	– Quantidade de íons Na ⁺ e Cl ⁻ na concentração 0.10 M, bem como de moléculas de água adicionadas em cada sistema scFv-CD22.....	79

Quadro 11	– Quantidade de íons Na ⁺ e Cl ⁻ na concentração 0.10 M, bem como de moléculas de água adicionadas em cada sistema scFv-CD22.....	81
Quadro 12	– Quantidade de íons Na ⁺ e Cl ⁻ na concentração 0.05 M, bem como de moléculas de água adicionadas em cada sistema scFv-CD22.....	82

LISTA DE TABELAS

Tabela 1	– Resultados obtidos por meio do programa MolProbity para o modelo de scFv com linker curto obtido pela modelagem molecular, antes e após a etapa de minimização de energia.....	85
Tabela 2	– Resultados obtidos por meio do programa MolProbity para o modelo de scFv com linker longo obtido pela modelagem molecular, antes e após a etapa de minimização de energia.....	85
Tabela 3	– Resultados obtidos por meio do programa PROCHECK para o modelo de scFv com linker curto obtido pela modelagem molecular, antes e após a etapa de minimização de energia.....	86
Tabela 4	– Resultados obtidos por meio do programa PROCHECK para o modelo de scFv com linker longo obtido pela modelagem molecular, antes e após a etapa de minimização de energia.....	86
Tabela 5	– Resultados obtidos por meio do programa MolProbity para o modelo de CD22.....	89
Tabela 6	– Resultados obtidos por meio do programa PROCHECK para o modelo de CD22.....	90
Tabela 7	– RMSDs médios calculados para VH+VL, VH e VL a partir do teq = 50 ns.....	93
Tabela 8	– RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 50 ns.....	94
Tabela 9	– Raio de Giro médio calculados para a estrutura VH+VL, VH e VL do scFv de linker curto a partir de 50 ns de simulação.....	96
Tabela 10	– RMSDs médios calculados para VH+VL, VH e VL a partir do teq = 110 ns.....	100
Tabela 11	– RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados do teq = 110 ns.....	101

Tabela 12 – Raio de Giro médio calculados para a estrutura VH+VL, VH e VL do scFv de linker longo a partir de 110 ns de simulação.....	103
Tabela 13 – RMSDs médio calculados para a estrutura completa do CD22, com a estrutura desconsiderando os glicanos (CD22-Glicanos) e desconsiderando os glicanos e os Loops modelados (CD22-Loops-Glicanos), teq = 100 ns de simulação.....	107
Tabela 14 – RMSDs médios calculados para domínios D6 e D7 de CD22 a partir do teq = 100 ns.....	109
Tabela 15 – RMSDs médios dos domínios D6 tendo como referência para sobreposição o D7, e vice-versa, calculados a partir do teq=100 ns.....	109
Tabela 16 – RMSDs médios calculados para CD22 e Transmembrana a partir do teq = 100 ns.....	114
Tabela 17 – Potencial de Interação Intermolecular entre VH+VL e CD22 em configurações scFv de linker curto e longo (olhar seções 8.1.1 e 8.1.2) coletadas para a realização dos dockings.....	117
Tabela 18 – Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL em ambos os sistemas scFv/CD22 linkers (curto e longo) a partir do teq = 50 ns.....	119
Tabela 19 – RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 50 ns.....	119
Tabela 20 – Médias e desvios de PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a glicosilação, no domínio D7 nos complexos scFv/CD22, nos sistemas com o scFv de linker curto e longo, dentro de um raio de corte de 0,5 nm.....	123
Tabela 21 – Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL em ambos os sistemas scFv/CD22 linkers (curto e longo) a partir do teq = 150 ns.....	125

Tabela 22 – RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 150 ns.....	126
Tabela 23 – Médias e desvios de PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a glicosilação, no domínio D7 nos complexos scFv/CD22, nos sistemas com o scFv de linker curto e longo, dentro de um raio de corte de 0,5 nm.....	130
Tabela 24 – Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL em ambos os sistemas scFv/CD22 linkers (curto e longo) a partir do teq = 50 ns.....	132
Tabela 25 – RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 50 ns.....	132
Tabela 26 – Médias e desvios de PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a glicosilação, no domínio D7 nos complexos scFv/CD22, nos sistemas com o scFv de linker curto e longo, dentro de um raio de corte de 0,5 nm.....	136
Tabela 27 – Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL da triplicata dos sistemas scFv/CD22 de linker curto e longo em membrana a partir de 75 ns	140
Tabela 28 – Médias e desvios de RMSD calculados para CD22 dos sistemas scFv/CD22-t curto e longo a partir de 75 ns de simulação.....	141
Tabela 29 – Médias e desvios de RMSD calculados para sobreposição CD22 e RMSD VH+VL dos sistemas scFv/CD22-t curto e longo a partir de 75 ns de simulação.....	142
Tabela 30 – Médias e desvios de PII entre os scFvs e a membrana nos complexos scFv/CD22-t de linker curto e longo, dentro de um raio de corte de 1,3 nm, calculadas a partir do teq = 75 ns.....	146

Tabela 31 – Médias e desvios de PII entre CDRs da cadeia pesada (H1, H2 e H3) e a membrana nos sistemas scFv/CD22-t da triplicata (n1), (n2) e (n3), dentro de um raio de corte de 1,3 nm, calculadas a partir do $teq = 75$ ns.....	147
Tabela 32 – Médias e desvios de PII entre a glicosilação, no domínio D7 de CD22-t, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22-t da triplicata (n1), (n2) e (n3), em um raio de corte de 0,5 nm, calculadas a partir do $teq.=75$ ns.....	149
Tabela 33 – Médias e desvios de PII entre a glicosilação N-ligada a D7 e a membrana nos complexos scFv/CD22-t de linker curto e longo, dentro de um raio de corte de 1,3 nm, calculadas a partir do $teq = 75$ ns.....	150
Tabela 34 – Resumo dos ΔG ligação médios dos sistemas scFv/CD22-t de linker curto e longo em triplicata n1, n2 e n3.....	151
Tabela 35 – Resumo dos de ΔG ligação médios aferidos por MM/PBSA dos sistemas scFv/CD22 de linker curto e longo na ausência da membrana (scFv/CD22-t).....	160
Tabela 36 – Resumo dos ΔG ligação médios dos sistemas scFv/CD22-t de linker curto e longo em triplicata n1, n2 e n3.....	162
Tabela 37 – Resumo dos K_d médios dos sistemas scFv/CD22-t de linker curto e longo em triplicata n1, n2 e n3.....	162

LISTA DE ABREVIATURAS E SIGLAS

3D	Tridimensional
C α	Carbono alfa
CAR	Receptor Quimérico de Antígeno
CD22	<i>Cluster of Differentiation 22</i>
CDR	Regiões Determinantes de Complementariedade
CH	Domínio Constante da Cadeia Pesada
CL	Domínio Constante da Cadeia Leve
DM	Dinâmica Molecular
Fab	<i>Fragment antigen-binding</i>
Fc	Fragmento cristalizado
FDA	<i>Food and Drug Administration</i>
Fiocruz/CE	Fundação Oswaldo Cruz/Ceará
Fv	Fragmento Variável
GROMACS	<i>Groningen Machine for Chemical Simulation</i>
Kcal	Quilocaloria
LLA	Leucemia Linfóide Aguda
mAb	Anticorpo monoclonal
MM/PBSA	Mecânica Molecular da Área Superficial de <i>Poisson-Boltzmann</i>
Nm	Nanômetro
NPT	Número de partículas, Pressão e Temperatura
Ns	Nanossegundo
NVT	Número de partículas, Volume e Temperatura
PCA	Análise de Componentes Principais
PDB	<i>Protein Data Bank</i>
PII	Potencial de Interação Intermolecular
RMSD	<i>Root Mean Square Deviation</i>
scFv	Fragmento Variável de Cadeia Única
VH	Domínio Variável da Cadeia Pesada
VL	Domínio Variável da Cadeia Leve

LISTA DE SÍMBOLOS

α	Alpha
δ	Delta
ε	Épsilon
γ	Gama
κ	Capa
\pm	Mais-menos
μ	Mi
λ	Lambda
σ	Sigma
θ	Teta
%	Porcentagem
®	Marca Registrada
Z	Zeta

SUMÁRIO

1	INTRODUÇÃO	29
2	REFERENCIAL TEÓRICO	31
2.1	Aspectos estruturais, funcionais e clínicos de CARS.....	31
2.2	Anticorpos.....	36
2.2.1	Aspectos Gerais.....	36
2.2.1	Desenvolvimento de MAbs murinos e humanizados.....	38
2.2.3	Estrutura de Anticorpos.....	40
2.3	Anticorpo M971	44
2.4	CD22: estrutura e função.....	47
2.5	Padrões de Glicosilação.....	51
2.6	Glicosilações em CD22.....	52
2.7	Simulação de Dinâmica Molecular.....	53
2.8	Cálculo de Energia Livre de Ligação	56
2.8.1	Mecânica Molecular de Poisson-Boltzmann com Solvatação da Área de Superfície (MM/PBSA)	57
2.9	Engenharia de Proteínas <i>in silico</i>.....	60
3	JUSTIFICATIVA	61
4	OBJETIVOS	62
4.1	Objetivo Geral	62
4.2	Objetivos Específicos.....	62
5	METODOLOGIA	63
5.1	Modeller.....	63
5.2	I-TASSER.....	64
5.3	Avaliação dos Modelos Obtidos.....	65
5.4	Simulação de Dinâmica Molecular e Análises.....	66
5.5	Docking Molecular e o PROGRAMA Haddock.....	68
6	PROCEDIMENTOS EXPERIMENTAIS.....	69
6.1	Modelagem Molecular.....	69
6.1.1	Modelagem dos scFvs de M971 anti-CD22.....	69
6.1.2	Modelagem dos domínios extracelulares D6-D7 glicosilados do CD22..	70
6.1.3	Modelagem do domínio transmembrana do CD22.....	72

6.1.4	Definição dos sistemas simulados	74
6.2	Montagem dos sistemas e simulação de Dinâmica Molecular	76
6.2.1	Montagem e simulação de DM dos sistemas scFv-água.....	76
6.2.2	Montagem e simulação de DM do sistema contendo os domínios extracelulares D6-D7 glicosilados em água.....	76
6.2.3	Montagem e simulação de DM dos sistemas contendo o domínio transmembrana de CD22.....	77
6.2.4	Montagem e simulação de DM do CD22 em membrana.....	78
6.2.5	Montagem e DM dos sistemas scFv/CD22 a partir do Docking Molecular.....	78
6.2.6	Montagem e simulação de DM dos sistemas scFv/CD22 com estruturas advindas da sobreposição de simulações anteriores à cristalográfica.....	79
6.2.7	Montagem e simulação de DM dos sistemas scFv/CD22 com estruturas advindas da cristalográfica.....	81
6.2.8	Montagem e simulação de DM dos sistemas scFv/CD22 em membrana.....	82
7	RESULTADOS.....	83
7.1	Modelagem Molecular.....	83
7.1.1	Modelagem e validação dos scFv de M971 anti-CD22	83
7.1.2	Modelagem e validação estrutural do CD22.....	86
7.1.3	Modelagem e validação do domínio transmembrana de CD22.....	89
7.2	Simulação de Dinâmica Molecular e Análises.....	92
7.2.1	Dinâmica Molecular do scFv curto.....	92
7.2.2	Dinâmica Molecular do scFv longo.....	99
7.2.3	Análise estrutural do CD22 em solução aquosa.....	106
7.2.4	DM do sistema domínio transmembrana do CD22 em membrana de 12 x 12 nm e 16 x 16 nm	110
7.2.5	Sistema composto por CD22 conectado ao domínio transmembrana, inserido em modelo de membrana celular.....	112
7.3	Estudos dos complexos scFv/CD22 em solução aquosa.....	116

7.3.1	Estudo da interface scFv/CD22 proveniente de docagem molecular do scFv ao CD22.....	116
7.3.2	Estudo da interface scFv/CD22 com o complexo formado por meio das estruturas de scFv e CD22 sobreposta à estrutura cristalográfica scFv/CD22.....	124
7.3.3	Estudo da interface scFv/CD22 usando as coordenadas cristalográficas.....	131
7.4	Estudo da interface scFv/CD22, com scFv de linker curto e longo e o CD22 inserido em membrana	136
8	DISCUSSÃO.....	155
9	CONCLUSÕES.....	164
10	PERSPECTIVAS FUTURAS.....	165
	REFERÊNCIAS.....	166
	APÊNDICE A – MM/PBSA PARA SISTEMAS NA AUSÊNCIA DA MEMBRANA.....	176
	APÊNDICE B – MM/PBSA PARA SISTEMAS NA PRESENÇA DA MEMBRANA.....	177
	ANEXO A – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv CURTO ANTES DA MINIMIZAÇÃO DE ENERGIA.....	178
	ANEXO B – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv CURTO APÓS A MINIMIZAÇÃO DE ENERGIA.....	179
	ANEXO C – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv LONGO ANTES DA MINIMIZAÇÃO DE ENERGIA.....	180
	ANEXO D – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv LONGO APÓS A MINIMIZAÇÃO DE ENERGIA.....	181
	ANEXO E – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE CD22 OBTIDO PELO MODELLER 10.1 ANTES DA MINIMIZAÇÃO DE ENERGIA.....	182

ANEXO F – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE CD22 OBTIDO APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E ANTES DA MINIMIZAÇÃO DE ENERGIA.....	183
ANEXO G – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE CD22 OBTIDO APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E SEGUIDA DE MINIMIZAÇÃO DE ENERGIA.....	184
ANEXO H – ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE TRANSMEMBRANA DE CD22 OBTIDO APÓS A MODELAGEM NO I-TASSER ANTES DA MINIMIZAÇÃO DE ENERGIA.....	185
ANEXO I – GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv CURTO ANTES DA MINIMIZAÇÃO DE ENERGIA.....	186
ANEXO J – GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv CURTO APÓS A MINIMIZAÇÃO DE ENERGIA.....	187
ANEXO K – GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv LONGO ANTES DA MINIMIZAÇÃO DE ENERGIA.....	188
ANEXO L – GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv LONGO APÓS A MINIMIZAÇÃO DE ENERGIA.....	189
ANEXO M – GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE CD22 OBTIDO PELO MODELLER 10.1 ANTES DA MINIMIZAÇÃO DE ENERGIA.....	190
ANEXO N – GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE CD22 OBTIDO APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E ANTES DA MINIMIZAÇÃO DE ENERGIA.....	191

**ANEXO O – GRÁFICO DE RAMACHANDRAN GERADO PELO
PROGRAMA PROCHECK PARA O MODELO DE CD22 OBTIDO
APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E
SEGUIDA DE MINIMIZAÇÃO DE ENERGIA..... 192**

1. INTRODUÇÃO

Câncer, ou neoplasia maligna é uma doença multifatorial que possui como característica o crescimento e multiplicação desordenada das células. Essas células que se multiplicam excessivamente podem invadir tecidos e órgãos, em um processo chamado de metástase, sendo a principal causa de morte por neoplasias malignas na atualidade (INCA, 2020).

Existem diversos tipos de neoplasias malignas, os quais podem ser agrupados em categorias, dependendo do tipo celular onde se originam. As principais categorias são: os carcinomas, que começam na pele ou nos tecidos que revestem ou cobrem órgãos; os sarcomas, originários do osso, cartilagem, gordura, músculo ou outro tecido conjuntivo; os cânceres do sistema nervoso central, que começam nos tecidos do cérebro e da medula espinhal; e leucemias, provenientes do tecido que produz sangue, como medula óssea (INCA, 2020).

No Brasil, estima-se que para cada ano do triênio 2023-2025 ocorrerão cerca de 704 mil novos casos de neoplasias malignas e, entre estes, aproximadamente 11.540 mil correspondem a algum tipo de leucemia. A leucemia é um tipo de neoplasia maligna que surge a partir de uma mutação genética em uma célula sanguínea. Sendo assim, as células sanguíneas que possuem a mutação genética multiplicam-se rapidamente, substituindo as células sanguíneas saudáveis da medula óssea (INCA, 2022).

A Leucemia Linfóide Aguda (LLA) é um tipo agressivo de leucemia que, em geral, surge a partir de uma mutação em células linfóides, especificamente nos linfócitos B (LLA-B) ou nos linfócitos T (LLA-T). A LLA é frequentemente diagnosticada em crianças e jovens de 0 a 19 anos. Na faixa pediátrica, cerca de 80% dos casos são LLA-B. Já na adolescência, apesar de maior prevalência continuar sendo LLA-B, são diagnosticadas mais LLA-T do que quando comparadas às crianças escolares e lactantes (INCA, 2020). Os tratamentos tradicionais mais comuns para a remissão da LLA-B é o transplante de medula óssea com sessões de quimioterapia e radioterapia como intervenções auxiliares. Ademais, outro tratamento tradicional utilizado como tentativa de retrocesso dessa enfermidade é representado por sessões de quimioterapia e transfusão de anticorpos monoclonais (mAbs). A eficácia dessas infusões de mAbs fornece uma justificativa para desenvolver outras imunoterapias celulares a fim de tratar as malignidades que acometem as células linfóides do tipo B

(KOCHENDERFER *et al.*, 2010).

Atualmente, vem sendo desenvolvidas inúmeras imunoterapias para o tratamento de diversos tipos de neoplasias malignas, inclusive para a remissão da LLA-B (BAYBUTT *et al.*, 2019). Dentre essas imunoterapias pode-se destacar a terapia por células CAR-T, em que um Receptor Quimérico de Antígeno (CAR) é expresso na superfície de linfócitos T (CAR-T), o qual possui a capacidade de redirecionar a ação da célula T, a fim de inibir e eliminar as células cancerígenas que expressam antígenos específicos (ABDO; ARAGÃO; BONAMINO, 2020).

Em geral, os CARs possuem três domínios ou porções, denominadas de domínio extracelular, transmembrana e intracelular, na qual a porção extracelular é responsável pelo reconhecimento antigênico (SERMER; BRENTJENS, 2019). Tal funcionalidade do domínio extracelular é proporcionada pela presença da estrutura do fragmento variável de cadeia única (scFv). O scFv, é um domínio de reconhecimento de um mAb, sendo, portanto, ele isoladamente considerado como um mAb (SALTER; MARGOT; STANLEY, 2018).

Em 2017, o *Food and Drug Administration* (FDA, órgão regulatório americano) aprovou dois CARs comerciais, Kymriah® da Novartis, para tratamento de leucemia linfocítica aguda recorrente/refratária de células B e Yescarta® da Kite Pharma, para o tratamento de linfoma não-Hodgkin de células B recorrente ou refratário. Esses CARs comerciais reconhecem o antígeno específico CD19, significativamente, expresso nos linfócitos B malignos, mas não encontrado em células pró-B, isto é, células que dão origem aos linfócitos B (BAYBUTT *et al.*, 2019). Apesar do avanço obtido para tratamento da LLA-B e de linfoma não-Hodgkin por meio do uso desses CARs comerciais, estudos demonstraram que alguns pacientes não conseguem obter uma resposta imune duradoura, devido ao escape de antígeno CD19 (ZHANG *et al.*, 2022). Sabendo disso, diversos grupos de pesquisa desenvolveram CARs capazes de reconhecer o antígeno CD22, alguns desses novos CARs continham como domínio de reconhecimento antigênico, o fragmento variável de cadeia única (scFv) originário de BL22¹ e HA22². Também foram realizados estudos com CARs anti-CD22 utilizando como scFv derivado do mAb denominado M971. Esses estudos demonstraram que a

¹ BL22 é uma imunotoxina recombinante composta pelo domínio Fv do mAb RFB4 unido a um fragmento de 38 kDa da exotoxina A de *Pseudomonas*.

² HA22 é um mutante de BL22 com mutações na CDR-H3, resultando em aumento da atividade citotóxica.

localização do epítopo na macromolécula CD22, com 7 domínios, possui grande influência na eficiência do CAR. Uma vez que foi observado que CARs com scFv de M971 resultaram em atividade antileucêmica mais potente do que CARs anti-CD22 de afinidade semelhante, visando epítomos distintos, em domínios de CD22 mais distais da membrana (HASO *et al.*, 2013). Diante dessa realidade, atualmente, estão sendo desenvolvidos CARs anti-CD22 e Bi específicos anti-CD19/CD22, utilizando o domínio de reconhecimento ao antígeno advindo do mAb humano M971, reconhecendo epítopo no domínio D7 do CD22 (DAI *et al.*, 2020; SINGH *et al.*, 2021).

2. REFERENCIAL TEÓRICO

2.1 Aspectos estruturais, funcionais e clínicos de CARS

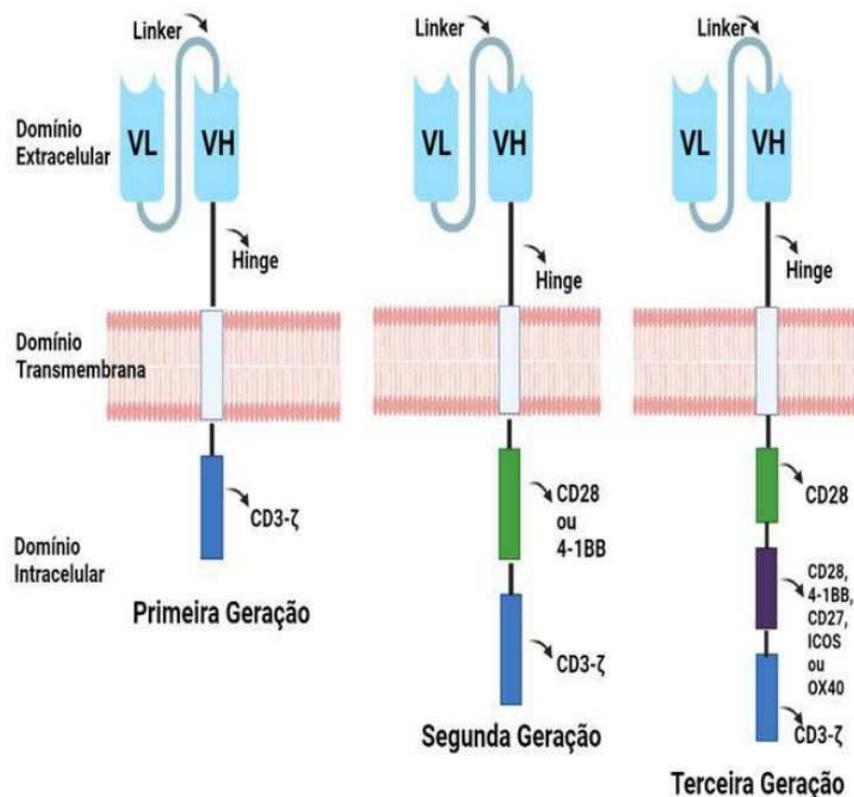
CARs são moléculas de proteína recombinante que podem ser expressas em células imunológicas, com a finalidade de direcionar a ação delas contra células alvo, como células cancerígenas, que expressam um marcador tumoral específico definido como antígeno para o CAR (TOMULEASA *et al.*, 2018). Tal redirecionamento deve-se a capacidade do CAR de conferir maior especificidade à célula T no reconhecimento de um antígeno tumoral (ABDO; ARAGÃO; BONAMINO, 2020).

A Figura 1, apresenta as estruturas de CAR relacionadas as suas gerações. Em geral, em todas as gerações de CARs tem-se o domínio de reconhecimento ao antígeno, que normalmente é composto por um scFv. O scFv é formado pelos domínios variáveis da cadeia leve (VL) e pesada (VH) de um anticorpo monoclonal (mAb), os quais são conectados por um *linker* (ABDO; ARAGÃO; BONAMINO, 2020). Além disso, tem-se o espaçador (*hinge*) e o domínio transmembrana. O *hinge* é responsável por conectar o scFv ao domínio transmembrana, além de proporcionar mobilidade à porção extracelular (GUEST *et al.* 2005). Os *hinges* mais comuns são derivados de IgG1, CD8 α e IgG4 (ALABANZA *et al.*, 2017). O domínio transmembrana é responsável por ancorar o CAR à membrana das células T, além de ter como função realizar a conexão entre o domínio extracelular e o domínio intracelular. Os tipos de domínios transmembranas mais utilizados nos CARs são derivados de CD4, CD28 ou CD8 α (IMAI *et al.*, 2004; KOCHENDERFER *et al.*, 2009).

Ainda conforme a Figura 1, o domínio intracelular é a porção ativadora de células T, formada geralmente por CD3- ζ (TASIAN; GARDNER, 2015). Entretanto, a

sinalização de apenas CD3- ζ , nos CARs de primeira geração, em muitos casos, não proporciona uma resposta imune prolongada, uma vez que sem um segundo sinal, as células T sofrem apoptose logo após o primeiro estímulo (SALOMON; BLUESTONE, 2001). Sabendo disso, os CARs de segunda e terceira geração foram desenvolvidos visando otimizar a sinalização.

Figura 1- Modelos estruturais de CARs.



Fonte: elaborada pela autora (2024). Estruturas de CARs de primeira geração, segunda e terceira, exemplificando os três domínios que o compõem: extracelular, transmembrana e intracelular.

Nos CARs de segunda geração, o CD3- ζ é acompanhado por domínio CD28 ou 4-1BB, ligado ao transmembrana, como segundo estímulo na região do endodomínio. Essa estruturação demonstrou melhor atividade antitumoral *in vivo*, em modelos pré-clínicos e ensaios clínicos para células B malignas, quando comparados com CAR de primeira geração. A explicação funcional é que o segundo sinal, o sinal coestimulador, produzido por uma molécula coestimuladora, como CD28, que promove a síntese de interleucina 2 (IL-2) para completar a ativação das células T e evitar a apoptose (PARK *et al.*, 2011).

Já em relação aos CARs de terceira geração, esses possuem dois domínios

coestimulatórios em sua porção intracelular, entre os domínios transmembrana e CD3- ζ . O segundo domínio coestimulatório é geralmente constituído por 4-1BB, CD28, CD27, ICOS ou OX40 (MAUS *et al.*, 2014). Apesar dos esforços realizados para construção de CARs de terceira geração, ainda não está claro quais os benefícios deste tipo de CAR quando comparados aos da geração anterior, sendo realizados vários estudos clínicos visando averiguar se a inserção do segundo domínio coestimulatório no CAR de terceira geração aumenta a eficiência da terapia (TANG *et al.*, 2016; DALAL *et al.*, 2018; GEORGE *et al.*, 2020).

Um dos desafios para o uso dos CARs comerciais Kymriah® e Yescarta® para o tratamento da LLA e LNH, respectivamente, refere-se ao aparecimento de diversos efeitos adversos em pacientes. Estudos sugerem que esses efeitos adversos ocorrem devido aos scFvs presentes nestes CARs serem derivados do anticorpo murino FMC63, um mAb anti-CD19. Diante dessa realidade, vem crescendo o interesse em desenvolver novos CARs utilizando scFvs derivados de anticorpos humanizados e, até mesmo, de anticorpos humanos, a fim de reduzir a imunogenicidade (ALABANZA *et al.*, 2017; SOMMERMEYER *et al.*, 2017). Um outro desafio para a utilização dos CARs comerciais Kymriah® e Yescarta® refere-se ao escape de antígeno CD19, assim, não favorecendo uma resposta imune duradoura (ZHANG *et al.*, 2022).

Diante desse contexto, tornou-se necessário a busca por novos antígenos-alvo para as células CAR-T destinadas ao tratamento da LLA. Um desses novos alvos é o CD22. Em geral, CARs anti-CD22 utilizam um scFv em sua porção extracelular proveniente do mAb humano M971 (FRY *et al.*, 2018; DAI *et al.*, 2020; SINGH *et al.*, 2021; TAN *et al.*, 2021).

Conforme a Figura 2A, Fry e colaboradores (2018) desenvolveram um CAR anti-CD22 com o domínio extracelular incluindo um scFv de M971 de *linker* curto (GGGGS) e *hinge* de CD8, um domínio transmembrana de CD8 TM e na porção intracelular contendo um domínio de sinalização CD3- ζ e um coestimulatório 4-1BB. Esse CAR está em fase de teste clínico I sob código NCT02315612, em pacientes com idade pediátrica e adulta, realizado pelo *National Cancer Institute* (NCI). Já Ruella e colaboradores (2017) desenvolveram um CAR anti-CD22 diferente do anterior, de *linker* curto, como descrito na Figura 2A, com o domínio extracelular incluindo um scFv de M971 de *linker* longo (GGGGS)₄ e *hinge* de CD8, um domínio transmembrana de CD8 TM e na porção intracelular contendo um domínio de sinalização CD3- ζ e um coestimulatório 4-1BB. O estudo está em fase I de teste clínico sob códigos

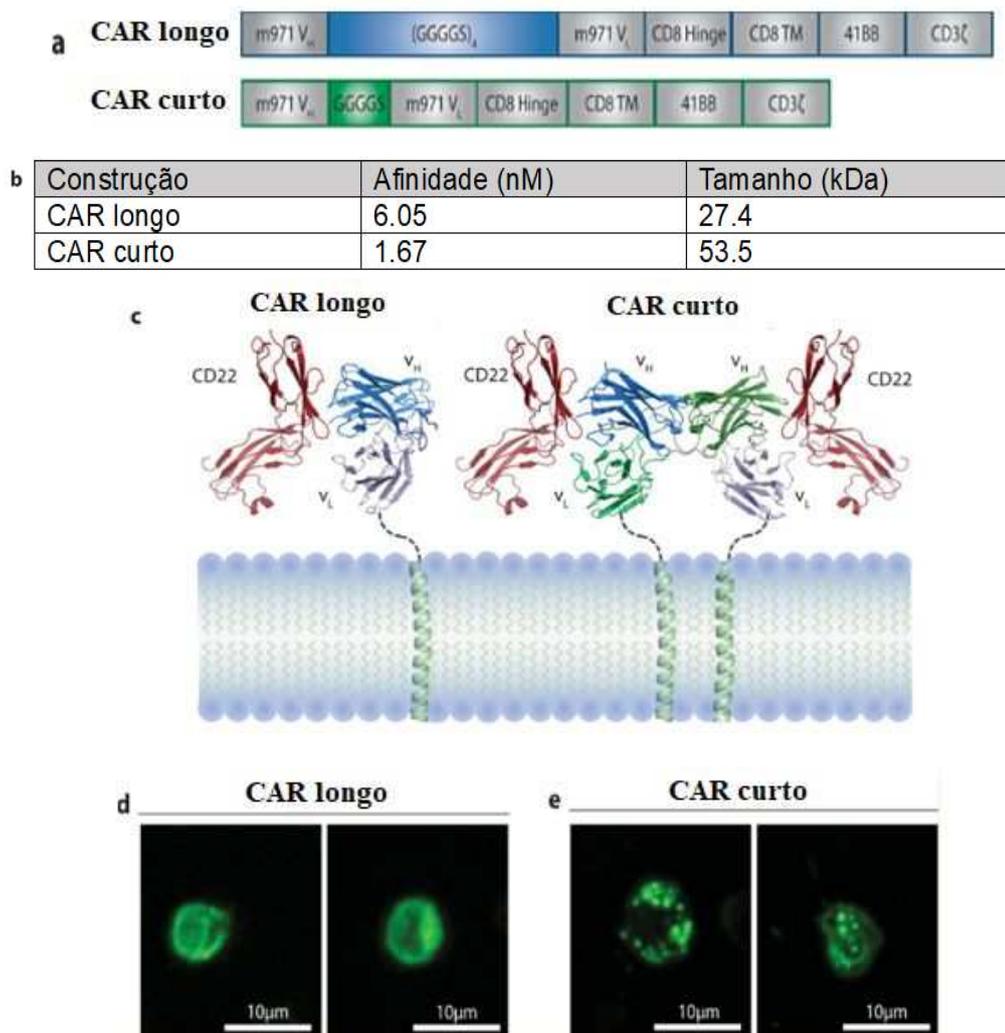
NCT02650414, em pacientes com idade pediátrica; e NCT02588456, em pacientes com idade adulta, realizado pela *University of Pennsylvania* (Penn/CHOP).

Nos ensaios pré-clínicos, em que os CARs anti-CD22 de *linker* curto e longo foram testados, em modelos de camundongos, constatou-se que o encurtamento do *linker* para 5 resíduos entre os domínios VH e VL do scFv de M971 aumentou significativamente a atividade antileucêmica da célula CAR-T. Além disso, o trabalho desses pesquisadores mostrou que o CAR de *linker* curto apresentou maior proliferação *in vivo* em comparação com o CAR de *linker* longo (RUELLA *et al.*, 2017).

Já nos ensaios clínicos, realizados por Singh e colaboradores (2021), em que os CARs anti-CD22 de *linker* curto e longo foram testados em pacientes humanos com LLA, observou-se que as células CAR-T que possuíam o CAR de *linker* curto eram menos móveis e permaneciam em contato com o alvo CD22 por mais tempo do que as células CAR-T que possuíam o CAR de *linker* longo, consistente com a formação de sinapses imunológicas mais duradouras. Visando maiores detalhes, a afinidade dos scFvs de M971 *linker* curto e longo pelo CD22 foi avaliada, na tentativa de justificar a formação de sinapses mais duradouras. As afinidades dos complexos scFv/CD22 de *linker* curto e longo apresentaram constantes de dissociação (K_d) de 1.67 e 6.05 nM, como descrito na Figura 2B, concluindo que as afinidades dos scFvs ao CD22 semelhantes, sendo assim, a afinidade de ligação não justifica a formação de sinapses mais produtivas no CAR de *linker* curto. O trabalho também discute sobre a formação de interface entre VH e VL dos scFvs de *linker* curto e longo. As células CAR-T de *linker* longo permitiram a formação de interface apropriada dos domínios variáveis VH e VL do scFv de M971, enquanto que as de *linker* curto impossibilitaram a formação de interface intramolecular correta de VH e VL do scFv. Buscando maiores esclarecimentos sobre a estruturação dos scFvs, a oligomerização do scFv de *linker* curto e longo foi avaliada usando cromatografia de exclusão molecular. Foi constatado que o scFv de *linker* longo se apresentou como monômero, enquanto o scFv de *linker* curto se apresentou na forma de dímero (Figura 2C). Conforme as Figuras 2D e 2E, os autores verificaram através de microscopia confocal que os CARs de *linker* longo se distribuíram uniformemente pela superfície celular, enquanto os CARs de *linker* curto apresentaram regiões bem definidas, sugerindo que se agregaram (SINGH *et al.*, 2021). A agregação do CAR de *linker* curto foi avaliada sem e com a presença do CD22 na célula. O agrupamento do CAR curto na ausência do CD22 resultou em maior fosforilação de moléculas críticas de sinalização de células T e aumento da

ativação da p53, uma proteína reguladora chave do ciclo celular (SINGH *et al.*, 2021). Os CARs de *linker* curto, agregados e na presença do CD22, expressaram mais genes envolvidos na regulação da função efetora de células T, como *TNFSF4*, *CCR7*, *TCF7*, *KLF2* e *Eomes*, entre outros. Os autores relatam ainda que a sinalização autônoma do CAR curto não resultou em expressão de proteínas associadas à exaustão como também não levou à proliferação autônoma de células T (SINGH *et al.*, 2019; SINGH *et al.*, 2021). Todos esses resultados indicam que a sinalização autônoma do CAR de *linker* curto leva a um estado de responsividade aumentada, sugerindo que o CAR curto é mais apto para iniciar, rapidamente, a sinalização intracelular e ativar programas de resposta imune após o engajamento com o alvo (SINGH *et al.*, 2021).

Figura 2- Representações dos CARs anti-CD22 de *linker* curto e longo.



Fonte: adaptado de SINGH *et al.* (2021). (A) Representação da composição do CAR curto e longo. (B) Correlação entre estruturas dos CARs anti-CD22, constante de dissociação e tamanho do agrupamento. (C) Representação do CAR longo monomérico e CAR curto na forma de dímero. (D) e (E) Ilustração do resultado de microscopia confocal de ambas as construções de CAR.

Em seguida, Singh e colaboradores (2021) avaliaram o papel do domínio coestimulatório 41BB, na geração de funções efetoras aprimoradas, substituindo o domínio coestimulatório de 41BB pelo domínio CD28 em CAR anti-CD22 curto e longo, com novos ensaios de citotoxicidade *in vitro*. Nesse caso, o CAR de *linker* curto, foi menos eficaz que o de *linker* longo, constituído pelo domínio coestimulatório de CD28. Esse resultado obtido é consistente com vários relatos anteriores demonstrando que a sinalização tônica de CARs baseados em CD28 leva à disfunção de células T (FRIGAULT *et al.*, 2015; LYNN *et al.*, 2019); e sinalização celular divergente entre as células CAR-T contendo o domínio coestimulatório de CD28 e 41BB (SINGH *et al.*, 2021).

Por fim, com o objetivo de obter um CAR anti-CD22 de *linker* curto livre de propriedade intelectual, Singh e colaboradores (2021) propuseram um CAR usando scFv, desenvolvido através de triagem em uma biblioteca de *phage-display* e com posterior avaliação de atividade antileucêmica a CD22 *in vivo*. Em seguida, desenvolveram CARs com scFvs de *linker* curto e longo e realizaram novos testes de citotoxicidade *in vitro* e *in vivo*. Os testes de citotoxicidade demonstraram que os CARs de *linker* curto, contendo scFv de M971 e da biblioteca de *phage-display*, tiveram atividade antileucêmica superior às suas contrapartes longas, sendo que a eficácia foi semelhante entre si. Já o CAR de scFv de *linker* longo da biblioteca de *phage-display* obteve eficácia superior ao de M971 (SINGH *et al.*, 2021).

2.2 Anticorpos

2.2.1 Aspectos Gerais

Anticorpos, ou Imunoglobulinas (Ig) são moléculas de glicoproteína plasmáticas que possuem uma importante função para defesa do organismo pelo sistema imune. Inicialmente, em 1890, os anticorpos foram descobertos no soro de animais que eram tratados com doenças infecciosas. O prêmio Nobel de Medicina e Fisiologia de 1901, foi resultado do primeiro indício da utilização de anticorpos de origem não humana, promovido pelo médico alemão Emil Adolf Von Behring em conjunto com Shibasaburo Kitasato, esses pesquisadores utilizaram anticorpos não humanos para neutralizar a toxina diftéria, o que gerou uma grande mudança no pensamento científico daquela época (CORDEIRO *et*

al., 2014). Após esse feito, os anticorpos de origem não-humana foram amplamente utilizados em soroterapia contra muitas doenças bacterianas. No final da década de 1930 a soroterapia foi abandonada devido à sua toxicidade e a descoberta do antibiótico penicilina (HWANG *et al.*, 2005).

Existem dois principais grupos de anticorpos: monoclonais e policlonais. Os anticorpos policlonais reconhecem vários epítomos de um mesmo antígeno, desencadeando respostas diversas. Já os anticorpos monoclonais reconhecem um mesmo epítomo de um antígeno (ABBAS *et al.*, 2015), assim obtendo resultados mais previsíveis quando aplicado para fins de diagnósticos e tratamentos de doenças (SINGH *et al.*, 2018). A maior parte das aplicações de mAbs está relacionada ao diagnóstico e tratamento de doenças infecciosas, genéticas e o câncer (HWANG *et al.*, 2005).

Os anticorpos monoclonais de camundongos recebem a nomenclatura de murinos, e são reconhecidos pelo sistema imunológico humano como agentes antígenos, provocando uma resposta imunológica de anticorpos humanos contra anticorpos murinos. Entretanto, foi observado que a maior parte dos mAbs murinos quando administrados em humanos ocasionavam reações alérgicas e choques anafiláticos na maioria dos pacientes. Esse conjunto de efeitos adversos originárias do próprio sistema imunológico do indivíduo foi chamado de *Human Anti-Mouse Antibody* (HAMA), responsável também pela diminuição da eficácia do tratamento por afetar as propriedades farmacocinéticas do mAb murino. Ciente dessas problemáticas, foi então que durante a década de 80, diversos pesquisadores através de técnicas biomoleculares e de hibridomas objetivaram desenvolver mAbs mais semelhantes aos anticorpos totalmente humanos, para reduzir o potencial imunogênico dos mAbs murinos (HWANG *et al.*, 2005).

Atualmente, os anticorpos e seus derivados constituem cerca de 20% dos biofármacos em desenvolvimento pela indústria biotecnológica. Apesar dos efeitos adversos aos pacientes pelo uso de anticorpos murinos, ou seja, anticorpos não humanos, avanços na biologia molecular associados ao aprimoramento de tecnologias de hibridonas possibilitaram resultados promissores nos estudos envolvendo este produto (CORDEIRO *et al.*, 2014).

Como sendo componentes essenciais do sistema imunológico, os anticorpos são glicoproteínas que reconhecem outras macromoléculas chamadas de antígenos. Os anticorpos circundam pelo organismo até encontrarem e reconhecerem um

antígeno. Uma vez ligado, podem recrutar outros componentes do sistema imune para destruírem as células que contém o antígeno. As principais formas de mecanismo de ação dos anticorpos é através do bloqueio de receptores, fatores de crescimento essenciais à célula, lesão física e lesão química (ABBAS *et al.*, 2015).

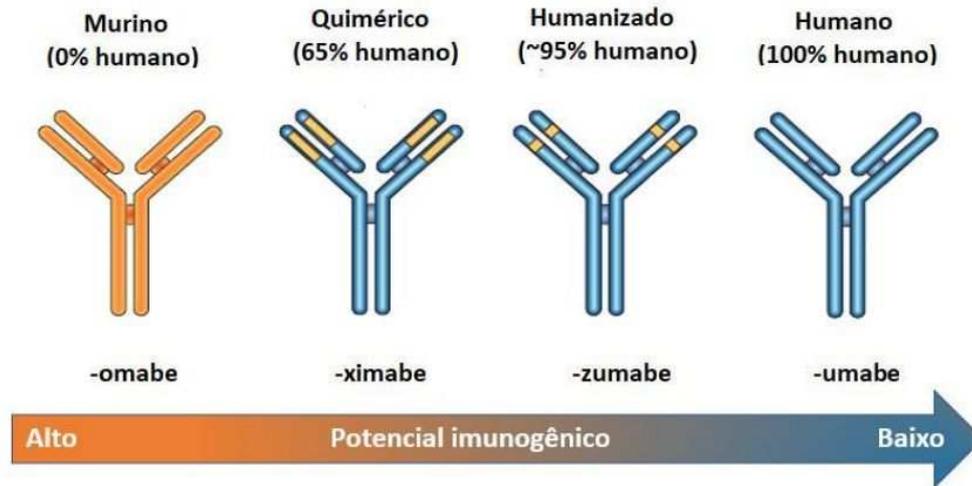
O sucesso do uso de mAbs contra o câncer se dá principalmente pela seleção do antígeno ideal. O antígeno ideal para anticorpos monoclonais deve ser altamente expresso por células tumorais e não por células normais; estável na superfície celular e estar funcionalmente envolvido no processo tumoral (SHORE, 2015).

Os principais mAbs conhecidos são das subclasses IgG (imunoglobulina tipo G) ou IgM (imunoglobulina tipo M), e podem ser utilizados isoladamente (não conjugados) ou em combinação (conjugados) com outros agentes terapêuticos (HWANG *et al.*, 2005). A aprovação de anticorpos monoclonais pela FDA para aplicação em oncologia é relativamente recente, sendo o primeiro mAb aprovado em 1997, Rituximabe. Desde então, vários outros anticorpos monoclonais foram aprovados para uso clínico em diferentes doenças oncológicas.

2.2.2 Desenvolvimento de MAbs murinos e humanizados

Os anticorpos monoclonais se dividem em 4 subclasses, de acordo com a sua estrutura e imunogenicidade sendo: murinos, quiméricos, humanizados e humanos, representados na Figura 3. De acordo com essa classificação, à medida que as regiões/fragmentos aminoácidos em determinadas posições murinas (em laranja) são substituídas pelas regiões humanas (em azul), o potencial imunogênico dos anticorpos vai diminuindo (WEINER, 2015).

Figura 3 - Classificação dos mAbs de acordo com a sua estrutura e imunogenicidade.



Fonte: MACEDO (2018).

Em 2002, o Adalimumab se tornou o primeiro mAb humano a ser aprovado pela FDA. Desde então, outros anticorpos monoclonais humanos receberam aprovação da FDA, a exemplo: Panitumumab, Golimumab, Canakinumab, Ustekinumab, Ofatumumab e Denosumab (NELSON *et al.*, 2010; FOLTZ *et al.*, 2013).

Com o auxílio do aprimoramento da engenharia genética na biologia molecular, foi possível obter anticorpos híbridos entre humanos e camundongos, gerando os primeiros anticorpos denominados quiméricos. O anticorpo quimérico apresenta expressão proteica do camundongo na sua região variável (VH e VL) e, expressão proteica humana na região constante onde ambas as junções proteicas estão unidas para formar um anticorpo pleno. Contudo, embora em menor grau, ainda se observaram reações imunológicas indesejadas denominadas *Human Anti-Chimeric Antibody* (HACA) (HWANG *et al.*, 2005).

Ainda com o objetivo de reduzir a imunogenicidade dos anticorpos murinos e quiméricos foram desenvolvidos os primeiros mAbs humanizados. O anticorpo monoclonal humanizado é originário de protocolos mais modernos de obtenção de mAbs. Em geral, o protocolo de obtenção desse tipo de mAb baseia-se no transplante de CDRs murinas para cadeias variáveis da estrutura da molécula de anticorpos humanos, como resultado, tem-se a obtenção de anticorpos monoclonais humanizados possuindo especificidade e afinidade ao alvo, mas com uma mínima porção murina em sua composição, reduzindo muito sua característica imunogênica devido ao fato deste anticorpo possuir apenas traços do anticorpo do camundongo

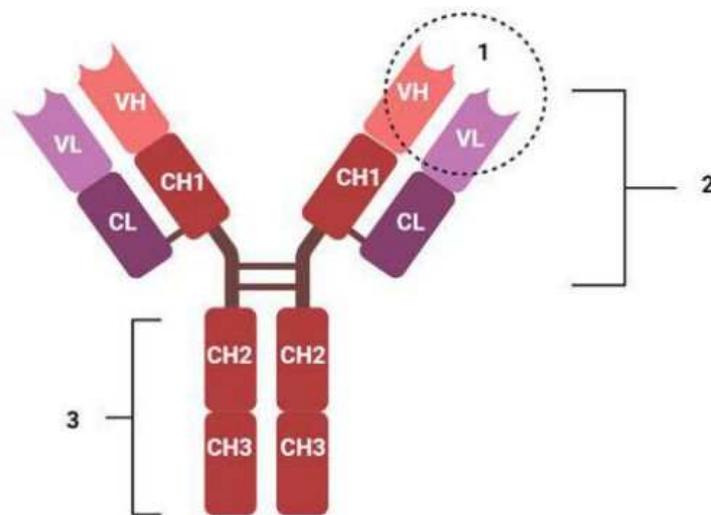
nas suas regiões hipervariáveis e o restante da molécula com características totalmente humana (SAFDARI *et al.*, 2013).

A partir da década de 2000, conforme exposto no trabalho de Nelson e colaboradores (2010), é observado um aumento substancial no estudo clínico de mAbs totalmente humanos e uma tendência para redução do uso de anticorpos quiméricos e humanizados; além da redução drástica no número de aparecimento de novos mAbs murinos para aplicação clínica, evidenciando que os mAbs totalmente humanos são uma crescente e promissora categoria de agentes terapêuticos.

2.2.3 Estrutura de Anticorpos

A Figura 4 mostra um anticorpo da classe IgG. Pode-se observar que todas as classes e subclasses de anticorpos conhecidas possuem uma estrutura básica em comum, que consiste em dois domínios de ligação ao antígeno, denominado de fragmento de ligação ao antígeno (Fab), e um sítio efetor, chamado de fragmento cristalizado (Fc) (ABBAS *et al.*, 2015).

Figura 4 - Representação estrutural de um anticorpo da classe IgG.



Fonte: elaborada pela autora (2024). Vermelho: Cadeia Pesada (VH: Domínio variável da cadeia pesada e CH: Domínio constante da cadeia pesada). Roxo: Cadeia Leve (VL: Domínio variável da cadeia leve e CL: Domínio constante da cadeia leve). Em que 1: Fragmento variável (Fv); 2: Fab e 3: Fc.

Estruturalmente, os anticorpos são constituídos por quatro cadeias unidas entre si por ligações dissulfeto e por ligações não covalentes: duas cadeias leves idênticas (isotipos κ ou λ); e, duas cadeias pesadas idênticas (isotipos γ , ϵ , δ , α ou μ). Como

apresentado na Figura 4, a cadeia leve é formada por dois domínios, chamados de domínio variável da cadeia leve (VL) e domínio constante da cadeia leve (CL), diferenciados pelo critério de variabilidade da sequência de aminoácidos. As cadeias pesadas são constituídas por um domínio variável da cadeia pesada (VH) e três ou quatro domínios constantes da cadeia pesada (CH1, CH2, CH3 e CH4), dependendo do isotipo da cadeia pesada. Já o fragmento variável (Fv) é constituído pelos domínios VH e VL, e contém o sítio de ligação ao antígeno (ALBERTS *et al.*, 2017).

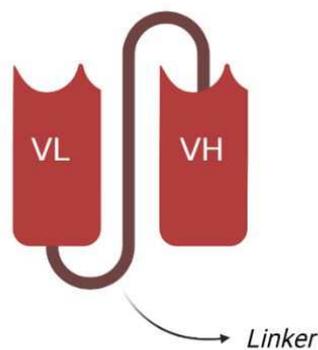
Ainda conforme a Figura 4, ao observar-se a estrutura básica de um anticorpo pode-se verificar que é uma molécula bifuncional, uma vez que contém uma porção composta pelos dois domínios variáveis, VL e VH, juntamente pelos domínios constantes CL e CH1, constituindo a região responsável pela ligação com o antígeno. A região de ligação ao antígeno é reconhecida de forma específica por meio de pequenas estruturas hipervariáveis presentes em ambos os domínios variáveis, conhecidas como Regiões Determinantes de Complementaridade (CDRs). Existem três CDRs no domínio variável da cadeia leve (L1, L2 e L3) como também três CDRs no domínio variável da cadeia pesada (H1, H2 e H3). Já a região composta pelo restante dos domínios constantes integra o chamado Fragmento Cristalizável (Fc), o qual é capaz de exercer funções efetoras (HOLLIGER, HUDSON, 2005; NELSON, REICHERT, 2009).

Os avanços obtidos na área de engenharia de proteínas, mais especificamente na área de engenharia de anticorpos, ocorreram a partir do progresso das técnicas de digestão proteolítica de imunoglobulinas, com ou sem redução das ligações dissulfeto que unem as cadeias peptídicas; como também a partir da evolução da área de expressão heteróloga de anticorpos (LERNER, 2016). O aperfeiçoamento da área de engenharia de imunoglobulinas também possibilitou compreender a funcionalidade dos fragmentos de anticorpos, não sendo necessário utilizar o anticorpo em sua totalidade para estudar o seu mecanismo de reconhecimento pelo antígeno (SAEED *et al.*, 2017; KENNEDY *et al.*, 2018).

A maior parte dos fragmentos de anticorpos conhecidos e aplicados em pesquisas na área de engenharia de imunoglobulinas são originários de IgG, pois são mais abundantes nos seres humanos (HOLLIGER; HUDSON, 2005). Conforme a Figura 5, dentre os diversos tipos de fragmentos de anticorpos que se podem obter destaca-se a porção Fab, podendo ser monovalente, bivalente e trivalente, reconhecendo um, dois e três antígenos distintos, respectivamente; e o scFv, cuja

composição contempla apenas os domínios variáveis tanto da cadeia pesada (VH), como da cadeia leve (VL), unidos por um pequeno *linker* peptídico bastante flexível (AHMAD *et al.*, 2012), que previne a quebra dos domínios (MOHAMMADI; NEJATOLLAHI, 2014). Sabe-se ainda que o scFv pode formar dímeros e trímeros, dando origem aos diacorpos e triacorpos, respectivamente (HOLLIGER; HUDSON, 2005; MUYLDERMANS; SMIDER, 2016).

Figura 5 - Representação estrutural de um scFv.



Fonte: elaborada pela autora (2024). Identificação da união dos domínios VL e VH de um anticorpo por uma cadeia polipeptídica flexível denominada de *linker*.

Geralmente, os *linkers* dos scFvs são constituídos de 10 a 25 resíduos que frequentemente possuem características hidrofílicas. Esses resíduos hidrofílicos nesse polipeptídeo flexível (*linker*) pode proporcionar o aumento da solubilidade do scFv durante a expressão heteróloga (SHEN *et al.*, 2009). Existem dois principais tipos de *linker*, um composto somente por resíduos de glicina e serina, (G4S)₃ (HUSTON *et al.*, 1988), e outro denominado de *Whitlow Linker* (WHITLOW *et al.*, 1993), que possui além dos resíduos de glicina e serina outros tipos de aminoácidos como lisina, glutamato, prolina e treonina (Quadro 1).

Existem outros tipos de *linker*, menos comuns, constituídos por 5 e 20 resíduos, respectivamente. Um exemplo são os *linkers* entre os domínios VH e VL que compõe o scFv curto e longo de M971 (SINGH *et al.*, 2021), conforme o Quadro 1.

Quadro 1 - Resíduos que fazem parte dos tipos de linker utilizados em construções de scFvs.

Tipo de Linker	Sequência
(G4S) ₃	GGGGSGGGGSGGGGS
<i>Whitlow Linker</i>	GSTSGSGKPGSGEGSTKG
<i>Linker curto</i>	GGGGS
<i>Linker longo</i>	GGGGSGGGGSGGGGSGGGGS

Fonte: elaborada pela autora (2024).

Em geral, o *Linker* de (G4S)₃ é mais utilizado para as construções de scFvs, uma vez que as cadeias laterais dos resíduos de glicina e serina interferem menos na estrutura dos domínios VH e VL (HUSTON *et al.*, 1991). Entretanto, os resíduos diferentes de glicina e serina no *Whitlow Linker* proporcionam a este tipo de *linker* uma maior resistência à proteólise *in vitro* (WHITLOW *et al.*, 1993). Já o *linker* curto favorece a formação de multímeros de scFv (SHEN *et al.*, 2009). Sabe-se ainda que o *linker* longo entre os domínios VH e VL forma scFvs funcionais e preferencialmente na forma estrutural monomérica (ARSLAN *et al.*, 2022).

Em geral os *linkers* não possuem estrutura secundária bem definida (AHMAD *et al.*, 2012), já que os principais resíduos de aminoácidos que os constituem podem assumir variados valores de ângulos diedros ϕ (*phi*) e ψ (*psi*), assim apresentando diversas conformações. Portanto, sendo necessário utilizar ferramentas de bioinformática, como protocolos de modelagem molecular, para a obtenção da sua estrutura tridimensional.

As principais vantagens do uso de scFvs em comparação ao anticorpo completo é devido ao seu tamanho. O menor tamanho dos scFvs promove uma diminuição na meia vida destas moléculas, visto que são eliminadas facilmente pelos rins, pois estes conseguem metabolizar compostos de até 55 kDa (HOLLIGER; HUDSON, 2005). Sabe-se ainda que os scFvs são ótimos para atingir alvos, pois possuem alta penetrância nos tecidos pouco vascularizados e, portanto, surgem como alternativa à terapia em tumores sólidos (YOKOTA *et al.*, 1992), além de serem menos imunogênicos que os anticorpos monoclonais completos (BEGENT *et al.*, 1996).

Uma outra vantagem dos scFvs em relação aos anticorpos completos deve-se ao fato de que uma vez que são moléculas pequenas e desprovidas de glicosilação podem ser expressas em sistemas mais simples, como os sistemas bacterianos. Sendo assim, a obtenção de scFvs ao invés de anticorpos monoclonais completos representa uma das diversas alternativas para baratear os custos de produção de

moléculas terapêuticas (PONTES, 2019).

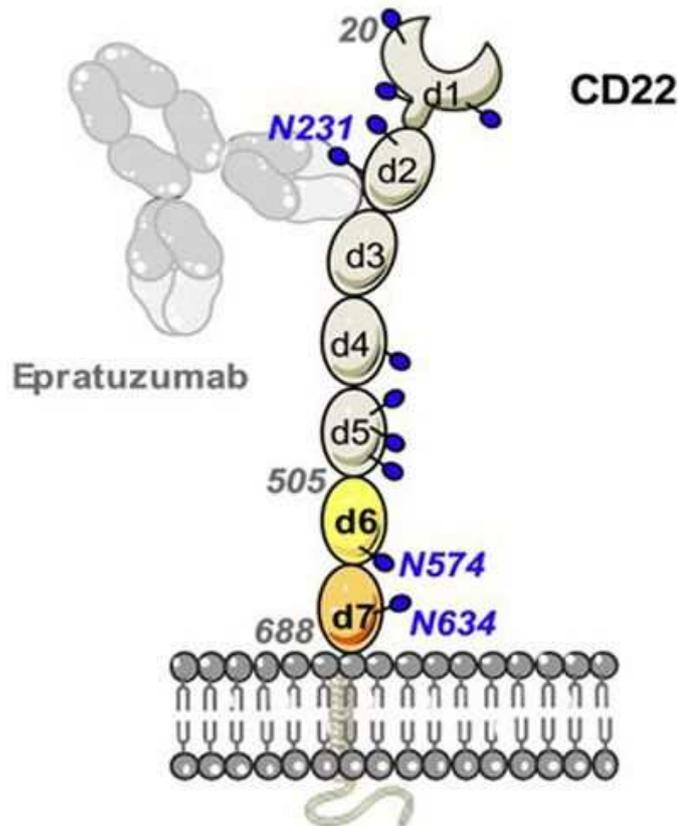
Uma desvantagem do scFv em relação aos anticorpos completos é que não possuem função efetora, uma vez que não apresentam o domínio Fc em sua constituição. Um exemplo de função efetora realizada pelos anticorpos é a sua ligação a vários tipos celulares, como células fagocitárias, linfócitos, plaquetas, células master, e basófilos que possuem receptores que se ligam a imunoglobulinas (HOLLIGER; HUDSON, 2005).

Apesar de não apresentarem função efetora, a obtenção de scFvs, através de técnicas de engenharia de proteínas é de extrema importância no desenvolvimento de novos biofármacos, principalmente no aprimoramento da terapia de células CAR-T, para a remissão de diversas malignidades, como as neoplasias malignas (MUÑOZ-LÓPEZ *et al.*, 2022). Exemplos de scFvs para aplicação em biofármacos são os derivados do mAb humano M971, proposto no CAR de *linker* curto e longo de Fry e colaboradores (2018) e Ruella e colaboradores (2017), respectivamente.

2.3 Anticorpo M971

O M971 é um anticorpo monoclonal humano anti-CD22 capaz de proporcionar a apoptose em células B leucêmicas. O mecanismo de interação do M971 com o antígeno CD22 difere de outros mAbs, pois o M971 interage com o epítipo próximo à membrana, localizado no domínio extracelular D7 da molécula CD22, como observado na Figura 6. Além disso, diferentemente do mAb de Epratuzumab, um anticorpo anti-CD22, o epítipo do CD22 reconhecido pelo M971 não inclui glicanos (Figura 6). Essas características tornam o M971 um candidato promissor para aplicação em CARs anti-CD22 e CARs Bi específicos anti-CD19/CD22 (EREÑO-ORBEA *et al.*, 2021).

Figura 6 - Representação esquemática da glicoproteína de membrana CD22 na superfície celular interagindo com anticorpo terapêutico Epratuzumab.



Fonte: EREÑO-ORBEA *et al.* (2021). O CD22 é uma glicoproteína de membrana com 12 glicanos ligados a N (esferas azuis) nos domínios extracelulares D1, D2, D4, D5, D6 e D7. Os domínios D1-D3 compreendem a região de interação com o anticorpo Epratuzumab e os domínios D6-D7 contém a região de interação com o mAb M971.

Outros anticorpos anti-CD22 que também já passaram por ensaios *in vitro* foram os Inotuzumab, ozogamicina e BL22. Esses anticorpos, assim como o Epratuzumab, possuem como característica ligarem-se às regiões N terminais de CD22. Essa área de ligação no CD22, para esses anticorpos, está próxima à região através da qual o CD22 interage com seus ligantes, como glicoproteínas terminadas em ácido siálico na superfície de linfócitos, neutrófilos, monócitos e eritrócitos (ENGEL *et al.*, 1995); e uma infinidade de outros receptores de superfície celular como o CD45. Devido aos complexos esquemas de interação em diferentes linhagens celulares, essa região no CD22 pode adotar diferentes conformações e, assim, expor vários epítomos em diferentes superfícies celulares. Isso pode levar à ligação inconsistente por esses anticorpos terapêuticos (XIAO *et al.*, 2009).

Os anticorpos anti-CD22 que se ligam a região mais proximal à membrana

possuem a vantagem de que essa região no CD22 não está vinculada em processos de *splicing* alternativo, ou seja, não há deleção dessa região em diferentes linhagens de CD22; e não está envolvida em múltiplas interações físicas e funcionais com outras macromoléculas biológicas. Sendo assim, anticorpos direcionados a essa região, como o anticorpo M971, podem exibir reconhecimento específico mais consistente a CD22 de várias linhagens celulares ou tecidos (XIAO *et al.*, 2009).

Nesse trabalho, as CDRs do anticorpo M971 foram determinadas pelo método de Kabat submetendo as sequências da porção VL (7O52-L) e VH (7O52-H) ao programa AbRSA (LI *et al.*, 2019) (Quadro 2).

Quadro 2 - CDRs dos domínios VH e VL do M971 determinadas pelo método de Kabat utilizando o programa AbRSA.

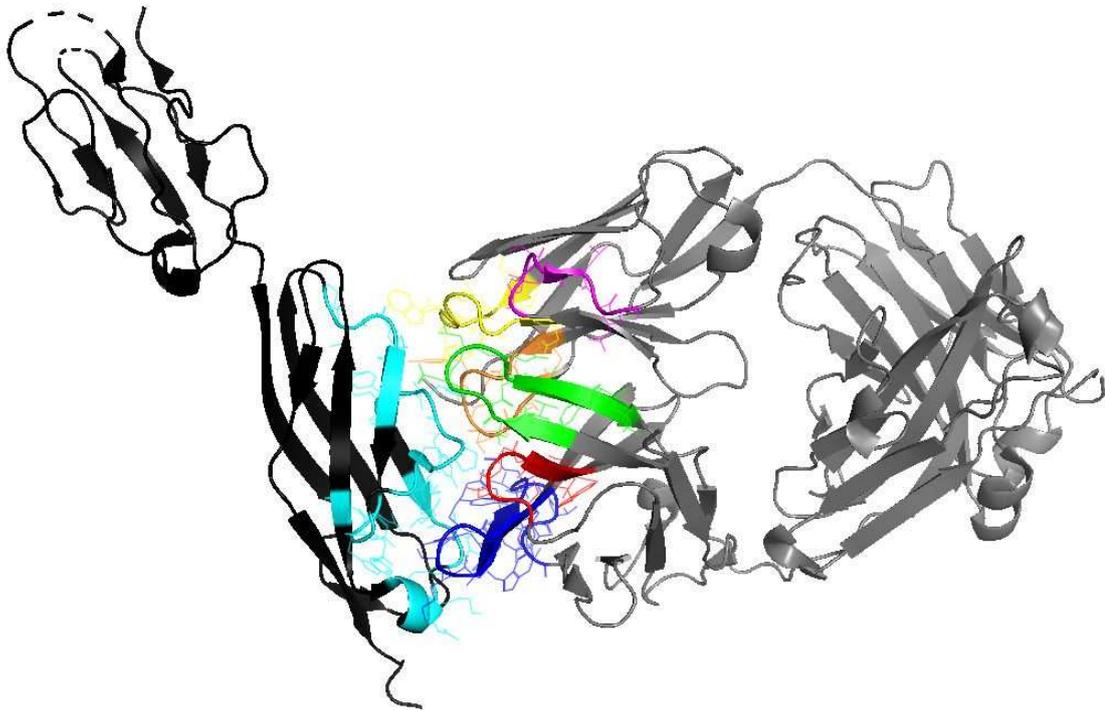
Domínio	CDR1	CDR2	CDR3
VH	SNSAAWN	RTYYRSKWYNDYAVSVKS	EVTGDLEDAFDI
VL	RASQTIWSYLN	AASSLQS	QQSYSIPQT

Fonte: elaborada pela autora (2024).

No trabalho de Ereño-Orbea e colaboradores (2021) foi obtida a estrutura cristalográfica do complexo Fab-CD22, (cód. PDB 7O52). Com base nesse trabalho foram determinados os resíduos-chave de CD22 para M971 sendo descritos como esses resíduos os: P623, V624, S625, H626, Y627, T628, D631, W632, N633, Q635, S636, L637, P638, Y639, H640, S641, Q642, R645, L646, E647, P648, K650, Q652, H653 e Y657. É importante destacar que esses resíduos são regiões potenciais de ligação, portanto, não sendo definidos como epítipo de CD22 para M971.

A Figura 7 mostra uma representação da interface entre as regiões CDR do anticorpo M971, de acordo com as cores já definidas no Quadro 2, e os resíduos-chave de CD22 para M971. A área de superfície de contato é extensa (990 Å²) e é mediada principalmente pela cadeia pesada (722 Å²) e menos pela cadeia leve (268 Å²). Os resíduos do M971 que fazem parte da área de superfície de contato já foram descritos por Ereño-Orbea e colaboradores (2021), conforme o Quadro 3.

Figura 7 - Representação da interface de interação Fab-CD22.



Fonte: elaborada pela autora (2024). Interface Fab-CD22, destacando em cores e linhas as regiões CDR da porção Fab do anticorpo M971 e os resíduos-chave do CD22. Em ciano e linhas, os resíduos-chave do CD22 para M971.

Quadro 3 - Em negrito e sublinhado os resíduos do M971 que constituem a área de superfície de contato descritos por Ereño-Orbea e colaboradores (2021).

Domínio	CDR1	CDR2	CDR3
VH	<u>S</u>NSA<u>A</u>WN	<u>R</u>TY<u>Y</u>R<u>S</u>K<u>W</u>Y<u>N</u>D<u>Y</u>AV<u>S</u>V<u>K</u>S	<u>E</u>VT<u>G</u>D<u>L</u>E<u>D</u>A<u>F</u>D<u>I</u>
VL	RASQT <u>I</u> <u>W</u> <u>S</u> <u>Y</u> <u>L</u> N	AASSLQS	QQ <u>S</u> <u>Y</u> S <u>I</u> P <u>Q</u> T

Fonte: elaborada pela autora (2024).

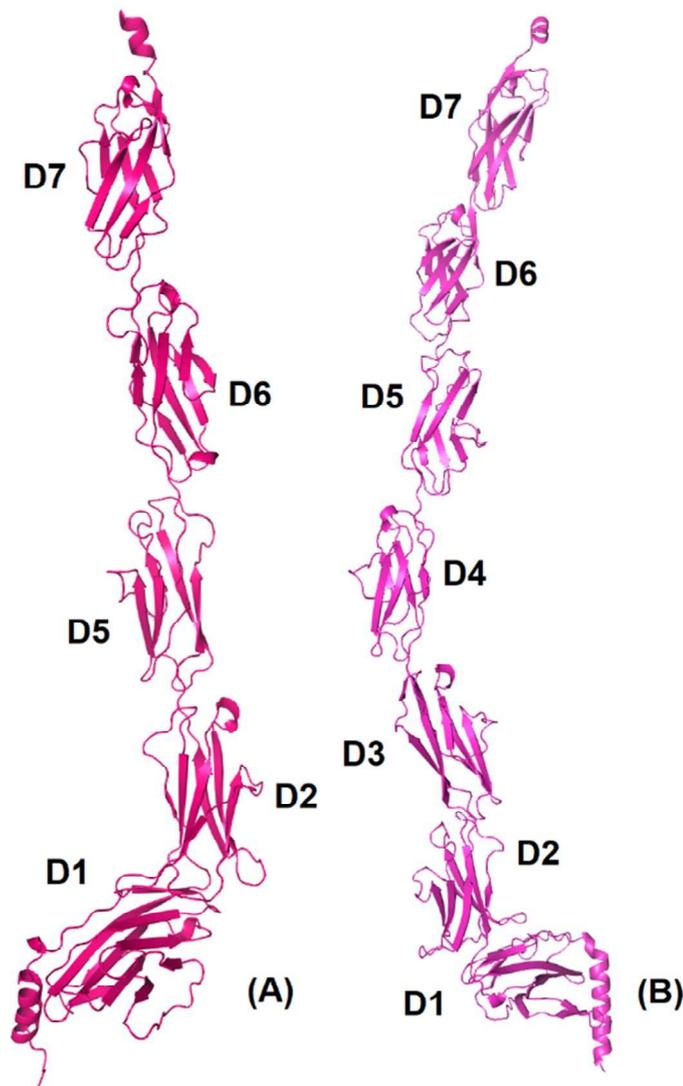
2.4 CD22: estrutura e função

O CD22 é uma glicoproteína transmembrana, com massa molecular de ~140 kDa, presente na superfície de células B. É comumente aceito que sua expressão é restrita em células B. Durante o desenvolvimento das células B, o CD22 está presente nas células pró-B e pré-B, mas nessas fases sendo restrito ao citoplasma. Nas células B maduras, o CD22 é expresso na superfície, entretanto, eventualmente, essa expressão é perdida quando as células B se diferenciam em células plasmáticas. Nos tecidos linfoides, o CD22 é expresso no manto folicular e nas células B da zona marginal, mas apenas fracamente nas células B germinativas, sendo por isso

considerado um alvo atrativo para a depleção de células B, em casos de doenças como a LLA e o Linfoma Difuso de Grandes Células B (LDGCB) (TERSTAPPEN *et al.*, 1990; BAGDONAITE *et al.*, 2015; EREÑO-ORBEA *et al.*, 2021).

Em humanos, a molécula CD22 apresenta duas isomorfias (BRAUTIGAN *et al.*, 1990; WILSON *et al.*, 1991) (Figura 8A e 8B). A isoforma maior, CD22 β , que contém sete domínios extracelulares semelhantes a Ig, dos quais todos, exceto um, são do tipo C. A única exceção é o domínio N-terminal, que é do tipo V (STAMENKOVIC *et al.*, 1991; EREÑO-ORBEA *et al.*, 2021). Além disso, essa isoforma possui um domínio citoplasmático contendo 141 resíduos. Já a isoforma menor, CD22 α , possui cinco domínios extracelulares semelhantes a Ig, domínios 3 e 4 são excluídos, e uma região citoplasmática, a qual dispõe de 23 resíduos a menos do que o domínio citoplasmático do CD22 β . A forma CD22 β é isoforma predominante e é encontrada em humanos e única forma encontrada em murinos e camundongos (STAMENKOVIC *et al.*, 1991; CHAN *et al.*, 1998). Além disso, o CD22 β é encontrado, predominantemente, na forma monomérica, entretanto, também pode ocorrer na forma de heterodímero $\alpha\beta$ (STAMENKOVIC *et al.*, 1991; SCHWARTZ-ALBIEZ *et al.*, 1991). Ambas as isoformas apresentam tirosinas em suas caudas citoplasmáticas, 3 resíduos na forma isomorfa α e 6 na forma isomorfa β (CHAN *et al.*, 1998). A maioria dessas tirosinas são arranjadas para formar motivos de inibição baseados em tirosina de imunorreceptores (ITIM), consistindo em um trecho de 6 resíduos com a sequência consenso (I/L/V)XYXX(L/V) e/ou potenciais motivos de ativação de base de tirosina de imunorreceptores (ITAM) (RETH, 1989; WILSON *et al.*, 1991; DOODY *et al.*, 1995; VÉLY; VIVIER, 1997). Assim, o CD22 tem a capacidade de regular a inibição das células B por meio de uma ampla gama de moléculas de sinalização dependentes de tirosina intracelulares (CHAN *et al.*, 1998).

Figura 8 (A) e (B) – Representação das estruturas de CD22 α e CD22 β , em *cartoon*, realizadas no Pymol, modeladas no programa de modelagem molecular Robetta.



Fonte: elaborada pela autora (2024). (A) Domínios D1-D7, excluídos D3 e D4, da isoforma CD22 α . (B) Domínios D1-D7 da isoforma CD22 β .

Estudos indicam que ainda não está bem compreendido a razão pela qual existe a diferença de conformação entre as isoformas CD22 α e CD22 β de humanos. Acredita-se que as proteínas chaperonas no citoplasma se ligam a polipeptídeos imaturos ou desdobrados e garantem o dobramento ou transporte correto e, portanto, sugere-se que tenham um papel importante na conformação tridimensional de proteínas de múltiplos domínios (FRYDMAN, 2001; TOBA *et al.*, 2002).

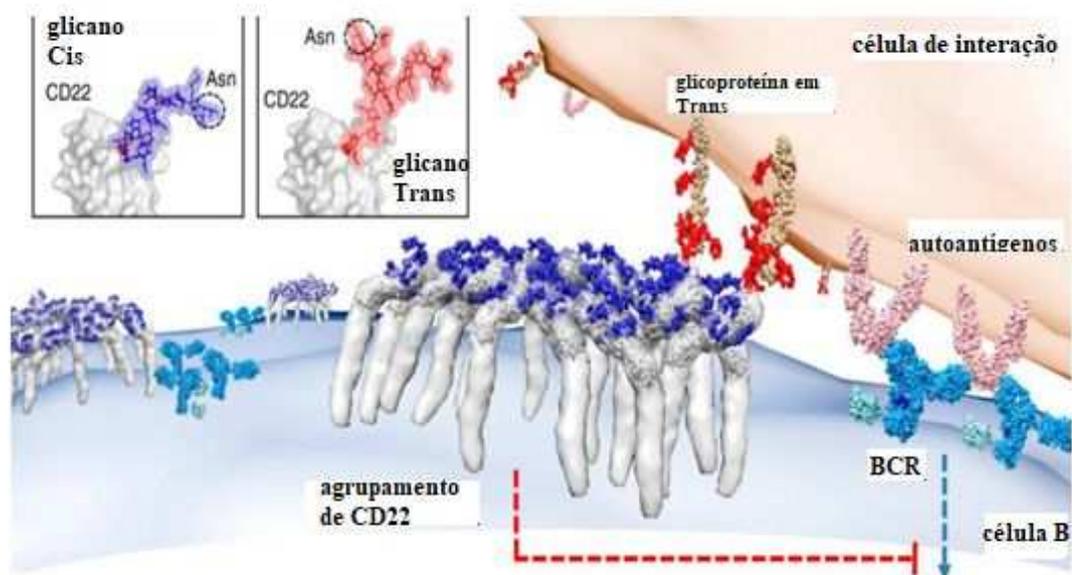
Ainda com relação a estrutura e função do CD22, essa molécula pode conter ou não as árvores de glicosilações terminadas em ácido siálico α 2,6, assim como o IgM de superfície (sIgM) e o CD45. Dessa forma, o CD22 pode se associar a si mesmo

ou a outras moléculas da superfície celular nas células B em uma configuração “cis” ou com ligantes em outras células em uma configuração “trans”. As interações “cis” CD22-CD22 endógenas podem “mascarar” o CD22, limitando sua capacidade de ligação a ligantes em “trans”. Como nem todo CD22 contém as árvores de glicosilações terminadas em ácido siálico $\alpha 2,6$, ele também é encontrado nas células B em uma forma “desmascarada” livre de ligante (HAN *et al.*, 2005; GASPARRINI *et al.*, 2016; CLARK; GILTIAY, 2018). O significado desse desmascaramento de CD22 ainda não é totalmente compreendido, mas acredita-se que promove a interação de CD22 com ligantes em trans, ou com um subconjunto distinto de ligantes cis, talvez incluindo o receptor do antígeno de célula B (BCR) (WALKER; SMITH, 2008).

Uma questão em aberto e importante é como as interações CD22 em “cis” ou “trans” afetam a sinalização com BCR. O trabalho de Gasparrini e colaboradores (2016) sugere que o CD22 existe na forma de *nanoclusters* pré-formados na superfície da membrana de células B virgens primárias e que a organização do CD22 depende da sua capacidade de se ligar ao ácido siálico ($\alpha 2,6$ -sialilado) e é influenciada pela presença da glicoproteína CD45. Ainda de acordo com o trabalho desses pesquisadores, estruturalmente as interações CD22 em “cis” mantêm CD22 dentro de *nanoclusters*, juntamente com CD45; e longe do BCR, sendo assim, o enfraquecimento das interações CD22-ligante aumenta a associação CD22-BCR e amortece sinalização BCR (GASPARRINI *et al.*, 2016; ENTERINA *et al.*, 2022).

Outro estudo que investiga as interações “cis” e “trans” de CD22 foi realizado por Meyer e colaboradores (2018), de acordo com trabalho desses pesquisadores, quando as células B estão em repouso, a conformação do BCR está fechada e o CD22 forma homoligômeros em “cis”, longe do BCR, conforme a Figura 9. A ligação do BCR a autoantígenos induz a sua abertura conformacional, seguida de ativação e fosforilação de ITAMs do complexo $Ig\alpha/Ig\beta$. Além disso, *clusters* CD22 são recrutados para o BCR. Após a ativação do BCR, o recrutamento de CD22 inibe a sinalização do BCR. Ainda conforme a Figura 9, o CD22 também pode ser recrutado para o BCR ligando seus ligantes em outras células por interações “trans”. Devido a aproximação do CD22 ao BCR, os motivos ITIMs do CD22 são fosforilados, resultando na sinapse imune e inibição da célula B.

Figura 9 - Organização celular do CD22 e interações “cis” e “trans” com glicoproteínas de células B e de outras células.



Fonte: adaptado de EREÑO-ORBEA *et al.* (2017). Modelo explicativo para formação de *nanoclusters* em “cis” de CD22 e inibição da sinalização de BCR e da célula B. Os BCRs são mostrados ligando-se a autoantígenos (rosa) em uma célula em interação (superfície salmão). A seta azul representa a ativação de BCR, enquanto a linha tracejada vermelha representa a inibição de BCR mediada por CD22.

2.5 Padrões de Glicosilação

A glicosilação de proteínas é uma das modificações pós-traducionais que consiste na adição de cadeias de oligossacarídeo de cadeia linear ou ramificada. As glicoproteínas resultantes são geralmente glicoproteínas de superfície celular ou glicoproteínas da via secretora (NELSON; COX, 2018).

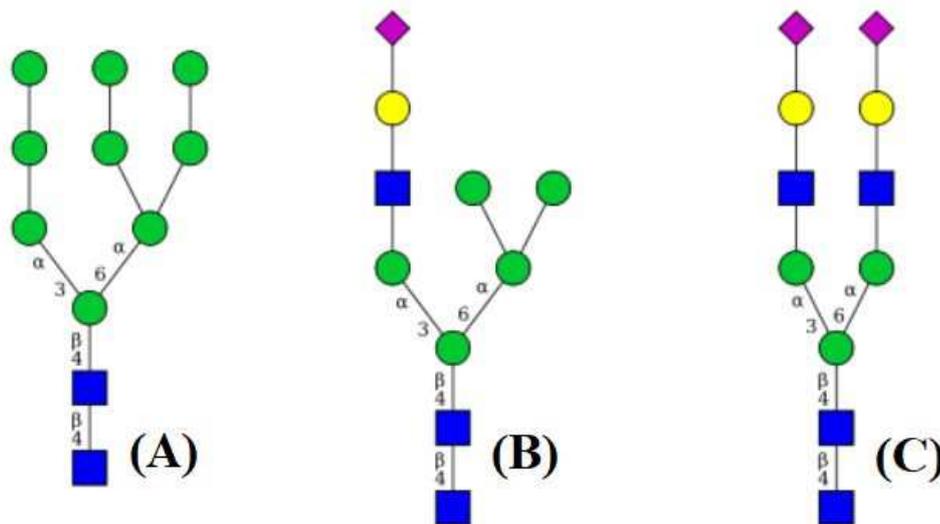
Dependendo do local de ligação do oligossacarídeo à proteína, a glicosilação pode ser classificada em 4 tipos:

- N-glicosilação: É o mais comum de todos e ocorre quando os oligossacarídeos se ligam ao nitrogênio do grupo amida dos resíduos de asparagina no motivo Asn-X-Ser / Thr, onde X pode ser qualquer aminoácido exceto prolina.
- O-glicosilação: Quando os carboidratos se ligam ao grupo hidroxila da serina, treonina, hidroxilisina ou tirosina. É uma modificação menos comum e exemplos são proteínas como colágeno, glicoforina e mucinas.
- C-manosilação: Consiste na adição de um resíduo de manose que se liga à proteína por uma ligação Carbono-Carbono com o átomo C2 do grupo indol nos resíduos de triptofano.

- Glicoproteína: Um polissacarídeo atua como uma ponte para ligar uma proteína a uma âncora de glicosilfosfatidilinositol (GPI) na membrana.

A N-glicosilação é a classificação de glicosilação que tem sido estudada em mais detalhes nos últimos anos (STOWELL *et al.*, 2015). Conforme a Figura 10, em geral, existem três tipos de N-glicosilação, sendo elas: (A) ricas em manose, (B) híbrida e (C) complexa.

Figura 10 - Tipos de N-glicosilação mais comuns em seres eucarióticos.



Fonte: elaborada pela autora (2024). : (A) ricas em manose, (B) híbrida e (C) complexa. Em azul (N-acetil-D-glucosamina), em verde (D-manose), em amarelo (D-galactose) e em roxo (N-acetil-D-neuramínico ácido). A: ligação do tipo alfa (α) glicosídica. B: ligação do tipo beta (β) glicosídica.

As N-glicosilações ligadas a glicoproteínas servem para diversas funcionalidades. Por exemplo, algumas proteínas requerem dessa modificação pós-traducional para conseguir dobrar adequadamente sua estrutura (SPIRO, 2002). Por fim, existe um interesse crescente em verificar se mutações ou alterações nos processos de glicosilação podem estar relacionados à alteração do microambiente das células tumorais e, assim, promover a progressão do desenvolvimento de tumores e metástases em pacientes com neoplasias malignas (STOWELL *et al.*, 2015).

2.6 Glicosilações em CD22

A glicoproteína CD22 é um tipo de Siglecs (lectinas semelhantes a

imunoglobulinas de ligação ao ácido siálico). Os Siglecs são membros da superfamília das imunoglobulinas (Ig), que reconhecem especificamente ácidos siálicos ligados a regiões terminais de glicoconjugados da superfície celular (CROCKER *et al.*, 2007).

Dois grupos de proteínas Siglec foram identificadas. O primeiro grupo inclui Siglecs que são bem conservados em mamíferos e consistem em sialoadesina (Siglec-1), CD22 (Siglec-2), associada à mielina glicoproteína (MAG) (Siglec-4) e Siglec-15. O segundo grupo de Siglecs é chamado de Siglecs relacionados ao receptor transmembrana expresso nas células de linhagem mielóide, o CD33. Este subconjunto compreende 10 membros em humanos, entre eles Siglec-10, mas apenas cinco membros em camundongos, incluindo Siglec-G. Com exceção do MAG, que é encontrado nas células da glia do sistema nervoso, todos os Siglecs são expressos nas células do sistema imunológico (KELM *et al.*, 1994).

Schwartz-Albiez e colaboradores (1991) realizaram uma série de estudos experimentais com o objetivo de investigar mais detalhes estruturais das glicosilações na molécula de CD22. Em conclusão, o trabalho desses pesquisadores sugere que a maioria das cadeias oligossacarídeos N-ligadas a CD22 são do tipo complexo com uma galactose terminal ou uma galactose terminal ligada ao ácido siálico ligado a(2-6) e uma fucose ligada a acetilglucosamina.

2.7 Simulação de Dinâmica Molecular

Para compreender como as moléculas desempenham a sua correta função, torna-se necessário obter um melhor entendimento acerca das suas interações com outras moléculas e com o solvente. Atualmente, já tem-se o conhecimento de técnicas tradicionais de caráter experimental que auxiliam na obtenção de informações estruturais de biomacromoléculas, tais como: cristalografia de raios-X, ressonância magnética nuclear, espalhamento de nêutrons ou espectroscopia de fluorescência (VAN GUNSTEREN *et al.*, 2008).

A simulação de Dinâmica Molecular (DM) é um técnica computacional capaz de estudar movimento das moléculas em solução. Sendo assim, as interações entre átomos e moléculas são empiricamente descritas como uma função do potencial, da qual derivam as forças que atuam sobre cada partícula, conforme a Equação 1. Após o conhecimento das forças que atuam no sistema é possível calcular a trajetória de um conjunto de átomos em simulação utilizando as equações de movimento de

Newton (FRENKEL; SMIT, 2002).

$$\vec{F}_i = -\frac{dV(r_1, r_2, r_3, \dots, r_N)}{dr_i} \quad (1)$$

Na simulação de DM através das equações de potencial que utilizam os diferentes parâmetros do campo de força é possível conhecer a energia potencial total, V , do sistema (Equação 2) a partir das coordenadas atômicas do sistema. A função da energia potencial é descrita pela energia de interação dos átomos ligados (forças de ligação, ângulos de ligação, torsionais e diedrais) e não ligados (interações de curto e longo alcance) (ABRAHAM *et al.*, 2013).

$$V = \frac{1}{2} \sum_{\text{ligações}} K_b (b_{eq} - b)^2 + \frac{1}{2} \sum_{\text{ângulos}} K_\theta (\theta_{eq} - \theta)^2 + \frac{1}{2} \sum_{\text{torsões}} K_\varphi [1 + \cos(n\varphi)] + \sum V_{\text{não-ligados}} \quad (2)$$

Na equação 2, tem-se o potencial total do sistema e as constituições de potenciais entre átomos ligados e não ligados. Em um determinado momento (t), b é o comprimento de ligação em um instante qualquer (instantâneo) que representa o desvio em relação ao comprimento de equilíbrio b_{eq} da ligação, assim como θ é o ângulo instantâneo entre duas ligações químicas consecutivas que descreve o desvio em relação ao ângulo de equilíbrio θ_{eq} entre as duas ligações químicas. As distâncias de ligações e as amplitudes dos ângulos podem oscilar, considerando um modelo de oscilador harmônico, entre as posições de deformação máxima e a posição de equilíbrio. Assim, K_b é a constante de Hooke associada à deformação da ligação química específica, e K_θ é a constante de Hooke para a restituição do ângulo de equilíbrio (θ_{eq}) entre as duas ligações químicas, φ representa o ângulo de torção e n é a multiplicidade do ângulo quando esta passa por 360° (ALLEN; TILDESLEY, 2017).

Em relação a energia potencial dos átomos não ligados (Equação 3), esta é

representada pela soma do potencial de Lennard-Jones (V_{L-J}) (Equação 4), que modela o potencial de *van der Waals* (vdW), responsável por interações de curto alcance, com o potencial de Coulomb (Equação 5), que modela as interações eletrostáticas (V_{elet}), consideradas de longo alcance.

$$V_{\text{n\~{a}o-ligados}} = V_{L-J} + V_{elet} \quad (3)$$

$$V_{L-J} = \sum 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (4)$$

$$V_{elet} = \sum \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \quad (5)$$

Na Equação 3, V_{L-J} é referente às interações de vdW, σ_{ij} corresponde a distância finita na qual o potencial interatômico entre os átomos i e j é zero e E_{ij} representa a profundidade do potencial entre a barreira repulsiva e atrativa entre os átomos. Na Equação 5, V_{elet} , é referente às interações eletrostáticas, onde r_{ij} corresponde à distância entre os átomos i e j com as cargas, q_i e q_j pontuais, ϵ_r é a constante dielétrica relativa do meio e ϵ_0 a permissividade do espaço livre.

Conhecendo-se a força (\mathbf{F}_i) que atua sobre determinado átomo (i) de massa conhecida (m_i), se faz possível calcular a aceleração desse átomo (\mathbf{a}_i) no sistema através da equação de movimento de Newton (Equação 6) (ABRAHAM *et al.*, 2015).

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} \quad (6)$$

Determinado o valor da aceleração, a integração das equações de movimento através de algoritmos como o de Verlet (VERLET, 1967) fornece a mudança de posição \mathbf{r} de cada átomo i . Na equação 7, tem-se as posições e acelerações dos átomos no tempo t , as quais são utilizadas para calcular $\mathbf{r}(t - dt)$, posição no tempo anterior e $(t + dt)$, a nova posição, sendo dt o intervalo temporal para evoluir o sistema.

$$r(t + dt) = 2r(t) - r(t - dt) + a(t)dt^2 \quad (7)$$

Sendo assim, conhecendo-se as posições iniciais para cada átomo que constitui o objeto de estudo da DM, as novas posições serão calculadas após intervalos de tempo (dt), até que a trajetória dos átomos ao longo do tempo total estimado para a simulação seja obtida. Após concluída a simulação de DM, as posições atômicas são gravadas em arquivos de trajetória para posterior análise de determinantes estruturais e energéticos de interesse.

Para se produzir a trajetória da DM algumas etapas devem ser realizadas. Em um primeiro momento, é necessário obter as posições atômicas iniciais das moléculas de interesse que irão compor a caixa de simulação, o sistema a ser simulado. As posições atômicas iniciais podem ser obtidas a partir de dados cristalográficos do banco de dados *Protein Data Bank* (PDB) ou por meio da modelagem das moléculas, utilizando-se de programas de modelagem por homologia, por exemplo (BERMAN *et al.*, 2000). Depois da obtenção das posições atômicas iniciais, é realizado o cálculo das forças exercidas sobre cada átomo do sistema, utilizando as informações de parâmetros do campo de força. Após esse processo, é fundamental realizar a minimização de energia do sistema otimizando a geometria da molécula e minimizando contatos de forma a evitar possíveis conflitos estereoquímicos. Em seguida, tem-se a etapa de termalização, em que a estrutura e a dinâmica do sistema são ajustadas para se obter o estado termodinâmico de interesse. A depender do tipo de estudo, essa etapa pode ser controlada em *ensemble canônico* (NVT), representado por número de moléculas e volume fixos e temperatura controlada ou em *ensemble isotérmico-isobárico* (NPT) (ALLEN; TILDESLEY, 1987), no qual o número de moléculas é fixo e a pressão e temperatura são controladas. Após esse processo, o sistema está pronto para seguir à etapa de produção da trajetória por DM, em *ensemble* NVT ou NPT a depender do estudo a ser realizado. Por fim, após a aquisição da trajetória obtida pela DM, pode-se realizar diversas análises estruturais e energéticas de interesse.

2.8 Cálculo de Energia Livre de Ligação

A energia livre de ligação, é uma grandeza termodinâmica que estima a energia de interação entre duas moléculas, como a interação antígeno(Ag)/anticorpo(Ab), por exemplo. A Equação 8 ilustra que o anticorpo e o antígeno podem se ligar,

reversivelmente, para formar um complexo Ag-Ab. No equilíbrio termodinâmico, a proporção entre a concentração do complexo Ag-Ab e as concentrações do antígeno e do anticorpo é constante, e essa proporção é conhecida como constante de equilíbrio (K_{eq}).

$$\frac{[Ag-Ab \text{ complexo}]}{[antígeno][anticorpo]} = \frac{K_a}{K_d} = K_{eq} \quad (8)$$

No equilíbrio termodinâmico, a variação da energia livre padrão de ligação (ΔG^0), pode ser relacionada com a afinidade entre um fragmento scFv de um anticorpo e o antígeno por meio da Equação 9.

$$\Delta G^0 = -k_B T \cdot \ln K \quad (9)$$

Onde, k_B é a constante de Boltzmann que relaciona temperatura e energia molar de moléculas, 0.0019872041, T é a temperatura associada a k_B e é expressa em Kelvin (K). A contante no logaritmo neperiano de K, pode ser K_a , constante de afinidade, ou K_d , constante de dissociação; de acordo como a ordem dos fatores de energia dos reagentes (G_{reag}^0) e dos produtos (G_{prod}^0), como mostrado na Equação 10.

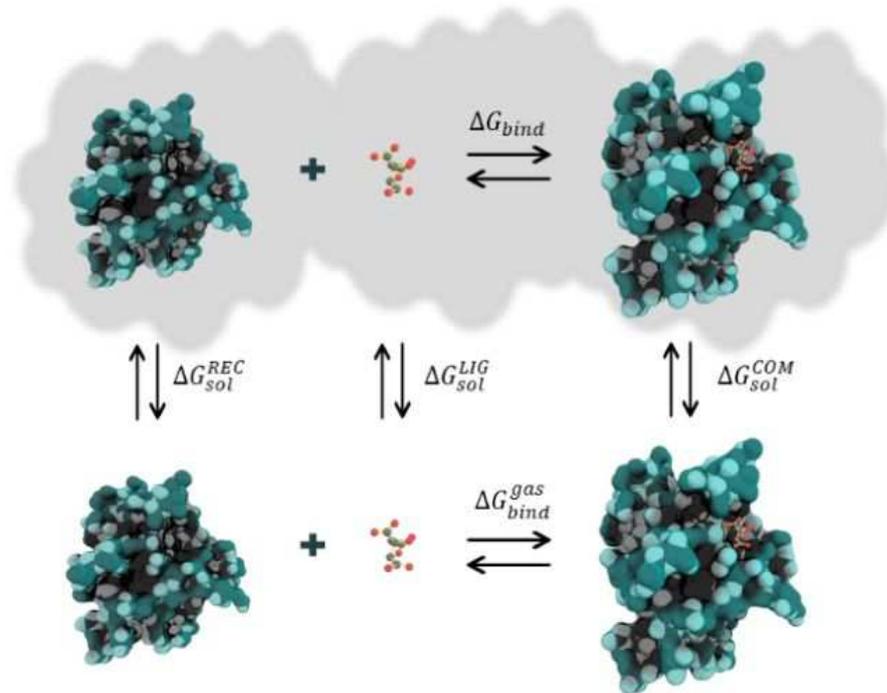
$$G_{reag}^0 - G_{prod}^0 = -K_b T \cdot \ln K_a \quad \text{ou} \quad G_{prod}^0 - G_{reag}^0 = -K_b T \cdot \ln K_d \quad (10)$$

2.8.1 Mecânica Molecular de Poisson-Boltzmann com Solvatação da Área de Superfície (MM/PBSA)

Sabe-se que as proteínas no ambiente biológico não são estáticas. Portanto, as trajetórias obtidas via simulação de DM de evolução temporal dos sistemas, a nível atomístico, são uma ferramenta preciosa para o cálculo de energia livre de ligação. O cálculo de energia livre de ligação baseado em MM/PBSA considera a energia livre ser uma função de estado. Dessa forma, para calcular a sua variação, é necessário, apenas, calcular o estado inicial e estado final, desprezando-se o caminho entre os dois estados, e, sendo assim, muito mais rápido computacionalmente (UL-HAQ; MADURA, 2015). Conforme a Figura 11, considerando a energia livre de ligação sendo uma grandeza de estado, é possível descrever sua variação através de um

ciclo termodinâmico, cujo $\Delta G_{total} = 0$ (VALDÉS-TRESANCO *et al.*, 2021).

Figura 11 - Ciclo termodinâmico para cálculo de energia livre de ligação por MM/PBSA.



Fonte: VALDÉS-TRESANCO *et al.* (2021). LIG: ligante; REC: receptor; e COM: complexo. Bind: Ligação. Gas: contribuição energética na fase gasosa.

A energia de ligação livre para um complexo (ΔG_{Bind}) que está na Figura 11, pode ser estimada conforme a Equação 11.

$$\Delta G_{Bind} = G_{COM} - G_{REC} - G_{LIG} \quad (11)$$

Onde a contribuição energética do complexo, receptor e ligante é dada por:

$$G = EMM + G_{sol} - T \cdot \Delta S$$

Considerando que a energia livre de ligação para um complexo pode ser estimada conforme exposto na Equação 12.

$$\Delta G_{Bind} = \Delta H - T \cdot \Delta S \quad (12)$$

Onde ΔH corresponde à entalpia de ligação e $-T\Delta S$ à entropia conformacional

após a ligação do ligante. O ΔH pode ser decomposto nos seguintes termos, conforme a Equação 13.

$$\Delta H = \Delta EMM + \Delta G_{sol} \quad (13)$$

Onde a componente ΔEMM faz referência à contribuição energética da associação das moléculas na fase gasosa e é obtida através do campo de força usado na simulação de DM (VALDÉS-TRESANCO *et al.*, 2021). Nela estão inclusas as energias referentes aos comprimentos, ângulos e torções de ligações, assim como as interações eletrostáticas e de *van der Waals*, conforme a Equação 14.

$$\Delta EMM = (\Delta E_{\text{ligação}} + \Delta E_{\text{ângulos}} + \Delta E_{\text{diedros}}) + (\Delta E_{\text{elet}} + \Delta E_{\text{vdW}}) \quad (14)$$

Já o termo ΔG_{sol} se refere soma da energia de solvatação polar e solvatação apolar, conforme a Equação 15.

$$\Delta G_{sol} = \Delta G_{\text{solpolar}} + \Delta G_{\text{solapolar}} \quad (15)$$

A componente polar do termo energético de energia de solvatação é originária da energia de transferência de uma partícula carregada entre dois meios – vácuo e solvente – com constantes dielétricas distintas, sendo o solvente modelado de forma implícita. Essa componente pode ser determinada por meio de diferentes métodos como *Poisson-Boltzmann* (PB) ou *Generalized Born* (GB). Já componente apolar advém da formação de uma cavidade no solvente que possa comportar o soluto. Essa componente é calculada como a soma entre energias “desfavoráveis” pela ruptura de interações solvente-solvente e energias “favoráveis” pela formação de interações soluto-solvente.

Há vários programas para cálculo de energia livre de ligação baseado em MM/PBSA, sendo eles: *g_mmpbsa*, *GMXPBSA2.1*, *MMPBSA.py* e *gmx_MMPBSA*. O *gmx_MMPBSA* (VALDÉS-TRESANCO *et al.*, 2021) foi utilizado por esse trabalho uma vez que nesse programa pode-se considerar uma constante dielétrica para membrana diferente do restante do soluto e solvente, assim, sendo mais adequado para o estudo dos complexos scFv/CD22.

2.9 Engenharia de Proteínas *in silico*

O grupo de pesquisa em Engenharia de Proteínas e Soluções para a Saúde (GEPeSS), da Fundação Oswaldo Cruz (FIOCRUZ), sede Ceará, possui diversas linhas de pesquisa voltadas para o desenvolvimento de proteínas de interesse biotecnológico, dentre as quais pode-se destacar: PD&I em Anticorpos Monoclonais e Fragmentos, Simulação Molecular e Engenharia de CAR para Terapia Celular. Um dos principais objetivos do GEPeSS é usar técnicas de engenharia de proteínas, *in vitro*, *in silico* (simulações de DM) e *in vivo* para desenvolver scFvs e outros domínios do CAR para aumentar sua eficácia em células CAR-T.

O GEPeSS conta com uma equipe de pesquisadores de excelência na área de biotecnologia e é liderado pelos professores Dr. Marcos Roberto Lourenzoni e Dr. Gilvan Pessoa Furtado. Os trabalhos envolvendo estudantes de pós-graduação, que estão vinculados ao Programa de Pós-Graduação em Biotecnologia de Recursos Naturais (PPGBRN) do Centro de Ciências Agrárias da Universidade Federal do Ceará (UFC), foram desenvolvidos pelos mestres Natália Fernandes Frota Oliveira, Ana Júlia Ferreira Lima, Mathias Coelho Batista e Ana Virgínia Frota Guimarães; e doutor Alison de Sousa Rebouças.

O Mestrado da Me. Ana Julia Ferreira Lima visou estudar a influência da sonda de Myc no scFv anti-CD19 para construção de Receptor Quimérico de Antígeno (CAR). Os dados de simulação de DM testaram outras duas sondas nas posições N e C terminal e os resultados corroboram o observado experimentalmente *in vivo* (LIMA, 2022). O trabalho do Me. Mathias Coelho Batista objetivou a construção de uma ferramenta computacional que trata os dados de sequenciamento NGS, provenientes de bibliotecas de scFv. A construção dessa ferramenta computacional buscou otimizar, entre outros fenótipos, a afinidade entre scFvs e receptores, para aplicação em CAR (BATISTA, 2023).

Já o trabalho de mestrado da Me. Natália Fernandes Frota Oliveira contribuiu para o entendimento do mecanismo de interação de um modelo de CAR com o CD19, por meio do uso da técnica de simulação de Dinâmica Molecular (DM) (OLIVEIRA, 2020). Entretanto, recentemente, estudos demonstraram que alguns pacientes tratados com CAR anti-CD19 não conseguem obter uma resposta imune duradoura, devido ao escape de antígeno CD19 (ZHANG *et al.*, 2022). Sabendo disso, desenvolveu-se o interesse no GEPeSS em estudar scFvs presente nos CARs

visando outros alvos como o CD22. Assim, surgindo a atual temática de estudo desta dissertação.

3. JUSTIFICATIVA

O governo federal brasileiro importa mais de 90% de todos os biofármacos utilizados para o tratamento da LLA. Em 2020, o custo dessas importações chegou a cerca de 13,7 bilhões de dólares. Diante desse contexto, recentemente, por intermédio de uma parceria entre o Ministério da Saúde e a Empresa Brasileira de Pesquisa e Inovação Industrial (Embrapii), o governo federal investiu aproximadamente vinte milhões de reais em pesquisas e formação de mão de obra qualificada na área da biotecnologia, a fim de aumentar a produção de biofármacos no território brasileiro e, assim, reduzir a dependência de importações para a aquisição desses produtos biotecnológicos (MINISTÉRIO DA SAÚDE, 2020).

A terapia por células CAR-T que usa um biofármaco chamado de CAR é uma das imunoterapias mais promissoras para o tratamento da LLA. Os resultados provenientes do uso do CAR comercial Kymriah® retratam que nos Estados Unidos (EUA) cerca de 70% a 90% dos enfermos apresentaram regressão da leucemia e alguns pacientes obtiveram a remissão completa dessa neoplasia maligna (BAYBUTT *et al.*, 2019). Por essas razões, a terapia por células CAR-T pode tornar-se um tratamento alternativo para a LLA de extrema relevância no Brasil.

Recentemente, começaram a ser publicados resultados do desenvolvimento de CARs anti-CD22. Dois CARs anti-CD22 que já estão em fase de testes clínicos usam o scFv derivado do mAb M971, um usando um *linker* curto (GGGGS) e o outro usando um scFv de *linker* longo (GGGGS)₄, desenvolvidos por Fry e colaboradores (2018) e Ruella e colaboradores (2017), respectivamente. O domínio sobre modificações na composição de resíduos dos scFvs presentes nesses tipos de CARs é essencial para a construção de CARs anti-CD22 em território brasileiro. A compreensão do mecanismo atômico-molecular envolvendo os scFvs com seu respectivo alvo molecular, por meio de técnicas computacionais, como a simulação de Dinâmica Molecular acoplada a metodologia de cálculo de energia livre de ligação por MM/PBSA, possibilita a proposição e avaliação mutações que poderão realçar a interação scFv-CD22, beneficiando a afinidade. Um aumento de afinidade, por mutações pontuais, dos scFvs de *linker* curto e longo de M971 pelo CD22, por exemplo, pode viabilizar a

obtenção dos CARs anti-CD22 completos e melhorados, com as mutações sugeridas. Portanto, torna-se fundamental desenvolver um protocolo, *in silico*, confiável, a fim de ser usado para proposição de mutações nos scFvs para realçar a sua afinidade ao CD22.

Após o estudo a nível atomístico-molecular, da interação entre os scFvs de M971 de *linker* curto e longo com e o antígeno CD22, geração do protocolo *in silico* e a proposição de novas mutações nos scFvs, pode-se aproveitar os resultados obtidos dentro do próprio GPeSS, que possui projeto vigente, fomentado pelo DECIT e INOVA Produtos-Fiocruz, para a produção de CARs anti-CD19, anti-CD20, anti-CD22 e anti CD19/CD22, a fim de inicialmente realizar testes *in vitro* em colaboração com a Dra. Andrea Maranhão e o Dr. Marcelo Brígido da Universidade de Brasília (UNB), assim como realizar testes *in vitro* em colaboração com o Dr. Gilvan Furtado da Fundação Oswaldo Cruz Ceará e, finalmente, testes *in vivo* em parceria com o Dr. Martin Bonamino do Instituto Nacional do Câncer (INCA).

4. OBJETIVOS

4.1 Objetivo Geral

Estudar, a partir de simulação de DM, a interação entre os scFvs de *linker* curto e longo de M971 e a glicoproteína de membrana CD22, a fim de gerar um protocolo confiável que possa ser usado para proposição de mutações, *in silico* e *in vitro*, nos scFvs para realçar a afinidade ao CD22 visando obter CARs mais efetivos para terapia celular *CAR-T Cell*.

4.2 Objetivos Específicos

- a) Obter as estruturas tridimensionais (3D) de fragmentos scFvs de *linker* curto e longo derivados do M971 a partir das coordenadas cristalográficas da porção Fab deste anticorpo (código PDB 7O52), e submetê-las às simulações de DM;
- b) Efetuar análises estruturais para os modelos de scFvs com base nas trajetórias provenientes das simulações de DM;
- c) Adicionar os loops e árvores de glicosilações ausentes nos domínios D6 e D7 na estrutura 3D da molécula de CD22 (código PDB 7O52);
- d) Montar o sistema CD22 em solução aquosa, contendo os domínios extracelulares glicosilados D6 e D7 da molécula CD22, e submetê-lo à simulação de DM;

- e) Executar análises de determinantes estruturais para o modelo de CD22 a partir da trajetória obtida pela simulação de DM;
- f) Modelar o domínio transmembrana da molécula de CD22 em modelo de membrana fosfolipídica com dimensões 12 x 12 e 16 x 16 nm e submeter os modelos obtidos à simulação de DM;
- g) Montar o modelo de CD22 desse estudo inserido em membrana, conectando as estruturas representativas, obtidas por DMs, dos domínios D7 de (D6-D7) e o transmembrana, inserido num modelo de membrana de fosfolipídio 12 x 12 nm.
- h) Montar os sistemas scFv/CD22, considerando o scFv de *linker* curto e longo, a partir das melhores configurações de docagem e submetê-los às simulações de DM;
- i) Montar os sistemas scFv/CD22, considerando o scFv de *linker* curto e longo, sobrepondo as estruturas presentes nos *clusters* mais representativos às iniciais dos dados cristalográficos (código PDB 7O52), e submetê-los às simulações de DM;
- j) Montar os sistemas scFv/CD22, considerando o scFv de *linker* curto e longo, a partir das coordenadas cristalográficas (código PDB 7O52), somente adicionando o *linker* curto e longo, *loops*, glicosilações ausentes, e submetê-los às simulações de DM;
- k) Realizar análises estruturais e energéticas, incluindo MM/PBSA, utilizando as trajetórias de DM, para verificar a estabilidade estrutural e o ΔG de ligação dos modelos de scFv/CD22.
- l) Montar os sistemas scFv/CD22, considerando o scFv de *linker* curto e longo, em um modelo de membrana fosfolipídica 16 x 16 nm, a partir da união do domínio transmembrana ao domínio D7 dos sistemas scFv/CD22 montados a partir das coordenadas cristalográficas, e submetê-los às simulações de DM, em triplicata;
- m) Realizar análises estruturais e energéticas, incluindo MM/PBSA e Potencial de Interação Intermolecular (PII) por resíduos, utilizando as trajetórias de DM, para verificar a estabilidade estrutural e o ΔG de ligação dos modelos construídos em membrana.

5. METODOLOGIA

5.1 Modeller

O *MODELLER* (WEBB; SALI, 2016) é um dos diversos programas de modelagem molecular comparativa e encontra-se disponível gratuitamente para

download na internet. A interação do usuário com esse programa se dá por meio do uso de *scripts* (ŠALI; BLUNDELL, 1993). Esse *software* realiza a modelagem molecular por homologia de proteínas utilizando como molde uma ou mais proteínas que já possuem sua estrutura tridimensional resolvida e depositada no banco de dados PDB. A partir do alinhamento entre a sequência da proteína alvo e a sequência da(s) proteína(s) molde é gerado um conjunto de restrições que são aplicadas para geração da estrutura 3D da proteína de interesse. Essas restrições realizadas a partir da proteína de referência são de ângulos, distâncias, ângulos diedros, pares de ângulos diedros e algumas outras características espaciais definidas por átomos ou pseudo-átomos (GINALSKI, 2006; LIU *et al.*, 2011; WEBB; SALI, 2016). A violação dessas restrições, por incongruências entre as sequências da(s) proteína(s) molde(s) e proteína alvo, é minimizada por meio de processos de otimização (MARTÍ-RENOM *et al.*, 2000). No *MODELLER*, após o processo de otimização, o melhor modelo gerado é selecionado por meio das funções objetivas *DOPE score* (SHEN; SALI, 2006) ou *SOAP* (DONG *et al.*, 2013).

Para este trabalho, uma vez que já se tem a estrutura cristalográfica do complexo Fab de M971 com os domínios D6-D7 de CD22, sob cód. PDB 7O52 (ERENO-ORBEA *et al.*, 2021), realizou-se a modelagem apenas do *linker* entre os domínios VH e VL de M971 para construir os modelos de scFv curto e longo. Da mesma forma, usou-se o *MODELLER* para CD22 apenas para modelar *loops* faltantes nos domínios D6-D7 do CD22. Posteriormente, os melhores modelos obtidos pelo *MODELLER* passaram por etapas de validação da estrutura tridimensional.

5.2 I-TASSER

O *I-TASSER* (*Iterative Threading ASSEmbly Refinement*) (YANG *et al.*, 2015) é um programa online gratuito de modelagem molecular que utiliza a metodologia *threading*. Neste programa, gera-se um modelo completo de proteína através da remoção de fragmentos contínuos de estruturas encontrados após o alinhamento com o template ou templates, remontando-os utilizando simulações de Monte Carlo *replica-exchanged*. Para cada modelo que se pretende obter, as simulações do *I-TASSER* geram um grande número de estruturas, chamadas de *decoys*. Por meio do programa *SPICKER* essas estruturas são agrupadas em *clusters*, com base em suas similaridades. Em seguida, os cinco principais clusters são selecionados para gerar

os modelos completos (ZHANG, 2008; ROY *et al.*, 2010; YANG *et al.*, 2015).

A confiabilidade de cada modelo gerado é verificada, inicialmente, por meio de parâmetros do próprio programa I-TASSER. A exemplo desses parâmetros tem-se o C-score, calculado com base no alinhamento dos *templates* com os modelos obtidos. O valor C-score compreende uma faixa de [-5, 2], em que um valor mais próximo de 2 indica um modelo de maior confiabilidade (ZHANG, 2008; ROY *et al.*, 2010; YANG *et al.*, 2015).

Para este trabalho, uma vez que não se tem estrutura 3D do domínio transmembrana de CD22 e nem de estruturas homólogas no banco de dados PDB para se realize a sua modelagem por homologia, o modelo de domínio transmembrana de CD22 foi obtido por meio do programa I-TASSER. Em seguida, o melhor modelo obtido passou por etapas de validação da estrutura tridimensional.

5.3 Avaliação dos Modelos Obtidos

Visto que proteínas podem assumir teoricamente diversas conformações devido à capacidade que essas macromoléculas possuem de realizar rotações em sua estrutura, sem que ocorra a quebra de ligações covalentes, tem-se a necessidade de avaliar o melhor modelo obtido por meio da modelagem molecular (FILHO; ALENCASTRO, 2003). A avaliação desses modelos, no caso dos scFvs de *linker* curto e longo de M971, domínios D6-D7 glicosilados de CD22 e domínio transmembrana de CD22, se deu pelo uso de programas como *MolProbity* (DAVIS *et al.*, 2007) e *PROCHECK* (LASKOWSKI *et al.*, 1993), por exemplo. No *MolProbity* foram avaliados os parâmetros como os valores de *clash score*, de *MolProbity score* de rotâmeros ruins e de ligações e ângulos em conformações não desejadas, ângulos torcionais (Φ , Ψ e Ψ) e o número de resíduos fora de regiões permitidas no gráfico de Ramachandran. Já no programa *PROCHECK*, foi verificado se os resíduos de cada modelo avaliado se encontram em regiões favoráveis ou permitidas. Um bom modelo deve apresentar a seguinte combinação de resultados: *clash score* e *MolProbity score* mais próximo de 100° *percentil*, menor número de rotâmeros ruins e menos de 0.05% de resíduos em regiões não permitidas.

Após as modelagens e validação das estruturas tridimensionais (3D) obtidas prossegue-se para montagem dos sistemas, simulação de DM e análises de determinantes estruturais e energéticos de interesse.

5.4 Simulação de Dinâmica Molecular e Análises

Foi utilizado o programa GROMACS versão 2018.3 (BERENDSEN *et al.*, 1995) para as simulações de DM, uma vez que contém as ferramentas necessárias para a montagem de todos os sistemas. O campo de força CHARMM36m foi escolhido para conduzir as simulações de DM (HUANG; MACKERELL, 2013). As árvores de glicosilação da molécula de CD22 e os arquivos de itps foram gerados com o programa CHARMM-GUI. Posteriormente, a molécula D6-D7 glicosilada, dois domínios do CD22, foi submetida ao programa H++ (GORDON *et al.*, 2005) a fim de estimar o estado de protonação dos resíduos de histidina em pH fisiológico, pH= 7,4.

Todos os sistemas (scFv-água, curto e longo, domínios D6-D7 glicosilados de CD22 em água; molécula D6-D7 glicosilada de CD22 em membrana; scFv-CD22, de *linker* curto e longo em água; e scFv/CD22 de *linker* curto e longo em membrana) foram preparados a partir da criação da caixa de simulação, solvatação (modelo de água TIP3P) e adição de íons. As proteínas, provenientes de coordenadas cristalográficas e modelagem, foram submetidas ao processo de minimização de energia para otimização da geometria. Adiante, procedeu-se à etapa de termalização a fim de equilibrar a energia total do sistema, os íons e o solvente em torno da proteína no estado termodinâmico de interesse. Essa fase aconteceu de duas formas: *ensemble* canônico ou NVT (número de partículas, volume e temperatura constante) (BERENDSEN *et al.*, 1984; BERENDSEN *et al.*, 1995) e *ensemble* isobárico-isotérmico ou NPT (número de partículas, pressão e temperatura constante) a 310 K. Por fim, os sistemas foram submetidos à etapa de produção da DM sob condições NPT (PARRINELLO, RAHMAN, 1981; BERENDSEN *et al.*, 1984; HUANG, 1987; BERENDSEN *et al.*, 1995) a 310 K.

Os parâmetros utilizados para a equilibração dos sistemas e para as simulações sem membrana foram: controle de temperatura: V-rescale (BUSSI; DONADIO; PARRINELLO, 2007); controle da pressão: Berendsen (BERENDSEN *et al.*, 1984, 1995); integrador: md; temperatura: 310K de modo a representar as condições fisiológicas na célula. Para os sistemas em membrana foram usados os mesmos parâmetros, somente alterando o controle da pressão para Parrinello-Rahman (BERENDSEN *et al.*, 1984, 1995; BUSSI; DONADIO; PARRINELLO, 2007). Sendo assim, para todos os sistemas que foram submetidos à etapa de termalização e a produção da DM, a pressão foi controlada a 1 atm. Ademais, com o objetivo de

controlar os comprimentos de ligação envolvendo os átomos de hidrogênio das proteínas e das águas foram utilizados os algoritmos *LINCS* (HESS *et al.*, 1997; HESS, 2008) e *SHAKE* (RYCKAERT; CICCOTTI; BERENDSEN, 1977), respectivamente.

As análises estruturais foram realizadas através do cálculo do perfil de RMSD e do Raio de giro (Rg). Para o cálculo do perfil de RMSD, foi utilizada a ferramenta *gmx rms* do pacote GROMACS 2018.3 (BERENDSEN *et al.*, 1995). Como visto na equação 16, média é calculada entre N átomos (*i*) de carbono alfa (C α) sobrepostos, nas posições cartesianas tridimensional (*r*), entre duas estruturas, uma coletada ao longo da simulação, no tempo *t*₂, e a de referência (no caso do início da DM, *t* = 0; *t*₁). Não foram consideradas as massas *m* e a normalização pela massa total (*M*), pois todos os átomos são idênticos, C α . Os RMSDs médios e os desvios serão calculados a partir do momento em que se considera que a estrutura alcança o equilíbrio estrutural (teq).

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_1) - r_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (16)$$

Na análise estrutural baseada no cálculo do Raio de giro (Rg), foi utilizada a ferramenta *gmx gyrate* do pacote GROMACS 2018.3 (BERENDSEN *et al.*, 1995). A curva de Rg em equilíbrio revela o alto grau de compacidade proteica, reflexo do enovelamento estável da proteína.

A ferramenta *gmx cluster*, também originária do pacote GROMACS 2018.3 (BERENDSEN *et al.*, 1995), foi utilizada visando verificar as estruturas tridimensionais de proteínas mais frequentes e aquelas que melhor representam cada sistema. Na análise de *cluster*, as estruturas 3D obtidas ao longo da simulação de DM são agrupadas com base nas distâncias entre essas estruturas.

Para os complexos scFv/CD22 curto e longo, em solução aquosa e em membrana, foi utilizada a ferramenta *gmx energy*. Na obtenção dos perfis de energia configuracional, o potencial de interação intermolecular (PII) é a soma dos potenciais de *Lennard-Jones* e *Coulomb*, o primeiro se refere à interação de *van der Waals* entre átomos não ligados e neutros, enquanto o último, está relacionado a interação eletrostática entre átomos não ligados e carregados (SPOEL *et al.*, 2005). O raio de corte de distância considerado para calcular os potenciais foi de 0,5 nm, em algumas situações o raio pode ser diferente, como será informado nos perfis. O raio de 0,5 nm

visa estabelecer análises de maior proximidade entre resíduos. PII também foi expresso como média por resíduos, entre os resíduos dos scFvs de *linker* curto e longo e os do CD22 dentro de um raio de corte de 0,5 nm. Foi utilizado um programa desenvolvido pelo grupo para calcular os potenciais acessando a estrutura de dados do GROMACS e lendo as trajetórias de DM.

Por fim, os sistemas scFv/CD22, em água e em membrana foram submetidos aos métodos de proposição de energia livre de ligação MM/PBSA, a fim de estimar a afinidade. Para realizar essa análise foi utilizado o *software gmx_MMPBSA* (VALDÉS-TRESANCO *et al.*, 2021) que pode ser instalado como um pacote do programa GROMACS 2018 (BERENDSEN *et al.*, 1995). Dessa forma, como as simulações de DM desses sistemas sem e com membrana foram realizadas no GROMACS 2018.3, utilizou-se as trajetórias geradas para submissão no programa *gmx_MMPBSA* (VALDÉS-TRESANCO *et al.*, 2021).

5.5 Docking Molecular e o PROGRAMA Haddock

O *Docking* Molecular é uma importante ferramenta computacional que permite prever a interface de interação entre uma proteína e um ligante. O processo de *Docking* Molecular pode ser dividido em duas etapas: (1) determinação da conformação do ligante no sítio de ligação da proteína e; (2) classificação de cada conformação por uma função de pontuação (MENG *et al.*, 2012).

O *Haddock* (*High Ambiguity Driven biomolecular Docking*) é um programa de *Docking* Molecular que se baseia nas seguintes etapas: minimização de energia de corpo rígido, refinamento semiflexível no espaço do ângulo de torção e refinamento final com solvente explícito. Os resultados gerados são agrupados em *clusters*, onde o melhor *cluster* é apresentado a partir de uma média ponderada dos termos eletrostáticos, dessolvatação e energia de *van der Waals*, juntamente com a contribuição energética das restrições utilizadas para impulsionar a docagem (DOMINGUEZ *et al.*, 2003; BONVIN *et al.*, 2010).

O programa *Haddock Sever Web* versão 2.4 (HONORATO *et al.*, 2021) possui interfaces voltadas para cada nível de usuário:

- Fácil: Requer apenas as estruturas e a lista de resíduos do sítio ativo que interagem com o ligante. Todos os outros parâmetros utilizados para o

processo da docagem são definidos por valores padrão ou determinados automaticamente.

- Especialista: Permite a personalização de alguns parâmetros, tais como: a definição dos segmentos flexíveis ou rígidos, distâncias entre os átomos, estado de protonação, dentre outros.
- Guru: Além de permitir as mesmas alterações de parâmetros do nível especialista, também dá acesso para configurar constantes, tais como: constantes de força, temperatura e pesos dos termos da função de pontuação (*score*).

Neste trabalho, a interface de interação dos complexos scFv/CD22 de *linker* curto e longo foi montada a partir de três diferentes posições de interação. Na primeira posição de interação, realizou-se *Docking* Molecular entre diferentes configurações obtidas de cada sistema scFv-água (*linker* curto e longo) durante a simulação de DM em solução aquosa (100, 390, 500 e 600 ns) e a configuração final da molécula de CD22 obtida pela DM, utilizando o nível de usuário especialista do programa *Haddock Sever Web* versão 2.4 (HONORATO *et al.*, 2021). Na segunda posição de interação, coletou-se as configurações presentes nos *clusters* mais representativos de cada sistema scFv-água de *linker* curto e longo; e CD22-água e realizou-se a sobreposição dessas estruturas coletadas à cristalográfica obtida por Ereño-Orbea e colaboradores (2021). Por fim, na última posição de interação coletou-se a interface de interação da estrutura cristalográfica (EREÑO-ORBEA *et al.*, 2021) e apenas adicionou-se os *linkers* ausentes nos scFvs; e *loops* faltantes e glicosilações ausentes na posição N70 e N130 na molécula de CD22.

6. PROCEDIMENTOS EXPERIMENTAIS

6.1 Modelagem Molecular

6.1.1 Modelagem dos scFvs de M971 anti-CD22

Os dois modelos de scFv deste estudo tiveram os domínios variáveis da cadeia pesada (VH, 124 resíduos) e leve (VL, 107 resíduos) derivados do anticorpo monoclonal humano M971, conectados por um polipeptídeo chamado de *linker* (Quadro 4). A porção Fab do anticorpo M971, que contém os domínios VH e VL, já

possui estrutura tridimensional (3D) depositada na plataforma *Protein Data Bank* (PDB), sob cód. 7O52 (ERENO-ORBEA *et al.*, 2021). Os *linkers* inseridos foram do tipo GGGGS e (GGGGS)₄, denominado de *linkers* curto e longo, respectivamente. Portanto, os scFvs resultantes foram chamados de scFv curto e scFv longo (SINGH *et al.*, 2021). Em geral, os *linkers* não possuem estrutura 3D depositada no banco de dados PDB. A estrutura 3D dos dois modelos de scFv foi obtida através do *Modeller 10.1* (WEBB; SALI, 2016), utilizando o *script* de modelagem de resíduos faltantes, apenas para inserir o *linker* ausente. Os *linkers* iniciaram no C-terminal do domínio VH e finalizaram no N-terminal do domínio VL. Assim, obtendo-se scFvs na orientação VH-VL. Foram gerados 1050 modelos para cada tipo de scFv e o melhor modelo de cada scFv foi selecionado inicialmente avaliando-se o valor da função objetiva *DOPE*, mais negativo, *score* do *Modeller 10.1*. Posteriormente, realizou-se a validação dos modelos de scFv selecionados utilizando os programas *MolProbity* (DAVIS *et al.*, 2007; CHEN *et al.*, 2010) e *PROCHECK* (LASKOWSKI *et al.*, 1993). Uma vez que as estruturas tridimensionais dos scFvs foram validadas prosseguiu-se para as etapas de montagem dos dois sistemas scFv-água a fim de posterior realização da DM.

Quadro 4 - Sequências de aminoácidos que foram utilizados para a modelagem dos dois modelos de scFv. As sequências dos domínios variáveis (VH e VL) foram derivadas do anticorpo M971.

Domínio VH*	QVQLQQSGPGLVKPSQTLTLTCAISGDSVSSNSAAWN WIRQSPSRGLEWLGRTYYRSKWYNDYAVSVKSRITINP DTSKNQFSLQLNSVTPEDTAVYYCAREVTGDLEDAFDI WGQGTMTVSS
Domínio VL*	DIQMTQSPSSLSASVGDRTITCRASQTIWSYLNWYQQ RPGKAPNLLIYAASSLQSGVPSRFSGRGSFTDFTLTSS LQAEDFATYYCQQSYISIPQTFGGGTKLEIK
Linker curto**	GGGGS
Linker longo**	(GGGGS) ₄

*Sequências obtidas do *Protein Data Bank* (PDB).

**Sequências obtidas do trabalho de Singh e colaboradores (2021).

Fonte: elaborada pela autora (2024).

6.1.2 Modelagem dos domínios extracelulares D6-D7 glicosilados do CD22

A estrutura 3D dos domínios extracelulares D6-D7 da proteína de membrana CD22 encontra-se co-cristalizada com a porção Fab do anticorpo monoclonal humano M971, no PDB, sob cód. 7O52 (ERENO-ORBEA *et al.*, 2021). Na estrutura de CD22

há algumas regiões, *loops*, três *loops* (A, B e C) que estão ausentes no domínio D6 e um *loop* (D) ausente logo após a sequência de aminoácidos do domínio extracelular D7 (Quadro 5). Foi feito alinhamento entre as sequências do CD22 *UniProt*, cód. P20273 (Figura 12) e dos domínios D6-D7 da estrutura cristalográfica para encontrar sequências faltantes em D6-D7. Neste trabalho, foi realizada a modelagem por homologia dos *loops* faltantes, utilizando o protocolo de modelagem de resíduos faltantes do programa *Modeller 10.1* (WEBB; SALI, 2016), de forma a obter a completa estrutura 3D proteica dos domínios extracelulares D6-D7 da molécula CD22.

Figura 12 - Alinhamento entre as sequências de aminoácidos dos domínios extracelulares D6-D7 do CD22 obtidas pelo Protein Data Bank e pelo UniProt.

```

aln.pos      10      20      30      40      50      60
CD22/PDB    --DVRVRKIKPLSEIHSNGNSVSLQCD-----VQFFWEKNGRLLGKESQLNFDSISPEDAGSYSCW
UNIPROT     PRDVRVRKIKPLSEIHSNGNSVSLQCDFSSSHPKEVQFFWEKNGRLLGKESQLNFDSISPEDAGSYSCW
_consrvd    *****
aln.p       70      80      90      100     110     120     130
CD22/PDB    V-----TASKAWTLEVLVYAPRRLRVSMSPGDQVMEGKSATLTCESDANPPVSHYTWFDWNNQSLPYH
UNIPROT     VNNSIGQTASKAWTLEVLVYAPRRLRVSMSPGDQVMEGKSATLTCESDANPPVSHYTWFDWNNQSLPYH
_consrvd    *
aln.pos     140     150     160     170     180
CD22/PDB    SQKLRLPEVKVQHSGAYWCQGTNSVGKGRSPLSTLTVYYSPE-----
UNIPROT     SQKLRLPEVKVQHSGAYWCQGTNSVGKGRSPLSTLTVYYSPETIGRR
_consrvd    *****

```

Fonte: elaborada pela autora (2024). O alinhamento foi realizado no *Modeller 10.1*. Asterisco (*) identifica aminoácidos conservados nas duas sequências.

Quadro 5 - Identificação da sequência de aminoácidos e da numeração dos loops faltantes na estrutura cristalográfica que foram modelados por meio do programa *Modeller 10.1*.

Loops Modelados	Sequência	Resíduos
Loop A*	PR	1-2
Loop B*	FSSSHPKE	27-34
Loop C*	NNSIGQ	70-75
Loop D*	TIGRR	179-183

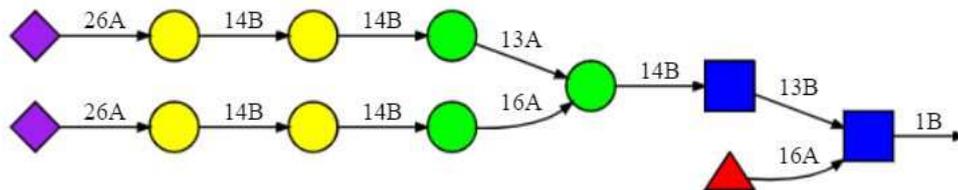
*Sequências obtidas do *UniProt*: <https://www.uniprot.org/uniprotkb/P20273/entry>

Fonte: elaborada pela autora (2024). A numeração dos resíduos está de acordo com o arquivo do sistema em formato PDB submetido à simulação de DM.

Na modelagem foram gerados 1050 modelos. O melhor modelo foi selecionado utilizando os programas *MolProbity* (DAVIS *et al.*, 2007; CHEN *et al.*, 2010) e *PROCHECK* (LASKOWSKI *et al.*, 1993). O modelo selecionado e validado de

domínios extracelulares de CD22 foi submetido ao programa *H++*, sendo executado num servidor *online* (GORDON *et al.*, 2005) para estimar o estado de protonação das histidinas em pH 7.4. Posteriormente, a estrutura foi submetida ao programa *CHARMM-GUI*, também executado num servidor *online* (LEE *et al.*, 2016) para adicionar as árvores de glicosilações. No *CHARMM-GUI* também foram adicionadas as ligações dissulfeto e foram gerados os arquivos itp para os domínios D6-D7 glicosilados de CD22. Além disso, no *CHARMM-GUI* foi alterado o estado de protonação das histidinas para estarem adequadas ao pH fisiológico de 7.4, como sugerido pelo *H++*. No caso das glicosilações foram adicionadas duas árvores de glicosilação idênticas, através da ferramenta *Glycan Reader & Modeler* do *CHARMM-GUI*, nos resíduos N70 e N130, que compreendem resíduos pertencentes aos domínios extracelulares D6 e D7 da molécula do CD22, respectivamente (Figura 13).

Figura 13 - Esquema da árvore de glicosilação adicionada aos resíduos N70 e N130 do CD22 obtida através do Modeller 10.1.



Fonte: elaborada pela autora (2024). Em vermelho (L-fucose), em azul (N-acetil-D-glucosamina), em verde (D-manose), em amarelo (D-galactose) e em roxo (N-acetil-D-neuramínico ácido). A: ligação do tipo alfa (α) glicosídica. B: ligação do tipo beta (β) glicosídica.

Também no programa *CHARMM-GUI* foram adicionadas as ligações dissulfeto localizadas entre os pares de resíduos de cisteína 25-67 e 112-155 nos domínios extracelulares D6 e D7 do CD22, respectivamente. Todas as histidinas (posição 16, 31, 122, 136 e 149) foram alteradas para estarem protonadas no nitrogênio epsilon (HSE), seguindo para montagem do sistema a ser simulado.

6.1.3 Modelagem do domínio transmembrana do CD22

Inicialmente foi feita uma busca em diversos bancos de dados de proteínas sobre a sequência de aminoácidos que compõe o domínio transmembrana da molécula de CD22. A sequência foi encontrada na plataforma *UniProt*, cód. P20273

(Quadro 6) e não foi obtido estrutura correlata no banco de dados PDB. Sendo assim, foi necessário utilizar ferramentas de modelagem molecular para a obtenção da sua estrutura 3D. Neste trabalho, a estrutura tridimensional do domínio transmembrana de CD22 foi obtida através do servidor *I-TASSER* (YANG *et al.*, 2015) que utiliza uma metodologia de modelagem molecular chamada de *threading*. Foram gerados quatro modelos, em que o melhor foi selecionado a partir da avaliação do valor de *C-score*. Posteriormente, realizou-se a validação do modelo de domínio transmembrana selecionado por meio do programa *MolProbity* (DAVIS *et al.*, 2007; CHEN *et al.*, 2010).

Quadro 6 - Sequência de aminoácidos do domínio transmembrana da proteína de membrana CD22.

	Sequência
Domínio transmembrana*	VAVGLGSCLAAILILAICGL

*Sequência obtida do *UniProt*: <https://www.uniprot.org/uniprotkb/P20273/entry>
 Fonte: elaborada pela autora (2024).

A estrutura do domínio transmembrana foi submetida à ferramenta *Membrane Builder* do *CHARMM-GUI* para ser inserido em um modelo de membrana com composição de lipídeos, semelhante à membrana de linfócitos B humana. Foram usadas porcentagens de cada tipo de ácido graxo que compõe os fosfolipídeos de membrana, sendo a composição de ácidos graxos da membrana de 43 % de fosfatidilcolina (PC); 28 % de fosfatidiletanolamina (PE), 9 % de fosfatidilserina (PS) e 8 % de fosfatidilinositol (PI) (BOUGNOUX *et al.*, 1985). Foi determinada também a relação entre ácidos graxos insaturados e saturados; sendo 0.64 para PC e 1.72 para PE. As contribuições de PS e PI foram ignoradas, pois além de pequenas comparadas a PC e PE, possuem carga negativa, o que pode gerar artefatos na simulação e maior custo computacional. Tendo como base as porcentagens de ácidos graxos e a relação entre ácidos graxos saturados e insaturados da membrana, a quantidade de cada tipo de fosfolipídeo das membranas de dimensões 12 x 12 nm e 16 x 16 nm foi capaz de ser determinada (Quadro 7 e 8).

Quadro 7 - Quantidade dos tipos de fosfolípídeo utilizado para a construção da membrana 12 x 12 nm de linfócitos B humana no *CHARMM-GUI*.

Tipo de fosfolípídeo	Quantidade
DPPC	120
DOPC	64
DSPE	72
DLIPE	16
DAPE	88
DSPC	64
DOPE	24
DLIPC	24
DAPC	40

Fonte: elaborada pela autora (2024). Nota: DPPC, DOPC, DSPC e DLIPC pertencem ao grupo da fosfatidilcolina; DSPE, DLIPE e DOPE pertencem ao grupo da fosfatidiletanolamina; DAPE e DAPC pertencem ao grupo de ácidos graxos poliinsaturados.

Quadro 8 - Quantidade dos tipos de fosfolípídeo utilizado para a construção da membrana 16 x 16 nm de linfócitos B humana no *CHARMM-GUI*.

Tipo de fosfolípídeo	Quantidade
DPPC	182
DOPC	92
DSPE	110
DLIPE	18
DAPE	128
DSPC	92
DOPE	36
DLIPC	36
DAPC	54

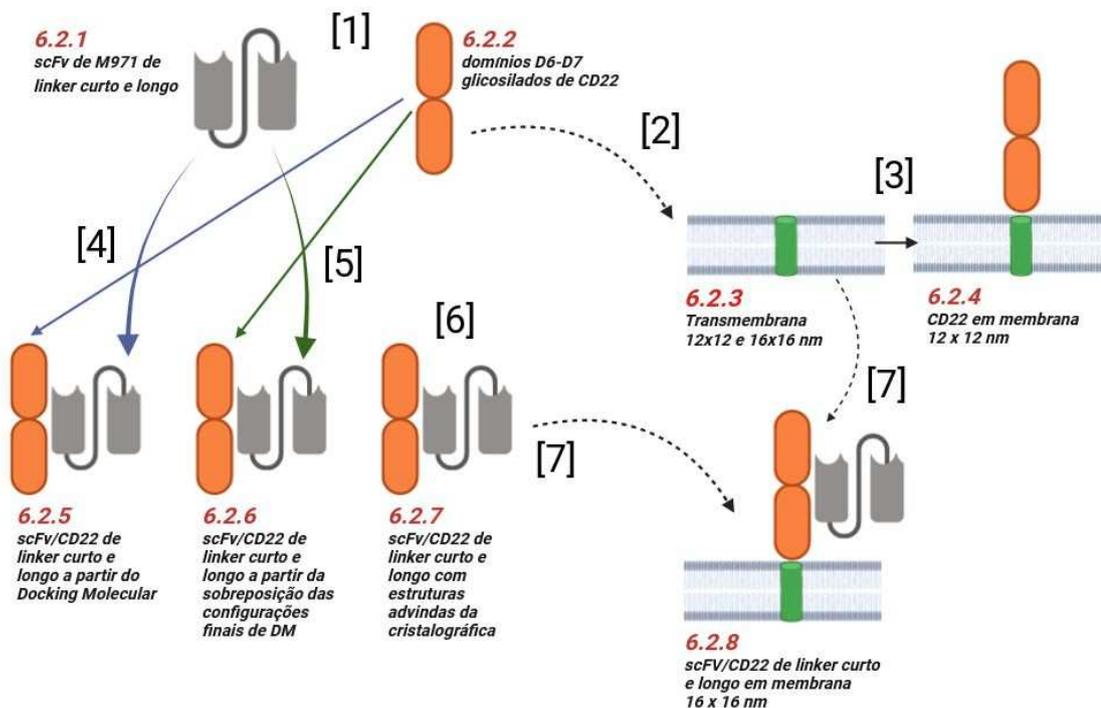
Fonte: elaborada pela autora (2024). Nota: DPPC, DOPC, DSPC e DLIPC pertencem ao grupo da fosfatidilcolina; DSPE, DLIPE e DOPE pertencem ao grupo da fosfatidiletanolamina; DAPE e DAPC pertencem ao grupo de ácidos graxos poliinsaturados.

6.1.4 Definição dos sistemas simulados

Para facilitar o entendimento dos processos experimentais serão descritos nos próximos tópicos e dos resultados obtidos por cada um deles, foi construído um fluxograma esquemático apresentado na Figura 14, contendo a definição dos sistemas submetidos à simulação de DM deste trabalho. Abaixo do esquema de cada sistema está a nomenclatura usada e a seção onde se encontra nos procedimentos experimentais. As setas em azul, verde e tracejadas indicam a sequência de eventos realizada. [1] As estruturas dos scFvs de M971 e do CD22 foram simuladas isoladamente em solução aquosa. [2] Realizou-se simulações de DM dos domínios

transmembrana de CD22 inseridos em modelos de membrana de tamanhos de 12 x 12 e 16 x 16 nm. [3] A última configuração de simulação do sistema de CD22 em solução aquosa foi conectada a última configuração do domínio transmembrana inserido na membrana de 12x12 nm e submetido à simulação de DM. [4] Foram coletadas quatro configurações de cada um dos scFvs de M971 e a última configuração do CD22 para a realização do *Docking Molecular*. [5] Coletou-se apenas as últimas configurações de simulação de cada scFv de M971 e do CD22, para realizar a sobreposição destas estruturas aos domínios VH+VL e CD22 obtidos experimentalmente. [6] Foram montados também sistemas scFv/CD22 de *linker* curto e longo, considerando a sobreposição das estruturas obtidas durante a modelagem (scFvs e CD22) aos domínios VH+VL e CD22 obtidos por cristalografia. [7] Por fim, a última configuração do domínio transmembrana de CD22 foi conectada a última configuração de DM do CD22 dos sistemas scFv/CD22, originários da sobreposição da modelagem à estrutura obtida experimentalmente, conservando também a posição de interação dos scFvs. Dessa forma, construindo-se dois sistemas scFv/CD22 de M971 de *linker* curto e longo considerando o CD22 imerso em um modelo de membrana.

Figura 14 - Fluxograma que esquematiza a definição dos sistemas simulados.



Fonte: elaborada pela autora (2024). As setas estão identificadas com número e a sequência de acontecimentos realizada em cada seção descrita nos procedimentos experimentais

6.2 Montagem dos sistemas e simulação de Dinâmica Molecular

6.2.1 Montagem e simulação de DM dos sistemas scFv-água

Os dois modelos de scFv (curto e longo) em solução aquosa, foram, inicialmente, submetidos a uma etapa de minimização de energia, em uma caixa cúbica com 9.8 e 10 nm de aresta, respectivamente. As caixas foram preenchidas com águas e íons para atingir concentração de 0.15 M com as quantidades descritas no Quadro 9. A etapa de minimização para cada um dos sistemas scFv-água foi realizada utilizando o seguinte protocolo: 10000 passos de minimização de energia para otimização da geometria e redução de conflitos estereoquímicos, por meio do algoritmo *steepest descent*. Posteriormente, com a finalidade de equilibrar a água e os íons em torno de cada modelo de scFv, foi realizada a etapa de termalização em cada um dos sistemas scFv-água. Essa etapa teve o seguinte protocolo: 12,5 ps com cada scFv congelado em *ensemble* NVT, 5 ps com cada scFv livre em *ensemble* NVT e, por fim, 80 ps com cada scFv livre no *ensemble* NPT. Em seguida, os sistemas scFv de *linker* curto e scFv de *linker* longo foram submetidos à 600 ns de simulação de DM em *ensemble* NPT.

Quadro 9 - Quantidade de íons Na⁺ e Cl⁻ na concentração 0.15 M, bem como de moléculas de água adicionadas em cada sistema scFv-água

Modelo de scFv	Na ⁺	Cl ⁻	H ₂ O
scFv curto	84	85	29832
scFv longo	88	89	31471

Fonte: elaborada pela autora (2024).

Após à simulação de DM dos dois sistemas scFv-água as trajetórias obtidas foram analisadas com a utilização de ferramentas do pacote *GROMACS* versão 2018.3, como *gmx rms*, *gmx gyrate*, *gmx energy*, *gmx cluster*.

6.2.2 Montagem e simulação de DM do sistema contendo os domínios extracelulares D6-D7 glicosilados em água

O CD22 foi inserido numa caixa dodecaédrica com 2.1 nm de distância entre a extremidade da proteína e as arestas em todos os eixos. Para se obter o estado mínimo de energia realizou-se uma minimização de energia utilizando o seguinte

protocolo: 5000 passos de minimização de energia para otimização da geometria e redução de conflitos estereoquímicos, por meio do algoritmo *steepest descent*. Foram adicionadas 73121 moléculas de água; e 207 íons Na⁺ e 207 íons Cl⁻ na concentração 0.15 M para manter a neutralidade das cargas. Posteriormente, com a finalidade de equilibrar a água e os íons em torno dos domínios D6-D7 glicosilados de CD22, foi realizada a etapa de termalização. Essa etapa teve o seguinte protocolo: 215 ps com proteína congelada em *ensemble* NVT, 215 ps com a proteína sem os *loops* modelados congelada em *ensemble* NVT, 215 ps com a proteína livre em *ensemble* NVT, 215 ps com a proteína considerando os *loops* modelados com movimentação restrita em *ensemble* NPT e, por fim, 200 ps com a proteína livre em *ensemble* NPT. Em seguida, o sistema foi submetido à 500 ns de simulação de DM em *ensemble* NPT. As trajetórias obtidas foram analisadas com a utilização de ferramentas do pacote GROMACS versão 2018.3, como *gmx rms*, *gmx gyrate* e *gmx cluster*.

6.2.3 Montagem e simulação de DM dos sistemas contendo o domínio transmembrana de CD22

Foram desenvolvidos dois sistemas com o domínio transmembrana inseridos em 2 modelos de membranas fosfolipídicas, usando o programa *CHARMM-GUI* tendo as dimensões, 12 x 12 nm e 16 x 16 nm no eixo xy (plano da membrana). O eixo z foi dimensionado a inserir 2,5 nm nas extremidades da proteína abaixo e acima da bicamada de fosfolípido. Foram adicionados 50 íons Na⁺ e 50 íons Cl⁻ na concentração de 0.10 M para membrana 12 x 12 nm; e 64 íons Na⁺ e 64 íons Cl⁻ na concentração 0.10 M para membrana 16 x 16 nm. Em seguida os sistemas seguiram etapas de minimização, utilizando o algoritmo *steepest descent* por 10000 passos em cada etapa, mantendo-se a posição da cadeia principal da proteína e da membrana restritas. Em seguida, foram realizadas simulações curtas, escalonando o tempo de integração, em NVT durante 22,5 ps, ainda com a cadeia principal da proteína e da membrana restritas, depois 225 ps em NPT, reduzindo a força de restrição na cadeia principal da proteína e da membrana, e, por fim, 200 ps em NPT com a proteína e membrana livres. Após as etapas de simulações curtas, os sistemas transmembrana-membrana-água foram simulados durante 200 ns para equilibrar em água.

As análises foram conduzidas com ferramentas do pacote GROMACS versão 2018.3, como *gmx rms* e *gmx gyrate*. A estrutura final da simulação do sistema transmembrana-membrana-água de 12 x 12 nm foi utilizada para a montagem e

simulação do sistema CD22-membrana-água e scFv/CD22-longo-membrana água. Já a estrutura final da simulação do sistema transmembrana-membrana-água de 16 x 16 nm foi utilizada para montagem e simulação do sistema scFv/CD22-curto-membrana-água.

6.2.4 Montagem e simulação de DM do CD22 em membrana

Após a simulação do CD22 e do domínio transmembrana inserido em um modelo de membrana, as duas estruturas finais de cada simulação foram conectadas através da ligação entre o resíduo arginina no C-terminal do CD22 e o resíduo valina no N-terminal do transmembrana no programa *Pymol 2.4.0 (DeLano Scientific)*, cuja distância de ligação e ângulos foram estabelecidos usando a etapa de minimização de energia com algoritmo *steepest descent*.

Em seguida, com o objetivo de manter o domínio transmembrana do CD22 recém minimizado centralizado e orientado em relação ao eixo z na membrana anteriormente equilibrada em solução aquosa foi realizada a sobreposição do domínio transmembrana de CD22 recém minimizada ao domínio transmembrana inserido em membrana da simulação, utilizando o programa *VMD (Visual Molecular Dynamics)* versão 1.9.3 (HUMPHREY; DALKE; SCHULTEN, 1996). Dessa forma, obteve-se o sistema CD22-membrana. Posteriormente, foi criada uma caixa de simulação com dimensões 12 x 12 x 18 nm. A caixa de simulação com o sistema CD22-membrana foi preenchida com 71693 moléculas de água; e 133 íons Na⁺ e 133 íons Cl⁻ na concentração de 0.10 M. Inicialmente, o sistema foi minimizado utilizando o algoritmo *steepest descent* por 20000 passos com as posições da cadeia principal da proteína, glicosilações e membrana restritas. Em seguida, foi termalizado por 22,5 ps em NVT e por 55 ps em NPT reduzindo a força de restrição sobre a cadeia principal da proteína, glicosilações e da membrana. Após a termalização, o sistema foi submetido à simulação de DM em *ensemble* NPT durante 1000 ns.

Por fim, as trajetórias obtidas foram analisadas com a utilização de ferramentas do pacote *GROMACS* versão 2018.3, como *gmx rms* e *gmx cluster*.

6.2.5 Montagem e DM dos sistemas scFv/CD22 a partir do Docking Molecular

O CD22 nas interfaces selecionadas pelos *Dockings* de scFv/CD22 de *linker* curto e longo foram submetidas ao *CHARMM-GUI* para recolocar as árvores de

glicosilação, conforme demonstrado na Figura 14. Posteriormente, os sistemas scFv/CD22 de *linker* curto e longo foram adicionados em caixas dodecaédricas de simulação com 1,8 e 2,0 nm de distância, entre as extremidades da proteína e as arestas em todos os eixos, respectivamente, sendo que as quantidades de moléculas de água e íons estão descritas no Quadro 10. A etapa de minimização, para cada um dos sistemas scFv/CD22, foi realizada utilizando o seguinte protocolo: 10000 passos de minimização de energia aplicando uma força de restrição sobre a cadeia principal da proteína e das glicosilações, a fim de otimizar a geometria e reduzir os conflitos estereoquímicos, por meio do algoritmo *steepest descent*. Posteriormente, com a finalidade de equilibrar a solução em torno de cada modelo de scFv/CD22, foi realizada a etapa de termalização em cada um dos sistemas scFv/CD22. Essa etapa teve o seguinte protocolo: 32,5 ps com a cadeia principal da proteína e das glicosilações restritas em *ensemble* NVT e 65 ps com a proteína e glicosilações livres em *ensemble* NPT. Em seguida, cada sistema scFv-CD22 foi submetido à 600 ns de simulação de DM em *ensemble* NPT.

Quadro 10 - Quantidade de íons Na⁺ e Cl⁻ na concentração 0.10 M, bem como de moléculas de água adicionadas em cada sistema scFv-CD22.

Modelo de scFv/CD22	Na ⁺	Cl ⁻	água
scFv curto/CD22	153	156	80788
scFv longo/CD22	146	149	75054

Fonte: elaborada pela autora (2024).

Após à simulação de DM dos sistemas scFv-CD22 de *linker* curto e longo as trajetórias obtidas foram analisadas com a utilização de ferramentas do pacote GROMACS versão 2018.3, como *gmx rms*, *gmx energy*. Também foi utilizado o programa *gmx_MMPBSA* (VALDÉS-TRESANCO *et al.*, 2021) para cálculo de energia livre de ligação entre os scFvs e o CD22, utilizando como *script* o Apêndice A.

6.2.6 Montagem e simulação de DM dos sistemas scFv/CD22 com estruturas advindas da sobreposição de simulações anteriores à cristalográfica

Após a simulação de Dinâmica Molecular dos dois modelos de scFv-água; e do modelo de CD22 em água, a estrutura final presente no *cluster* mais representativo de cada sistema foi selecionada para a montagem de dois sistemas scFv/CD22,

obtendo-se assim o sistema scFv/CD22, com scFv de *linker* curto; e o sistema scFv/CD22, com scFv de *linker* longo. Sendo assim, utilizando o programa *VMD* (*Visual Molecular Dynamics*) versão 1.9.3 (HUMPHREY; DALKE; SCHULTEN, 1996), a estrutura final do CD22, equilibrada em água, foi sobreposta a molécula CD22 obtida experimentalmente; e, os dois modelos finais de scFvs equilibrados em água foram sobrepostos aos domínios VH e VL presentes na porção Fab do anticorpo M971. Dessa forma, os scFvs obtiveram o melhor posicionamento com o CD22 e respeitaram a conformação do sítio de interação, conforme descrito por Ereno-Orbea e colaboradores (2021). Em seguida, nos dois sistemas scFv/CD22, realizou-se a mudança do estado de protonação das histidinas no CD22, nas posições 122 e 136 para estarem protonadas no nitrogênio epsilon e delta (HSP) usando o *CHARMM-GUI*, assim como foram reinseridas as árvores de glicosilação, conforme demonstrado na Figura 14. Posteriormente, as proteínas foram alocadas em caixas dodecaédricas de simulação com 1,8 e 2,0 nm de distância entre a extremidade da proteína e as arestas em todos os eixos, respectivamente; e cuja as quantidades de moléculas de água e íons estão descritas no Quadro 11. A etapa de minimização para cada um dos sistemas scFv/CD22 foi realizada utilizando o seguinte protocolo: 10000 passos de minimização de energia aplicando uma força de restrição para a proteína e as glicosilações, a fim de otimizar a geometria e reduzir os conflitos estereoquímicos, por meio do algoritmo *steepest descent*. Posteriormente, com a finalidade de otimizar as interações com o solvente em torno de cada modelo de scFv/CD22, foi realizada a etapa de termalização em cada um dos sistemas scFv/CD22. Essa etapa teve o seguinte protocolo: 32,5 ps com a cadeia principal da proteína e das glicosilações restritas em *ensemble* NVT e 65 ps com a proteína e glicosilações livres em *ensemble* NPT. Em seguida, cada sistema scFv/CD22 foi submetido à 600 ns de simulação de DM em *ensemble* NPT.

Quadro 11 - Quantidade de íons Na⁺ e Cl⁻ na concentração 0.10 M, bem como de moléculas de água adicionadas em cada sistema scFv-CD22.

Modelo de scFv/CD22	Na ⁺	Cl ⁻	água
scFv curto/CD22	148	151	78348
scFv longo/CD22	159	162	83464

Fonte: elaborada pela autora (2024).

Após à simulação de DM dos sistemas scFv-CD22 de *linker* curto e longo as trajetórias obtidas foram analisadas com a utilização de ferramentas do pacote GROMACS versão 2018.3, como *gmx rms*, *gmx energy*. Também foi utilizado o programa *gmx_MMPBSA* (VALDÉS-TRESANCO *et al.*, 2021) para cálculo de energia livre de ligação entre os scFvs e o CD22, utilizando como *script* o Apêndice A.

6.2.7 Montagem e simulação de DM dos sistemas scFv/CD22 com estruturas advindas da cristalografia

Outros dois sistemas scFv-CD22 foram montados, a partir da sobreposição do CD22 modelado ao obtido experimentalmente (considerando VH e VL do Fab M971 da estrutura cristalográfica; PDB cód. 7O52); e, sobreposição dos modelos de scFv, curto e longo, modelados no *Modeller 10.1* aos domínios VH e VL obtidos experimentalmente (ERENO-ORBEA *et al.*, 2021), utilizando o programa *VMD (Visual Molecular Dynamics)* versão 1.9.3 (HUMPHREY; DALKE; SCHULTEN, 1996). Dessa forma, os modelos de scFv/CD22 obtidos tiveram a posição de interação conservada, conforme descrito por Ereno-Orbea e colaboradores (2021).

Nos dois novos sistemas scFv/CD22 montados foi promovida a mudança do estado de protonação das histidinas da molécula de CD22 nas posições 122 e 136 para estarem protonadas no nitrogênio epsilon e delta (HSP) usando *CHARMM-GUI*, que também foi usado para reinserir as árvores de glicosilação conforme demonstrado na Figura 14. Posteriormente, os sistemas foram adicionados em caixas cúbicas de simulação com 17 nm de aresta; e cujo as quantidades de moléculas de água e de íons então descritas no Quadro 12. Os protocolos de minimização e de termalização utilizados para esses dois novos sistemas foi o mesmo descrito para os dois sistemas scFv/CD22 citados anteriormente. Em seguida, cada sistema scFv/CD22 foi submetido à 400 ns de simulação de DM em *ensemble* NPT.

Quadro 12 - Quantidade de íons Na⁺ e Cl⁻ na concentração 0.05 M, bem como de moléculas de água adicionadas em cada sistema scFv-CD22.

Modelo de scFv/CD22	Na ⁺	Cl ⁻	água
scFv curto/CD22	148	151	158403
scFv longo/CD22	148	151	158385

Fonte: elaborada pela autora (2024).

Após à simulação de DM dos sistemas scFv/CD22 de *linker* curto e longo as trajetórias obtidas foram analisadas com a utilização de ferramentas do pacote *GROMACS versão 2018.3*, como *gmx rms*, *gmx energy* e *gmx cluster*. Também foi utilizado o programa *gmx_MMPBSA* (VALDÉS-TRESANCO *et al.*, 2021) para cálculo de energia livre de ligação entre os scFvs e o CD22, utilizando como *script* o Apêndice A.

6.2.8 Montagem e simulação de DM dos sistemas scFv/CD22 em membrana

Após a simulação dos sistemas scFv/CD22 de *linker* curto e longo advindos da sobreposição das estruturas modeladas à cristalográfica; e do domínio transmembrana de CD22 inserido em um modelo de membrana de 16 x 16 nm, as estruturas de domínio D6-D7 glicosilada do complexo scFv/CD22 curto e de domínio transmembrana finais de cada simulação foram conectadas através da ligação entre o resíduo arginina no C-terminal do CD22 e o resíduo valina no N-terminal do transmembrana no programa *Pymol 2.4.0 (DeLano Scientific)*, cuja distância de ligação e ângulos foram estabelecidos usando a etapa de minimização de energia com algoritmo *steepest descent*, formando assim o CD22-transmembrana.

Em seguida, com o objetivo de manter centralizado e orientado em relação ao eixo z na membrana o domínio transmembrana do CD22-transmembrana recém minimizado foi realizada a sobreposição do domínio transmembrana do CD22-transmembrana, recém minimizada, ao domínio transmembrana inserido em membrana 16 x 16 nm. Posteriormente, utilizando o *VMD versão 1.9.3* foi realizada a sobreposição da última configuração da simulação do complexo scFv/CD22 de *linker* curto e longo ao CD22-transmembrana já orientado na membrana 16 x 16 nm apenas para conservar a posição de interação dos scFvs. Dessa forma, obteve-se os sistemas scFv/CD22 de *linker* curto e longo em membrana.

A caixa de simulação para os sistemas scFv/CD22 de *linker* curto e longo tiveram dimensões de 15 x 15 x 19 nm. A caixa de simulação no sistema de *linker* curto foi preenchida com 113057 moléculas de água e 267 íons Na⁺ e 270 íons Cl⁻ na concentração de 0.10 M. Já a caixa de simulação no sistema de *linker* longo foi preenchida com 113018 moléculas de água, 267 íons Na⁺ e 270 íons Cl⁻ na concentração de 0.10 M. Os sistemas foram minimizados utilizando o algoritmo *steepest descent* por 20000 passos com as posições da cadeia principal da proteína,

glicosilações e da membrana restritas. Em seguida, foram termalizados por 22,5 ps em NVT e por 55 ps em NPT reduzindo a força de restrição sobre a cadeia principal da proteína, glicosilações e da membrana. Após a termalização, os sistemas scFv/CD22 foram submetidos à simulação de DM em *ensemble* NPT durante 400 ns, em triplicata, partindo de velocidades iniciais diferentes.

Por fim, as todas as trajetórias obtidas foram analisadas com a utilização de ferramentas do pacote *GROMACS* versão 2018.3, como *gmx rms* e *gmx energy*. Foi utilizado o programa *gmx_MMPBSA* (VALDÉS-TRESANCO *et al.*, 2021) para cálculo de energia livre de ligação entre os scFvs e o CD22, utilizando como *script* o Apêndice B.

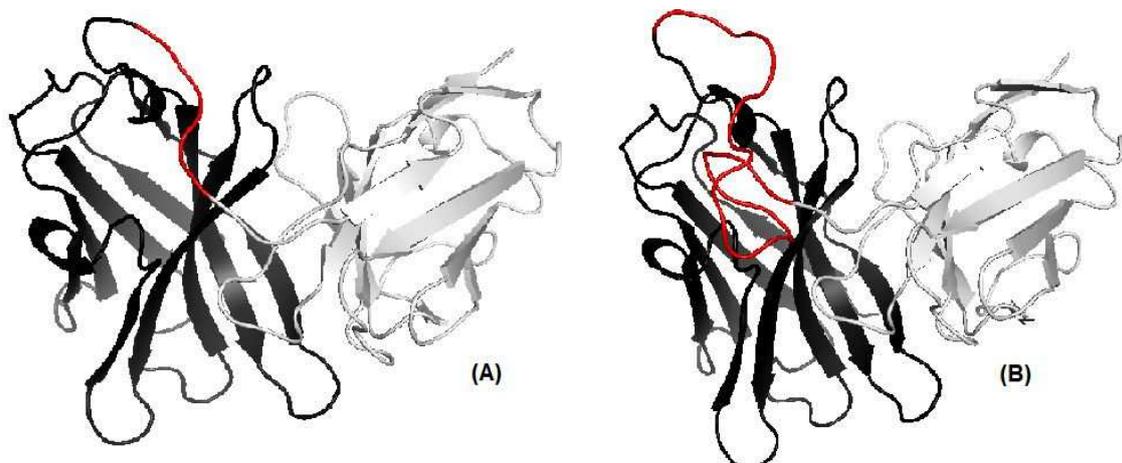
7. RESULTADOS

7.1 Modelagem Molecular

7.1.1 Modelagem e validação dos scFv de M971 anti-CD22

As Figuras 15A e 15B representam os modelos de scFv curto e longo, respectivamente, obtidos a partir da união do *linker* curto aos domínios VH e VL, derivados da porção Fab do anticorpo humano M971.

Figura 15 - Estrutura tridimensional dos modelos de scFv curto e longo na orientação VH-VL.



Fonte: elaborada pela autora (2024). (A) scFv de linker curto. (B) scFv de linker longo. Identificação do domínio VH (em preto), o domínio VL (em cinza) e o Linker (em vermelho) para ambos scFvs.

Na Tabela 1 e 2, estão representados os resultados de validação obtidos por

meio do programa *MolProbity* para cada modelo tridimensional de scFv (curto e longo) antes e após a etapa de minimização de energia, como os valores de *clash score* em percentil, de *MolProbity score*, de rotâmeros ruins, de ligações e ângulos em conformações não desejadas (Anexos A, B, C e D). Para ambos os modelos de scFv, após a etapa de minimização de energia, tem-se valores de *clash score* e de *MolProbity score* próximos à 100° e 0, respectivamente, indicando que são modelos tridimensionais de boa qualidade. Conforme observado nas Tabelas 1 e 2, os valores de avaliação de cada parâmetro analisado no programa *MolProbity* para os modelos de scFv em estudo se alteram, uma vez que durante a etapa de minimização de energia os modelos de scFv (curto e longo) estão em solução aquosa, possibilitando o movimento de seus ângulos e ligações, assim melhorando a qualidade estereoquímica das estruturas tridimensionais.

Tabela 1 - Resultados obtidos por meio do programa *MolProbity* para o modelo de scFv com *linker* curto obtido pela modelagem molecular, antes e após a etapa de minimização de energia.

Modelo	Clash score (percentil)	MolProbity score	Rotâmeros ruins	Ângulos ruins	Ligações ruins	Outliers
scFv antes*	3°	2.66	3	30	6	3
scFv após**	96°	1.77	3	10	2	2

Fonte: elaborada pela autora (2024).

*scFv antes da etapa de minimização de energia

** scFv após da etapa de minimização de energia

Tabela 2 - Resultados obtidos por meio do programa *MolProbity* para o modelo de scFv com *linker* longo obtido pela modelagem molecular, antes e após a etapa de minimização de energia.

Modelo	Clash score (percentil)	MolProbity score	Rotâmeros ruins	Ângulos ruins	Ligações ruins	Outliers
scFv antes*	5°	2.58	3	28	6	2
scFv após**	99°	1.66	3	13	0	2

Fonte: elaborada pela autora (2024).

*scFv antes da etapa de minimização de energia

** scFv após da etapa de minimização de energia

Os dados de validação de cada modelo tridimensional de scFv (curto e longo)

no programa *PROCHECK* estão representados nos Anexos I, J, K e L. O número de resíduos presentes em regiões mais favoráveis, regiões permitidas e em regiões não permitidas para cada modelo está apresentado nas Tabelas 3 e 4. Para o scFv com *linker* curto, antes da etapa de minimização de energia, 91,2% dos aminoácidos estavam em regiões muito favoráveis, 8,4% em regiões favoráveis, 0,5% em regiões pouco favoráveis, representado pelo resíduo de Ser124. Após a etapa de minimização de energia, para esse mesmo modelo de scFv, 89,7% dos aminoácidos encontraram-se em regiões muito favoráveis, 9,8% em regiões favoráveis, 0,5% em regiões pouco favoráveis, representado pelo resíduo de Ser124.

Já para o scFv com o *linker* longo, antes da minimização de energia, os resultados obtidos por meio do programa *PROCHECK* revelam que 90,8% dos aminoácidos estavam em regiões muito favoráveis, 9,2% em regiões favoráveis e 0% em regiões pouco favoráveis. Após a etapa de minimização de energia, para esse mesmo modelo de scFv, 87,9% dos aminoácidos encontraram-se em regiões muito favoráveis, 11,5% em regiões favoráveis, 0,5% em regiões pouco favoráveis, representado pelo resíduo de Trp174.

Todos esses resultados corroboram para a validação dos dois modelos tridimensionais de scFv (curto e longo), tendo em vista que mesmo antes e após a etapa de minimização de energia mais de 99% dos aminoácidos de ambos scFvs encontraram-se em regiões favoráveis ou permitidas.

Tabela 3 - Resultados obtidos por meio do programa *PROCHECK* para o modelo de scFv com *linker* curto obtido pela modelagem molecular, antes e após a etapa de minimização de energia.

Modelo	Regiões mais favoráveis (%)	Regiões permitidas (%)	Regiões não permitidas (%)	Resíduos em regiões não permitidas
scFv antes*	91,2	8,4	0,5	Ser124
scFv após**	89,7	9,8	0,5	Ser124

Fonte: elaborada pela autora (2024).

*scFv antes da etapa de minimização de energia

** scFv após da etapa de minimização de energia

Tabela 4 - Resultados obtidos por meio do programa *PROCHECK* para o modelo de scFv com *linker* longo obtido pela modelagem molecular, antes e após a etapa de minimização de energia.

Modelo	Regiões mais favoráveis (%)	Regiões permitidas (%)	Regiões não permitidas (%)	Resíduos em regiões não permitidas
scFv antes*	90,8	9,2	0	-
scFv após**	87,9	11,5	0,5	Trp174

Fonte: elaborada pela autora (2024).

*scFv antes da etapa de minimização de energia

** scFv após da etapa de minimização de energia

7.1.2 Modelagem e validação estrutural do CD22

A Figura 16 representa o modelo proteico completo de estrutura tridimensional de CD22, utilizando-se como molde a estrutura cristalográfica CD22 (PDB 7O52).

Figura 16 - Estrutura tridimensional do CD22, obtido a partir de modelagem por homologia.



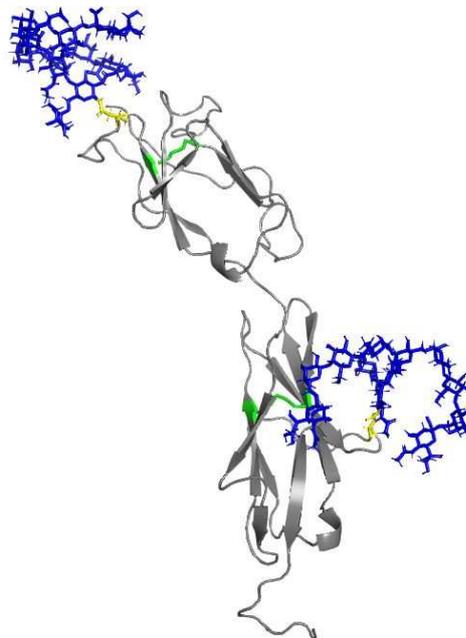
Fonte: elaborada pela autora (2024). Na coloração azul: *loop A*; verde: *loop B*; vermelho: *loop C*; e rosa: *loop D*. Nota: a identificação da sequência e numeração dos *loops* modelados foi mencionada anteriormente no Quadro 5.

A estrutura tridimensional modelada do CD22 foi submetida ao programa

MolProbity para avaliar os valores de *clash score* em percentil, de *MolProbity score*, de rotâmeros ruins, de ligações e ângulos em conformações não desejadas, como descrito. Conforme apresentado no Anexo E e Tabela 5. O modelo estrutural de CD22 apresentou *clash score* e de *MolProbity score* 3° e 2.71, respectivamente. Foi constatado 2 rotâmeros ruins, 22 ângulos ruins e 5 resíduos *outliers*. Não foi identificada a presença de ligações ruins. O número de resíduos em regiões favoráveis/permitidas e proibitivas, apresentado no Anexo M e Tabela 6, 90,5% dos aminoácidos encontraram-se em regiões muito favoráveis, 8,2% em regiões favoráveis, 1,3% em regiões pouco favoráveis, representado pelos resíduos de Lys10 e Gln36.

Após a primeira avaliação da qualidade da estrutura tridimensional do modelo de CD22, ligações dissulfeto e glicosilações na posição N70 e N130 foram inseridas, de acordo com o esquema da Figura 17. Dessa forma, obteve-se o modelo completo proteico de CD22 glicosilado, conforme representado pela Figura 18.

Figura 17 - Estrutura tridimensional do modelo completo de CD22 glicosilado, obtido a partir da adição das ligações dissulfeto e glicosilações previstas nos resíduos N70 e N130, utilizando o programa CHARMM-GUI.



Fonte: elaborada pela autora (2024). Na coloração azul e verde, glicosilações e ligações dissulfeto adicionadas no CHARMM-GUI. Em amarelo é destacada a ligação entre a proteína e os carboidratos.

Após a obtenção do modelo completo de CD22 glicosilado foi realizada a etapa de validação da estrutura tridimensional obtida pelo CHARMM-GUI. Inicialmente, foi

utilizado o programa *MolProbity* e *PROCHECK*. Conforme apresentado no anexo F e N; e tabelas 5 e 6, os resultados obtidos foram iguais aos já descritos anteriormente.

Após a segunda avaliação da qualidade da estrutura tridimensional do modelo de CD22, realizou-se a etapa de minimização de energia. Nesse caso, o modelo completo de CD22 glicosilado, inserido em uma caixa de simulação dodecaédrica, teve sua estrutura tridimensional minimizada, a fim de reduzir os conflitos estereoquímicos. Após a obtenção do modelo completo de CD22 glicosilado minimizado foi realizada a validação da estrutura tridimensional. Inicialmente, foi utilizado o programa *MolProbity*. Conforme apresentado no Anexo G e Tabela 5, para esse modelo de CD22 tem-se valores de *clash score* e de *MolProbity score* 99° e 1.66, respectivamente. Também pode-se constatar a presença de 4 rotâmeros ruins, 14 ângulos ruins e 5 resíduos *outliers*. Não foi identificada a presença de ligações ruins para esse modelo de CD22.

Os valores dos parâmetros obtidos pelo programa *MolProbity* após a etapa de minimização de energia foram diferentes dos que obtidos pela modelagem molecular, usando *Modeller 10.1*; e *CHARMM-GUI*, uma vez que durante a etapa de minimização de energia o modelo completo de CD22 glicosilado está em solução aquosa, possibilitando o movimento de seus ângulos e ligações, assim melhorando a qualidade estereoquímica da estrutura tridimensional.

Por fim, submeteu-se o modelo completo de CD22 glicosilado minimizado ao programa *PROCHECK* para averiguar o número de resíduos em regiões favoráveis/permitidas e proibitivas. Conforme apresentado no Anexo O e Tabela 6, para esse modelo de CD22 90,5% dos aminoácidos encontraram-se em regiões muito favoráveis, 8,8% em regiões favoráveis, 0,6% em regiões pouco favoráveis, representado pelo resíduo de Lys10.

Todos esses resultados corroboram para a validação do modelo completo de CD22 glicosilado, tendo em vista que mesmo antes e após a etapa de minimização de energia mais de 99% dos aminoácidos encontraram-se em regiões favoráveis ou permitidas.

Tabela 5 - Resultados obtidos por meio do programa *MolProbity* para o modelo de CD22.

Modelo	Clash score (percentil)	MolProbity score	Rotâmeros ruins	Ângulos ruins	Outliers
CD22 modelagem*	3°	2.72	2	22	5
CD22 CHARMM-GUI**	2°	2.77	2	16	4
CD22 minimizado***	99°	1.66	4	14	5

Fonte: elaborada pela autora (2024).

* CD22 obtido após a modelagem dos *loops* faltantes

** CD22 após inserção das ligações dissulfeto e glicosilações

*** CD22 após a etapa de minimização de energia

Tabela 6 - Resultados obtidos por meio do programa *PROCHECK* para o modelo de CD22.

Modelo	Regiões mais favoráveis (%)	Regiões permitidas (%)	Regiões não permitidas (%)	Resíduos em regiões não permitidas
CD22 modelagem*	90,5	8,2	1,3	Lys10 e Gln36
CD22 CHARMM-GUI**	90,5	8,2	1,3	Lys10 e Gln36
CD22 minimizado***	90,5	8,8	0,6	Lys10

Fonte: elaborada pela autora (2024).

Fonte: elaborada pela autora (2024).

* CD22 obtido após a modelagem dos *loops* faltantes

** CD22 após inserção das ligações dissulfeto e glicosilações

*** CD22 após a etapa de minimização de energia

7.1.3 Modelagem e validação do domínio transmembrana de CD22

A Figura 18 representa o modelo de domínio transmembrana obtido através do programa *I-TASSER*, utilizando-se como sequência alvo a originária do *UniProt* (cód. P20273).

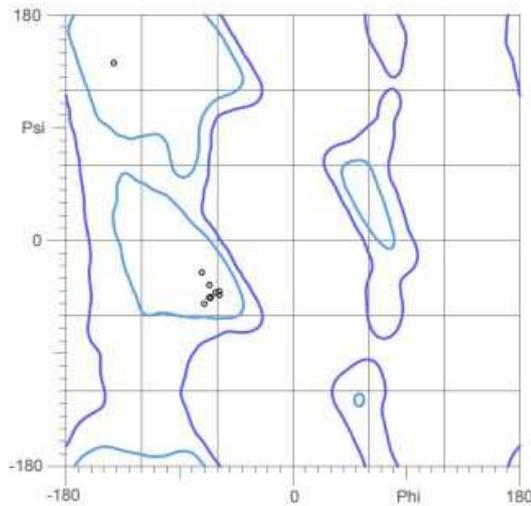
Figura 18 - Estrutura tridimensional do modelo de domínio transmembrana do obtido através do programa I-TASSER.



Fonte: elaborada pela autora (2024). Representação em *cartoon* de uma alfa hélice, na cor roxa.

Após a obtenção do modelo de domínio transmembrana foi realizada a etapa de validação da estrutura tridimensional modelada. Foi utilizado o programa *MolProbity* para avaliar os valores de *clash score* em *percentil*, de *MolProbity score*, de rotâmeros ruins, de ligações e ângulos em conformações não desejadas. Conforme apresentado no Anexo H, para esse modelo de domínio transmembrana tem-se valores de *clash score* e de *MolProbity score* 96° e 1.83, respectivamente. Também pode-se constatar a presença de 1 rotâmero insatisfatório. Não foi identificada a presença de ligações ruins, ângulos ruins e resíduos *outliers*. Como pode ser observado pela Figura 19, o modelo de domínio transmembrana obtido no programa *I-TASSER* não possui aminoácidos fora de regiões permitidas no gráfico de *Ramachandran*.

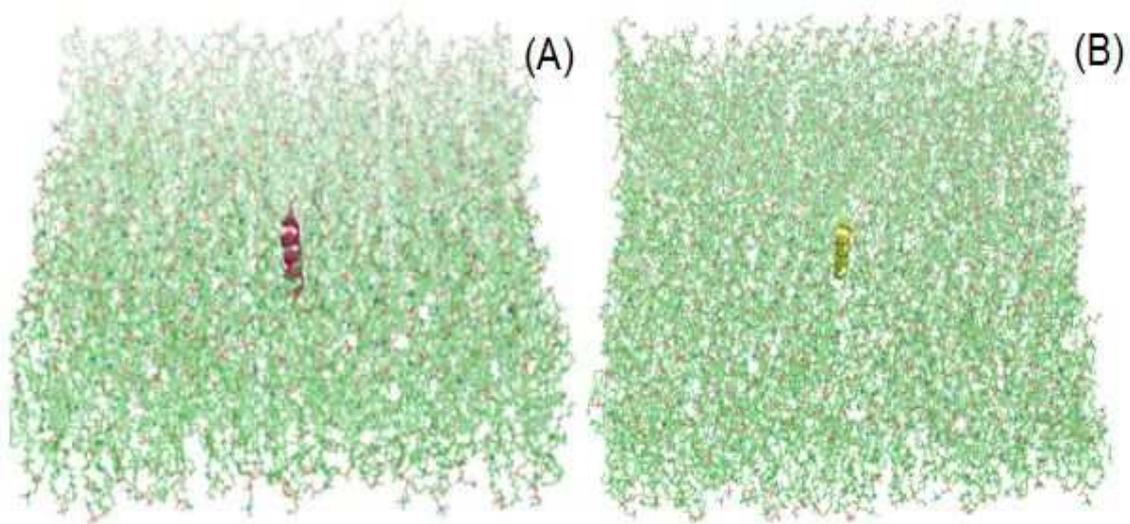
Figura 19 - Gráfico Ramachandran gerado pelo programa MolProbity para o modelo do domínio transmembrana obtido pelo programa I-TASSER.



Fonte: elaborada pela autora (2024).

Após a primeira avaliação da qualidade da estrutura tridimensional do modelo de domínio transmembrana, realizou-se a sua submissão no programa *CHARMM-GUI* para sua inserção em dois modelos de membrana fosfolipídica de mesma composição com dimensões de 12 x 12 nm e 16 x 16 nm. Dessa forma, obteve-se os modelos de domínio transmembrana inserido nas membranas conforme as Figuras 20A e 20B.

Figura 20 - Estrutura tridimensional do modelo de domínio transmembrana inserido na membrana.



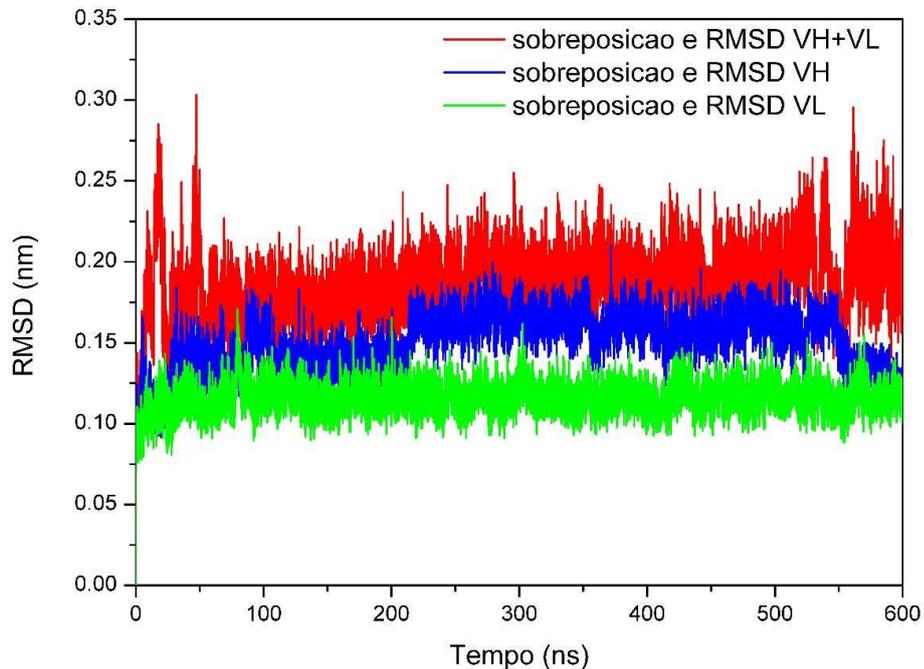
Fonte: elaborada pela autora (2024). (A) inserido na membrana fosfolipídica (12 x 12 nm) obtido pelo *CHARMM-GUI*. (B) inserido na membrana fosfolipídica (16 x 16 nm) obtido pelo *CHARMM-GUI*.

7.2 Simulação de Dinâmica Molecular e análises

7.2.1 Dinâmica Molecular do scFv curto

As variações estruturais dos domínios VH+VL, VH e VL do scFv curto foram avaliadas ao longo da simulação de 600 ns por meio da análise dos perfis de RMSD (Figura 21). Os RMSD para os scFv estão considerando sempre domínios VH e VL e VH+VL, sem a presença do linker flexível de 5 resíduos, pois a sua presença nos cálculos proverá RMSD alto, sobrepondo os dos demais domínios citados. Observa-se nos perfis dos domínios VH+VL, VH e VL que o RMSD está abaixo de $\sim 0,20$ nm, foi considerado que as estruturas alcançaram um equilíbrio estrutural (teq) a partir de 50 ns de simulação. Nota-se também nos perfis, ~ 550 ns uma redução abrupta no RMSD, tanto para VH quanto para VH+VL. Os RMSDs são apresentados na Tabela 7, as médias são $0,19 \pm 0,02$ nm para VH+VL; $0,15 \pm 0,01$ nm para VH e $0,12 \pm 0,01$ nm para VL. Os RMSDs médios do VH+VL e do VH são próximos entre si e superiores aos de VL, indicando que as mudanças na conformação do VH+VL são decorrentes de alterações estruturais do domínio VH ou movimentação entre eles.

Figura 21 - Perfil de RMSD dos C α dos componentes do scFv curto em solução aquosa durante 600 ns de simulação.



Fonte: elaborada pela autora (2024). Curva de RMSD entre as configurações estruturais do VH+VL (sobreposição e cálculo de RMSD de VH+VL), VH (sobreposição e cálculo de RMSD de VH) e VL (sobreposição e cálculo de RMSD de VL), comparadas com a estrutura de referência inicial.

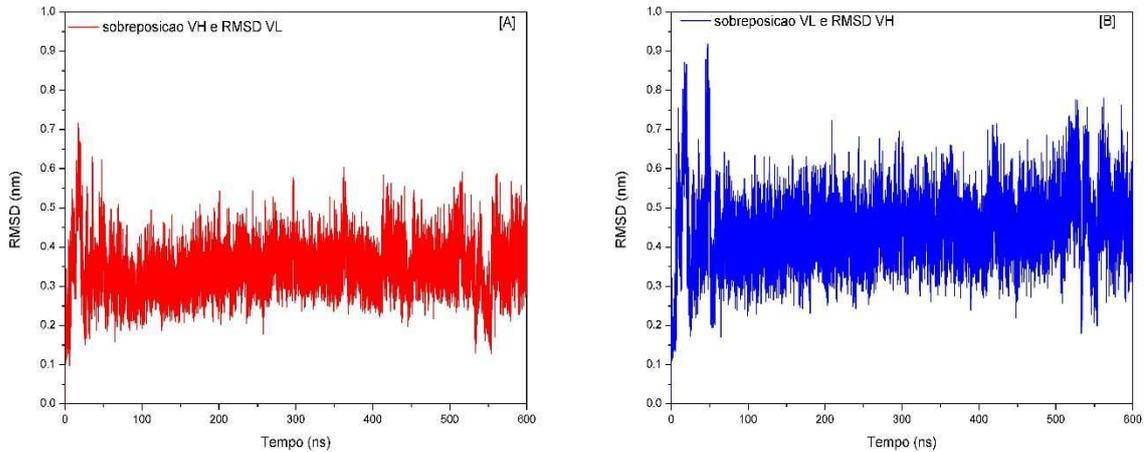
Tabela 7 - RMSDs médios calculados para VH+VL, VH e VL a partir do $t_{eq} = 50$ ns.

Estrutura	RMSD médio (nm)
VH+VL	$0,19 \pm 0,02$
VH	$0,15 \pm 0,01$
VL	$0,12 \pm 0,01$

Fonte: elaborada pela autora (2024).

Realizou-se análise dos perfis de RMSD de VL tendo como referência VH para sobreposição e vice-versa, com objetivo de investigar se existe movimentação relativa entre VH e VL (Figura 22). Os RMSDs médios do VH com sobreposição de VL, e do VL com sobreposição do VH, foram $0,44 \pm 0,08$ nm e $0,34 \pm 0,06$ nm, como visto na Tabela 8. As curvas de RMSD os altos desvios relevam que há movimentação rápida de orientação entre VH e VL. Da mesma forma que observado para RMSD (Figura 21) há uma redução abrupta do RMSD, no perfil de VH ~ 550 ns, com uma amplitude considerável (entre 0,2-0,7 nm em curto intervalo de tempo), ratificando que foi um momento de maior alteração estrutural de $\sim 0,15$ para 0,12 nm), como visto na Figura 22, considera uma discreta alteração estrutural.

Figura 22 - Curvas de RMSD relativo entre as estruturas tridimensionais dos domínios VH e VL do scFv curto submetido à 600 ns de simulação em solução aquosa.



Fonte: elaborada pela autora (2024). (A) sobreposição VH e RMSD VL. (B) sobreposição VL e RMSD VH.

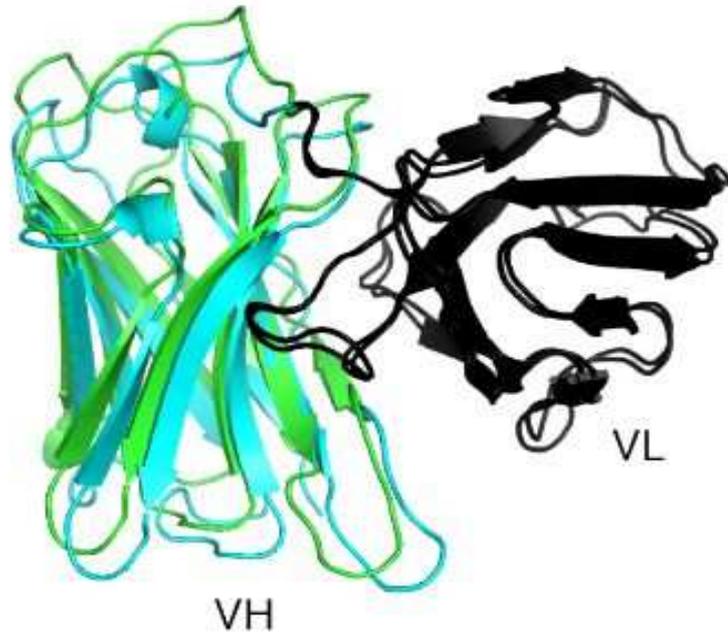
Tabela 8 - RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 50 ns.

Estrutura	RMSD médio (nm)
VH Ref VL	$0,44 \pm 0,08$
VL Ref VH	$0,34 \pm 0,06$

Fonte: elaborada pela autora (2024). VH Ref VL = sobreposição de VL e RMSD de VH e VL Ref VH = sobreposição de VH e RMSD de VL.

Foram coletadas estruturas do scFv em 50 ns e 450 ns sobrepondo o domínio VL de ambas como mostrado na Figura 23. Nota-se que há um deslocamento da estrutura de VH em relação a de VL, evidenciando a mudança de orientação entre VH e VL, como indicado pelo RMSD.

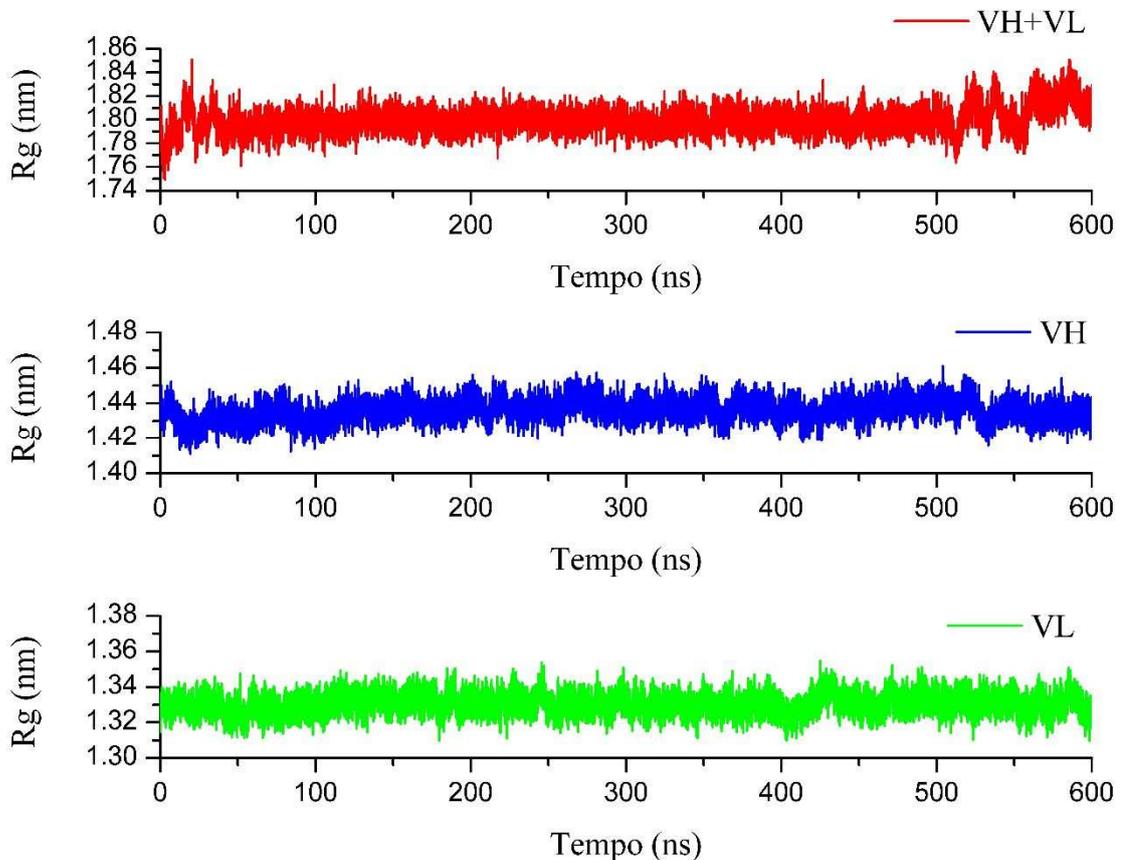
Figura 23 - Conformações do VH+VL coletadas em 50 ns e em 450 ns de simulação evidenciando as diferenças estruturais no domínio VH que foram observadas no gráfico do perfil de RMSD relativo.



Fonte: elaborada pela autora (2024). Preto: domínio VL das duas estruturas sobrepostas. Verde: domínio VH em 50 ns. Azul: domínio VH em 450 ns.

A Figura 24 mostra a curva de Raio de giro (Rg) de VH+VL, VH e VL. Os resultados apresentados demonstram que o enovelamento proteico se mantém estável durante os 600 ns de simulação, como mostrado na Tabela 9, onde os desvios das medias estão em 0,01 nm.

Figura 24 - Curvas de Rg dos domínios VH+VL, VH e VL do scFv de linker curto em solução aquosa ao longo de 600 ns de simulação.



Fonte: elaborada pela autora (2024).

Tabela 9 - Raio de Giro médio calculados para a estrutura VH+VL, VH e VL do scFv de linker curto a partir de 50 ns de simulação.

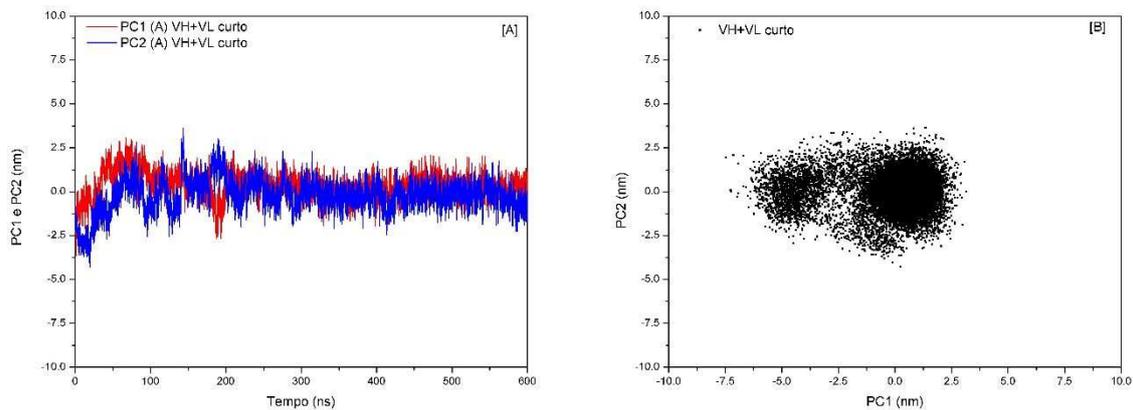
Estrutura	Raio de Giro médio (nm)
VH+VL	$1,80 \pm 0,01$
VH	$1,44 \pm 0,01$
VL	$1,33 \pm 0,01$

Fonte: elaborada pela autora (2024).

A Figura 25 representa os resultados obtidos por meio da análise de PCA do VH+VL de *linker* curto. Com base na Figura 25A, observa-se que os perfis apresentam PC1 e PC2 oscilando ao redor da amplitude de ~ -2.0 a ~ 2.0 nm até ~ 200 ns. A partir de 200 ns, as projeções PC1 e PC2 são reduzidas, chegando a praticamente zero. Pode-se observar ainda que na Figura 25B que o VH+VL de *linker* curto apresenta pouca flexibilidade estrutural durante a simulação, uma vez que suas projeções nos subespaços comuns definidos por PC1/PC2 possuem duas regiões, uma bem mais intensa que a outra. A região menos intensa, deslocada para esquerda, é decorrente

da alteração de PC1 e PC2 no começo da simulação. Esse resultado, em conjunto com o observado nos perfis de RMSD, ratificam a mudança estrutural em VH+VL em relação a configuração inicial, mas as análises de PCA mostram maior estabilidade após 200 ns.

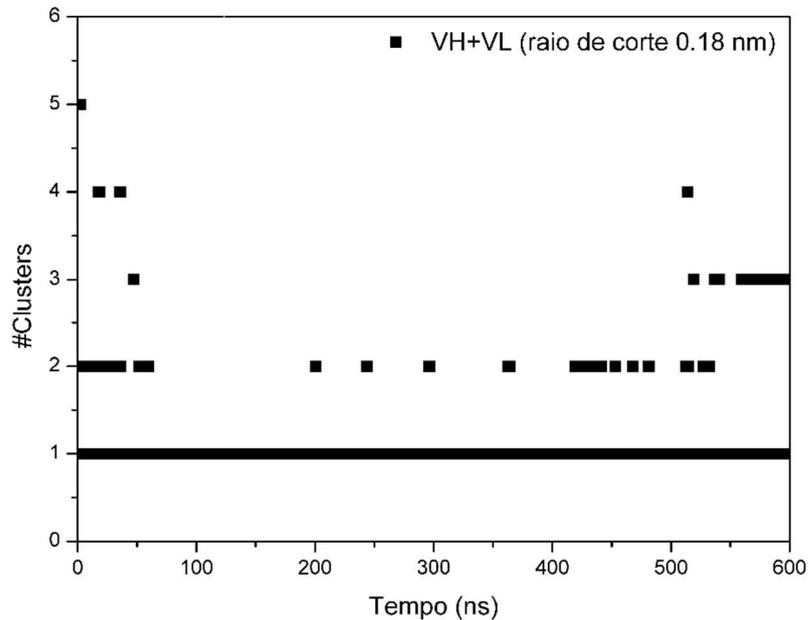
Figura 25 - Curvas de Análise de Componentes Principais.(A) PC1 e PC2 em função do tempo durante 600 ns de simulação. (B) Projeções da simulação do scFv de linker curto nos subespaços comuns definidos por PC1/PC2.



Fonte: elaborada pela autora (2024).

Para a análise *cluster* foi utilizado um raio de corte de 0.18 nm. Os dados da clusterização (Figura 26) revelam que ocorrem cinco grupos de estruturas, sendo o grupo de estruturas referente ao *#Cluster 1* o mais representativo, preponderante desde o início da simulação. Contudo, como foi observado duas regiões na análise de PCA, a alteração se deve a PC1 no final da DM, o *#Cluster 3* pode prover estrutura diferente da média observada na DM.

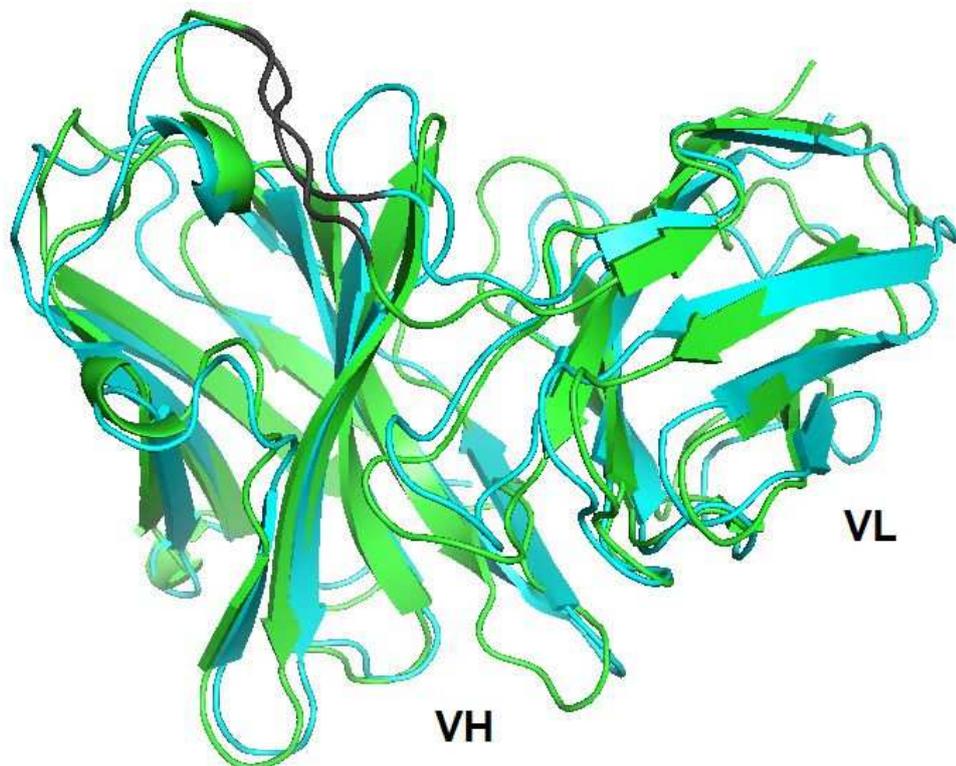
Figura 26 - Resultado análise de cluster para a estrutura VH+VL do scFv com linker curto.



Fonte: elaborada pela autora (2024). *Clusters* de configurações de VH+VL do scFv curto no decorrer de 600 ns simulação, utilizando um raio de corte de 0.18 nm.

Com base na análise de *cluster*, a estrutura do scFv escolhida para a montagem do sistema scFv/CD22 foi a configuração final da simulação, uma vez que essa configuração está presente no *#Cluster1*, o *cluster* mais representativo. Foi colhida uma estrutura no final da DM do *#Cluster1* e do *#Cluster3* para efeito de comparação, em função dos dados de PCA e RMSD. O RMSD após a sobreposição do domínio VH+VL do scFv curto do *#Cluster1* e *#Cluster3* foi de 0,20 nm (Figura 27), mostrando que as estruturas estão muito próximas, dentro da variação de RMSD de VH+VL, como visto na Figura 22.

Figura 27 - Representação estrutural das configurações do scFv curto após a sobreposição do domínio VH+VL do scFv de linker curto.

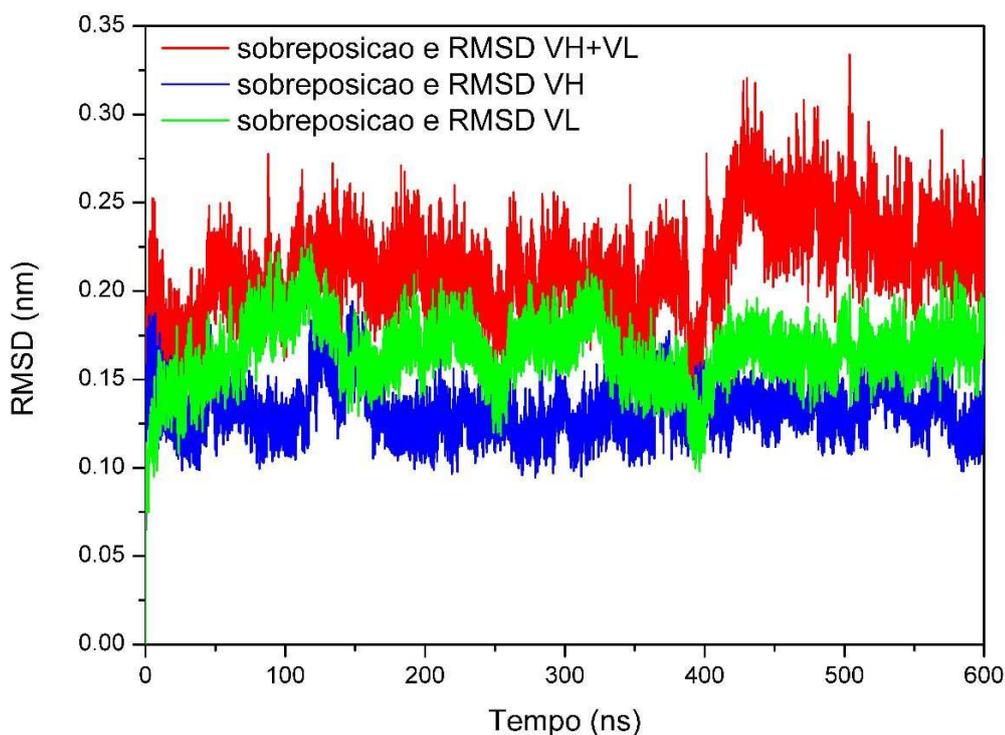


Fonte: elaborada pela autora (2024). #Cluster1 (em verde); #Cluster3 (em ciano). O linker está destacado em preto.

7.2.2 Dinâmica Molecular do scFv longo

Os perfis de RMSD para o sistema scFv de *linker* longo, em solução aquosa, entre os domínios VH+VL, VL e VH são mostrados na Figura 28. As curvas apresentam oscilações para todos os domínios, com RMSD máximo para VH+VL, $\sim 0,30$ nm. As oscilações dificultam em escolher um momento de alcance de equilíbrio estrutural, contudo a curva de RMSD de VL apresenta um crescimento de alcançando máximo em 110 ns ($\sim 0,22$ nm) e esse tempo foi considerado como $t_{eq}=110$ ns para cálculo das médias. Os RMSDs médios estão descritos na Tabela 10, com $0,22 \pm 0,02$ nm para VH+VL, $0,13 \pm 0,01$ nm para VH e $0,17 \pm 0,02$ nm para VL. Os perfis de RMSDs são distintos entre si, assim como as médias. Nota-se no perfil de VH+VL um acréscimo abrupto de RMSD em ~ 400 ns, indicando mudança na conformação do VH+VL. Uma vez que os RMSD de VH e VL estão abaixo VH+VL e não foi observado acréscimo no RMSD em 400 ns para VH e VL, sugere-se que há um reordenamento na estrutura entre VH e VL.

Figura 28 - Perfil de RMSD dos C α dos componentes do scFv longo em solução aquosa, durante 600 ns de simulação.



Fonte: elaborada pela autora (2024). Curva de RMSD entre as configurações estruturais do VH+VL (sobreposição e cálculo de RMSD de VH+VL), VH (sobreposição e cálculo de RMSD de VH) e VL (sobreposição e cálculo de RMSD de VL), comparadas com a estrutura de referência inicial.

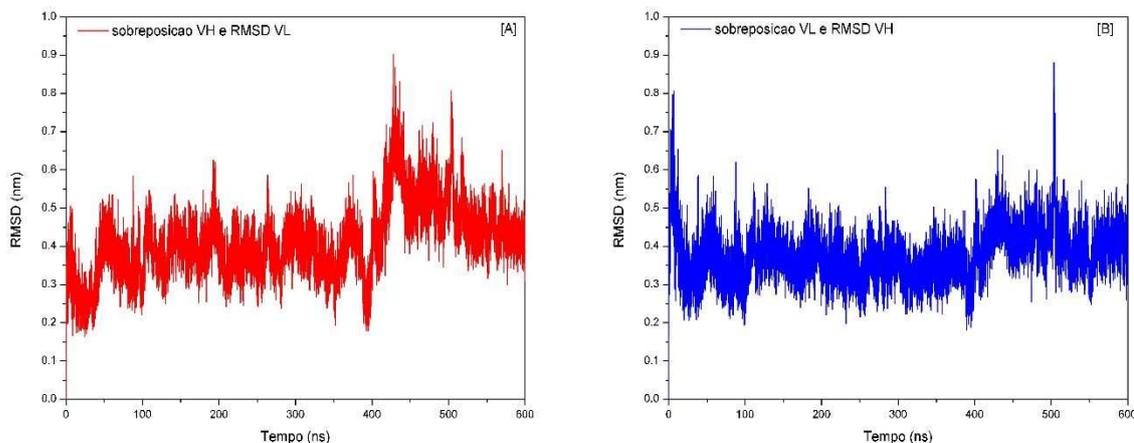
Tabela 10 - RMSDs médios calculados para VH+VL, VH e VL a partir do teq = 110 ns.

Estrutura	RMSD médio (nm)
VH+VL	0,22 ± 0,02
VH	0,13 ± 0,01
VL	0,17 ± 0,02

Fonte: elaborada pela autora (2024).

Realizou-se análise dos perfis de RMSD de VH tendo como referência VL e vice-versa, com objetivo de investigar se existe movimentação relativa de um domínio em relação ao outro (Figura 29). Os perfis são semelhantes entre si, o que demonstra que os movimentos são associados, quase que exclusivamente a movimentação entre VH e VL. Os RMSDs médios do VH em relação ao VL, e do VL em relação ao VH, foram $0,38 \pm 0,06$ e $0,43 \pm 0,09$ nm (Tabela 11). Em 400 ns foi observado um acréscimo nos perfis do RMSDs, assim como foi observado na figura 28, ratificando a movimentação relativa entre os domínios VH e VL.

Figura 29 - Curvas de RMSD relativo entre as estruturas tridimensionais dos domínios VH e VL do scFv longo submetido à 600 ns de simulação em solução aquosa.



Fonte: elaborada pela autora (2024). (A) sobreposição VH e RMSD VL. (B) sobreposição VL e RMSD VH.

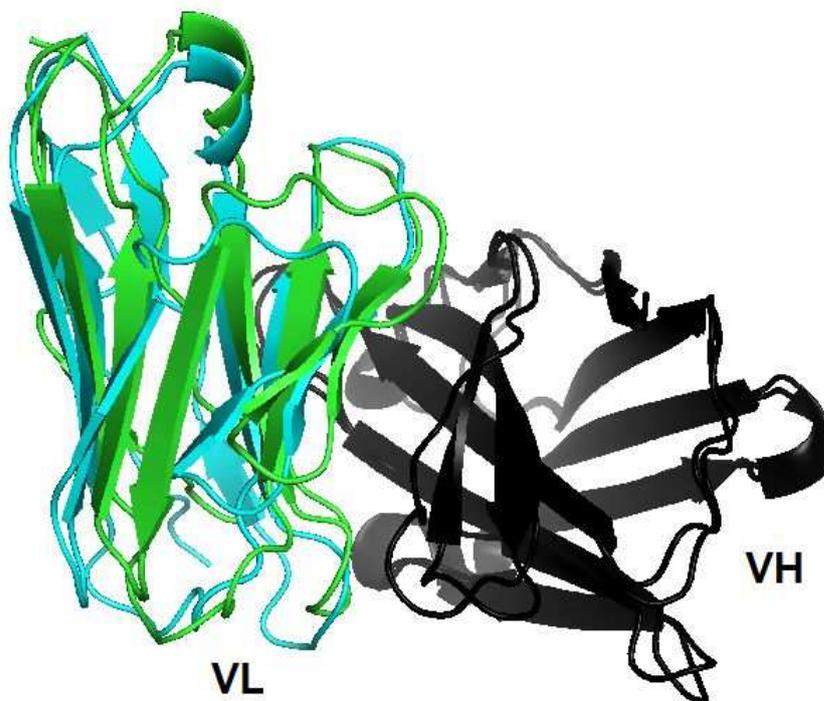
Tabela 11 - RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados do $t_{eq} = 110$ ns.

Estrutura	RMSD médio (nm)
VH Ref VL	$0,38 \pm 0,06$
VL Ref VH	$0,43 \pm 0,09$

Fonte: elaborada pela autora (2024). VH Ref VL = sobreposição de VL e RMSD de VH e VL Ref VH = sobreposição de VH e RMSD de VL.

Foram coletadas estruturas do VH+VL em 0 ns e 500 ns de simulação, a fim de visualizar a alteração da posição do domínio VL em relação a VH. A Figura 30 apresenta as duas estruturas tridimensionais do scFv, sendo que os domínios VH das estruturas foram sobrepostas.

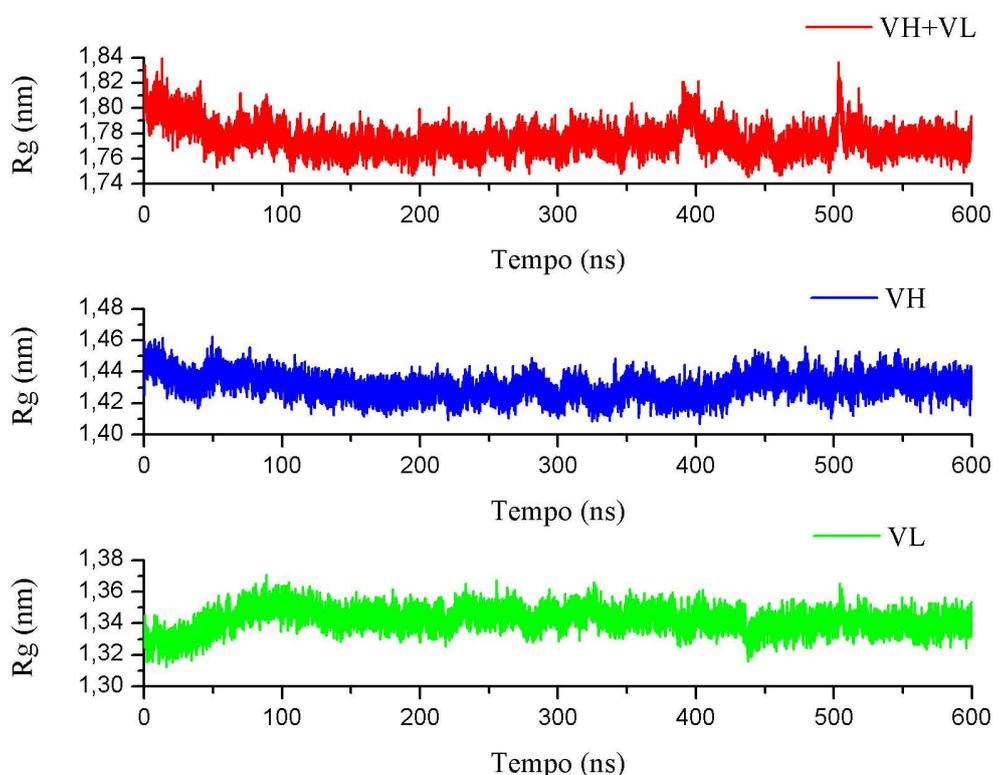
Figura 30 - Conformações do VH+VL coletadas em 0 e 500 ns de simulação evidenciando as diferenças estruturais no domínio VL que foram observadas no gráfico do perfil de RMSD relativo (Figura 29).



Fonte: elaborada pela autora (2024). Preto: domínio VH das duas estruturas sobrepostas. Verde: domínio VL em 0 ns. Azul: domínio VL em 500 ns.

A Figura 31 representa os resultados obtidos por meio da análise de Raio de giro (R_g) dos domínios VH+VL, VH e VL. Os resultados apresentados demonstram que o enovelamento proteico se mantém estável a partir de 110 ns de simulação, revelando alto grau de compacidade proteica (Tabela 12).

Figura 31 - Curvas de Rg dos domínios VH+VL, VH e VL do scFv de linker longo em solução aquosa ao longo de 600 ns de simulação.



Fonte: elaborada pela autora (2024).

Tabela 12 - Raio de Giro médio calculados para a estrutura VH+VL, VH e VL do scFv de linker longo a partir de 110 ns de simulação.

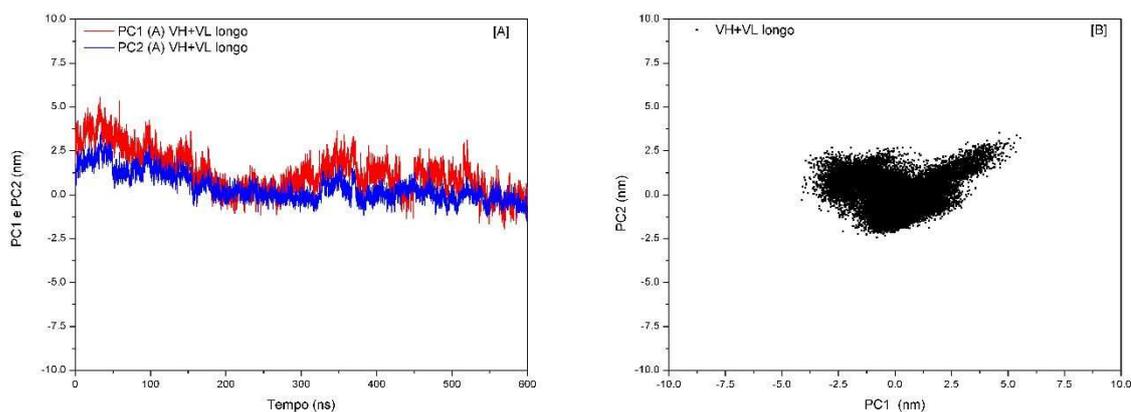
Estrutura	Raio de Giro médio (nm)
VH+VL	$1,77 \pm 0,01$
VH	$1,43 \pm 0,01$
VL	$1,34 \pm 0,01$

Fonte: elaborada pela autora (2024).

A Figura 32 representa os resultados obtidos por meio da análise de PCA do VH+VL de *linker* longo. Conforme a Figura 32A observa-se que as curvas de PC1 e PC2 variam até ~200 ns de simulação, com oscilação de ~- 0 a ~ 5 nm. Após esse período, as curvas tendem a estabilizarem próximo de 2.5 nm. Pode-se observar ainda que na Figura 32B que o VH+VL de *linker* longo apresenta pouca flexibilidade estrutural durante a simulação, uma vez que suas projeções nos subespaços comuns definidos por PC1/PC2 possuem duas regiões, muitas próximas, ambas de mesma intensidade. Esse resultado, em conjunto com o observado nos perfis de RMSD,

ratificam a mudança estrutural conformacional entre os domínios VH e VL, considerando a configuração inicial.

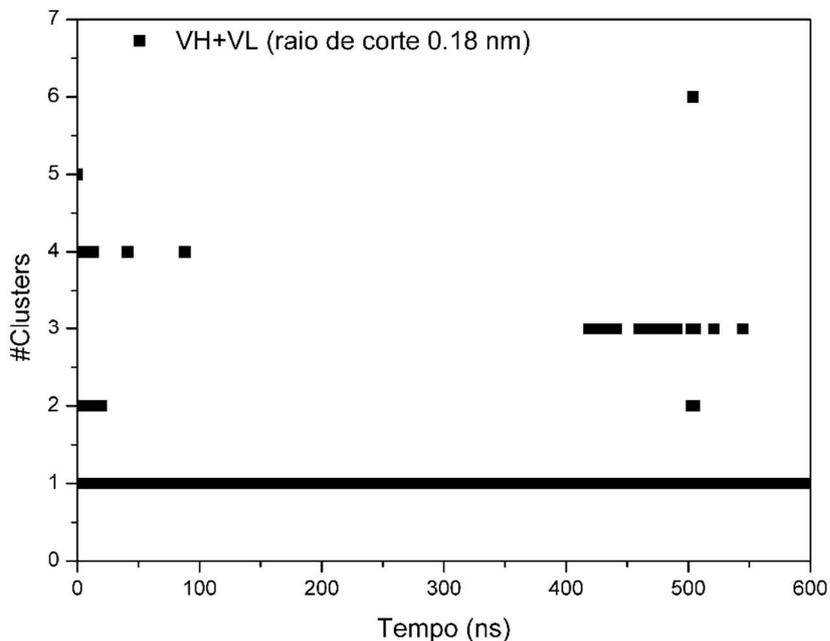
Figura 32 - Curvas de Análise de Componentes Principais. (A) PC1 e PC2 em função do tempo durante 600 ns de simulação. (B) Projeções da simulação do scFv de linker longo nos subespaços comuns definidos por PC1/PC2.



Fonte: elaborada pela autora (2024).

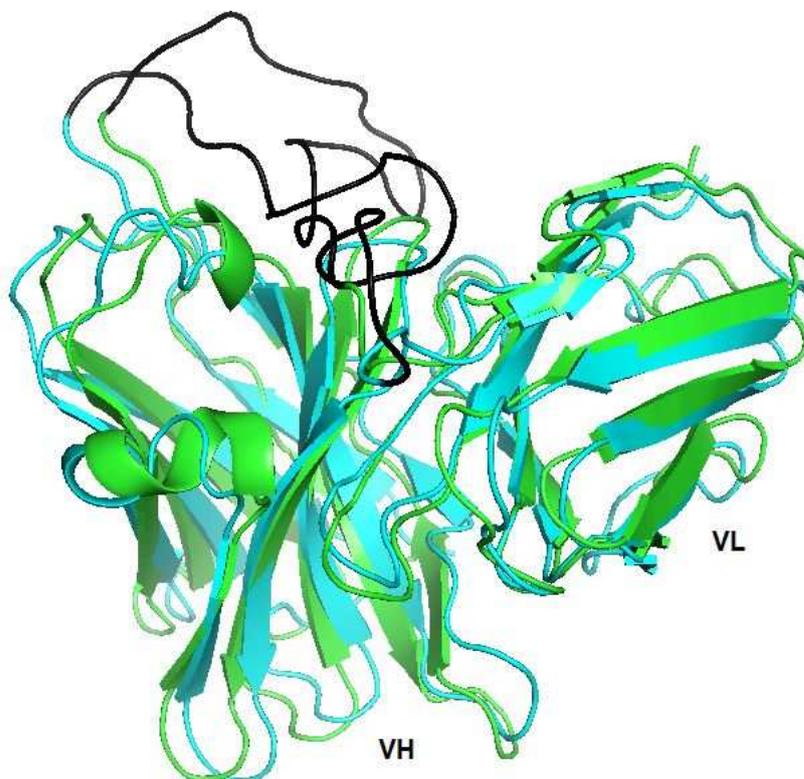
Para a análise *cluster* foi utilizado um raio de corte de 0.18 nm, como mostrado na Figura 33. Os dados da clusterização revelam seis grupos de estruturas, sendo o *#Cluster 1* o mais representativo, preponderante desde o início da simulação. Foi colhida uma estrutura no final da DM do *#Cluster1* e do *#Cluster3* para efeito de comparação, em função dos dados de PCA e RMSD. O RMSD após a sobreposição do domínio VH+VL do scFv curto do *#Cluster1* e *#Cluster3* foi de 0,20 nm (Figura 34), mostrando que as estruturas estão muito próximas, dentro da variação de RMSD de VH+VL, como visto na Figura 29.

Figura 33 - Resultado análise de cluster para a estrutura VH+VL do scFv com linker longo.



Fonte: elaborada pela autora (2024). *Clusters* de configurações de VH+VL do scFv longo ao longo de 600 ns simulação, utilizando um raio de corte de 0.18 nm.

Figura 34 - Representação estrutural das configurações do scFv longo, após a sobreposição do domínio VH+VL, do scFv de linker longo.

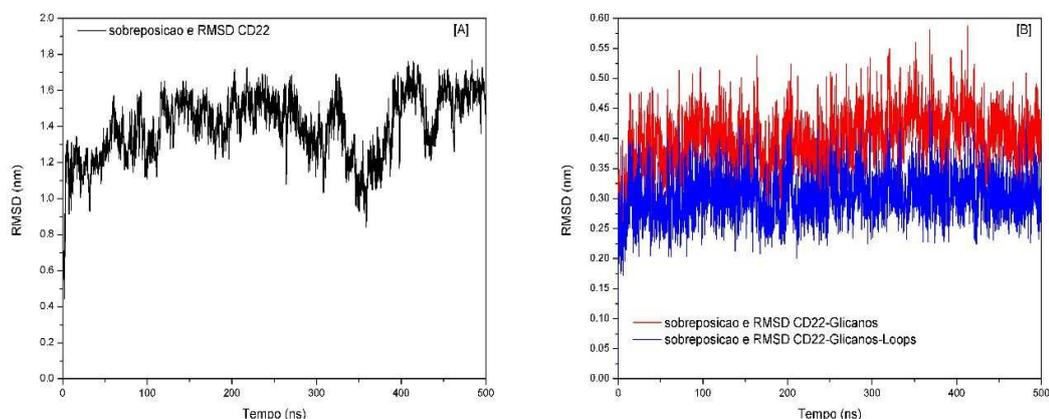


Fonte: elaborada pela autora (2024). #Cluster1 (em verde); #Cluster3 (em ciano). O linker está destacado em preto.

7.2.3 Análise estrutural do CD22 em solução aquosa

Os perfis de RMSDs gerados no estudo da estrutura do CD22, durante 500 ns de DM, é mostrado nas Figuras 35A-B e Tabela 13. Observa-se no perfil de RMSD do CD22 um aumento rápido de RMSD nos 2 ns iniciais ($\sim 1,2$ nm) e um acréscimo suave na curva até 100 ns, onde se estabeleceu $t_{eq} = 100$ ns. O RMSD médio do CD22 foi de $1.43 \pm 0,16$ nm enquanto o do CD22 descontando as árvores de glicosilação (CD22-glicanos) foi de $0,40 \pm 0,04$ nm. Portanto, a diferença de RMSD infere a grande movimentação das árvores de glicosilação. Além das árvores de glicosilação o CD22 possui *loops*, já conhecidos pela grande mobilidade, que tiveram que ser modelados por não possibilitarem a resolução através de cristalografia. Nesse sentido, o perfil de RMSD do CD22 sem os *loops* e sem as árvores de glicosilação (CD22-Loops-Glicanos) foi gerado, como visto na Figura 35B. Nota-se uma queda no perfil, em relação ao CD22-glicanos, demonstrando que os *loops* desempenham movimentação na estrutura do CD22 e que o restante da estrutura possui RMSD médio de $0,30 \pm 0,03$ nm. O perfil de RMSD para CD22 apresentou uma depressão ($\sim 0,8$ nm) em ~ 310 ns. Foram coletadas as configurações iniciais e no tempo 310 ns para confrontar as estruturas, como mostrado na Figura 36. Nota-se que as árvores de glicosilação estão bem distintas da posição inicial, além disso estão numa conformação distendida em relação aos domínios D6 e D7 do CD22. No entanto fica claro que a estruturas de D6-D7 é mantida próxima a de referência, 0 ns, com visto no RMSD em Figura 35.

Figura 35 - Perfis de RMSD dos C α das configurações estruturais do CD22, ao longo de 500 ns de simulação.



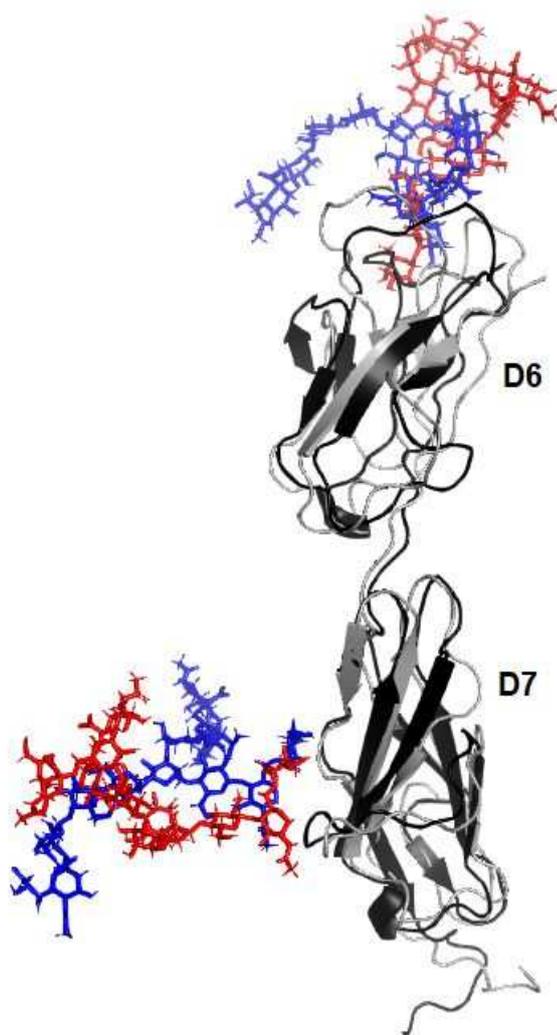
Fonte: elaborada pela autora (2024). (A) Perfil de RMSD dos C α das configurações estruturais do CD22, ao longo do tempo de simulação, comparadas a estrutura inicial como referência. (B) RMSDs médios do CD22 com a estrutura desconsiderando os glicanos (CD22-Glicanos) e desconsiderando os glicanos e os *Loops* modelados (CD22-Glicanos-Loops), ao longo do tempo de simulação, comparadas a estrutura inicial, como referência.

Tabela 13 - RMSDs médio calculados para a estrutura completa do CD22, com a estrutura desconsiderando os glicanos (CD22-Glicanos) e desconsiderando os glicanos e os Loops modelados (CD22-Loops-Glicanos), teq = 100 ns de simulação.

Estrutura	RMSDs médio (nm)
CD22	1,43 ± 0,16
CD22-Glicanos	0,40 ± 0,04
CD22-Glicanos-Loops	0,30 ± 0,03

Fonte: elaborada pela autora (2024).

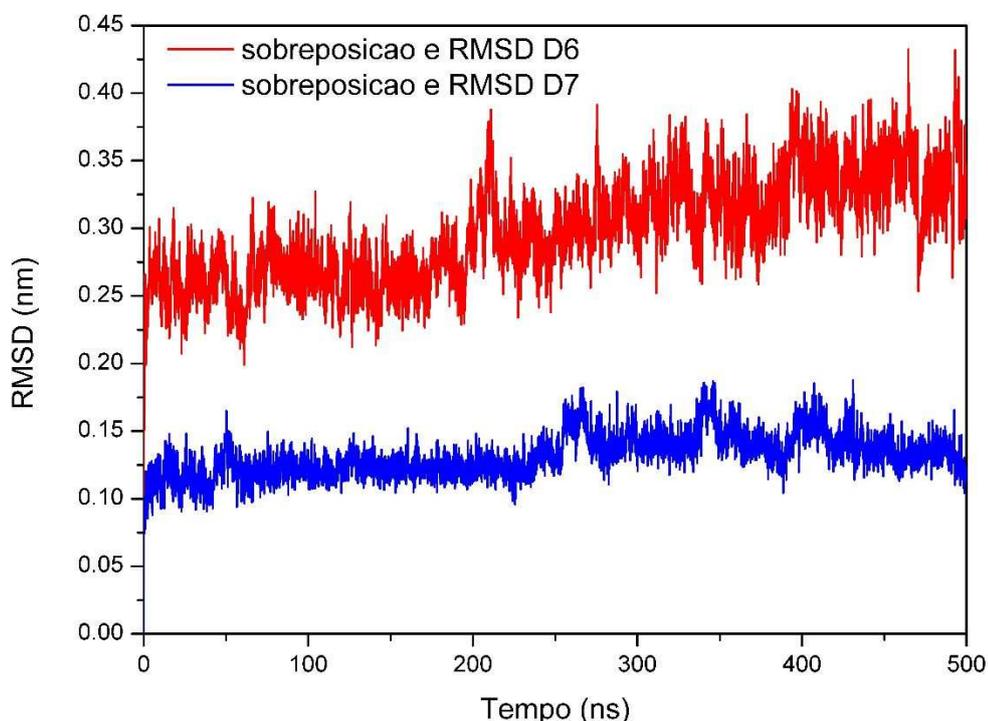
Figura 36 - Conformações de CD22 coletadas em 0 e 310 ns de simulação evidenciando que as árvores de glicosilação estão bem distintas da posição inicial.



Fonte: elaborada pela autora (2024). Estruturas representadas em *cartoon* e glicosilações em *sticks* sendo na cor azul para estrutura em 0 ns e na cor vermelha para a estrutura em 310 ns. O fundo da figura é branco. Configurações de CD22 em 0 (cinza) e 310 ns (preto) de simulação, evidenciando os glicanos em uma conformação mais estendida em comparação com a estrutura inicial de simulação.

Buscando ratificar a estabilidade das estruturas dos domínios D6 e D7, no CD22, foi avaliada através de RMSD de cada um deles, sem as glicosilações, como mostrado na Figura 37. Nota-se que as curvas possuem comportamentos distintos, a de D7 o RMSD é mais baixo e oscila com $0,14 \pm 0,02$ nm (Tabela 14) indicando maior estabilidade estrutural. Enquanto para D6 a curva de RMSD apresenta uma elevação a partir de ~ 200 ns e a flutuação é maior que em D7, como observado pelo desvio ($0,31 \pm 0,04$ nm). Uma vez que o RMSD do CD22-Glicanos é $0,40 \pm 0,01$ nm, o RMSD está bem maior que o de D7 ($\sim 0,14$ nm) e próximo a D6 ($\sim 0,31$ nm), é indicativo que há variações na estrutura do CD22, desconsiderando a glicosilação, decorrente da alteração estrutural em D6. Ressalta-se que o baixo RMSD em D7 ($0,14 \pm 0,02$ nm) é importante para constitui uma região de epítipo. A estrutura cristalográfica mostra que o contato entre CD22 e o Fab ocorrem sobremaneira no D7 (EREÑO-ORBEA *et al.*, 2021).

Figura 37 - Perfil de RMSD para os C α do CD22-Glicanos dos domínios D6 e D7, comparadas com a estrutura de referência inicial.



Fonte: elaborada pela autora (2024).

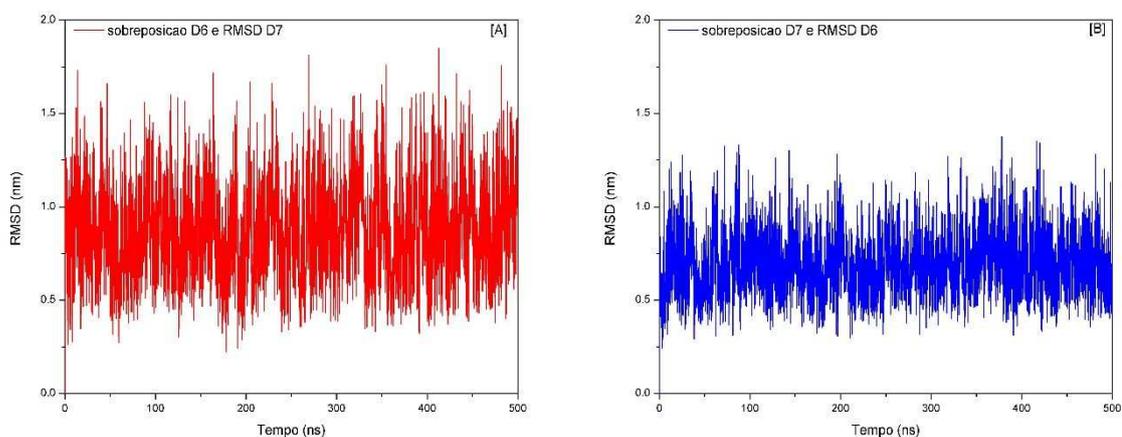
Tabela 14 - RMSDs médios calculados para domínios D6 e D7 de CD22 a partir do $t_{eq} = 100$ ns.

Estrutura	RMSD médio (nm)
D6	$0,31 \pm 0,04$
D7	$0,14 \pm 0,02$

Fonte: elaborada pela autora (2024).

A Figura 38 mostra os perfis de RMSD de D6 tendo feito a sobreposição de D7 e vice-versa com objetivo de investigar se há movimentação relativa entre domínios D6 e D7. Nota-se RMSDs altos e alta frequência nas oscilações, indicando que há movimentação relativa e que as mesmas ocorrem rapidamente (Tabela 15). As oscilações na curva de RMSD para D7 são maiores pois no cálculo do RMSD inclui a posição relativa a D6, que possui maior movimentação intrínseca em sua estrutura ($0,31 \pm 0,04$ nm) que em D7 ($0,14 \pm 0,02$ nm) (Figura 37) e, nesse caso a sobreposição varia entre as configurações temporais de D6 e de referência, além de ocorrer a movimentação relativa entre os domínios.

Figura 38 - Perfil de RMSD para os C α do CD22-Loops-Glicanos, considerando RMSD de D6 e sobreposição de D7 e vice-versa.



Fonte: elaborada pela autora (2024). (A) sobreposição D6 e RMSD D7. (B) sobreposição D7 e RMSD D6.

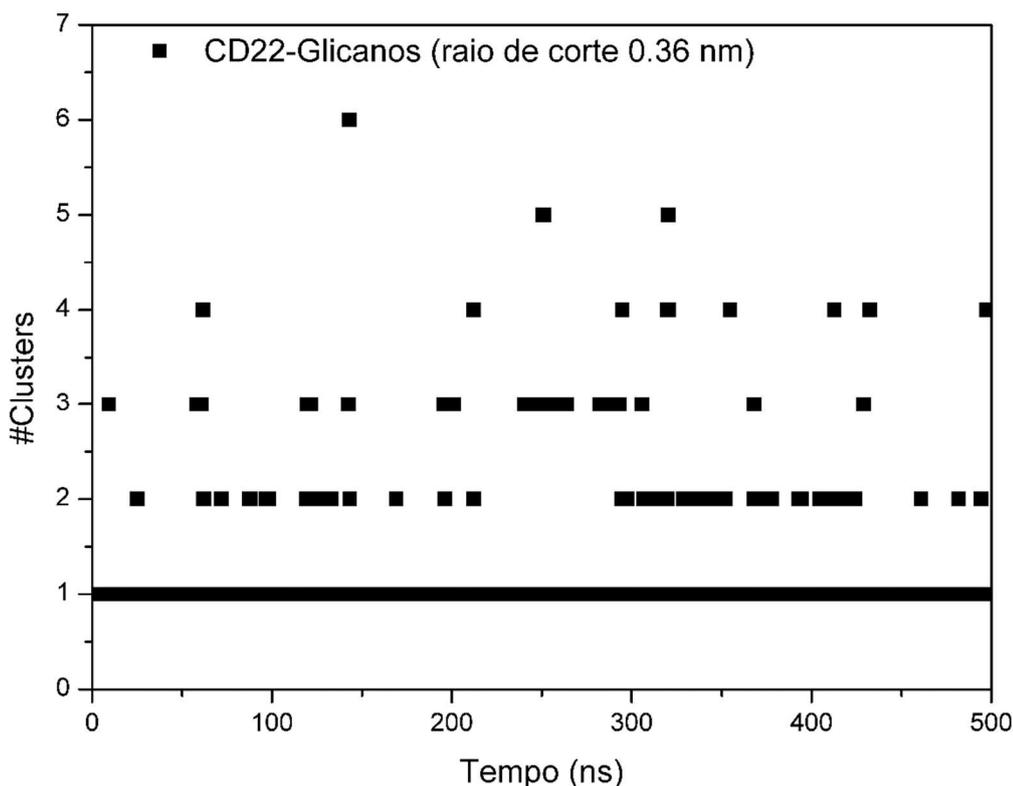
Tabela 15 - RMSDs médios dos domínios D6 tendo como referência para sobreposição o D7, e vice-versa, calculados a partir do $t_{eq}=100$ ns.

Estrutura	RMSD médio (nm)
Sobreposição de D6 RMSD de D7	$0,87 \pm 0,26$
Sobreposição de D7 RMSD de D6	$0,69 \pm 0,17$

Fonte: elaborada pela autora (2024).

A Figura 39 mostra a análise de *cluster* com raio de corte de 0,36 nm para o CD22-Glicanos. Foram identificados 6 *clusters*, o #*Cluster 1* é o mais representativo, apresentando maior número de estruturas. A última configuração do #*Cluster 1* foi escolhida para dar continuidade aos estudos.

Figura 39 - Clusters de configurações do CD22 ao longo de 500 ns simulação, utilizando raio de corte de 0,36 nm.

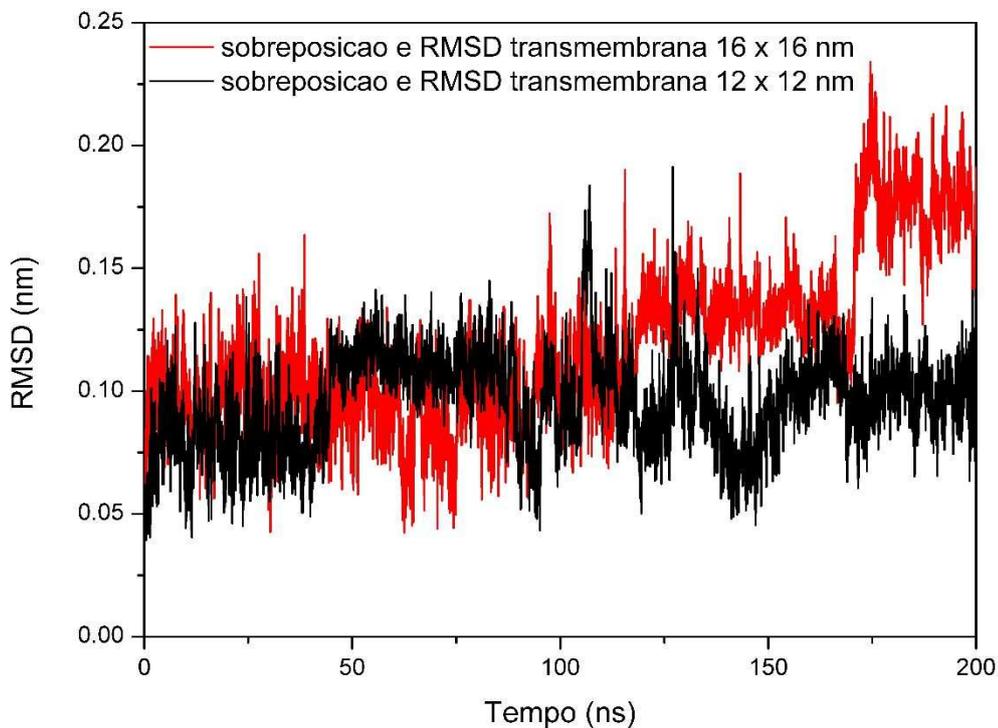


Fonte: elaborada pela autora (2024).

7.2.4 DM do sistema domínio transmembrana do CD22 em membrana de 12 x 12 nm e 16 x 16 nm

O RMSD do domínio transmembrana foi monitorado na simulação dos dois sistemas como mostrado na Figura 40. Os perfis de RMSD mostram pouca alteração na estrutura do domínio transmembrana em ambos os sistemas. Os RMSDs médios na membrana 12 x 12 nm foram de $0,10 \pm 0,02$ nm e na de 16 x 16 nm foi de $0,15 \pm 0,02$ nm. O perfil de RMSD do domínio transmembrana em membrana de 16 x 16 nm apresentou um acréscimo no final do perfil

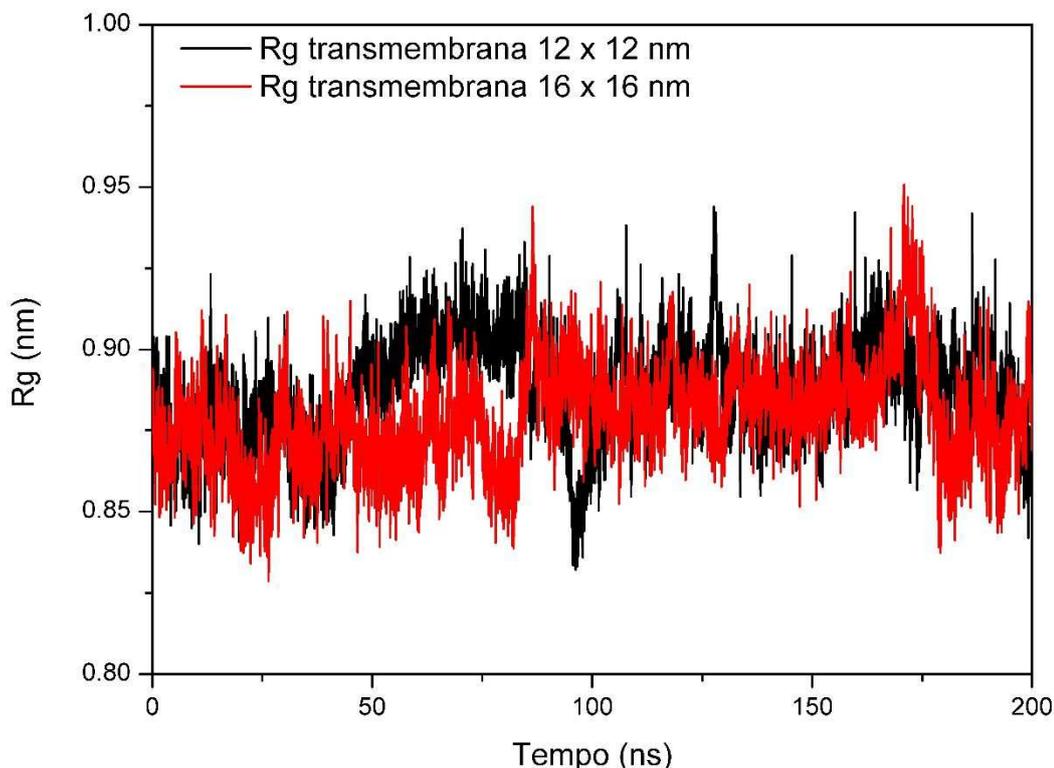
Figura 40 - Perfil de RMSD dos C α das configurações estruturais do domínio transmembrana de CD22, ao longo de 200 ns de simulação, comparadas a estrutura inicial como referência.



Fonte: elaborada pela autora (2024).

A Figura 41 representa o resultado obtido por meio da análise de Raio de giro das estruturas dos domínios transmembranas, visando identificar algum estiramento na estrutura que explica o acréscimo de RMSD quando em membrana de 16 x 16 nm. Os perfis não mostram alterações relevantes e os Rg estão próximos entre as simulações.

Figura 41 - Raio de Giro do domínio transmembrana de CD22, ao longo do tempo de simulação, comparadas a estrutura inicial como referência, nos sistemas 12 x 12 e 16 x 16 nm.



Fonte: elaborada pela autora (2024).

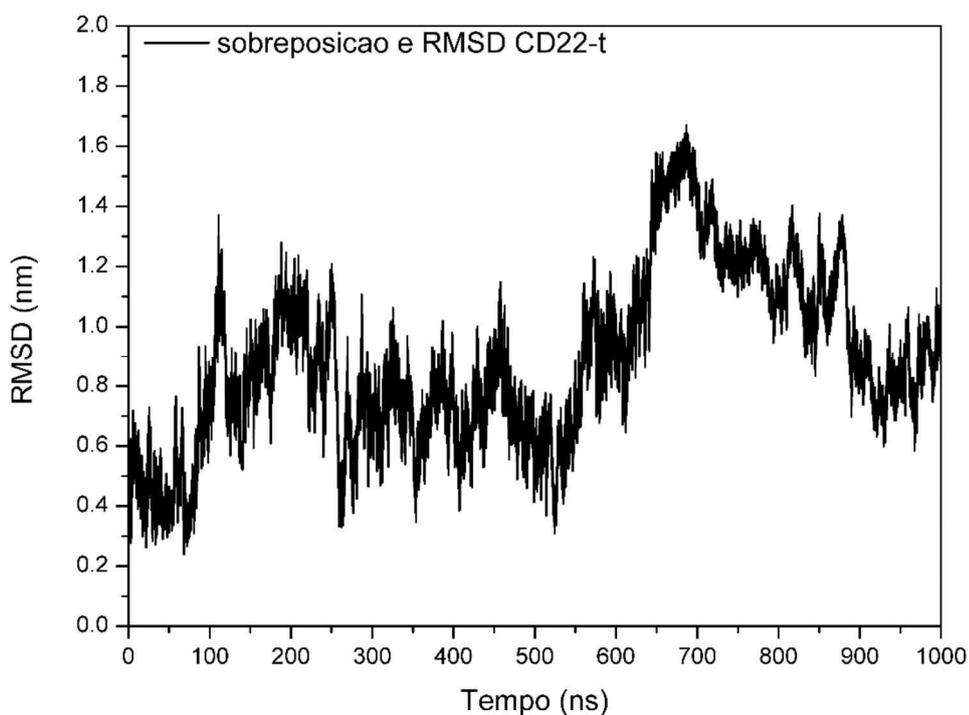
A estrutura final do domínio transmembrana 16 x 16 nm foi utilizada para a montagem dos sistemas scFv/CD22 de *linker* curto e longo em membrana e a estrutura final do domínio transmembrana 12 x 12 nm foi utilizada para a montagem do sistema CD22 em membrana.

7.2.5 Sistema composto por CD22 conectado ao domínio transmembrana, inserido em modelo de membrana celular

Nessa seção, o CD22 incorpora o domínio transmembrana e será denominado CD22-t. O perfil de RMSD do CD22-t mostra mínimos e máximo de ~0,4 a 1,6 nm, ao longo da DM, como visto na Figura 42. Essa amplitude de movimento sugere que o CD22 pode estar se movimentando sob o plano da membrana, enquanto o domínio transmembrana está inserido na bicamada lipídica (Figura 44). Foi analisada o quanto a estrutura do domínio transmembrana é modificada na Figura 43. Da mesma forma, o CD22 possui duas árvores de glicosilação e *loops* modelados que se movimentam

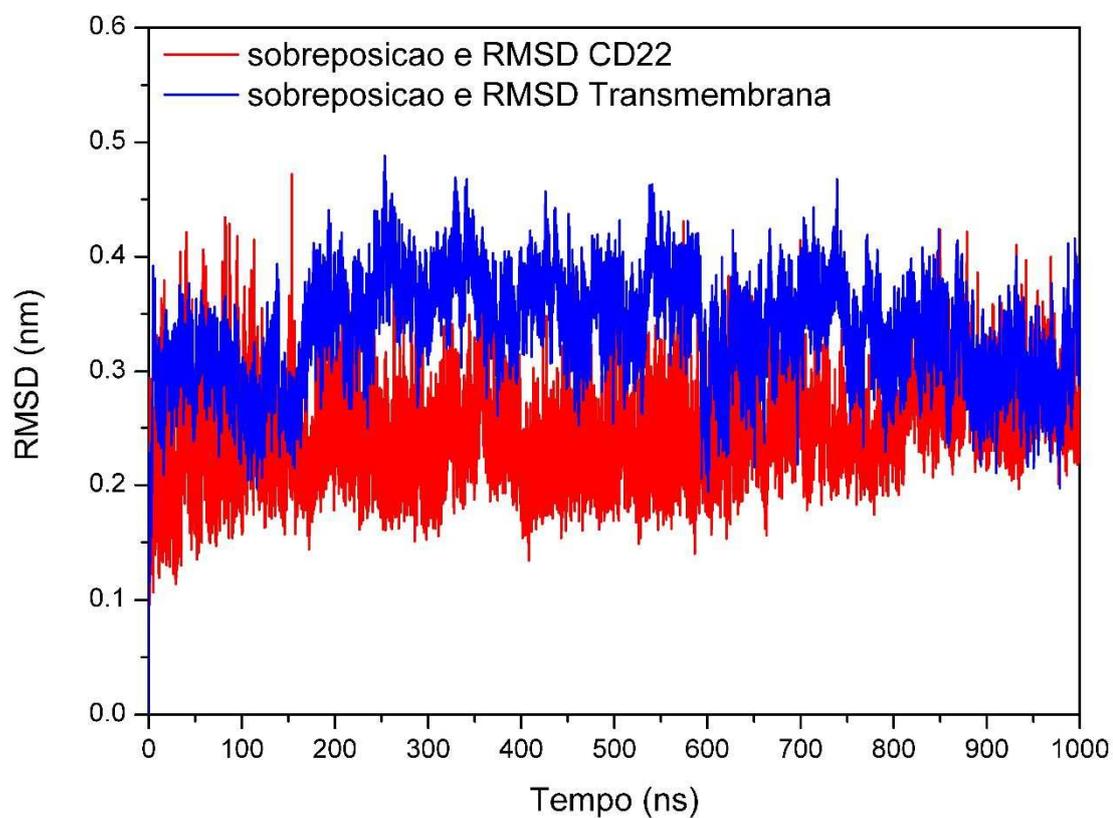
muito, portanto o RMSD também foi gerado descontando as árvores e os *loops* modelados. Os perfis em ambos os casos apresentam RMSD próximos como observado na Figura 43, sendo que para transmembrana o RMSD é $0,34 \pm 0,05$ nm, acima de CD22, $0,25 \pm 0,04$ nm (Tabela 16). Em relação ao domínio transmembrana ele alcança RMSD $\sim 0,3$ nm no final da DM, um patamar não elevado considerando que uma da extremidade está ligada ao CD22 e a outra está livre.

Figura 42 - Perfil de RMSD do CD22-t considerando a sobreposição do dos Ca das configurações em comparação com a estrutura inicial de referência.



Fonte: elaborada pela autora (2024).

Figura 43 - Perfil de RMSD dos C α das configurações estruturais do CD22 e Transmembrana, ao longo de 1000 ns de simulação, comparadas a estrutura inicial como referência.



Fonte: elaborada pela autora (2024).

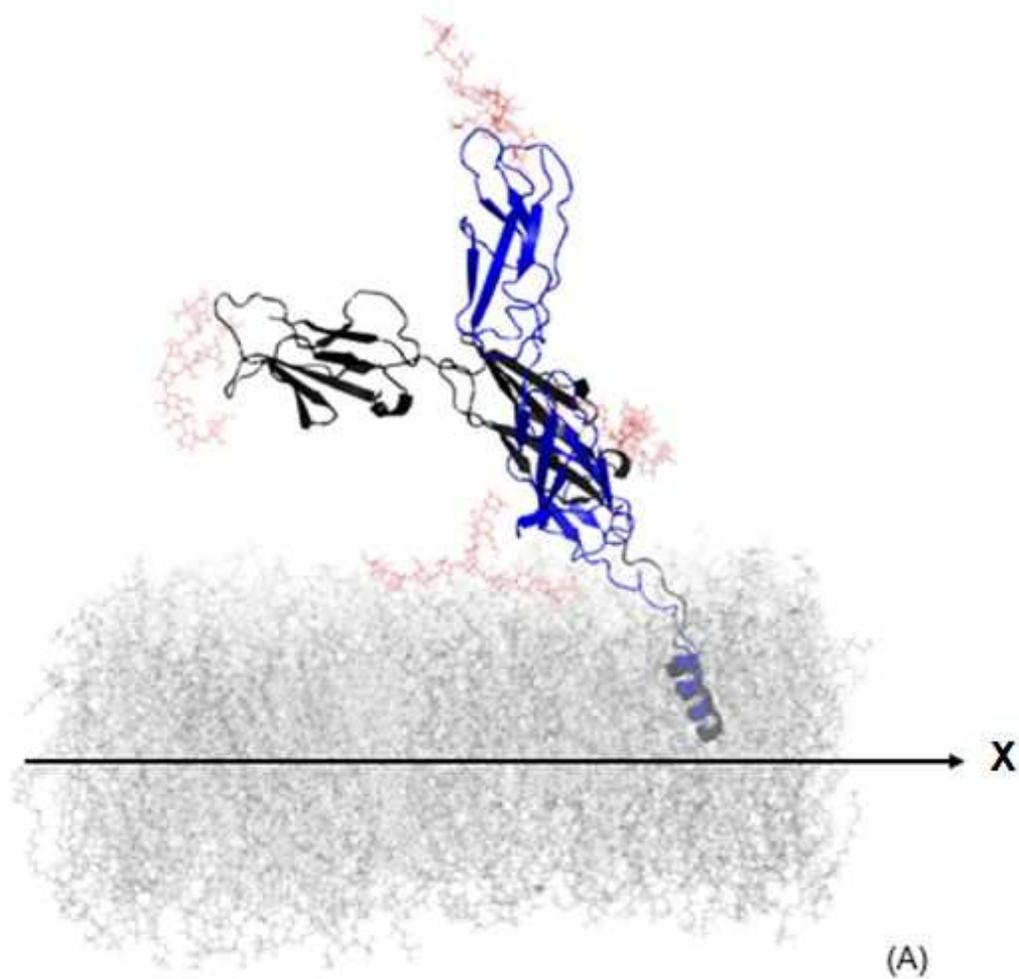
Tabela 16 - RMSDs médios calculados para CD22 e Transmembrana a partir do teq = 100 ns.

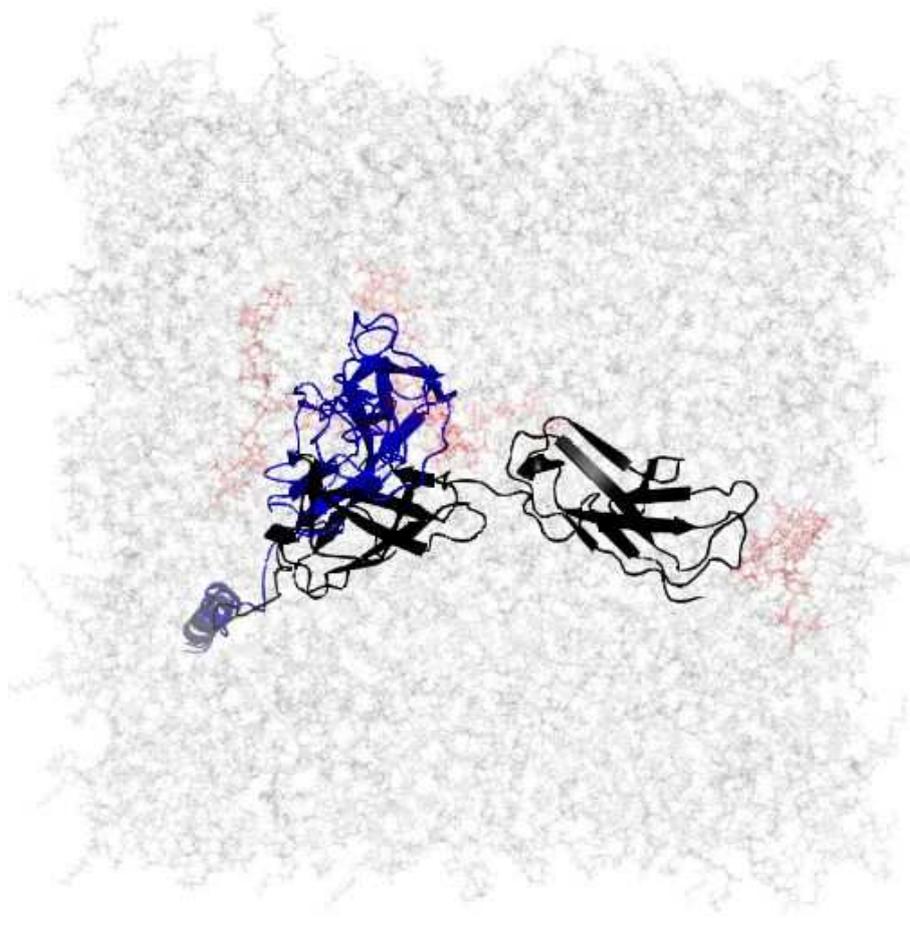
Estrutura	RMSD médio (nm)
CD22	0,25 ± 0,04
Transmembrana	0,34 ± 0,05

Fonte: elaborada pela autora (2024).

A Figura 44 corrobora com a Figura 42 à medida que são apresentadas duas diferentes configurações de CD22-t 100 e 900 ns evidenciando as conformações que podem estar orientadas ao plano da membrana e perpendicularmente.

Figura 44 - Representação estrutural do CD22-t após a sobreposição do domínio transmembrana.





Fonte: elaborada pela autora (2024). A) Representação do sistema tendo a bicamada lipídica na horizontal e transmembrana atravessando a bicamada, na superfície o CD22. B) Representação vista de cima, com rotação de 90° no eixo x, que está sob o plano da bicamada, na horizontal. Em 100 ns (azul); em 900 ns (preto) de simulação.

7.3 Estudos dos complexos scFv/CD22 em solução aquosa

7.3.1 Estudo da interface scFv/CD22 proveniente de docagem molecular do scFv ao CD22

Após a simulação de DM dos dois modelos de scFv-água; e, do modelo de CD22 em água, foram coletadas as estruturas de 100, 390, 500 e 600 ns de simulação de cada sistema scFv-água, assim como também foi coletada a configuração de 500 ns do sistema CD22 em água. Em seguida, as configurações de scFv de *linker* curto e longo e CD22 coletadas foram submetidas a *docking* molecular no *Haddock Docking*. Para todos os *dockings* realizados foram considerados os estados das protonação das histidinas na posição 16, 31 e 129 protonadas no nitrogênio epsilon (HSE); e 122

e 136 protonadas no nitrogênio epsilon e delta (HSP) na molécula de CD22. Com relação aos resíduos ativos no CD22, foram utilizados os *hotspots* de acordo com o trabalho realizado por Ereño-orbea e colaboradores (2021). Já com relação aos resíduos ativos nos scFvs de *linker* curto e longo, foram selecionadas as CDRs das variáveis VH e VL (Quadro 2). Para a realização do *docking* molecular definiu-se a estrutura do CD22 para permanecer rígida. Dando continuidade, foram realizados um total de 8 *dockings*, sendo 4 para o scFv de *linker* curto e 4 para o scFv de *linker* longo, sendo selecionados os melhores *clusters* gerados pelos *dockings* a partir da avaliação de um conjunto de parâmetros como: média ponderada dos termos eletrostáticos, dessolvatação e energia de *van der Waals*, juntamente com a contribuição energética das restrições utilizadas para impulsionar o *docking*. Por fim, foi avaliado, através da ferramenta *gmx energy*, o PII, dentro de um raio de corte de 1,3 nm, entre os scFvs e o CD22, dos complexos selecionados pelos diferentes *dockings*, conforme exposto pela Tabela 17. Em seguida, foram selecionados os complexos que apresentaram PII mais negativos para realização da DM.

Tabela 17 - Potencial de Interação Intermolecular entre VH+VL e CD22 em configurações scFv de linker curto e longo (olhar seções 8.1.1 e 8.1.2) coletadas para a realização dos dockings.

Configurações coletadas nos instantes (ns)	scFv/CD22 curto (kcal/mol)	scFv/CD22 longo (kcal/mol)
100	246,66	-149,62
390	-106,18	-111,54
500	-118,87	-106,75
600	-146,59	428,36

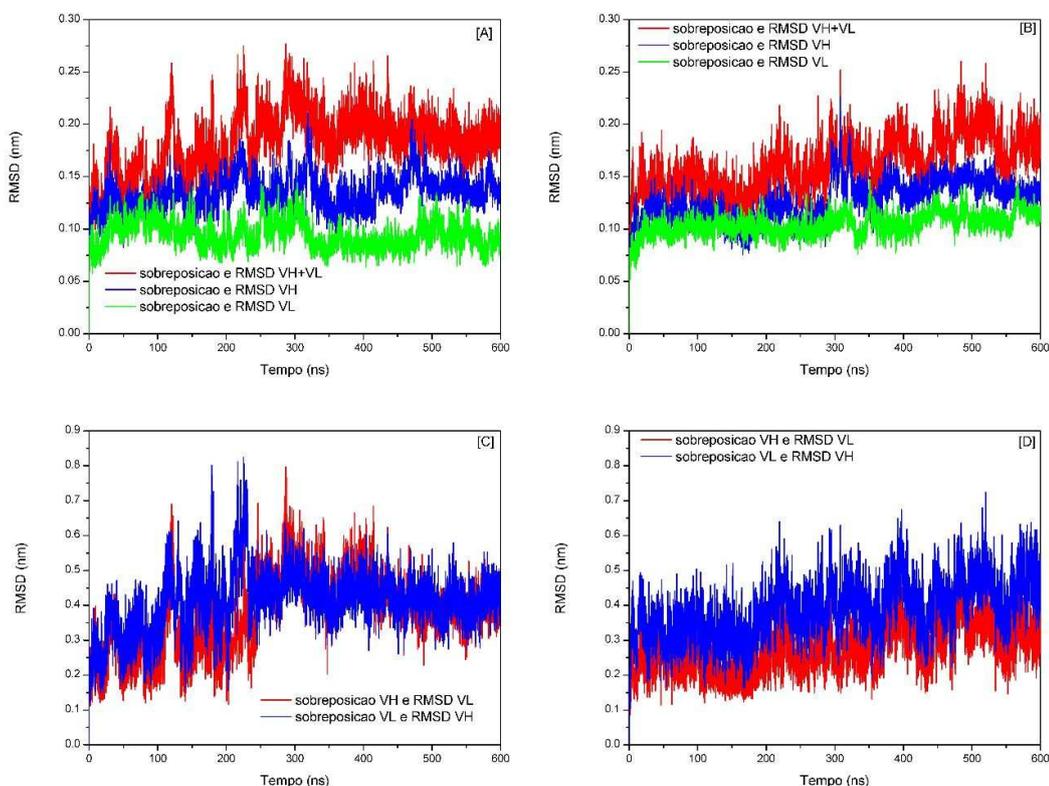
Fonte: elaborada pela autora (2024).

Os perfis de RMSD para o scFv de *linker* curto e longo nos sistemas scFv/CD22 foram gerados e descritos na Figura 45A e 45B. A curva de RMSD de VH+VL está acima das de VH e VL, respectivamente. Além disso a forma das curvas entre VH+VL e VH são similares, indicando que os movimentos atribuídos a VH+VL podem estar associados aos de VH. Nota-se que o RMSD de VH e VL aumenta rapidamente nos primeiros ns de DM e se mantem oscilando ao longo da simulação, sendo que foi

escolhido $\text{teq} = 50$ ns para o cálculo das médias. Os RMSDs médios dos domínios do scFv no sistema scFv/CD22 de *linker* curto e longo estão descritos na Tabela 18. Os desvios da média denotam oscilação estrutural no intervalo temporal analisado.

Visando investigar se o movimento de VH-VL está associado ao VH, foi gerado perfis de RMSD de VH após sobreposição das estruturas de VL e vice-versa como mostrado na Figura 45C e 45D. As curvas de VH e VL mostram RMSD maiores que observado na Figura 45A-B e curvas similares, ratificando que há movimentação relativa entre VH e VL. No entanto, em comparação entre os scFvs, nota-se que no *linker* curto, após ~ 250 ns as curvas estão sobrepostas, demonstrando que a movimentação entre VH e VL está fortemente correlacionada, com menor influência das movimentações intrínsecas desses domínios, enquanto para *linker* longo há uma correlação, mas com uma clara influência de movimentação do VH, com RMSD maior que VL. Os RMSD médios estão relacionados na Tabela 19 e ratificam essas diferenças de RMSD observadas nas curvas.

Figura 45 - Perfis de RMSD para os domínios do scFv nos sistemas scFv/CD22, sendo que o scFv pode ser de linker curto ou longo.



Fonte: elaborada pela autora (2024). A) curto. B) longo. C) curto com sobreposição de VH e RMSD de VL e vice-versa. D) longo com sobreposição de VH e RMSD de VL e vice-versa.

Tabela 18 - Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL em ambos os sistemas scFv/CD22 linkers (curto e longo) a partir do teq = 50 ns.

Sistema	RMSD (nm)		
	VH+VL	VH	VL
curto	0,19 ± 0,03	0.14 ± 0.02	0.09 ± 0.01
longo	0.16 ± 0.03	0.13 ± 0.02	0.11 ± 0.01

Fonte: elaborada pela autora (2024).

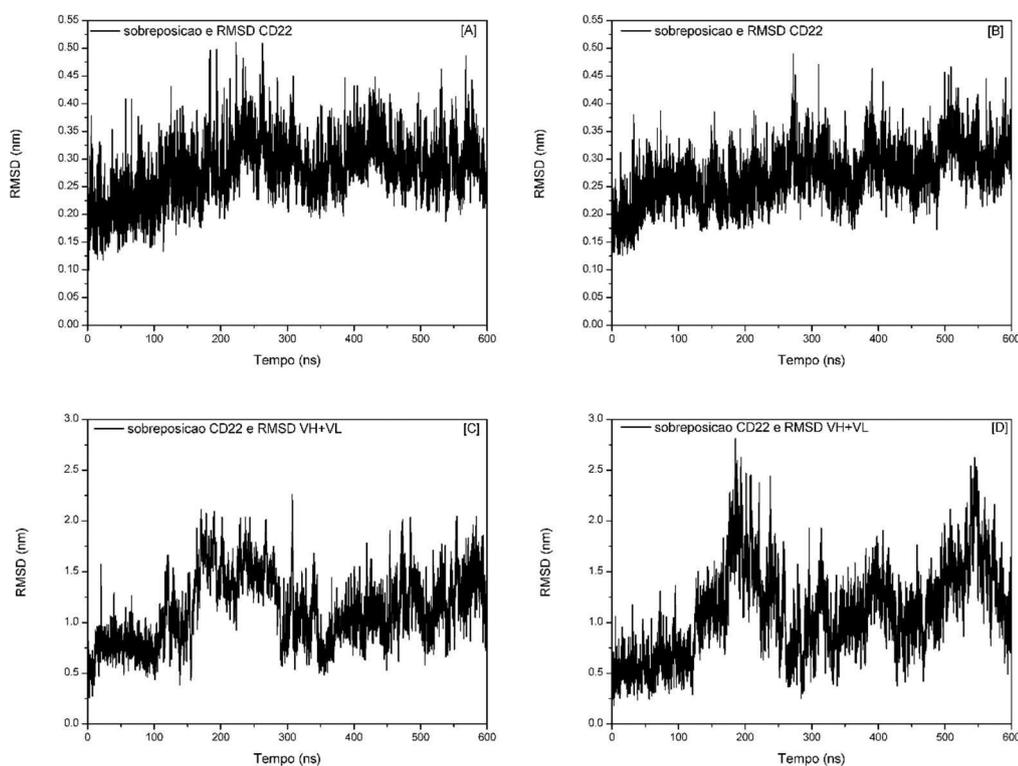
Tabela 19 - RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 50 ns.

Sistema	RMSD (nm)	
	VH Ref VL	VL Ref VH
curto	0,42 ± 0,08	0,39 ± 0,11
longo	0,39 ± 0,08	0,27 ± 0,07

Fonte: elaborada pela autora (2024). VH Ref VL = sobreposição de VL e RMSD de VH e VL Ref VH = sobreposição de VH e RMSD de VL

As Figuras 46A e 46B apresentam os perfis de RMSD do CD22, excluindo os *loops* e glicosilações, no sistema scFv/CD22 de *linker* curto e longo, no decorrer de 600 ns de simulação, respectivamente. O RMSD para o CD22 oscilando longo do tempo acima de 0,30 nm, com variações altas nas amplitudes em poucos ns (entre ~ 0,51 nm e ~ 0,17 nm para os sistemas scFv/CD22 de *linker* curto e longo). As médias de RMSD para o CD22 foram de $0,29 \pm 0,05$ e $0,28 \pm 0,05$ nm para scFv/CD22 de *linker* curto e longo. As Figuras 46C e 46D apresentam os perfis de RMSD do VH+VL em relação a CD22. Os perfis mostram RMSD alto indicando ampla mudança estrutural em ambos os complexos. As amplitudes de RMSD são maiores para o caso de scFv de *linker* longo (Figura 48D) que do *linker* curto (Figura 48C) indicando que há maior variação na posição do scFv de *linker* longo em relação ao CD22. Neste caso, as médias de RMSD foram de $1,16 \pm 0,33$ e $1,15 \pm 0,43$ nm para scFv/CD22 de *linker* curto e longo.

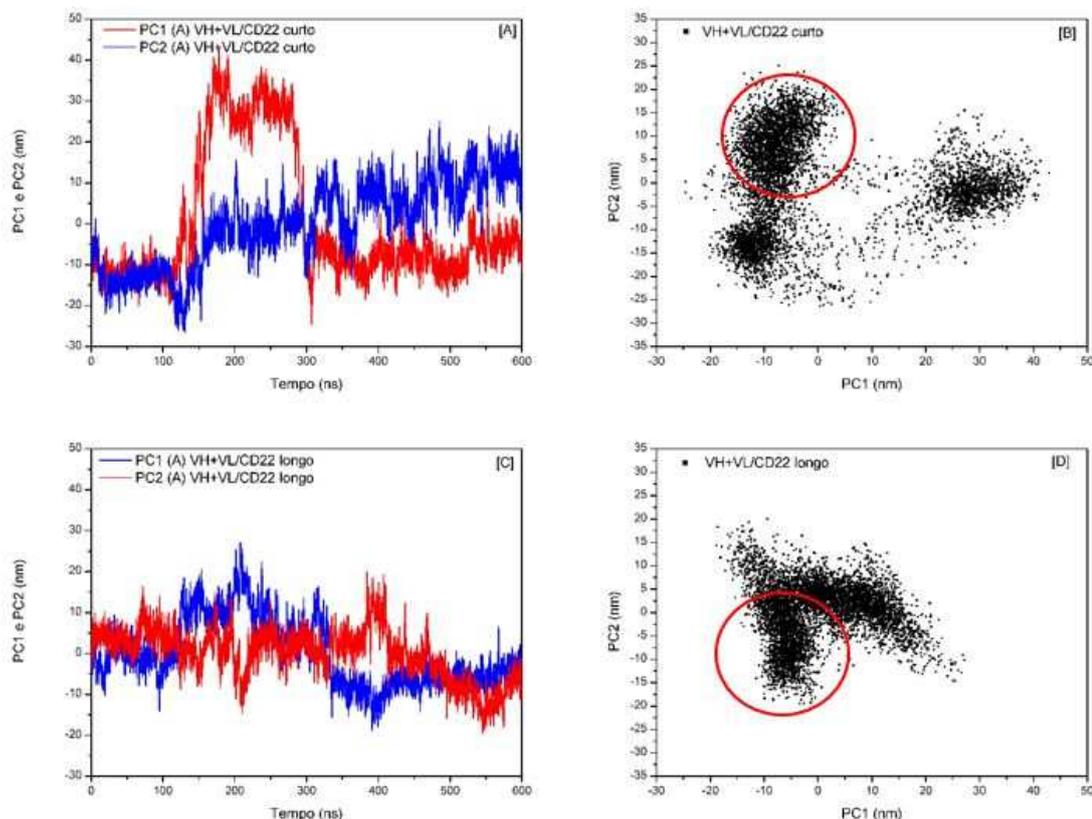
Figura 46 - Perfis de RMSD do CD22 e do VH+VL com sobreposição de CD22 dos sistemas scFv/CD22.



Fonte: elaborada pela autora (2024). A) CD22 e C) VH+VL do scFv de *linker* curto. B) CD22 e D) VH+VL do scFv de *linker* longo. Nota: CD22 desconsiderando glicosilações e *loops* modelados definidos na seção 8.1.3.

A Figura 47 apresenta os resultados obtidos das análises de PCA dos sistemas scFv/CD22 de *linker* curto e longo. Conforme a Figura 47A observa-se que as curvas de PC1 e PC2, para scFv de *linker* curto, variam bastante até ~300 ns de simulação. Após esse período, as curvas tendem a estabilizarem próximo de -10 nm para PC1 e 10 nm para PC2, indicando que o complexo scFv/CD22 se torna mais estável a partir desse período. Já na Figura 47C as curvas de PC1 e PC2 para scFv de *linker* longo variam bastante do começo ao fim da simulação, mas com amplitude menor que o de *linker* curto, sobretudo no início da DM, com oscilações entre -20 a 20 nm para ambas as projeções. As projeções nas Figuras 47B e 47D mostram que o complexo scFv/CD22 de *linker* curto e longo apresentam grande flexibilidade estrutural, uma vez que suas projeções nos subespaços comuns definidos por PC1/PC2 possuem três regiões bastante distintas. Contudo, as regiões para PC1 e PC2 > ~350 ns mostram grande concentração de pontos, indicando que ambos os sistemas, scFv/CD22 de *linker* curto e longo, possuem movimentações em níveis similares.

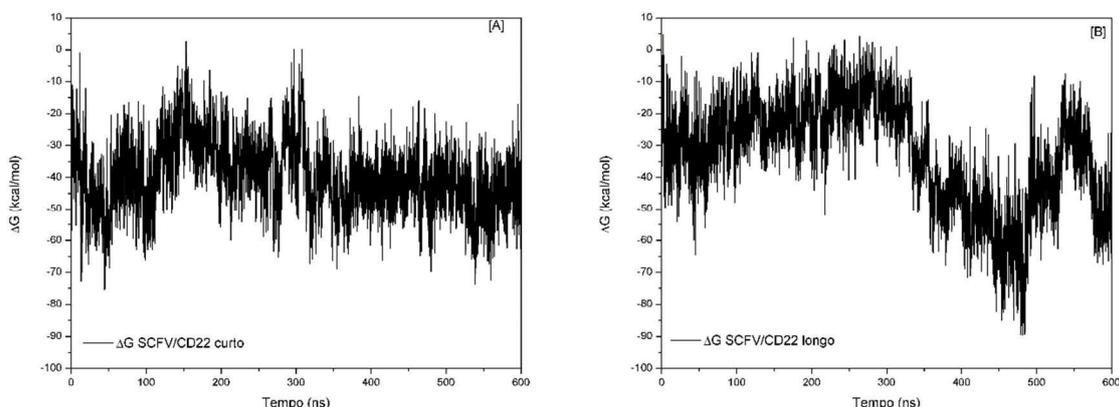
Figura 47 - Curvas de PCA para os sistemas scFv/CD22 curto e longo. A) e C) PC1 e PC2 em função do tempo durante 600 ns de simulação. B) e D) Projeções das curvas PC1/PC2.



Fonte: elaborada pela autora (2024). Foram considerados VH+VL do scFv e o CD22 sem as glicosilações e *loops* modelados definidos na seção 8.1.3.

Uma vez que os resultados de PCA indicam mudança de posicionamento de scFv em relação ao CD22, foi calculado o perfil de energia livre de ligação ($\Delta G_{\text{ligação}}$) entre eles, ao longo da DM, como mostrado na Figura 48. Sabendo que a organização estrutural das CDRs é fundamental para o resultado no cálculo do $\Delta G_{\text{ligação}}$ de ligação efetivo, os ΔG s foram calculados a partir de 350 ns. Os perfis de $\Delta G_{\text{ligação}}$ são similares entre os scFvs, com médias de $-44,4 \pm 8,5$ e $-46,9 \pm 14,0$ kcal mol⁻¹. Contudo, o perfil de $\Delta G_{\text{ligação}}$ para scFv de *linker* longo mostra uma acentuada queda na energia em ~490 ns, com ~ -90 kcal/mol, indicando que o scFv de *linker* longo encontra melhores posições de interação com o CD22.

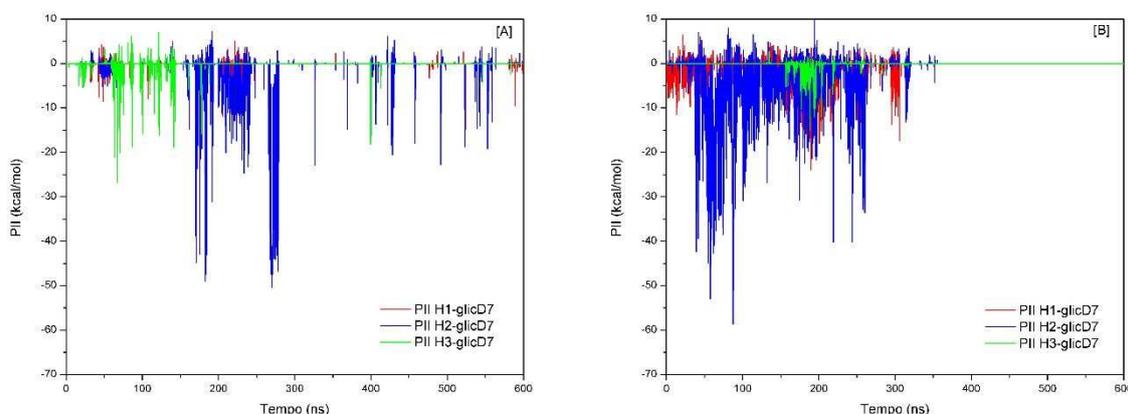
Figura 48 - Curvas de ΔG ligação dos complexos durante 600 ns de simulação entre o (A) scFv e o CD22 de linker curto e (B) scFv de linker longo.



Fonte: elaborada pela autora (2024).

Uma vez que o CD22 é glicosilado, mesmo as árvores estando em posição, inicialmente, oposta a interface scFv/CD22, com scFv em contato com o domínio D7 do CD22, os perfis mostram que em curto alcance, num raio de corte de 0,5 nm (RC = 0,5 nm) (Figura 49 A-B), há indicação de interação atrativa entre a glicosilação de D7 e as CDRs H dos scFvs. O PII não resulta em interação extremamente atrativa, e fica abaixo de -10 kcal/mol para *linker* curto e chega a zero para o longo, após ~350 ns. Importante observar que a partir de 350 ns também foi observado mudança no complexo scFv/CD22 por análise PCA e nos comportamentos das curvas de ΔG ligação, de forma que a árvore de glicosilação exerce influência na estabilidade da interface nesses sistemas, sendo que para o scFv de *linker* longo há maior flexibilidade na posição de interface com o CD22. Os perfis de PII, mostram também que a CDRs que mais interagem atrativamente como a glicosilação é a H2, como ratificado pelas médias na Tabela 20.

Figura 49 - PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22, A) curto. B) longo.



Fonte: elaborada pela autora (2024). Utilizando um raio de corte de 0,5 nm.

Tabela 20 - Médias e desvios de PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a glicosilação, no domínio D7 nos complexos scFv/CD22, nos sistemas com o scFv de linker curto e longo, dentro de um raio de corte de 0,5 nm.

scFv de <i>linker</i>	PII das CDRs (kcal/mol)		
	CDR-H1	CDR-H2	CDR-H3
curto	-0,5 ± 3,4	-5,7 ± 23,4	-0,7 ± 5,4
longo	-4,5 ± 11,4	-10 ± 26,4	-0,5 ± 3,6

Fonte: elaborada pela autora (2024).

Com o objetivo de verificar se a árvore de glicosilação ligada ao domínio extracelular D7 da molécula CD22 interage com as CDRs da cadeia leve do scFv de M971 de *linker* curto e longo, realizou-se a análise de PII entre a glicosilação do domínio extracelular D7 de CD22 e as CDRs da cadeia leve do scFv de M971 de *linker* curto e longo, dentro de um raio de corte de 0,5 nm. Para ambos os sistemas os valores de PII médio entre a glicosilação do domínio D7 de CD22 e as CDRs L1, L2 e L3 foram praticamente nulos (dados não mostrados), indicando que essas CDRs não contribuem para a mudança de posição dos scFvs em relação ao CD22.

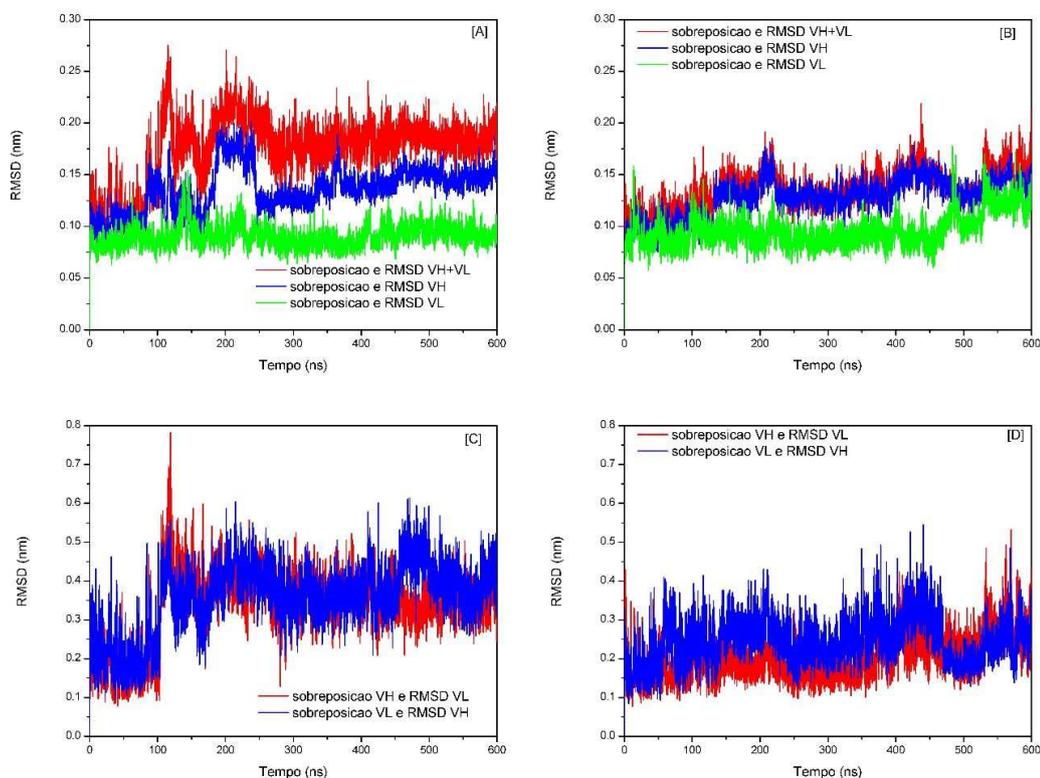
7.3.2 Estudo da interface scFv/CD22 com o complexo formado por meio das estruturas de scFv e CD22 sobreposta à estrutura cristalográfica scFv/CD22

Nesse sistema scFv/CD22 os scFvs e CD22 foram simulados separadamente em água e um processo de análise e escolha da melhor estrutura foi executado nas seções 8.1.3 (CD22) e 8.1.1 scFv de *linker* curto e 8.1.2 scFv de *linker* longo. Com as estruturas coletadas do equilíbrio, em solução, foram utilizadas para serem sobrepostas a estrutura cristalográfica PDB 7O52. Os passos de minimização e termalização seguiram como descrito na seção 6.2.6.

Na análise da trajetória, foi calculado os perfis de RMSD para o scFv de *linker* curto e longo nos sistemas scFv/CD22, como descritos na Figura 50A e 50B. Observa-se que a curva de RMSD de VH+VL está acima da de VH e VL, respectivamente. Além disso a forma das curvas entre VH+VL e VH são similares, indicando que movimentos atribuídos a VH+VL podem estar associados aos de VH. Foi escolhido $t_{eq} = 150$ ns para o cálculo das médias. Os RMSDs médios dos domínios do scFv, no sistema scFv/CD22 de *linker* curto e longo, estão descritos na Tabela 21.

Visando investigar se o movimento de VH-VL está relacionado ao VH, foi gerado perfis de RMSD de VH sobrepondo as estruturas de VL e vice-versa como mostrado na Figura 50C e 50D. As curvas de VH e VL mostram muita similaridade ratificando que há movimentação relativa entre VH e VL. No entanto, na comparação entre os scFvs nota-se que no *linker* curto, após ~ 100 ns, as curvas estão sobrepostas, demonstrando que a movimentação entre VH e VL está fortemente correlacionada, enquanto para *linker* longo há uma correlação, mas com uma clara influência de movimentação do VH. Os RMSD médios estão relacionados na Tabela 22.

Figura 50 - Perfis de RMSD para os domínios do scFv nos sistemas scFv/CD22, sendo que o scFv pode ser de linker curto ou longo.



Fonte: elaborada pela autora (2023). A) curto. B) longo. C) curto com sobreposição de VH e RMSD de VL e vice-versa. D) longo com sobreposição de VH e RMSD de VL e vice-versa.

Tabela 21 - Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL em ambos os sistemas scFv/CD22 linkers (curto e longo) a partir do $t_{eq} = 150$ ns.

Sistema	RMSD (nm)		
	VH+VL	VH	VL
curto	$0,19 \pm 0,02$	$0,14 \pm 0,02$	$0,09 \pm 0,01$
longo	$0,14 \pm 0,02$	$0,14 \pm 0,01$	$0,10 \pm 0,02$

Fonte: elaborada pela autora (2024).

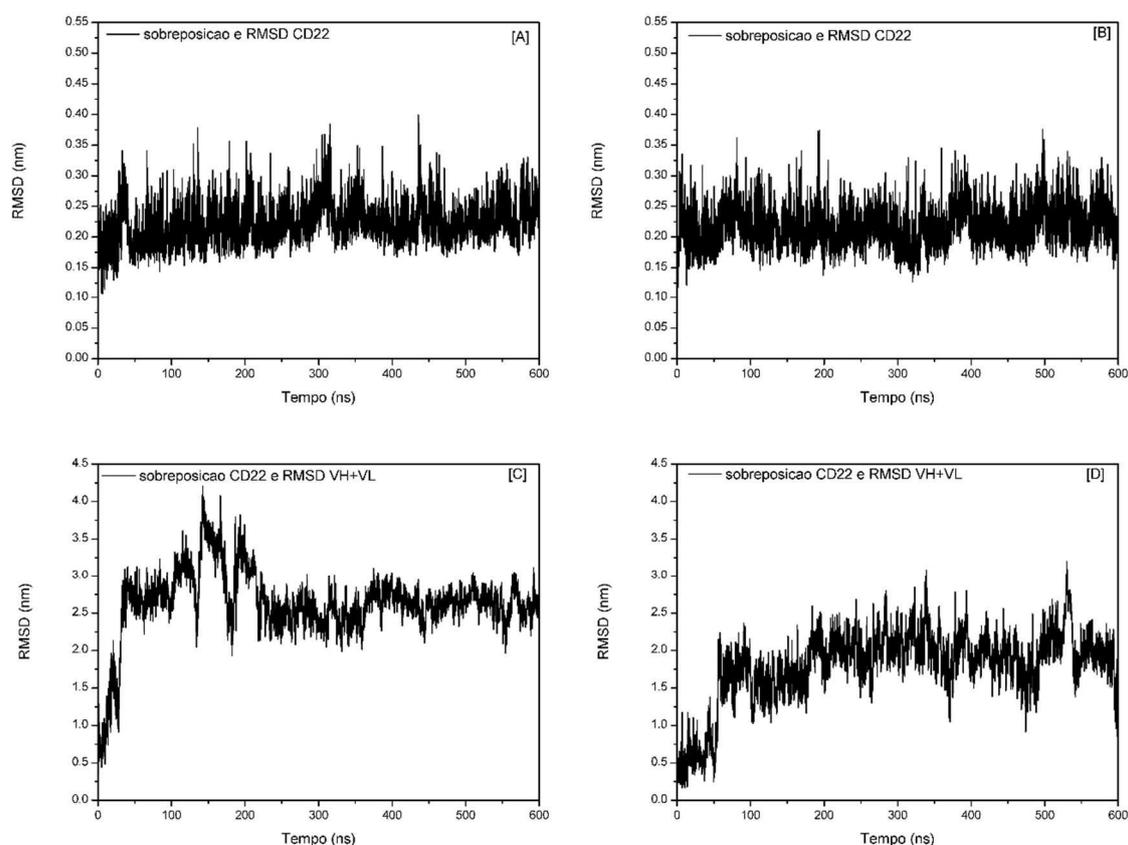
Tabela 22 - RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 150 ns.

Sistema	RMSD (nm)	
	VH Ref VL	VL Ref VH
curto	0,39 ± 0,06	0,36 ± 0,05
longo	0,26 ± 0,06	0,21 ± 0,06

Fonte: elaborada pela autora (2024). VH Ref VL = sobreposição de VL e RMSD de VH e VL Ref VH = sobreposição de VH e RMSD de VL.

As Figuras 51A e 51B apresentam os perfis de RMSD do CD22, excluindo os *loops* e glicosilações, no sistema scFv/CD22 de *linker* curto e longo, no decorrer de 600 ns de simulação, respectivamente. Os perfis de RMSD do CD22 oscilam ~ 0,25 nm. As médias de RMSD para o CD22 foram de $0,23 \pm 0,03$ e $0,22 \pm 0,03$ nm para scFv/CD22 de *linker* curto e longo. Os RMSDs mostram que embora o CD22 seja uma proteína com dois domínios (D6 e D7) ela não indica movimentações bruscas ao longo da DM. As Figuras 51C e 51D apresentam os perfis de RMSD do VH+VL em relação a CD22. Os perfis mostram RMSD alto indicando ampla mudança estrutural em ambos os complexos. As amplitudes de RMSD são maiores para o caso de scFv de *linker* curto (Figura 51C) que do *linker* longo (Figura 51D) indicando que há maior variação na posição do scFv de *linker* curto em relação ao CD22. Neste caso, as médias de RMSD foram de $2,66 \pm 0,31$ e $1,98 \pm 0,30$ para scFv/CD22 de *linker* curto e longo. Os altos RMSD sugerem que o scFv se deslocam ao longo da estrutura do CD22, mudando a interface. Contudo como é uma associação de 4 domínios, (D6-D7) e (VH-VL) movimentos coletivos entre todos eles pode elevar o RMSD. A estabilidade da interface deve ser estudada com maior profundidade.

Figura 51 - Perfis de RMSD do CD22 e do VH+VL com sobreposição de CD22 dos sistemas scFv/CD22.

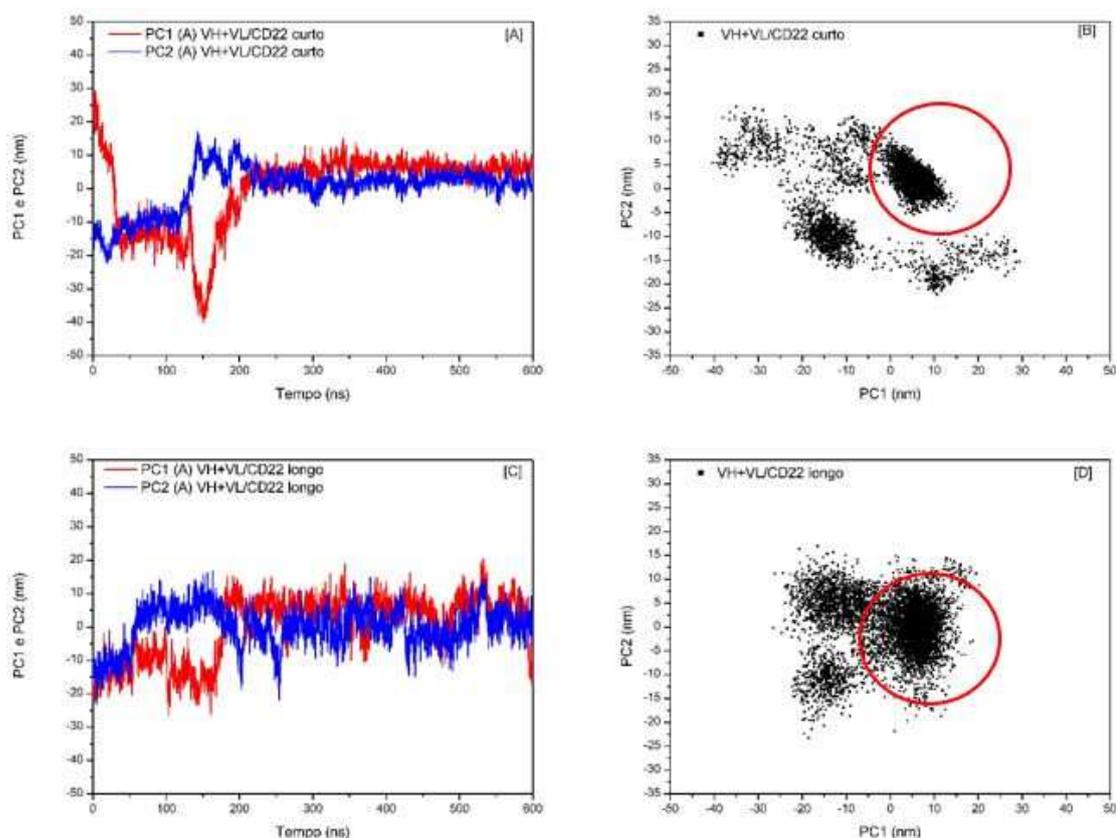


Fonte: elaborada pela autora (2024). A) CD22 e C) VH+VL do scFv de *linker* curto. B) CD22 e D) VH+VL do scFv de *linker* longo. Nota: CD22 desconsiderando glicosilações e *loops* modelados definidos na seção 8.1.3.

A Figura 52 representa os resultados obtidos por meio da análise de PCA dos complexos VH+VL/CD22 de *linker* curto e longo. Conforme a Figura 52A e 52C observa-se que as curvas de PC1 e PC2 variam bastante até ~200 ns de simulação. Após esse período, as curvas tendem a zero para PC1 e PC2, indicando que os complexos VH+VL/CD22 curto e longo se tornam mais estáveis a partir desse período. Pode-se observar ainda que na Figura 52B e 52D que o VH+VL de *linker* curto e longo apresenta grande flexibilidade estrutural em relação ao CD22, principalmente no início da simulação, uma vez que suas projeções nos subespaços comuns definidos por PC1/PC2 possuem três ou mais regiões bastante distintas. Essas regiões distintas correspondem as mudanças estruturais que ocorreram nos primeiros 200 ns de simulação em ambos os complexos. Contudo, as regiões para PC1 e PC2 > ~250 ns (Figura 52A e 52C) mostram grande concentração de pontos, indicando que ambos os sistemas, scFv/CD22 de *linker* curto e longo, possuem movimentações em níveis

similares, com indicação de maior flexibilidade para sistema com scFv de linker longo, uma região mais difusa na Figura 52B e D.

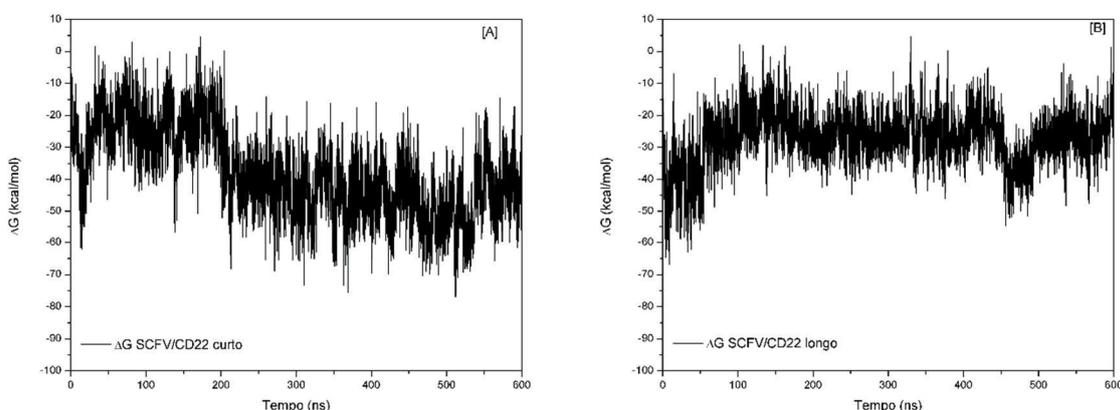
Figura 52 - Curvas de PCA para os sistemas scFv/CD22 curto e longo. A) e C) PC1 e PC2 em função do tempo durante 600 ns de simulação. B) e D) Projeções das curvas PC1/PC2.



Fonte: elaborada pela autora (2024).

Uma vez que os resultados de PCA indicam posicionamento estável entre de scFv/CD22 após 250 ns, a energia livre de ligação ($\Delta G_{\text{ligação}}$) média foi calculada a partir do perfil gerado ao longo da DM, como mostrado na Figura 53. Os perfis, para os dois sistemas, mostram $\Delta G_{\text{ligação}}$ diferentes, principalmente nos primeiros 200 ns. O que corrobora os períodos limitantes de movimentos indicados para estabilização da interface entre scFv/CD22, nas análises de PCA (Figura 53A e 53C). Após 250 ns as médias dos $\Delta G_{\text{ligação}}$ foram $-45,4 \pm 9,9$ e $-26,9 \pm 7,4$ kcal mol⁻¹, calculadas a partir de 250 ns. O perfil para scFv de *linker* curto mostra uma acentuada queda na energia em ~200 ns, com ~ -40 kcal/mol, indicando que o scFv de *linker* curto encontra melhor posição de interação com o CD22.

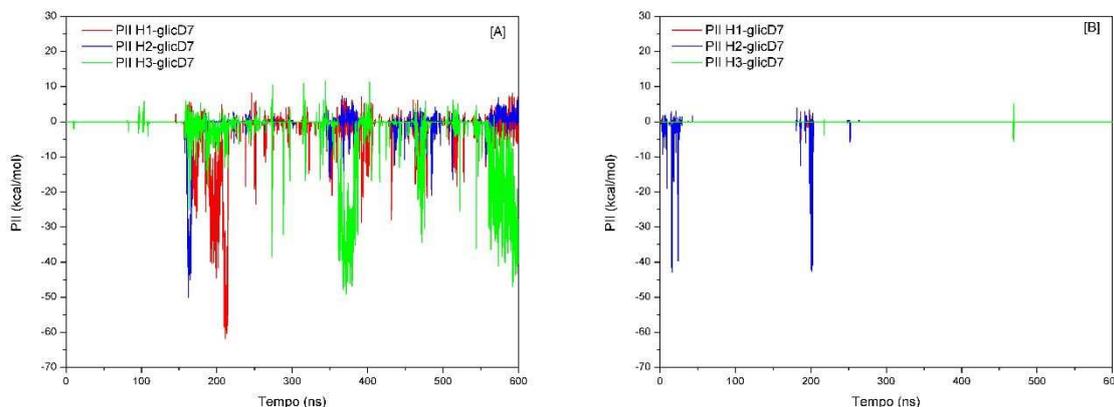
Figura 53 - Curvas de ΔG ligação dos complexos durante 600 ns de simulação entre o (A) scFv e o CD22 de linker curto e (B) scFv de linker longo.



Fonte: elaborada pela autora (2024).

O CD22 é glicosilado e do lado aposto a interface scFv/CD22, no domínio D7 do CD22, há uma glicosilação. Foi monitorado o PII entre a glicosilação e as CDRs do scFv como mostrado na Figura 54 A-B, os perfis de PII foram gerados para curto alcance (RC = 0,5 nm) há indicação de interação atrativa entre a glicosilação e as CDRs dos scFvs. O PII não resulta em interação extremamente atrativa, e fica abaixo de -50 kcal/mol para *linker* longo e chega a zero, após ~200 ns. No caso do scFv de *linker* curto o perfil mostra PII < ~ -70 kcal/mol para o *linker* curto, com manutenção da interação atrativa entre 150 à 600 ns. Todas as CDRs do VH no sistema contendo o scFv curto interagem atrativamente como a glicosilação como ratificado pelas médias na Tabela 23. Importante observar que a partir de 150 ns também foi conclusiva a mudança na interface scFv/CD22, por análise PCA e dos comportamentos das curvas de ΔG ligação. Portanto, a glicosilação ajuda a estabilizar a interface scFv/CD22 com *linker* curto. O scFv de *linker* longo não mostrou estabilidade na interface, sugerindo que a estrutura VH+VL seja mais flexível e, por esse motivo, não alcança estabilidade energia de ligação mais negativa.

Figura 54 - PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22. A) curto. B) longo.



Fonte: elaborada pela autora (2024). Utilizando um raio de corte de 0,5 nm.

Tabela 23 - Médias e desvios de PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a glicosilação, no domínio D7 nos complexos scFv/CD22, nos sistemas com o scFv de linker curto e longo, dentro de um raio de corte de 0,5 nm.

scFv de <i>linker</i>	PII das CDRs (kcal/mol)		
	CDR-H1	CDR-H2	CDR-H3
curto	-8,9 ± 28,9	-2,1 ± 15,7	-14,6 ± 38,1
longo	-0,1 ± 1,2	-1,5 ± 11,0	-0,1 ± 0,6

Fonte: elaborada pela autora (2024).

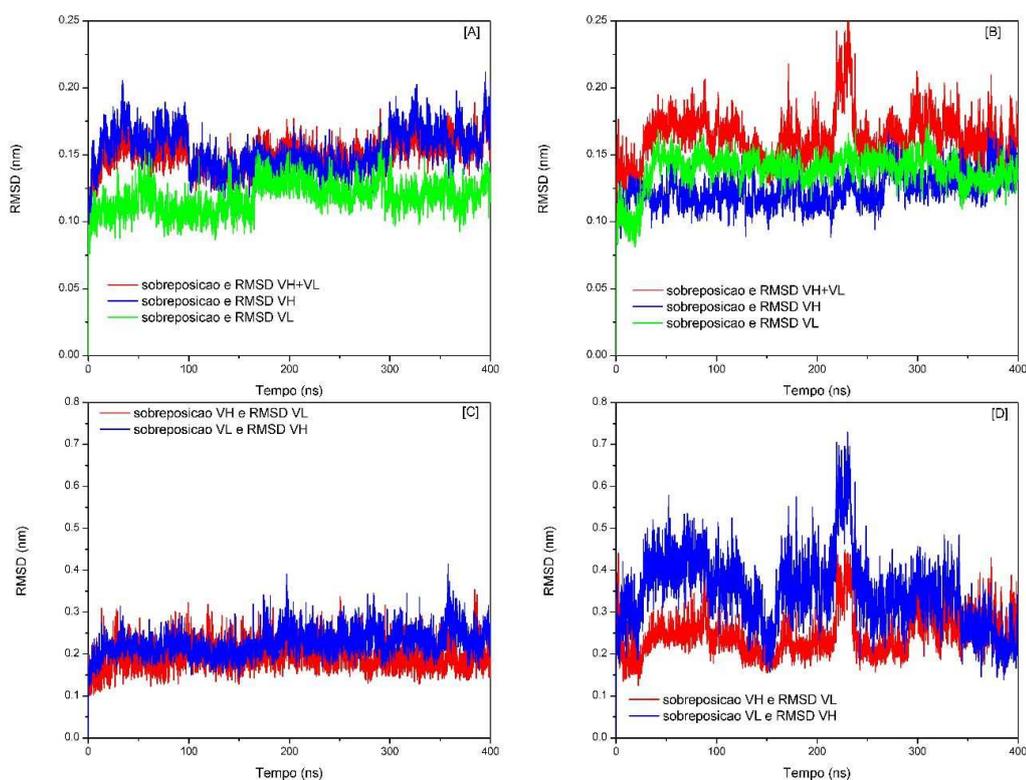
Com o objetivo de verificar se a árvore de glicosilação ligada ao domínio extracelular D7 da molécula CD22 interage com as CDRs da cadeia leve do scFv de M971 de *linker* curto e longo, realizou-se a análise de PII entre a glicosilação do domínio extracelular D7 de CD22 e as CDRs da cadeia leve do scFv de M971 de *linker* curto e longo, dentro de um raio de corte de 0,5 nm. Para ambos os sistemas os valores de PII médio entre a glicosilação do domínio D7 de CD22 e as CDRs L1, L2 e L3 foram praticamente nulos (dados não mostrados), indicando que essas CDRs não contribuem para a mudança de posição ou estabilização da interface para os scFvs em relação ao CD22.

7.3.3 Estudo da interface scFv/CD22 usando as coordenadas cristalográficas

Nesse sistema, foram usadas as coordenadas cristalográficas do CD22 e do scFv (scFv/CD22) no código PDB, segundo a seção 2.3. A minimização e termalização ocorreram em seguida, como descrito na seção 6.2.7. O RMSD foi calculado para avaliar mudanças estruturais do scFv de *linker* curto e longo, nos sistemas scFv/CD22, como descrito na Figura 55A e B. Nota-se que as curvas para VH e VL oscilam ao redor de ~0,13 nm. As curvas mostram acréscimos rápidos de RMSD nos primeiros ns de DM, indicando um $\tau = 50$ ns para ocorrer o equilíbrio estrutural do scFv. A curva de VH+VL mostra oscilações mais bruscas que as demais no RMSD. Os RMSDs médios dos domínios do scFv no sistema scFv/CD22 de *linker* curto e longo estão descritos na Tabela 24.

Visando investigar se o movimento de VH-VL está relacionado ao VH, foi gerado perfis de RMSD de VH sobrepondo as estruturas de VL e vice-versa como mostrado na Figura 55C e 55D. As curvas de VH e VL mostram muita similaridade ratificando que há movimentação relativa entre VH e VL. No entanto, em comparação entre os scFvs, nota-se que no *linker* curto, após ~100 ns as curvas estão sobrepostas, com baixa amplitude RMSD enquanto no longo a amplitude é maior. Isso demonstra que o scFv de *linker* curto mantém as estruturas de VH e VL e que elas se movimentam pouco entre elas, enquanto no *linker* longo a movimentação é mais intensa. Os RMSD médios estão relacionados na Tabela 25.

Figura 55 - Perfis de RMSD para os domínios do scFv nos sistemas scFv/CD22, sendo que o scFv pode ser de linker curto ou longo.



Fonte: elaborada pela autora (2024). A) curto. B) longo. C) curto com sobreposição de VH e RMSD de VL e vice-versa. D) longo com sobreposição de VH e RMSD de VL e vice-versa.

Tabela 24 - Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL em ambos os sistemas scFv/CD22 linkers (curto e longo) a partir do $t_{eq} = 50$ ns.

Sistema	RMSD (nm)		
	VH+VL	VH	VL
curto	$0,15 \pm 0,01$	$0,15 \pm 0,02$	$0,12 \pm 0,01$
longo	$0,16 \pm 0,02$	$0,13 \pm 0,01$	$0,14 \pm 0,01$

Fonte: elaborada pela autora (2024).

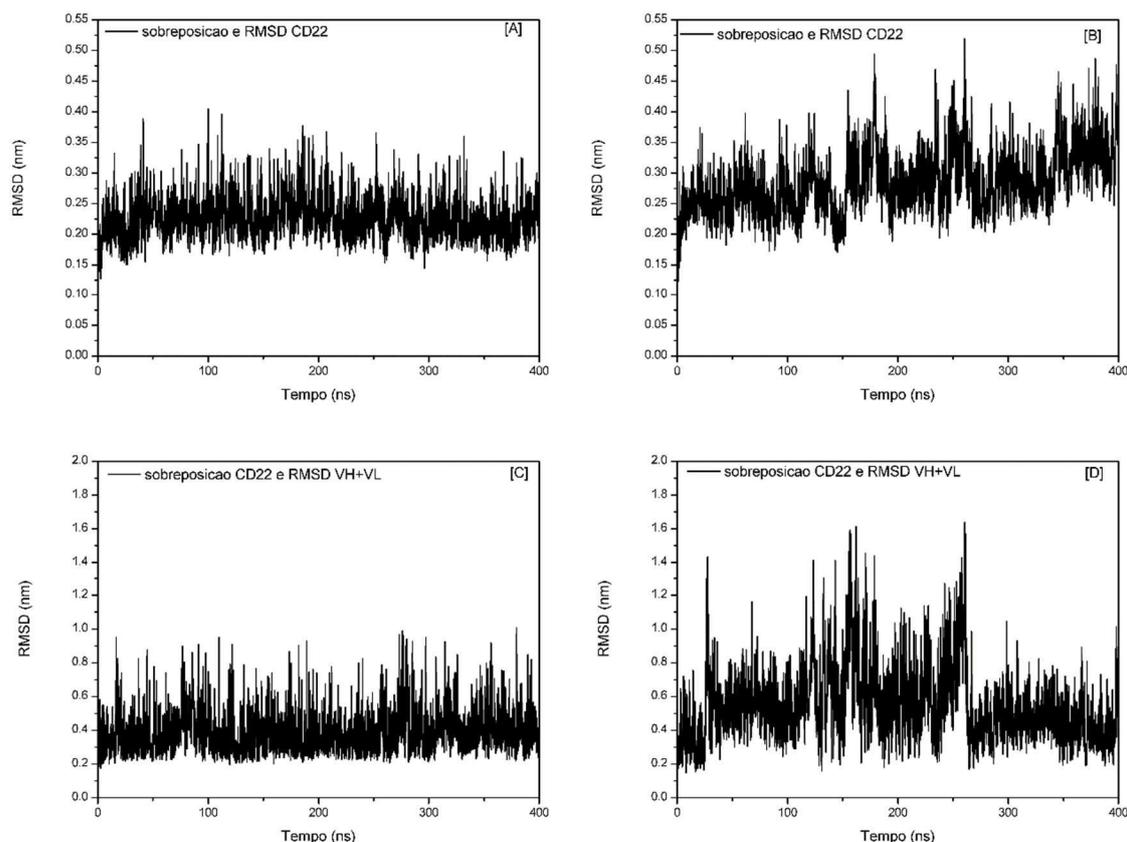
Tabela 25 - RMSDs médios do domínio VH tendo como referência VL, e vice-versa, calculados a partir do tempo de 50 ns.

Sistema	RMSD (nm)	
	VH Ref VL	VL Ref VH
curto	$0,23 \pm 0,03$	$0,20 \pm 0,03$
longo	$0,35 \pm 0,09$	$0,25 \pm 0,05$

Fonte: elaborada pela autora (2024). VH Ref VL = sobreposição de VL e RMSD de VH e VL Ref VH = sobreposição de VH e RMSD de VL.

As Figuras 56A e 56B apresentam os perfis de RMSD do CD22, excluindo os *loops* e glicosilações, no sistema scFv/CD22 de *linker* curto e longo, no decorrer de 400 ns de simulação, respectivamente. O perfil de RMSD para o CD22 está em $\sim 0,20$ nm, com altas nas amplitudes, principalmente para o sistema com scFv longo. As médias de RMSD para o CD22 foram de $0,23 \pm 0,04$ e $0,29 \pm 0,05$ nm para scFv/CD22 de *linker* curto e longo. As Figuras 56C e 56D apresentam os perfis de RMSD do VH+VL em relação a CD22. Os perfis mostram RMSD alto indicando ampla mudança estrutural em ambos os complexos. As amplitudes de RMSD são maiores para o caso de scFv de *linker* longo (Figura 56D) que do *linker* curto (Figura 56C) indicando que há maior variação na posição do scFv de *linker* longo em relação ao CD22. Neste caso, as médias de RMSD foram de $0,40 \pm 0,13$ e $0,57 \pm 0,23$ para scFv/CD22 de *linker* curto e longo.

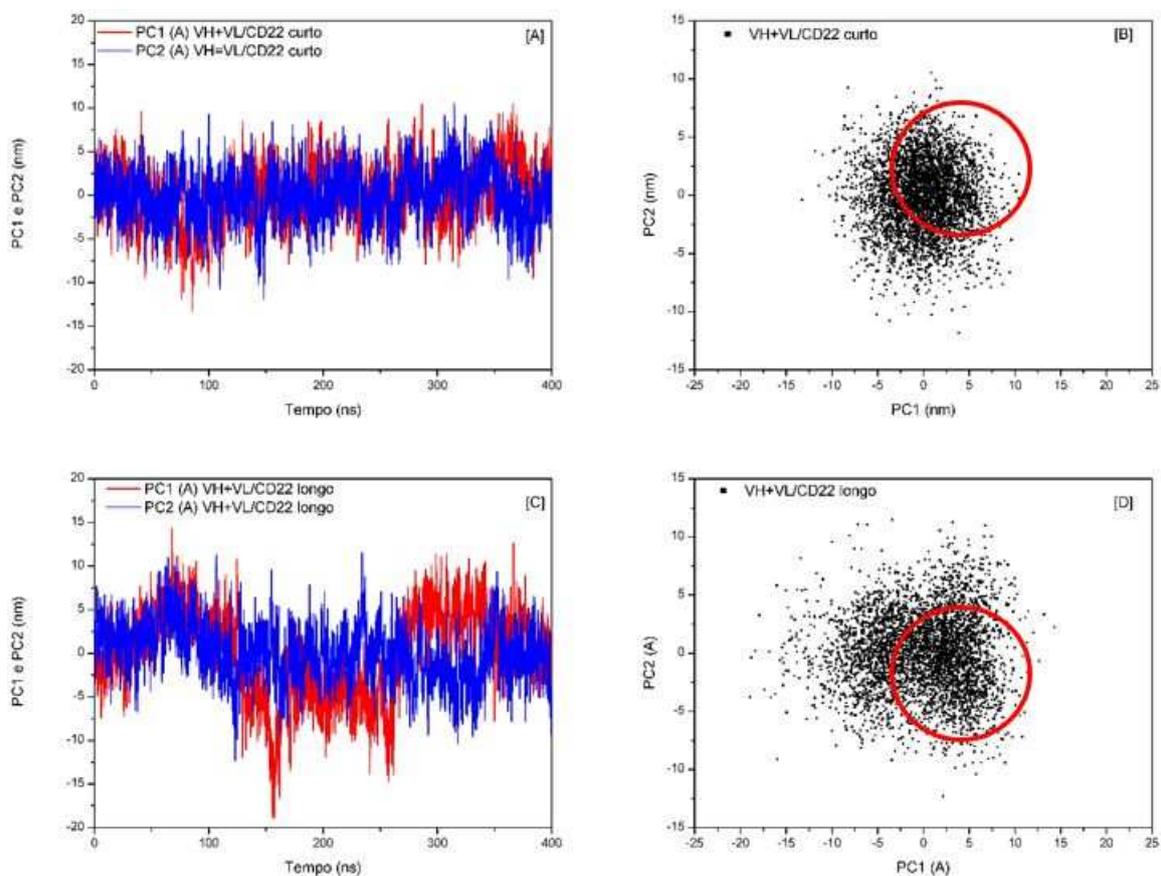
Figura 56 - Perfis de RMSD do CD22 e do VH+VL com sobreposição de CD22 dos sistemas scFv/CD22.



Fonte: elaborada pela autora (2024). A) CD22 e C) VH+VL do scFv de *linker* curto. B) CD22 e D) VH+VL do scFv de *linker* longo. Nota: CD22 desconsiderando glicosilações e *loops* modelados definidos na seção 8.1.3.

A Figura 57 representa os resultados obtidos por meio da análise de PCA dos complexos VH+VL/CD22 de *linker* curto e longo. Conforme a Figura 57A e 57C observa-se que as curvas de PC1 e PC2 variam mais no sistema de *linker* longo que no sistema de *linker* curto. No sistema de *linker* curto as projeções PC1 e PC2 oscilam de ~ 5 a ~ -5 nm do começo ao fim da simulação. Já no sistema de *linker* longo, as projeções PC1 e PC2 oscilam de ~ 10 a ~ -15 nm do começo ao fim da simulação. Isso é ratificado observando as projeções nos subespaços comuns definidos por PC1/PC2, com pontos mais difusos para o sistema com scFv de *linker* longo. As movimentações intrínsecas entre os scFv de *linker* curto e longo, mostram ser distintas e que exerce influência na estabilidade da interface, como visto nas análises de PCA e indicado por RMSD. Nesse sentido, a estabilidade da interface scFv/CD22 foi investigada usando o cálculo de energia livre de ligação.

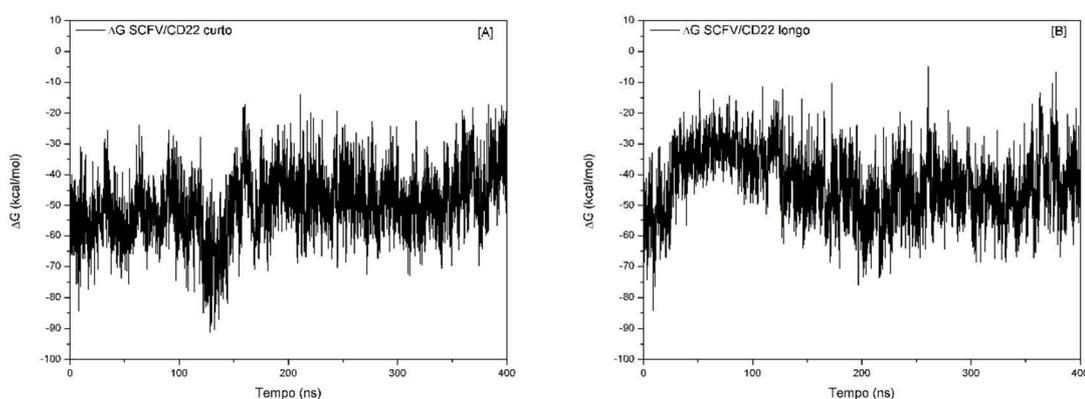
Figura 57 - Curvas de PCA para os sistemas scFv/CD22 curto e longo. A) e C) PC1 e PC2 em função do tempo durante 400 ns de simulação. B) e D) Projeções das curvas PC1/PC2.



Fonte: elaborada pela autora (2024).

Foi usado MM/PBSA para calcular a energia livre de ligação ($\Delta G_{\text{ligação}}$) entre scFv e CD22 de *linker* curto e longo. Os perfis de $\Delta G_{\text{ligação}}$ dos sistemas scFv/CD22 de *linker* curto e longo são mostrados na Figura 58. Nota-se oscilações de $\Delta G_{\text{ligação}}$ rápidas e de amplitudes similares, em ambos os casos. As médias de $\Delta G_{\text{ligação}}$ ratificam que as curvas se mostram similares, $-49,3 \pm 11$ kcal/mol para sistema com scFv de linker curto e $-42,8 \pm 10,7$ kcal/mol para o longo.

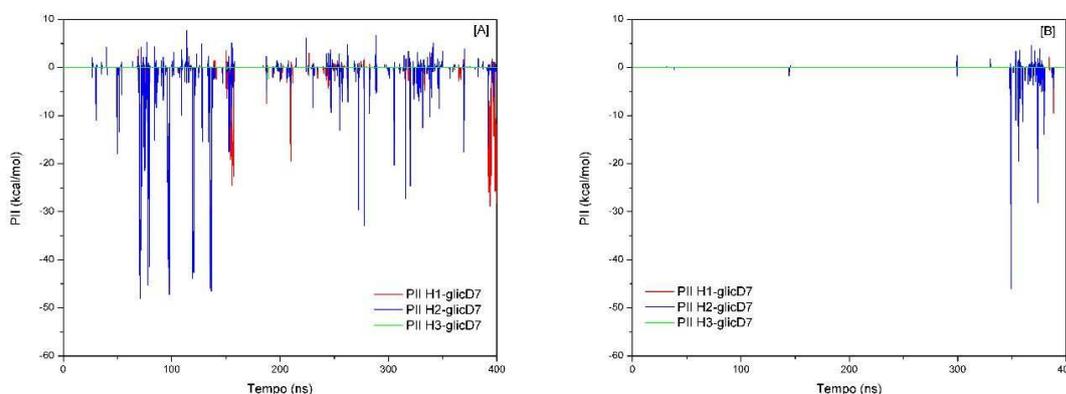
Figura 58 - Curvas de $\Delta G_{\text{ligação}}$ dos complexos durante 400 ns de simulação entre o (A) scFv e o CD22 de linker curto e (B) scFv de linker longo.



Fonte: elaborada pela autora (2024).

Foi investigado se a glicosilação no domínio D7, do CD22, interage atrativamente com o scFV. Foi calculado o PII de curto alcance, num raio de corte de 0,5 nm (RC = 0,5 nm) (Figura 59 A-B). Para o sistema de *linker* curto, há interação atrativa da CDR H1, mas principalmente H2, durante toda a DM e para *linker* longo a interação atrativa aparece no final da DM. O PII é ~ -50 kcal/mol para a CDR H2, como ratificado pelas médias na Tabela 26. Nesse caso, não houve indicação de que a interação atrativa contribísse para a mudança de posição do scFv de *linker* curto e longo em relação ao CD22. Contudo, pode explicar por que o CD22 no sistema de *linker* curto apresentou menores RMSDs ao longo da DM ver Figura 56 A e C.

Figura 59 - PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22. A) curto. B) longo.



Fonte: elaborada pela autora (2024). Utilizando um raio de corte de 0,5 nm.

Tabela 26 - Médias e desvios de PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a glicosilação, no domínio D7 nos complexos scFv/CD22, nos sistemas com o scFv de linker curto e longo, dentro de um raio de corte de 0,5 nm.

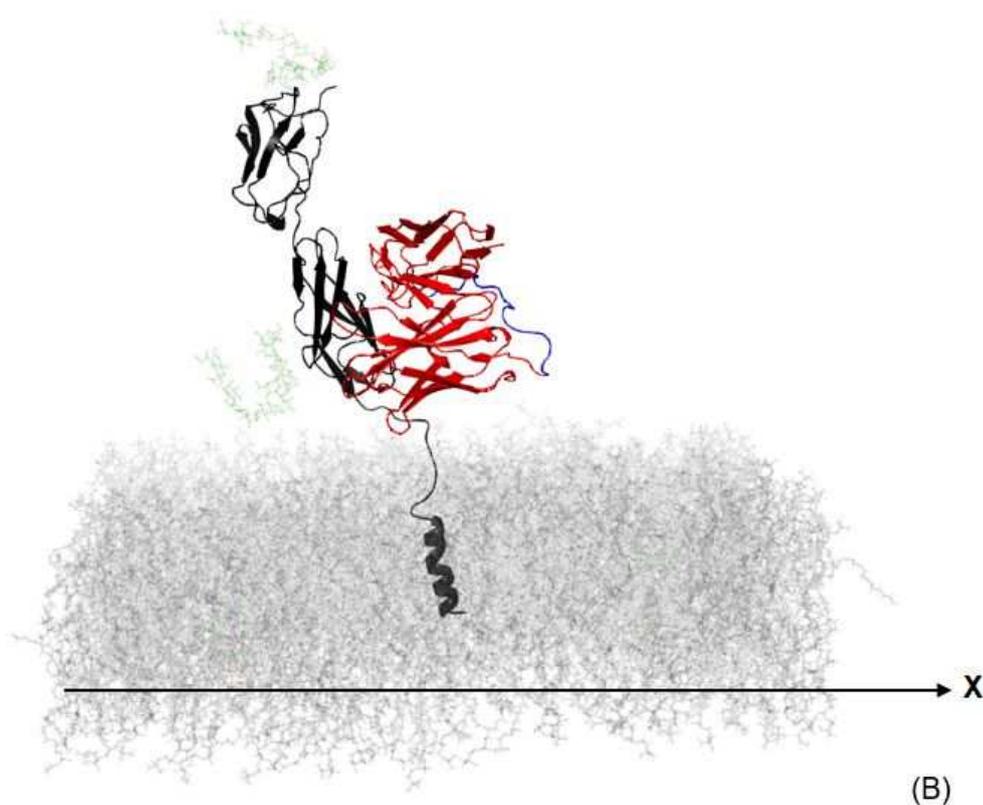
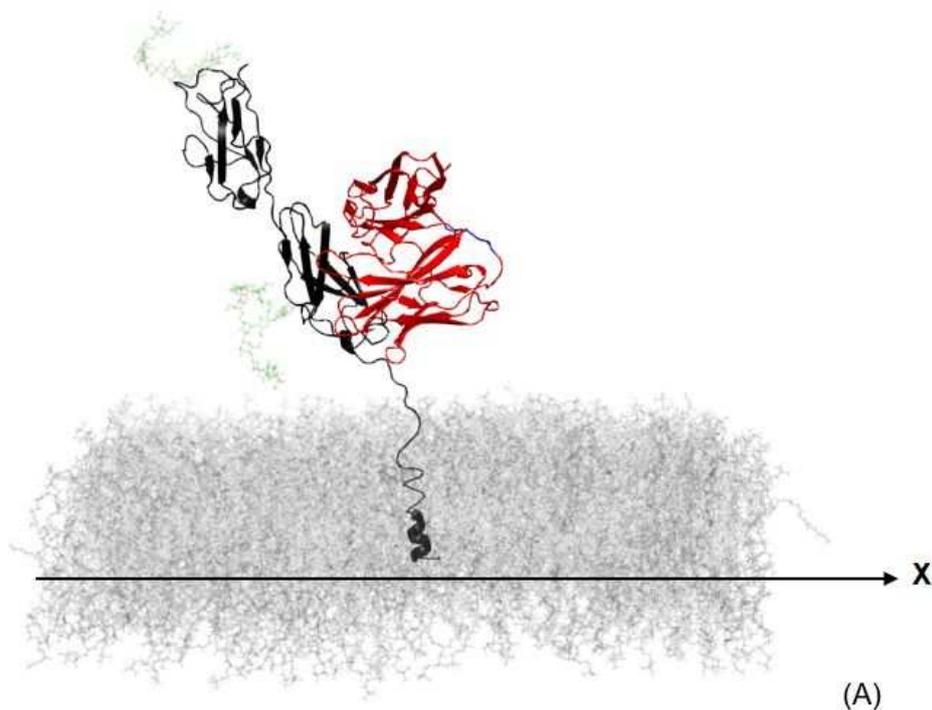
scFv de <i>linker</i>	PII das CDRs (kcal/mol)		
	CDR-H1	CDR-H2	CDR-H3
curto	-1,9 ± 11,6	-4 ± 21,5	-0,1 ± 0,2
longo	-0,1 ± 0,8	-0,5 ± 6,9	0

Fonte: elaborada pela autora (2024).

7.4 Estudo da interface scFv/CD22, com scFv de linker curto e longo e o CD22 inserido em membrana

Os sistemas scFv/CD22-t foram montados conforme descrito na seção 6.2.8. Os sistemas scFv/CD22-t em membrana, com scFvs de *linker* curto e longo, foram simulados em triplica (n1, n2 e n3) durante 400 ns. A configuração inicial (Figura 60) foi termalizada e criadas as triplicatas dos sistemas, que iniciaram a partir de velocidades atômicas distintas.

Figura 60 - Representação estrutural da configuração inicial dos sistemas scFv/CD22-t curto e longo.

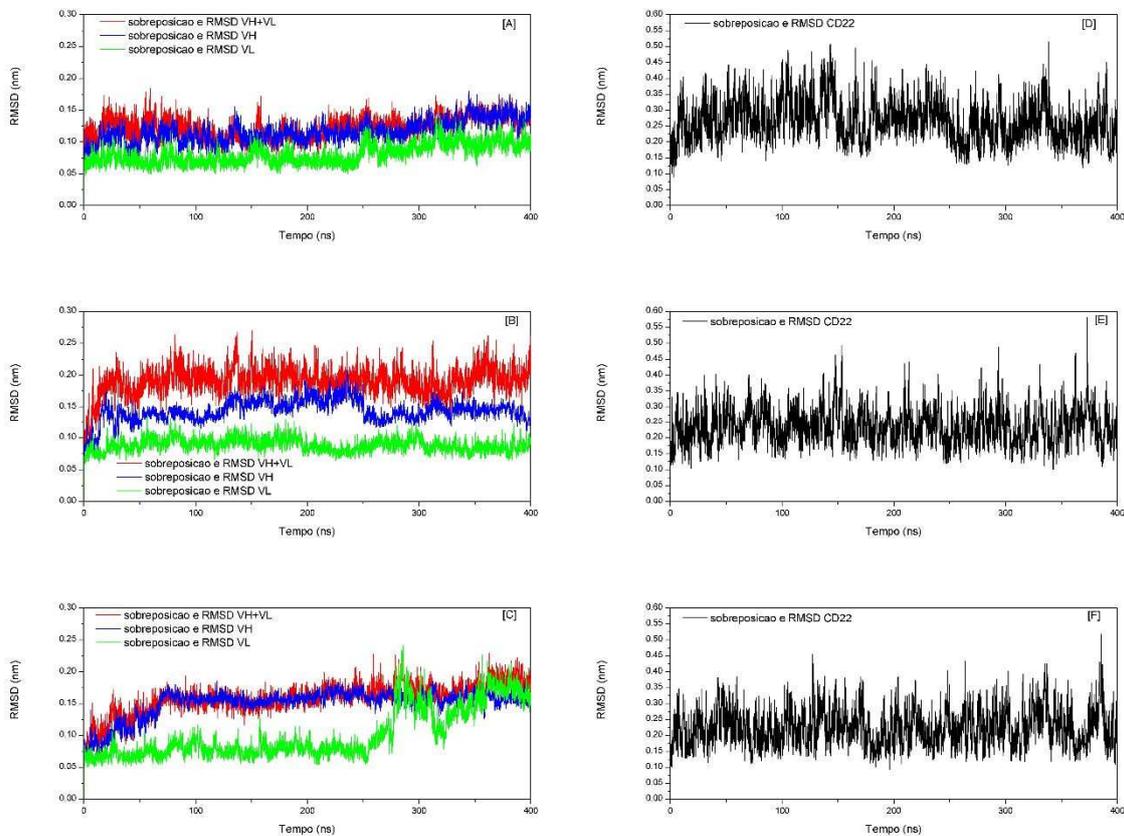


Fonte: elaborada pela autora (2024). (A) complexo de *linker* curto. (B) complexo de *linker* longo. Em ambos os complexos na cor verde as árvores de glicosilação, na cor preto o CD22-t (CD22 com o transmembrana), em vermelho os domínios VH e VL dos scFvs e em azul o *linker*. Nota: Representação dos sistemas tendo a bicamada lipídica na horizontal (eixo X) e transmembrana atravessando a bicamada, na superfície o CD22. Nota: as moléculas de água e íons foram retiradas da imagem para melhor visualização dos sistemas.

Foram gerados perfis de RMSD para o sistema scFv/CD22-t de *linker* curto (Figura 61) e longo (Figura 62), a denominação CD22-t é decorrente do CD22 estar com o domínio transmembrana inserido no modelo de membrana (Ver Figura 42). Para o sistema *linker* curto nota-se as curvas com RMSD máximo abaixo de ~ 0,20 nm e RMSD do VL com os menores RMSDs, entre a triplicata. No caso do scFv de *linker* longo, as curvas, apresentam RMSDs abaixo de 0,15 nm. Os RMSDs indicam uma necessidade das estruturas VH+VL se acomodarem no scFv de linker curto, enquanto para o longo esse ajuste é mínimo. Foi considerado $t_{eq} = 75$ ns no cálculo das médias, sendo que em algumas curvas de VH e VL a estrutura atinge um equilíbrio antes desse tempo. As médias, na tabela 27, ratificam a observação de que o scFv de *linker* longo apresenta uma estruturação mais estável que o curto, na interface com CD22 em membrana.

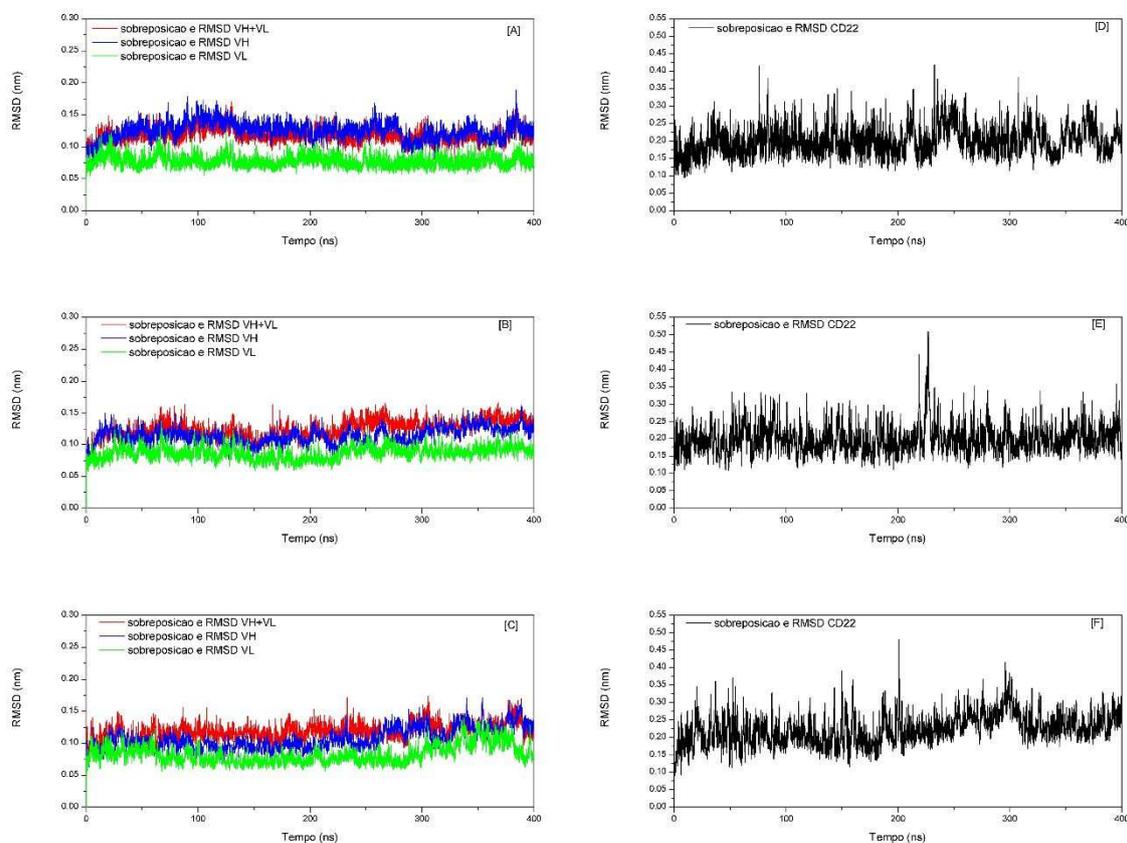
Os perfis para CD22 mostram alta frequência nas mudanças de RMSD e amplitude, para os dois sistemas. No caso, do sistemas com scFv de *linker* curto (Figura 61D, E, F), os RMSDs são minimamente maiores que o longo (Figura 62D, E, F), As médias e desvios de RMSD para o CD22 estão descritos na Tabela 28, ratificando a observação de que o RMSD do CD22 é superior no sistema de scFv com *linker* longo.

Figura 61 - Perfis de RMSD para os domínios do scFv e CD22 do sistema scFv/CD22-t, scFv de linker curto. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas os RMSDs VH+VL, VH e VL (A, B e C) e CD22 (D, E e F).



Fonte: elaborada pela autora (2024). Vermelho: sobreposição e cálculo de RMSD do VH+VL; Azul: sobreposição e cálculo de RMSD do VH; Verde: sobreposição e cálculo de RMSD do VL e Preto: sobreposição. Nota: cálculo de CD22 sem glicosilações, *loops* modelados e domínio transmembrana.

Figura 62 - Perfis de RMSD para os domínios do scFv e CD22 do sistema scFv/CD22-t, scFv de linker longo. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas os RMSDs VH+VL, VH e VL (A, B e C) e CD22 (D, E e F).



Fonte: elaborada pela autora (2024). Vermelho: sobreposição e cálculo de RMSD do VH+VL; Azul: sobreposição e cálculo de RMSD do VH; Verde: sobreposição e cálculo de RMSD do VL e Preto: sobreposição e cálculo de CD22-t. Nota: cálculo de CD22 sem glicosilações, *loops* modelados e domínio transmembrana.

Tabela 27 - Médias e desvios de RMSD calculados para os domínios VH+VL, VH e VL da triplicata dos sistemas scFv/CD22 de linker curto e longo em membrana a partir de 75 ns.

scFv(linker curto)/CD22-t				scFv(linker longo)/CD22-t		
RMSD/nm				RMSD/nm		
N*	VH+VL	VH	VL	VH+VL	VH	VL
n1	0,12 ± 0,02	0,12 ± 0,02	0,08 ± 0,02	0,12 ± 0,01	0,13 ± 0,01	0,08 ± 0,01
n2	0,19 ± 0,02	0,15 ± 0,01	0,09 ± 0,01	0,13 ± 0,01	0,11 ± 0,01	0,09 ± 0,01
n3	0,17 ± 0,02	0,16 ± 0,01	0,11 ± 0,04	0,12 ± 0,01	0,11 ± 0,02	0,08 ± 0,02

Fonte: elaborada pela autora (2024). *N – simulações em triplicata (n1, n2 e n3).

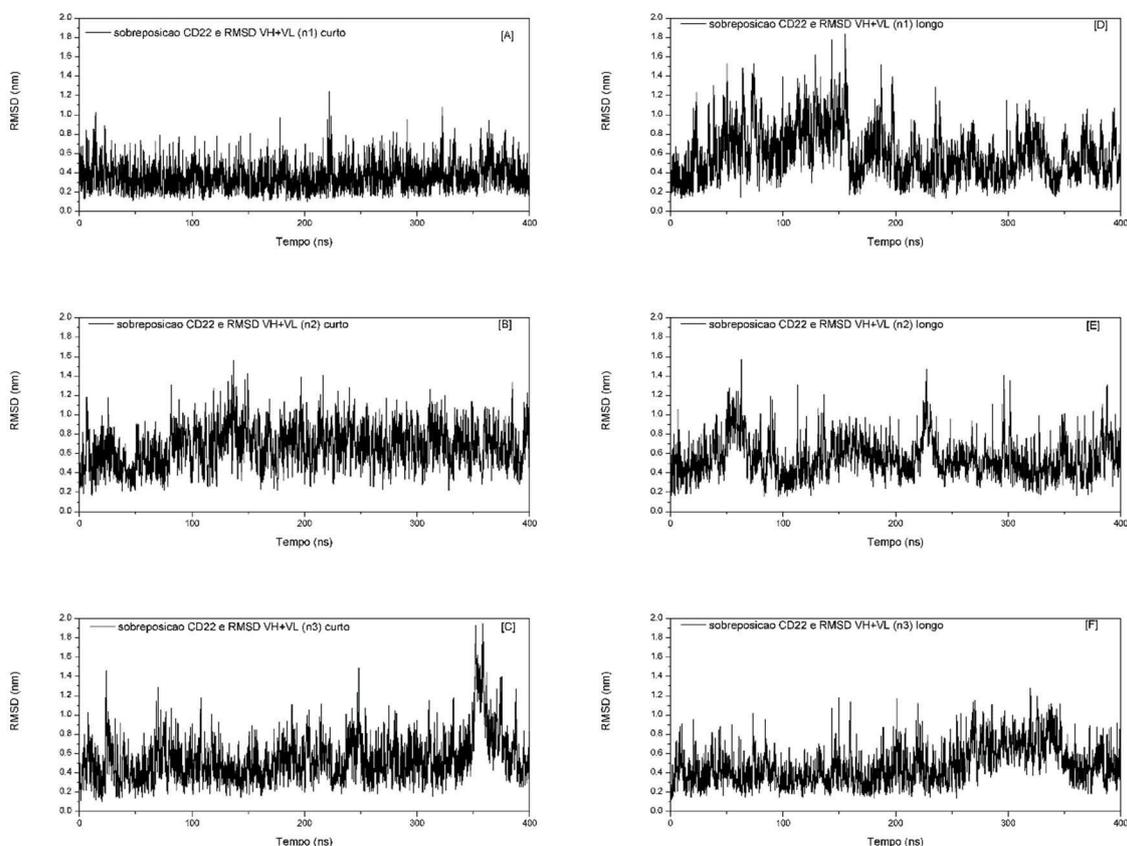
Tabela 28 - Médias e desvios de RMSD calculados para CD22 dos sistemas scFv/CD22-t curto e longo a partir de 75 ns de simulação.

	scFv (curto)/CD22-t RMSD/nm	scFv (longo)/CD22-t RMSD/nm
N*	CD22*	CD22*
n1	0,27 ± 0,06	0,21 ± 0,04
n2	0,24 ± 0,06	0,20 ± 0,04
n3	0,23 ± 0,06	0,23 ± 0,04

Fonte: elaborada pela autora (2024). *Nota: cálculo de CD22 sem glicosilações, *loops* e domínio transmembrana.

Foi investigado se o scFv mantém uma posição estável na interface com CD22-t. Foram geradas curvas de RMSD de VH+VL, sobrepondo as estruturas do CD22, como mostra as curvas na Figura 63. É importante salientar que os RMSDs do CD22 (sem glicosilações, *loops* modelados e domínio transmembrana) estão $< \sim 0,27$ nm com alto desvio (Tabela 28). Os perfis, na Figura 81, mostram grandes amplitudes de RMSDs, associados as mudanças frequentes ao longo do tempo, com exceção para n1 do sistema com scFv de linker curto, cuja o RMSD é menor que os demais. As médias mostradas na tabela 29 não permitem afirmar uma tendência de mudança ou maior modificação da interface inicial, entre os scFvs (linker curto ou longo). Os perfis na Figura 63 têm formas próximas e RMSDs bem mais elevados, aos perfis observados no RMSD somente do CD22 (Figuras 61 e 62), o que indica muita interferência da movimentação do CD22, inviabilizando a análise com clareza. Visando esclarecer melhor a movimentação na interface, foi feita a análise de PCA para os scFvs na interface com CD22, como mostra as Figuras 64 e 65.

Figura 63 - Perfis de RMSD de VH+VL após sobreposição do CD22 dos sistemas scFv/CD22-t curto e longo.



Fonte: elaborada pela autora (2024). Nota: CD22 desconsiderando *loops* modelados, glicosilações e o domínio transmembrana. O scFv desconsiderando o *linker*. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas os RMSDs de VH+VL com sobreposição de CD22 para o sistema de *linker* curto (A, B e C) e RMSDs de VH+VL com sobreposição de CD22 para o sistema de *linker* longo (D, E e F).

Tabela 29 - Médias e desvios de RMSD calculados para sobreposição CD22 e RMSD VH+VL dos sistemas scFv/CD22-t curto e longo a partir de 75 ns de simulação.

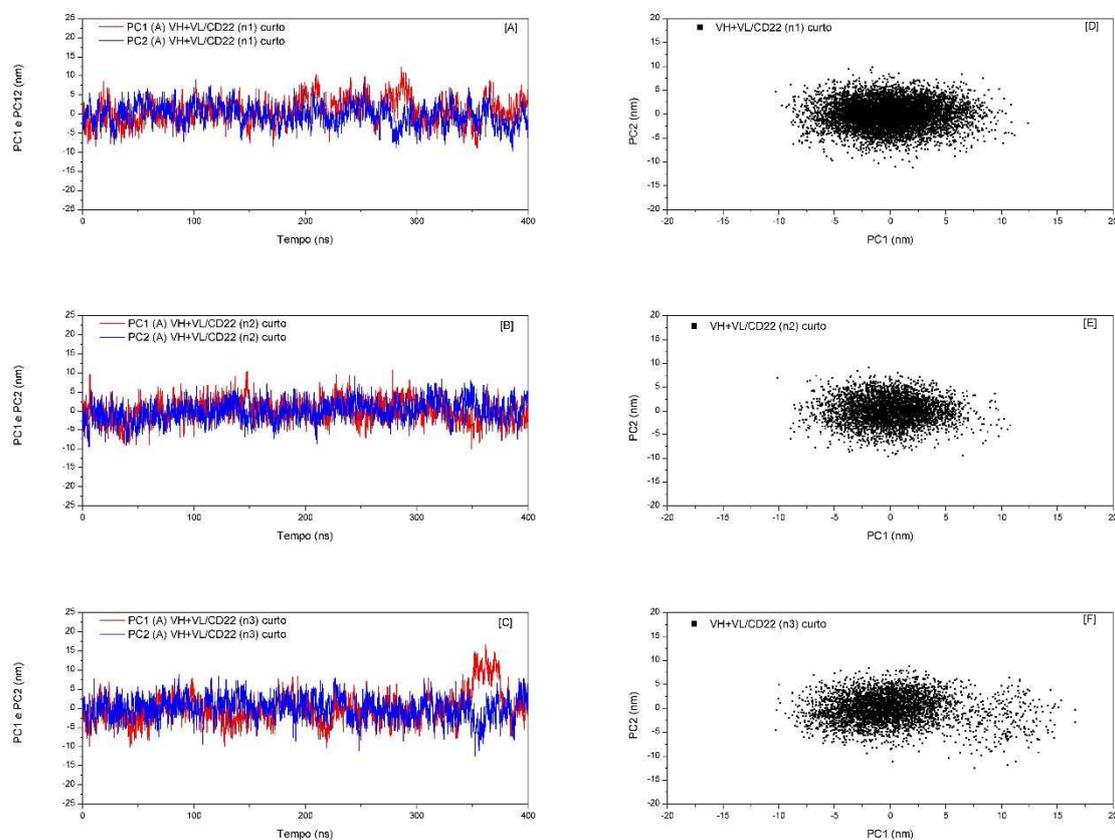
	scFv (curto)/CD22-t RMSD/nm	scFv (longo)/CD22-t RMSD/nm
N*	VH+VL	VH+VL
n1	0,36 ± 0,13	0,59 ± 0,25
n2	0,71 ± 0,20	0,54 ± 0,19
n3	0,56 ± 0,25	0,50 ± 0,20

Fonte: elaborada pela autora (2024).

As Figuras 64 e 65 mostram as curvas de PCA obtidas para os complexos scFv/CD22-t de *linker* curto e longo. Conforme a Figura 64A-C e 65A-C observa-se

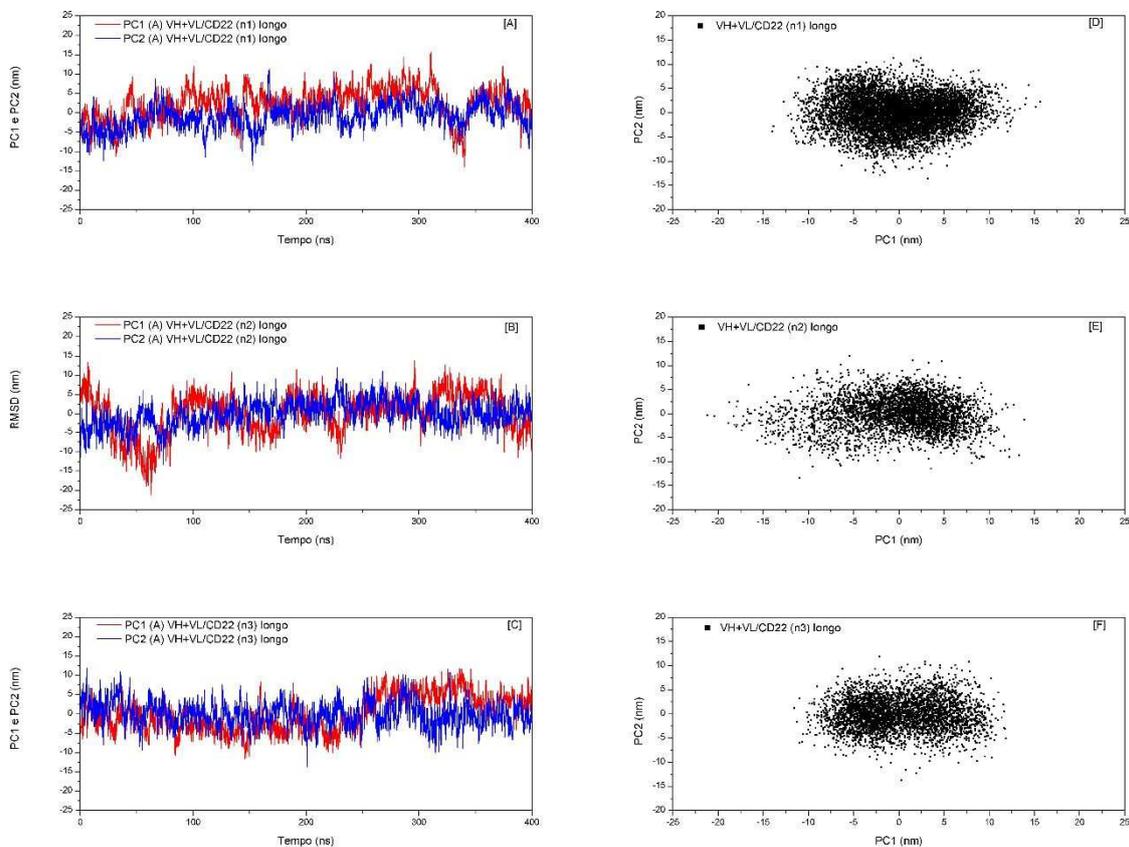
que as curvas de PC1 e PC2 se mantiveram estáveis em ambos os sistemas nas triplicatas. Nota-se nas curvas, em ambas as triplicatas, que as projeções PC1 e PC2 oscilam de ~ 5 a ~ -5 nm. Contudo, nota-se que tanto nas curvas quanto nas projeções nos subespaços comuns, que o scFv de linker longo alcança maiores PC1 e PC2, sendo mais difuso nas projeções, comparando as Figuras 65D-F e 65D-F. As movimentações intrínsecas dos scFv de *linker* curto e longo, mostram sutis diferenças.

Figura 64 - Curvas de PCA de VH+VL/CD22 para os sistemas scFv/CD22-t curto em triplicata de 400 ns de simulação. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas PC1 e PC2 em função do tempo (A, B e C) e Projeções das curvas PC1/PC2 (D, E e F).



Fonte: elaborada pela autora (2024). Nota: CD22 desconsiderando *loops* modelados, glicosilações e o domínio transmembrana. O scFv desconsiderando o *linker*.

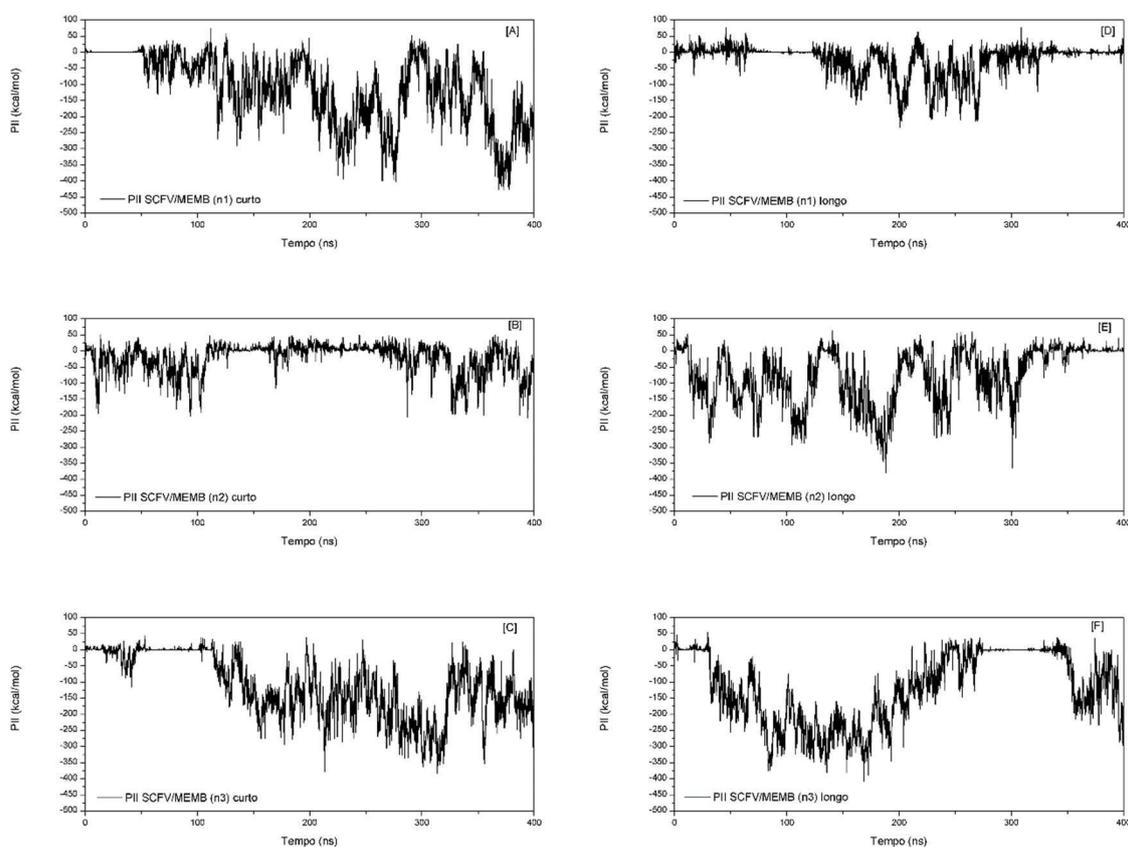
Figura 65 - Curvas de PCA de VH+VL/CD22 para os sistemas scFv/CD22-t longo em triplicata de 400 ns de simulação. A triplicata (n1, n2 e n3) está disposta nas linhas e nas colunas PC1 e PC2 em função do tempo (A, B e C) e Projeções das curvas PC1/PC2 (D, E e F).



Fonte: elaborada pela autora (2024). Nota: CD22 desconsiderando *loops* modelados, glicosilações e o domínio transmembrana. O scFv desconsiderando o *linker*.

Foi investigado se os scFvs interagem com a membrana e a natureza da interação (repulsiva ou atrativa). Foi calculado o PII de longo alcance, num raio de corte de 1,3 nm (RC = 1,3 nm) (Figura 66 A-F). Em ambos os sistemas o PII é atrativo como ratificado pelas médias na Tabela 30. Entretanto, em alguns momentos o PII vai a zero, nos dois sistemas, indicando que em alguns momentos o scFvs e a membrana estão distantes. A interação é majoritariamente atrativa em ambos os sistemas, com intensidades similares. Não há prevalência nessa interação atrativa, entre os scFvs de *linker* curto ou longo.

Figura 66 - PII entre os scFvs e a membrana, dentro de um raio de corte de 1,3 nm, nos sistemas scFv/CD22-t.



Fonte: elaborada pela autora (2024). A), B) e C) triplicata do sistema de *linker* curto. D), E) e F) triplicata do sistema de *linker* longo. Nota: MEMB é a nomenclatura dada a membrana.

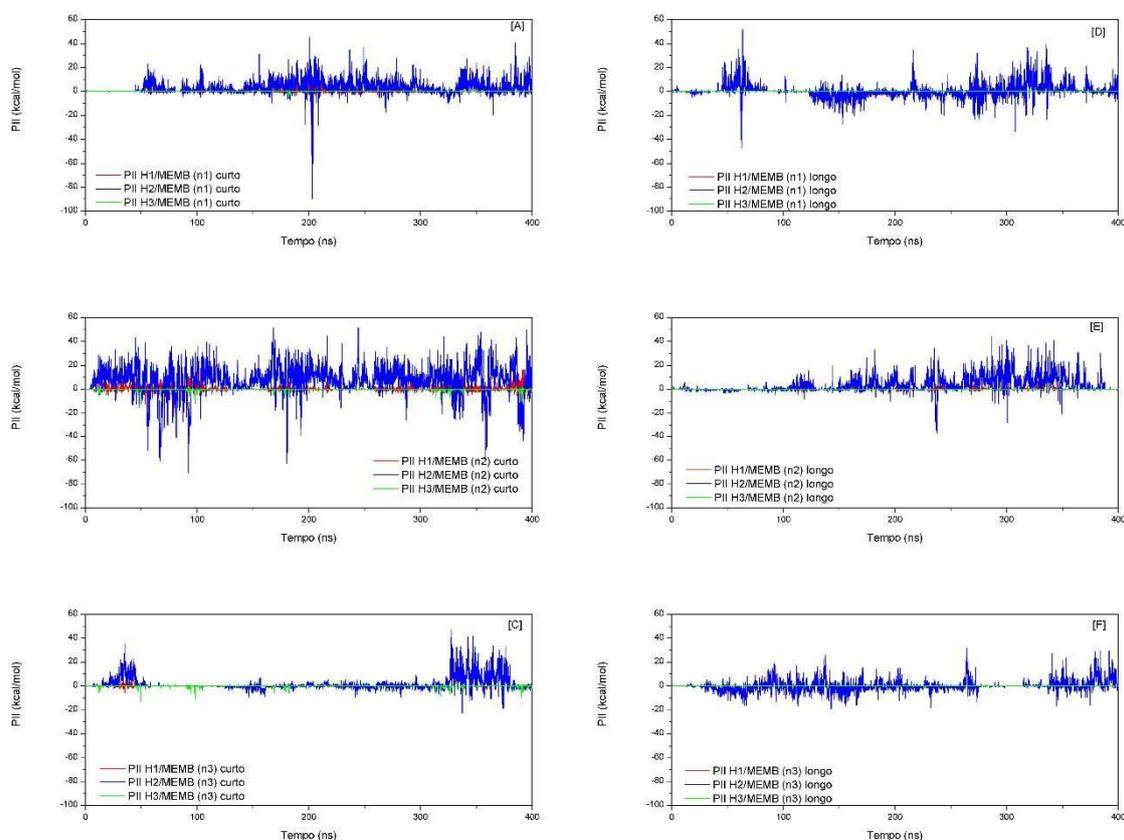
Tabela 30 - Médias e desvios de PII entre os scFvs e a membrana nos complexos scFv/CD22-t de linker curto e longo, dentro de um raio de corte de 1,3 nm, calculadas a partir do $\tau_{eq} = 75$ ns.

	scFv (curto)/CD22-t	scFv (longo)/CD22-t
	PII (kcal/mol)	PII (kcal/mol)
N*	SCFV-MEMB	SCFV-MEMB
n1	-135,7 ± 102,0	-28,8 ± 49,8
n2	-21,2 ± 48,1	-73,7 ± 86,8
n3	-133,3 ± 89,4	-128,3 ± 114,1

Fonte: elaborada pela autora (2024). *N – simulações em triplicata (n1, n2 e n3). Nota: MEMB é a nomenclatura dada a membrana.

Também foi investigado a interação entre as CDRs dos domínios VH (H1, H2 e H3) e VL (L1, L2 e L3) e a membrana. Foi calculado o PII de longo alcance, num raio de corte de 1,3 nm (RC = 1,3 nm) (Figura 67 A-F). As curvas para VL não foram mostradas pois o potencial foi nulo para os dois sistemas. Como observado pela Tabela 31, os valores de PII entre VH e a membrana são baixos e pouco negativo em ambos os sistemas, indicando que há uma fraca interação atrativa entre as CDRs e a membrana. Observa-se ainda que a CDR H2 é a que mais interage com a membrana e, majoritariamente, repulsiva, entretanto, no sistema scFv/CD22-t curto a interação é um pouco mais repulsiva, uma vez que possuem médias e desvios mais altos que no sistema scFv/CD22-t longo (Tabela 31). A natureza dessa interação repulsiva pode ser explicada pelos resíduos de arginina e lisina que fazem parte dessa CDR, conforme exposto no Quadro 3, que podem interagir com fosfolipídeos carregados, positivamente, da membrana celular.

Figura 67 - PII entre as CDRs da cadeia pesada (H1, H2 e H3) e a membrana, dentro de um raio de corte de 1,3 nm, nos sistemas scFv/CD22-t.



Fonte: elaborada pela autora (2024). A), B) e C) triplicata do sistema de *linker* curto. D), E) e F) triplicata do sistema de *linker* longo. Nota: MEMB é a nomenclatura dada a membrana. H1, H2 e H3 são as CDRs do domínio VH.

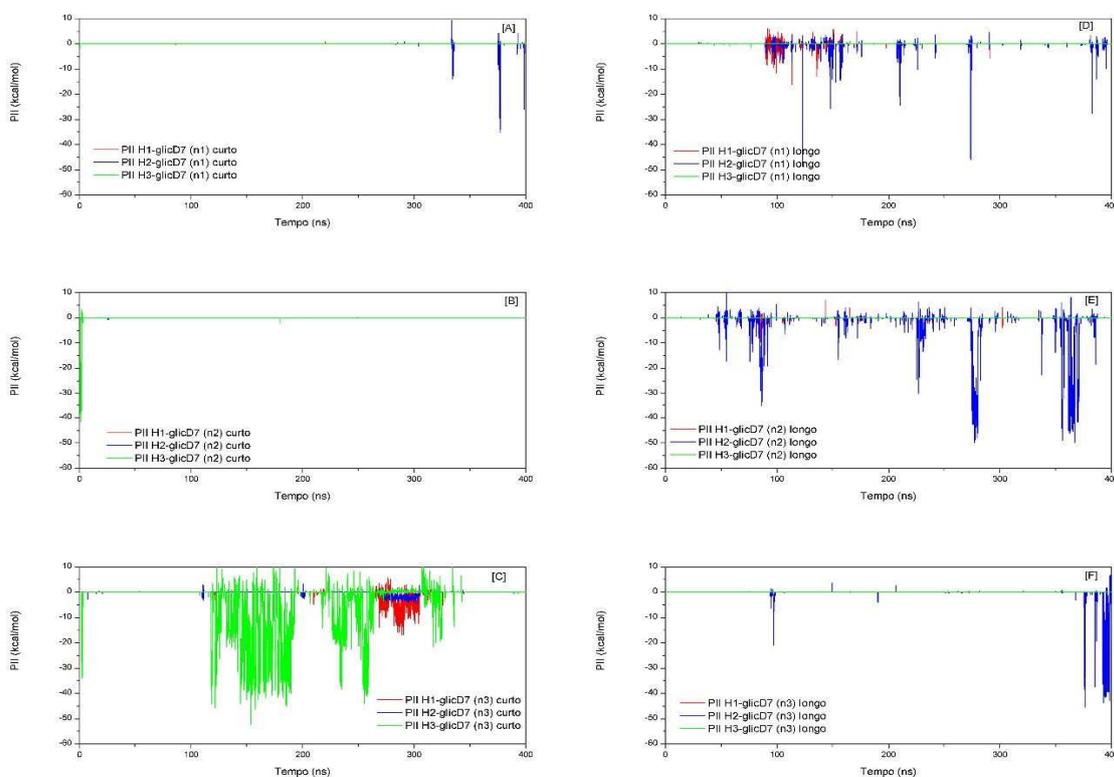
Tabela 31 - Médias e desvios de PII entre CDRs da cadeia pesada (H1, H2 e H3) e a membrana nos sistemas scFv/CD22-t da triplicata (n1), (n2) e (n3), dentro de um raio de corte de 1,3 nm, calculadas a partir do $t_{eq} = 75$ ns.

scFv(curto)/CD22-t				scFv(longo)/CD22-t		
PII (kcal/mol)				PII (kcal/mol)		
CDRs-MEMB				CDRs-MEMB		
N*	H1	H2	H3	H1	H2	H3
n1	0,4 ± 1,1	3,3 ± 5,7	0,1 ± 0,2	-0,1 ± 0,1	0,3 ± 5,6	0
n2	0,6 ± 2,6	7,9 ± 12,7	-0,1 ± 0,7	0,2 ± 0,7	3,9 ± 6,6	0,1 ± 0,1
n3	0,1 ± 0,2	1,4 ± 5,4	0,1 ± 0,1	0	-0,1 ± 4,6	0

Fonte: elaborada pela autora (2024). *N – simulações em triplicata (n1, n2 e n3). Nota: MEMB é a nomenclatura dada a membrana. H1, H2 e H3 são as CDRs do domínio VH.

O CD22 é glicosilado do lado aposto a interface scFv/CD22, no domínio D7. Foi monitorado o PII entre a glicosilação, em D7, e as CDRs. O potencial das CDRs do domínio VL-glicosilação foi nulo (dados não mostrados), enquanto para as CDRs do domínio VH as curvas são mostradas na Figura 68 A-F. Os perfis de PII foram gerados para curto alcance (RC = 0,5 nm) e há indicação de interação atrativa entre a glicosilação e as CDRs dos scFvs (Tabela 32). Entretanto, o PII não resulta em interação extremamente atrativa, e fica abaixo de -50 kcal/mol em ambos os scFvs. Foi possível verificar que a árvore de glicosilação interage atrativamente com a membrana (Tabela 33), de forma que não exerce influência direta na estabilidade da interface scFv/CD22, pois não contribui com o potencial entre eles. Contudo, a interação entre a árvore de glicosilação e a membrana pode estabilizar a posição do CD22, facilitando o alinhamento epítopo-parátopo, estabilizando a interface (Figuras 68A-F e 69A-F).

Figura 68 - PII entre a glicosilação, no domínio D7 de CD22, e as CDRs da cadeia pesada (H1, H2 e H3), dentro de um raio de corte de 0,5 nm, nos sistemas scFv/CD22-t.



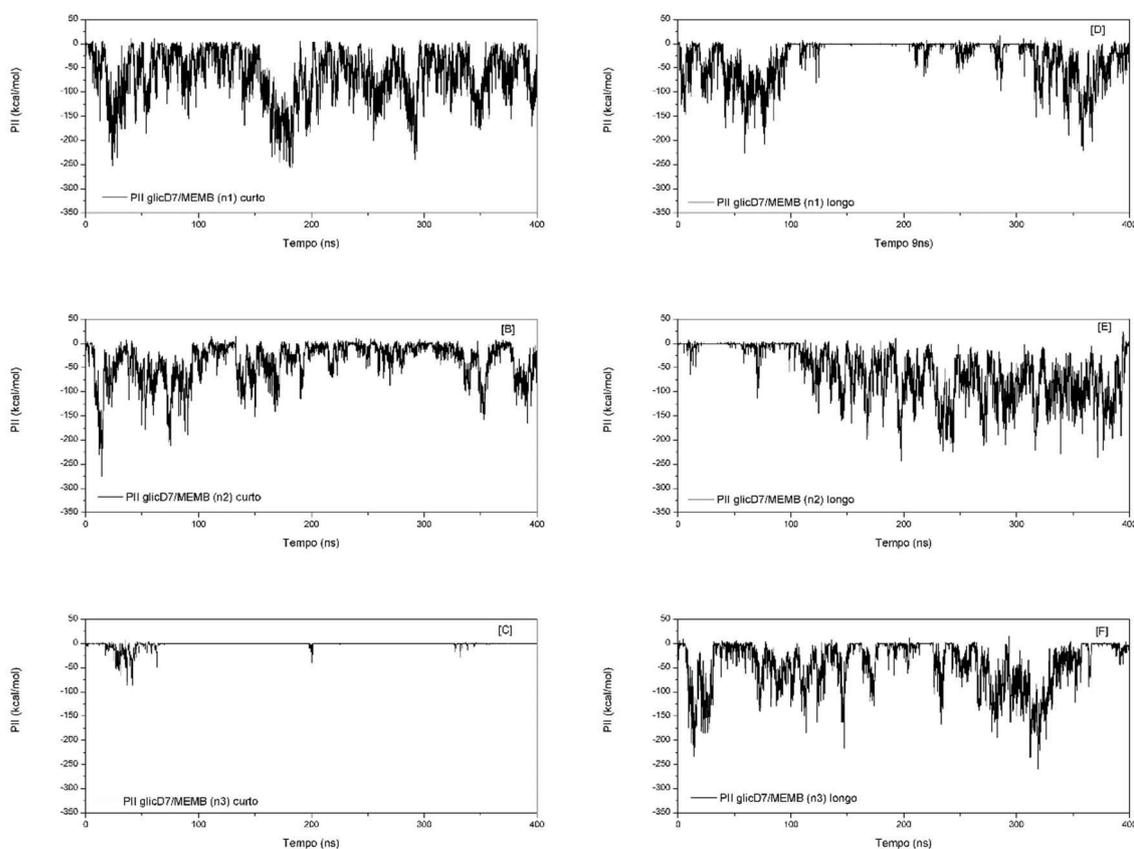
Fonte: elaborada pela autora (2024). A), B) e C) triplicata do sistema de *linker* curto. D), E) e F) triplicata do sistema de *linker* longo. Nota: glicD7 é a glicosilação ligada a D7. H1, H2 e H3 são as CDRs do domínio VH.

Tabela 32 - Médias e desvios de PII entre a glicosilação, no domínio D7 de CD22-t, e as CDRs da cadeia pesada (H1, H2 e H3) nos sistemas scFv/CD22-t da triplicata (n1), (n2) e (n3), em um raio de corte de 0,5 nm, calculadas a partir do $t_{eq}=75$ ns.

scFv(curto)/CD22-t				scFv(longo)/CD22-t		
PII (kcal/mol)				PII (kcal/mol)		
CDRs-glicD7				CDRs-glicD7		
N*	H1	H2	H3	H1	H2	H3
n1	0	$-0,9 \pm 1,4$	0	$-0,1 \pm 0,8$	$-0,3 \pm 0,2$	$-0,1 \pm 0,2$
n2	0	0	$-0,1 \pm 0,1$	$-0,1 \pm 0,4$	$-2,2 \pm 7,9$	0
n3	$-0,5 \pm 2,0$	$-0,2 \pm 0,7$	$-4,9 \pm 10,5$	0	$-0,6 \pm 4,3$	0

Fonte: elaborada pela autora (2024). *N – simulações em triplicata (n1, n2 e n3). Nota: glicD7 é a glicosilação ligada a D7. H1, H2 e H3 são as CDRs do domínio VH.

Figura 69 - PII entre a glicosilação, no domínio D7 de CD22, e a membrana, dentro de um raio de corte de 1,3 nm, nos sistemas scFv/CD22-t.



Fonte: elaborada pela autora (2024). A), B) e C) triplicata do sistema de *linker* curto. D), E) e F) triplicata do sistema de *linker* longo. Nota: glicD7 é a glicosilação ligada a D7. MEMB é a nomenclatura dada a membrana.

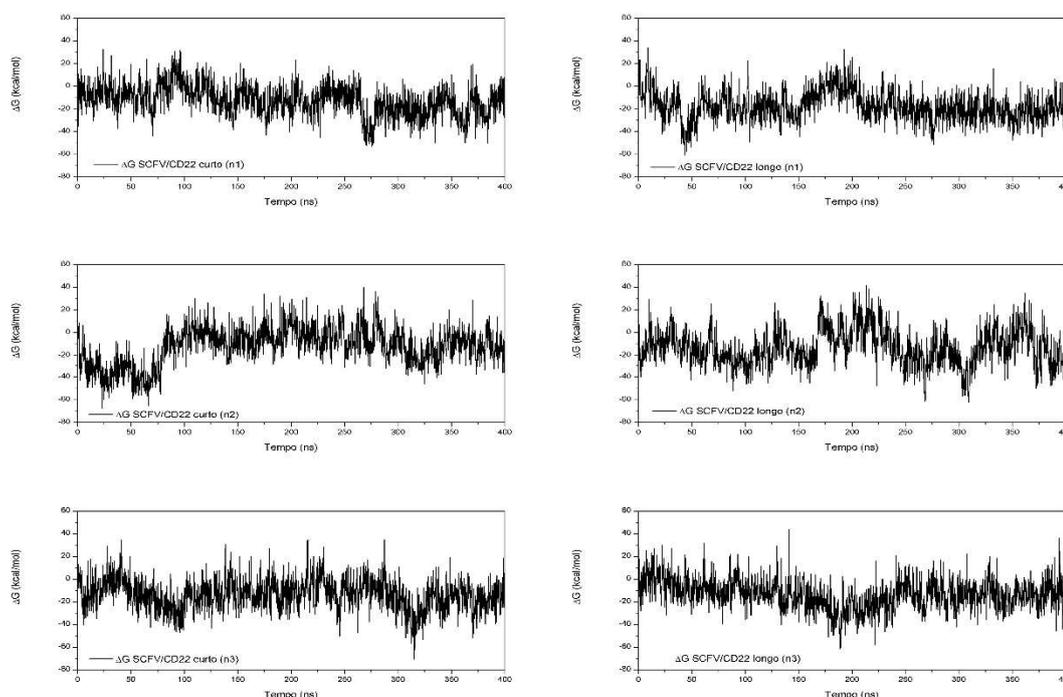
Tabela 33 - Médias e desvios de PII entre a glicosilação N-ligada a D7 e a membrana nos complexos scFv/CD22-t de linker curto e longo, dentro de um raio de corte de 1,3 nm, calculadas a partir do $\tau_{eq} = 75$ ns.

	scFv (curto)/CD22-t PII (kcal/mol)	scFv (longo)/CD22-t PII (kcal/mol)
N*	GlicD7/MEMB	GlicD7/MEMB
n1	-65,4 ± 50,1	-22,5 ± 37,4
n2	-37,5 ± 33,7	-71,6 ± 50,7
n3	-0,2 ± 2,1	-43,8 ± 46,4

Fonte: elaborada pela autora (2024). *N – simulações em triplicata (n1, n2 e n3). Nota: glicD7 é a glicosilação ligada a D7. MEMB é a nomenclatura dada a membrana.

A energia livre de ligação ($\Delta G_{\text{ligação}}$) média foi calculada a partir dos perfis de $\Delta G_{\text{ligação}}$ gerados ao longo da DM, como mostrado na Figura 70A-F. As médias de $\Delta G_{\text{ligação}}$, calculadas a partir de 75 ns de simulação, estão descritas na Tabela 34.

Figura 70 - Curvas de $\Delta G_{\text{ligação}}$ das triplicatas de 400 ns de simulação dos complexos scFv/CD22-t curto e longo.



Fonte: elaborada pela autora (2024). (A), (B) e (C) triplicata do scFv/CD22-t curto. (D), (E) e (F) triplicata do scFv/CD22-t longo. Nota: CD22-t considerando os *loops* modelados, glicosilações e o domínio transmembrana. O scFv considerando o *linker*.

Tabela 34 - Resumo dos ΔG ligação médios dos sistemas scFv/CD22-t de *linker* curto e longo em triplicata n1, n2 e n3.

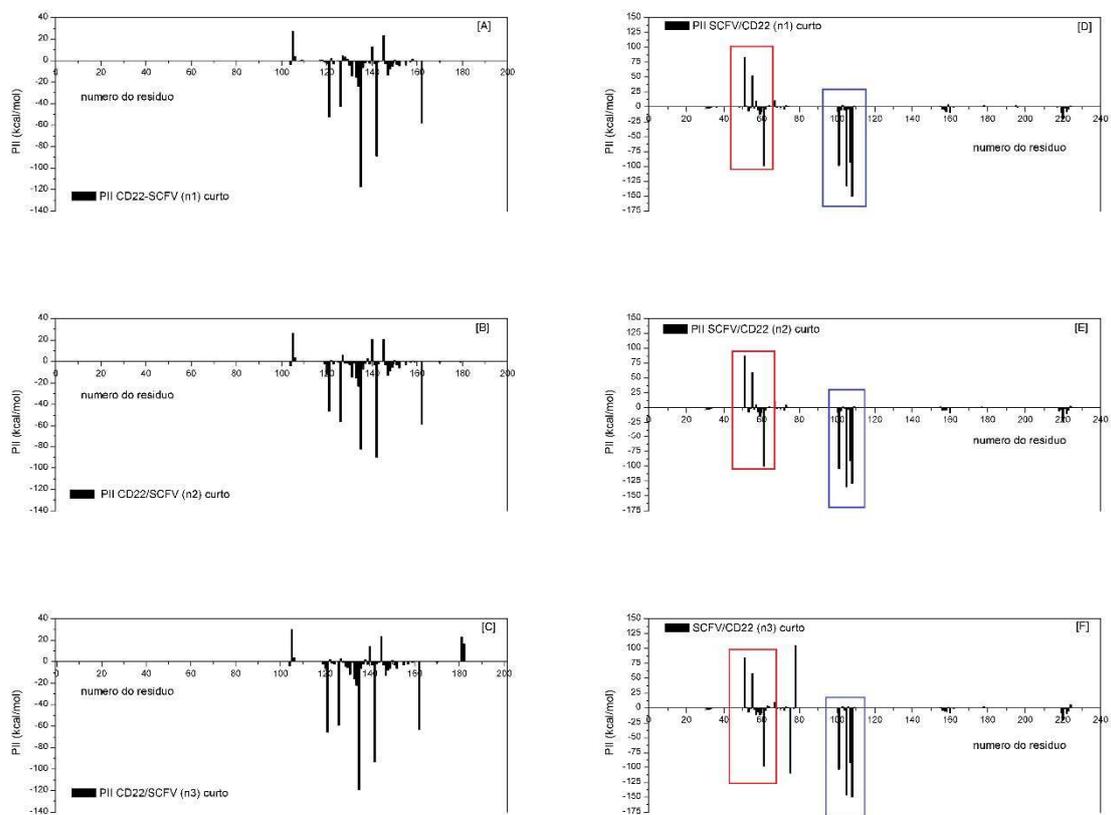
Sistema	Valor de ΔG ligação médio (kcal/mol)			ΔG ligação médio (kcal/mol)
	(n1)	(n2)	(n3)	
curto	-13,1 ± 13,9	-7,1 ± 13,2	-14,8 ± 14	-11,7 ± 3,03*
longo	-17,5 ± 12,6	-13,5 ± 17,1	-15,5 ± 12,6	-15,5 ± 1,33*

Fonte: elaborada pela autora (2024). Nota: N – simulações em triplicata (n1, n2 e n3). *Desvio da média.

Por fim, para determinar a contribuição de cada resíduo do CD22 e dos scFvs curto e longo na interação scFv/CD22-t foi realizada a análise de PII entre os resíduos do scFv e o CD22, dentro de um raio de corte de 0,5 nm, visando um mapeamento dos que estão envolvidos na interface (Figura 73). As Figuras 71A-C e 72A-C apresentam a distribuição de PII médio por resíduo (sem apresentar os desvios), para não atrapalhar a visualização dos dados. Nota-se que o grupo de resíduos do CD22 que interagem atrativamente com o scFvs, são reprodutíveis na triplicata. Destacam-se os resíduos His122 e His136; Asp127; Glu143; e Lys163 no CD22 que interagem atrativamente com o scFv de *linker* curto. Esses mesmos resíduos também interagem atrativamente com o scFv de *linker* longo. Entretanto, a Arg141 se mostrou numa interação repulsiva com o scFv curto, mas atrativa com o scFv de *linker* longo.

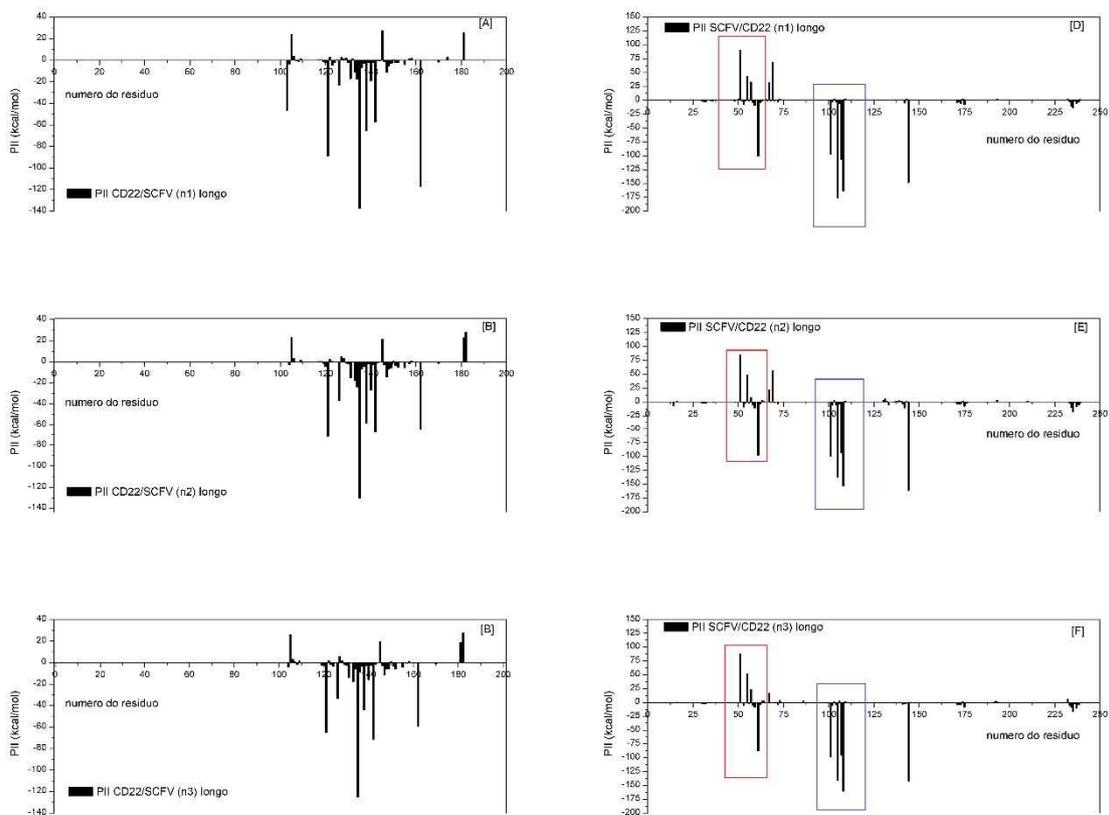
Foi realizado uma análise inversa para mapear os resíduos dos scFvs (*linker* curto e longo) em interação na interface com CD22. Destacam-se os resíduos carregados negativamente (Glu102, Asp106, Glu108 e Asp109) da CDR H3, que mais contribuem para a interação atrativa. Já os resíduos Arg52 e Arg56 da CDR-H2 são os que mais contribuem para a interação repulsiva em ambos tanto para o scFv de *linker* curto quanto longo, sendo os iniciais principais alvos para mutação (Figura 71D-F e Figura 72D-F).

Figura 71 - PII entre resíduos do complexo scFv/CD22-t curto dentro de um raio de corte de 0,5 nm.



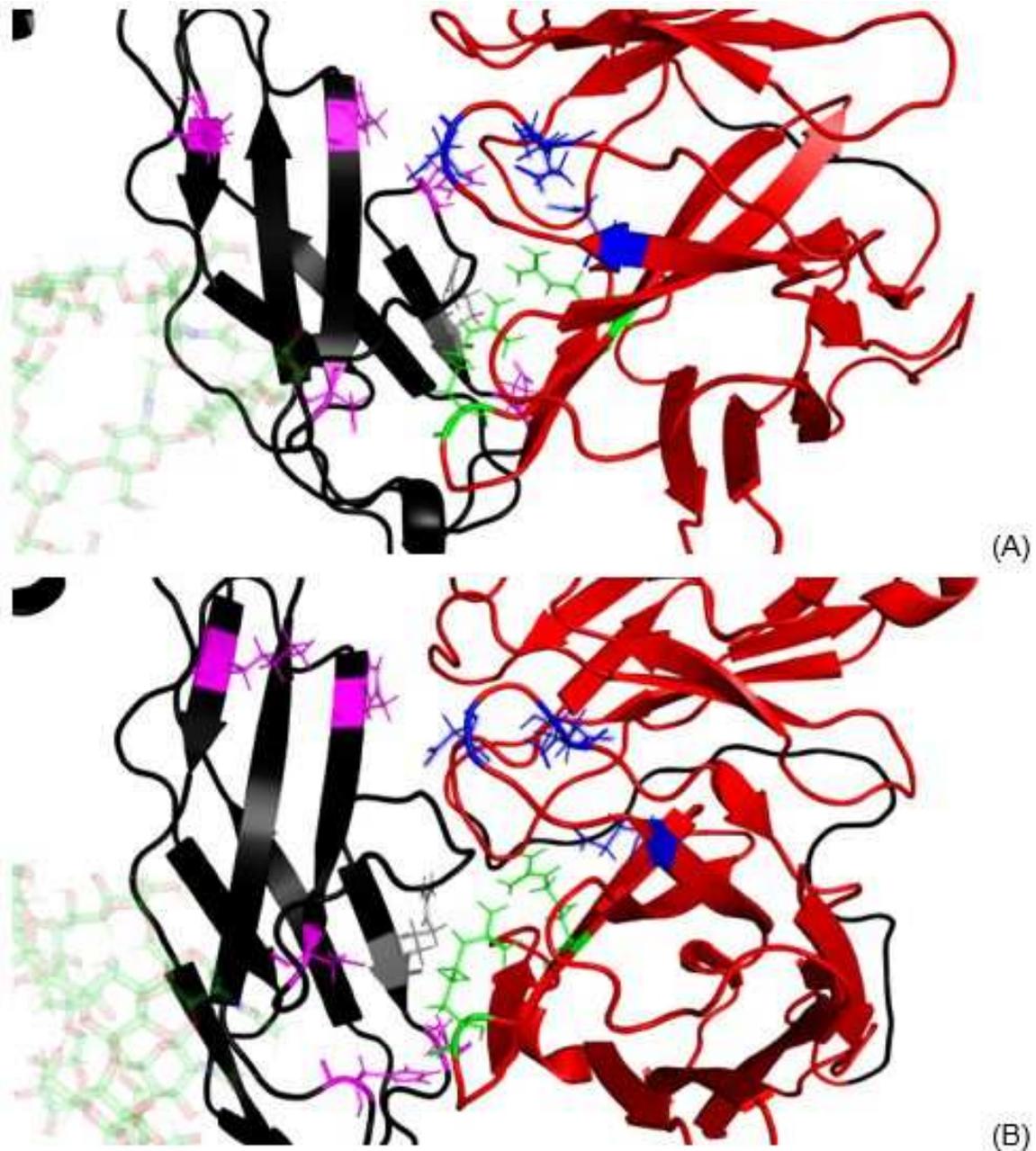
Fonte: elaborada pela autora (2024). (A-C) PII entre resíduos do CD22 e scFv curto. (D-E) PII entre resíduos do scFv curto e o CD22. *N – simulações em triplicata (n1, n2 e n3). Nota: marcado com quadrado vermelho os resíduos pertencentes a CDR-H2 e marcados com quadrado azul os resíduos pertencentes a CDR-H3.

Figura 72 - PII entre resíduos do complexo scFv/CD22-t longo dentro de um raio de corte de 0,5 nm.



Fonte: elaborada pela autora (2024). (A-C) PII entre resíduos do CD22 e scFv longo. (D-E) PII entre resíduos do scFv curto e o CD22. *N – simulações em triplicata (n1, n2 e n3). Nota: marcado com quadrado vermelho os resíduos pertencentes a CDR-H2 e marcados com quadrado azul os resíduos pertencentes a CDR-H3.

Figura 73 - Representação em cartoon da interface de interação scFv-CD22-t curto e longo evidenciando os resíduos (em sticks) que contribuem para interação atrativa mostrado pelas Figuras 71 e 72.



Fonte: elaborada pela autora (2024). (A) sistema scFv/CD22-t curto. (B) sistema scFv/CD22-t longo. Nota: Em preto o domínio D7 de CD22 e o *linker* curto e longo. Em vermelho os scFvs. Em magenta os resíduos His122 e His136; Asp127; Glu143; e Lys163 do CD22. Em cinza o resíduo Arg141 do CD22. Em azul os resíduos Glu102, Asp106, Glu108 e Asp109 da CDR-H3. Em verde os resíduos R52 e R56 da CDR-H2. No fundo da figura, em *sticks*, a árvore de glicosilação ligada ao resíduo N130 do D7 de CD22.

8 DISCUSSÃO

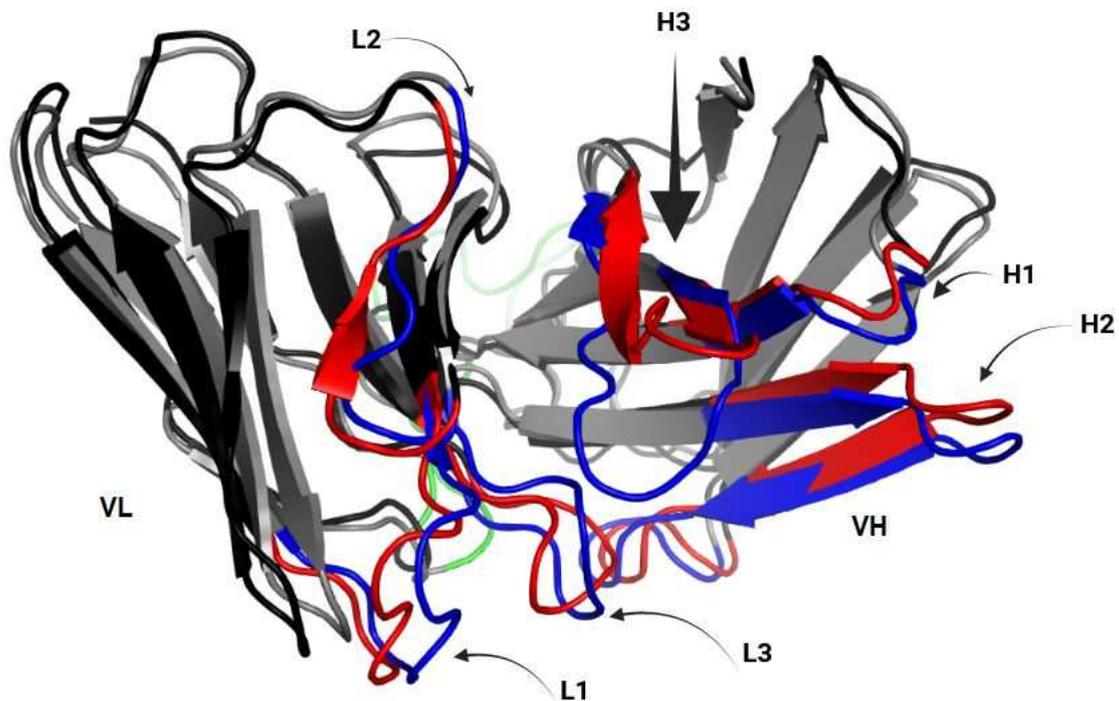
O sistema estudado contempla dois scFvs, com os mesmos domínios, ligados por *linkers* distintos (curto e longo). Os scFv foram simulados numa interface com o CD22, cuja estrutura foi a obtida através de coordenadas atômicas do arquivo PDB 7O52, através da resolução da estrutura. O CD22 teve partes de loops faltante modelados e inserida duas arvores de glicosilações, uma em cada domínio, D6 e D7. Nesse estudo, o CD22 é formado apenas pelos dois domínios. A estrutura do CD22 foi validada e avaliada pelo programa *PROCHECK*, com parâmetros de validação satisfatórios.

Foram simulados vários sistemas como preparação para o sistema completo, um CD22 inserido com o domínio transmembrana num modelo de bicamada lipídica e em interface com scFv. Além de serem preparatórios os sistemas individuais, simulados, possibilitaram extrair informações estruturais que permitem entender como ocorre a estabilização do complexo.

Os dois sistemas com scFv em solução, o de *linker* curto e longo, apresentaram diferenças estruturais sutis. O scFv de *linker* longo se mostrou mais flexível que de *linker* curto (comparar Figuras 25A e 32A), como resultado da longa conexão resíduos entre os domínios. No caso do scFv de *linker* curto, uma sequência de GGGGS, provocou uma alteração na interface, como é observado pela projeção de PC1/PC2 que apresenta duas regiões definidas (Figuras 25A e B). A estrutura inicial é modificada e passa a ser estável depois de um período de ~ 100 ns. A configuração estrutural do scFv com uma sequência de *linker* curto é conhecida por formar *diabody*. Essa modificação estrutural é bem conhecida e estabelecida na literatura, e ocorre com esse tamanho de *linker* (TANG *et al.*, 1996; KNOW *et al.*, 2018). Os perfis de raio de giro são distintos no início da simulação, considerando que a estrutura de partida é a cristalográfica e a mesma para ambos, mudando somente o tamanho do *linker*. O perfil de R_g para scFv de *linker* curto (Figura 24) mostra um decréscimo no início da DM, a partir de $\sim 1,84$ nm, enquanto no longo, a curva mostra um acréscimo, partindo de ($R_g \sim 1,74$ nm) (Figura 31), sendo que convergem para $1,80 \pm 0,01$ nm (curto) e $1,77 \pm 0,01$ (longo). A Figura 74 mostra a representação das estruturas do scFv de *linker* curto e longo e é destacado a posição da CDR H3, que mostra uma mudança conformacional de H3 entre os scFvs. Na Figura 75A, observa-se ainda que há mudança de posição de CDR-H3 em ambos os scFvs, sendo que no scFv longo inicia

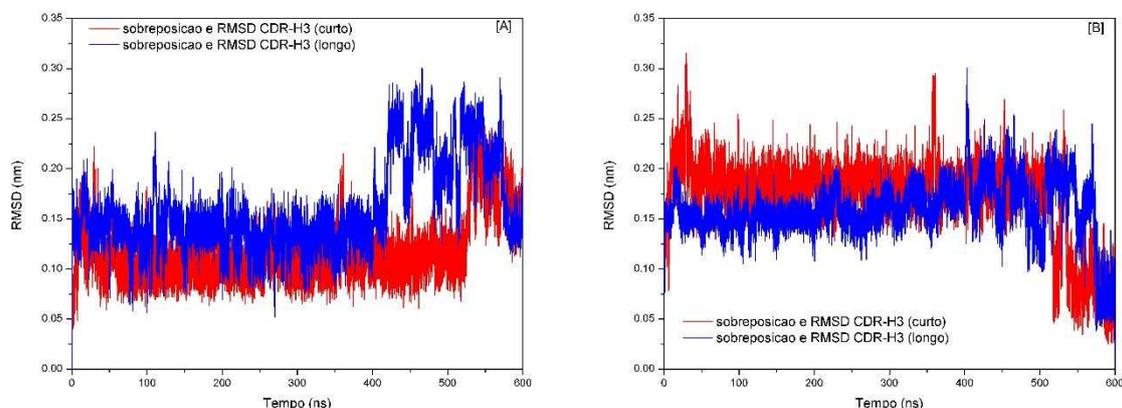
a partir de ~400 ns e no scFv curto inicia a partir de ~550 ns, tendo como referência a configuração inicial de simulação. Já tendo como referência a configuração final de simulação, Figura 75B, observa-se mudança de posição da CDR-H3 de ambos os scFvs a partir de ~500 ns. As Figuras 75A e B podem explicar a mudança de posição relativa entre os domínios VH e VL observadas nas Figuras 22A-B e 29A-B, uma vez que podem estar associadas a mudança de posição da CDR-H3 em ambos os scFvs.

Figura 74 - Representação em cartoon da sobreposição dos scFvs curto e longo evidenciando as diferenças de posicionamento das CDRs, em especial, da CDR-H3.



Fonte: elaborada pela autora (2024). Nota: O scFv de *linker* curto em preto coletado em 390 ns de simulação. O scFv de *linker* longo em cinza coletado em 390 ns de simulação. Em vermelho as CDRs do domínio VH e VL do scFv curto. Em azul as CDRs do domínio VL e VH do scFv longo. Na seta em preto mais grossa é evidenciado a diferença de posicionamento da CDR-H3 em ambos os scFvs.

Figura 75 - Perfil de RMSD dos C α das configurações estruturais da CDR-H3 nos sistemas scFv curto e longo, no decorrer de 600 ns de simulação, comparadas a estrutura inicial e final como referência.



Fonte: elaborada pela autora (2024). (A) sobreposição e RMSD da CDR-H3 nos sistemas scFv curto e longo comparadas a estrutura inicial como referência. (B) sobreposição e RMSD da CDR-H3 nos sistemas scFv curto e longo comparadas a estrutura final como referência.

A simulação do CD22 em solução aquosa revelou que a estrutura apresentou mudanças conformacionais intensas. A comparação entre as curvas de RMSD na Figura 35 A e B possibilitam definir que essas movimentações são decorrentes das duas árvores de glicosilação, presentes nos domínios D6 e D7, e dos *loops* que estavam ausentes na resolução da estrutura cristalográfica (EREÑO-ORBEA *et al.*, 2021). Em relação a estrutura dos domínios D6 e D7, essas são mantidas estáveis estruturalmente (Figura 37), e movimentos relativos entre eles de amplitude de RMSD elevada (Figura 38A e B). Quando inserido em membrana, o CD22 ancorado pelo domínio transmembrana, foi denominado como sistema CD22-t, mostrou que os movimentos das glicosilações e *loops* são mantidos. Da mesma forma, o CD22 também se movimenta em relação ao plano da membrana, como indicado na curva de RMSD (Figura 42), como visto nas Figuras 44A e B.

Foram simulados três sistemas interfaces (scFv/CD22), partindo de estratégias iniciais distintas: docking do scFv ao CD22, uso das estruturas mais representativas dos scFvs e CD22 simulados e sobreposição às estruturas cristalográfica e a partir da estrutura cristalográfica.

O sistema com interface, scFv/CD22, obtida por *docking* considera alguns resíduos descritos na literatura, como *hotspots*. O estudo de Ereno-orbea e colaboradores (2021) relata a importância, principalmente, dos seguintes resíduos da

cadeia pesada e leve³. Apesar de relatos na literatura de definição desses resíduos importantes no anticorpo M971 como eles não constituem o parátipo no reconhecimento da glicoproteína CD22, nem todos foram escolhidos como resíduos chaves para execução do *Docking* Molecular. Portanto, os resíduos de todas as CDRs do scFv foram definidos como resíduos ativos no *Docking*. Já para a definição dos resíduos ativos do CD22⁴, apesar dos resíduos chaves de interação, já descritos por Ereno-orbea e colaboradores (2021), não constituírem o epítipo de interação, esses resíduos foram utilizados para a realização do *Docking* Molecular. Para os sistemas scFv/CD22 de *linker* curto e longo obtidos a partir do *Docking* Molecular foi possível observar, por meio de análise de PCA, Figuras 47A-D que os scFvs apresentaram oscilações na estrutura acima do observado para quando simulados em solução aquosa, partindo da configuração de estrutura cristalográfica (Figuras 25A-B e 32A-B). A movimentações foram distintas entre os scFvs, o scFv de *linker* curto apresentou três regiões mais bem definidas, nas projeções de PC1/PC2, que o longo, para o qual apresentou duas regiões, mais difusas. Foi definido que somente a partir de 350 ns é que o scFv alcança uma região mais estável para os dois casos. A partir desse instante, foi calculado o ΔG ligação, em $-44,4 \pm 8,5$ e $-46,9 \pm 14,0$ kcal/mol para o scFv de *linker* curto e longo. O CD22 apresenta movimentação intensa, incluindo movimentação relativa, entre os domínios D6 e D7, o que pode influenciar na estabilidade da interface.

Na simulação do modelo de interface scFv/CD22, montado a partir da sobreposição das estruturas coletadas, na fase de equilíbrio estrutural, das DMs isoladas do scFv e do CD22, em solução. As estruturas dos scFvs visitaram várias conformações distintas até alcançar, após ~250 ns, fase de equilíbrio estrutural na interface, como é ratificado pelos gráficos de PCA Figura 52 A-D. As mudanças estruturais são mais evidentes para o scFv de *linker* curto que longo, como mostra o RMSD Figura 51 A e C em comparação com Figura 51B e D. Por outro lado, a estrutura possui maior flutuação estrutural, como na frequência de alteração nos perfis de PCA. Os complexos obtiveram valores de ΔG ligação médio bem distintos $-45,4 \pm 9,9$

³ Da cadeia pesada: Asn32, Ala35, Arg52, Tyr54, Arg56, Ser53, Lys58, Trp59, Tyr60, Asn61, Asp62, Tyr63, Lys68, Thr104, Gly105, Asp106, Leu107 e Asp109; e da cadeia leve: Trp159, Tyr161, Ser220, Tyr221, Ser22 e Ile223 (para scFv curto) e Trp174, Tyr176, Ser235, Tyr236, Ser237 e Ile238 (para scFv longo).

⁴ Pro119, Val120, Ser121, His122, Tyr123, Thr124, Asp127, Trp128, Asp129, Gln131, Ser132, Leu133, Pro134, Tyr135, His136, Ser137, Gln138, Arg141, Leu142, Glu143, Pro144, Lys146, Gln148, His149 e Tyr153.

kcal/mol e $-26,9 \pm 7,4$ kcal/mol para o scFv de linker curto e longo, respectivamente. O $\Delta G_{\text{ligação}}$ do scFv de linker longo é maior em função da movimentação intrínseca observada. Contudo, foi observado que o CD22 apresenta movimentação intensa, embora os domínios D6 e D7 tenham estrutura mantidas. Essa movimentação pode prejudicar a manutenção da interface.

Já para os sistemas scFv/CD22 de *linker* curto e longo, usando as coordenadas cristalográficas no início da DM, pode-se verificar, através da análise dos perfis de RMSD, que não há ocorrências de transições no perfil que indicam mudanças proeminentes nas estruturas. Da mesma forma, a análise de PCA (Figura 57A-D) não mostra alterações efetivas na estrutura, contudo as projeções de PC1/PC2 indicam que o scFv de *linker* longo altera a estrutura com maior frequência que o de *linker* curto. O $\Delta G_{\text{ligação}}$ entre scFv/CD22 de linker curto é $-49,3 \pm 11,0$ kcal/mol, mais negativo que o de linker longo $-42,8 \pm 10,7$ kcal/mol. As mudanças sutis na estrutura, observada para scFv de linker *longo* pode ter contribuído para maior $\Delta G_{\text{ligação}}$. A análise do perfil de RMSD do CD22 (Figura 56A-B) mostra movimentação relativa entre D6 e D7 menores que nos outros sistemas. Os perfis sugerem que há um efeito cooperativo na interface, uma vez que o CD22 apresenta maior oscilações nos perfis de RMSD quando ligado ao scFv de linker longo.

Singh e colaboradores (2021) mediram a afinidade dos complexos scFv/CD22 de *linker* curto e longo, com constantes de dissociação (K_d) de 1.67 e 6.05 nM, respectivamente. Nesse caso, o scFv de *linker* curto tem maior afinidade ao CD22 que o scFv de *linker* longo. Utilizando a equação que relaciona a constante de dissociação e $\Delta G_{\text{ligação}}$ ($\Delta G^0 = -k_B T \cdot \ln K$). O $\Delta G_{\text{ligação}}$ é -12,45 e -11,65 kcal/mol para o scFv de *linker* curto e longo, respectivamente. Comparando os $\Delta G_{\text{ligação}}$ experimentais com os obtidos pelos MM/PBSA, descritos na Tabela 35, a partir de simulações de diferentes posições de interação em solução aquosa, pode-se concluir que não são semelhantes.

Tabela 35 - Resumo dos de ΔG ligação médios aferidos por MM/PBSA dos sistemas scFv/CD22 de linker curto e longo na ausência da membrana.

Interface	ΔG ligação médio (kcal/mol)		Tempo (ns)
	curto	longo	
Docking	-44,4 \pm 8,5	-46,9 \pm 14,0	350
Sobreposição	-45,4 \pm 9,9	-26,9 \pm 7,4	250
Coordenadas cristalográficas	-49,3 \pm 11,0	-42,8 \pm 10,7	50

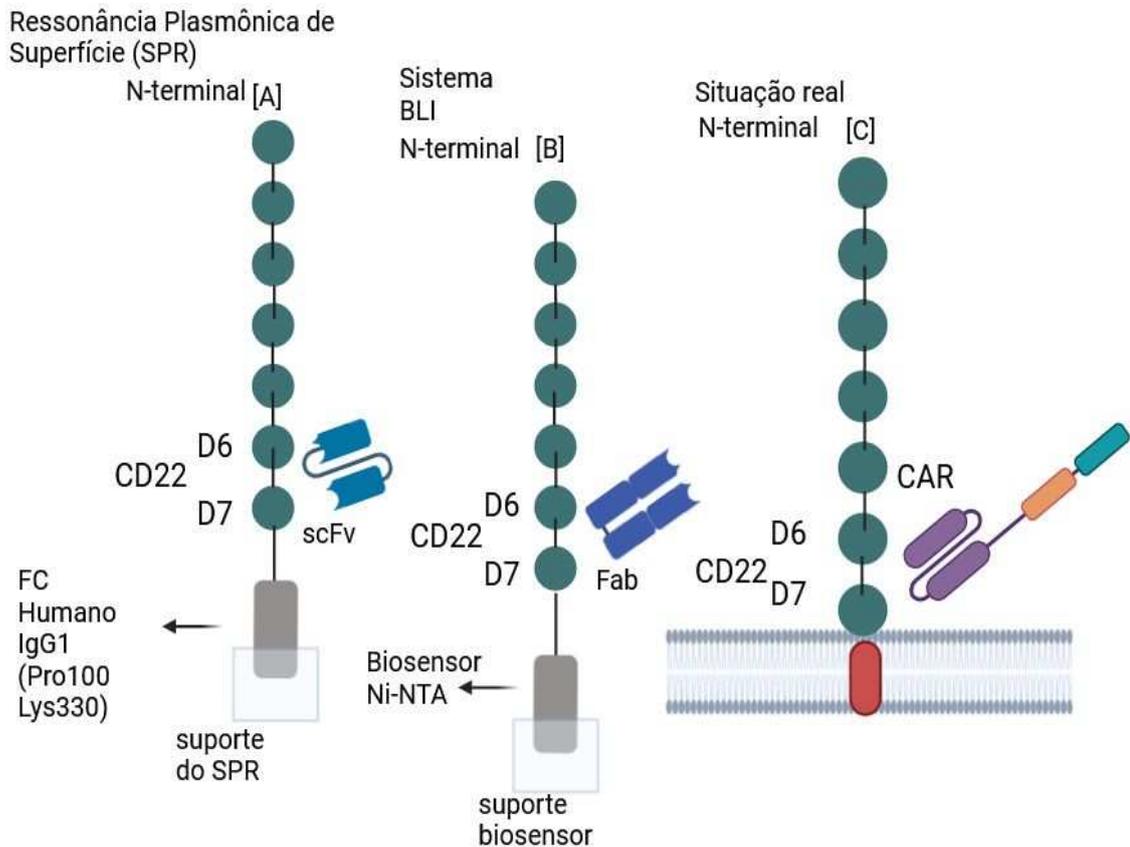
Fonte: elaborada pela autora (2024). Nota: Tempo (ns) é período inicial do cálculo das médias e desvios.

Singh e colaboradores (2021), realizaram a aferição da afinidade entre os scFvs de *linker* curto e longo, usando todo o domínio extracelular de CD22 glicosilado, proveniente da expressão em células de mieloma de camundongo. O tipo de fragmento de anticorpo usado nos estudos é o scFv de *linker* curto e longo derivados do anticorpo M971. A Ressonância Plasmônica de Superfície (SPR) foi usada para medir a afinidade entre o scFv e o CD22. Nesse caso, como visto na Figura 76, um fragmento Fc de um anticorpo humano está covalentemente ligado a região D7 da molécula de CD22 e o CD22-Fc é imobilizado na superfície de um chip CM5.

Ereno-orbea e colaboradores (2021) utilizaram *Biolayer Interferometry* (BLI) para medir a afinidade. Nesse caso o CD22 foi adsorvido no sensor pelo domínio D7 para interação com Fab do M971 (o tipo de fragmento de anticorpo que foi usado), deixando a região D1 mais exposta. O CD22 utilizado foi expresso em células de rim embrionário humano (HEK) e está glicosilada. A medição de afinidade foi realizada a temperatura ambiente. A constante de dissociação obtida foi de 24,8 \pm 1,8 nM, o que equivale a ΔG ligação de -10,37 kcal/mol.

Nota-se no esquema descrito na Figura 76 que as duas abordagens experimentais, para medir as afinidades, são distintas, o que explica a inferência de afinidades distintas. O domínio D7, onde encontra-se a região de ligação (epítipo) do CD22 encontra-se próximo a membrana celular e, potencialmente, há influência dessa região na aproximação e ligação de anticorpos, fragmentos e o CAR. No estudo de Singh e colaboradores (2021), o D7 está próximo a superfície para as médias de SPR e apresentou Kd de 1,67 e 6,05 nM, enquanto o trabalho de Ereno-orbea e colaboradores (2021) o D7 também está próximo a membrana, entretanto o Kd foi de 24,8 \pm 1,8 nM. Nesse sentido, o tamanho dos fragmentos (Fab e scFv) pode influenciar a afinidade (CRIVIANU-GAITA; THOMPSON, 2016).

Figura 76 - Comparação dos experimentos realizados por Singh e colaboradores (2021) e Ereno-orbea e colaboradores (2021) com a situação realista de inserção dos scFvs em uma célula CAR-T anti-CD22.



Fonte: elaborada pela autora (2023). (A) Experimento de Singh e colaboradores (2021). (B) Experimento de Ereno-orbea e colaboradores (2021). (C) Situação real de uma célula CAR-T contendo scFv de M971.

O sistema que considera as coordenadas atômicas determinadas por cristalografia de Raios-X foi usada para prover simulações em triplicata. A última configuração (após 400 ns de DM) do sistema scFv/CD22 em solução aquosa foi usada para unir ao domínio transmembrana, em membrana, seguido para minimização, termalização e simulações em triplicata, mudando as velocidades iniciais do sistema. A opção por esse sistema foi a estabilidade estrutural apresentada nas análises com RMSD e PCA. Os perfis de RMSDs das estruturas dos scFvs e os RMSDs médios mostraram uma estabilização de VH e VL em relação ao obtido para scFv/CD22 em solução aquosa (comparar Tabelas 21 e 27). Contudo, a movimentação relativa de VH/VL permaneceu para o scFvs e no linker curto, mudou de $\sim 0,15$ nm solução aquosa para $\sim 0,16$ nm e para linker longo saiu de $0,16$ nm em solução aquosa para $\sim 0,12$ nm em membrana. Da mesma forma, o CD22 no sistema scFv/CD22, com scFv de linker curto, manteve o RMSD em mesmos patamares que

em solução aquosa ($\sim 0,24$ nm), enquanto no sistema com *linker* longo foi observado uma estabilização no CD22, com RMSD caindo de $\sim 0,29$ nm para $\sim 0,22$ nm.

A estabilidade da interface scFv/CD22 foi monitorada, principalmente, pelas análises de PCA. Em solução aquosa, a associação scFv/CD22 de *linker* curto apresentou-se mais estável que longo. Essa estabilidade foi realçada quando eles estão na presença da bicamada lipídica, embora seja mantida a diferença sutil de maior estabilidade para *linker* curto, como observado comparando as curvas e projeções PC1/PC2, nas Figuras 64 e 65.

Uma vez que os complexos scFv/CD22 estão estáveis no sistema com membrana, foi possível avaliar o ΔG ligação, como descritos na Tabela 36. Nota-se que na triplicata o n2 no sistema scFv/CD22 *linker* curto apresenta um ΔG ligação que destoa dos outros (n1 e n3). Os outros ΔG ligação estão próximos do experimental (Tabelas 36 e 37) e bem distante do observado no sistema com scFv/CD22 aquoso (Tabelas 35), ratificando a importância em simular o sistema com a presença da bicamada lipídica.

Tabela 36 – Resumo dos ΔG ligação médios dos sistemas scFv/CD22-t de *linker* curto e longo em triplicata n1, n2 e n3.

Sistema	Valor de ΔG ligação médio (kcal/mol)			ΔG ligação médio da triplicata
	n1	n2	n3	
curto	-13,1	-7,1	-14,8	$-11,7 \pm 3,03^*$
longo	-17,5	-13,5	-15,5	$-15,5 \pm 1,33^*$

Fonte: elaborada pela autora (2024). Nota: N – simulações em triplicata (n1, n2 e n3). *Desvio da média.

Tabela 37 - Resumo dos Kd médios dos sistemas scFv/CD22-t de *linker* curto e longo em triplicata n1, n2 e n3.

Sistema	Kd médio (nM)			Kd médio da triplicata
	n1	n2	n3	
curto	0,2498	6.246	0,014	2,6536
longo	$1,49 \times 10^{-4}$	0,127	$4,35 \times 10^{-3}$	$4,35 \times 10^{-3}$

Fonte: elaborada pela autora (2024). N – simulações em triplicata (n1, n2 e n3). Nota: Kd obtido por meio da Equação $\Delta G^0 = -k_B T \cdot \ln K$, onde k_B é 0,0019872041 e T é 298,15 K.

ΔG ligação e Kd experimental para scFv curto: -12,45 kcal/mol e 1,67 nM

ΔG ligação e Kd experimental para scFv longo: -11,65 kcal/mol e 6,05 nM

Na análise de PII entre a glicosilação no domínio extracelular D7 do CD22 e as CDRs da cadeia pesada e leve foi possível observar que em ambos os sistemas, todas as CDRs do domínio VH pouco interagem atrativamente e VL não interage. Também foi realizada a análise de PII entre a glicosilação, no domínio extracelular D7 do CD22 e a membrana em ambos os sistemas. Nesta análise, observou-se que em ambos os casos essa árvore de glicosilação interage atrativamente com a membrana, mais do que qualquer CDR em ambos os scFvs. Esses resultados juntos corroboram os dados experimentais apresentados por Ereno-orbea e colaboradores (2021), uma vez que esses pesquisadores sugerem que essa glicosilação não faz parte do epítipo de interação.

Foi feita também a análise de PII entre as CDRs dos domínios VH e VL de cada scFv e a membrana. Nesta análise foi observado que a CDR-H2 é a CDR que mais interage com a membrana e a natureza da interação é majoritariamente repulsiva, sugerindo que os resíduos carregados positivamente dessa CDR podem ser importantes para mediar interações adicionais com grupos fosfolipídeos na superfície da membrana celular (ERENO-ORBEA *et al.*, 2021). Também foi realizado o cálculo de PII, dentro de um raio de corte de 1,3 nm, entre cada scFv e a membrana. Neste caso, foi possível observar que ambos os scFvs interagem atrativamente com a membrana. Todos esses resultados, se relacionam com a hipótese levantada por Ereno-orbea e colaboradores (2021) de que interação com lipídios da membrana celular pode influenciar na atividade do CAR.

Também foi possível determinar quais CDRs mais contribuem para uma interação atrativa com o CD22 e, também, quais resíduos do CD22 mais contribuem para uma interação atrativa aos scFvs. Sendo os resíduos carregados positivamente e negativamente do CD22 que possuem PII mais negativos, portanto, mais atrativos aos scFvs. Em contrapartida os resíduos dos scFvs que mais contribuem com a interação atrativa são carregados negativamente. Esses resíduos carregados interagem por meio de pontes salinas e ligações de hidrogênio, as quais desempenham importante papel nesse reconhecimento e na manutenção da afinidade. Essas interações eletrostáticas fundamentais para a interação entre os scFvs desse estudo e CD22, também são observadas no reconhecimento do CD22 pelo anticorpo M971 (ERENO-ORBEA *et al.*, 2021). Foram observados resíduos de interação semelhantes para ambos os sistemas, de *linker* curto e longo (Figuras 71 e 72). Entretanto, no sistema com *linker* longo foram observados maiores desvios médios.

Por fim, também foi possível determinar quais os prováveis resíduos dos scFvs passíveis de mutação, sendo os resíduos que apresentaram PII positivos mais promissores, como os carregados positivamente (Arg52 e Arg56) principais candidatos para mutação em ambos os scFvs. Todavia, os locais possíveis para mutações precisam ser confirmados por meio de outras análises, como *scanning* de alanina.

9. CONCLUSÕES

Os scFvs se mostraram estruturalmente diferentes o que os diferenciou quanto a estabilidade na interface scFv/CD22 em solução aquosa. Entretanto, quando os complexos scFv/CD22 curto e longo estiveram inseridos em um ambiente de membrana celular apresentaram estabilidade estrutural semelhantes. Também pode-se concluir que a glicosilação ligada ao domínio D7 não tem influência direta na estabilidade da interface scFv/CD22 curto e longo em ambiente de membrana e, portanto, não fazem parte do epítipo de interação. Os valores de Δ Gligação médios dos sistemas scFv/CD22 de *linker* curto e longo se aproximaram mais do Δ Gligação experimental, quando foi considerada a presença da membrana, indicando que a membrana tem influência na interação scFv/CD22.

Em ambos os sistemas a CDR-H2 é a que mais interage com a membrana e a natureza da interação é fracamente repulsiva, sugerindo que os resíduos arginina e lisina dessa podem ser importantes para mediar interações adicionais com grupos fosfolipídeos carregados positivamente na superfície da membrana celular. Foi possível observar ainda que ambos os scFvs interagiram atrativamente com a membrana, corroborando a hipótese de que interação com fosfolipídeos da membrana celular pode influenciar a ação do CAR. Foi possível identificar resíduos nos scFvs que mais contribuem para as interações atrativas e repulsivas, constituindo os resíduos-chaves para proposição de mutações para realçar afinidade e especificidade.

Por fim, o procedimento experimental produziu resultados que permitem afirmar que foi gerado um protocolo, *in silico*, que usa simulações de DM, para analisar a interface e produzir dados confiáveis que corroboram os obtidos *in vitro* para proposição de mutações nos scFvs, possibilitando testá-las *in vivo* com efetividade no formato de CARs e alcançando mais rapidamente o desenvolvimento de nova *CAR-T Cell anti-CD22* para o tratamento da LLA.

10. PERSPECTIVAS FUTURAS

- Confirmar os locais possíveis para mutações por meio de outras análises como *scanning* de alanina;
- Testar novas mutações ainda não descritas na literatura;
- Realizar novos cálculos de ΔG ligação considerando a entropia dos sistemas;
- Realizar simulações de DM considerando o scFv curto como *diabody*;
- Realizar simulações de DM em pH controlado.

REFERÊNCIAS

- ABBAS, ABUL K.; LICHTMAN, ANDREW H.; PILAI, S. **Imunologia Celular e Molecular**. 2015.
- ABDO, L.; ARAGÃO, E.A.; BONAMINO, M. H. Structural Determinants of Chimeric Antigen Receptor Design. **Critical Reviews™ in Immunology**. v. 41, n.1. p 89-104, 2021.
- ABRAHAM, M. J.; VAN DER SPOEL, D.; LINDAHL, E.; HESS, B. **GROMACS User Manual version 2018**. Disponível em: www.gromacs.org, 2015. Acesso em: 23 de out. 2023.
- AHMAD, Z. A.; YEAP, S. K.; ALI, A. M.; HO, W. Y.; ALITHEEN, N. B. M.; HAMID, M. scFv Antibody: Principles and Clinical Application. **Clinical and Developmental Immunology**. v. 2012, p. 1–15, 2012.
- ALABANZA, L.; PEGUES, M.; GELDRES, C.; *et al.* Function of Novel Anti-CD19 Chimeric Antigen Receptors with Human Variable Regions Is Affected by Hinge and Transmembrane Domains. **Molecular Therapy**. v. 25, n. 11. p. 2452–2465, 2017. Elsevier Inc..
- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; *et al.* **Biologia Molecular da Célula**. 6° ed. 2017.
- ALLEN, M. P.; TILDESLEY, D. J. **Computer simulation of liquids**. 1987.
- ARSLAN, M. *et al.* Effect of non-repetitive linker on in vitro and in vivo properties of an anti-VEGF scFv. **Scientific Reports**. v. 12, n. 5449, p. 1-8, 2022.
- BAGDONAITE, I. *et al.* Ectopic expression of a novel CD22 splice-variant regulates survival and proliferation in malignant T cells from cutaneous T cell lymphoma (CTCL) patients. **Oncotarget**. v. 6, n. 16, p. 14374- 14384, 2015.
- BATISTA, M. C. **Desenvolvimento de uma ferramenta para análise de dados de sequenciamento NGS de bibliotecas de scFvs selecionados por Phage Display**. 2023. 82 f. Dissertação (Mestrado em Biotecnologia de Recursos Naturais) – Universidade Federal do Ceará, Fortaleza, CE, 2023. Disponível em: <http://repositorio.ufc.br/handle/riufc/71372>. Acesso em: 31 de out. de 2023.
- BAYBUTT, T. R. *et al.* Advances in Chimeric Antigen Receptor T-Cell Therapies for Solid Tumors. **Clinical Pharmacology & Therapeutics**. v. 105, n. 1. p. 71-78, 2019.
- BEGENT, R. H. J.; VERHAAR, M. J.; CHESTER, K. A.; CASEY, J. L.; GREEN, A. J.; NAPIER, M. P.; HOPE-STONE, L. D.; CUSHEN, N.; KEEP, P. A.; JOHNSON, C. J.; HAWKINS, R. E.; HILSON, A. J.; ROBSON, L. Clinical evidence of efficient tumor targeting based on single-chain Fv antibody selected from a combinatorial library. **Nature Medicine**. v. 2, n. 9, p. 979–984, 1996.
- BERENDSEN, H J C; POSTMA, J. P. M.; GUNSTEREN, W. F. VAN; 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. 1984.

BERENDSEN, H. J. C.; SPOEL, D. VAN DER; DRUNEN, R. VAN. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**. v. 91. n. 1–3, p. 43–56, 1995.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The protein data bank. **Nucleic acids research**. v. 28, n. 1. p. 235–242, 2000.

BONVIN, A. M. *et al.* Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. **Molecular and Cellular Proteomics**. [S.l.], p. 1784-1794, 2010.

BOUGNOUX, P.; SALEM, N.; LYONS, C.; HOFFMAN, T. Alteration in the membrane fatty acid composition of human lymphocytes and cultured transformed cells induced by interferon. **Molecular Immunology**. v. 22, n. 9, p. 1107–1113, 1985.

BRAUTIGAN, D. L. *et al.* Cell cycle oscillation of phosphatase inhibitor-2 in rat fibroblasts coincident with p34cdc2 restriction. **Nature**. v. 6261, n. 344, p. 74-78, 1990.

BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **Journal of Chemical Physics**. v. 126, n. 1. 2007.

CHAN, C. H. *et al.* Internalization of the lymphocytic surface protein CD22 is controlled by a novel membrane proximal cytoplasmic motif. **Cell Biology and Metabolism**. v. 273, n. 43, p. 27809-27815, 1998.

CLARK, E. A.; GILTIAY, N. V. CD22: A Regulator of Innate and Adaptive B Cell Responses and Autoimmunity. **Frontiers in Immunology**. v. 9, n. 2235, p. 1-13, 2018.

CORDEIRO, M. L. S. *et al.* Anticorpos Monoclonais: Implicações terapêuticas no Câncer. **Revista Saúde e Ciência**. n 3, p. 252-262, 2014.

CROCKER, P. R. *et al.* Siglecs and their roles in the immune system. **Nature reviews immunology**. v. 7, p. 255–266, 2007.

DAI, H. *et al.* Bispecific CAR-T cells targeting both CD19 and CD22 for therapy of adults with relapsed or refractory B cell acute lymphoblastic leukemia. **Journal of Hematology & Oncology**. v. 13, n. 30, p. 1-11, 2020.

DALAL, A. *et al.* Third-Generation Human Epidermal Growth Factor Receptor 2 Chimeric Antigen Receptor Expression on Human T Cells Improves with Two-Signal Activation. **Human Gene Therapy**. v. 29, n. 8, p. 845-852, 2018.

DAVIS, I. W.; LEAVER-FAY, A.; CHEN, V. B.; BLOCK, J. N.; KAPRAL, G. J.; WANG, X.; MURRAY, L. W.; ARENDALL 3RD, W. B.; SNOEYINK, J.; RICHARDSON, J. S.;

RICHARDSON, D. C. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. **Nucleic Acids Research**. v. 35, n. SUPPL.2, p. 375–383, 2007.

DOMINGUEZ, C. *et al.* HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. **Journal of the American Chemical Society**. [S.I.], v. 125, p. 1731-1737, 2003.

DONG, G. Q.; FAN, H.; SCHNEIDMAN-DUHOVNY, D.; WEBB, B.; SALI, A. Optimized atomic statistical potentials: Assessment of protein interfaces and loops. **Bioinformatics**. v. 29, n. 24, p. 3158–3166, 2013.

DOODY, G. M. *et al.* A role in B cell activation for CD22 and the protein tyrosine phosphatase SHP. **Science**. v. 269, n. 5221, p. 242-244, 1995.

ENGEL, P. *et al.* Identification of the ligand-binding domains of CD22, a member of the immunoglobulin superfamily that uniquely binds a sialic acid-dependent ligand. **The Journal of Experimental Medicine**. v. 181, n.4, p. 1581–1586, 1995.

ENTERINA, J. R. *et al.* Coordinated changes in glycosylation regulate the germinal center through CD22. **Cell Reports**. v. 38, n. 11, p. 1-43, 2022.

EREÑO-ORBEA, J. *et al.* Structural details of monoclonal antibody m971 recognition of the membrane-proximal domain of CD22. **Journal of Biological Chemistry**. v. 297, p. 1-10, 2021.

FILHO, O. A. S.; ALENCASTRO, R. B. Modelagem de Proteínas por Homologia. **Química Nova**. v. 26, n. 2, p. 253–259, 2003.

FOLTZ, I. N.; KAROW, M.; WASSERMAN, S. M. Evolution and emergence of therapeutic monoclonal antibodies what cardiologists need to know. **Circulation**. v. 127, n. 22, p. 2222–2230, 2013.

FRENKEL, D.; SMIT, B. **Understanding Molecular Simulation**. 2002.

FRIGAULT, M. J. *et al.* Identification of chimeric antigen receptors that mediate constitutive or inducible proliferation of T cells. **Cancer Immunol Res**. v. 3, v. 4, p. 356-367, 2015.

FRY, T. J. *et al.* CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. **Nature medicine**. v. 24, n. 1, p. 20-28, 2018.

FRYDMAN, J. Folding of newly translated proteins in vivo: the role of molecular chaperones. **Annual Review of Biochemistry**. v. 70, p. 603-647, 2001.

GASPARRINI, F. *et al.* Nanoscale organization and dynamics of the siglec CD22 cooperate with the cytoskeleton in restraining BCR signalling. **The EMBO Journal**. v. 35, n. 3, p. 258-280, 2016.

GEORGE, P. *et al.* Third-generation anti-CD19 chimeric antigen receptor T-cells incorporating a TLR2 domain for relapsed or refractory B-cell lymphoma: a phase I clinical trial protocol (ENABLE). **BMJ Open**. v. 10, n.2, p. 1-9, 2020.

GINALSKI, K. Comparative modeling for protein structure prediction. **Current Opinion in Structural Biology**. v. 16, n. 2, p. 172-177, 2006.

GORDON, J. C. *et al.* H++: A server for estimating pKas and adding missing hydrogens to macromolecules. **Nucleic Acids Research**. v. 33, n. 2, p. 368–371. 2005.

GUEST, R. D.; HAWKINS, R. E.; KIRILLOVA, N.; *et al.* The Role of Extracellular Spacer Regions in the Optimal Design of Chimeric Immune Receptors. **Journal of Immunotherapy**. v. 28, n. 3, p. 203–211. 2005.

HAN, S. *et al.* Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking. **Nature Chemical Biology**. v.2, n.1, p. 93-97, 2005.

HASO, W. *et al.* Anti-CD22–chimeric antigen receptors targeting B-cell precursor acute lymphoblastic leukemia. **Blood**. v. 121, n. 7, p. 1165-1174, 2013.

HESS, B. P-LINCS: A parallel linear constraint solver for molecular simulation. **Journal of Chemical Theory and Computation**, v. 4, n. 1. p. 116–122, 2008.

HESS, B.; BEKKER, H.; BERENDSEN, H. J. C.; FRAAIJE, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**. v. 18, n. 12, p. 1463–1472, 1997.

HESS, B.; KUTZNER, C.; SPOEL, D. V. D.; LINDAHL, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. **Journal of Chemical Theory and Computation**. v. 4, n. 3, p. 435– 447, 2008.

HOLLIGER, P.; HUDSON, P. J. Engineered antibody fragments and the rise of single domains. **Nature biotechnology**. v. 23, n. 9, p. 1126–1136, 2005.

HONORATO, R. V., KOUKOS, P. I., JIMÉNEZ-GARCÍA, B. *et al.* Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. **Frontiers in molecular biosciences**. v. 8, 2021.

HUANG, J.; MACKERELL, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. **J Comput Chem**. v. 34, n. 25, p. 2135–2145, 2013.

HUANG, K. **Statistical Mechanics, 2nd Edition**. 2nd ed. John Wiley & Sons, 1987.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: visual molecular dynamics. **J Mol Graph**. v. 14, n. 1, p. 33-38, 1996.

HUSTON, J. S.; LEVINSON, D.; MUDGETT-HUNTER, M.; TAI, M. S.; NOVOTNY, J.; MARGOLIES, M. N.; RIDGE, R. J.; BRUCCOLERI, R. E.; HABER, E.; CREA, R.

Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. **Proc. Natl. Acad. Sci. USA**. v. 85, n. 16, p. 5879 – 5883, 1988.

HUSTON, J. S.; MUDGETT-HUNTER, M.; TAI, M. S.; MCCARTNEY, J.; WARREN, F.; HABER, E.; OPPERMANN, H. Protein engineering of single-chain Fv analogs and fusion proteins. **Methods Enzymol.** v. 203, p. 46 – 52, 1991.

HWANG, W. Y. K. *et al.* Immunogenicity of engineered antibodies. **Methods**. v. 36, n. 1, p. 3-10, 2005.

IMAI, C.; MIHARA, K.; ANDREANSKY, M.; *et al.* Chimeric receptors with 4-1BB signaling capacity provoke potent cytotoxicity against acute lymphoblastic leukemia. **Leukemia**. v. 18, n. 4, p. 676–684, 2004.

INCA. **ABC do Câncer: Abordagens básicas para o controle de Câncer.**

Disponível em:

<https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/livro-abc-6-edicao-2020.pdf> . Acesso em: 20 de out. de 2023.

INCA. **Estimativa 2023: Incidência de Câncer no Brasil.** Disponível em:

<https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2023.pdf> . Acesso em: 20 de out. de 2023.

KELM, S. *et al.* Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. **Current Biology**. v. 4, n. 11, p. 965-972, 1994.

KENNEDY, P. J.; OLIVEIRA, C.; GRANJA, P. L.; *et al.* Monoclonal antibodies: technologies for early discovery and engineering. **Critical Reviews in Biotechnology**. v. 38, n. 3, p. 1– 15, 2018.

KNOW, N. KIM, Y. LEE, J. Structural diversity and flexibility of diabodies. **Methods**. v. 154, p. 136-142, 2018.

KOCHENDERFER, J. N. *et al.* Eradication of B-lineage cells and regression of lymphoma in a patient treated with autologous T cells genetically engineered to recognize Brief report Eradication of B-lineage cells and regression of lymphoma in a patient treated with autologous T cells. **Blood**. v. 116, n. 20. p. 4099–4102, 2010.

KOCHENDERFER, J. N.; FELDMAN, S. A.; ZHAO, Y.; *et al.* Construction and Pre-clinical Evaluation of an Anti-CD19 Chimeric Antigen Receptor. **J Immunother.** v. 32, n. 7, p. 689– 702, 2009.

LASKOWSKI, R. A. *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**. v. 26, n. 2, p. 283–291. 1993.

LEE, J.; CHENG, X.; SWAILS, J. M.; *et al.* 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, 2016.

LERNER, R. A. Combinatorial antibody libraries: new advances, new immunological insights. **Nature Reviews Immunology**. v. 16, n. 8, p. 498–508, 2016.

LI, L. *et al.* AbRSA: A robust tool for antibody numbering. **Protein Science : A Publication of the Protein Society**. v. 28, n.8, p. 1524–1531, 2019.

LIMA, A. J. F. **Estudo da influência do epítipo c-myc na formação da interface entre o scFv e o CD19 para aplicação em CAR**. 2022. 99 f. Dissertação (Mestrado em Biotecnologia de Recursos Naturais) – Universidade Federal do Ceará, Fortaleza, CE, 2022. Disponível em: <http://repositorio.ufc.br/handle/riufc/65870> . Acesso em: 31 de out. de 2023.

LIU, T.; TANG, G. W.; CAPRIOTTI, E. Comparative modeling: the state of the art and protein drug target structure prediction. **Combinatorial Chemistry & High Throughput Screening**. v.14, n. 6, p. 532 – 547, 2011.

LYNN, R. C. *et al.* c-Jun overexpression in CAR T cells induces exhaustion resistance. **Nature**. v. 576, n. 7786, p. 293-300, 2019.

MACEDO, M. J. P. B. **Estratégias e inovações aplicadas ao desenvolvimento de anticorpos monoclonais**. 2018. 48 f. TCC (Graduação) - Curso de Farmácia, Faculdade de Ciências Farmacêuticas da Universidade de São Paulo, São Paulo, 2018. Disponível em: <https://bdta.abcd.usp.br/directbitstream/3f02fedb-97c8-4325-ad61-05018ad79352/2954353.pdf> . Acesso em: 31 de out. de 2023.

MARTÍ-RENO, M. A. *et al.* Comparative protein structure modeling of genes and genomes. **Annu Rev Biophys Biomol Struct**. v. 29, p. 291-325, 2000.

MAUS, M. V.; GRUPP, S. A.; PORTER, D. L.; JUNE, C. H. Antibody-modified T cells: CARs take the front seat for hematologic malignancies. **Blood**. v. 123, n. 17, p. 2625 – 2635, 2014.

MENG, X-Y. *et al.* Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. **Current Computer Aided-Drug Design**. v. 7, n. 2, p. 146–157, 2012.

MEYER, S. J. *et al.* B Cell Siglecs—News on Signaling and Its Interplay With Ligand Binding. **Frontiers in Immunology**. v. 9, n. 2820, p. 1-14, 2018.

MINISTÉRIO DA SAÚDE. **Plano Nacional de Saúde 2020-2023**. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/plano_nacional_saude_2020_2023.pdf . Acesso em: 24 de set. de 2023.

MOHAMMADI, M.; NEJATOLLAHI, F. 3D structural modeling of neutralizing SCFV against glycoprotein-D of HSV-1 and evaluation of antigen-antibody interactions by bioinformatic methods. **International Journal of Pharma and Bio Sciences**. v. 5, n. 4, p. 835–847, 2014.

MUÑOZ-LÓPEZ, P. *et al.* Single-chain fragment variable: recent progress in cancer diagnosis and therapy. **Cancers**. v. 14, n. 17, p. 1-26, 2022.

MUYLDERMANS, S.; SMIDER, V. V. Distinct antibody species: Structural differences creating therapeutic opportunities. **Current Opinion in Immunology**. v. 40, p. 7–13, 2016.

NELSON, A. L. *et al.* Development trends for human monoclonal antibody therapeutics. **Nature reviews drug discovery**. v. 9, n.10, p. 767-774, 2010.

NELSON, A. L.; REICHERT, J. M. Development trends for therapeutic antibody fragments. **Nature Biotechnology**. v. 27, n. 4, p. 331–337, 2009.

NELSON, D. L.; COX, M. M. **Princípios de Bioquímica de Lehninger**. 7 o ed. Brasil: Artmed, 2018.

OLIVEIRA, N. F. F. **Simulação de dinâmica molecular de um modelo de CAR em interação com o CD19, marcador de células cancerosas**. 2020. 128 f. Dissertação (Mestrado em Biotecnologia de Recursos Naturais) – Universidade Federal do Ceará, Fortaleza, CE, 2020. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/50807> . Acesso em: 31 de out. de 2023.

PARK, T. S.; ROSENBERG, S. A.; MORGAN, R. A. Treating Cancer with Genetically Engineered T Cells. **Trends Biotechnol.** v. 29, n. 11. p. 550–557, 2011.

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, 1981.

PONTES, L. Q. **Construção, por desenho racional, e caracterização de variantes de afinidade da região variável do anticorpo monoclonal Rituximabe**. 2019. 76 f. Dissertação (Mestrado em Biotecnologia de Recursos Naturais) – Universidade Federal do Ceará, Fortaleza, CE, 2019. Disponível em: <https://repositorio.ufc.br/handle/riufc/40330>. Acesso em: 31 de out. de 2023.

RETH, M. Antigen receptor tail clue. **Nature**. v. 338, n. 6214, p. 383-384, 1989.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: A unified platform for automated protein structure and function prediction. **Nature Protocols**. v. 5, n. 4, p. 725–738, 2010.

RUELLA, M. *et al.* Clinical Efficacy of Anti-CD22 Chimeric Antigen Receptor T Cells for B-Cell Acute Lymphoblastic Leukemia Is Correlated with the Length of the Scfv Linker and Can be Predicted Using Xenograft Models. **Blood**. v. 130, p. 807-807, 2017.

RYCKAERT, J. P.; CICCOTTI, G.; BERENDSEN, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. **Journal of Computational Physics**. v. 23, n. 3, p. 327–341, 1977.

SAEED, A. F. U. H.; WANG, R.; LING, S.; WANG, S. Antibody engineering for

pursuing a healthier future. **Frontiers in Microbiology**. v. 8, n. MAR, p. 1–28, 2017.

SAFDARI, Y. *et al.* Antibody humanization methods - a review and update. **Biotechnology and Genetic Engineering Reviews**. v. 29, n. 2, p. 175-86, 2013.

ŠALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. **Journal of Molecular Biology**. v. 234, p. 779–815, 1993.

SALOMON, B.; BLUESTONE, J. A. Complexities of CD28/B7: CTLA-4 costimulatory pathways in autoimmunity and transplantation. **Annual Review of Immunology**. v. 19, n. 1. p. 225–252, 2001.

SALTER, A. L.; MARGOT, J. P.; STANLEY, R. R. Chimeric antigen receptor– modified T cells: CD19 and the road beyond. **Blood**. v. 131, p. 2621-2629, 2018.

SCHWARTZ-ALBIEZ, R. *et al.* CD22 antigen: biosynthesis, glycosylation and surface expression of a B lymphocyte protein involved in B cell activation and adhesion. **International Immunology**. v. 3, n. 7, p. 623-633, 1991.

SERMER, D.; BRENTJENS, R. CAR T-cell therapy: Full speed ahead. **Hematol Oncol**. v.1, n.37, p.95-100, 2019.

SHEN, M.-Y.; SALI, A. Statistical potential for assessment and prediction of protein structures. **Protein Science**. v. 11, p. 2507–2524, 2006.

SHEN, Z.; YAN, H.; ZHANG, Y.; MERNAUGH, R. L.; ZENG, X. Engineering peptide linkers for scFv immunosensors. **Anal Chem**. v. 80, n. 6, p. 1910 -1917, 2009.

SHORE, N.D. **Avanços no entendimento da imunoterapia do câncer**. BJU Int. 2015.

SINGH, N. *et al.* Single Chain Variable Fragment Linker Length Regulates CAR Biology and T Cell Efficacy. **Blood**. v. 134, p. 1-5, 2019.

SINGH, N.; FREY, N. V.; RUELLA, M. Antigen-independent activation enhances the efficacy of 4-1BB-costimulated CD22 CAR T cells. **Nature Medicine**. v.1, n.27, p. 842–850, 2021.

SINGH, S. *et al.* Monoclonal antibodies: a review. **Current Clinical Pharmacology**. v. 13, n. 2, p. 85-99, 2018.

SOMMERMEYER, D.; HILL, T.; SHAMAH, S. M.; SALTER, A. I.; CHEN, Y.; MOHLER, K. M.; RIDDELL, S. R. Fully human CD19-specific chimeric antigen receptors for T-cell therapy. **Leukemia**. v. 31, n. 10, p. 2191-2199, 2017.

SPIRO, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. **Glycobiology**. v. 12, n.4, p. 43-53, 2002.

SPOEL, D. V. D.; LINDAHL, E.; HESS, B.; GROENHOF, G.; MARK, A. E.; BERENDSEN, H. J. C. GROMACS: Fast, flexible, and free. **Journal of**

Computational Chemistry. v. 26, n. 16, p. 1701–1718, 2005.

STAMENKOVIC, I. *et al.* The B lymphocyte adhesion molecule CD22 interacts with leukocyte common antigen CD45RO on T cells and alpha 2-6 sialyltransferase, CD75, on B cells. **Cell.** v. 66, n. 6, p. 1133-1144, 1991.

STOWELL, S. R. *et al.* Protein glycosylation in cancer. **Annu Rev Pathol.** v. 10, p. 473–510, 2015.

TAN, Y. *et al.* A novel full-human CD22-CAR T cell therapy with potent activity against CD22low B-ALL. **Blood Cancer Journal.** v. 11, n. 4, p. 71-77, 2021.

TANG, X. *et al.* Third-generation CD28/4-1BB chimeric antigen receptor T cells for chemotherapy relapsed or refractory acute lymphoblastic leukaemia: a non-randomised, open-label phase I trial protocol. **BMJ Open.** v. 6, n. 12, p. 1-8, 2016.

TANG, Y. *et al.* Selection of Linkers for a Catalytic Single-chain Antibody Using Phage Display Technology. **The Journal of Biological Chemistry.** v. 271, n. 26, p. 15682–15686, 1996.

TASIAN, S. K.; GARDNER, R. A. CD19-redirected chimeric antigen receptor-modified T cells: A promising immunotherapy for children and adults with B-cell acute lymphoblastic leukemia (ALL). **Therapeutic Advances in Hematology.** v. 6, n. 5, p. 228–241. 2015.

TERSTAPPEN, L. W. M. M. *et al.* Identification and Characterization of Plasma Cells in Normal Human Bone Marrow by High-Resolution Flow Cytometry. **Blood.** v. 76, n. 9, p. 1739-1747, 1990.

TOBA, K. *et al.* Difference in CD22 molecules in human B cells and basophils. **Experimental Hematology.** v. 30, n. 3, p. 205-211, 2002.

TOMULEASA, C.; FUJI, S.; BERCE, C.; *et al.* Chimeric antigen receptor T-cells for the treatment of B-cell acute lymphoblastic leukemia. **Frontiers in Immunology.** v. 9, n. FEB, p. 1–14, 2018.

UL-HAQ, Z.; MADURA, J. D. **Frontiers in computational chemistry.** Shariqah: Bentham Science Publishers, 2015. v.1, 362 p.

VALDÉS-TRESANCO, M. S. *et al.* gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. **J. Chem. Theory Comput.** v. 17, n. 10, p. 6281–6291, 2021.

VAN GUNSTEREN, W. F.; DOLENC, J.; MARK, A. E. Molecular simulation as an aid to experimentalists. **Current Opinion in Structural Biology.** v. 18, n. 2, p. 149-153, 2008.

VÉLY, F.; VIVIER, E. Conservation of structural features reveals the existence of a large family of inhibitory cell surface receptors and noninhibitory/activatory counterparts. **Journal of immunology.** v. 159, n. 5, p. 2075-2077, 1997.

VERLET, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. **Physical Review**. v. 159, n. 1, p. 98-103, 1967.

WALKER, J. A.; SMITH, K. G. C. CD22: an inhibitory enigma. **Immunology**. v. 123, n. 3, p. 314-325, 2008.

WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. **Current Protocols in Bioinformatics**. 2016.

WEINER, G. J. Building better monoclonal antibody-based therapeutics. **Nature Reviews Cancer**. v. 15, n. 6, p. 361-370, 2015.

WHITLOW, M.; BELL, B. A.; FENG, S. L.; FILPULA, D.; HARDMAN, K. D.; HUBERT, S. L.; ROLLENCE, M. L.; WOOD, J. F.; SCHOTT, M. E.; MILENIC, D. E.; et al. An improved linker for single-chain Fv with reduced aggregation and enhanced proteolytic stability. **Protein Eng**. v. 6, n. 8, p. 989 – 895, 1993.

WILSON, G. L. *et al.* cDNA cloning of the B cell membrane protein CD22: a mediator of B-B cell interactions. **The Journal of Experimental Medicine**. v. 173, n. 1, p. 137-146, 1991.

XIAO, X. *et al.* Identification and characterization of fully human anti-CD22 monoclonal antibodies. **mAbs**. v.1, n.3, p. 297-303, 2009.

YANG, J.; YAN, R.; ROY, A.; XU, D.; POISSON, J.; ZHANG, Y. The I-TASSER Suite: protein structure and function prediction. **Nature Methods**. v. 12, n. 1. p. 1–5, 2015.

YOKOTA, T.; MILENIC, D. E.; WHITLOW, M.; SCHLOM, J. Rapid tumor penetration of a single-chain Fv and comparison with other immunoglobulin forms. **Cancer Res**. v. 52, p. 3402–3408, 1992.

ZHANG, Y. *et al.* A novel and efficient CD22 CAR-T therapy induced a robust antitumor effect in relapsed/refractory leukemia patients when combined with CD19 CAR-T treatment as a sequential therapy. **Experimental Hematology & Oncology**. v. 11, n. 15, p. 1-18, 2022.

ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC Bioinformatics**. v. 9, n. 40. p. 1–8, 2008.

APÊNDICE A – MM/PBSA PARA SISTEMAS NA AUSÊNCIA DA MEMBRANA

Sample input file for PB calculation

This input file is meant to show only that `gmx_MMPBSA` works. Although, we tried to use the input files as recommended in the Amber manual, some parameters have been changed to perform more expensive calculations in a reasonable amount of time. Feel free to change the parameters according to what is better for your system.

```
&general
sys_name="Prot-Lig-CHARMM",
startframe=1,
endframe=4000,
# In gmx_MMPBSA v1.5.0 we have added a new PB radii set named charmm_radii.
# This radii set should be used only with systems prepared with CHARMM force fields.
# Uncomment the line below to use charmm_radii set
#PBRadii=7,
/
&pb
# radiopt=0 is recommended which means using radii from the prmtop file for both the PB calculation and for the NP
# calculation
istrng=0.05, fillratio=4.0, radiopt=0
/
```

Fonte: elaborada pela autora (2024).

APÊNDICE B – MM/PBSA PARA SISTEMAS NA PRESENÇA DA MEMBRANA

Sample input file for MMPBSA with membrane proteins

This input file is meant to show only that gmx_MMPBSA works. Although, we tried to use the input files as recommended in the Amber manual, some parameters have been changed to perform more expensive calculations in a reasonable amount of time. Feel free to change the parameters according to what is better for your system.

```
&general
sys_name="Prot-Memb-CHARMM",
startframe=1,
endframe=2000,
# In gmx_MMPBSA v1.5.0 we have added a new PB radii set named charmm_radii.
# This radii set should be used only with systems prepared with CHARMM force fields.
# Uncomment the line below to use charmm_radii set
#PBRadii=7,
/
&pb
memopt=1, emem=7.0, indi=1.0,
mctrdz=80, mthick=50, poretype=1,
radiopt=0, istrng=0.150, fillratio=1.25, inp=2,
sasopt=0, solvopt=2, ipb=1, bcopt=10, nfocus=1, linit=1000,
eneopt=1, cutfd=7.0, cutnb=99.0,
maxarcdot=15000,
npbverb=1,
/
```

Fonte: elaborada pela autora (2024).

ANEXO A - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv CURTO ANTES DA MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	53.03	3 rd percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	3	1.49% Goal: <0.3%
	Favored rotamers	194	96.04% Goal: >98%
	Ramachandran outliers	3	1.28% Goal: <0.05%
	Ramachandran favored	223	95.30% Goal: >98%
	Rama distribution Z-score	-1.21 ± 0.51	Goal: abs(Z score) < 2
	MolProbity score ¹	2.66	39 th percentile* (N=27675, 0Å - 99Å)
	CB deviations >0.25Å	2	0.93% Goal: 0
	Bad bonds:	6 / 1859	0.32% Goal: 0%
	Bad angles:	30 / 2531	1.19% Goal: <0.1%
Peptide Omegas	Cis Prolines:	2 / 10	20.00% Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	4	1.7% Goal: <1.0%
	CA Geometry outliers	3	1.29% Goal: <0.5%
Additional validations	Chiral volume outliers	0/278	
	Waters with clashes	0/0	0.00% See UnDowser table for details

Fonte: MolProbity (2024).

ANEXO B - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv CURTO APÓS A MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	3.71	96 th percentile* (N=1784, all resolutions)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	1.58%	Goal: <0.3%
	Favored rotamers	174	91.58%	Goal: >98%
	Ramachandran outliers	2	0.85%	Goal: <0.05%
	Ramachandran favored	217	92.74%	Goal: >98%
	Rama distribution Z-score	-1.70 ± 0.50	Goal: abs(Z score) < 2	
	MolProbity score [^]	1.77	87 th percentile* (N=27675, 0Å - 99Å)	
	Cβ deviations >0.25Å	3	1.39%	Goal: 0
	Bad bonds:	2 / 1859	0.11%	Goal: 0%
Bad angles:	10 / 2531	0.40%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	2 / 10	20.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	7	3.0%	Goal: <1.0%
	CA Geometry outliers	2	0.86%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/278		
	Waters with clashes	0/0	0.00%	See UnDowser table for details

Fonte: MolProbity (2024).

ANEXO C - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv LONGO ANTES DA MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	47	5 th percentile* (N=1784, all resolutions)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	1.46%	Goal: <0.3%
	Favored rotamers	197	96.10%	Goal: >98%
	Ramachandran outliers	2	0.80%	Goal: <0.05%
	Ramachandran favored	238	95.58%	Goal: >98%
	Rama distribution Z-score	-1.21 ± 0.47		Goal: abs(Z score) < 2
	MolProbity score [^]	2.58		42 nd percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	1	0.46%	Goal: 0
	Bad bonds:	6 / 1925	0.31%	Goal: 0%
	Bad angles:	28 / 2615	1.07%	Goal: <0.1%
	Peptide Omegas	Cis Prolines:	2 / 10	20.00%
Low-resolution Criteria	CaBLAM outliers	6	2.4%	Goal: <1.0%
	CA Geometry outliers	3	1.21%	Goal: <0.5%
Additional validations	Tetrahedral geometry outliers	1		
	Waters with clashes	0/0	0.00%	See UnDowser table for details

Fonte: MolProbity (2024).

ANEXO D - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE scFv LONGO APÓS A MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	2.49		99 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	1.55%	Goal: <0.3%
	Favored rotamers	173	89.64%	Goal: >98%
	Ramachandran outliers	2	0.80%	Goal: <0.05%
	Ramachandran favored	229	91.97%	Goal: >98%
	Rama distribution Z-score	-1.56 ± 0.49		Goal: abs(Z score) < 2
	MolProbity score [^]	1.66		90 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	3	1.37%	Goal: 0
	Bad bonds:	0 / 1925	0.00%	Goal: 0%
	Bad angles:	13 / 2615	0.50%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	2 / 10	20.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	8	3.2%	Goal: <1.0%
	CA Geometry outliers	3	1.21%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/281		
	Waters with clashes	0/0	0.00%	See UnDowser table for details

Fonte: MolProbity (2024).

ANEXO E - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE CD22 OBTIDO PELO MODELLER 10.1 ANTES DA MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	55.73		3 rd percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	1.22%	Goal: <0.3%
	Favored rotamers	150	91.46%	Goal: >98%
	Ramachandran outliers	5	2.76%	Goal: <0.05%
	Ramachandran favored	169	93.37%	Goal: >98%
	Rama distribution Z-score	-1.61 ± 0.56		Goal: abs(Z score) < 2
	MolProbity score [^]	2.72		36 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	1	0.58%	Goal: 0
	Bad bonds:	0 / 1493	0.00%	Goal: 0%
	Bad angles:	22 / 2027	1.09%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	3 / 11	27.27%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	12	6.7%	Goal: <1.0%
	CA Geometry outliers	2	1.12%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/212		
	Waters with clashes	0/0	0.00%	See UnDowser table for details

Fonte: *MolProbity* (2024).

ANEXO F - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE CD22 OBTIDO APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E ANTES DA MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	60.51		2 nd percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	1.30%	Goal: <0.3%
	Favored rotamers	141	91.56%	Goal: >98%
	Ramachandran outliers	4	2.41%	Goal: <0.05%
	Ramachandran favored	155	93.37%	Goal: >98%
	Rama distribution Z-score	-1.59 ± 0.55		Goal: abs(Z score) < 2
	MolProbity score [^]	2.77		33 rd percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	1	0.58%	Goal: 0
	Bad bonds:	0 / 1433	0.00%	Goal: 0%
	Bad angles:	16 / 1936	0.83%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	3 / 11	27.27%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	12	6.7%	Goal: <1.0%
	CA Geometry outliers	2	1.12%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/207		
	Waters with clashes	0/0	0.00%	See UnDowser table for details

Fonte: MolProbity (2024).

ANEXO G - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE CD22 OBTIDO APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E SEGUIDA DE MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	1.79	99 th percentile * (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	4	2.60% Goal: <0.3%
	Favored rotamers	139	90.26% Goal: >98%
	Ramachandran outliers	5	3.01% Goal: <0.05%
	Ramachandran favored	156	93.98% Goal: >98%
	Rama distribution Z-score	-1.14 ± 0.60	Goal: abs(Z score) < 2
	MolProbity score [^]	1.66	91 st percentile * (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	5	2.92% Goal: 0
	Bad bonds:	0 / 1433	0.00% Goal: 0%
Bad angles:	14 / 1936	0.72% Goal: <0.1%	
Peptide Omegas	Cis Prolines:	3 / 11	27.27% Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	8	4.5% Goal: <1.0%
	CA Geometry outliers	3	1.68% Goal: <0.5%
Additional validations	Chiral volume outliers	0/207	
	Waters with clashes	0/0	0.00% See UnDowser table for details

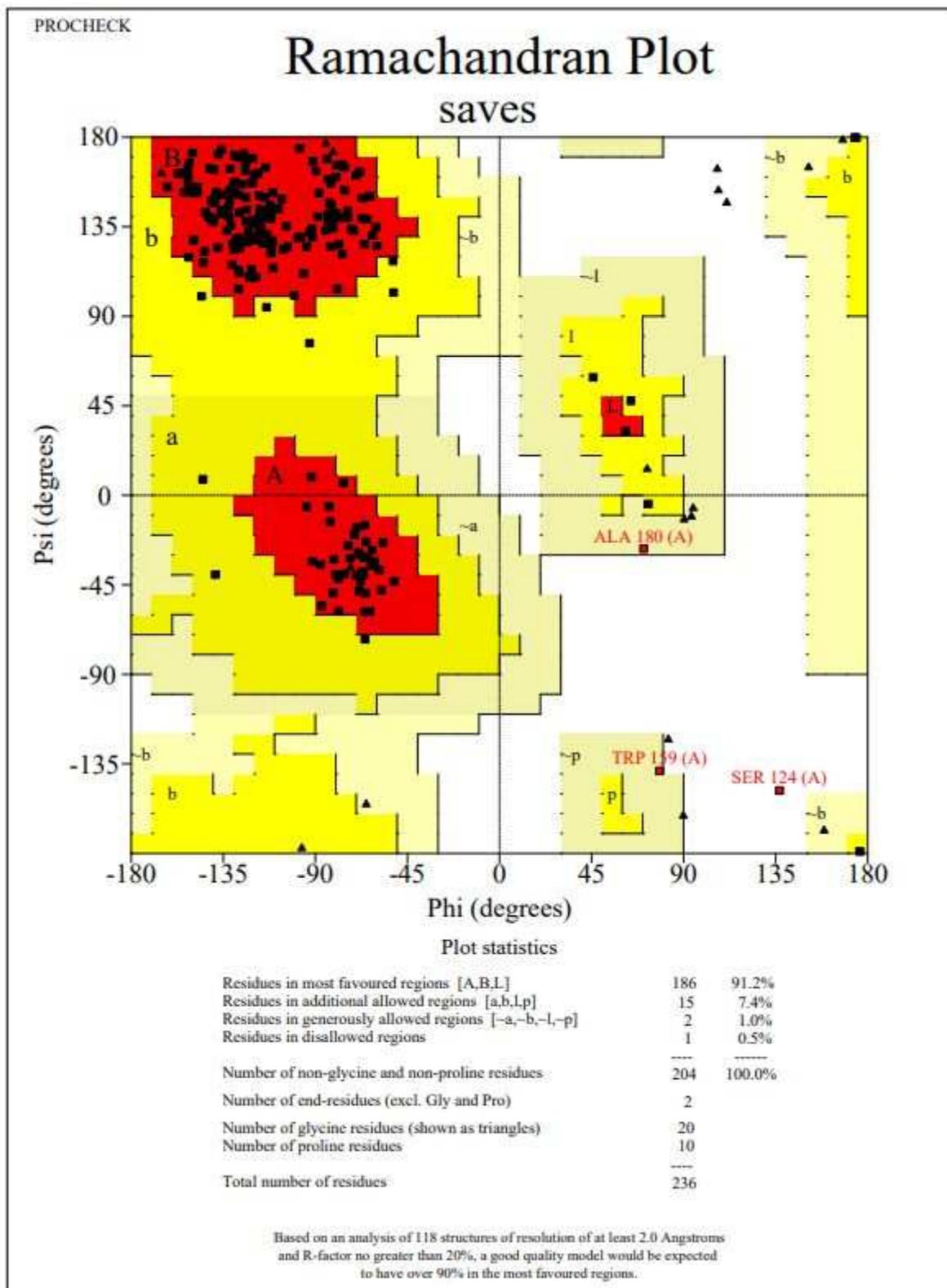
Fonte: MolProbity (2024).

ANEXO H - ANÁLISE DA QUALIDADE GEOMÉTRICA PELO PROGRAMA MOLPROBITY PARA O MODELO DE TRANSMEMBRANA DE CD22 OBTIDO APÓS A MODELAGEM NO I-TASSER ANTES DA MINIMIZAÇÃO DE ENERGIA

All-Atom Contacts	Clashscore, all atoms:	3.7		96 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	7.69%	Goal: <0.3%
	Favored rotamers	11	84.62%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	17	100.00%	Goal: >98%
	Rama distribution Z-score	-1.09 ± 1.74		Goal: abs(Z score) < 2
	MolProbity score [^]	1.83		84 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 122	0.00%	Goal: 0%
	Bad angles:	0 / 165	0.00%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 0	0%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	0	0.0%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/26		
	Waters with clashes	0/0	0.00%	See UnDownser table for details

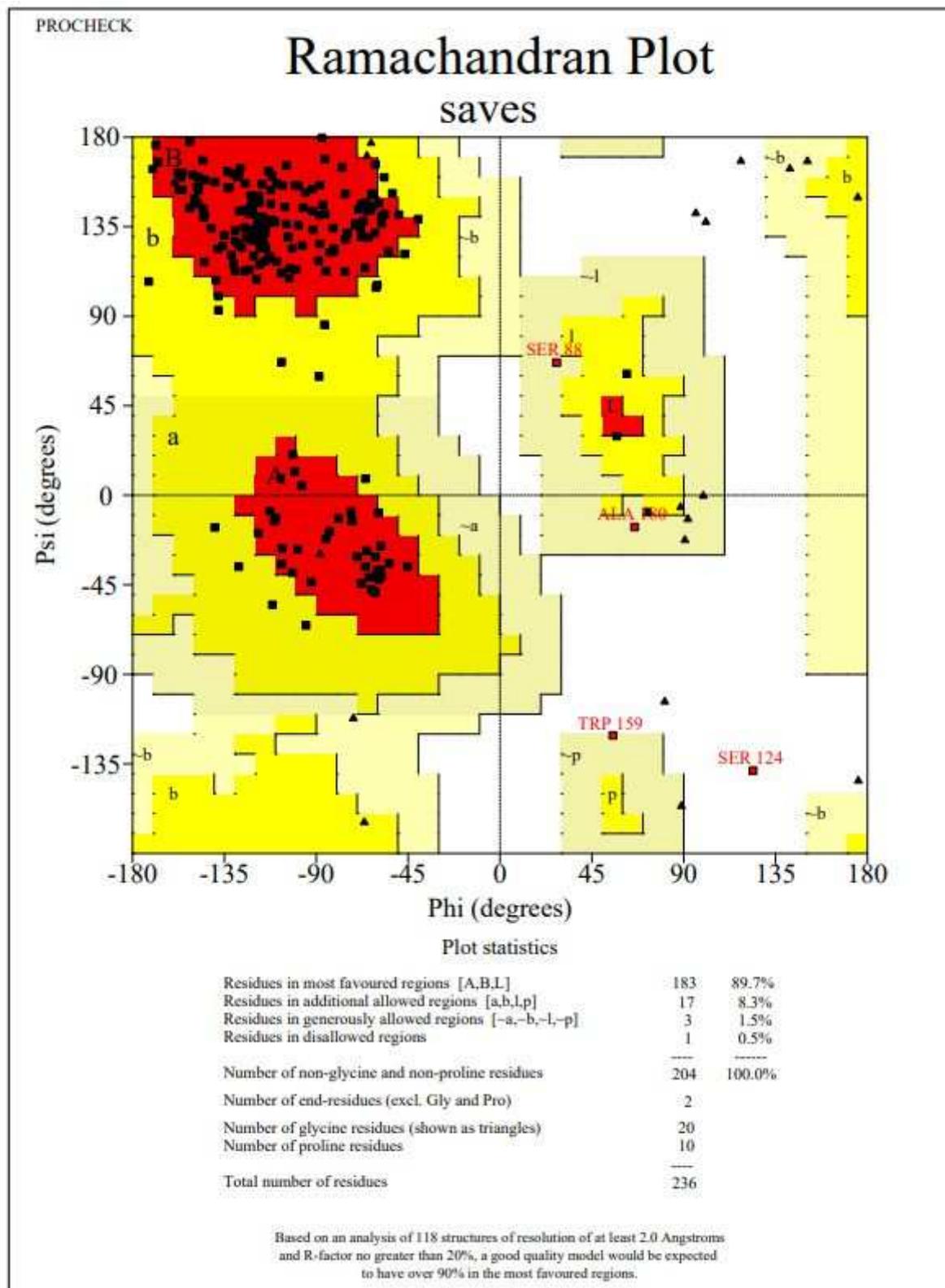
Fonte: MolProbity (2024).

ANEXO I - GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv CURTO ANTES DA MINIMIZAÇÃO DE ENERGIA



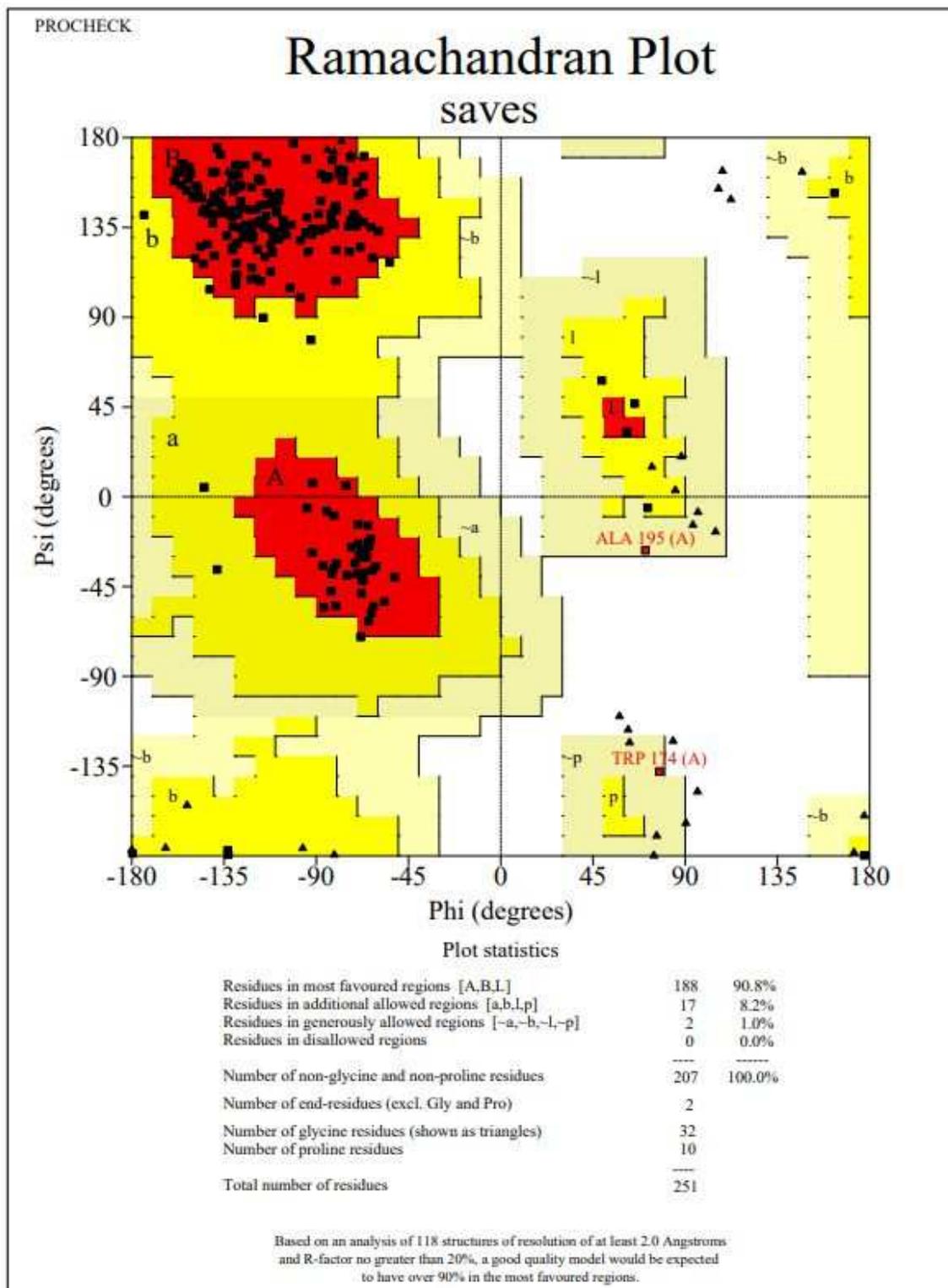
Fonte: PROCHECK (2024).

ANEXO J - GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv CURTO APÓS A MINIMIZAÇÃO DE ENERGIA



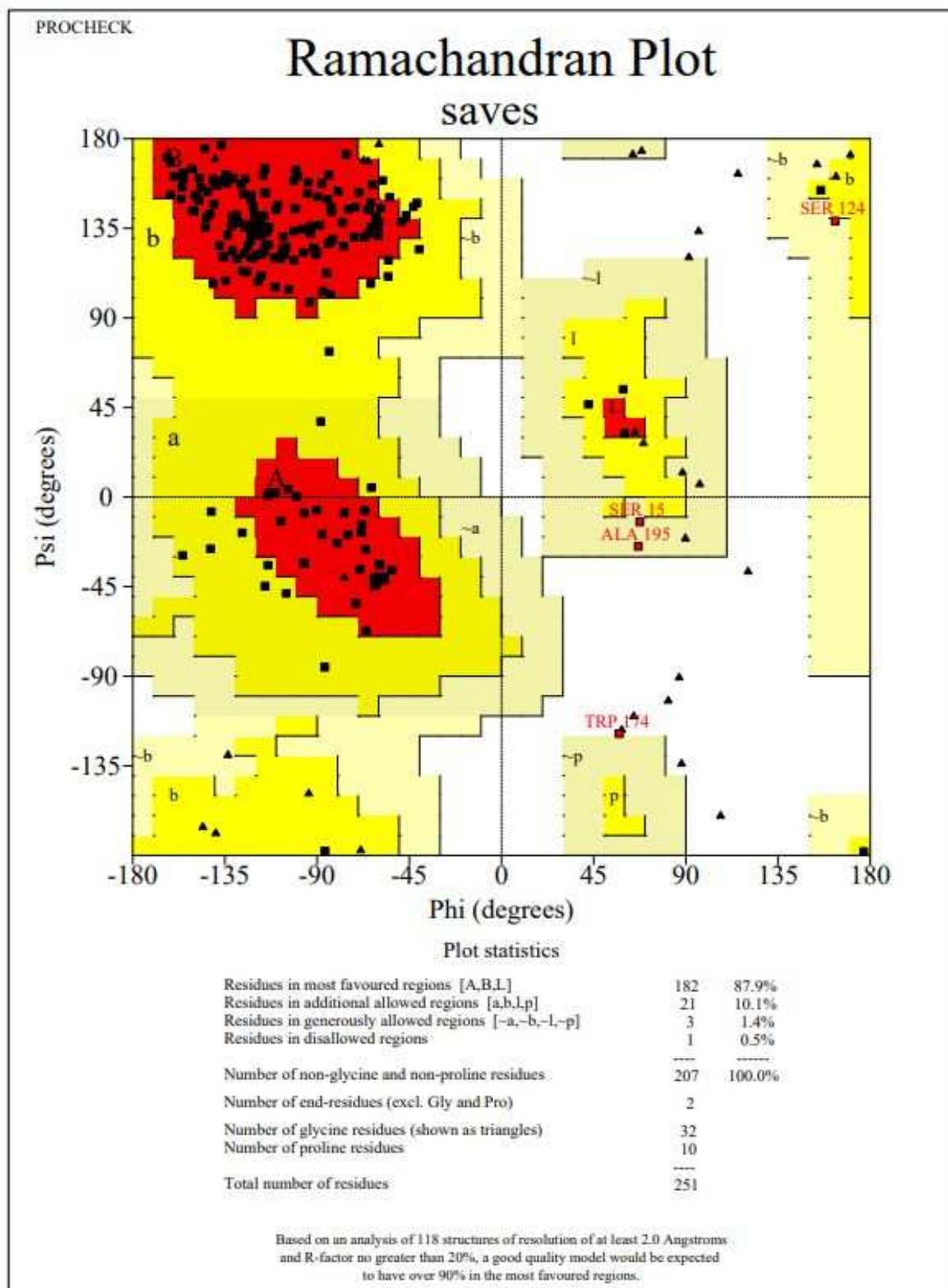
Fonte: PROCHECK (2024).

ANEXO K - GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv LONGO ANTES DA MINIMIZAÇÃO DE ENERGIA



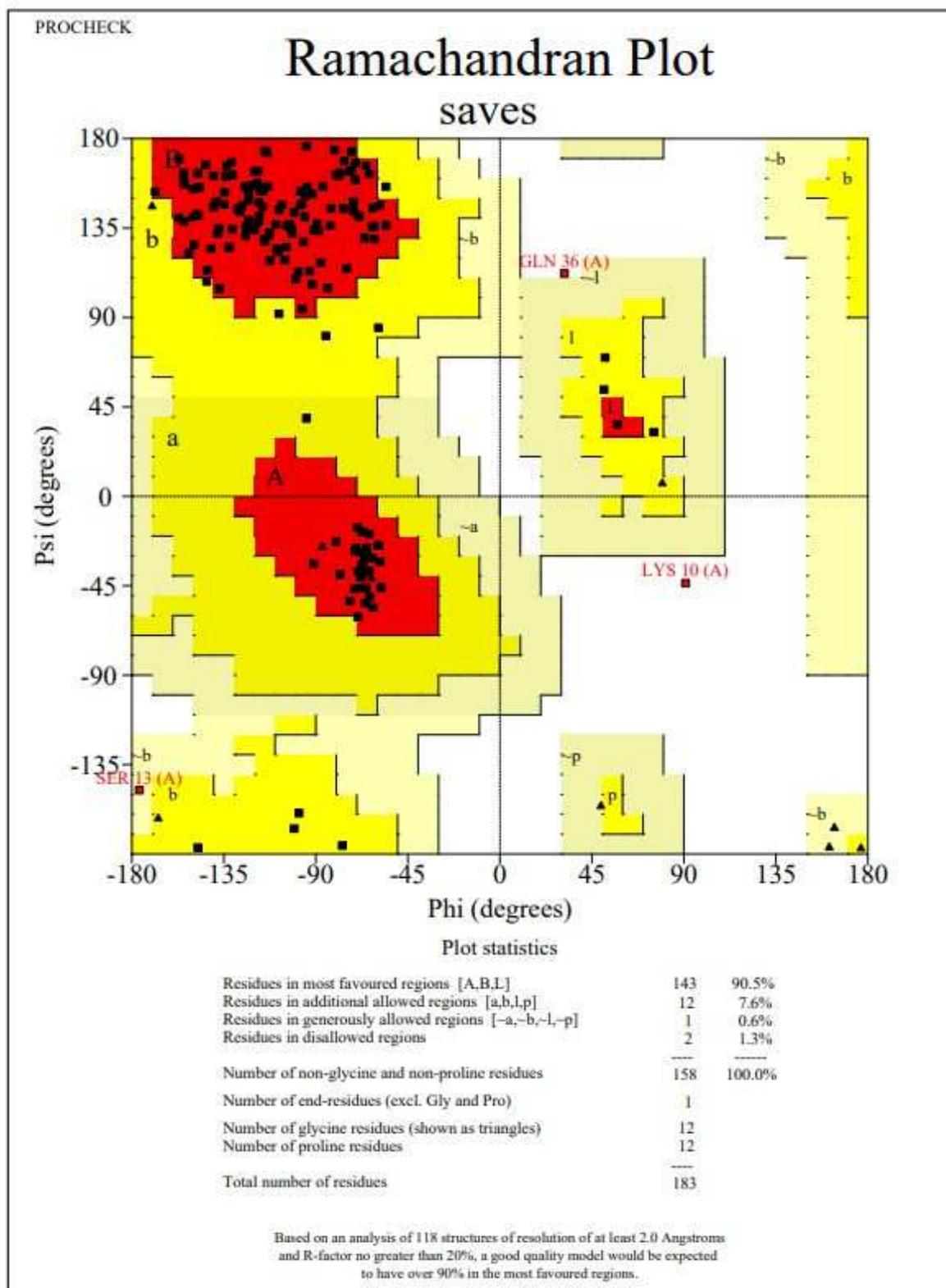
Fonte: PROCHECK (2024).

ANEXO L - GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE scFv LONGO APÓS A MINIMIZAÇÃO DE ENERGIA



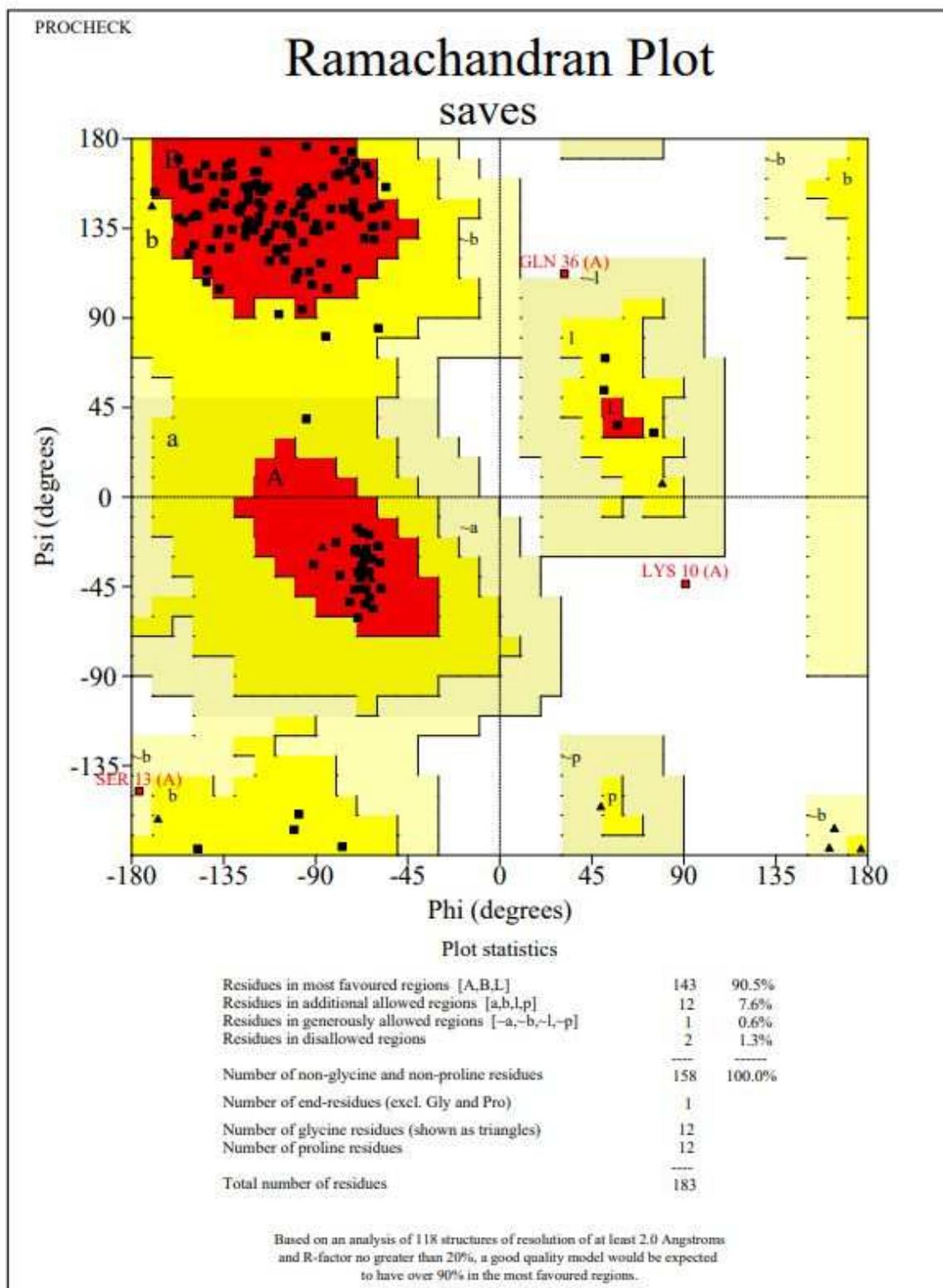
Fonte: PROCHECK (2024).

ANEXO M - GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE CD22 OBTIDO PELO MODELLER 10.1 ANTES DA MINIMIZAÇÃO DE ENERGIA



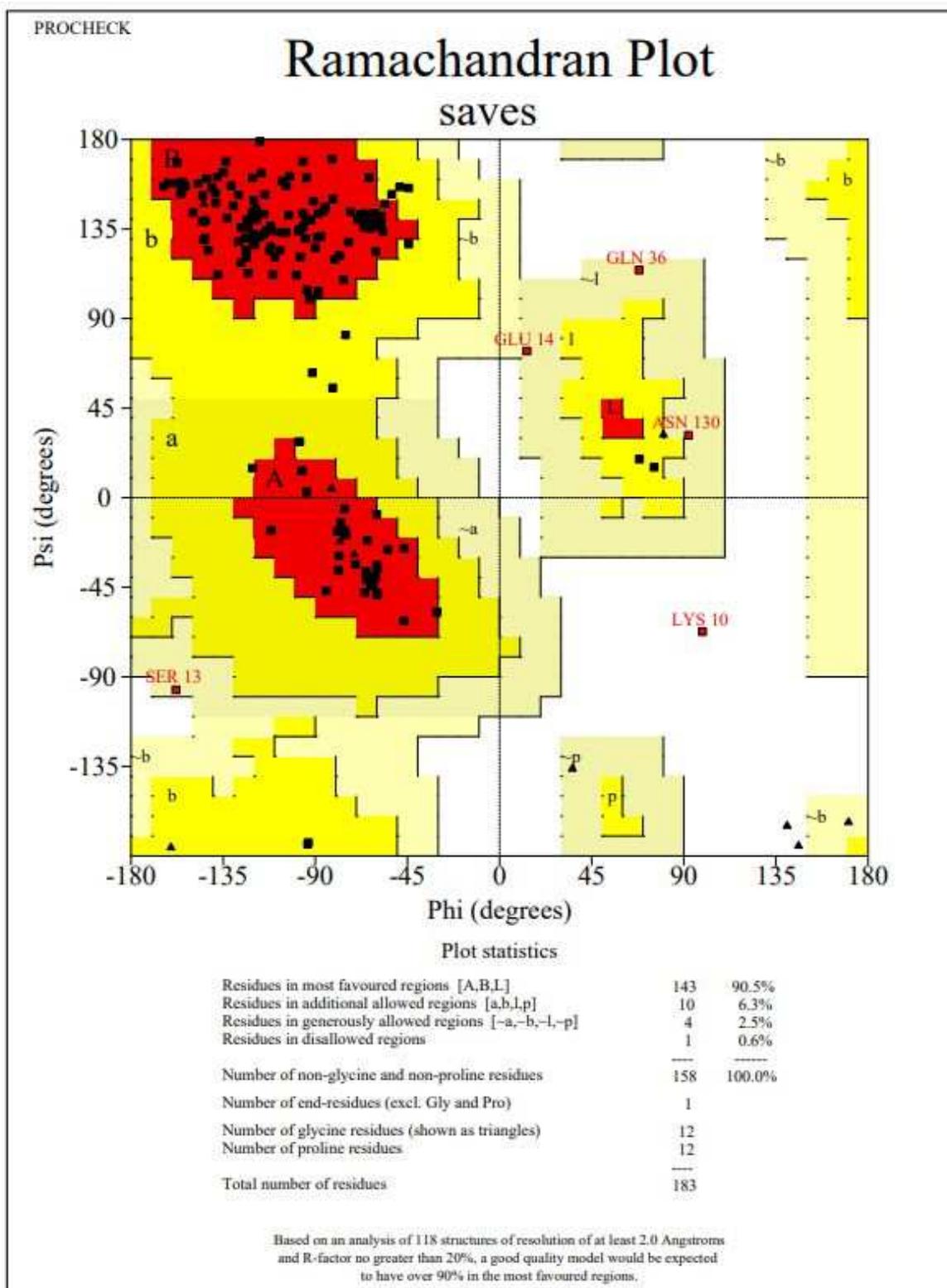
Fonte: PROCHECK (2024).

Anexo N - GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE CD22 OBTIDO APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E ANTES DA MINIMIZAÇÃO DE ENERGIA



Fonte: PROCHECK (2024).

Anexo O - GRÁFICO DE RAMACHANDRAN GERADO PELO PROGRAMA PROCHECK PARA O MODELO DE CD22 OBTIDO APÓS A INSERÇÃO DAS GLICOSILAÇÕES NO CHARMM-GUI E SEGUIDA DE MINIMIZAÇÃO DE ENERGIA



Fonte: PROCHECK (2024).